Translating and the Computer 28 Conference 2006

Integrated bilingual specialist dictionaries :

The LexTerm initiative

Marie-Jeanne Derouin, Langenscheidt Fachverlag, Munich, Germany

André Le Meur, Université de Rennes 2, France

Content

This paper about integrated specialist bilingual dictionaries in Translation Memories will have two parts:

- **The challenge for the specialist dictionary publishers:**
  - ➢ Meeting professional dictionary users' need for global electronic solutions.
  - ➢ Integration of their lexicographical contents in language tools such as Translation Memories and automatic Translation machines

- **The methodology for reusing lexicographical data:**

  LexTerm, a bridge between lemma-oriented lexicography and concept-oriented terminology

Actual Translators' needs

- Solutions for optimising their work-flow and time-sparing tools
- A quick access to the language ressources they currently use
- A unique tool with « à la carte » integrated specialist dictionaries or other bilingual resources

Future developments of Specialist Dictionary programs

- **The current situation**: large collections of bilingual specialist dictionaries in technical, scientific and economic fields in print and electronic versions.
- **Two main issues**
  - Keeping accommodating the needs of a large dictionary users community for traditional printed or electronic dictionaries
  - And meeting the needs of the professional users (Translators and technical writers) for tools beyond Machine Readable Dictionaries on CD-Rom or Online. The specialist dictionary being <u>one component</u> of multifunctional tools for CAT

A joint challenge for specialist dictionary Publishers and TMS providers

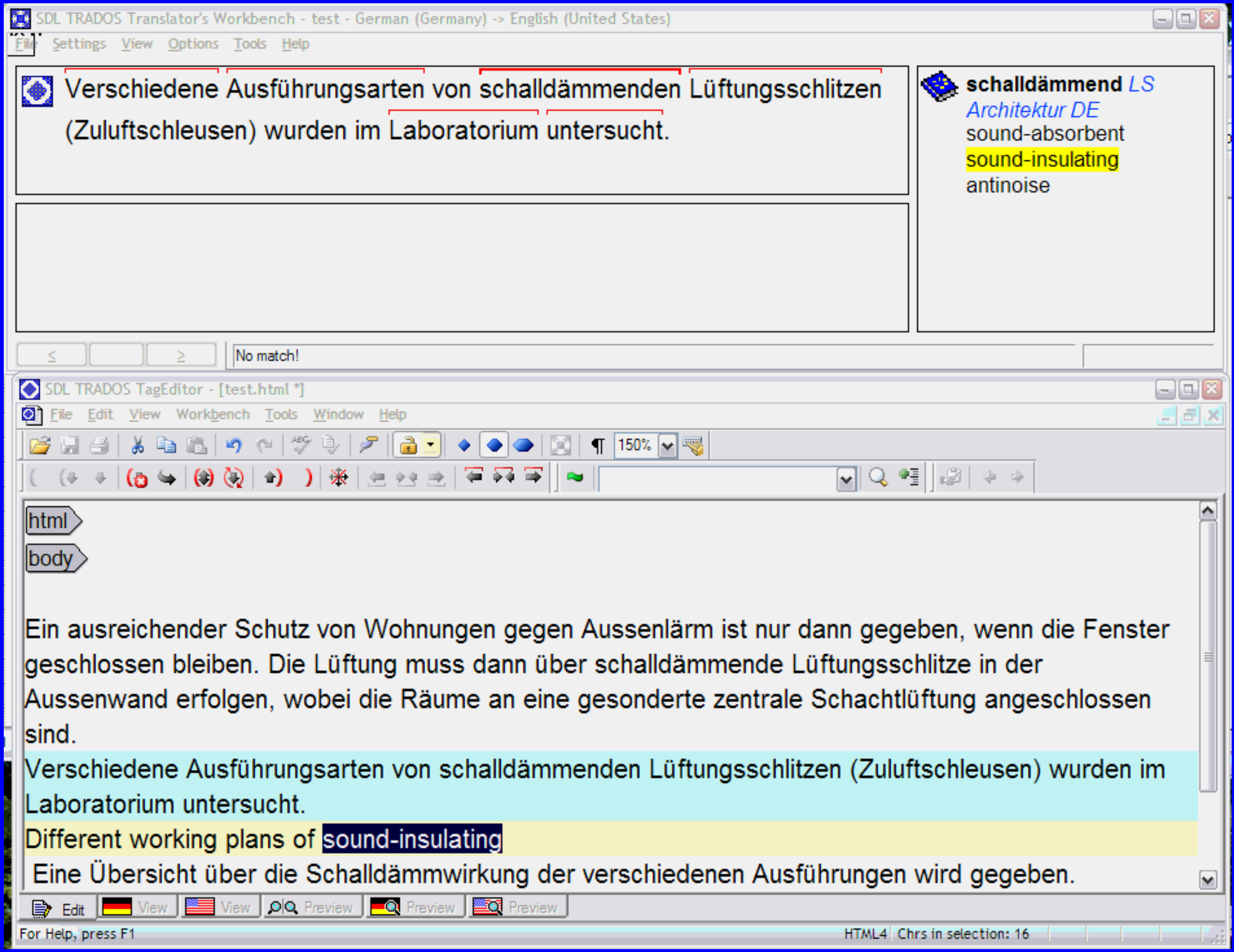
- **For Publishers**: An extended data-workflow in order to reuse the existing lexicographical data for the new purposes (A big issue for over a million data)
  - **The challenge**: Two versions of a single source for every specialist bilingual dictionary: a lemma-oriented one for print and electronic dictionaries and a concept-oriented one for integration in other language tools
  - **The Aim**: A better adequacy to the translation market demand

- **For TMS providers:** A new product concept for a unique tool with „à la carte“ specialist dictionaries
  - **The challenge**: the integration of the dictionaries in the translator‘s CAT-tool
  - **The Aim**: An added value for TMS-tools

The solution and the product

- An application of terminological research with data-modelling experts from the Université of Rennes 2
- A longstanding cooperation between Publisher and University:
  - 1998: European project for merging dictionaries and the first DTD for our specialist dictionaries
  - 2000-2006: Major contribution to the new version of ISO 1951. Awareness of the importance of lexicographical standards for reuse and integration of dictionary data
- The product:
  - 10 specialist bilingual dictionaries in over 100 subject fields in the main European languages and in combination with German will be integrated in Translation Memory tools
  - And available on the market in 2007

Application : Integration in Trados Translator‘s workbench


SDL TRADOS Translator's Workbench - test - German (Germany) -> English (United States)
File Settings View Options Tools Help
SDL TRADOS TagEditor - [test.html *]
File Edit View Workbench Tools Window Help
150%
html
body
Edit
View
View
Preview
Preview
Preview

Verschiedene Ausführungsarten von schalldämmenden Lüftungsschlitzen (Zuluftschleusen) wurden im Laboratorium untersucht.

schalldämmend *LS*
Architektur DE
sound-absorbent
sound-insulating
antinoise

No match!

Ein ausreichender Schutz von Wohnungen gegen Aussenlärm ist nur dann gegeben, wenn die Fenster geschlossen bleiben. Die Lüftung muss dann über schalldämmende Lüftungsschlitze in der Aussenwand erfolgen, wobei die Räume an eine gesonderte zentrale Schachtlüftung angeschlossen sind.

Verschiedene Ausführungsarten von schalldämmenden Lüftungsschlitzen (Zuluftschleusen) wurden im Laboratorium untersucht.

