Prototype of the multilingual Balkanet semantic network

Deliverable D.7.1, WP7, BalkaNet,
IST-2000-29388

BalkaNet
March 2004

Identification Number:	IST-2000-29388
Type:	Report
Title:	Loading the individual Wordnets into the multilingual Balkanet lexical database
Workpackage:	WP7
Deliverable:	D.7.1
Date:	April 2004
Status:	Final
Deliverable Responsible:	MEMODATA
EC Project Officer:	Erwin Valentini
Project Coordinator:	Prof. Dimitrios Christodoulakis Director, DBLAB Computer Engineering and Informatics Department Patras University GR – 265 00 PATRA Phone: +30 610 960385 Fax: +30 610 960438 e-mail: dxri@cti.gr
Actual Distribution:	Project Consortium, Project Officer, EC

CONTENTS

PREFACE

In this task, we will see how to use the different monolingual wordnets to merge them in one global multilingual WordNet.

These wordnets developped by the different partners as seen in the previous tasks and are available in Bulgarian, Czech, Greek, Romanian, Serbian and Turkish. This subtask describes the different steps to create a multilingual WordNet from these individual WordNets.

To evaluate the reusability and the robustness of the Balkanet model, derived from the EuroWordNet model, it was important to use a different tool than the one used to build the different WordNets (WMS, VisDic) and developped by a partner not involved in their constructions.

That's why we will load the different WordNets in The Integral Dictionary, TID, developped by Memodata, to create this global multilingual WordNet. The Integral Dictionary is a semantic network associated to a lexicon. Its size is comparable with WordNet. The Integral Dictionary organizes words into a variety of concepts and uses semantic lexical functions. Concept definitions are based on the componential semantic theory, the decomposition of the words into a set of smaller units of meaning, and the lexical functions are inspired by the Meaning-Text theory.

This model, being different, will allow to evaluate the flexibility and, hence, the power of the Balkanet one.

PART I
Structure of a WordNet file

Each WordNet has been built by the Wordnet Management System and is then stored in one individual XML file. To handle the different languages, these XML files will use the Unicode Charset (UTF8). Those are the files used by VisDic.

Here is an extract of the Romanian Wordnet XML file :

```
<SYNSET><ID>ENG20-00004609-n</ID><POS>n</POS><SYNONYM><LITERAL>via
a<SENSE>1</SENSE></LITERAL></SYNONYM><DEF>forme de viata, vazute in mod
global; "Nu exista viata pe Marte"</DEF><STAMP>Dan
Cristea</STAMP><BCS>1</BCS><ILR>ENG20-00003009-
n<TYPE>hypernym</TYPE></ILR></SYNSET><SYNSET><ID>ENG20-00004824-
n</ID><POS>n</POS><SYNONYM><LITERAL>celula<SENSE>1</SENSE></LITERAL></S
YNONYM><DEF>Element constitutiv fundamental al organismelor vii, alcatuit din membrana,
citoplasma si nucleu, reprezentand cea mai simpla unitate anatomica.</DEF><STAMP>Dan
Cristea</STAMP><BCS>1</BCS><ILR>ENG20-00003009-
n<TYPE>hypernym</TYPE></ILR><ILR>ENG20-00003226-
n<TYPE>holo_part</TYPE></ILR><ILR>ENG20-05681603-
n<TYPE>category_domain</TYPE></ILR></SYNSET>
```

Description of the different tags :

- SYNSET : contains all the data relative to Synset.
- ID : identifier of the ILI. The prefix ENG20 means that it had been created by the Princeton WordNet, version 2.0.
- POS : part of speech. The possible values are :
    - n : noun
    - v : verbe
    - b : adverb
    - a : adjective
- SYNONYM : list of the literals of this synset. At least one literal is mandatory.
    - LITERAL : wording of the literal
    - SENSE : number used for the sense differentiation.
    - LNOTE : note about this literal
- Def : gloss of the synset. This wording allows to describe the synset. It's not mandatory.
- STAMP : gives some additional information about this synset : author, date...
- USAGE : gives an example of use of the synset
- BCS : number of the base concept associated with this synset. The possible values are 1, 2 or 3.

- ILR : Interlingua relation. Gives a relation between this synset and the specified Ili.
    - TYPE : type of this relation. The possible values are : be_in_state, category_domain, causes, derived, eng_derivative, holo_member, holo_part, holo_portion, hypernym, near_antonym, particle, region_domain, similar_to, subevent, usage_domain, verb_group

PART II
Structure of the Integral Dictionary

General presentation

As said above, The Integral Dictionary (TID) organizes words into a variety of concepts and uses semantic lexical functions. Concept definitions are based on the componential semantic theory, the decomposition of the words into a set of smaller units of meaning, and the lexical functions are inspired by the Meaning-Text theory.
The basic component of TID is called a "concept". Each concept is annotated by a gloss written in mostly in French that describes intentionally its content. It consists of three main ontologies:

A first ontology is based on the relations generic or specific. When a concept is entirely lexicalized, a particular relation between the concept and the literal is used : *generic*. When the word does not describe the concept entirely, the relation is said to be *specific*.
A second one is based on a thesaurus, similar to the Roget's, but more linguistically restricted. It includes thousands of themes (domains or small conceptual worlds).
The third ontology describes lexical-syntactic patterns.

The Integral Dictionary also contains a large number of lexical functions that generate word senses from another word sense given as an input.
One important property of the Integral Dictionary is its structure: merging several approaches (hence its name), the Integral Dictionary is fundamentally an acyclic oriented graph instead of a tree.

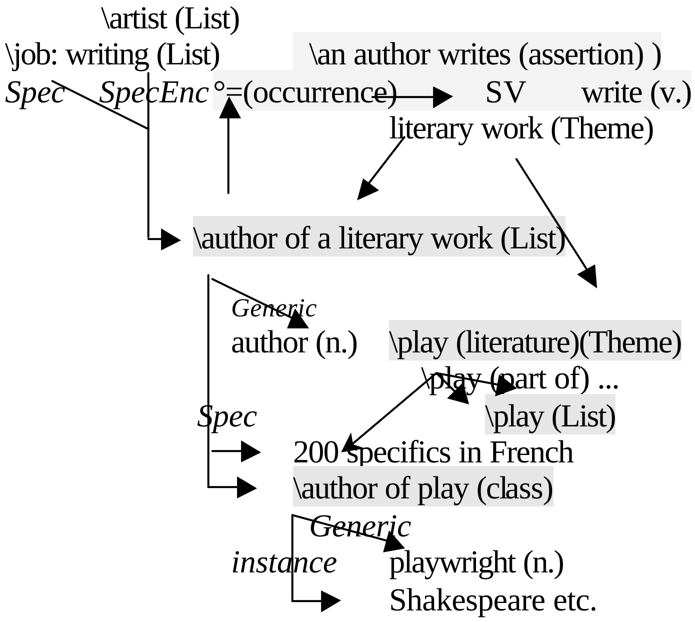


Figure 1: An excerpt of TID.

Figure 1 is an excerpt of TID, which shows that:

The class \author (List) is possibly subsumed by the class: \artist (List). (Enc means potentiality, Spec Enc means "is a" potentially.
In this class, the generic word in English is author.n.
The class contains a subclass labeled "\author of play", which is a specific.
Shakespeare is an instance of the previous class.
The class \author (List) belongs to a theme, a possible topic called \literary work (theme).
This theme contains the subtheme \play (literature) (Theme).
Finally, the \author (List) is directly connected to a part of its preferred assertion: write (a literary text).

We call relation a link from a node to another node and we never count the symmetrical links. For French, TID contains around 220,000 relations similar to that of the example in Figure 1. Concerning the lexical function borrowed from the Meaning-text theory, we have also 150,000 occurrences of relations for French. A part of them, 15,000, is not validated yet.
The multilingual part, English, Italian, Spanish, German, Dutch, and Portuguese, represents 300,000 relations to add to the previous number.

Architecture

The data of TID are stored in a relational database : Firebird 1.5. This free RDBMS has all the needed characteristics for such a project :

- Capacity (in terms of number of records, columns, tables, indexes...)
- Support of Unicode : this requirement was of course mandatory to store so many different languages with their different character sets.
- Simplicity
- Speed

- Multi platform (Windows, Linux)
- Stored procedures

Basically, the architecture of TID is very simple because relying mainly in two tables :
the LEAF table and the RELATION table.

The LEAF table :

This table contains all of the nodes of the graph : the concepts, the word senses (or the literals, using the WordNet terminology), the glosses, the ILIs.
Structure of this table :

Field	Type	Length	description
LANGUAGE	Char	1	language of the Word Sense or the gloss. The different values are : E English F French I Italian D German H Dutch S Espagnol P Portuguese G Greek T Turkish R Rumanian C Czech Y Serbian B Bulgarian W Swedish For a concept or an ILI, we use the character 'M' as metalanguage
SITE	Char	1	Code which indicates where the node has been created (this information is mandatory for the unicity of the key)
NUMBER	Integer		Numerical element of the key
WORDING	Varchar	400	Wording of a word sense or a gloss. In the case of a concept or an ILI, this field can be blank : the wording of these nodes would then be assured by other nodes of type Gloss, linked to them.
GRAMMAR	Integer		This field contains the code of the POS for a Word Sense or the type of a Concept. These informations are stored in an additional table.
DATE	Date		Date of creation or modification of the node
MODEL	Integer		Number of inflection model for the nouns, adjectives or verbs
ARTICLE	Varchar	80	Contains the 80 first characters of the wording in uppercase. This allows to retrieve a node by it's wording. In some cases this field may be different of the beginning of the wording. This

			field is indexed.

The three first fields compose the primary key of the node (ex MA15224, EW566711...)

The RELATION table :

Each row of this table contains the relation between two nodes. We call, conventionally, the first node the child node and the second node the parent node.
Structure of this table :

Field	Type	Length	description
LANGUAGE_CHILD	Char	1	Language of the child node
SITE_CHILD	Char	1	Site of the child node
NUMBER_CHILD	Integer		Number of the child node
LANGUAGE_PARENT	Char	1	Language of the parent node
SITE_PARENT	Char	1	Site of the parent node
NUMBER_PARENT	Integer		Number of the parent node
DATE	Date		Date of creation of the relation
TYPE	Integer		Type of the relation (specific, generic, etc). These informations are stored in an additional table.
LANGUAGE_CONTEXT	Char	1	Language of the context node (see below)
SITE_CONTEXT	Char	1	Site of the context node
NUMBER_CONTEXT	Integer		Number of the context node

The context is a node which allows to precise the context of a relation.

Let's explain this notion of Context.

Let’s consider the relations in Figure 1 again. Figure 2 shows the initial data format that TID used to represent it.

Child	Parent	KindOfRel
Author (n)	\author of a lit…	Generic
\author of play	\author of a lit…	Specific
etc.		

Figure 2: A general record in the table RELATION in TID.

Although this format was satisfactory for hierarchical data, it reached its limits when we introduced syntactical relations. Let’s consider the syntactic definition in Figure 1:

\author of a literary work (List)	SV	\write
	VO	\texts

Figure 3 shows the table in TID using the same formalism.

Child	Parent	KindOfRel
\author of a literary work (List)	\write	SV
\write	\texts	VO
etc.		

Figure 3: A part of TID.

However, it not possible to consider that \author of a literary work (List) is the child of \write and the grandchild of \text in Figure 3 in the same way it is the child of \author of a lit… in Figure 2. In addition, in terms of graph, Figure 3 cannot record the syntactic paths without ambiguity, for example if write exists in many different assertions.

