WordNet Extensions

German Rigau i Claramunt

german.rigau@ehu.es

IXA group

Departamento de Lenguajes y Sistemas Informáticos

UPV/EHU

Outline

- Multilingual Central Repository
  - MEANING
  - KNOW

MEANING

Developing Multilingual Web-scale Language Technologies

IST-2001-34460

http://www.lsi.upc.es/~nlp/meaning/meaning.html

German Rigau i Claramunt

MEANING: Introduction

- From Financial Times
  - US officials has expected Basra to fall early
  - Music sales will fall by up to 15% this year
  - No missiles have fallen and ...

MEANING: Introduction

Sense 10

fall -- (be captured; "The cities fell to the enemy")

=> yield -- (cease opposition; stop fighting)

Sense 2

descend, fall, go down, come down -- (move downward but not necessarily all the way; "The temperature is going down"; "The barometer is falling"; "Real estate prices are coming down")

=> travel, go, move, locomote -- (change location; …)

Sense 1

fall -- (descend in free fall under the influence of gravity; "The branch fell from the tree"; "The unfortunate hiker fell into a crevasse")

=> travel, go, move, locomote -- (change location; …)

MEANING: Introduction

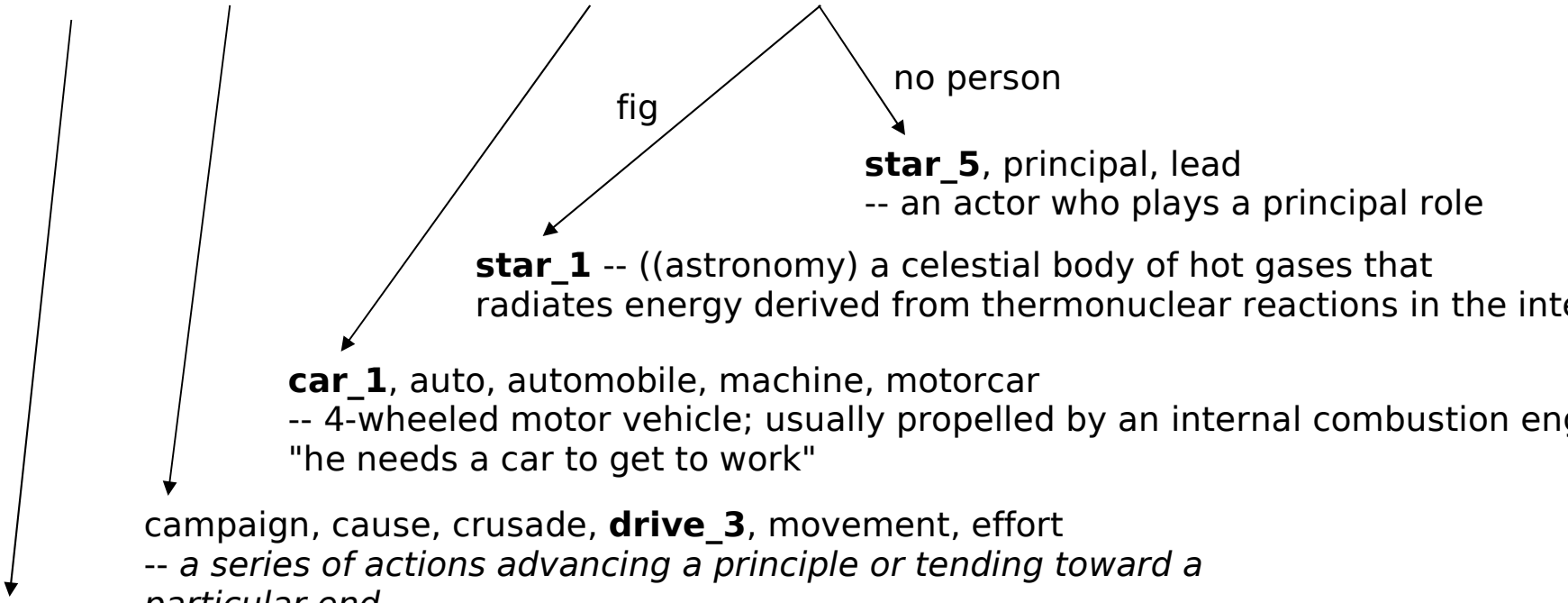
- From Financial Times

  GM’s drive to make Saturn a star again

MEANING: Introduction

- From Financial Times

GM’s drive to make Saturn a star again



MEANING: Introduction

- From NLP to NLU
- Large-scale Semantic Processing dealing with concepts (senses) rather than words
- Two complementary OPEN problems:

- Acquisition bottleneck
  - Autonomous large-scale knowledge acquisition systems
- Ambiguity bottleneck
  - Highly accurate WSD systems

MEANING: Introduction

Dealing with the ACQ/WSD deadlock

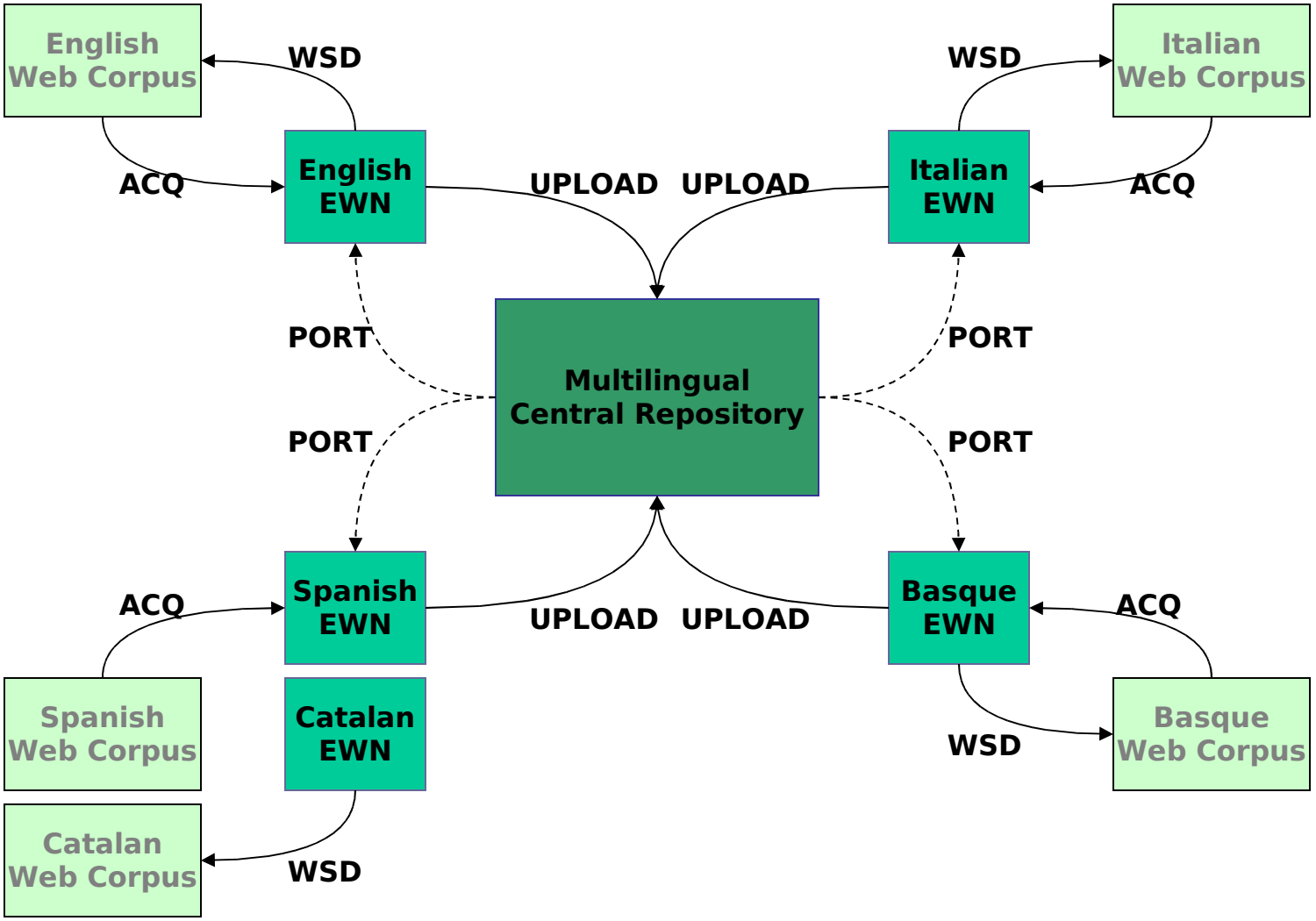
- Dealing with knowledge acquisition
  - Need of texts automatically sense tagged
  - Current state-of-the-art 60%-70% accuracy!
- Dealing with concepts
  - Need of knowledge not currently available:
    - Subcategorization frequencies for predicates
    - Selectional Preferences, etc.
- Dealing with multilingualism
  - Need of compatibility across resources

MEANING: Introduction

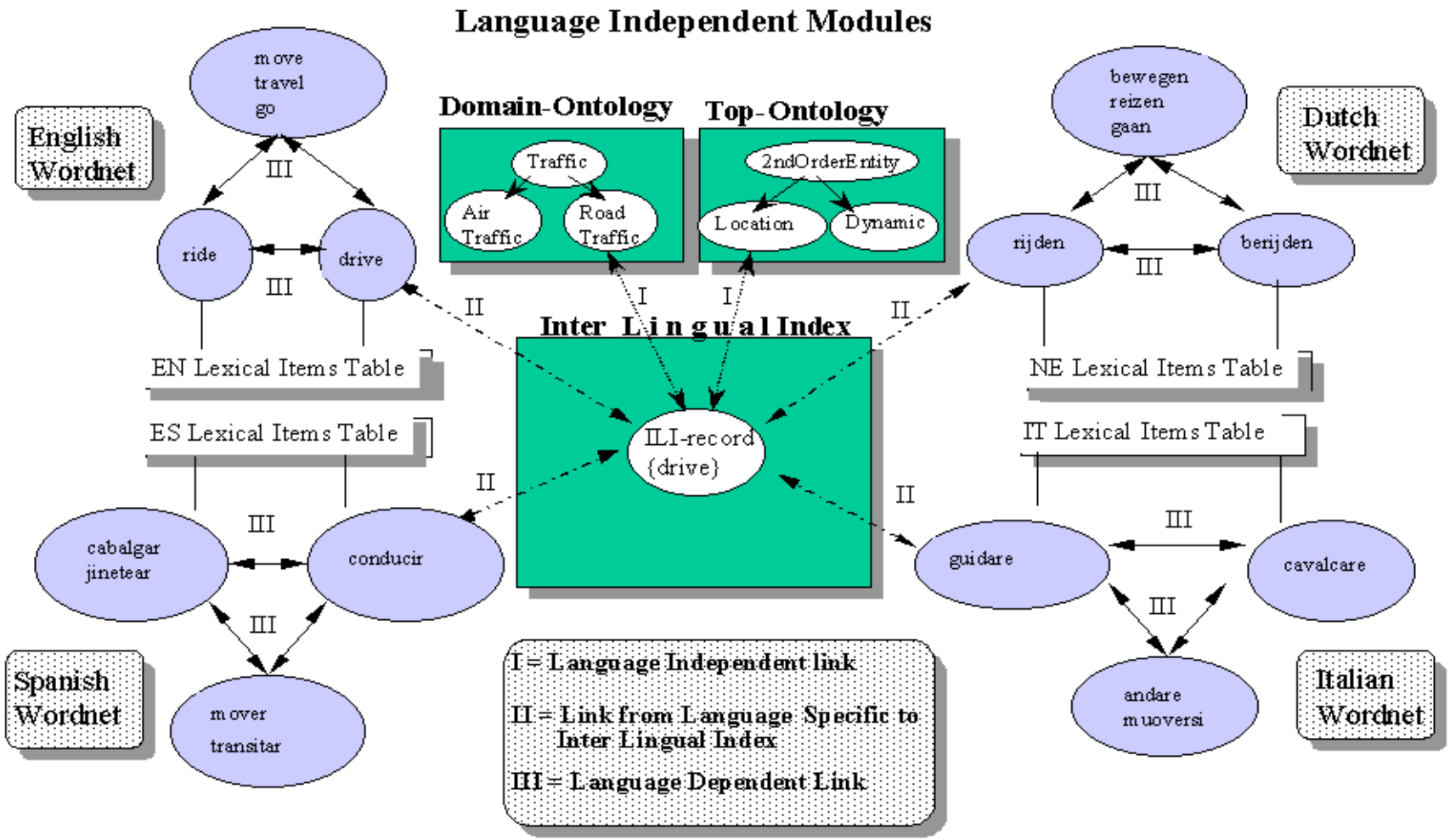
Dealing with the ACQ/WSD deadlock

- Addressing Acquisition and WSD simultaneously
  - three consecutive MEANING cycles
- Language is highly polysemous
  - but also highly redundant
- Multilingualism
  - maybe is part of the solution using EuroWordNet
- Reuse of incompatible large-scale resources
  - Mapping technology to connect already available data

MEANING: Architecture



Architecture of the EuroWordNet Data Structure



EuroWordNet Architecture

- Core
  - Inter-Lingual-Index (ILI)
  - Top Concept Ontology (TCO)
  - Domain Ontology (DO)
- Extensions
  - Local wordnets
  - Domain wordnets

Interlingual Index of EuroWordNet

- Set of synsets from WN 1.5
- Base concepts connected to TCO and DO

Top Concept Ontology of EuroWordNet

- Hierarchy of language independent concepts
  - Semantic distinctions: object, place, …
  - abstract (not lexical)
  - Connected to the ILI
- Three types of concepts:
  - First order: *entities*
  - Second order: *estatic or dinamic situations*
  - Third order: *abstract prepositions*

WordNet & EuroWordNet

Top Concept Ontology of EuroWordNet

Top[0]	
1stOrderEntity[1]	**2ndOrderEntity**[0]
Origin[0] Natural[21] Living[30] Plant[18] Human[106] Creature[2] Anima[123] Artifact[144] **Form**[0] Substance[32] Solid[63] Liquid[13] Gas[1] Object1[62] **Composition**[0] Part[86] Group[63] **Function**[55] Vehicle[8]	**SituationType**[6] Dynamic[134] BoundedEvent[183] UnboundedEvent[48] Static[28] Property[61] Relation[38] **SituationComponent**[0] Cause[67] Agentive[170] Phenomenal[17] Stimulating[25] Communication[50] Condition[62] Existence[27] Experience[43] Location[76] Manner[21] Mental[90]

WordNet & EuroWordNet

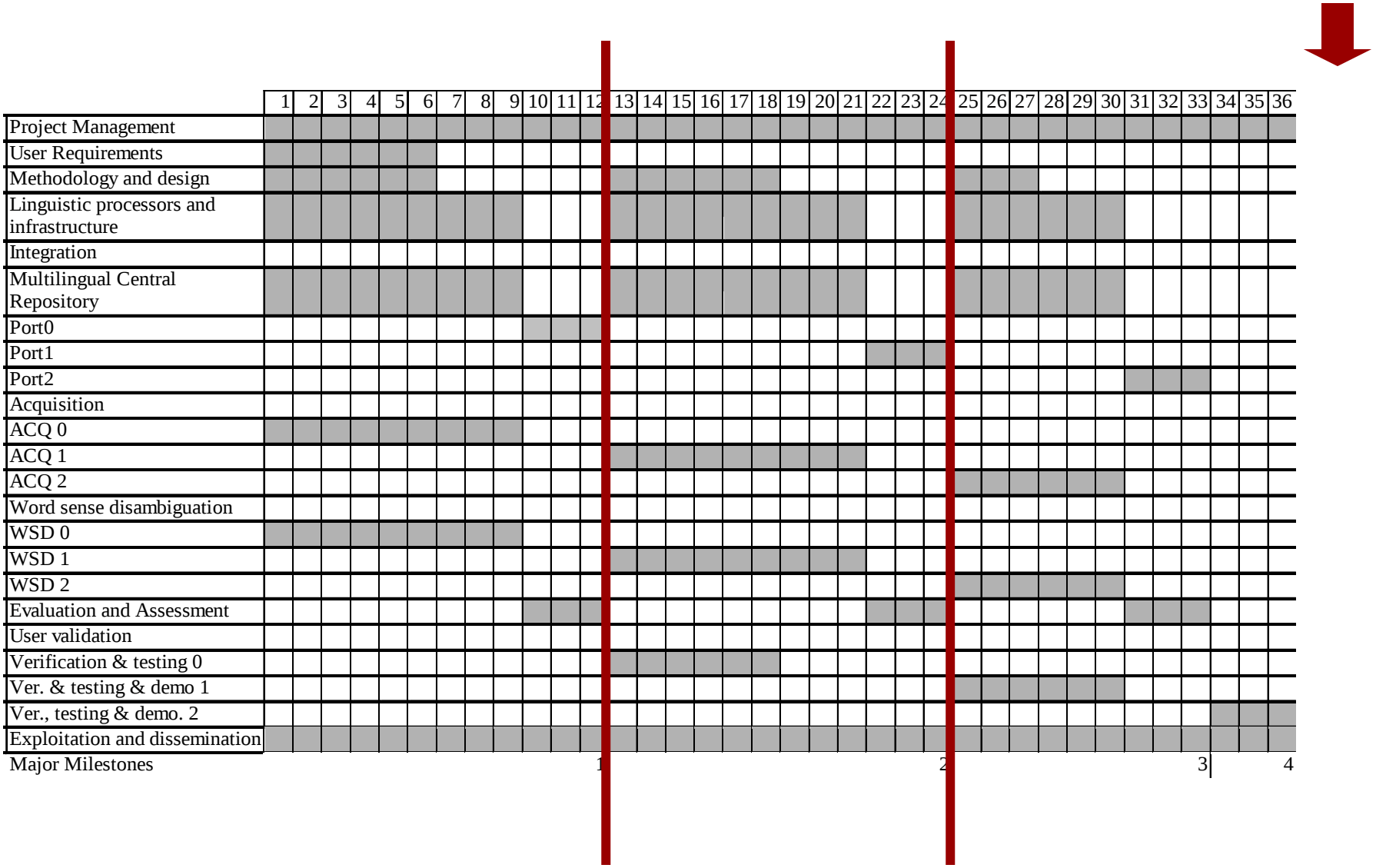
Domain Ontology of EuroWordNet

- Hierarchy of domains
  - Traffic:Road Traffic, Air traffic, etc.
  - Medicine
  - ...
- Domains label different parts of the hierarchies:
  - Medicine: doctor, nurse, operation, etc.
- Domains label different POS categories:
  - Medicine: doctor, to operate, etc.

