Lexical Ressources -Examples

DBR-MAT

lex_entry_g(40,'Oelphase',term,_,
crossref([],[35,89]),25,none,
sem([29],[],none))

VERBMOBIL

Entry 10788 matches

String key verkehrsg"unstiger

    Orth: verkehrsg"unstiger

    Phon:f6ke:6sgYnstIg6

    OrthSeg:ver+kehr#+s#g"unst+ig+er

    PhonSeg:f6.+k'e:6#+s+g''Yns.t+I.g#+6

    OrthStem:ver+kehr#+s#g"unst+ig

        ver+kehr#+s#g"unst+ig

    PhonStem: ...

    Flex: A,mixed,sg,nom,mask,pos

        A,strong,pl,gen,fem,pos

        .....

Limitations of actual standards

impose saving
format

➡️

- the lexicons were produced with
the help of systems that are not any
longer maintained: no export
facility

- lexicons contain procedural
elements dependent on hosting
system

no support of joining
two lexicons

➡️

- formats do not match

- differences in linguistic classes

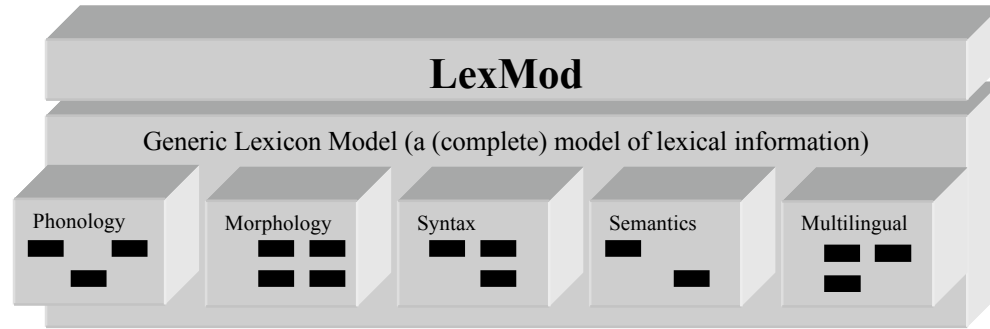- inconsistency or different
granularities for entries

MANAGELEX

- Generic lexicon management model
- Reads, converts and combines lexicons independent of :
  - format
  - language
  - system requirements
- Architecture levels (following ANSI specifications):

  - meta model level

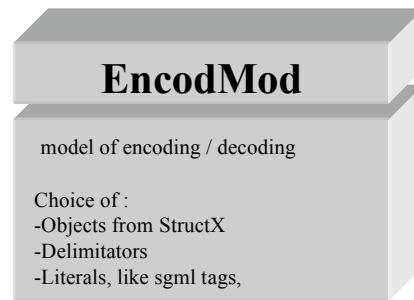 - model level

  - real world level
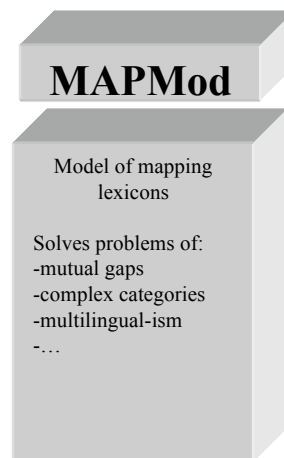
MANAGELEX-Meta-Model level

LexMod

Generic Lexicon Model (a (complete) model of lexical information)

| Phonology | Morphology | Syntax | Semantics | Multilingual |

EncodMod

model of encoding / decoding

Choice of :
-Objects from StructX
-Delimitators
-Literals, like sgml tags,
-...

MAPMod

Model of mapping
lexicons

Solves problems of:
-mutual gaps
-complex categories
-multilingual-ism
-…

Module LexMod is a rich model of possible lexical information

- MULTILEX, MILE

Module Encod specifies the data structure for a specific entry and for the lexicon

- OLIF, SALT

Module Map specifies the mapping of two lexicons

- takes into account mutual gaps, complex categories, etc.