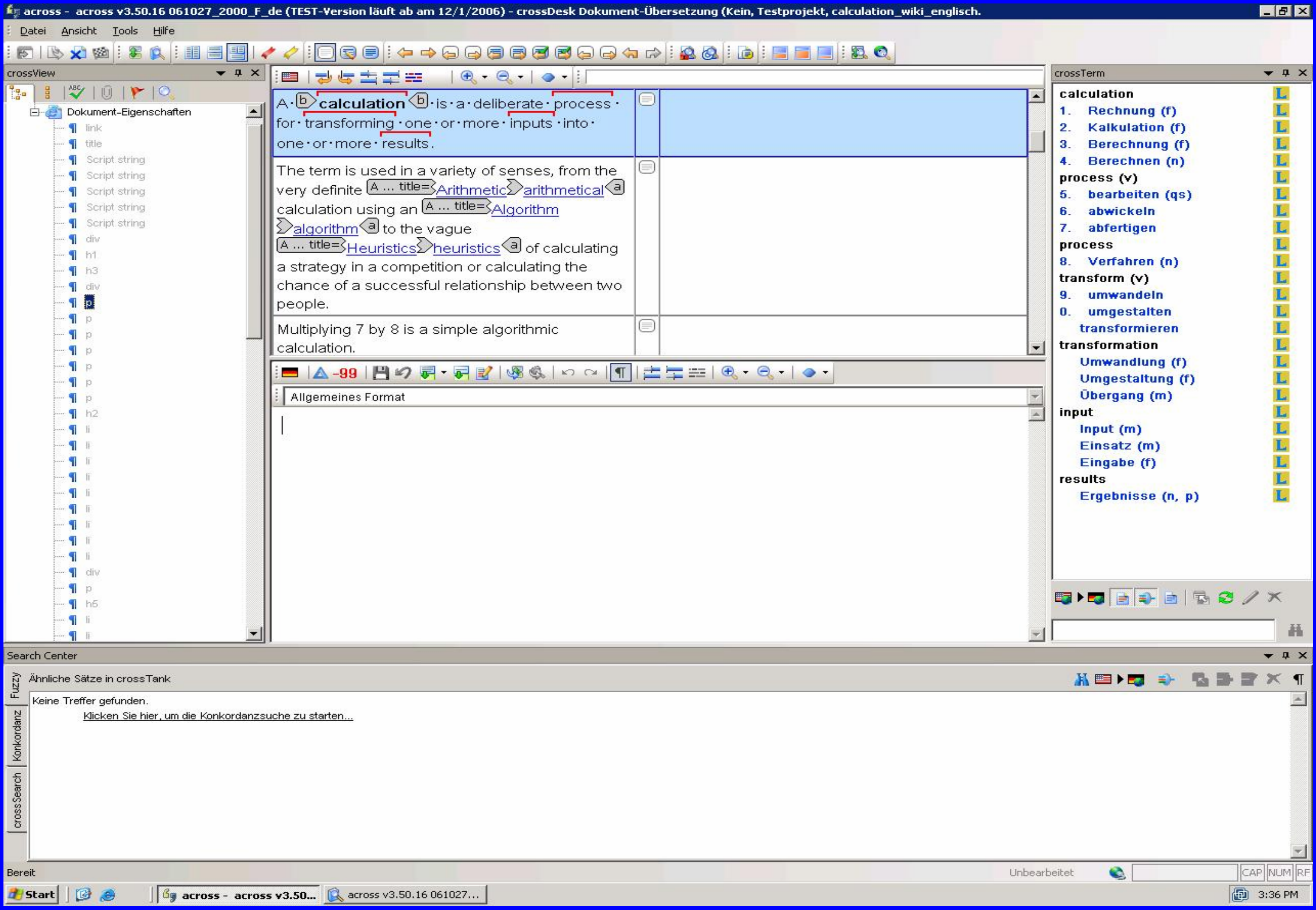
Different working plans of sound-insulating

Eine Übersicht über die Schalldämmwirkung der verschiedenen Ausführungen wird gegeben.

For Help, press F1

HTML4 Chrs in selection: 16

across- Integration in Translator‘s crossDesk
En-Ge



A·calculation·is·a·deliberate·process·for·transforming·one·or·more·inputs·into·one·or·more·results.

The term is used in a variety of senses, from the very definite A ... title=Arithmetic arithmetical a calculation using an A ... title=Algorithm algorithm a to the vague A ... title=Heuristics heuristics a of calculating a strategy in a competition or calculating the chance of a successful relationship between two people.

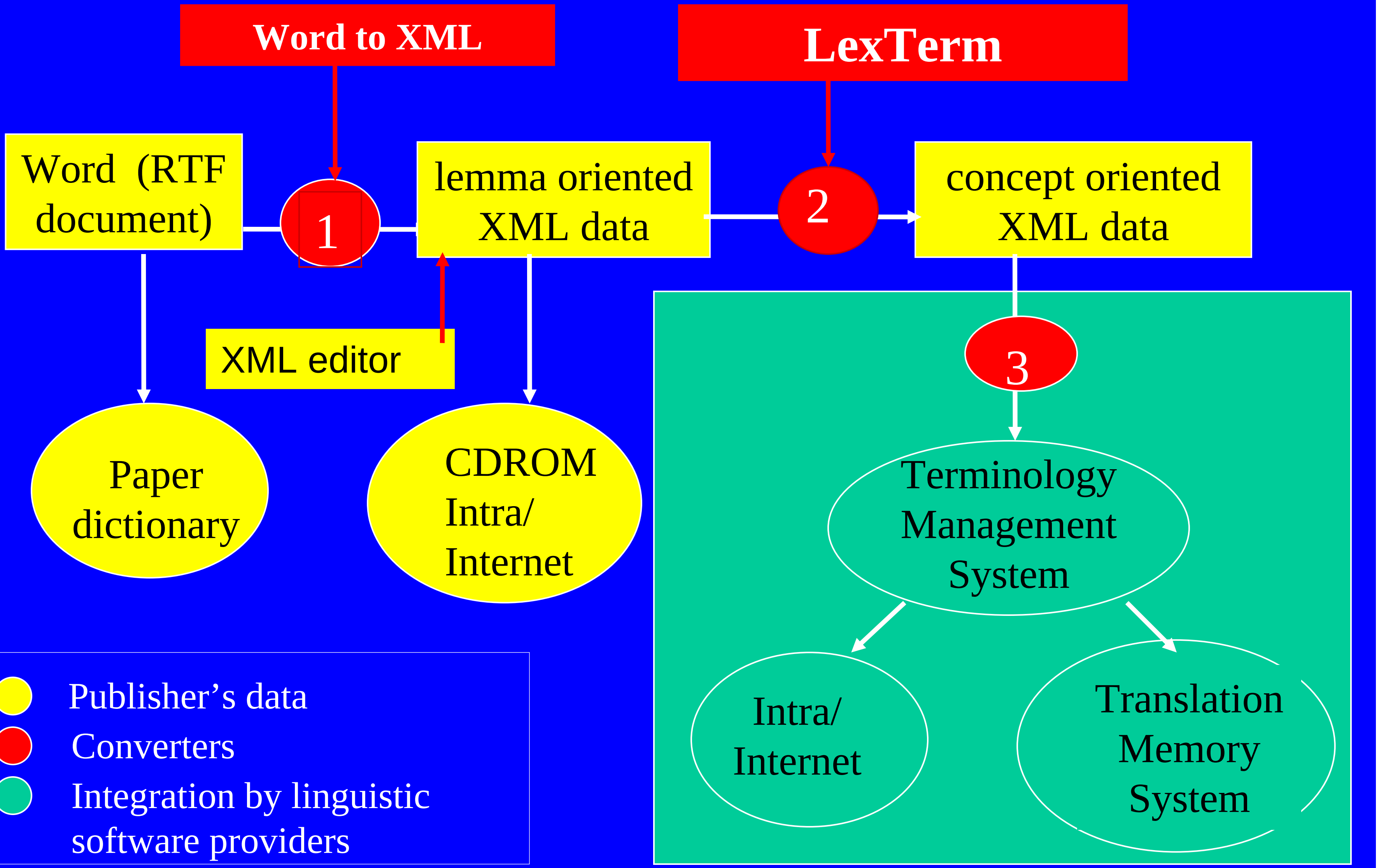
Multiplying 7 by 8 is a simple algorithmic calculation.

calculation
1. Rechnung (f)
2. Kalkulation (f)
3. Berechnung (f)
4. Berechnen (n)
process (v)
5. bearbeiten (qs)
6. abwickeln
7. abfertigen
process
8. Verfahren (n)
transform (v)
9. umwandeln
0. umgestalten
transformieren
transformation
Umwandlung (f)
Umgestaltung (f)
Übergang (m)
input
Input (m)
Einsatz (m)
Eingabe (f)
results
Ergebnisse (n, p)

Ähnliche Sätze in crossTank

Keine Treffer gefunden.
Klicken Sie hier, um die Konkordanzsuche zu starten...