MEANING: Overview

- 3 years research project (2002-2005)
- 1.610 Million Euro
- Consortium
  - TALP Research Center, UPC
  - ITC-IRST
  - IXA group, UPV/EHU
  - University of Sussex
    - Irion Technologies

MEANING: Workplan



MEANING: Workplan

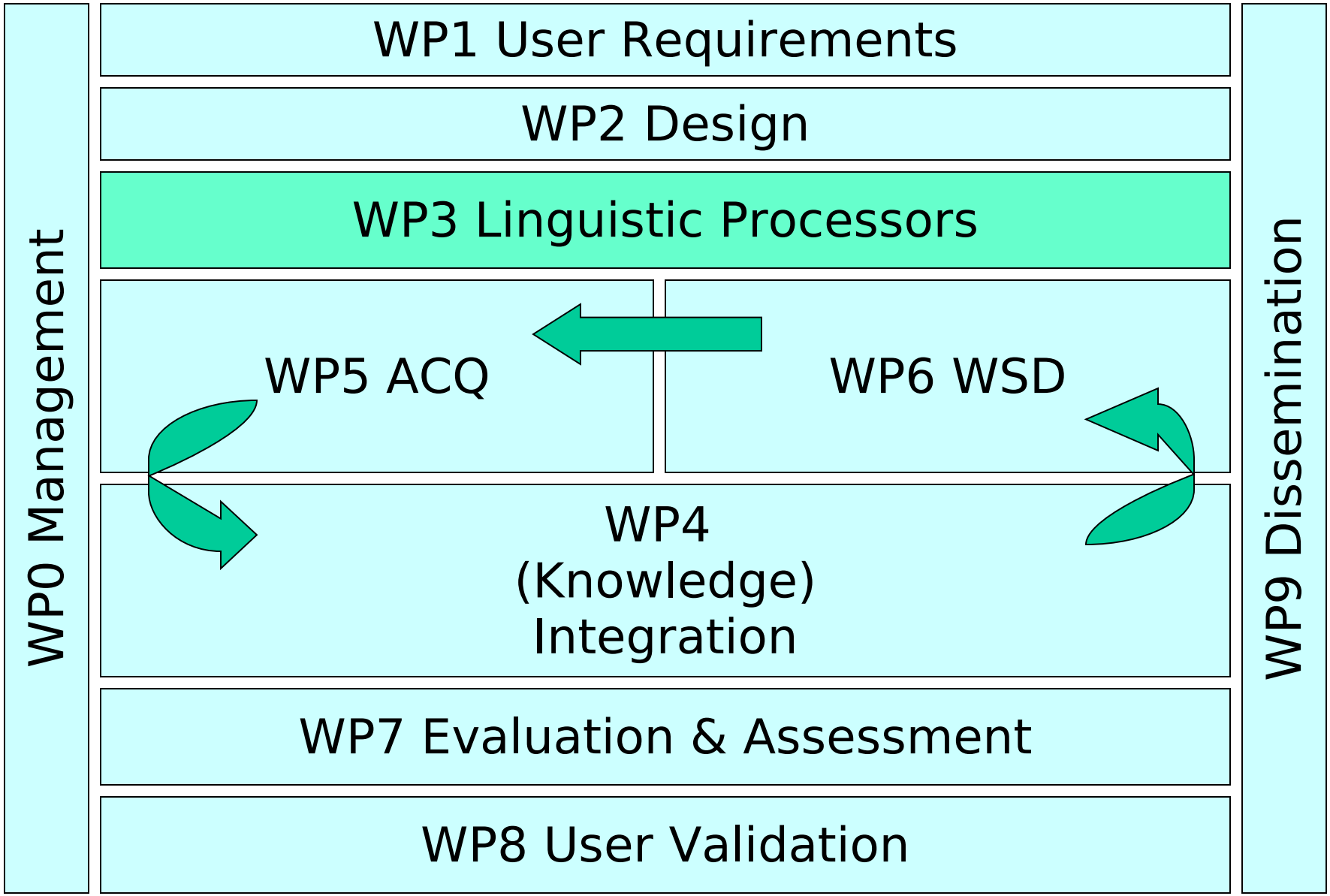
- **WP3** (**Linguistic Processors**)
- Three development cycles:
  - **WP5** (**Acquisition**): (ACQ0, ACQ1, ACQ2)

    Local acquisition of knowledge using specially designed tools and resources, corpus and wordnets
  - **WP4** (**Integration**): (PORT0, PORT1, PORT2)

    Uploading the acquired knowledge from each language into the Multilingual Central Repository and porting to the local wordnets
  - **WP6** (**WSD**): (WSD0, WSD1, WSD2)

    Word Sense Disambiguation using the local wordnets and the enriched knowledge ported from the MCR
- **WP7** (**evaluation and assessment**) of the software tools and resources produced

MEANING: Workplan

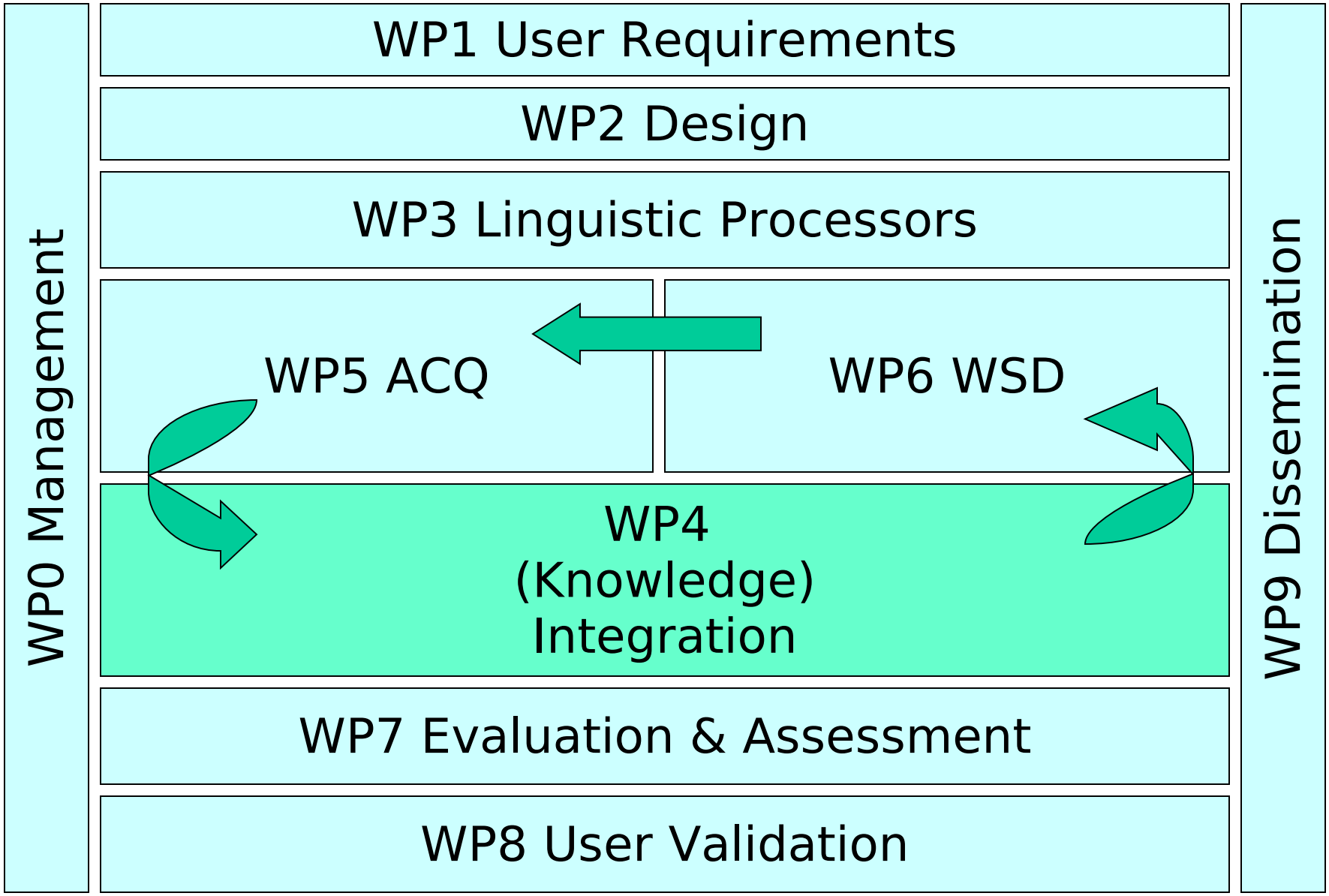


MEANING: WP5 Linguistic Processors & Infrastructure

- ITC-IRST
- Basque, Catalan, English, Italian, Spanish

- Tokenization and sentence boundary detection
- Lemmatization
- Part of Speech tagging
- Noun-group chunking
- Robust-shallow parsing
- NERC
- Keyword, topic and terminology detection
- Text Classification (e.g. FINANCE, SPORT, etc.)
- Direct access to web Search Engines

MEANING: Workplan



MEANING: WP4 (Knowledge) Integration

- TALP-UPC
- The **Multilingual Central Repository** acts as a multilingual interface for uploading, integrating and porting all the knowledge produced by MEANING

- **Uploading** the knowledge acquired from one language to the MCR
- **Integrating** and validating the knowledge uploaded
- **Porting** all the knowledge acquired to the local wordnets, balancing resources and technological advances across languages

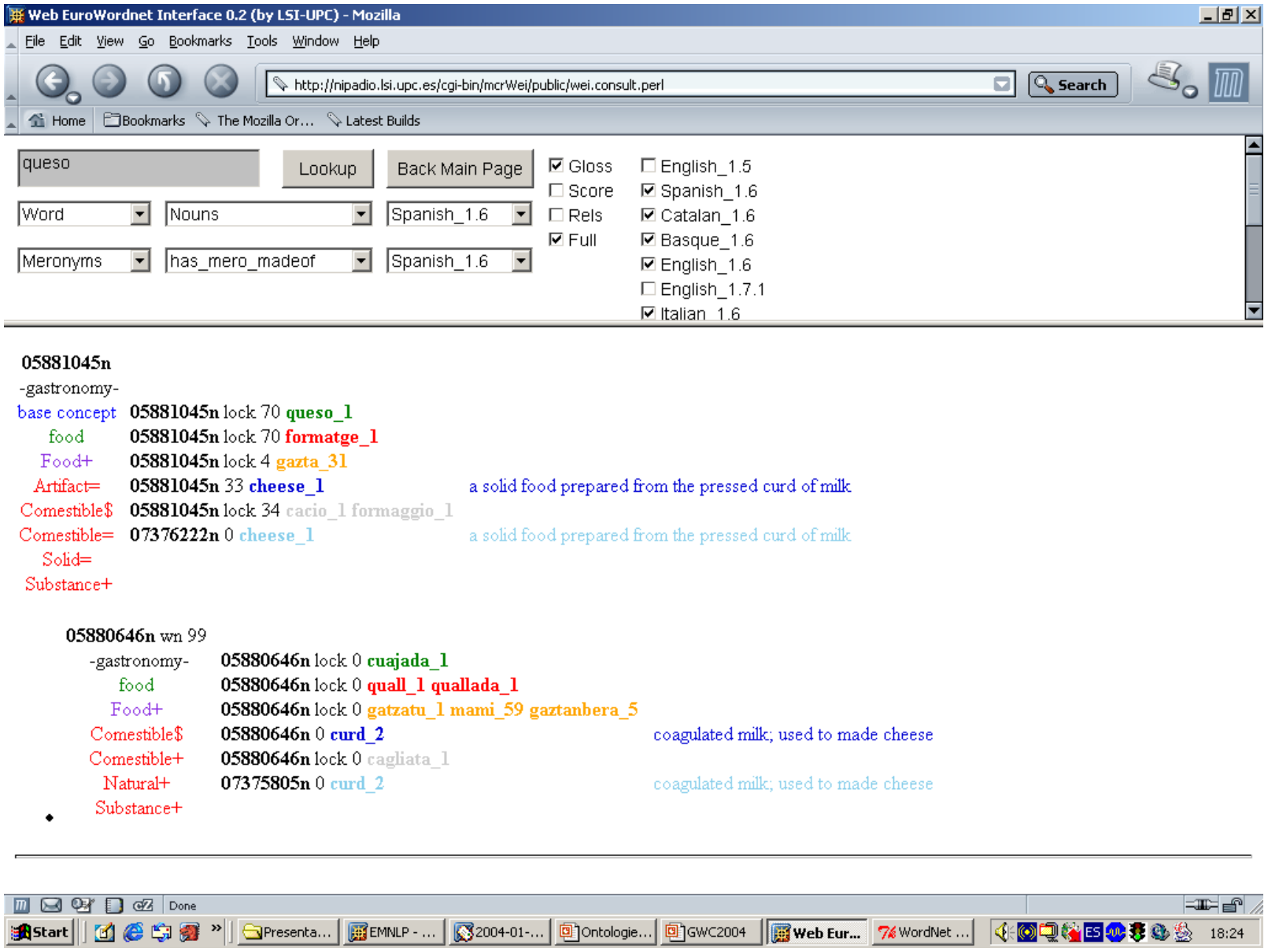
MEANING: MCR Software

- Web Interface to the MCR
  - Based on Web EuroWordNet Interface (WEI)
- APIs
  - SOAP
  - Perl, C++
- Import/Export facilities
  - XML
- Advanced Analysis Module
  - Provides different views of the multilingual data

MEANING: MCR Content

- ILI
  - WordNet1.6
  - EuroWordNet Base Concepts
  - EuroWordNet Top Ontology
  - Multiwordnet Domains
  - SUMO
- Local wordnets
  - Wordnets of five Languages
    - Basque, Catalan, English, Italian, Spanish
    - Five WordNet versions (1.5, 1.6, 1.7, 1.7.1, 2.0)
  - eXtended WordNet
- Large collections of Semantic Preferences
  - Acquired from SemCor (179,942)
  - Acquired from BNC (295,422)
- Instances
  - Named Instances

MEANING: MCR



WordNet Extensions

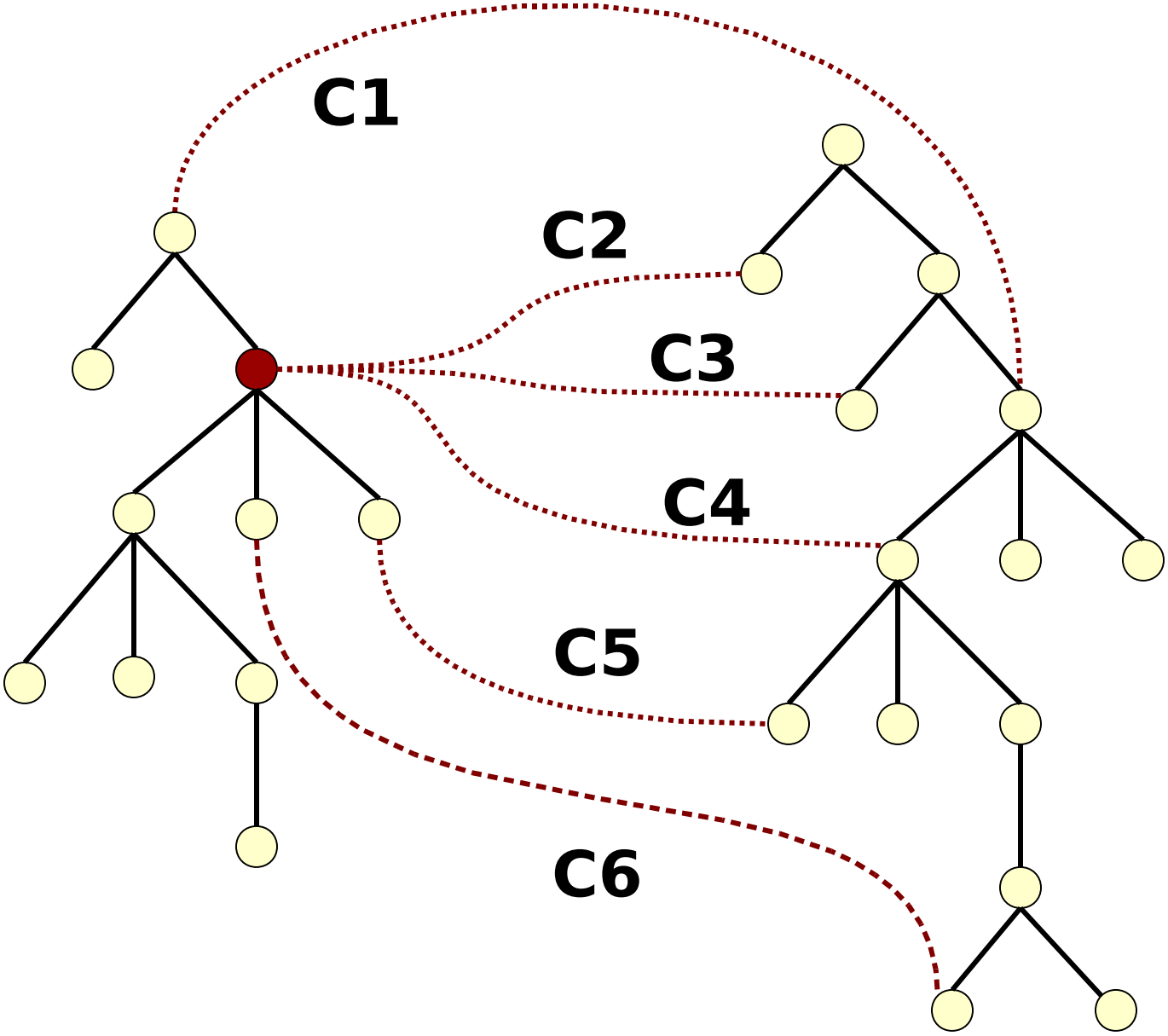
MEANING: Porting Process

- Uploading process
  - Checking errors and inconsistencies
  - Coherent integration of every piece of information
  - Dealing with several WordNet versions
- Integration process
  - Consistency checking and direct inference
  - Making explicit all knowledge contained into the MCR
    - Realisation (top-down)
    - Generalisation (bottom-up)
- Porting process
  - Direct porting to local wordnets or
  - New inference rules
    - When detecting particular semantic patterns