© v.Hahn 2004

MANAGELEX-Model level

| EditTool | Writes/ reads/edits entries |

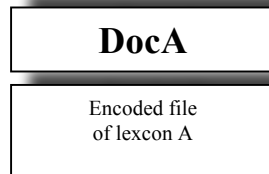| StructTool | Defines/changes linguistic structures |

| CodeTool | - Encodes entries+structures into files

- Decodes Lexicon files |

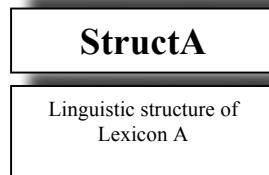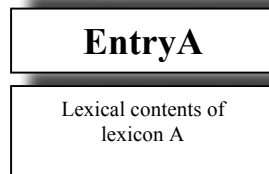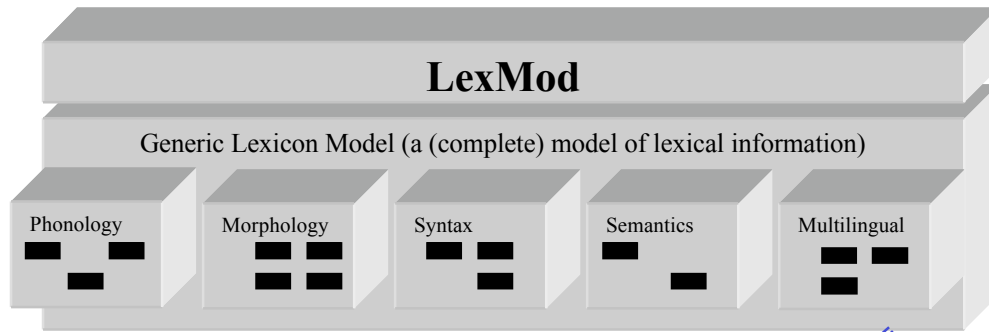| MapTool | Merges two lexicons with possibly different structure |