The editorial workflow



🟡 Publisher’s data

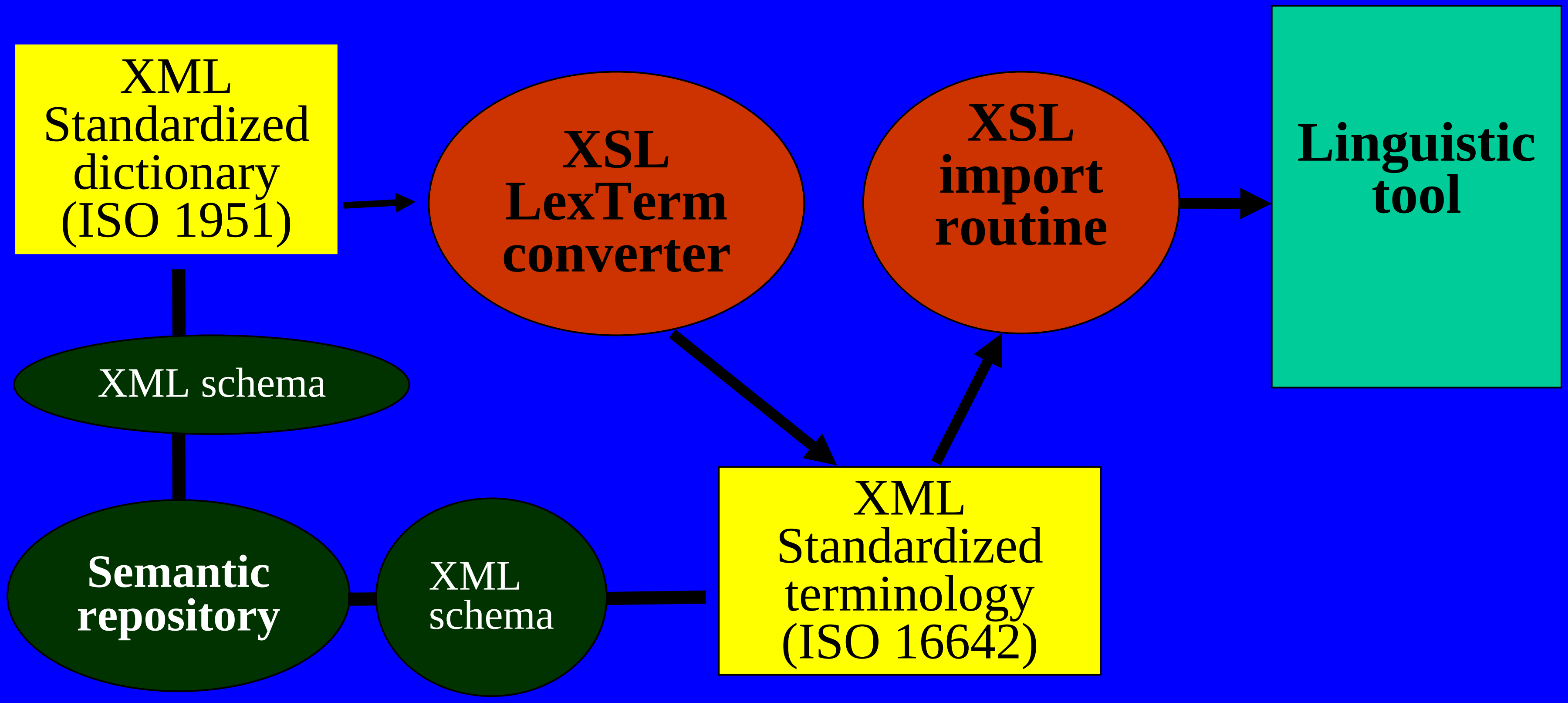
🔴 Converters

🟢 Integration by linguistic software providers

LexTerm: a methodology for reusing lexicographical data

- **Principle: a bridge between**
  - Dictionaries: based on ISO 1951 (XmLex)
  - Terminology: based on ISO 16642. Terminology Markup Language : Geneter (Annex C)
- **Methodology: semantic and syntactic interoperability**
  - Mapping data elements and structures
  - Resolving structural issues
  - Example)
- **Implementation: an experimental workbench (XML + XSL)**
  - LexTermLib an XSL library for transforming dictionaries into monosemic entries
  - A public and online demonstrator

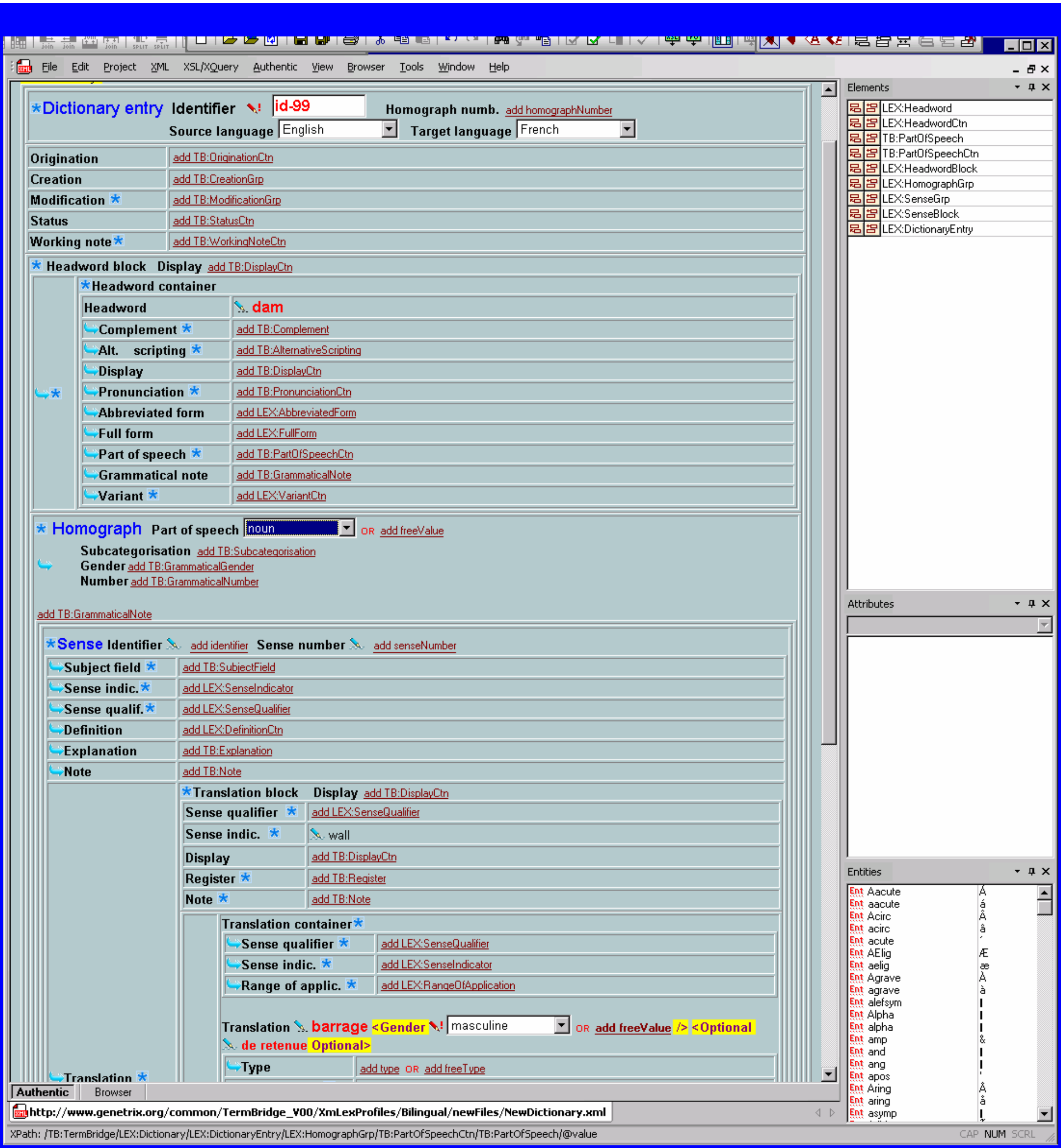
LexTerm: a XSL bridge between standardized XML formats



Second example: a bilingual general dictionary

- **dam**^{1} [dæm] ***1*** *n* ***a*** (wall) *[river]* barrage *m* (de retenue), digue *f* ;*[lake]* barrage (de retenue). ***b*** *(water)* réservoir *m*, lac *m* de retenue.***2*** *vt* ***a*** (also **~ up**) *river* endiguer ; *lake* construire un barrage sur. **to ~the waters of yhe Nile** faire or construire un barrage pour contenir les eaux du Nil. b *flow of words, oaths* endiguer. ***3 comp*** **dambuster** (*bomb*) bombe *f* à ricochets ; *(person)* (aviateur *m*) briseur *m* de barrages *(se refère à un épisode de la seconde guerre mondiale)*.
- **dam**^{2} [dæm] *n* (animal) mère *f*.

Editing an XmLex entry with XmlSpy



*Dictionary entry Identifier id-99 Homograph numb. add homographNumber

Source language English Target language French

Origination	add TB:OriginationCtn
Creation	add TB:CreationGrp
Modification *	add TB:ModificationGrp
Status	add TB:StatusCtn
Working note*	add TB:WorkingNoteCtn

* Headword block Display add TB:DisplayCtn

*Headword container

Headword	dam
Complement *	add TB:Complement
Alt. scripting *	add TB:AlternativeScripting
Display	add TB:DisplayCtn
Pronunciation *	add TB:PronunciationCtn
Abbreviated form	add LEX:AbbreviatedForm
Full form	add LEX:FullForm
Part of speech *	add TB:PartOfSpeechCtn
Grammatical note	add TB:GrammaticalNote
Variant *	add LEX:VariantCtn

* Homograph Part of speech noun OR add freeValue

Subcategorisation add TB:Subcategorisation

Gender add TB:GrammaticalGender

Number add TB:GrammaticalNumber

add TB:GrammaticalNote

*Sense Identifier add identifier Sense number add senseNumber

Subject field *	add TB:SubjectField
Sense indic.*	add LEX:SenseIndicator
Sense qualif.*	add LEX:SenseQualifier
Definition	add LEX:DefinitionCtn
Explanation	add TB:Explanation
Note	add TB:Note

*Translation block Display add TB:DisplayCtn

Sense qualifier *	add LEX:SenseQualifier
Sense indic. *	wall
Display	add TB:DisplayCtn
Register *	add TB:Register
Note *	add TB:Note

Translation container*

Sense qualifier *	add LEX:SenseQualifier
Sense indic. *	add LEX:SenseIndicator
Range of applic. *	add LEX:RangeOfApplication

Translation barrage <Gender masculine OR add freeValue /> <Optional de retenue Optional>