Syntactic patterns and lexical ontology represent two different viewpoints that are not necessarily related. To represent them with a relational database, we must take into account that these two dimensions (syntactic/paradigmatic) are different. Figure 4 shows the integration results where

OntoTID means ontology of TID and SyntTID means Syntactical Pattern of TID.
The index (1) is the key of the complete pattern.
The two last records indicate that OntoTID and SyntTID are parts of TID.

This format is more flexible and provides rich new possibilities. Firstly, the format can record any kind of hypergraph in a relational database. Secondly, it enables us to extend the group theory approach to a more general mereology.

Child	Parent	KindOfRel	Context
Author (n)	\author of a lit…	Generic	OntoTID
\author of play	\author of a lit…	Specific	OntoTID
etc.			
\author …(List)	\write	SV (1)	SyntTID
\write	\texts	VO (1)	SyntTID
etc.			
OntoTID	PartOfTID	part of	TID
SyntTID	PartOfTID	part of	TID

Figure 4: A part of TID.

We have used this format to integrate a set of ontological resources. Concerning EuroWordNet and Balkanet, the format allows us to upload data from xml files to a relational database. Figure 5 shows an excerpt of records where (1) is a key identifying a synset.

Since a synset has its gloss and literal, we have the English gloss {writes (books or stories or articles or the like) professionally (for pay)…} and the English literal *author* located in the *English WordNet.* We notice that in this case, *auteur (n)* is placed in the synset (1) in the *French WordNet.* In the end, it's also possible to generate the complete list of InterLingua index (ILI).

Child	*Parent*	*KindOfRel*	*Context*
Author (n)	*(ILI 1)*	*Literal*	*EnWordNet*
{writes (books or stories or articles or the like) professionally (for pay)…}	*(ILI 1)*	*Gloss*	*EnWordNet*
auteur (n)	*(ILI 1)*	*Litteral*	*FrWordNet*
(ILI 1)	*Interlingua*	*Elementof*	*ILIs*

Figure 5: The WordNets.

We will see in part four that the relations between synsets may occur in some WordNets and not in other ones. Context will allow to represent that.

Indexes :

The primary key is made up of the following fields :

LANGUAGE_CHILD
SITE_CHILD
NUMBER_CHILD
LANGUAGE_PARENT
SITE_PARENT
NUMBER_PARENT
TYPE
LANGUAGE_CONTEXT
SITE_CONTEXT
NUMBER_CONTEXT

There are two secondary indexes. One on the fields LANGUAGE_CHILD, SITE_CHILD, NUMBER_CHILD and one on the fields LANGUAGE_PARENT, SITE_PARENT, NUMBER_PARENT.
These two indexes allow to get the childs or the parents of a node.

PART III
Loading of a WordNet in the Integral Dictionary Architecture

The first task to load a WordNet in the Integral Dictionary architecture will be to translate the XML file in SQL queries to fill the tables LEAF and RELATION.
We have used a XSLT program to perform this transformation. XSL is a language which allows to transform an XML file in an other format.

The xsl script :

Here is the (simplified) XSL script to load the Greek Wordnet :

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
<xsl:output method="text" indent="yes"/>

<xsl:variable name="date">2004-2-27</xsl:variable>
<xsl:variable name="site">G</xsl:variable>
<xsl:variable name="langue">G</xsl:variable>
<xsl:variable name="extfic">_GREC</xsl:variable>
<xsl:variable name="wordnetName">Greek Wordnet</xsl:variable>


<xsl:template match="SYNSETS">

CREATE TABLE LEAF<xsl:value-of select ="$extfic "/> (
   LANGUE       VARCHAR(1) CHARACTER SET UNICODE_FSS NOT NULL ,
   SITE         VA RCHAR(1) CHARACTER SET UNICODE_FSS NOT NULL ,
   NUMERO       DECIMAL(15,0) NOT NULL,
...
);
...
/*Creation of other tables and indexes*/
...
INSERT INTO LEAF<xsl:value-of select ="$extfic"/> VALUES ('M','<xsl:value-of select
="$site"/>',0, '<xsl:value-of select ="$wordnetName"/>',0, '<xsl:value-of select
="$date"/>',NULL,NULL,32500,NULL,NULL,0,1,NULL,NULL,NULL);
INSERT INTO RELATION<xsl:value-of select ="$extfic"/> VALUES ('M','<xsl:value-of select
="$site"/>',0,'M','A',0,'<xsl:value-of select ="$date"/>',0,NULL,NULL,NULL,'M','A',0);

INSERT INTO LEAF<xsl:value-of select ="$extfic"/> VALUES ('<xsl:value-of select
="$langue"/>','<xsl:value-of select ="$site"/>',1, '<xsl:value-of select ="$wordnetName"/>',10,
'<xsl:value-of select ="$date"/>',NULL,NULL,32500,NULL,NULL,0,1,NULL,NULL,NULL);
INSERT INTO RELATION<xsl:value-of select ="$extfic"/> VALUES ('<xsl:value-of select
="$langue"/>','<xsl:value-of select ="$site"/>',1,'M','<xsl:value-of select ="$site"/>',0,'<xsl:value-of
select ="$date"/>',10,NULL,NULL,NULL,'M','A',1);


<xsl:for-each select= "SYNSET">

<xsl:variable name="position">
<xsl:value-of select ="position()*40"/>
</xsl:variable>
```

```
/* creation of the POS code*/
<xsl:variable name="gram">
<xsl:choose>
<xsl:when test="POS='a'">20</xsl:when>
<xsl:when test="POS='b'">30</xsl:when>
<xsl:when test="POS='v'">40</xsl:when>
<xsl:when test="POS='n'">50</xsl:when>
<xsl:otherwise>-1</xsl:otherwise>
</xsl:choose>
</xsl:variable>
<xsl:variable name="gloss">
<xsl:value-of select ="DEF"/>
</xsl:variable>
<xsl:variable name="glosstronc">
<xsl:choose>
    <xsl:when test='substring($gloss, 80,1) ="&apos;"'>
      <xsl:value-of select="substring($gloss,1,78)"/>
    </xsl:when>
    <xsl:otherwise>
      <xsl:value-of select="substring($gloss,1,80)"/>
    </xsl:otherwise>
  </xsl:choose></xsl:variable>

/* insertion of the ILI */
INSERT INTO LEAF<xsl:value-of select ="$extfic "/> VALUES ('M','<xsl:value-of select
="$site"/>',<xsl:value-of select ="$position"/>, '<xsl:value-of select = "ID"/>',1, '<xsl:value-of select
="$date"/>',NULL,NULL,32500,NULL,NULL,2,1,NULL,NULL, '<xsl:value-of select = "ID"/>');
/* insertion of the gloss */
INSERT INTO LEAF<xsl:value-of select ="$extfic "/> VALUES ('<xsl:value-of select
="$langue"/>','<xsl:value-of select ="$site"/>',<xsl:value-of select ="$position"/>,'<xsl:value-of select
="$gloss"/>',10,'<xsl:value-of select
="$date"/>',NULL,NULL,32500,NULL,NULL,0,1,NULL,NULL,UPPER('<xsl:value-of select
="$glosstronc"/>'));

/* Extraction of other informations : STAMP, USAGE, etc*/

/* relation between the gloss and the ILI */
INSERT INTO RELATION<xsl:value-of select ="$extfic "/> VALUES ('<xsl:value-of select
="$langue"/>','<xsl:value-of select ="$site"/>',<xsl:value-of select ="$position"/>,'M','<xsl:value-of
select ="$site"/>',<xsl:value-of select ="$position"/>,'<xsl:value-of select
="$date"/>',10,NULL,NULL,NULL,'M','A',1);

/* extraction of the literals */
<xsl:for-each select= "SYNONYM/LITERAL">
<xsl:variable name="numLiteral">
<xsl:value-of select ="$position+position()"/>
</xsl:variable>

INSERT INTO LEAF<xsl:value-of select ="$extfic "/> VALUES ('<xsl:value-of select
="$langue"/>','<xsl:value-of select ="$site"/>',<xsl:value-of select ="$numLiteral"/>, '<xsl:value-of
select = "text()"/>',<xsl:value-of select = "$gram"/>,'<xsl:value-of select
="$date"/>',NULL,NULL,32500,NULL,NULL,0,1,NULL,NULL, UPPER('<xsl:value-of select =
"text()"/>'));
INSERT INTO RELATION<xsl:value-of select ="$extfic "/> VALUES ('<xsl:value-of select
="$langue"/>','<xsl:value-of select ="$site"/>',<xsl:value-of select ="$numLiteral"/>,'M','<xsl:value-of
select ="$site"/>',<xsl:value-of select ="$position"/>,'<xsl:value-of select
="$date"/>',20,NULL,NULL,NULL,'M','A',1);
<xsl:for-each select= "LNOTE">
```

```
INSERT INTO SUPPL<xsl:value-of select ="$extfic "/> VALUES ('<xsl:value-of select
="$langue"/>','<xsl:value-of select ="$site"/>',<xsl:value-of select ="$numLiteral"/>, 'LNOTE',
'<xsl:value-of select = "text()"/>');
</xsl:for-each>
<xsl:for-each select= "SENSE">
INSERT INTO SUPPL<xsl:value-of select ="$extfic "/> VALUES ('<xsl:value-of select
="$langue"/>','<xsl:value-of select ="$site"/>',<xsl:value-of select ="$numLiteral"/>, 'SENSE',
'<xsl:value-of select = "text()"/>');
</xsl:for-each>
</xsl:for-each>

/* Extraction of the ILRs */
<xsl:for-each select= "ILR">
INSERT INTO LIENS<xsl:value-of select ="$extfic "/> VALUES ('<xsl:value-of select
="text()"/>','<xsl:value-of select ="TYPE"/>','M','<xsl:value-of select ="$site"/>',<xsl:value-of select
="$position"/>,'<xsl:value-of select ="$date"/>', NULL, NULL, NULL, NULL, 'M','<xsl:value-of
select ="$site"/>', 0 );
</xsl:for-each>
</xsl:for-each>

</xsl:template>
</xsl:stylesheet>
```

At this stage, all the data from the different WordNets are not merged yet. The problem is that the key for a node in TID (ex : MW1522450), as seen above, has a different structure of the identifier used in WordNet (ex : ENG20-00004609-n). Moreover, each XSLT transformation performs independently for each Wordnet and, thus, gives different keys to the same ILIs.

The SQL queries

To handle this issue, we first create a new table called CORRESP, which will contain the old key (from the individual WordNets) and the new key (from WordNet 2.0). If an ILI has been created in a monolingual Wordnet, we keep its key. We then fill this table with an SQL query, making the update by matching on the ILI identifier (ex : ENG20-00004609-n).

By some additional SQL queries, the monolingual tables are then merged in the multilingual tables.

Example : sql queries to import the Romanian WordNet

```
/*construction of the correspondance between the monolingual synsets
keys and the WordNet 2.0 synsets keys */
insert into corresp
select f1.langue, f1.site, f1.numero, f2.langue, f2.site, f2.numero
from leaf_roum f1
left join leaf f2
on f1.article = f2.article
and f1.langue='M';

update corresp
set langue_nouv = langue_anc,
site_nouv = site_anc,
numero_nouv=numero_anc
```

```
where langue_nouv is null
and site_nouv is null
and numero_nouv is null;
commit;

/*insertion of the leaves*/
insert into leaf
SELECT distinct f.*
from leaf_roum f, corresp c
WHERE
 (f.langue = c.langue_anc)
 AND (f.site = c.Site_anc)
 AND (f.Numero = c.Numero_anc)
 AND (c.site_nouv <> 'W')
 and (f.libelle <> '');
commit;