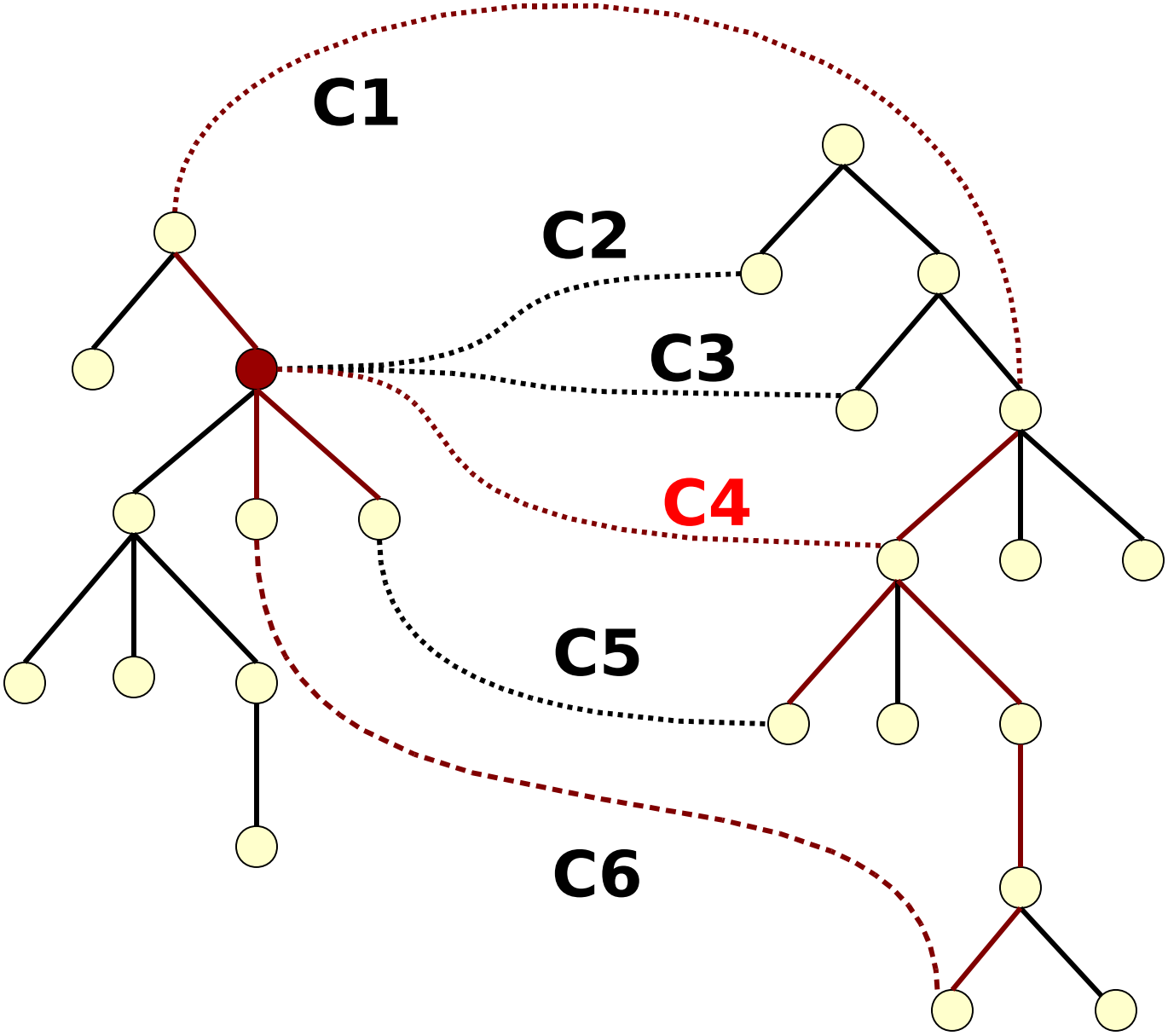
MEANING: MCR Content

- ILI
  - WordNet1.6
  - EuroWordNet Base Concepts => WN1.5
  - EuroWordNet Top Ontology => WN1.5
  - Multiwordnet Domains => WN1.6
  - SUMO => WN1.6
- Local wordnets
  - Wordnets of five European Languages
    - Basque, Catalan, English, Italian, Spanish
    - Five WordNet versions (1.5, 1.6, 1.7, 1.7.1, 2.0, 3.0+)
  - eXtended WordNet => WN1.7
- Large collections of Semantic Preferences
  - Acquired from SemCor (179,942) => WN1.6
  - Acquired from BNC (295,422) => WN1.6
- Instances
  - Named Instances => WN1.6

MEANING: Mapping technology



MEANING: Mapping technology



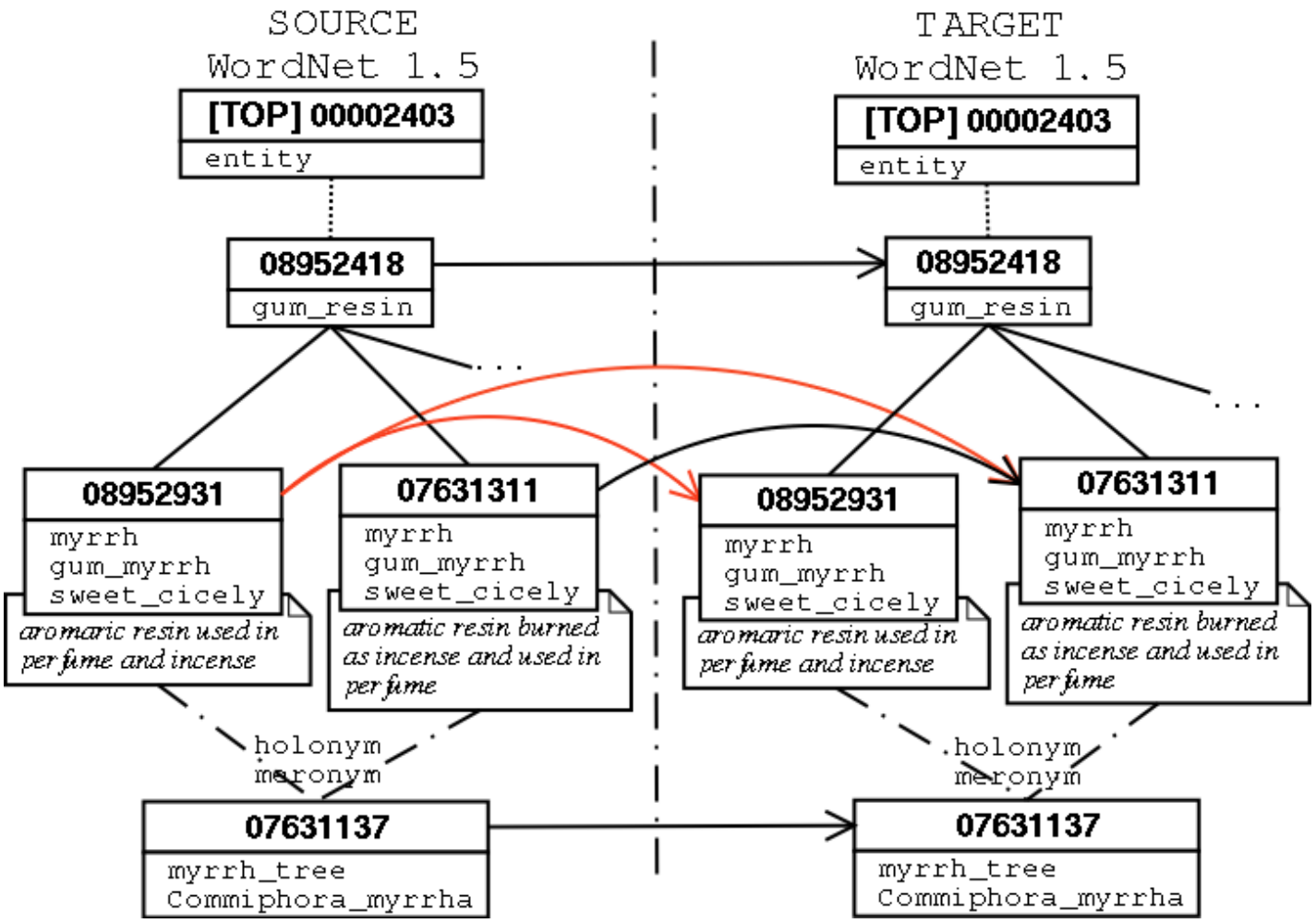
MEANING: Mapping Technology

- Mapping technology for connecting already existing semantic networks (i.e. wordnets)

- Relaxation Labelling Algorithm (Daudé et al. 2003)
- Iterative algorithm for function optimisation based on local information
- Local constraints with global effects!
  - Structural Constraints (hierarchical and non hierarchical)
  - Non structural constraints (synonym words, gloss, etc.)

- Given a set of constraints, provides de best possible mapping!

MEANING: Mapping Technology



MEANING: Porting Process

		UPLOAD0	PORT0
Relations	Spanish	53,272	=
	English	59,951	+4,246
	Italian	18,175	+763
	Catalan	53,272	=
	Basque	53,272	=
Role	Spanish	0	+162,212
	English	390,109	=
	Italian	0	+103,002
	Catalan	0	+125,997
	Basque	0	+161,807

MEANING: Porting Process

		UPLOAD0	PORT0
Instance	Spanish	0	+1,599
	English	0	+2,128
	Italian	+791	=
	Catalan	0	+1,599
	Basque	0	+365
Domain	Spanish	0	+48,053
	English	96,067	=
	Italian	30,607	=
	Catalan	0	+35,177
	Basque	0	+25,860

MEANING: Porting Process

		UPLOAD0	PORT0
Top Ontology	Spanish	1,290	=
	English	0	+1,554
	Italian	0	+946
	Catalan	1,180	=
	Basque	1,126	=

MEANING: MCR0

vaso_1 02755829n 06-NOUN.ARTIFACT FACTOTUM
GLOSS: a glass container for holding liquids while drinking

TO: 1stOrderEntity-Form-Object
TO: 1stOrderEntity-Origin-Artifact
TO: 1stOrderEntity-Function-Container
TO: 1stOrderEntity-Function-Instrument

EN: drinking_glass glass
IT: bicchiere
BA: edontzi baso edalontzi
CA: got vas

DOBJ SemCor
00849393v 0.0074 polish shine smooth ...
00201878v 0.0013 beautify embellish prettify
00826635v 0.0010 get_hold_of take
00140937v 0.0001 ameliorate amend ...
00083947v 0.0000 alter change

MEANING: MCR0

vaso_2 04195626n 08-NOUN.BODY ANATOMY
GLOSS: a tube in which a body fluid circulates

TO: 1stOrderEntity-Form-Substance-Solid
TO: 1stOrderEntity-Origin-Natural-Living
TO: 1stOrderEntity-Composition-Part
TO: 1stOrderEntity-Function-Container

EN: vessel vas
IT: vaso canale
BA: hodi baso
CA: vas

DOBJ SemCor	SUBJ SemCor
01781222v 0.0334 be occur	01831830v 0.0133 stop terminate
00058757v 0.0072 inject shoot	01357963v 0.0127 flow travel_along
01357963v 0.0068 flow ...	01830886v 0.0043 discontinue
00055849v 0.0045 administer ...	01779664v 0.0008 cease end finish ...

MEANING: MCR0

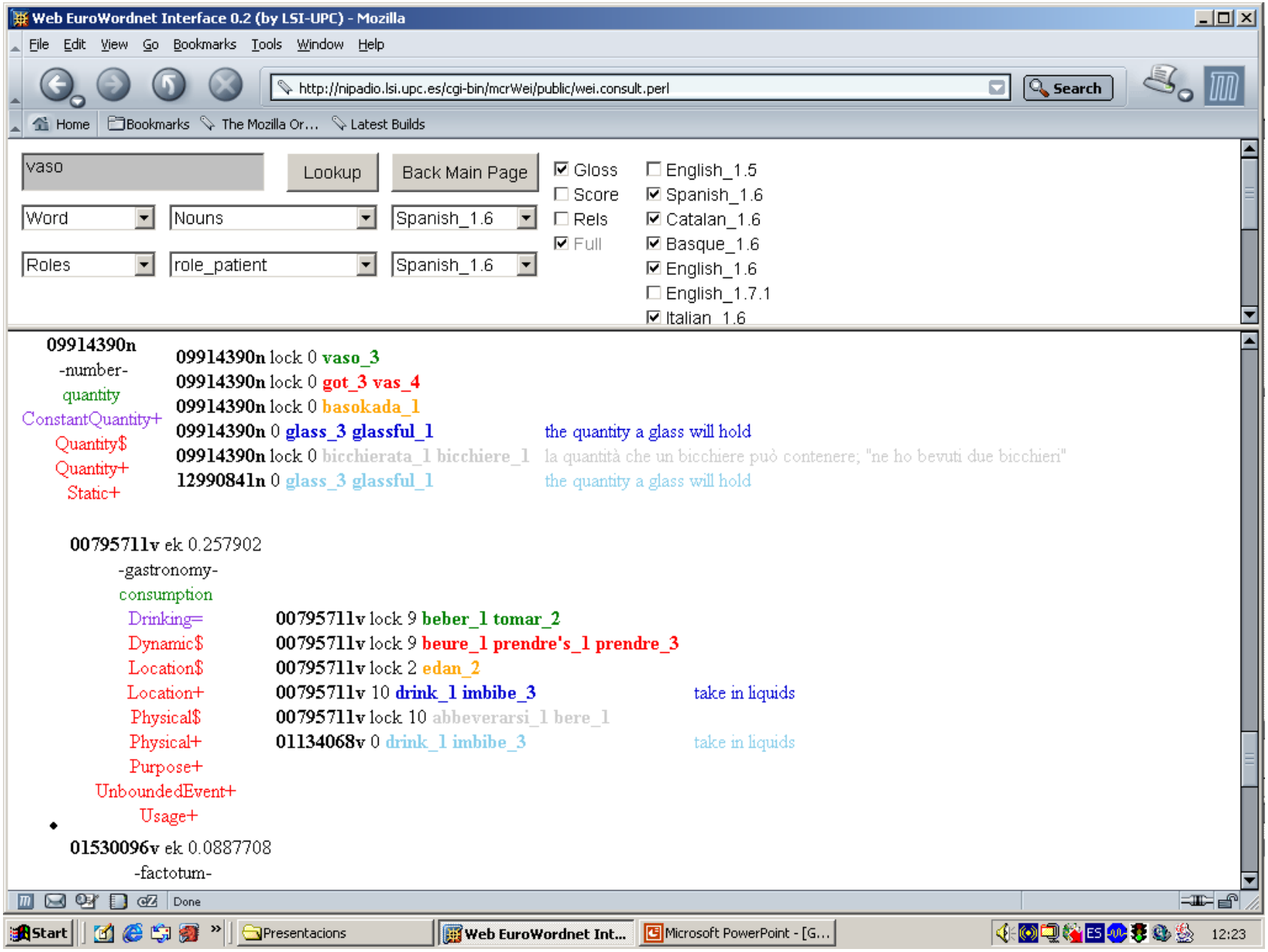
vaso_3 09914390n 23-NOUN.QUANTITY NUMBER
GLOSS: the quantity a glass will hold

TO: 1stOrderEntity-Composition-Part
TO: 2ndOrderEntity-SituationType-Static
TO: 2ndOrderEntity-SituationComponent-Quantity

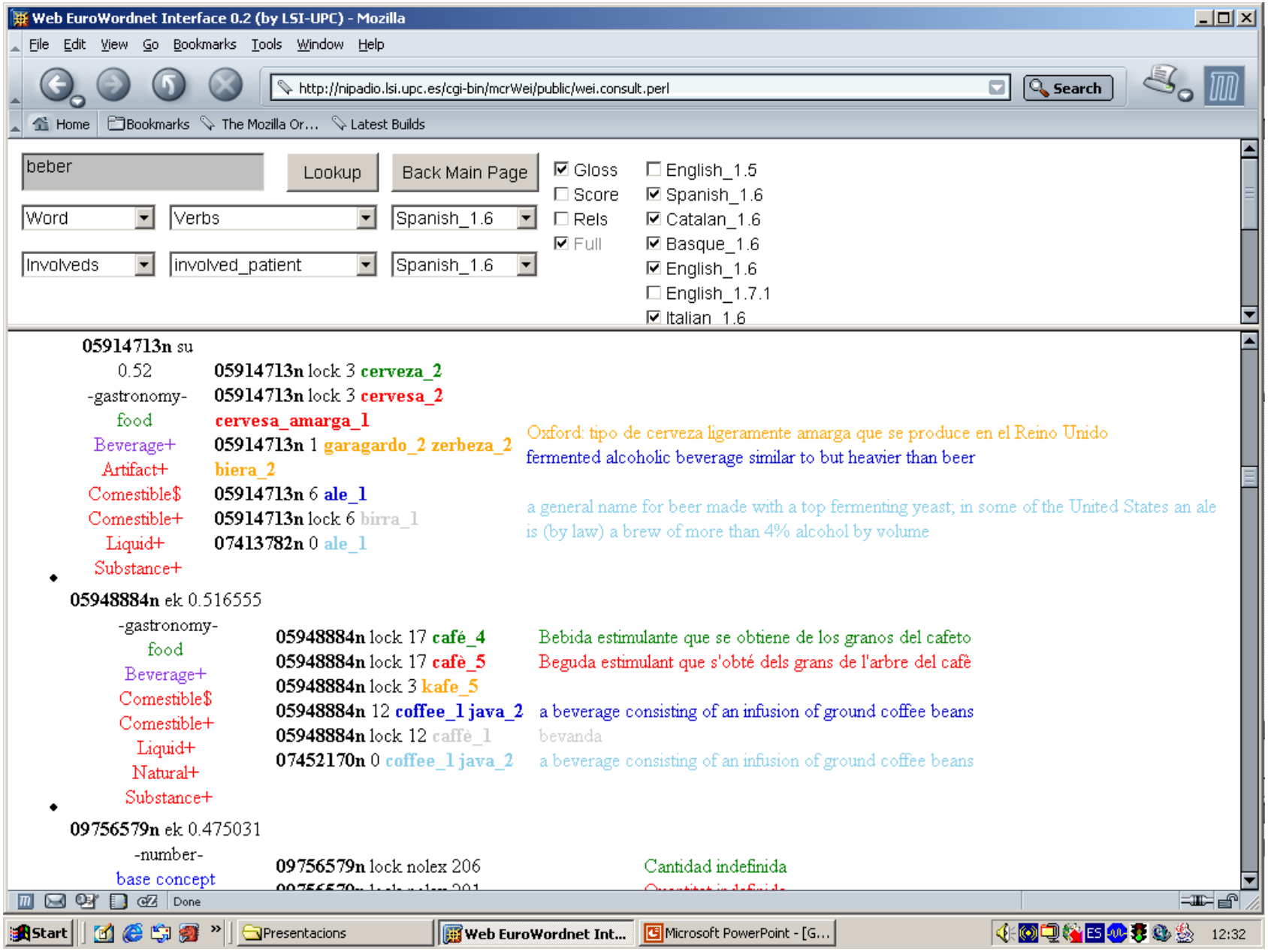
EN: glassful glass
IT: bicchierata bicchiere
BA: basokada
CA: got vas

DOBJ SemCor
00795711v 0.0026 drink imbibe
01530096v 0.0009 accept have take
00786286v 0.0009 consume have ingest take take_in
01513874v 0.0001 acquire get