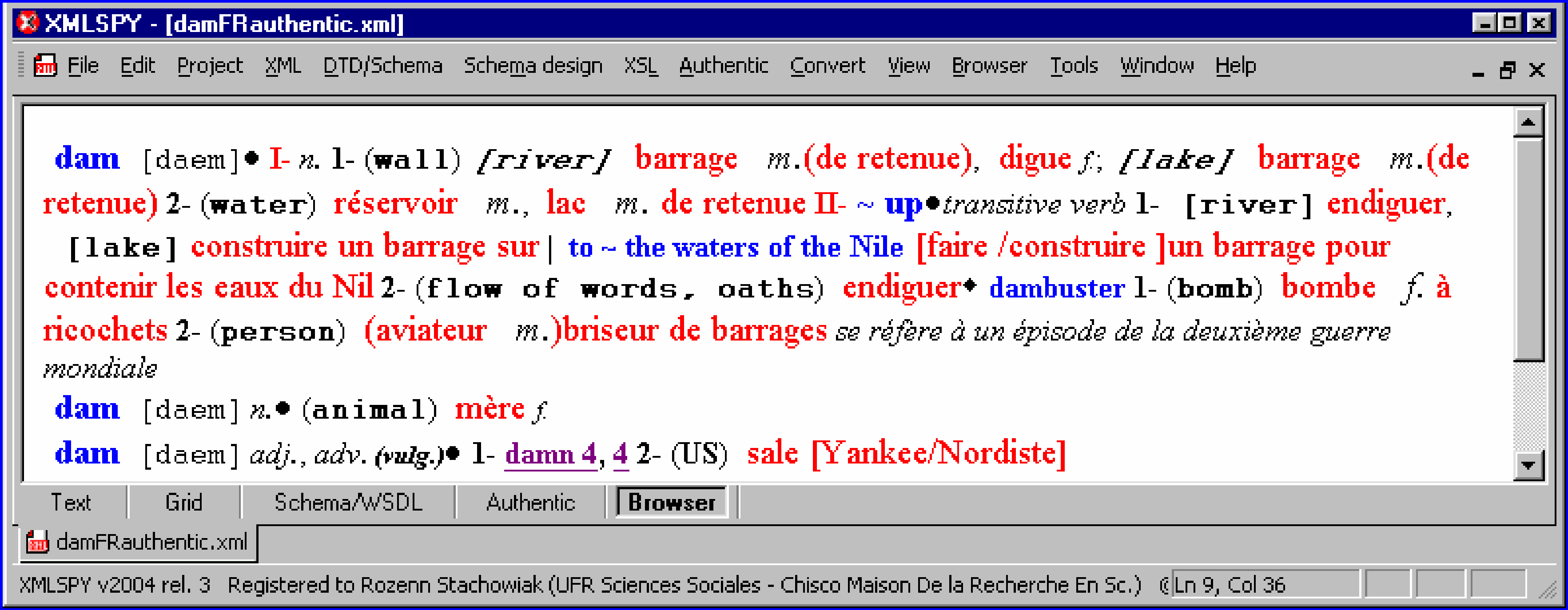
Type	add type OR add freeType

Translation *

http://www.genetrix.org/common/TermBridge_V00/XmLexProfiles/Bilingual/newFiles/NewDictionary.xml

XPath: /TB:TermBridge/LEX:Dictionary/LEX:DictionaryEntry/LEX:HomographGrp/TB:PartOfSpeechCtn/TB:PartOfSpeech/@value

HTML view of an XmLex entry

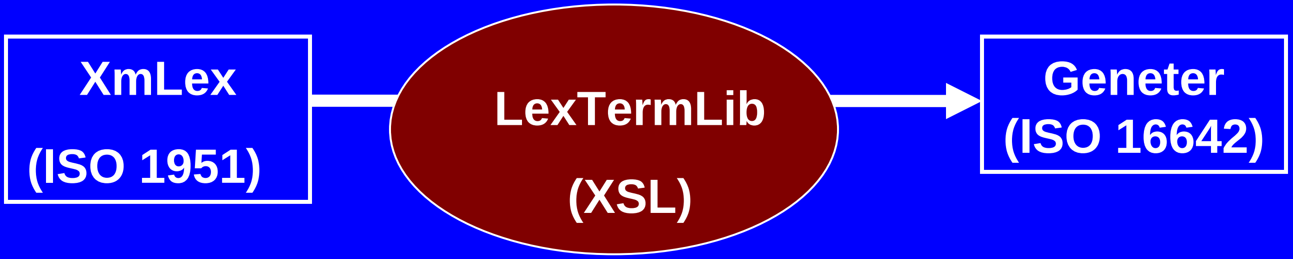

XMLSPY - [damFRauthentic.xml]
File
Edit
Project
XML
DTD/Schema
Schema design
XSL
Authentic
Convert
View
Browser
Tools
Window
Help
Text
Grid
Schema/WSDL
Authentic
Browser
damFRauthentic.xml
XMLSPY v2004 rel. 3 Registered to Rozenn Stachowiak (UFR Sciences Sociales - Chisco Maison De la Recherche En Sc.) ©Ln 9, Col 36

dam [daem]• I- *n.* 1- (**wall**) *[river]* barrage *m.*(de retenue), digue *f.*; *[lake]* barrage *m.*(de retenue) 2- (**water**) réservoir *m.*, lac *m.* de retenue II- ~ **up**•*transitive verb* 1- [**river**] endiguer, [**lake**] construire un barrage sur | to ~ the waters of the Nile [faire /construire]un barrage pour contenir les eaux du Nil 2- (**flow of words, oaths**) endiguer• **dambuster** 1- (**bomb**) bombe *f.* à ricochets 2- (**person**) (aviateur *m.*)briseur de barrages *se réfère à un épisode de la deuxième guerre mondiale*

dam [daem] *n.*• (**animal**) mère *f.*

dam [daem] *adj., adv.* ***(vulg.)***• 1- damn 4, 4 2- (US) sale [Yankee/Nordiste]

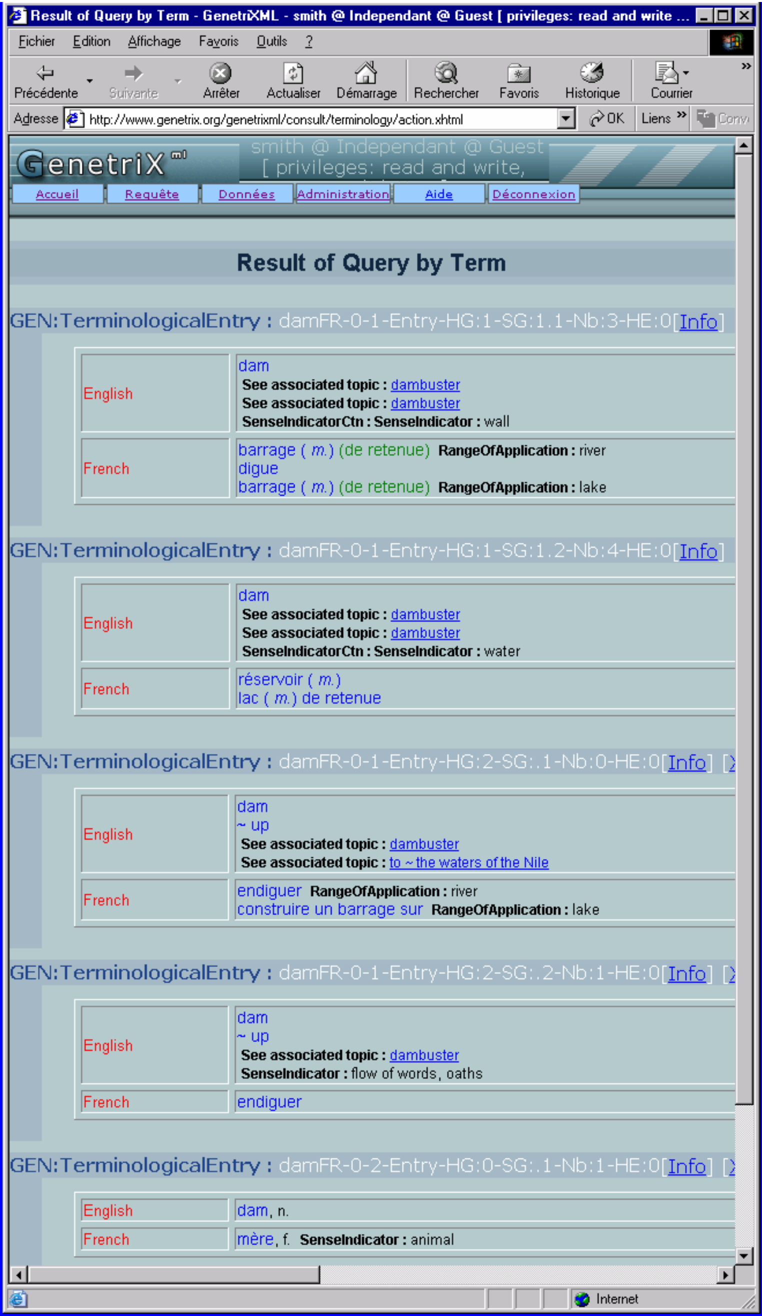
LexTermLib: a XSL library for automatic transformation