 /*insertion of the Synsets - gloss, literal relations */
insert into relation
SELECT DISTINCT c1.langue_nouv, c1.site_nouv, c1.numero_nouv,
c2.langue_nouv, c2.site_nouv, c2.numero_nouv, r.date_rel, r.type_rel,
r.ordre_rel,
r.periode, r.selection, r.ctx_langue, r.ctx_site, r.ctx_numero
FROM relation_roum r, corresp c1, corresp c2
WHERE
(c1.Langue_anc = r.Langue_fils)
 AND (c1.site_anc = r.Site_fils)
 AND (c1.Numero_anc = r.Numero_fils)
 AND (c2.Langue_anc = r.Langue_pere)
 AND (c2.Site_anc = r.Site_pere)
 AND (c2.Numero_anc = r.Numero_pere);
commit;

 /*insertion the relations Synsets - Synsets relations (ILR)*/
insert into relation
SELECT distinct c2.langue_nouv, c2.site_nouv, c2.numero_nouv,
c1.langue_nouv, c1.site_nouv, c1.numero_nouv, r.date_rel,
rl.code_rel, r.ordre_rel,
r.periode, r.selection, r.ctx_langue, r.ctx_site, r.ctx_numero
FROM liens_roum r, corresp c1, corresp c2, leaf_roum f, rel_lib rl
WHERE
(r.cle = f.libelle)
 AND (f.langue = c1.langue_anc)
 AND (f.site = c1.Site_anc)
 AND (f.Numero = c1.Numero_anc)
 AND (c2.Langue_anc = r.Langue_pere)
 AND (c2.Site_anc = r.Site_pere)
 AND (c2.Numero_anc = r.Numero_pere)
 AND (r.chaine_type_rel=rl.cat_fr);
commit;

/*insertion of the Synsets – BCS relations*/
insert into relation
SELECT distinct c.langue_nouv, c.site_nouv, c.numero_nouv,
r.langue_fils, r.site_fils, r.numero_fils, r.date_rel, r.type_rel,
r.ordre_rel,
r.periode, r.selection, r.ctx_langue, s.site, r.ctx_numero
FROM suppl_roum s, relation r, corresp c
where
r.langue_fils='M'
and r.site_fils='A'
and r.numero_fils=5
and s.cat='BCS'
and s.libelle='1'
and (s.langue = c.langue_anc)
and (s.site = c.Site_anc)
and (s.Numero = c.Numero_anc);

insert into relation
SELECT distinct c.langue_nouv, c.site_nouv, c.numero_nouv,
r.langue_fils, r.site_fils, r.numero_fils, r.date_rel, r.type_rel,
r.ordre_rel,
r.periode, r.selection, r.ctx_langue, s.site, r.ctx_numero
FROM suppl_roum s, relation r, corresp c
where
r.langue_fils='M'
and r.site_fils='A'
and r.numero_fils=6
and s.cat='BCS'
and s.libelle='2'
and (s.langue = c.langue_anc)
and (s.site = c.Site_anc)
and (s.Numero = c.Numero_anc);