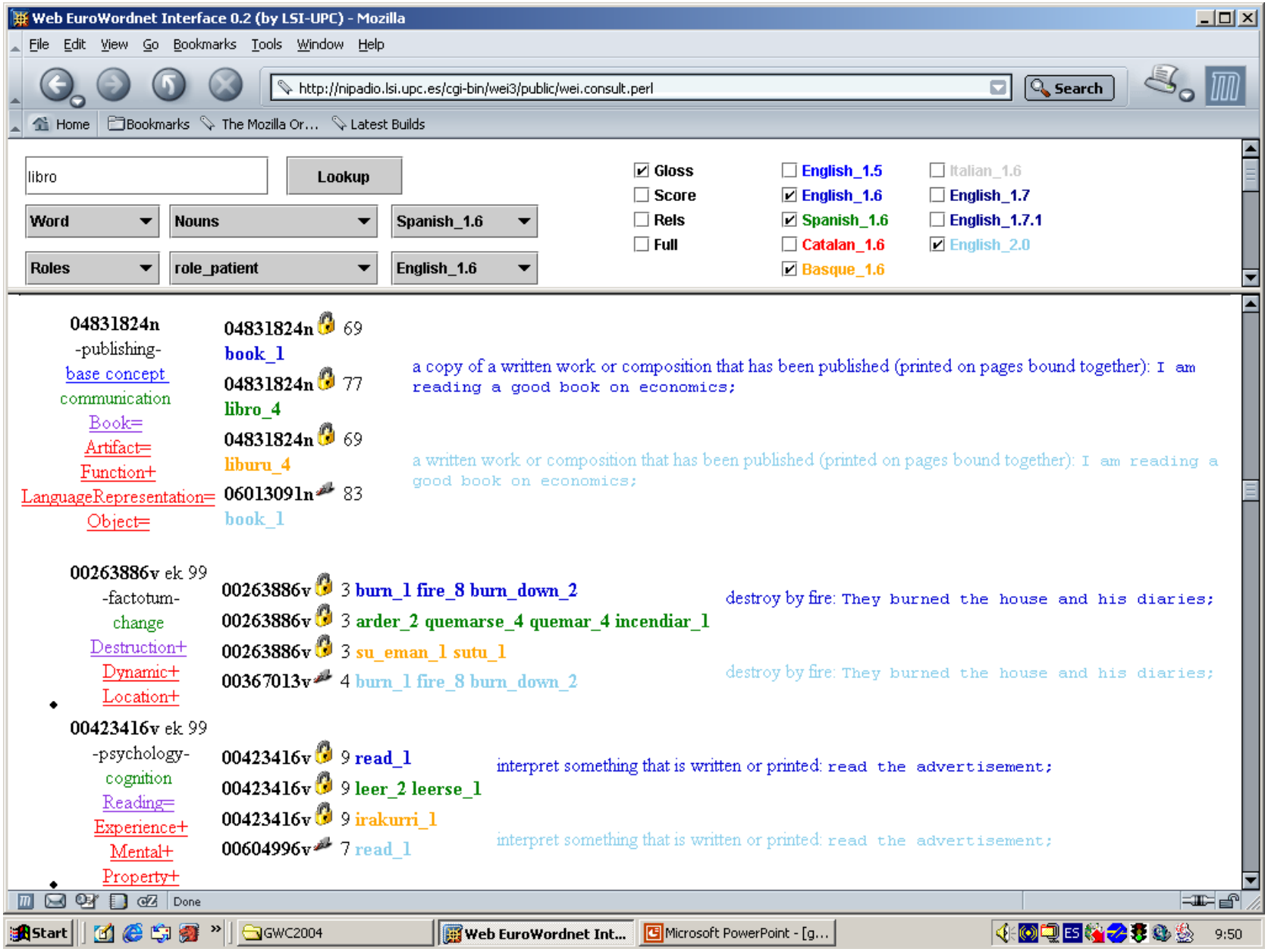
MEANING: MCR


Web EuroWordnet Interface 0.2 (by LSI-UPC) - Mozilla
File Edit View Go Bookmarks Tools Window Help
http://nipadio.lsi.upc.es/cgi-bin/mcrWei/public/wei.consult.perl
Search
Home Bookmarks The Mozilla Or... Latest Builds
vaso
Lookup
Back Main Page
Word
Nouns
Spanish_1.6
Roles
role_patient
Spanish_1.6
Gloss
Score
Rels
Full
English_1.5
Spanish_1.6
Catalan_1.6
Basque_1.6
English_1.6
English_1.7.1
Italian_1.6
09914390n
-number-
quantity
ConstantQuantity+
Quantity$
Quantity+
Static+
09914390n lock 0 vaso_3
09914390n lock 0 got_3 vas_4
09914390n lock 0 basokada_1
09914390n 0 glass_3 glassful_1 the quantity a glass will hold
09914390n lock 0 bicchierata_1 bicchiere_1 la quantità che un bicchiere può contenere; "ne ho bevuti due bicchieri"
12990841n 0 glass_3 glassful_1 the quantity a glass will hold
00795711v ek 0.257902
-gastronomy-
consumption
Drinking=
Dynamic$
Location$
Location+
Physical$
Physical+
Purpose+
UnboundedEvent+
Usage+
00795711v lock 9 beber_1 tomar_2
00795711v lock 9 beure_1 prendre's_1 prendre_3
00795711v lock 2 edan_2
00795711v 10 drink_1 imbibe_3 take in liquids
00795711v lock 10 abbeverarsi_1 bere_1
01134068v 0 drink_1 imbibe_3 take in liquids
01530096v ek 0.0887708
-factotum-
Done
Start
Presentacions
Web EuroWordnet Int...
Microsoft PowerPoint - [G...
12:23

MEANING: MCR


Web EuroWordnet Interface 0.2 (by LSI-UPC) - Mozilla
File Edit View Go Bookmarks Tools Window Help
http://nipadio.lsi.upc.es/cgi-bin/mcrWei/public/wei.consult.perl
Search
Home Bookmarks The Mozilla Or... Latest Builds
beber
Lookup
Back Main Page
Word
Verbs
Spanish_1.6
Involveds
involved_patient
Spanish_1.6
Gloss
Score
Rels
Full
English_1.5
Spanish_1.6
Catalan_1.6
Basque_1.6
English_1.6
English_1.7.1
Italian_1.6
05914713n su
0.52
-gastronomy-
food
Beverage+
Artifact+
Comestible$
Comestible+
Liquid+
Substance+
05914713n lock 3 cerveza_2
05914713n lock 3 cervesa_2
cervesa_amarga_1
05914713n 1 garagardo_2 zerbeza_2
biera_2
05914713n 6 ale_1
05914713n lock 6 birra_1
07413782n 0 ale_1
Oxford: tipo de cerveza ligeramente amarga que se produce en el Reino Unido
fermented alcoholic beverage similar to but heavier than beer
a general name for beer made with a top fermenting yeast; in some of the United States an ale is (by law) a brew of more than 4% alcohol by volume
05948884n ek 0.516555
-gastronomy-
food
Beverage+
Comestible$
Comestible+
Liquid+
Natural+
Substance+
05948884n lock 17 café_4
05948884n lock 17 cafè_5
05948884n lock 3 kafe_5
05948884n 12 coffee_1 java_2
05948884n lock 12 caffè_1
07452170n 0 coffee_1 java_2
Bebida estimulante que se obtiene de los granos del cafeto
Beguda estimulant que s'obté dels grans de l'arbre del cafè
a beverage consisting of an infusion of ground coffee beans
bevanda
a beverage consisting of an infusion of ground coffee beans
09756579n ek 0.475031
-number-
base concept
09756579n lock nolex 206
Cantidad indefinida
Done
Start
Presentacions
Web EuroWordnet Int...
Microsoft PowerPoint - [G...
12:32

MEANING: MCR


Web EuroWordnet Interface 0.2 (by LSI-UPC) - Mozilla
File Edit View Go Bookmarks Tools Window Help
http://nipadio.lsi.upc.es/cgi-bin/wei3/public/wei.consult.perl
Search
Home Bookmarks The Mozilla Or... Latest Builds
libro
Lookup
Word
Nouns
Spanish_1.6
Roles
role_patient
English_1.6
Gloss
Score
Rels
Full
English_1.5
English_1.6
Spanish_1.6
Catalan_1.6
Basque_1.6
Italian_1.6
English_1.7
English_1.7.1
English_2.0
04831824n
-publishing-
base concept
communication
Book=
Artifact=
Function+
LanguageRepresentation=
Object=
04831824n 69
book_1
04831824n 77
libro_4
04831824n 69
liburu_4
06013091n 83
book_1
a copy of a written work or composition that has been published (printed on pages bound together): I am reading a good book on economics;
a written work or composition that has been published (printed on pages bound together): I am reading a good book on economics;
00263886v ek 99
-factotum-
change
Destruction+
Dynamic+
Location+
00263886v 3 burn_1 fire_8 burn_down_2
00263886v 3 arder_2 quemarse_4 quemar_4 incendiar_1
00263886v 3 su_eman_1 sutu_1
00367013v 4 burn_1 fire_8 burn_down_2
destroy by fire: They burned the house and his diaries;
destroy by fire: They burned the house and his diaries;
00423416v ek 99
-psychology-
cognition
Reading=
Experience+
Mental+
Property+
00423416v 9 read_1
00423416v 9 leer_2 leerse_1
00423416v 9 irakurri_1
00604996v 7 read_1
interpret something that is written or printed: read the advertisement;
interpret something that is written or printed: read the advertisement;
Done
Start
GWC2004
Web EuroWordnet Int...
Microsoft PowerPoint - [g...
9:50

MEANING: MCR1

vaso_1 02755829n 06-NOUN.ARTIFACT FACTOTUM

SUMO: &%Artifact+

LOGICAL FORMULA:

glass:NN(x1) ->

glass:NN(x1) container:NN(x2) for:IN(x1, e1) hold:VB(e1, x1, x3)
liquid:NN(x3) while:IN(e0, e2) drink:VB(e2, x1)

PARSING:

```
(TOP (S (NP (NN glass) )
        (VP (VBZ is)
            (NP (NP (DT a) (NN glass) (NN container) )
                (PP (IN for)
                    (S (VP (VBG holding)
                           (PP (NP (NNS liquids) )
                               (IN while) )
                           (VBG drinking) ) ) ) ) )
        (. .) ) )
```

WSD:

```
<wf pos="DT" >a</wf>
<wf pos="NN" lemma="glass" quality="silver" wnsn="2" >glass</wf>
<wf pos="NN" lemma="container" quality="silver" wnsn="1" >container</
wf>
<wf pos="IN" >for</wf>
<wf pos="VBG" lemma="hold" quality="normal" wnsn="8" >holding</wf>
<wf pos="NNS" lemma="liquid" quality="normal" wnsn="1" >liquids</wf>

<wf pos="IN" >while</wf>
<wf pos="VBG" lemma="drink" quality="normal" wnsn="1"
>drinking</wf>
```

MEANING: MCR1

vaso_2 04195626n 08-NOUN.BODY ANATOMY

SUMO: &%BodyVessel+

LOGICAL FORMULA:

vessel:NN(x1) -> tube:NN(x1) in:IN(x2, x3) body_fluid:NN(x2) circulate:VB(e1, x2)

PARSING:

```
(TOP (S (NP (NN vessel) )
        (VP (VBZ is)
            (NP (NP (DT a) (NN tube) )
                (SBAR (WHPP (IN in)
                            (WHNP (WDT which) ) )
                      (S (NP (DT a) (NN body) (NN fluid) )
                         (VP (VBZ circulates) ) ) ) ) )
        (. .) ) )
```

WSD:

```
<wf pos="DT" >a</wf>
<wf pos="NN" lemma="tube" quality="gold" wnsn="4" wnsn="4" >tube</wf>

<wf pos="IN" >in</wf>
<wf pos="WDT" >which</wf>
<wf pos="DT" >a</wf>
<wf pos="NN" lemma="body_fluid" quality="silver" wnsn="1"
>body_fluid</wf>
<wf pos="VBZ" lemma="circulate" quality="gold" wnsn="4" wnsn="4“
>circulates</wf>
```

MEANING: MCR1

vaso_3 09914390n 23-NOUN.QUANTITY NUMBER

SUMO: &%ConstantQuantity+

LOGICAL FORMULA:

glass:NN(x1) -> quantity:NN(x1) glass:NN(x2) hold:VB(e1, x2)

PARSING:

```
(TOP (S (NP (NN glass) )
        (VP (VP (VBZ is)
                (NP (DT the) (NN quantity) )
                (NP (DT a) (NN glass) ) )
            (VP (MD will)
                (VP (VB hold) ) ) )
        (. .) ) )
```

WSD:

```
<wf pos="DT" >the</wf>
<wf pos="NN" lemma="quantity" quality="silver" wnsn="1"
>quantity</wf>
<wf pos="DT" >a</wf>
<wf pos="NN" lemma="glass" quality="normal" wnsn="2" >glass</wf>
<wf pos="MD" >will</wf>
<wf pos="VB" lemma="hold" quality="normal" wnsn="1" >hold</wf>
```

MEANING: MCR (on English LS SE3)

Source	#relations
Princeton WN1.6	138,091
Selectional Preferences from SemCor	203,546
New relations from Princeton WN2.0	42,212
Gold relations from eXtended WN	17,185
Silver relations from eXtended WN	239,249
Normal relations from eXtended WN	294,488
Total English	934,771
Total Spanish	517,279

Table 1: Semantic relations uploaded into the MCR

MEANING: MCR (on English LS SE3)

democratic	0.0126	socialist	0.0062
tammany	0.0124	organization	0.0060
alinement	0.0122	conservative	0.0059
federalist	0.0115	populist	0.0053
missionary	0.0103	dixiecrats	0.0051
whig	0.0099	know-nothing	0.0049
greenback	0.0089	constitutional	0.0045
anti-masonic	0.0083	pecking	0.0043
nazi	0.0081	democratic-republican	0.0040
republican	0.0074	republicans	0.0039
alcoholics	0.0073	labor	0.0039
bull	0.0070	salvation	0.0038

Table 2: Topic Signatures for party#n#1 using TSWEB (24 out of 15881 total words)

MEANING: MCR (on English LS SE3)

<instance id="party.n.bnc.00008131" docsrc="BNC"> <context> Up to the **late** 1960s , catholic nationalists were **split** between two **main political** groupings . There was the Nationalist Party , a weak **organization** for which **local** priests had to **provide** some **kind** of legitimation . As a <head>party</head> , it really only exercised a modicum of **power** in relation to the Stormont administration . Then there were the republican **parties** who focused their attention on Westminster **elections** . The disorganized **nature** of catholic nationalist **politics** was only turned round with the emergence of the **civil rights movement** of 1968 and the subsequent forming of the SDLP in 1970 . </context> </instance>

Table 4: Example of test num. 00008131 for party#n which its correct sense is 1

MEANING: MCR (on English LS SE3)

Baselines	P	R	F1
TRAIN	65.1	65.1	65.1
TRAIN-MFS	54.5	54.5	54.5
WN-MFS	53.0	53.0	53.0
SEMCOR-MFS	49.0	49.1	49.0
RANDOM	19.1	19.1	19.1

Table 3: P, R and F1 results for English Lexical Sample Baselines

KB	P	R	F1	Av. Size
TSSEM	**52.5**	**52.4**	**52.4**	103
MCR^2	45.1	45.1	45.1	26,429
MCR	45.3	43.7	44.5	129
spSemCor	43.1	38.7	40.8	56
$(WN+XWN)^2$	38.5	38	38.25	5,730
$WN+XWN$	40.0	34.2	36.8	74
TSWEB	36.1	35.9	36.0	1,721
XWN	38.8	32.5	35.4	69
WN^3	35.0	34.7	34.8	503
WN^4	33.2	33.1	33.2	2,346
WN^2	33.1	27.5	30.0	105
spBNC	36.3	25.4	29.9	128
WN	44.9	18.4	26.1	14

Table 4: P, R and F1 fine-grained results for the resources evaluated individually on English.

MEANING: MCR (on English LS SE3)

KB	Sum	Direct	Rank
MCR+TSSEM	52.3	45.4	**52.7**
MCR+(WN+XWN)^2	47.8	37.8	51.5
(WN+XWN)^2+TSSEM	51.0	41.7	50.5
TSSEM+TSWEB	51.0	42.2	49.4
MCR+TSWEB	48.9	37.6	48.6
(WN+XWN)^2+TSWEB	41.5	34.3	45.4

Table 5: F1 fine-grained results for the 2 system-combinations

KB	Sum	Direct	Rank
MCR+TSSEM+(WN+XWN)^2	52.6	37.9	**54.6**
MCR+TSWEB+TSSEM	54.1	37.2	53.3
MCR+TSWEB+(WN+XWN)^2	49.8	33.3	52.1
(WN+XWN)^2+TSSEM+TSWEB	51.5	36.1	51.5

Table 6: F1 fine-grained results for the 3 system-combinations

KB	Sum	Direct	Rank
MCR+(WN+XWN)^2+TSWEB+TSSEM	53.1	32.7	**55.5**

Table 7: F1 fine-grained results for the 4 system-combinations

MEANING

MCR volumes

Knowledge Resources	#relations
Princeton WN3.0	235,402
Selectional Preferences from SemCor	203,546
eXtended WN	550,922
Co-occurring relations from SemCor	932,008
New KnowNet-5	231,163
New KnowNet-10	689,610
New KnowNet-15	1,378,286
New KnowNet-20	2,358,927
New KnowNet-5 (es)	144,493
New KnowNet-10 (es)	447,317
New KnowNet-15 (es)	922,256
New KnowNet-20 (es)	1,606,893

Table 1: Number of synset relations

MEANING: MCR (from airport)

8 16	4 6	121 5
17 15	4530 5	27161 4
80 14	64713 4	68502 3
234 13	29767 3	3821 2
768 12	597 2	26 1
3091 11	20 1	1 0
10392 10	1 0	MCR
24094 9	WN+XWN	
26929 8		
19264 7		
7612 6		
2518 5		
552 4		
131 3		
11 2		
5 1		
3149 0		
WN		

MEANING: MCR and consistency checking

- body_covering_1
  - skin_4
  - plumage_1 feather_1
    - down_1
    - sickle_feather_1
  - protective_covering_2
  - skin_1
    - pellicle_1
    - dewlap_1
    - prepuce_2
    - scalp_1
    - animal_skin_1
      - parchment_2
      - leather_1
        - piece_of_leather_1
          - heel_4
          - toe_2
        - cordovan_1
      - fur_1
        - bearskin_1
        - lapin_1
  - hair_1
    - coat_3
    - hairball_2
    - mane_1
    - beard_3
    - postiche_1
    - hairdo_1
      - afro_1
    - pubic_hair_1
    - eyebrow_1
    - eyelash_1

MEANING: MCR and consistency checking

{body_covering_1 [*Living= Part= Covering=*]}
--- {skin_4 pelt_2 [*Living+ Part+ Covering+ Object=*]}
--- {plumage_1 feather_1 [*Living:Animal= Part+ Covering+ Substance:Solid=*]}
--- {down_1 [*Living:Animal+ Part+ Covering+ Substance:Solid+*]}
-x- {sickle_feather_1 [*Living:Animal= Part= Covering= Object=*]}
--- {protective_covering_2 [*Living+ Part+ Covering+ Object=*]}
--- {skin_1 tegument_1 [*Living+ Part+ Covering+ Substance:Solid =*]}
--- {pellicle_1 [*Living+ Part+ Covering+ Substance:Solid =*]}
-x- {dewlap_1 [*Object= Living:Animal= Part=*]}
-x- {prepuce_2 [*Object= Living:Animal= Part=*]}
-x- {scalp_1 [*Object= Living:Animal= Part=*]}
--- {animal_skin_1 [*Living+ Part+ Covering+ Substance:Solid =*]}
-x- {parchment_2 [*Substance:Solid= Artifact=*]}
-x- {leather_1 [*Substance:Solid= Artifact=*]}
-x- {piece_of_leather_1 [*Object= Artifact=*]}
--- heel_4 [*Object+ Artifact+ Garment= Part= *]}
--- toe_2 [*Object+ Artifact+ Garment= Part= *]}
--- {cordovan_1 [*Substance:Solid+ Artifact+*]}
-x- {fur_1[*Object= Artifact=*]}
--- {bearskin_1 [*Object+ Artifact+*]}
--- {lapin_1 [*Object+ Artifact+*]}
--- {hair_1 [*Living+ Part+ Covering+ Substance:Solid= *]}
--- {coat_3 [*Living+ Part+ Covering+ Substance:Solid= *]}
-x- {hairball_2 [*Object= Living=*]
-x- {mane_1 [*Object= Living:Animal= Part=*]
-x- {beard_3 [*Object= Living:Animal= Part= Covering=*]
-x- {postiche_1 [*Object+ Artifact+ Covering+ Garment+*][1]
------> {disguise_2}
-x- {hairdo_1 [*Property= Manner=*]}
--- afro_1 [*Property+ Manner+*]}
--- {pubic_hair_1 [*Living+ Part+ Covering+ Substance:Solid+*]}
-x- {eyebrow_1 [*Object= Living:Human= Part=*]}
-x- {eyelash_1 [*Object= Living= Part=*]}

MEANING: MCR and consistency checking

00536235n blow &%Breathing+ anatomy
00005052v blow &%Breathing+ medicine

00003430v exhale &%Breathing+ biology
00003142v exhale &%Breathing+ medicine
00899001a exhaled &%Breathing+ factotum
00263355a exhaling &%Breathing+ factotum