MANAGELEX- Real world level

DocA

Encoded file
of lexcon A

StructA

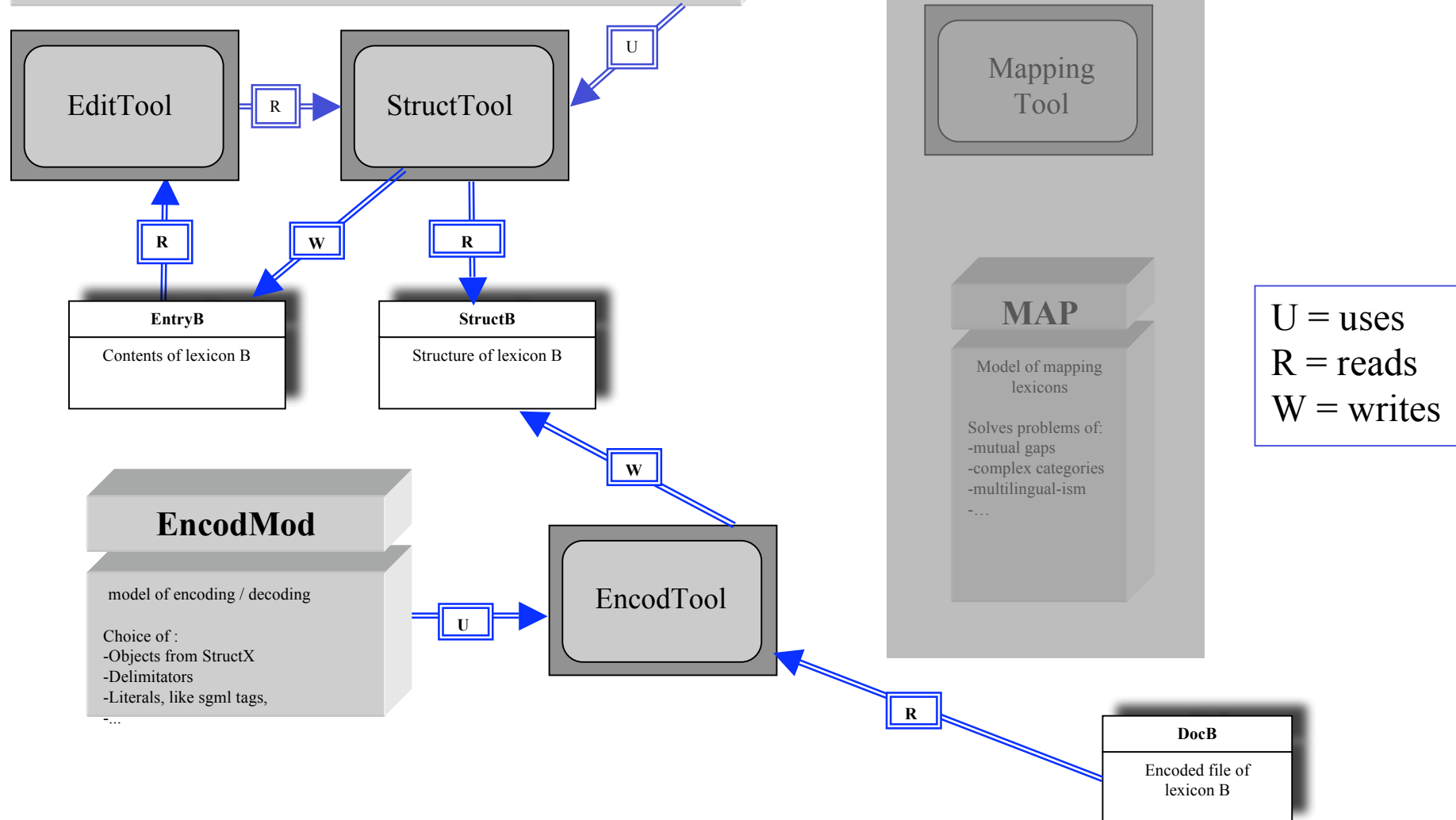Linguistic structure of
Lexicon A

EntryA

Lexical contents of
lexicon A

Encoded lexicons

Lexicon structure

lexical contents

Real, present,
distinct objects

LexMod

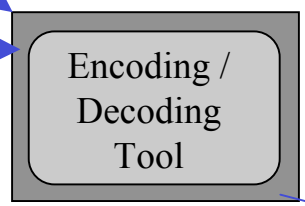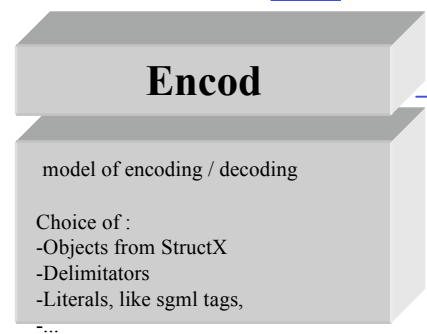Generic Lexicon Model (a (complete) model of lexical information)

| Phonology | Morphology | Syntax | Semantics | Multilingual |

Function: Reading a lexicon

EditTool → **R** → StructTool ← **U**

R (EditTool → EntryB)

W (StructTool → EntryB)

R (StructTool → StructB)

EntryB
Contents of lexicon B

StructB
Structure of lexicon B

Mapping Tool

MAP

Model of mapping lexicons

Solves problems of:
-mutual gaps
-complex categories
-multilingual-ism
-…

U = uses
R = reads
W = writes

W (EncodTool → StructB)

EncodMod

model of encoding / decoding

Choice of :
-Objects from StructX
-Delimitators
-Literals, like sgml tags,
-...

U → EncodTool ← **R**

DocB
Encoded file of lexicon B

© v.Hahn 2004

7

LexMod

Generic Lexicon Model (a (complete) model of lexical information)

| Phonology | Morphology | Syntax | Semantics | Multilingual |

Function:
Building / updating a lexicon

EditTool

StructTool

U

R

W

Mapping
Tool

StructA

Structure of Lexicon A

MAP

Model of mapping
lexicons

Solves problems of:
-mutual gaps
-complex categories
-multilingual-ism
-…

R

Encod

model of encoding / decoding

Choice of :
-Objects from StructX
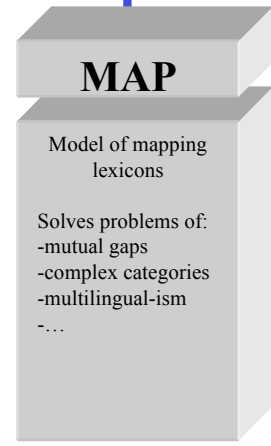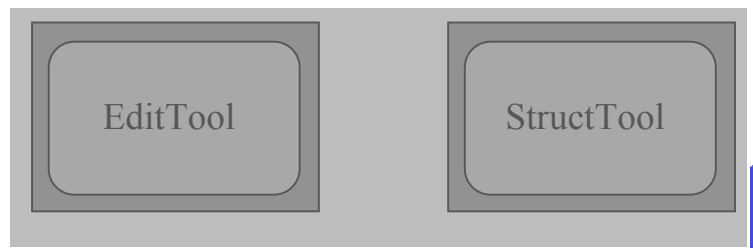-Delimitators
-Literals, like sgml tags,
-...