insert into relation
SELECT distinct c.langue_nouv, c.site_nouv, c.numero_nouv,
r.langue_fils, r.site_fils, r.numero_fils, r.date_rel, r.type_rel,
r.ordre_rel,
r.periode, r.selection, r.ctx_langue, s.site, r.ctx_numero
FROM suppl_roum s, relation r, corresp c
where
r.langue_fils='M'
and r.site_fils='A'
and r.numero_fils=7
and s.cat='BCS'
and s.libelle='3'
and (s.langue = c.langue_anc)
and (s.site = c.Site_anc)
and (s.Numero = c.Numero_anc);
commit;
```

statistics on the Balkanet Wordnets :

Once loaded in the multilingual model, we have computed some statistics on the Balkanet data, that we can compare with the EuroWordNet ones.
Those statistics are displayed in the table below.

		Synsets	*No. of senses*	*Sens./ syns.*	*Entries*	*Sens./ entry*
EuroWordNet						
Dutch Wordnet	Nouns	34455	54428	1,58	45972	1,18
	Verbs	9040	14151	1,57	8826	1,6
	Other	520	1622	3,12	1485	1,09
	Total	44015	70201	1,59	56283	1,25

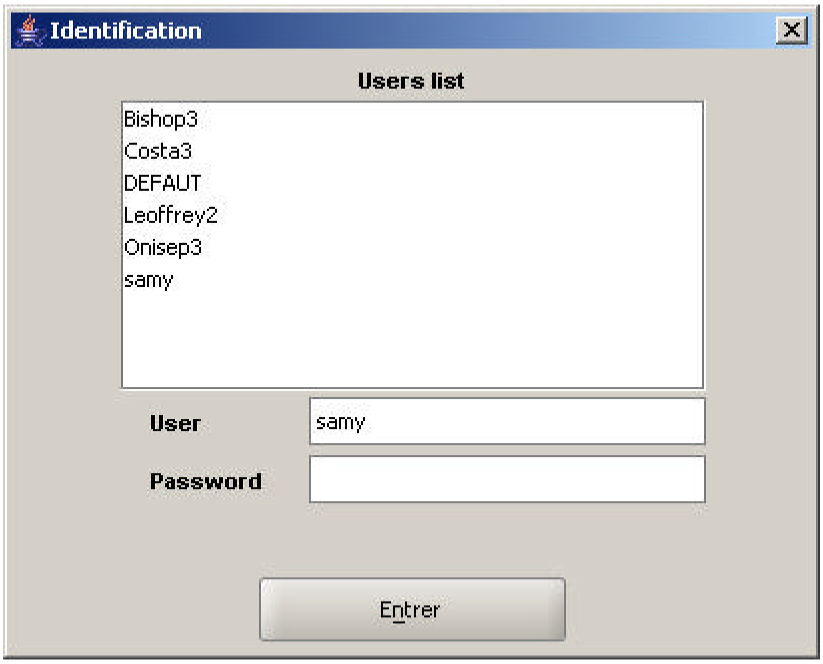
Spanish Wordnet	Nouns	18577	41292	2,22	23216	1,78
	Verbs	2602	6795	2,61	2278	2,98
	Other	2191	2439	1,11	2439	1
	Total	23370	50526	2,16	27933	1,81
Italian Wordnet	Nouns	30169	34552	1,15	24903	1,39
	Verbs	8796	12473	1,42	6607	1,89
	Other	1463	1474	1,01	1468	1
	Total	40428	48499	1,2	32978	1,47
French Wordnet	Nouns	17826	24499	1.37	14879	1.65
	Verbs	4919	8310	1.69	3898	2.13
	Other	0	0	0	0	0
	Total	22745	32809	1.44	18777	1.75
German Wordnet	Nouns	9951	13656	1.37	12746	1.07
	Verbs	5166	6778	1.31	4333	1.56
	Other	15	19	1.27	19	1
	Total	15132	20453	1.35	17098	1.20
Czech Wordnet	Nouns	9727	13829	1.42	9277	1.49
	Verbs	3097	6120	1.98	3006	2.04
	Other	0	0	0	0	0
	Total	12824	19949	1.56	12283	1.62
Estonian Wordnet	Nouns	5028	8226	1.64	7209	1.14
	Verbs	2650	5613	2.12	3752	1.50
	Other	0	0	0	0	0
	Total	7678	13839	1.80	10961	1.26
English WordNet Addition	Nouns	4751	14188	2,99	2524	5,62
	Verbs	11363	25761	2,27	14726	1,75
	Other	247	639	2,59	70	9,13
	Total	16361	40588	2,48	17320	2,34
WordNet1.5	Nouns	60521	107428	1,78	88175	1,22
	Verbs	11363	25768	2,27	14734	1,75
	Other	22631	54406	2,4	23708	2,29
	Total	94515	187602	1,98	126617	1,48
Balkanet						
WordNet2.0	Nouns	79689	141691	1,78	115775	1,22
	Verbs	13508	24632	1,82	11306	2,18
	Adjectives	18563	31016	1,67	21495	1,44
	Other	3664	5808	1,59	4660	1,25
	Total	115424	203147	1,76	153236	1,33
	Gloss	115431	115427	1,00	114419	1,01
Greek WordNet	Nouns	14185	18191	1,28	14531	1,25
	Verbs	3055	4658	1,52	2697	1,73
	Adjectives	490	673	1,37	558	1,21
	Other	14	21	1,50	21	1,00
	Total	17744	23543	1,33	17807	1,32
	Gloss	17713	17721	1,00	17614	1,01
Czech WordNet	Nouns	20604	30825	1,50	23485	1,31
	Verbs	4967	8764	1,76	5990	1,46
	Adjectives	2128	3015	1,42	1903	1,58
	Other	164	255	1,55	213	1,20
	Total	27863	42859	1,54	31591	1,36
	Gloss	864	864	1,00	844	1,02

Romanian	Nouns	10419	18243	1,75	12160	1,50
	Verbs	2878	6813	2,37	3119	2,18
	Adjectives	832	1228	1,48	987	1,24
	Other	200	378	1,89	258	1,47
WordNet	Total	14329	26662	1,86	16524	1,61
	Gloss	14554	14554	1,00	14201	1,02
Serbian	Nouns	4584	7355	1,60	6210	1,18
	Verbs	1494	2973	1,99	2199	1,35
	Adjectives	228	294	1,29	263	1,12
	Other	7	10	1,43	10	1,00
WordNet	Total	6313	10632	1,68	8682	1,22
	Gloss	6134	6134	1,00	6120	1,00
Turkish	Nouns	7729	11498	1,49	9125	1,26
	Verbs	2202	3511	1,59	2088	1,68
	Adjectives	340	580	1,71	448	1,29
	Other	0	0	/	0	/
WordNet	Total	10271	15589	1,52	11661	1,34
	Gloss	4507	4507	1,00	4324	1,04
Bulgarian	Nouns	12226	21898	1,79	17946	1,22
	Verbs	3403	6901	2,03	4304	1,60
	Adjectives	1656	2357	1,42	2002	1,18
	Other	4	6	1,50	6	1,00
WordNet	Total	17289	31162	1,80	24258	1,28
	Gloss	17292	17292	1,00	17269	1,00

PART IV
Viewing and editing the multilingual database with Lexidiom

General presentation

Lexidiom is a tool which enables to browse and edit the data of the Integral Dictionnary. Lexidiom is a program developed in Java thus making it independent of the system platform (Windows, Linux etc) it is used. This program, through a jdbc driver, can retrieve the data of the multilingual database, display them and edit them. Since Lexidiom is designed to be used by several users, the first step is to log in to enter the user name and the password. For each user, it would be given certain rights to create, delete, modify or view the data.

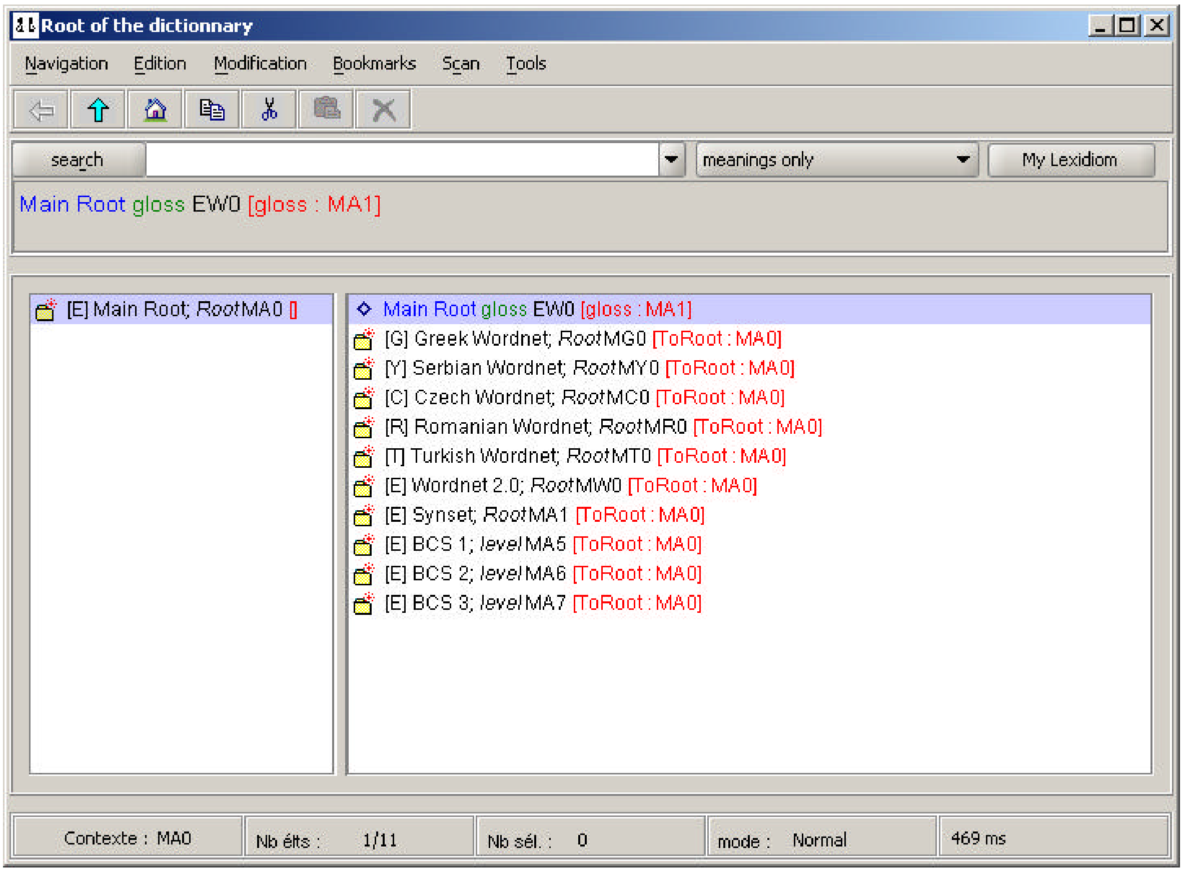


Bishop3
Costa3
DEFAUT
Leoffrey2
Onisep3
samy

User samy

Password

Once logged, the user discovers the following screen :



Main Root gloss EW0 [gloss : MA1]

[E] Main Root; *Root* MA0 []

◇ Main Root gloss EW0 [gloss : MA1]
[G] Greek Wordnet; *Root* MG0 [ToRoot : MA0]
[Y] Serbian Wordnet; *Root* MY0 [ToRoot : MA0]
[C] Czech Wordnet; *Root* MC0 [ToRoot : MA0]
[R] Romanian Wordnet; *Root* MR0 [ToRoot : MA0]
[T] Turkish Wordnet; *Root* MT0 [ToRoot : MA0]
[E] Wordnet 2.0; *Root* MW0 [ToRoot : MA0]
[E] Synset; *Root* MA1 [ToRoot : MA0]
[E] BCS 1; *level* MA5 [ToRoot : MA0]
[E] BCS 2; *level* MA6 [ToRoot : MA0]
[E] BCS 3; *level* MA7 [ToRoot : MA0]

Contexte : MA0 | Nb élts : 1/11 | Nb sél. : 0 | mode : Normal | 469 ms

The left window displays here the root of the graph and the right window the direct children of this root. Let's explain the different symbols of this screen.

🗀 : means that this node is a concept. This notion is not used in WordNet but is used in TID.

◇ means that this node is a literal or a gloss. In this case the node EW0 is the gloss of the node MA0 (Main Root). This node EW0 is the child of the node MA0 and gives this way its wording.

✳ means that this node is an ILI.

[E] gives the language of the literal, the gloss or the wording of a concept or an ILI. (here, Romanian Wordnet should have been translated in Romanian, though).

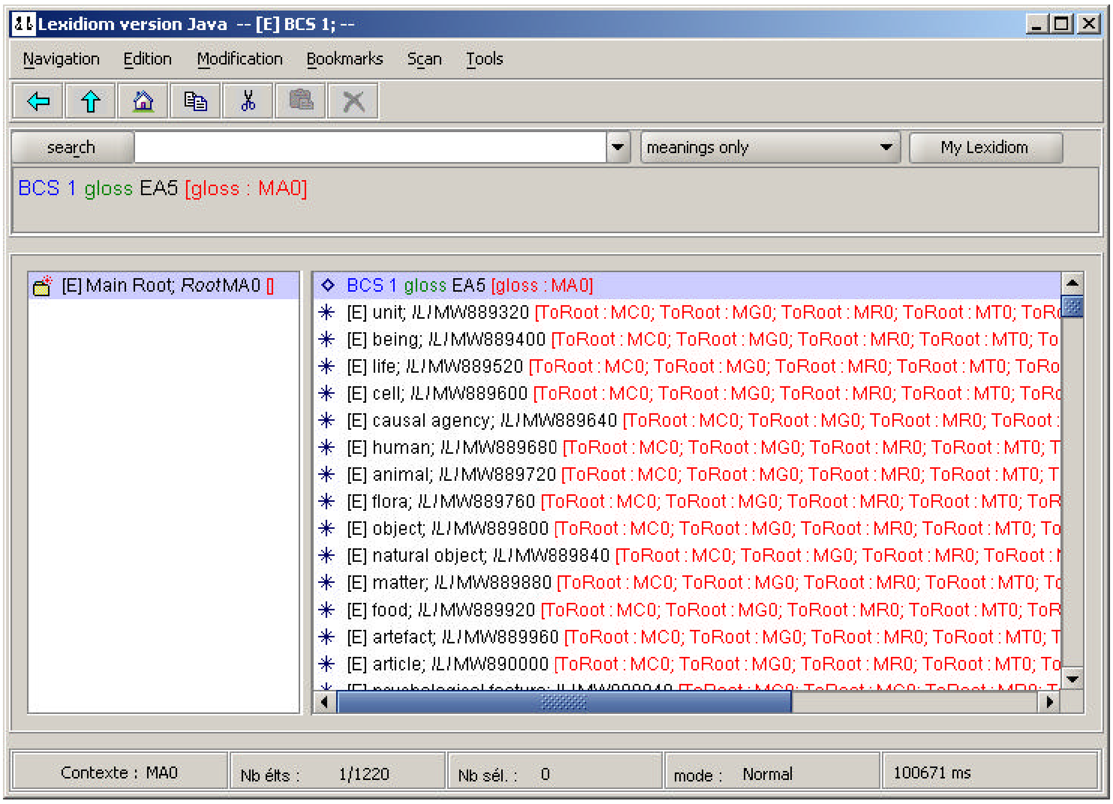
gloss : gives the grammar code of the node.

Root, *level* : these informations give the type of a concept.

[ToRoot : MA0] : type of the relation (ToRoot) and the context of this relation (MA0).

Data browsing :

Let's consider the concept BCS1 (MA5). This concept contains all the synsets marked BCS1 in the WordNet XML file. To know all the synsets contained in the node BCS1, we have just to get all the children of this node. This is done by double-clicking on it. We then have the following screen :



BCS 1 gloss EA5 [gloss : MA0]

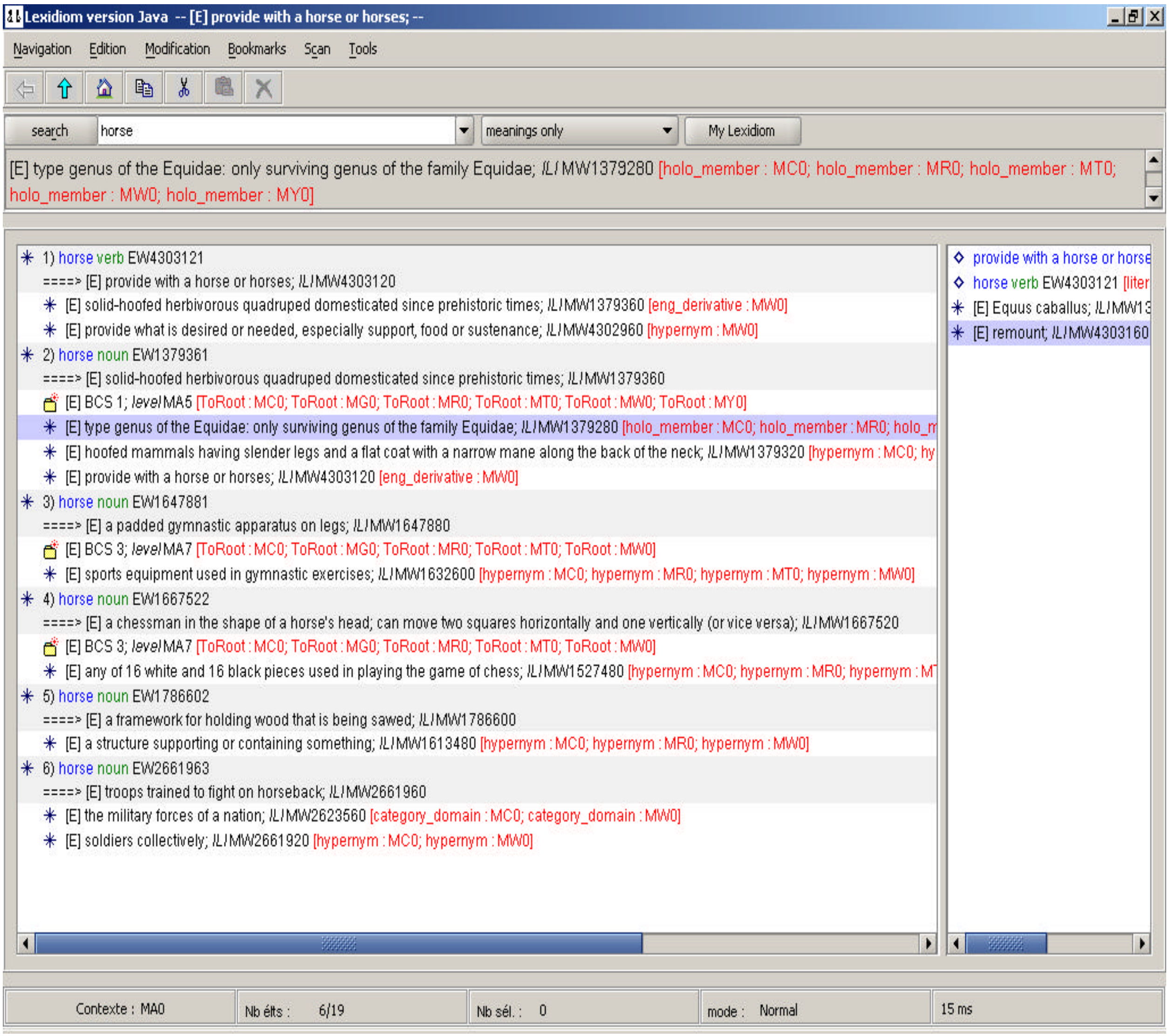
[E] Main Root; RootMA0 []