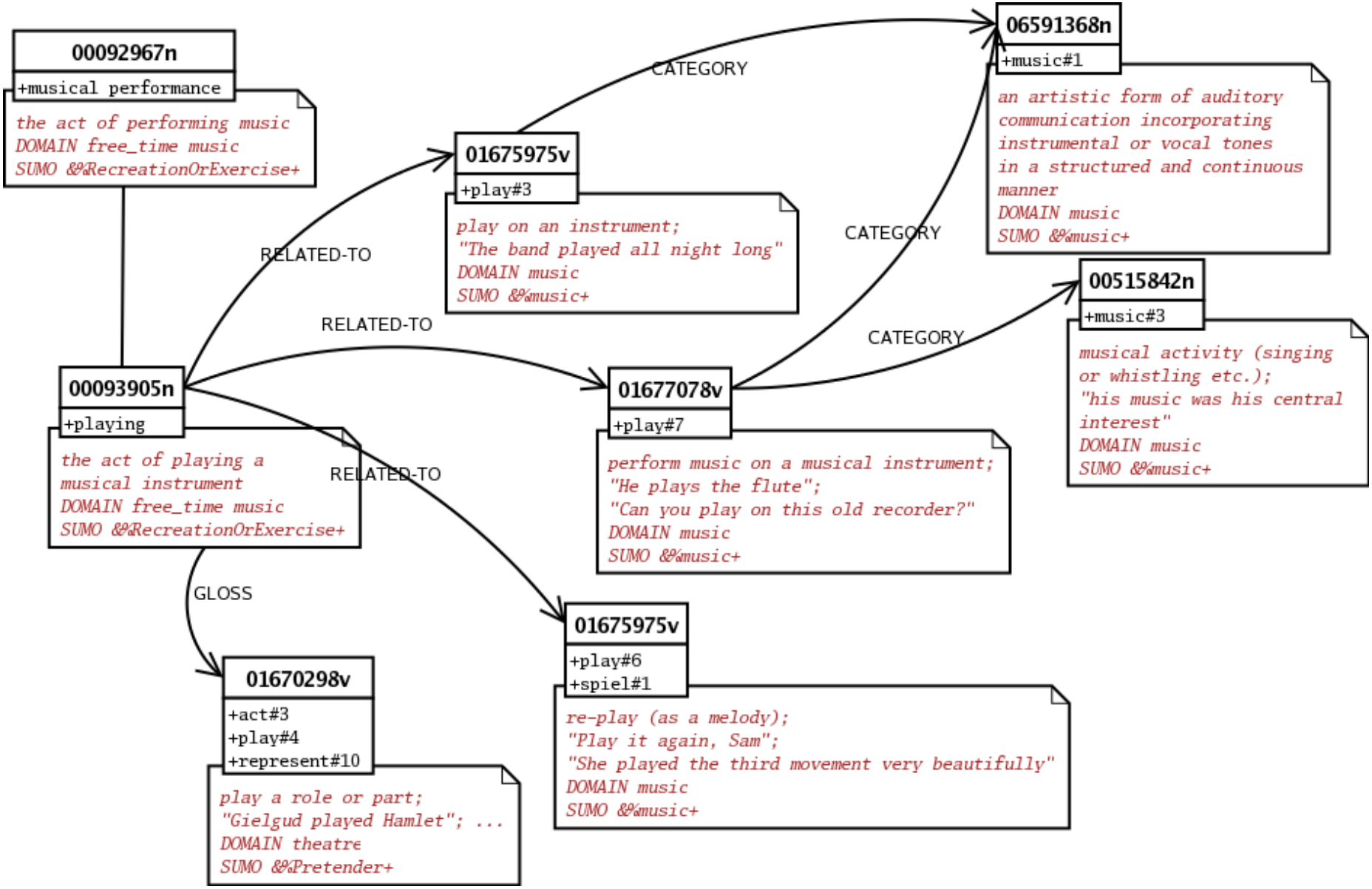
00536039n expiration &%Breathing+ anatomy
02849508a expiratory &%Breathing+ anatomy
00003142v expire &%Breathing+ medicine

02579534a inhalant &%Breathing+ anatomy
00536863n inhalation &%Breathing+ anatomy
00003763v inhale &%Breathing+ medicine
00898664a inhaled &%Breathing+ factotum
00263512a inhaling &%Breathing+ factotum

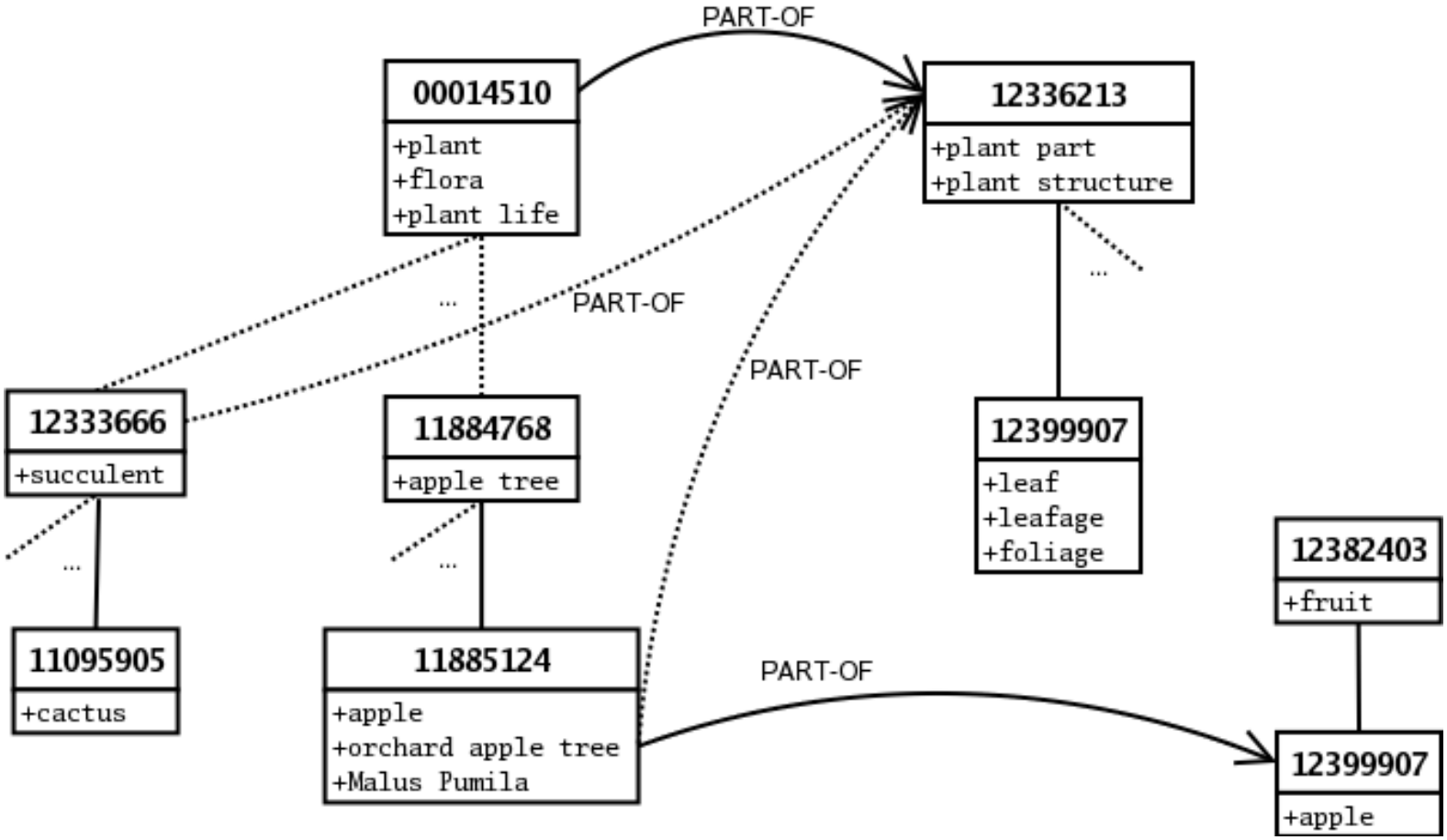
00537041n pant &%Breathing+ anatomy
00004002v pant &%Breathing+ medicine
00535106n panting &%Breathing+ anatomy
00264603a panting &%Breathing+ factotum
00411482r pantingly &%Breathing+ factotum

...

MEANING: MCR and consistency checking



MEANING: MCR and consistency checking



- Does an orchard apple tree have leaves?
- Does an orchad apple tree have fruits?
- Does a cactus have leaves?

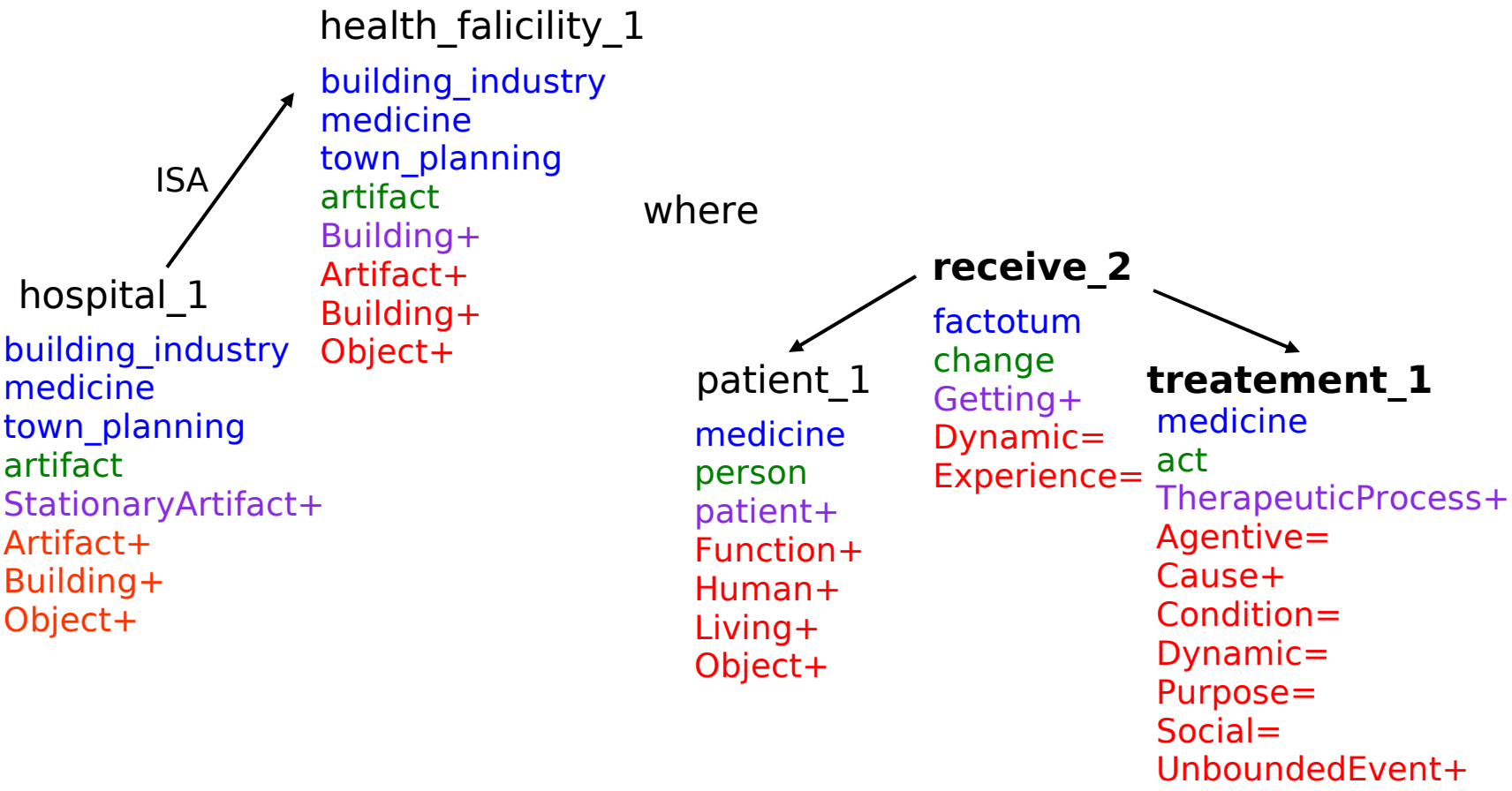
MEANING: MCR and consistency checking

Example SUMO: Boiling

- (subclass Boiling StateChange)
- (documentation Boiling "The Class of Processes where an Object is heated and converted from a Liquid to a Gas.")
- (=>
  (instance ?BOIL Boiling)
  (exists
  (?HEAT)
  (and
  (instance ?HEAT Heating)
  (subProcess ?HEAT ?BOIL))))
- "if instance BOIL Boiling, then there exists HEAT such that instance HEAT Heating and subProcess HEAT BOIL"

MEANING

hospital_1 a health facility where patients receive treatment



MEANING

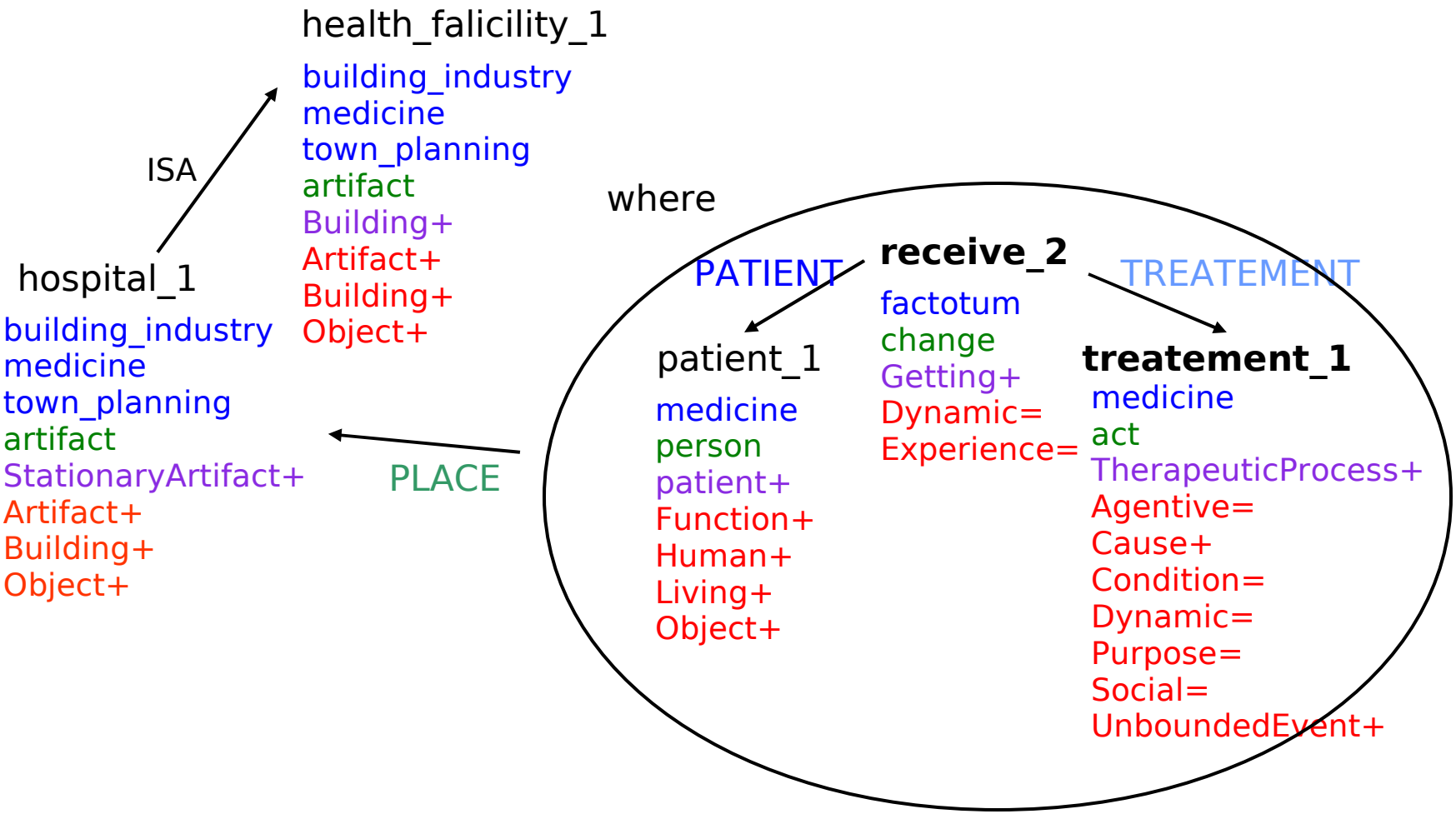
FRAMENET: cure.n

Frame Elements	Core Type
Affliction	Core
Body_part	Core
Degree	Peripheral
Duration	Extra-Thematic
Healer	Core
Manner	Peripheral
Medication	Core
Motivation	Extra-Thematic
Patient	Core
Place	Peripheral
Purpose	Extra-Thematic
Time	Peripheral
Treatment	Core

MEANING

hospital_1 a health facility where patients receive treatment

PLACE PATIENT TREATEMENT



MEANING: MCR

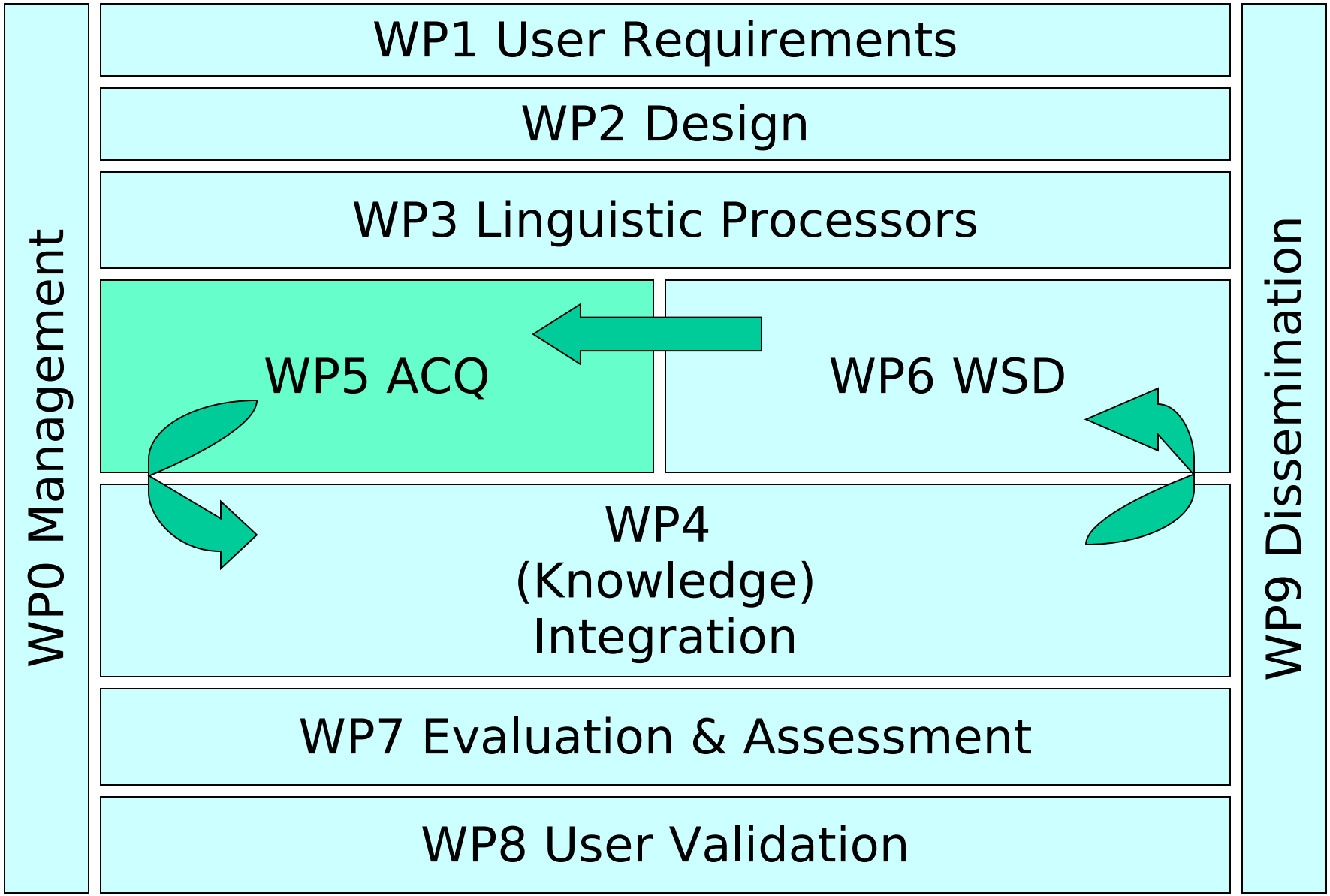
- MCR produced by Meaning is going to constitute the natural multilingual large-scale linguistic resource for a number of semantic processes that need large amounts of linguistic knowledge to be effective tools (e.g. Web ontologies).