15 steps

- Dissociate lemma, multiword units and compositional phrases
- Resolve structural issues
- Deal with synonymy (by grouping referring and main entries)
- Split meanings
- Take morpho-syntactic and pragmatic information apart from semantics

LexTerm Monosemic entries


Result of Query by Term - GenetriXML - smith @ Independant @ Guest [privileges: read and write ...
Fichier
Edition
Affichage
Favoris
Outils
?
Précédente
Suivante
Arrêter
Actualiser
Démarrage
Rechercher
Favoris
Historique
Courrier
Adresse
http://www.genetrix.org/genetrixml/consult/terminology/action.xhtml
OK
Liens
Conv
GenetriX ml
smith @ Independant @ Guest
[privileges: read and write,
Accueil
Requête
Données
Administration
Aide
Déconnexion
Internet

Result of Query by Term

GEN:TerminologicalEntry : damFR-0-1-Entry-HG:1-SG:1.1-Nb:3-HE:0[Info]

English	dam **See associated topic :** dambuster **See associated topic :** dambuster **SenseIndicatorCtn : SenseIndicator :** wall
French	barrage (*m.*) (de retenue) **RangeOfApplication :** river digue barrage (*m.*) (de retenue) **RangeOfApplication :** lake

GEN:TerminologicalEntry : damFR-0-1-Entry-HG:1-SG:1.2-Nb:4-HE:0[Info]

English	dam **See associated topic :** dambuster **See associated topic :** dambuster **SenseIndicatorCtn : SenseIndicator :** water
French	réservoir (*m.*) lac (*m.*) de retenue

GEN:TerminologicalEntry : damFR-0-1-Entry-HG:2-SG:.1-Nb:0-HE:0[Info] [

English	dam ~ up **See associated topic :** dambuster **See associated topic :** to ~ the waters of the Nile
French	endiguer **RangeOfApplication :** river construire un barrage sur **RangeOfApplication :** lake

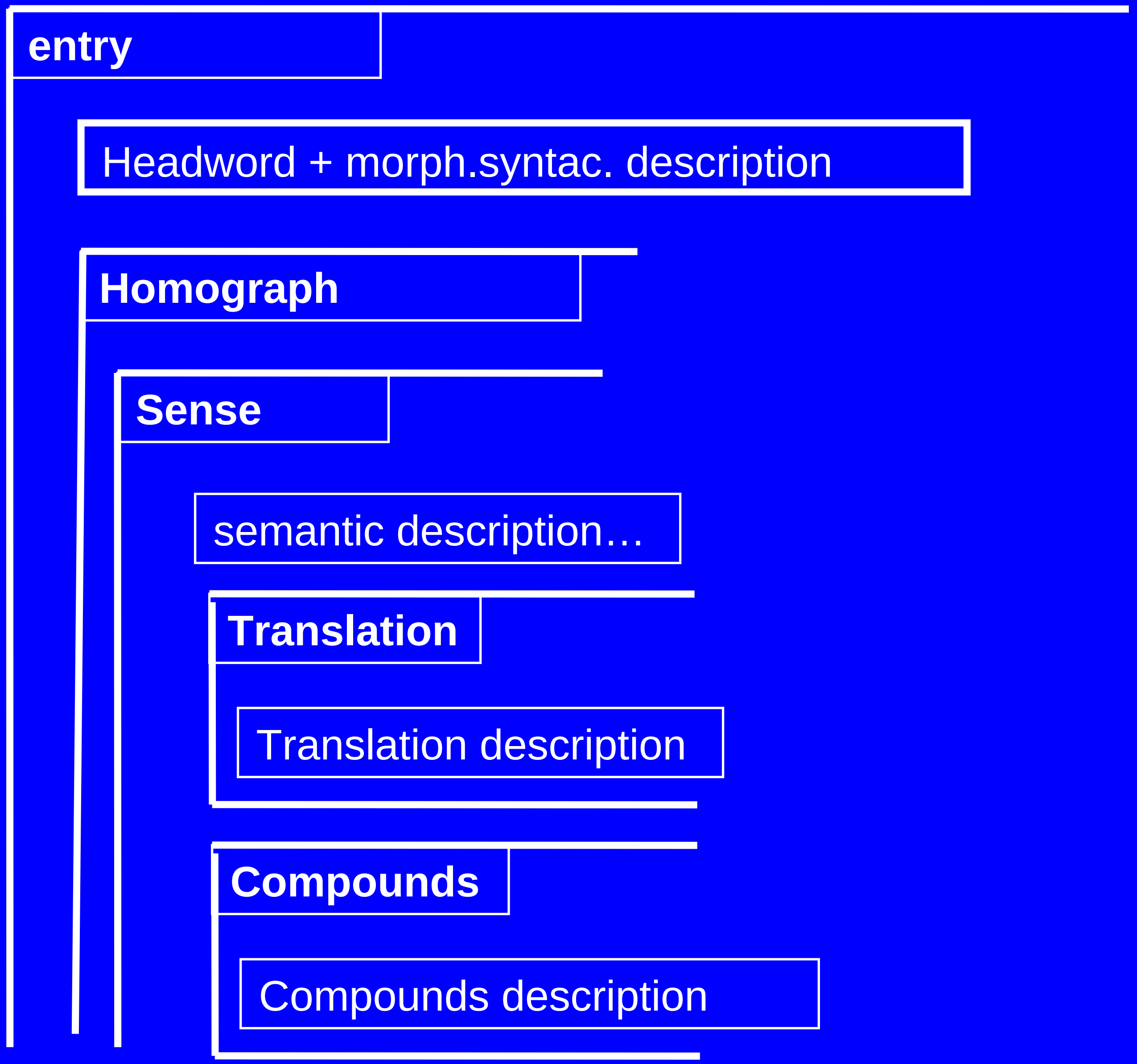
GEN:TerminologicalEntry : damFR-0-1-Entry-HG:2-SG:.2-Nb:1-HE:0[Info] [

English	dam ~ up **See associated topic :** dambuster **SenseIndicator :** flow of words, oaths
French	endiguer

GEN:TerminologicalEntry : damFR-0-2-Entry-HG:0-SG:.1-Nb:1-HE:0[Info] [

English	dam, n.
French	mère, f. **SenseIndicator :** animal

The ISO 1951 lemma-oriented meta model



entry

Headword + morph.syntac. description

Homograph

Sense

semantic description…

Translation

Translation description

Compounds

Compounds description

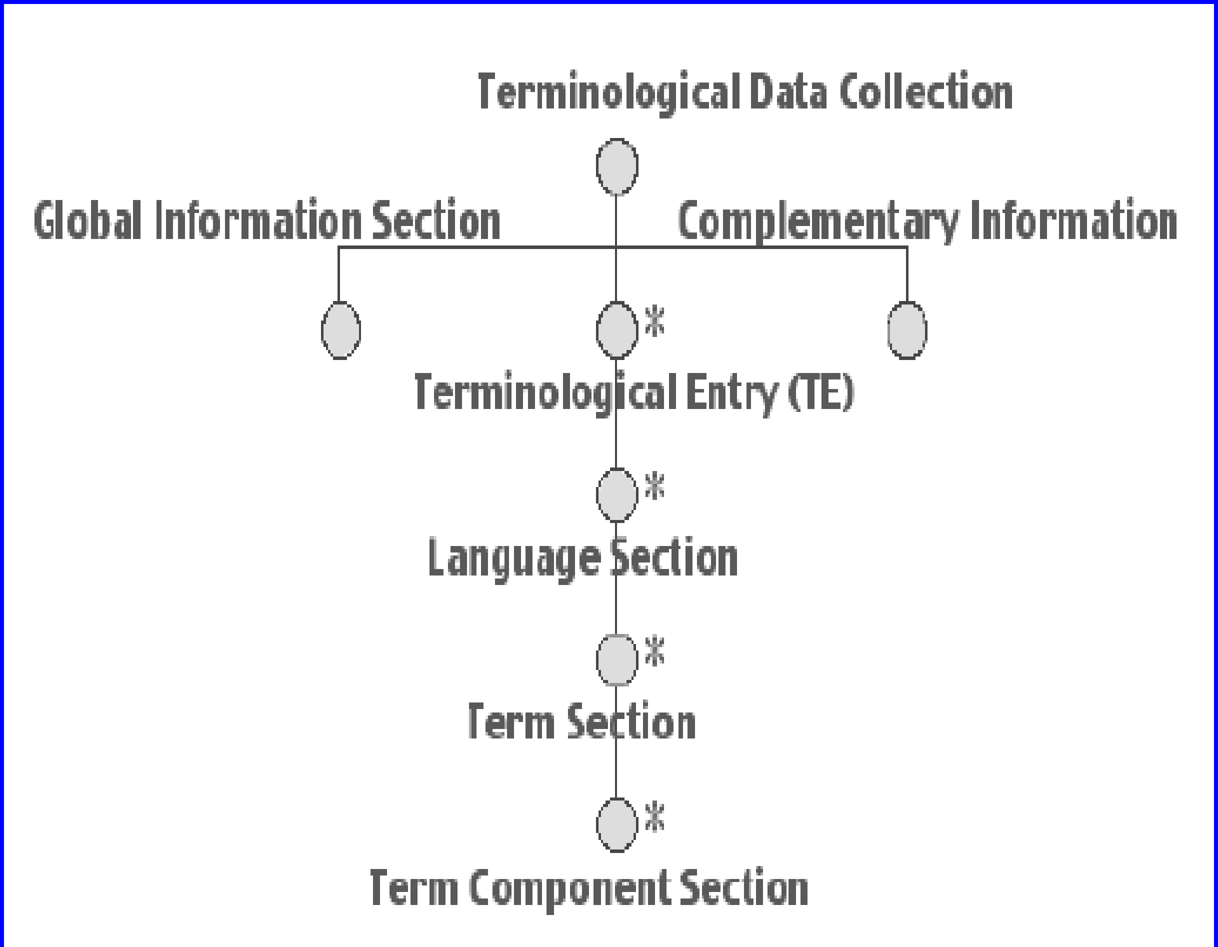
A lexicographical mark up language : XmLex