```
BCS 1 gloss EA5 [gloss : MA0]
[E] unit; ILI MW889320 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; ToRo
[E] being; ILI MW889400 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; To
[E] life; ILI MW889520 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; ToRo
[E] cell; ILI MW889600 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; ToRo
[E] causal agency; ILI MW889640 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot :
[E] human; ILI MW889680 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; T
[E] animal; ILI MW889720 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; T
[E] flora; ILI MW889760 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; ToR
[E] object; ILI MW889800 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; To
[E] natural object; ILI MW889840 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : M
[E] matter; ILI MW889880 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; To
[E] food; ILI MW889920 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; ToR
[E] artefact; ILI MW889960 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; T
[E] article; ILI MW890000 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; To
[E] psychological feature; ILI MW890040 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; T
```

Contexte : MA0 | Nb élts : 1/1220 | Nb sél. : 0 | mode : Normal | 100671 ms

We discover 1219 ILIs belonging to the BCS1 concept, plus the gloss of this concept.

Of course it's possible to search directly for a word. Let's search, for example, the word "horse".



search: horse
meanings only

```
[E] type genus of the Equidae: only surviving genus of the family Equidae; ILI MW1379280 [holo_member : MC0; holo_member : MR0; holo_member : MT0;
holo_member : MW0; holo_member : MY0]
```

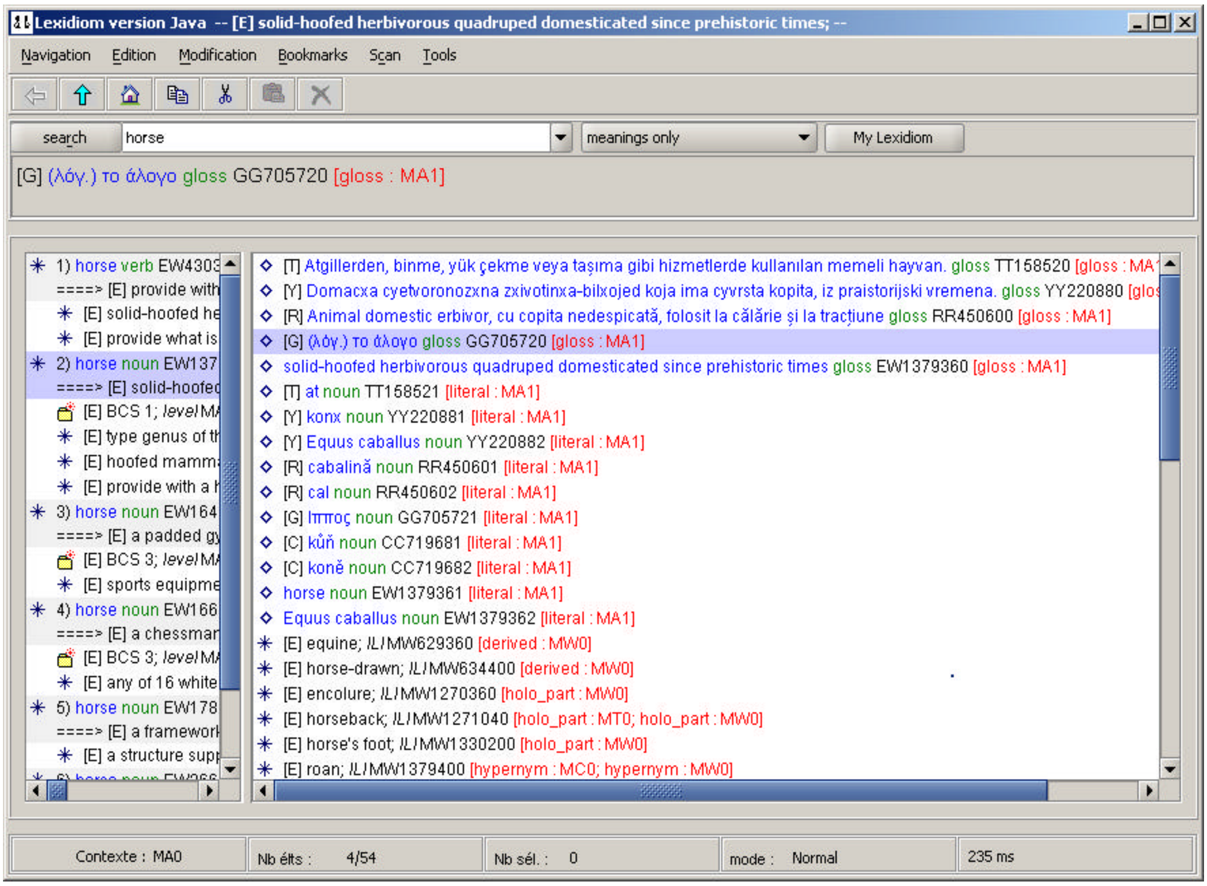
```
* 1) horse verb EW4303121
  ====> [E] provide with a horse or horses; ILI MW4303120
    * [E] solid-hoofed herbivorous quadruped domesticated since prehistoric times; ILI MW1379360 [eng_derivative : MW0]
    * [E] provide what is desired or needed, especially support, food or sustenance; ILI MW4302960 [hypernym : MW0]
* 2) horse noun EW1379361
  ====> [E] solid-hoofed herbivorous quadruped domesticated since prehistoric times; ILI MW1379360
    [E] BCS 1; level MA5 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; ToRoot : MW0; ToRoot : MY0]
    * [E] type genus of the Equidae: only surviving genus of the family Equidae; ILI MW1379280 [holo_member : MC0; holo_member : MR0; holo_m
    * [E] hoofed mammals having slender legs and a flat coat with a narrow mane along the back of the neck; ILI MW1379320 [hypernym : MC0; hy
    * [E] provide with a horse or horses; ILI MW4303120 [eng_derivative : MW0]
* 3) horse noun EW1647881
  ====> [E] a padded gymnastic apparatus on legs; ILI MW1647880
    [E] BCS 3; level MA7 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; ToRoot : MW0]
    * [E] sports equipment used in gymnastic exercises; ILI MW1632600 [hypernym : MC0; hypernym : MR0; hypernym : MT0; hypernym : MW0]
* 4) horse noun EW1667522
  ====> [E] a chessman in the shape of a horse's head; can move two squares horizontally and one vertically (or vice versa); ILI MW1667520
    [E] BCS 3; level MA7 [ToRoot : MC0; ToRoot : MG0; ToRoot : MR0; ToRoot : MT0; ToRoot : MW0]
    * [E] any of 16 white and 16 black pieces used in playing the game of chess; ILI MW1527480 [hypernym : MC0; hypernym : MR0; hypernym : M
* 5) horse noun EW1786602
  ====> [E] a framework for holding wood that is being sawed; ILI MW1786600
    * [E] a structure supporting or containing something; ILI MW1613480 [hypernym : MC0; hypernym : MR0; hypernym : MW0]
* 6) horse noun EW2661963
  ====> [E] troops trained to fight on horseback; ILI MW2661960
    * [E] the military forces of a nation; ILI MW2623560 [category_domain : MC0; category_domain : MW0]
    * [E] soldiers collectively; ILI MW2661920 [hypernym : MC0; hypernym : MW0]
```

```
◇ provide with a horse or horse
◇ horse verb EW4303121 [liter
* [E] Equus caballus; ILI MW13
* [E] remount; ILI MW4303160
```