- All wordnets gained some kind of new knowledge coming from other wordnets by means of the three porting processes.

- The resulting MCR is one of the largest and richest multilingual knowledge base ever built.
  - 1,642,386 semantic relations (ILI) (138,091 from WN1.6)
  - 466,937 properties (SUMO, TO, Domains)

- http://nipadio.lsi.upc.es/cgi-bin/wei3/public/wei.consult.perl

MEANING: Workplan



MEANING: WP5 Acquisition

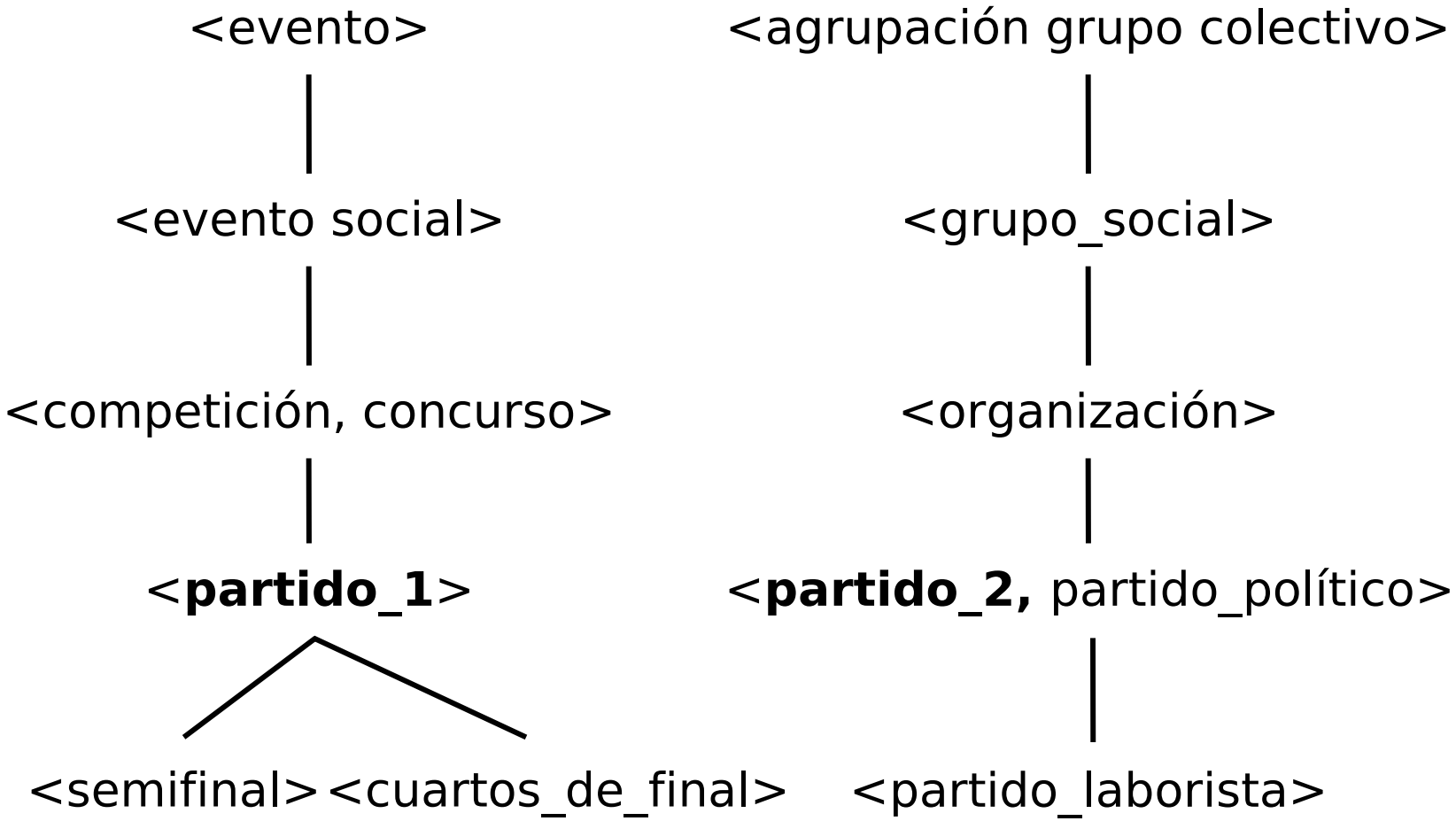
- University of Sussex

- ACQ0
  - Subcategorisation frequencies
  - Topic signatures
  - Domain Information for Named Entities
  - Sense examples
- ACQ1
  - New senses
  - Coarser-grained sense distinctions
  - Selectional Preferences
- ACQ2
  - Specific lexico-semantic relations
  - Thematic role assignments for nominalisations
  - Diathesis alternations

MEANING: WP5 Acquisition

- 11 ongoing experiments
  - A Multilingual Acquisition for predicates
  - B Collocations
  - C Domain information for NEs
  - D Topic signatures
  - E Sense Examples
  - F MRDs
  - G Selectional Preferences
  - H Coarse-grained senses
  - I Multiword Acquisition
  - J Enriching WordNet with collocations
  - K New senses

MEANING: WP5 Acquisition E: Sense Examples



MEANING: WP5 Acquisition E: Sense Examples

partido 1

Pero España puso al **partido** intensidad, ritmo y coraje.
El seleccionador cree que el **partido** de hoy contra Italia dará la medida de España
El Racing no gana en su campo desde hace seis **partidos**.

partido 2

Todos los **partidos** piden reformas legales para TV3.
La derecha planea agruparse en un **partido**.
El diputado reiteró que ni él ni UDC, “como **partido**”, han recibido dinero de Pellerols.

MEANING: WP5 Acquisition E: Sense Examples

partido 1

Rivera pide el soporte de la afición para encarrilar las **semifinales**.
Sólo el equipo de Valero Ribera puede sentenciar una **semifinal** como lo hizo ayer en un Palau Blaugrana completamente entregado.
El Racing ganó los **cuartos de final** en su campo.

partido 2

No negociaremos nunca com un **partido político** que sea partidario de la independencia de Taiwan.
Una vez más es noticia la desviación de fondos destinados a la formación ocupacional hacia la financiación de un **partido político**.
Estas lleyes fueron votadas gracias a un consenso general de los **partidos políticos**.

MEANING: WP5 Acquisition E: Sense Examples

	Senseval-2	BNC	Google
art%1:04:00:: ->	61 (48+13)	26	37.400
art%1:06:00:: ->	88 (70+18)	146	1.260.000
art%1:09:00:: ->	37 (29+8)	368	542.000
art%1:10:00:: ->	1 (1+0)	275	2.920.050
arts%1:09:00:: ->	32 (25+7)	311	3.289.320

	BNC	Google
art	9.989	56.000.000

MEANING: WP5 Acquisition E: Sense Examples

- Goal of Experiment E:

  automatically produce training data for WSD systems of size and coverage orders of magnitude larger than currently available (manually produced) resources

- First release of ExRetriever (Desember 2003)

- Experiments (February 2004)

- Future work (February 2005 and beyond …)

MEANING: WP5 Acquisition E: Sense Examples

First release of ExRetriever

- ExRetriever is able to use MCR and different corpora (SemCor, BNC, Google) through a common API.
- ExRetriever has been powered with a declarative language for query construction.
- A tool for performance evaluation and summarization (P/R/F-meassures)

MEANING: WP5 Acquisition E: Sense Examples

- Experiments
  - The experiment has been devoted to test the first prototype of ExRetriever.
  - Direct evaluation of accuracy and productivity of the different approaches for building queries have been performed for English on SemCor.
  - Words from Senseval 2 (lexical sample)
  - Different queries inspired by (Leacock et al. 98), (Mihalcea and Moldovan 99), etc.

MEANING: WP5 Acquisition E: Sense Examples

Query set using a declarative language

- Lea1Semcor
  query=or(nrel(1,syns)) or or(nrel(1,hypo)) or or(nrel(1,hype));
- Meaning1Semcor
  query=Glos(or,and,noempty) or or(nrel(1,syns)) or or(nrel(1,hypo));
- Meaning2Semcor
  query=Glos(or,and,noempty) or Glos(or,and,or,rel(hypo),noempty) or Glos(or,and,or,rel(syns),noempty);
- Moldo1Semcor
  query=or(nrel(1,syns));
- Moldo2Semcor
  query=or(rel(glos));
- Moldo3Semcor
  query=Glos(or,and,noempty);

MEANING: WP5 Acquisition E: Sense Examples

Example

Using LDB: WordNet

Using Indexer: Swish

Using Corpus: Semcor

Base on which the query is made (lemma#POS): grip#n

Query for sense (1): (clutches) or (embracing or "wrestling hold") or ("taking hold“ or prehension)

<Example Sentences="1" src="brownv/tagfiles/br-e03#1112"
Chars="60" size_tagged_Semcor="399" Words="12">
The pulsating vibration of energy
<MEANING synsetPOS="n" baseSense="1" baseLema="grip"
origPOS="n" rel="syns" synsetSense="1"
synsetLema="clutches" basePOS="n">
clutches
</MEANING>
at the_pit of your stomach.
</Example>

MEANING: WP5 Acquisition E: Sense Examples

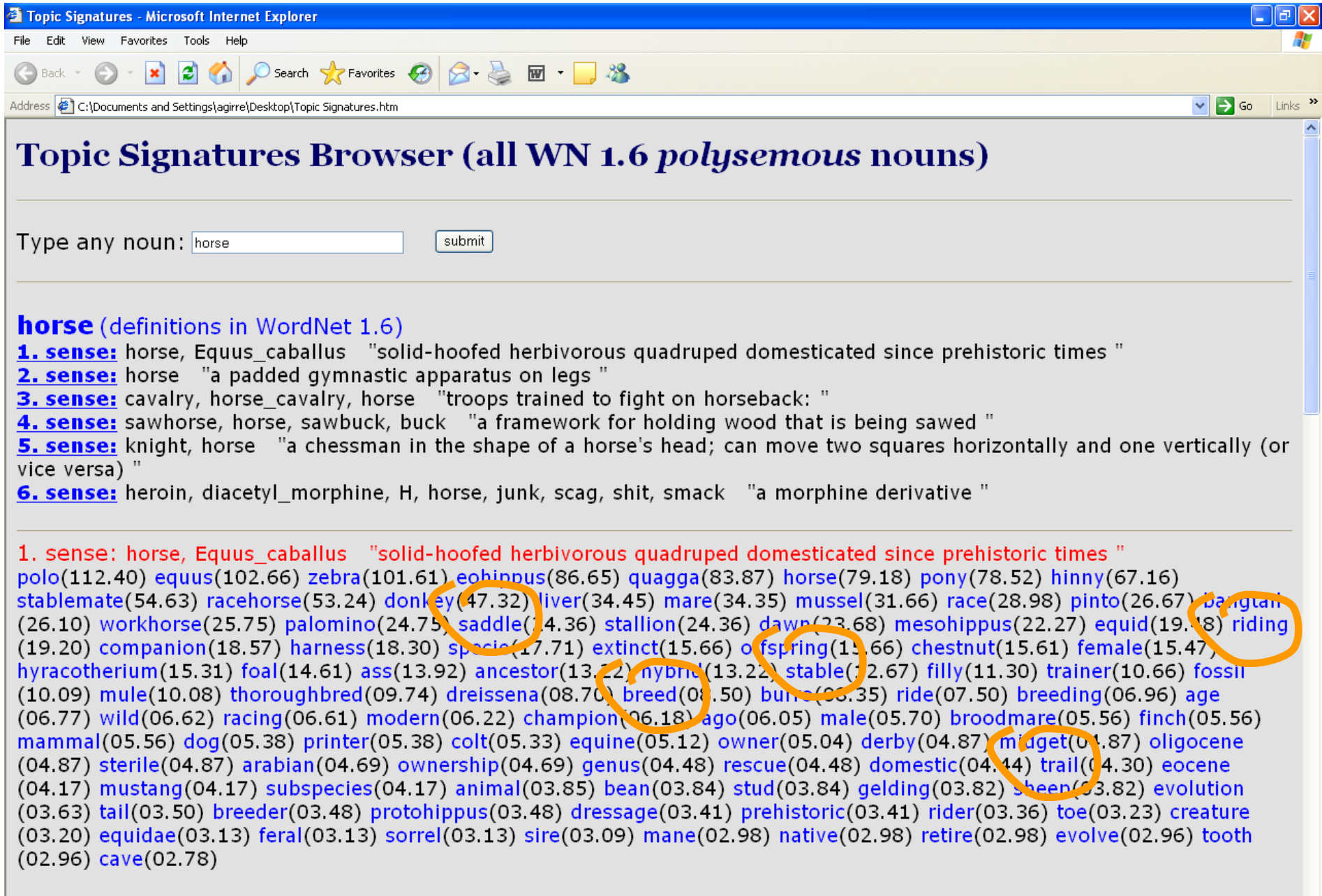
Future work (February 2004 and beyond …)

- Analysis of the Results (which query is best in which conditions)
- Designing New Queries using more knowledge (Domains, EWN Top ontology, SUMO, new relations, …)
- Latent Semantic Analisis and logic operations with vectors (Widdows et al. 2003)
- Indirect evaluation using BNC …

MEANING: WP6 WSD D: Topic Signatures

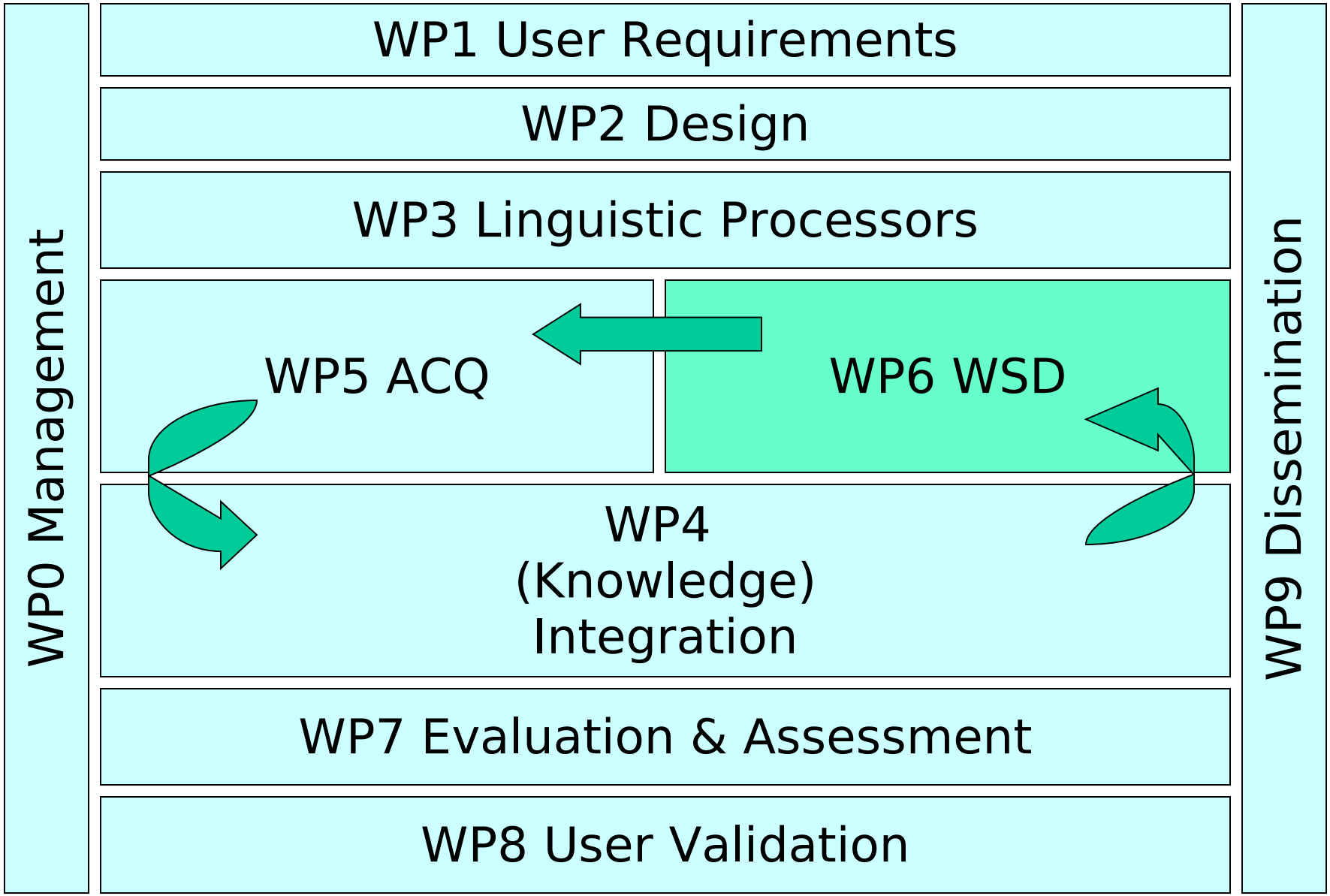
- words are ambiguos, but some are not: "transmission_channel"
- use monosemous relatives (synonyms, hypernyms, hyponyms) to gather context
- get 1000 Google snippets for each monosemous relative
- compare context of the word senses for each different word
- topic signatures for all (97.7%) noun word senses in WordNet (106.000)
- large amount of context for each word sense
  - sentences (4.7 – 7.2G)
  - processed topic signatures (2G)

MEANING: WP5 Acquisition D: Topic Signatures



MEANING: WP5 Acquisition D: Topic Signatures

MEANING: Workplan



MEANING: WP6 WSD

- IXA group, UPV/EHU
- Overall WP6 objective:
  - high precision system for all open-class words for all languages
  - Combining unsupervised knowledge-based systems with supervised Machine Learning algorithms
- Current state-of-the-art:
  - 69% in Senseval-2 all-words for English
  - Based on supervised ML on Semcor (500 Kw) as training data
  - No baseline for other languages

MEANING: WP6 WSD

- Main problem:
  - Need of dozens of manually tagged examples for each word sense (how many?)
- MEANING strategy:
  - Automatically acquiring a **huge number of examples** per sense from the web (ACQ, MCR, bootstrapping, sense ranking, ...)
  - Improve current supervised and unsupervised systems
    - Using **sophisticated linguistic information**, such as, syntactic relations, semantic classes, selectional restrictions, subcategorisation information, domains, etc.
    - Efficient margin-based **Machine Learning algorithms**
    - Novel algorithms that combine tagged examples with huge amounts of **untagged examples** in order to increase the precision of the system

MEANING: WP6 WSD

- IXA group, UPV/EHU

- WSD0
  - State-of-the-art all words systems
  - Explore improvements of current supervised systems
- WSD1
  - Improved all words systems using
    - richer linguistic features (better Linguistic Processors, MCR0)
- WSD2
  - Improved all words systems using
    - richer linguistic features (better Linguistic Processors, MCR1)
    - examples automatically acquired from the web

MEANING: WP6 WSD

- 9 ongoing experiments
  - A All-words for English
    - *B High precision WSD for Boostrapping => H*
    - *C High quality sense examples => H*
    - *D TSVM => H*
  - E All-words for non-English
  - F More informed features
  - G Unsupervised WSD
  - H Boostrapping
  - I Effect of sense clusters
  - J Semantic class classifiers
  - K Ranking senses automatically
  - L Disambiguating WN glosses

MEANING: WP6 WSD K: Ranking Senses Automatically

- The first sense heuristic (FSH) is a powerful one
- Usually, unsupervised WSD systems perform worse!
- Sense distributions change according to the type of text (Escudero et al. 2000, Martínez and Eneko 2000)
- Supervised systems only work if we do change the type of text!