U

Encoding /
Decoding
Tool

U = uses
R = reads
W = writes

W

DocA

Encoded file
of lexiconA

© v.Hahn 2004

8
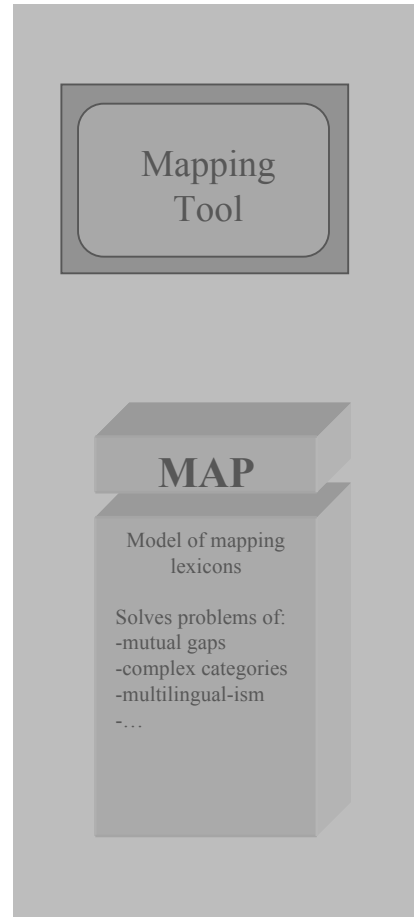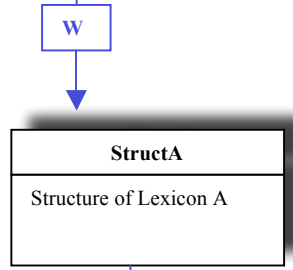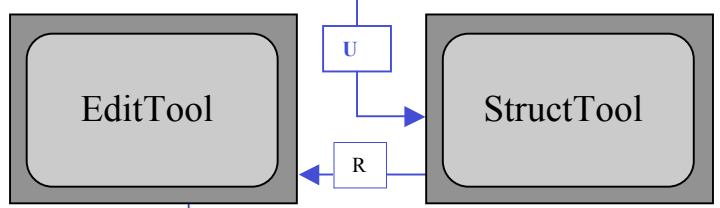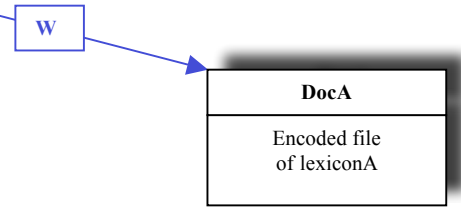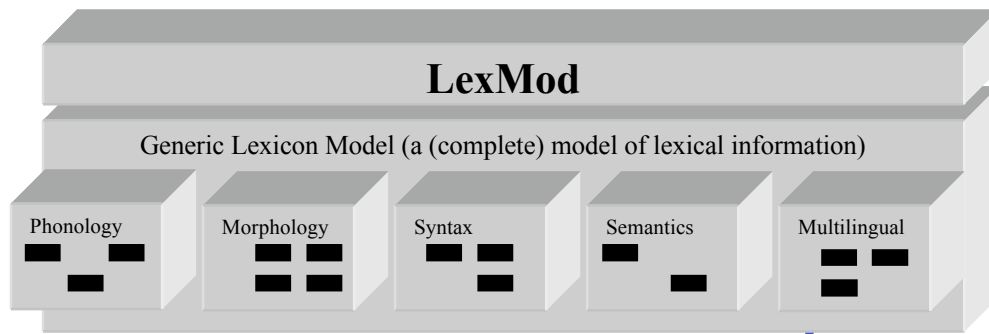
Function: Join two lexicons

LexMod

Generic Lexicon Model (a (complete) model of lexical information)

Phonology Morphology Syntax Semantics Multilingual

EditTool StructTool

Mapping Tool

U

R

R

W

U

MAP

Model of mapping lexicons

Solves problems of:
-mutual gaps
-complex categories
-multilingual-ism
-…

StructA

Structure of Lexicon A

StructB

Structure of Lexicon B

StructA+B

Structure of Lexicon A+B

Editing surface

U = uses
R = reads
W = writes

Encod

model of encoding / decoding

Choice of :
-Objects from StructX
-Delimitators
-Literals, like sgml tags,
-...

Encoding / Decoding Tool

© v.Hahn 2004

9

State of development

- Prototyping phase

- LexMod, StructTool and EditTools are implemented

- for the moment only European languages are modeled

- no replacement of actual standards but

- supply the use and reuse of already developed non-standard resources