```
Contexte : MA0    Nb élts :  6/19    Nb sél. :  0    mode :  Normal    15 ms
```

We got 6 meanings for "horse", five nouns and one verb. We will consider the meaning number 2 (EW1379361) which belongs to the synset "solid-hoofed herbivorious..." (the main meaning of "horse", in fact).
Below, we can see that this synset belongs to the BCS1 set, and to the other synsets MW1379280 (with the type holo_member), MW1379320 (whith the type hypernym) and the synset MW4303120 (eng_derivative).
In the last case, we can see that this relation exists only in Wordnet 2.0 and, thus, is not present in the other monolingual WordNets.

If we want to discover the content of the synset MW1379360 ("solid-hoofed herbivorious..."), we double-click on it to obtain :


Lexidiom version Java -- [E] solid-hoofed herbivorous quadruped domesticated since prehistoric times; --
Navigation
Edition
Modification
Bookmarks
Scan
Tools
search
My Lexidiom

search: horse

meanings only

[G] (λόγ.) το άλογο gloss GG705720 [gloss : MA1]

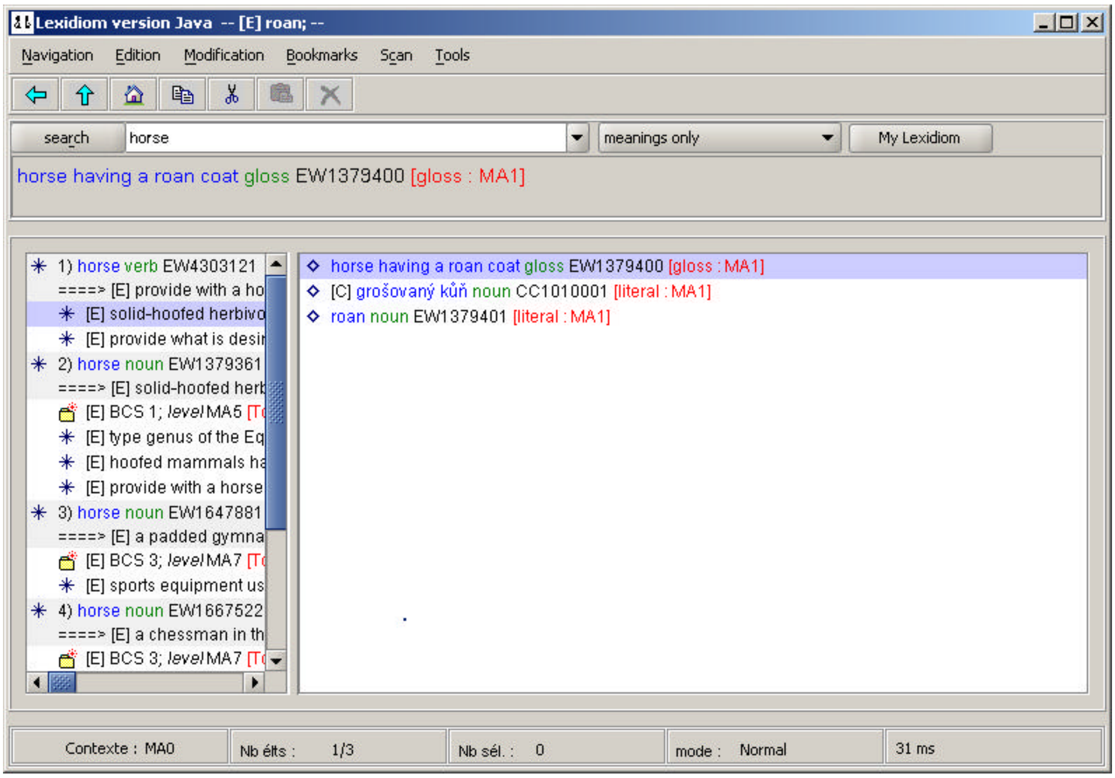
1) horse verb EW4303
====> [E] provide with
[E] solid-hoofed he
[E] provide what is
2) horse noun EW137
====> [E] solid-hoofed
[E] BCS 1; *level* MA
[E] type genus of th
[E] hoofed mamm
[E] provide with a h
3) horse noun EW164
====> [E] a padded gy
[E] BCS 3; *level* MA
[E] sports equipme
4) horse noun EW166
====> [E] a chessman
[E] BCS 3; *level* MA
[E] any of 16 white
5) horse noun EW178
====> [E] a framework
[E] a structure supp

◇ [T] Atgillerden, binme, yük çekme veya taşıma gibi hizmetlerde kullanılan memeli hayvan. gloss TT158520 [gloss : MA1
◇ [Y] Domacxa cyetvoronozxna zxivotinxa-bilxojed koja ima cyvrsta kopita, iz praistorijski vremena. gloss YY220880 [glos
◇ [R] Animal domestic erbivor, cu copita nedespicată, folosit la călărie și la tracțiune gloss RR450600 [gloss : MA1]
◇ [G] (λόγ.) το άλογο gloss GG705720 [gloss : MA1]
◇ solid-hoofed herbivorous quadruped domesticated since prehistoric times gloss EW1379360 [gloss : MA1]
◇ [T] at noun TT158521 [literal : MA1]
◇ [Y] konx noun YY220881 [literal : MA1]
◇ [Y] Equus caballus noun YY220882 [literal : MA1]
◇ [R] cabalină noun RR450601 [literal : MA1]
◇ [R] cal noun RR450602 [literal : MA1]
◇ [G] Ιππος noun GG705721 [literal : MA1]
◇ [C] kůň noun CC719681 [literal : MA1]
◇ [C] koně noun CC719682 [literal : MA1]
◇ horse noun EW1379361 [literal : MA1]
◇ Equus caballus noun EW1379362 [literal : MA1]
✳ [E] equine; *ILI* MW629360 [derived : MW0]
✳ [E] horse-drawn; *ILI* MW634400 [derived : MW0]
✳ [E] encolure; *ILI* MW1270360 [holo_part : MW0]
✳ [E] horseback; *ILI* MW1271040 [holo_part : MT0; holo_part : MW0]
✳ [E] horse's foot; *ILI* MW1330200 [holo_part : MW0]
✳ [E] roan; *ILI* MW1379400 [hypernym : MC0; hypernym : MW0]

Contexte : MA0 | Nb élts : 4/54 | Nb sél. : 0 | mode : Normal | 235 ms

The content of this synset are the glosses and the literals in the different languages, and the other synsets it is in relation with (in this case, the literals are displayed instead of the glosses). To see the content of one of these synsets, we can continue to double-click on it.

As explained above, we can see that these relations with the other synsets (ILRs) are dependant of the context. For instance, the link with the ILI MW1379400 (with the literal "roan") has been noticed in the Czech Wornet (MC0) and in the Princeton WordNet 2.0 (MW0) but not in the other monolingual WordNet.
In general that means that there are no literal of these languages in this synset. That's the case with the synset MW1379400 ("roan"), as we can see looking for its content :



horse

meanings only

horse having a roan coat gloss EW1379400 [gloss : MA1]

- 1) horse verb EW4303121
  - ====> [E] provide with a ho
  - [E] solid-hoofed herbivo
  - [E] provide what is desi
- 2) horse noun EW1379361
  - ====> [E] solid-hoofed herb
  - [E] BCS 1; *level* MA5 [T
  - [E] type genus of the Eq
  - [E] hoofed mammals ha
  - [E] provide with a horse
- 3) horse noun EW1647881
  - ====> [E] a padded gymna
  - [E] BCS 3; *level* MA7 [T
  - [E] sports equipment us
- 4) horse noun EW1667522
  - ====> [E] a chessman in th
  - [E] BCS 3; *level* MA7 [T

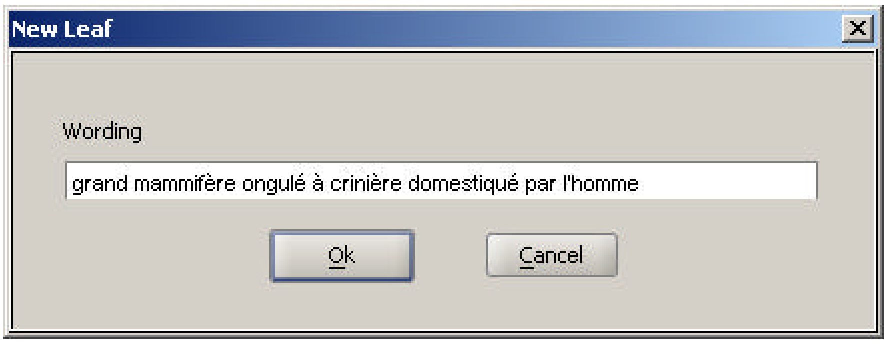
◇ horse having a roan coat gloss EW1379400 [gloss : MA1]
◇ [C] grošovaný kůň noun CC1010001 [literal : MA1]
◇ roan noun EW1379401 [literal : MA1]

Contexte : MA0 | Nb élts : 1/3 | Nb sél. : 0 | mode : Normal | 31 ms

note : in this synset there is a Czech literal but not a Czech gloss.

Data editing

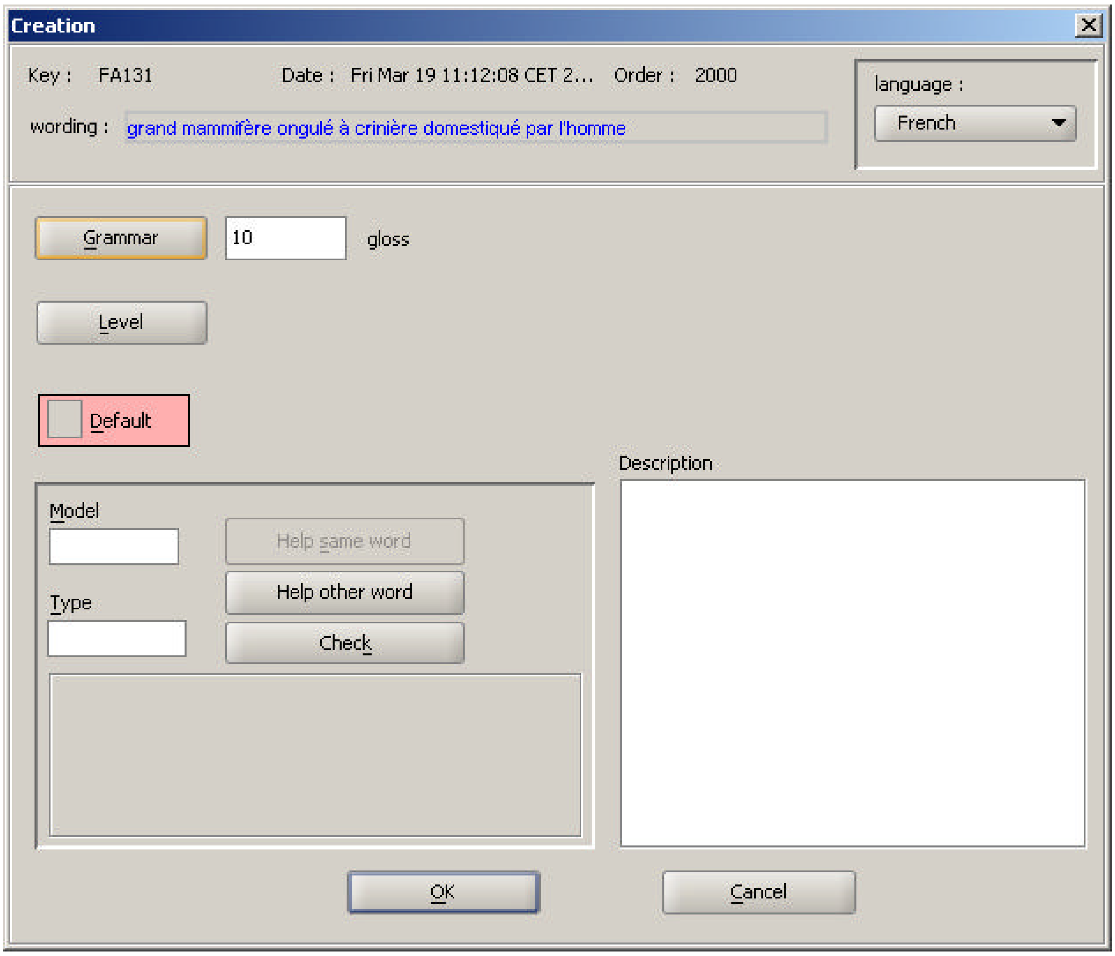
More than just browsing the data, Lexidiom allows to modify them.
For example, if we want to add a French gloss for the synset above, we launch the command "Create leaf" to obtain the following dialog box :



Wording

grand mammifère ongulé à crinière domestiqué par l'homme

We have then a second dialog box in which we choose the language (French) and the grammar code (10 : gloss)



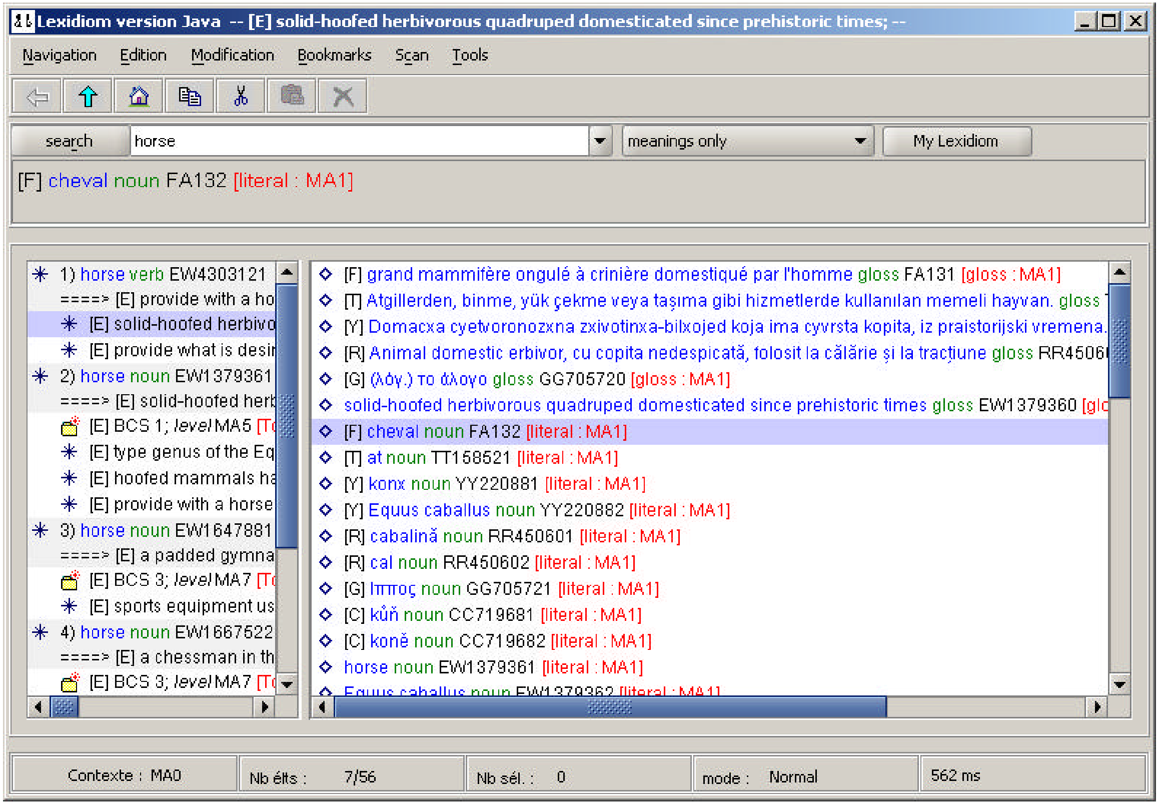