MEANING: WP6 WSD k: Ranking Senses Automatically

- Ranking Method
  - Use nearest neighbours acquired from corpora using distributional similarity (e.g. Lin 1998)
  - *star*: *superstar 0.1666, player (0.157), teammate (0.121), actor (0.121) ... galaxy (0.078), sun (0.077), world (0.063), planet (0,061) ...*
  - The dominance of a given sense is related to the distributional similarity of their neighbours
  - Disambiguate the neighbours using the WordNet Similarity package

MEANING: WP6 WSD k: Ranking Senses Automatically

- Ranking Experiments
  - Ranking from different corpora: pipe
    - Semcor: tobacco pipe
    - BNC: underground pipe
  - Ranking from domain specific corpora: tie
    - BNC: necktie
    - Reuters Finance: affiliation
    - Reuters sport: draw
  - Senseval-2 all nouns task:
    - 65% precission, 60% recall

MEANING: WP6 WSD J: Semantic Class Classifiers

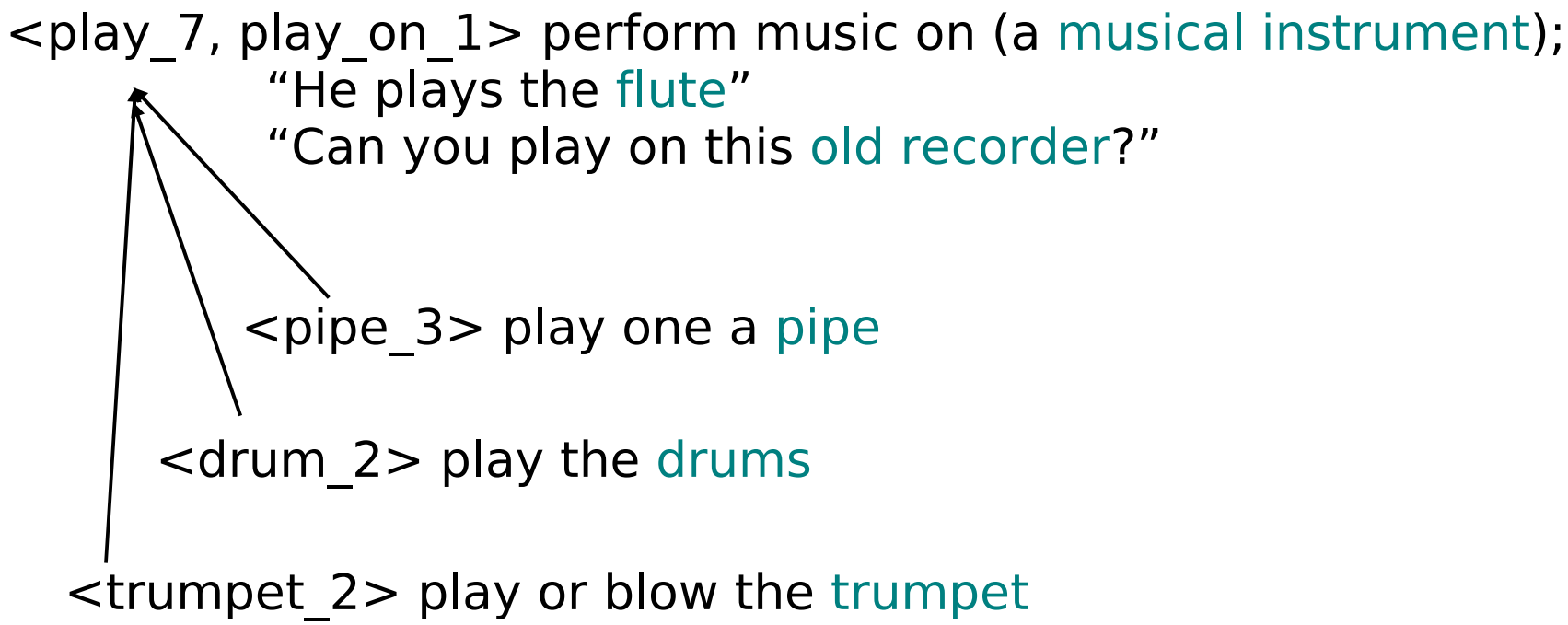
- From Financial Times
  - US officials has expected Basra to fall early

    (3) v.possession UnilateralGetting+
  - Music sales will fall by up to 15% this year

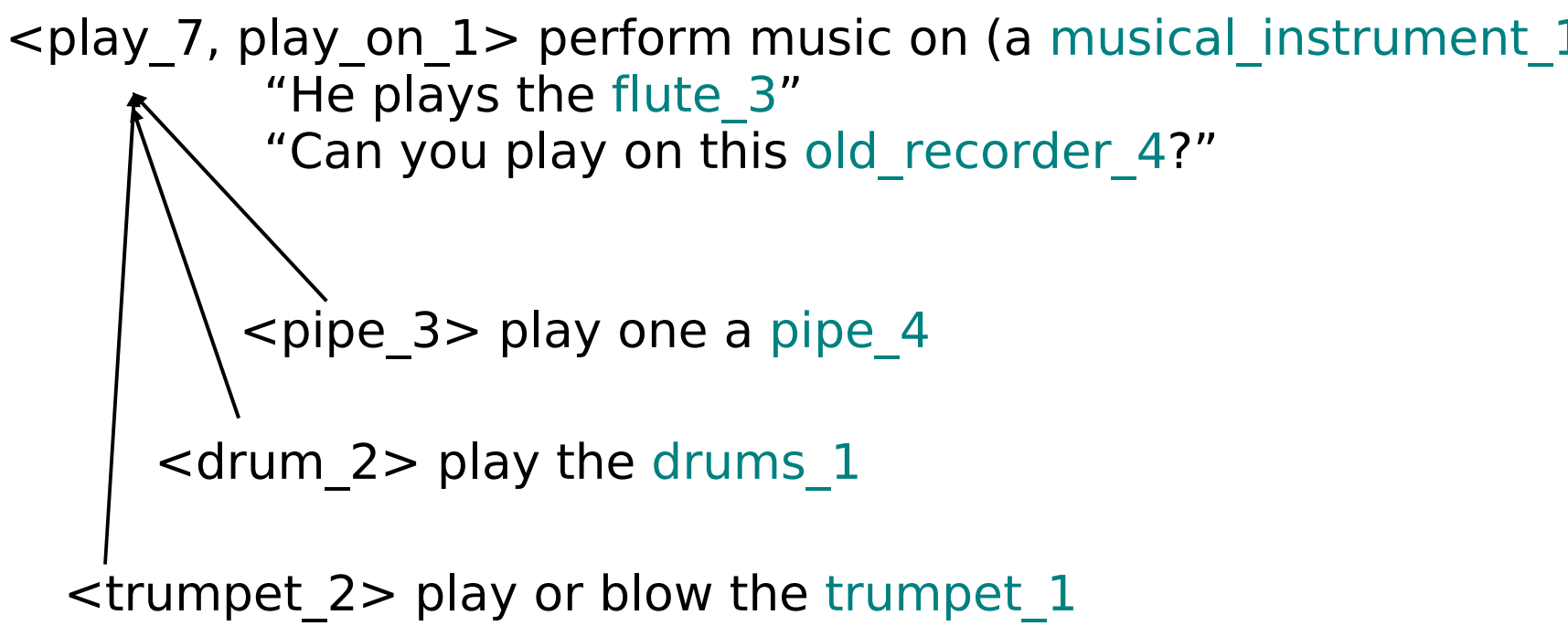
    (46) v.motion Decreasing+
  - No missiles have fallen and ...

    (21) v.motion Motion+

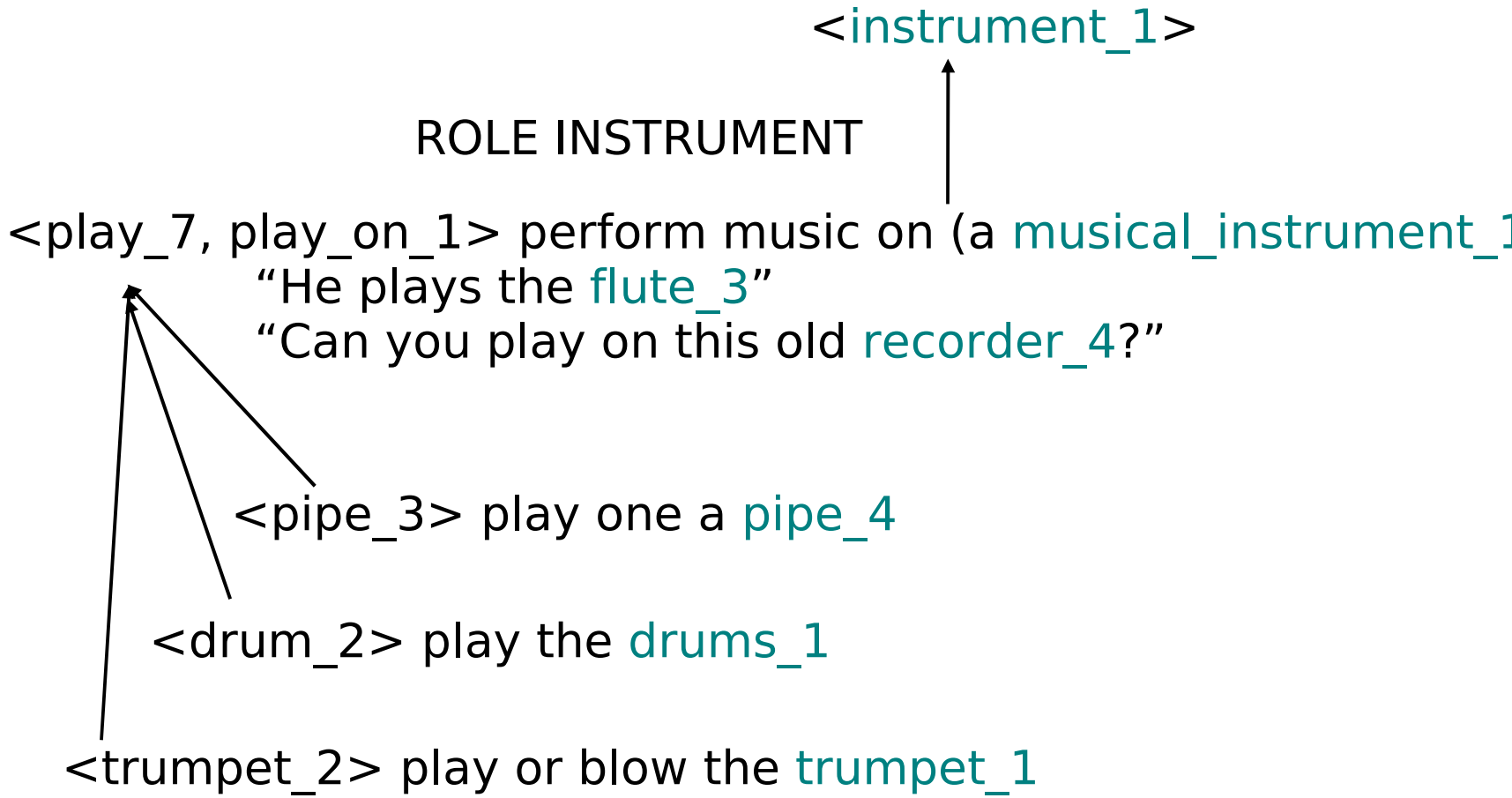
MEANING: WP6 WSD L: Disambiguating WN glosses



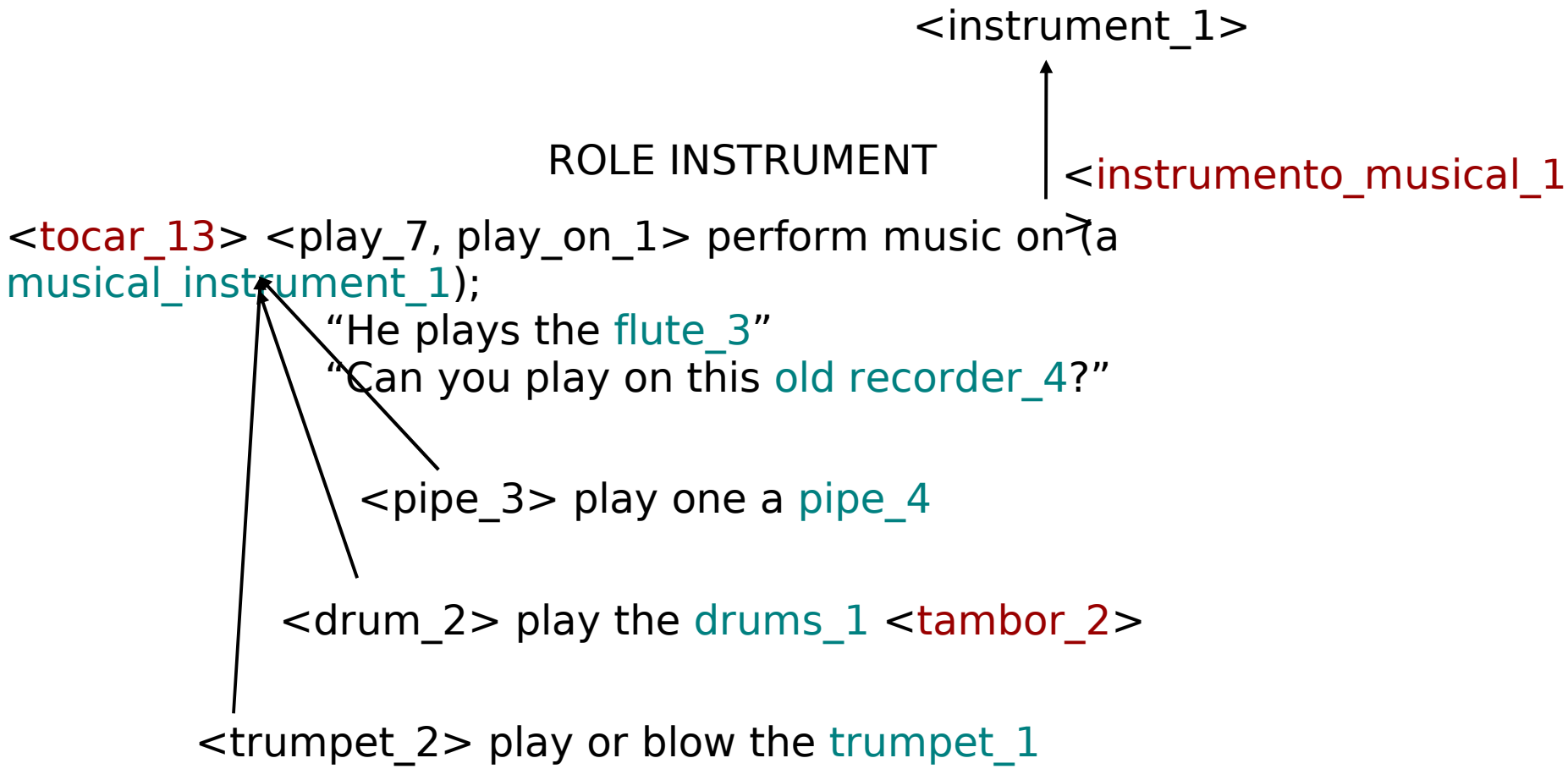
MEANING: WP6 WSD L: Disambiguating WN glosses



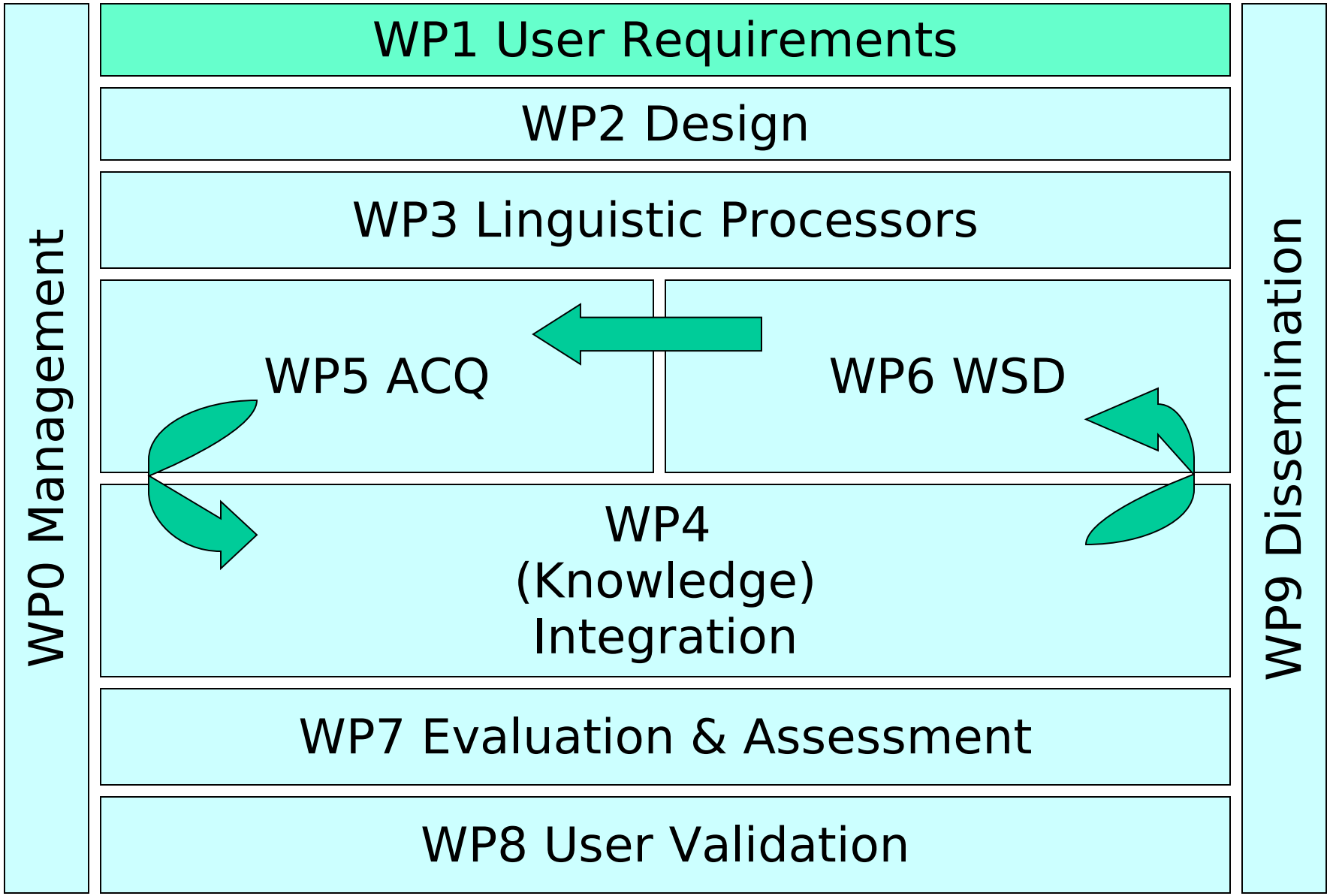
MEANING: WP6 WSD 2: Disambiguating WN glosses