- **ISO FDIS 1951**
  - To be published in 2006-2007
  - A « generic model »
  - Structures + data elements
  - Rules of subsetting
  - Examples
  - Non normative DTDs : XmLex, XmLexForBilingualDictionaries

ISO 1951 is made for Human Readable and Machine processable dictionaries

It is compatible with LMF (ISO CD 24613) and OLIF (Open Lexicon Interchange Format)

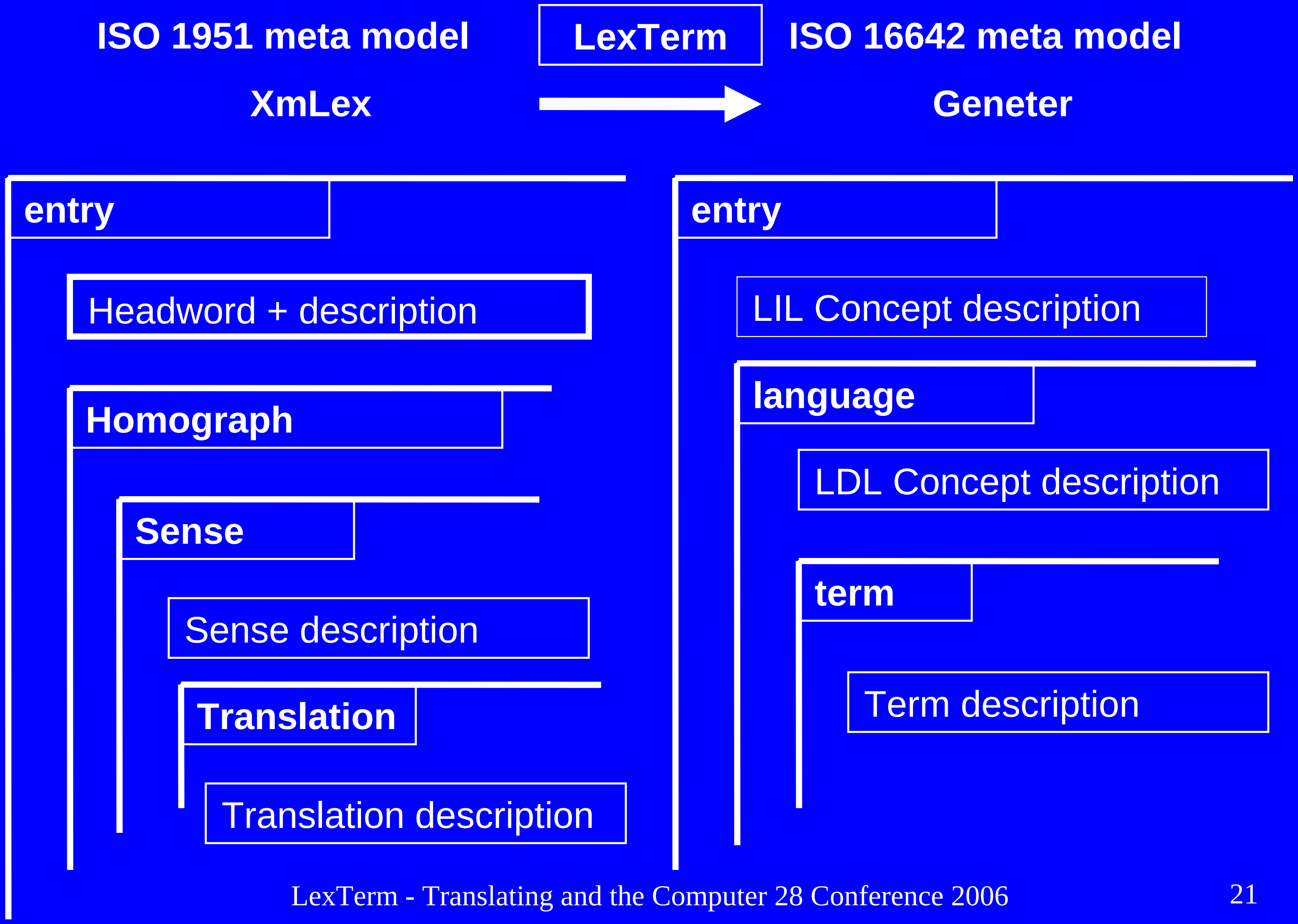
The concept-oriented meta model (ISO 16642 TMF)


Terminological Data Collection
Global Information Section
Complementary Information
*
Terminological Entry (TE)
*
Language Section
*
Term Section
*
Term Component Section

Principles of the concept-oriented approach

- **Principle and methods**: ISO 704
- **Data elements**: ISO 12620
- ISO 16642: Terminology Markup Framework (TMF)
  - Meta model
  - Terminology Markup Languages (TML), GMT, MSC (=TBX), Geneter (Annex C : Geneter)

The two models



ISO 1951 meta model

LexTerm

ISO 16642 meta model

XmLex

Geneter

entry

Headword + description

Homograph

Sense

Sense description

Translation

Translation description

entry

LIL Concept description

language

LDL Concept description

term

Term description

Methodological aspects of the conversion

- Semantic interoperability (what a data element means)
- Syntactic interoperability (how data elements are combined)
- An example

Semantic interoperability

- **Mapping of data elements**
  - Common elements (part of speech)
  - Corresponding („mappable“) elements (headword = term)
- **Solution: a common semantic repository**
  - ISO 11179 model
  - About 2000 elements and permissible values coming from ISO 12620, ISO 1087, ISO 16642, etc.

... that refers to the ISO TC 37 Data Category Registry

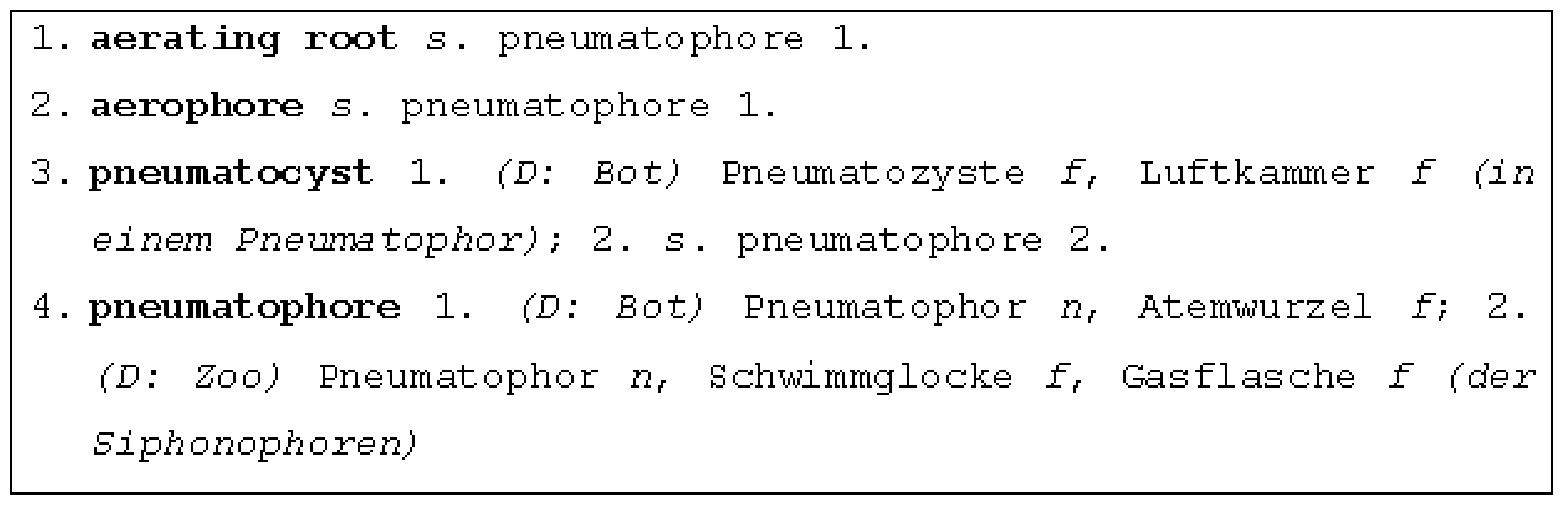
Syntactic interoperability

Mapping structures, taking into account:

- Synonymy (referring entries)
- Homography (homograph numbers)
- Polysemy (sense numbers)
- Factorization

Methodology of convertion
An example

A practical case : Langenscheidt specialised dictionaries



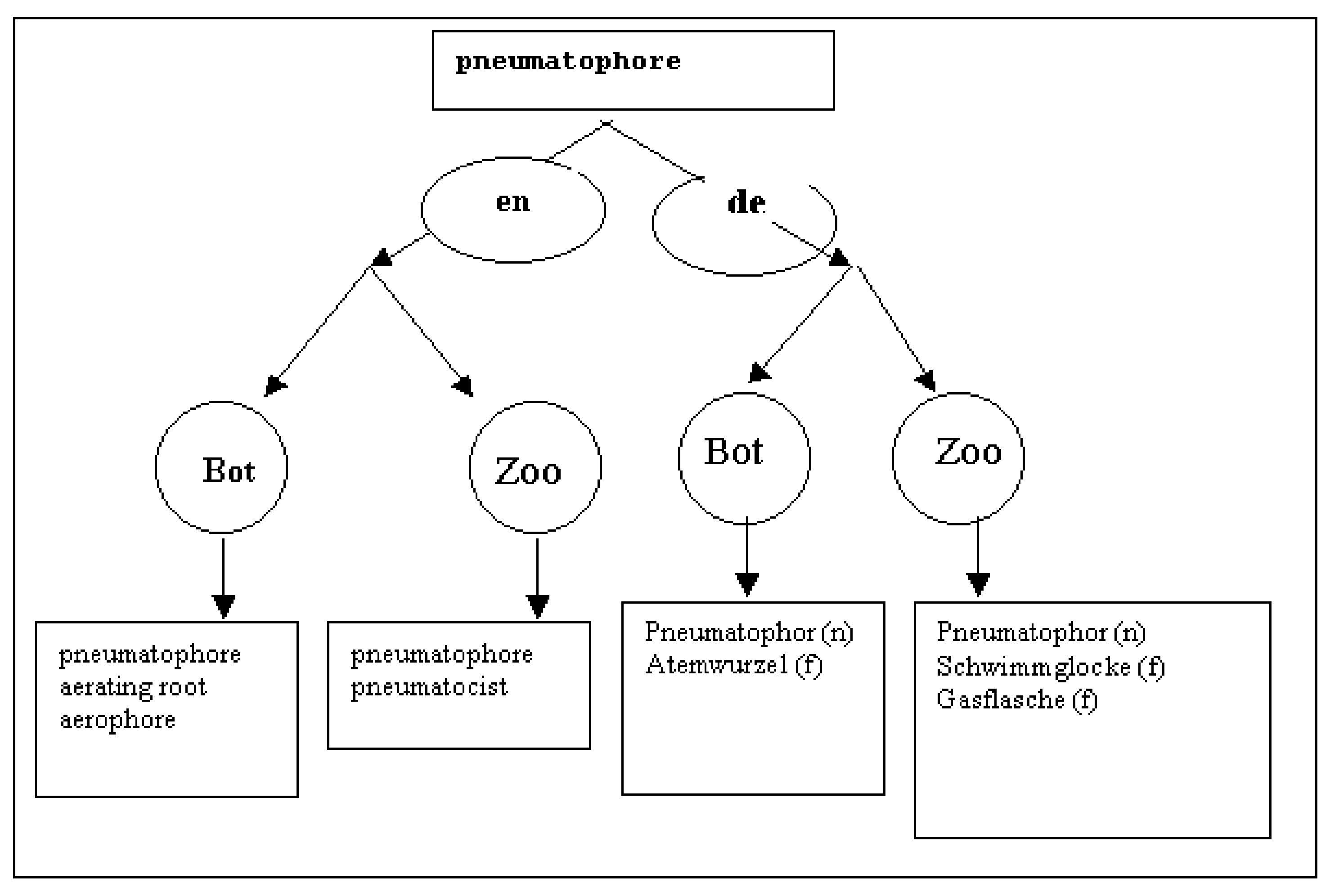
1. **aerating root** *s.* pneumatophore 1.

2. **aerophore** *s.* pneumatophore 1.

3. **pneumatocyst** 1. *(D: Bot)* Pneumatozyste *f*, Luftkammer *f (in einem Pneumatophor)*; 2. *s.* pneumatophore 2.

4. **pneumatophore** 1. *(D: Bot)* Pneumatophor *n*, Atemwurzel *f*; 2. *(D: Zoo)* Pneumatophor *n*, Schwimmglocke *f*, Gasflasche *f (der Siphonophoren)*

First step : Clustering synonyms



pneumatophore

en

de

Bot

Zoo

Bot

Zoo

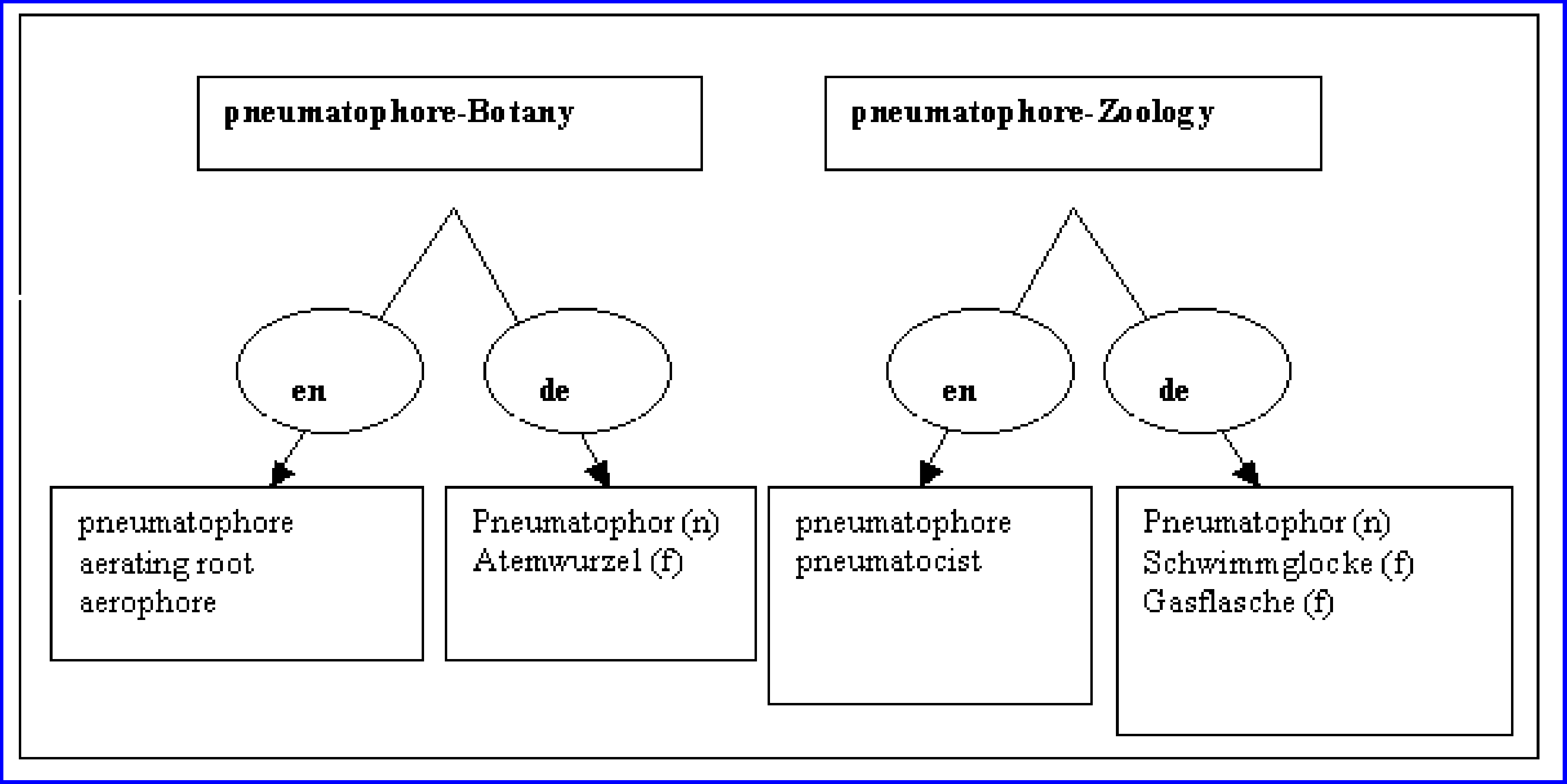
pneumatophore
aerating root
aerophore

pneumatophore
pneumatocist

Pneumatophor (n)
Atemwurzel (f)

Pneumatophor (n)
Schwimmglocke (f)
Gasflasche (f)