Key : FA131 Date : Fri Mar 19 11:12:08 CET 2... Order : 2000

wording : grand mammifère ongulé à crinière domestiqué par l'homme

language : French

Grammar 10 gloss

In the same way we can add the literal "cheval" to this synset. These two new nodes (FA131 and FA132) appear then in the synset :


Lexidiom version Java -- [E] solid-hoofed herbivorous quadruped domesticated since prehistoric times; --
Navigation
Edition
Modification
Bookmarks
Scan
Tools
search
My Lexidiom

search: horse — meanings only

[F] cheval noun FA132 [literal : MA1]

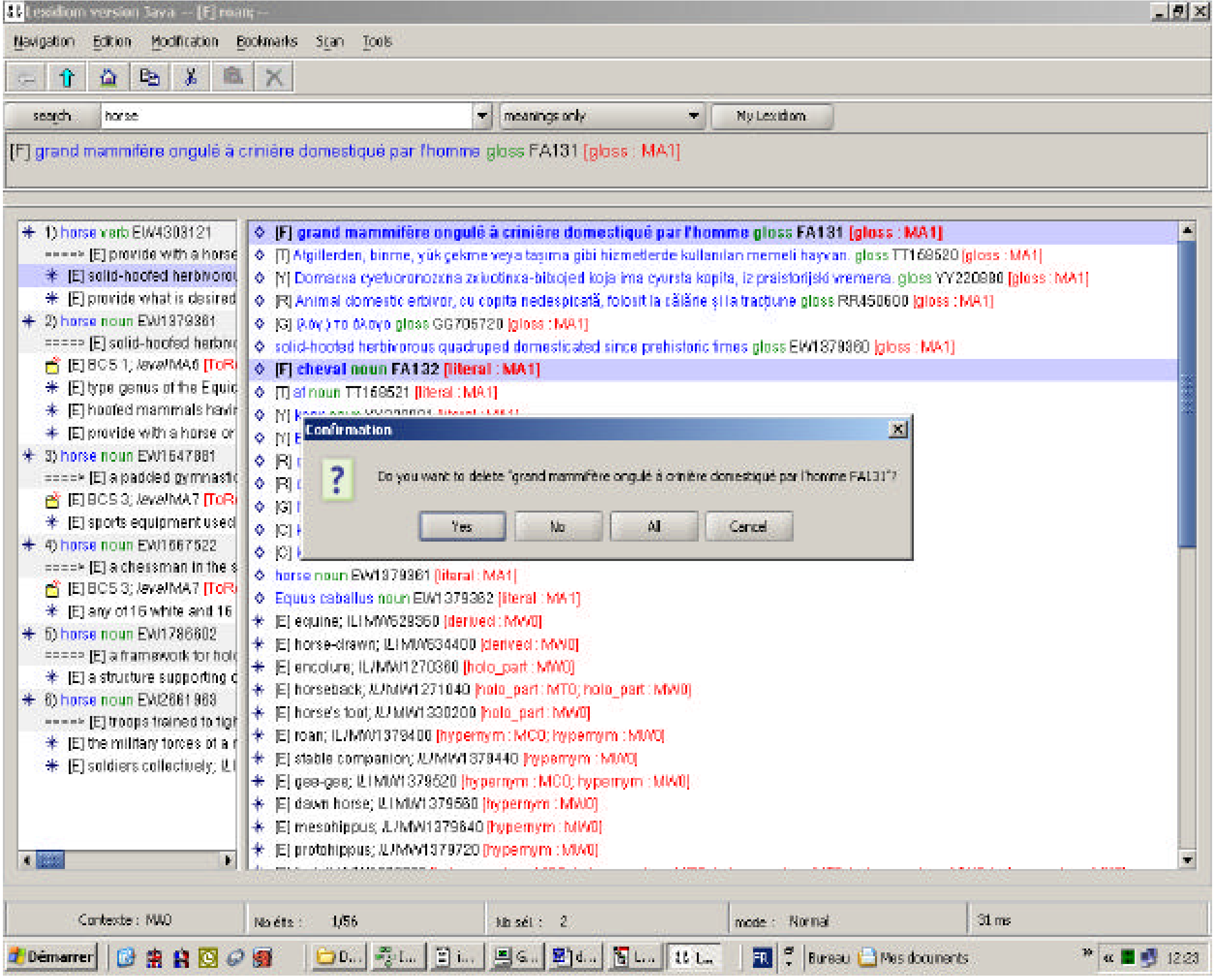
1) horse verb EW4303121
====> [E] provide with a ho
[E] solid-hoofed herbivo
[E] provide what is desi
2) horse noun EW1379361
====> [E] solid-hoofed herb
[E] BCS 1; *level* MA5 [T
[E] type genus of the Eq
[E] hoofed mammals ha
[E] provide with a horse
3) horse noun EW1647881
====> [E] a padded gymna
[E] BCS 3; *level* MA7 [T
[E] sports equipment us
4) horse noun EW1667522
====> [E] a chessman in th
[E] BCS 3; *level* MA7 [T

[F] grand mammifère ongulé à crinière domestiqué par l'homme gloss FA131 [gloss : MA1]
[T] Atgillerden, binme, yük çekme veya taşıma gibi hizmetlerde kullanılan memeli hayvan. gloss T
[Y] Domacxa cyetvoronozxna zxivotinxa-bilxojed koja ima cyvrsta kopita, iz praistorijski vremena.
[R] Animal domestic erbivor, cu copita nedespicată, folosit la călărie și la tracțiune gloss RR4506
[G] (λόγ.) το άλογο gloss GG705720 [gloss : MA1]
solid-hoofed herbivorous quadruped domesticated since prehistoric times gloss EW1379360 [glo
[F] cheval noun FA132 [literal : MA1]
[T] at noun TT158521 [literal : MA1]
[Y] konx noun YY220881 [literal : MA1]
[Y] Equus caballus noun YY220882 [literal : MA1]
[R] cabalină noun RR450601 [literal : MA1]
[R] cal noun RR450602 [literal : MA1]
[G] Ιπποç noun GG705721 [literal : MA1]
[C] kůň noun CC719681 [literal : MA1]
[C] koně noun CC719682 [literal : MA1]
horse noun EW1379361 [literal : MA1]
Equus caballus noun EW1379362 [literal : MA1]

Contexte : MA0 Nb élts : 7/56 Nb sél. : 0 mode : Normal 562 ms

It is of course also possible to modify the different nodes and to suppress them.

Example of the suppression of the two French nodes.



[F] grand mammifère ongulé à crinière domestiqué par l'homme gloss FA131 [gloss : MA1]

1) horse verb EW4303121
=====> [E] provide with a horse
[E] solid-hoofed herbivorou
[E] provide what is desired
2) horse noun EW1379381
=====> [E] solid-hoofed herbivo
[E] BCS 1; levelMA6 [ToR
[E] type genus of the Equic
[E] hoofed mammals havir
[E] provide with a horse or
3) horse noun EW1547881
=====> [E] a padded gymnastic
[E] BCS 3; levelMA7 [ToR
[E] sports equipment used
4) horse noun EW1667522
=====> [E] a chessman in the s
[E] BCS 3; levelMA7 [ToR
[E] any of 16 white and 16
5) horse noun EW1786602
=====> [E] a framework for hold
[E] a structure supporting c
6) horse noun EW2861983
=====> [E] troops trained to figh
[E] the military forces of a r
[E] soldiers collectively; ILI

[F] grand mammifère ongulé à crinière domestiqué par l'homme gloss FA131 [gloss : MA1]
[T] Atgillerden, binme, yük çekme veya taşıma gibi hizmetlerde kullanılan memeli hayvan. gloss TT168520 [gloss : MA1]
[Y] Domaćka cyetuoronožna zxivotinxa-biljojed koja ima cyursta kopita, iz praistorijski vremena. gloss YY220880 [gloss : MA1]
[R] Animal domestic erbivor, cu copita nedespicată, folosit la călărie şi la tracţiune gloss RR450600 [gloss : MA1]
[G] (Άλογ) το άλογο gloss GG705720 [gloss : MA1]
solid-hoofed herbivorous quadruped domesticated since prehistoric times gloss EW1379380 [gloss : MA1]
[F] cheval noun FA132 [literal : MA1]
[T] at noun TT169521 [literal : MA1]
[Y] k
[Y] E
[R]
[R]
[G]
[C]
[C]
horse noun EW1379381 [literal : MA1]
Equus caballus noun EW1379382 [literal : MA1]
[E] equine; ILI MW629360 [derived : MW0]
[E] horse-drawn; ILI MW634400 [derived : MW0]
[E] encolure; IL/MW1270360 [holo_part : MW0]
[E] horseback; IL/MW1271040 [holo_part : MT0; holo_part : MW0]
[E] horse's foot; IL/MW1330200 [holo_part : MW0]
[E] roan; IL/MW1379400 [hypernym : MC0; hypernym : MW0]
[E] stable companion; IL/MW1379440 [hypernym : MW0]
[E] gee-gee; ILI MW1379520 [hypernym : MC0; hypernym : MW0]
[E] dawn horse; ILI MW1379580 [hypernym : MW0]
[E] mesohippus; IL/MW1379640 [hypernym : MW0]
[E] protohippus; IL/MW1379720 [hypernym : MW0]

Confirmation

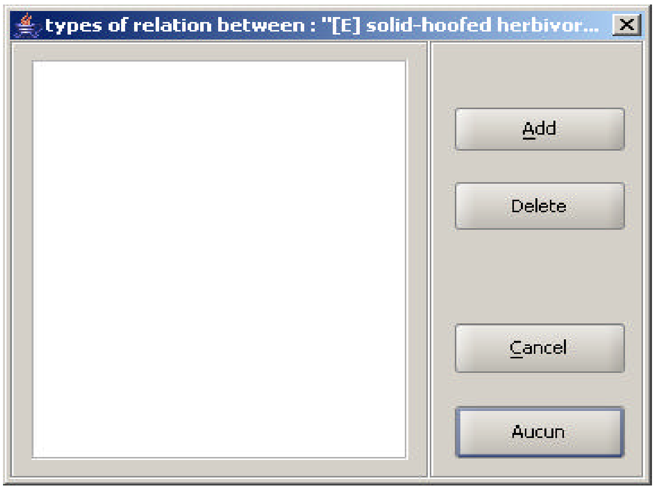
Do you want to delete "grand mammifère ongulé à crinière domestiqué par l'homme FA131"?

Contexte : MAO
Nb élts : 1/56
Nb sél : 2
mode : Normal
31 ms

Modification of a relation :

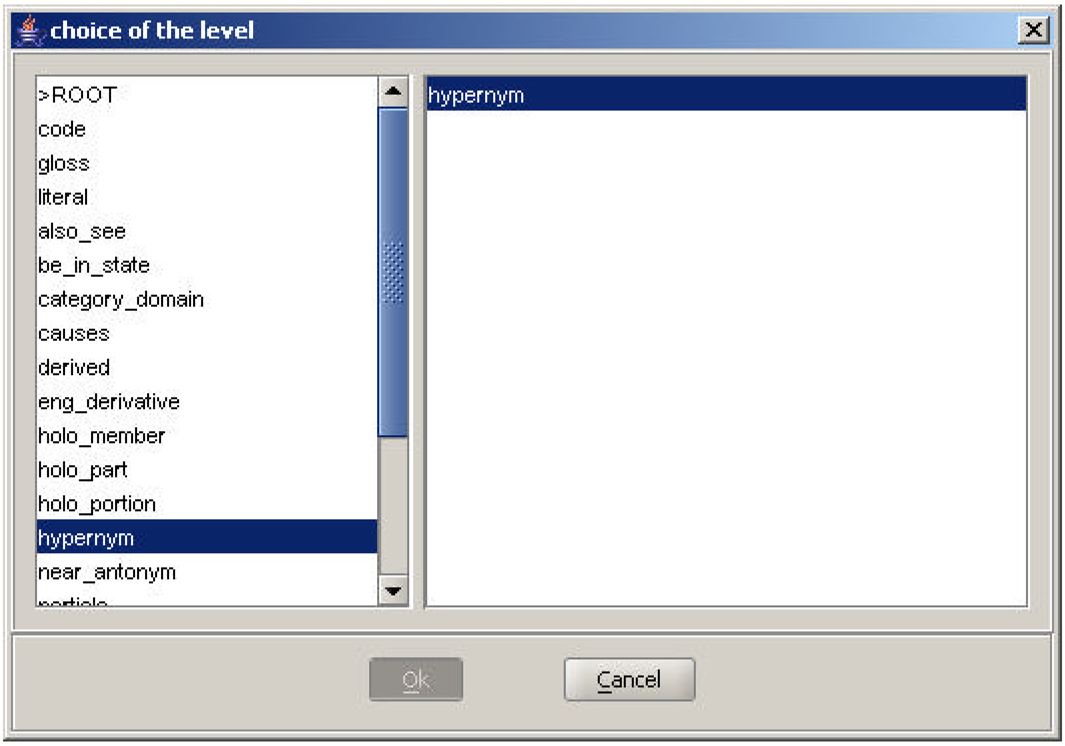
It is possible to add a relation, to modify it or to suppress it. For example, if we are in the context of TID (which will be represented by the key MAO), we can add the hypernym relation for this context.

We have first :



types of relation between : "[E] solid-hoofed herbivor...

We then click on the Add button and choose the type of relation :

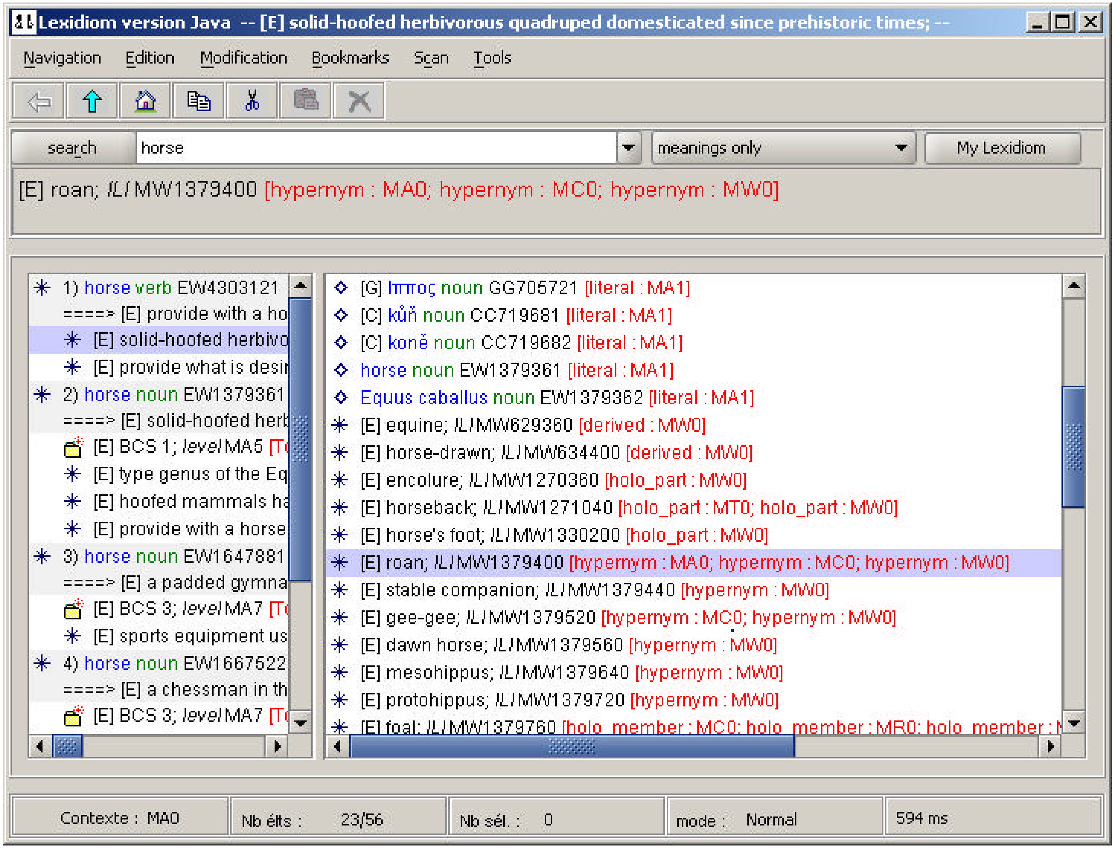


```
>ROOT
code
gloss
literal
also_see
be_in_state
category_domain
causes
derived
eng_derivative
holo_member
holo_part
holo_portion
hypernym
near_antonym
particle
```

```
hypernym
```

Note : it would have been possible here to add several types.

The type of relation is now added for the context MA0 (TID)


Lexidiom version Java -- [E] solid-hoofed herbivorous quadruped domesticated since prehistoric times; --
Navigation
Edition
Modification
Bookmarks
Scan
Tools
search
My Lexidiom

search: horse — meanings only

```
[E] roan; ILI MW1379400 [hypernym : MA0; hypernym : MC0; hypernym : MW0]
```

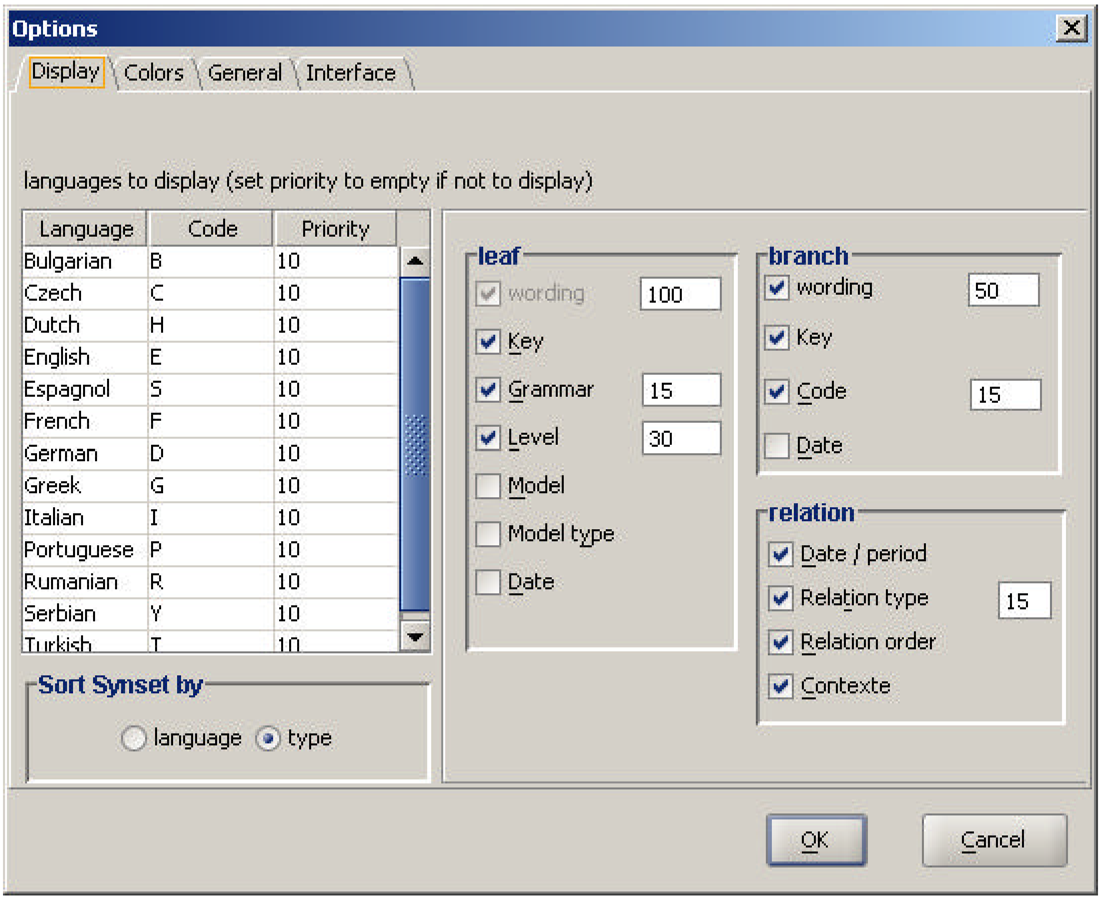
```
1) horse verb EW4303121
  ====> [E] provide with a ho
    [E] solid-hoofed herbivo
    [E] provide what is desi
2) horse noun EW1379361
  ====> [E] solid-hoofed herb
    [E] BCS 1; level MA5 [T
    [E] type genus of the Eq
    [E] hoofed mammals ha
    [E] provide with a horse
3) horse noun EW1647881
  ====> [E] a padded gymna
    [E] BCS 3; level MA7 [T
    [E] sports equipment us
4) horse noun EW1667522
  ====> [E] a chessman in th
    [E] BCS 3; level MA7 [T
```

```
[G] Ιππος noun GG705721 [literal : MA1]
[C] kůň noun CC719681 [literal : MA1]
[C] koně noun CC719682 [literal : MA1]
horse noun EW1379361 [literal : MA1]
Equus caballus noun EW1379362 [literal : MA1]
[E] equine; ILI MW629360 [derived : MW0]
[E] horse-drawn; ILI MW634400 [derived : MW0]
[E] encolure; ILI MW1270360 [holo_part : MW0]
[E] horseback; ILI MW1271040 [holo_part : MT0; holo_part : MW0]
[E] horse's foot; ILI MW1330200 [holo_part : MW0]
[E] roan; ILI MW1379400 [hypernym : MA0; hypernym : MC0; hypernym : MW0]
[E] stable companion; ILI MW1379440 [hypernym : MW0]
[E] gee-gee; ILI MW1379520 [hypernym : MC0; hypernym : MW0]
[E] dawn horse; ILI MW1379560 [hypernym : MW0]
[E] mesohippus; ILI MW1379640 [hypernym : MW0]
[E] protohippus; ILI MW1379720 [hypernym : MW0]
[E] foal; ILI MW1379760 [holo_member : MC0; holo_member : MR0; holo_member : M
```

Contexte : MA0 | Nb élts : 23/56 | Nb sél. : 0 | mode : Normal | 594 ms

Besides all these functions for browsing or editing, Lexidiom can be configured by an options Dialog Box, to choose the languages to display, their order, the characteristics of the nodes, etc.

For example, if we want to see only the Turkish Wordnet, we would just have to empty the Priority field. This way WordNets can be see alone or in combination with another language, using the same tool (and in the same session)



languages to display (set priority to empty if not to display)

Language	Code	Priority
Bulgarian	B	10
Czech	C	10
Dutch	H	10
English	E	10
Espagnol	S	10
French	F	10
German	D	10
Greek	G	10
Italian	I	10
Portuguese	P	10
Rumanian	R	10
Serbian	Y	10
Turkish	T	10

Sort Synset by

○ language ● type

leaf

☑ wording 100

☑ Key

☑ Grammar 15

☑ Level 30

□ Model

□ Model type

□ Date

branch

☑ wording 50

☑ Key

☑ Code 15

□ Date

relation

☑ Date / period

☑ Relation type 15

☑ Relation order

☑ Contexte

Conclusion

In this subtask, we have described the structure of a WordNet XML file. We have seen how to load the data of this file and to mount them in a different architecture, in this case, the Integral Dictionary one.

We have seen that it is really possible, and even easy, to merge those different wordnets using the ILI records of the EuroWordNet project.

The merge with The Integral Dictionary is not done yet, but since we have used the same architecture to build the multilingual Balkanet, we are sure that the two models could merge easily, even if the matching across the common senses could be more difficult.