Step 2 : Splitting concepts

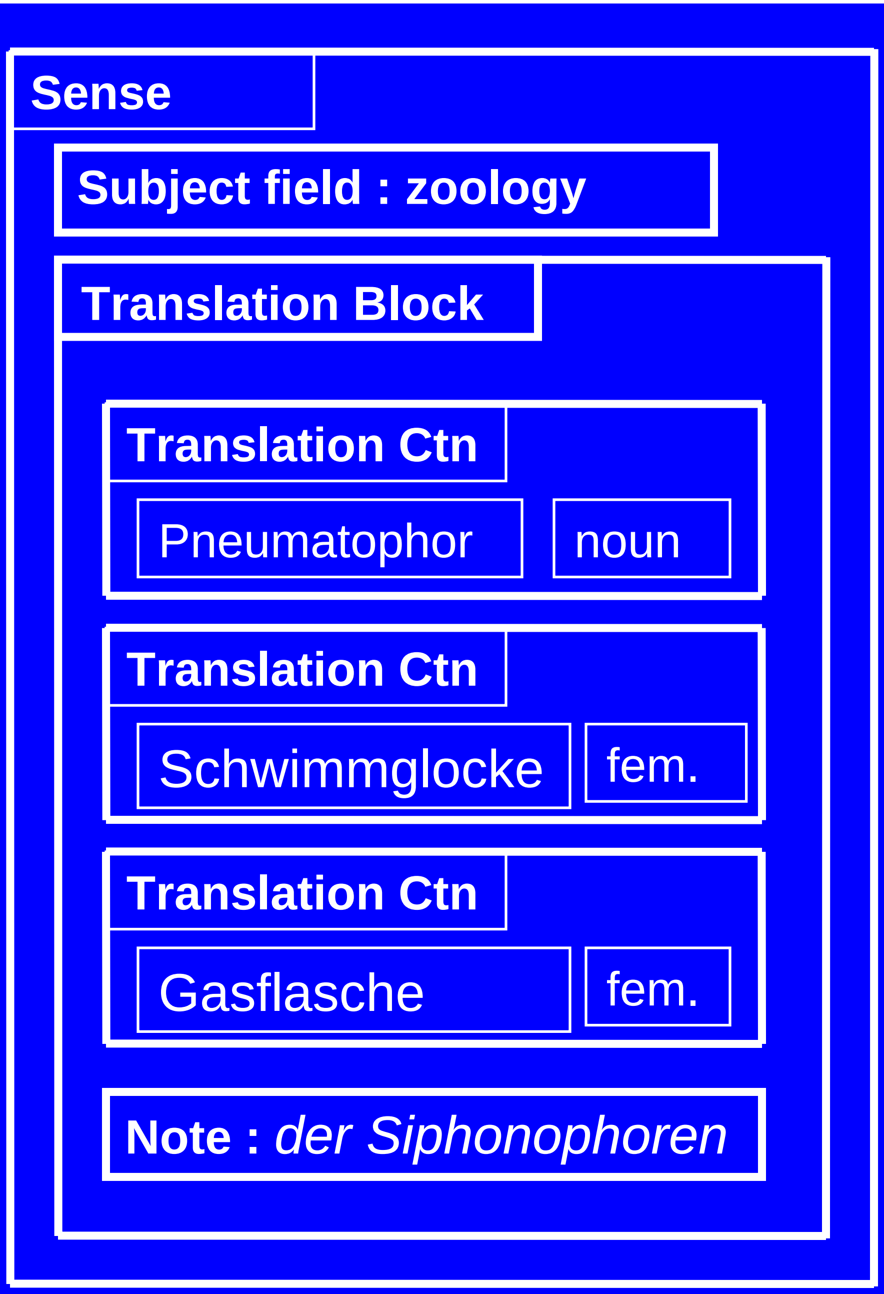


2006-05-24

Syntactic interoperability issues: structural factorization

1*(D: Zoo)* Pneumatophor *n*, Schwimmglocke *f*, Gasflasche *f (der Siphonophoren)*

2006-05-24



Sense

Subject field : zoology

Translation Block

Translation Ctn

Pneumatophor | noun

Translation Ctn

Schwimmglocke | fem.

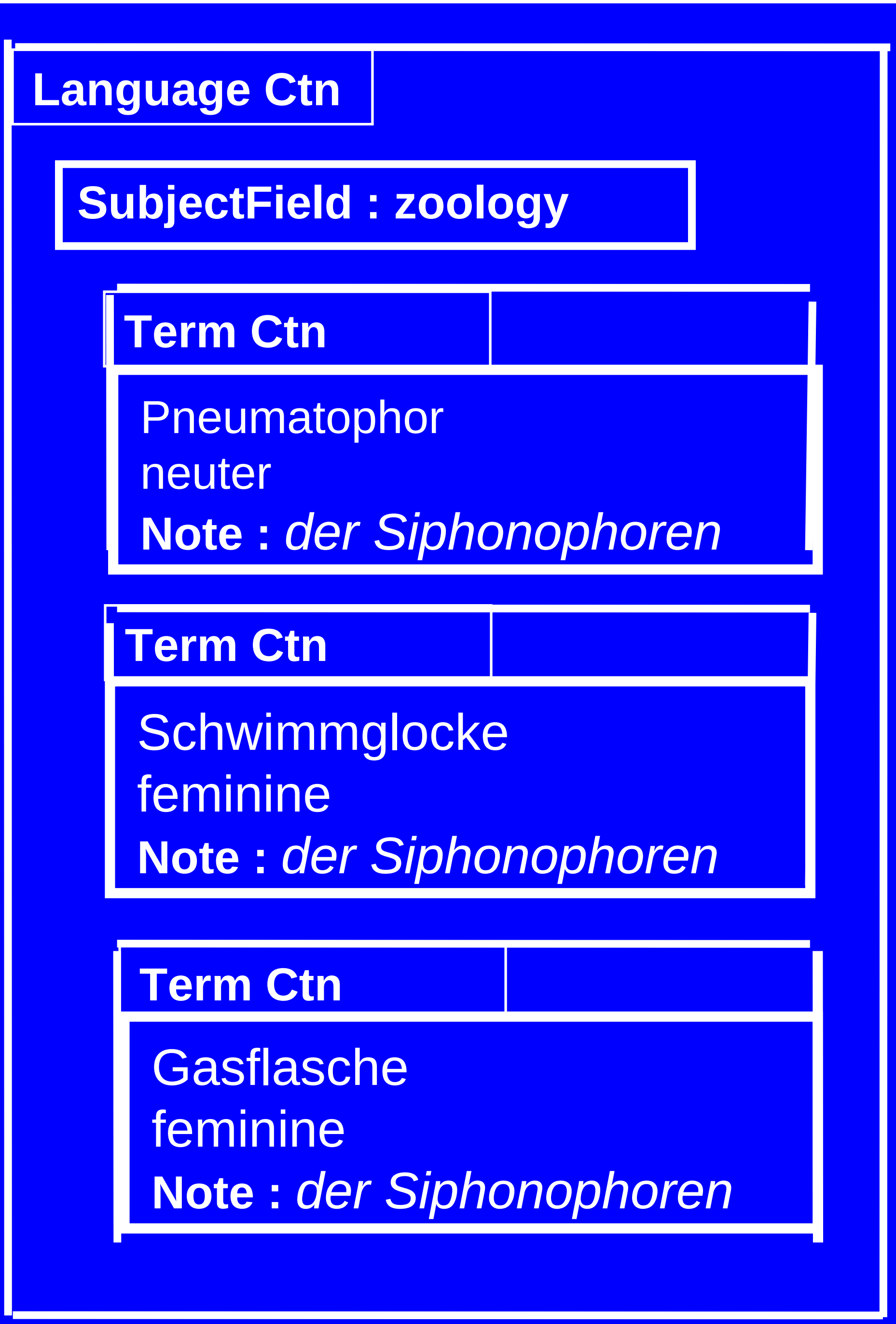
Translation Ctn

Gasflasche | fem.

Note : *der Siphonophoren*

Syntactic interoperability issues: structural factorization

1*(D: Zoo)* Pneumatophor *n*, Schwimmglocke *f*, Gasflasche *f (der Siphonophoren)*



Language Ctn

SubjectField : zoology

Term Ctn

Pneumatophor
neuter
Note : *der Siphonophoren*

Term Ctn

Schwimmglocke
feminine
Note : *der Siphonophoren*

Term Ctn

Gasflasche
feminine
Note : *der Siphonophoren*

How to test ?

- Read TermBridge home page : htp://www.genetrix.org (TermBridge is an XML framework for lexicography, terminology, and all their related informations)

- Read XmLex introduction :

  ttp://www.xmlex.net/lexicography/xmlexintro.pdf

- Download, unpack the LexTermLib.rar : ttp://ww.XmLex.net/lexicography/XmLexWorkbench.rar

Conclusions

- Specialist dictionary publishers will have to act in the future as content providers for language tools in order to meet their actual needs and will therefore have to concentrate more and more on the life cycle of data (production, maintenance, reusability)

- Lexical data bases hosting various semantically and syntactically compatible human-readable formats are the future of linguistic data management based on single sourcing strategies

- Technically, it is important to rely on publicly available standards (syntax = models, semantics = data elements) and to be compatible with XML and XSL methodology and tools.

- Experience shows that standardization is not a danger for quality and originality of content. In any case it is a guarantee for long life investment

Thank you for your attention !

anre.lemeur@uhb.fr

htp://www.genetrix.org

maie-jeanne.derouin@langenscheidt.de

htp://www.langenscheidt.de