MEANING: WI 6 WSD 2: Disambiguating WN glosses



MEANING: Workplan



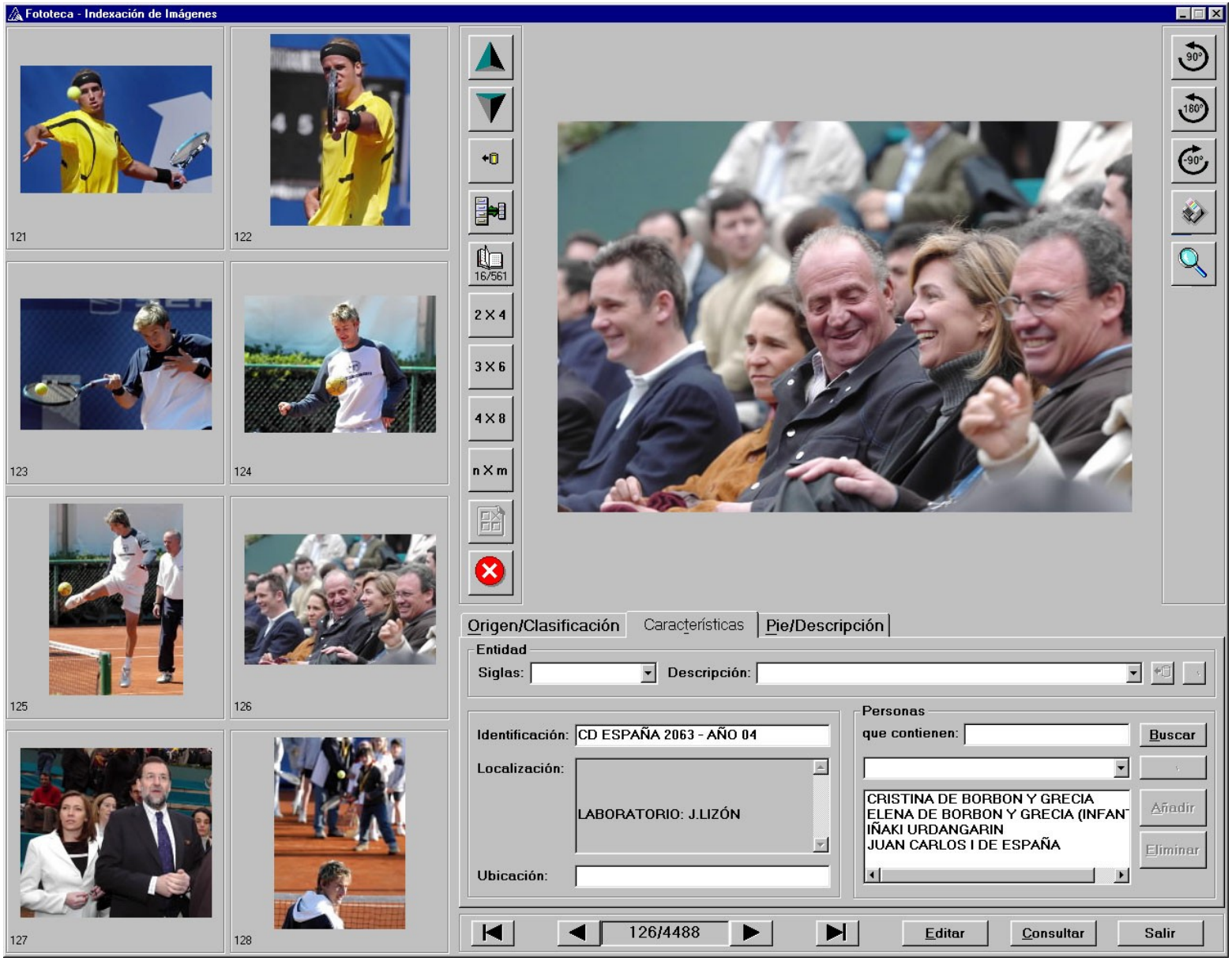
MEANING: WP1 User Requirements

- EFE FOTOTECA scenario

  - Spanish EFE News Agency
  - They receive around 800 pictures every day.
  - Mainly Spanish texts (from EFE)
  - Also English texts (from EPA and AP).
  - EFE is translating manualy most of the English texts.
  - Each caption is enriched using inconsistent coding
  - Small captions (50 words per text on average)
  - The text is in XML format

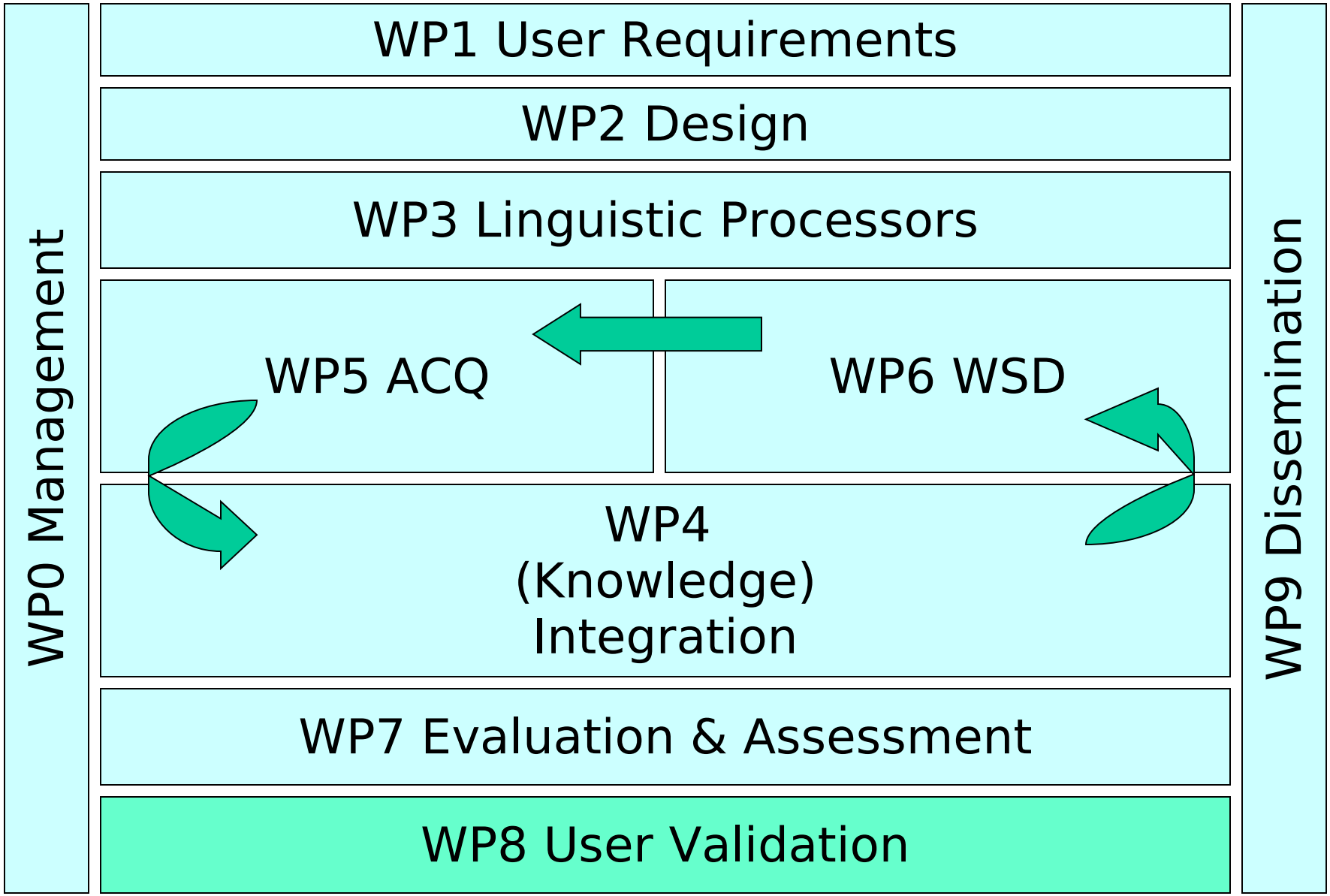
  - Costumers ask EFE for particular photographs

EFE scenario



WordNet Extensions

MEANING: Workplan



MEANING: WP8 User validation

- Baselines of Irion applications
  - Cross-lingual retrieval system: English, Dutch, German, French, Spanish and Italian
  - Document classification system
- Resources
  - SemNet
  - WordNet & WordNet Domains
  - Linking between SemNet and WordNet
- Test collection
  - Reuters News Archive 1996-1997, English
  - CLIR: 100 ambiguous queries extracted from NPs and translated
  - Document classification: 125 categories

MEANING: WP8 User validation

Irion system

- TwentyOne TNO system
- Two steps:
  - Vector-space model for document retrieval
  - Best matching phrases (NPs) from relevant documents
- Best matching between NPs and queries
  - Number of matching concepts (SemNet or MCR)
  - Degree of fuzziness mismatch
  - Degree of derivational, compounding, etc. mismatch
  - Synonymy mismatch
  - Language mismatch

MEANING: WP8 User validation

WSD system

- Based on WN Domains (Magnini et al. 2002)
- Unsupervised, Multilingual
- Microworld (documents), nanoworld (NPs)

Systems for Reuters

- NP: string matching
- FULL: Full synonymy expansion + translations (SemNet)
- WSD: FULL + WSD (retain senses of same domain)

MEANING: WP8 User validation

CLIR on the Reuters Collection

Table 2: Cross-lingual retrieval results on the Reuters collection

	English original "police cell"			English paraphrase "detention cell"			Dutch "politie-cel			German "Polizei-zelle"			French "cellule de police"			Italian "cella della polizia			Spanish "celda de la policía"		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	96	76	79	96	24	25	96	8	8	96	8	8	95	10	11	94	4	4	96	4	4
FULL	96	61	64	96	28	29	96	35	36	96	38	40	95	42	44	94	20	21	96	18	19
WSD	96	68	71	96	30	31	96	34	35	96	30	31	95	36	38	94	17	18	96	15	16

- Q: number of queries
- R: number of times the document appear in the top 10 results
- % Proportional recall

MEANING: WP8 User validation

CLIR on the Reuters Collection

Table 2: Cross-lingual retrieval results on the Reuters collection

	English original "police cell"			English paraphrase "detention cell"			Dutch "politie-cel			German "Polizei-zelle"			French "cellule de police"			Italian "cella della polizia			Spanish "celda de la policía"		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	96	76	79	96	24	25	96	8	8	96	8	8	95	10	11	94	4	4	96	4	4
FULL	96	61	64	96	28	29	96	35	36	96	38	40	95	42	44	94	20	21	96	18	19
WSD	96	68	71	96	30	31	96	34	35	96	30	31	95	36	38	94	17	18	96	15	16

- Q: number of queries
- R: number of times the document appear in the top 10 results
- % Proportional recall

MEANING: WP8 User validation

CLIR on the Reuters Collection

Table 2: Cross-lingual retrieval results on the Reuters collection

	English original "police cell"			English paraphrase "detention cell"			Dutch "politie-cel			German "Polizei-zelle"			French "cellule de police"			Italian "cella della polizia			Spanish "celda de la policía"		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	96	76	79	96	24	25	96	8	8	96	8	8	95	10	11	94	4	4	96	4	4
FULL	96	61	64	96	28	29	96	35	36	96	38	40	95	42	44	94	20	21	96	18	19
WSD	96	68	71	96	30	31	96	34	35	96	30	31	95	36	38	94	17	18	96	15	16

- Q: number of queries
- R: number of times the document appear in the top 10 results
- % Proportional recall

MEANING: WP8 User validation

CLIR on the Reuters Collection

Table 2: Cross-lingual retrieval results on the Reuters collection

	English original "police cell"			English paraphrase "detention cell"			Dutch "politie-cel			German "Polizei-zelle"			French "cellule de police"			Italian "cella della polizia			Spanish "celda de la policía"		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	96	76	79	96	24	25	96	8	8	96	8	8	95	10	11	94	4	4	96	4	4
FULL	96	61	64	96	28	29	96	35	36	96	38	40	95	42	44	94	20	21	96	18	19
WSD	96	68	71	96	30	31	96	34	35	96	30	31	95	36	38	94	17	18	96	15	16

- Q: number of queries
- R: number of times the document appear in the top 10 results
- % Proportional recall

MEANING: WP8 User validation

CLIR on the Reuters Collection

Table 2: Cross-lingual retrieval results on the Reuters collection

	English original "police cell"			English paraphrase "detention cell"			Dutch "politie-cel			German "Polizei-zelle"			French "cellule de police"			Italian "cella della polizia			Spanish "celda de la policía"		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	96	76	79	96	24	25	96	8	8	96	8	8	95	10	11	94	4	4	96	4	4
FULL	96	61	64	96	28	29	96	35	36	96	38	40	95	42	44	94	20	21	96	18	19
WSD	96	68	71	96	30	31	96	34	35	96	30	31	95	36	38	94	17	18	96	15	16

- Q: number of queries
- R: number of times the document appear in the top 10 results
- % Proportional recall

MEANING: WP8 User validation

Irion system preferences

- 1) documents with NPs having more concepts
- 2) most similar wording

Systems for Reuters

- NP: string matching
- FULL: Full synonymy expansion + translations (MCR)
- WSD: FULL + WSD (removing 50% of senses)

MEANING: WP8 User validation

CLIR on the EFE data

Table 3: Retrieval results for multi word queries

	Spanish original			Spanish paraphrase			English			Catalan			Basque			Italian		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	105	99	94	94	14	15	105	2	2	105	31	3	104	1	1	105	3	3
p1		60	57		9	1		0	0		21	2		1	1		2	2
p2		30	29		5	5		1	1		8	8		0	0		1	1
p3		9	9		0	0		1	1		2	2		0	0		0	0
FULL	105	96	91	94	71	76	105	39	37	105	70	67	104	50	48	105	39	37
p1		55	52		38	4		16	15		44	42		27	26		19	18
p2		33	31		27	29		17	16		22	21		19	18		15	14
p3		8	8		6	6		6	6		4	4		4	4		5	5
WSD	105	97	92	94	61	65	105	39	37	105	68	65	104	46	44	105	32	30
p1		60	57		39	41		21	2		48	46		27	26		20	19
p2		31	3		18	19		13	12		16	15		15	14		6	6
p3		6	6		4	4		5	5		4	4		4	4		6	6

MEANING: WP8 User validation

CLIR on the EFE data

Table 3: Retrieval results for multi word queries

	Spanish original			Spanish paraphrase			English			Catalan			Basque			Italian		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	105	99	94	94	14	15	105	2	2	105	31	3	104	1	1	105	3	3
p1		60	57		9	1		0	0		21	2		1	1		2	2
p2		30	29		5	5		1	1		8	8		0	0		1	1
p3		9	9		0	0		1	1		2	2		0	0		0	0
FULL	105	96	91	94	71	76	105	39	37	105	70	67	104	50	48	105	39	37
p1		55	52		38	4		16	15		44	42		27	26		19	18
p2		33	31		27	29		17	16		22	21		19	18		15	14
p3		8	8		6	6		6	6		4	4		4	4		5	5
WSD	105	97	92	94	61	65	105	39	37	105	68	65	104	46	44	105	32	30
p1		60	57		39	41		21	2		48	46		27	26		20	19
p2		31	3		18	19		13	12		16	15		15	14		6	6
p3		6	6		4	4		5	5		4	4		4	4		6	6

MEANING: WP8 User validation

CLIR on the EFE data

Table 3: Retrieval results for multi word queries

	Spanish original			Spanish paraphrase			English			Catalan			Basque			Italian		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	105	99	94	94	14	15	105	2	2	105	31	3	104	1	1	105	3	3
p1		60	57		9	1		0	0		21	2		1	1		2	2
p2		30	29		5	5		1	1		8	8		0	0		1	1
p3		9	9		0	0		1	1		2	2		0	0		0	0
FULL	105	96	91	94	71	76	105	39	37	105	70	67	104	50	48	105	39	37
p1		55	52		38	4		16	15		44	42		27	26		19	18
p2		33	31		27	29		17	16		22	21		19	18		15	14
p3		8	8		6	6		6	6		4	4		4	4		5	5
WSD	105	97	92	94	61	65	105	39	37	105	68	65	104	46	44	105	32	30
p1		60	57		39	41		21	2		48	46		27	26		20	19
p2		31	3		18	19		13	12		16	15		15	14		6	6
p3		6	6		4	4		5	5		4	4		4	4		6	6

MEANING: WP8 User validation

CLIR on the EFE data

Table 3: Retrieval results for multi word queries

	Spanish original			Spanish paraphrase			English			Catalan			Basque			Italian		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	105	99	94	94	14	15	105	2	2	105	31	3	104	1	1	105	3	3
p1		60	57		9	1		0	0		21	2		1	1		2	2
p2		30	29		5	5		1	1		8	8		0	0		1	1
p3		9	9		0	0		1	1		2	2		0	0		0	0
FULL	105	96	91	94	71	76	105	39	37	105	70	67	104	50	48	105	39	37
p1		55	52		38	4		16	15		44	42		27	26		19	18
p2		33	31		27	29		17	16		22	21		19	18		15	14
p3		8	8		6	6		6	6		4	4		4	4		5	5
WSD	105	97	92	94	61	65	105	39	37	105	68	65	104	46	44	105	32	30
p1		60	57		39	41		21	2		48	46		27	26		20	19
p2		31	3		18	19		13	12		16	15		15	14		6	6
p3		6	6		4	4		5	5		4	4		4	4		6	6

MEANING: WP8 User validation

CLIR on the EFE data

Table 3: Retrieval results for multi word queries

	Spanish original			Spanish paraphrase			English			Catalan			Basque			Italian		
	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%	Q	R	%
NP	105	99	94	94	14	15	105	2	2	105	31	3	104	1	1	105	3	3
p1		60	57		9	1		0	0		21	2		1	1		2	2
p2		30	29		5	5		1	1		8	8		0	0		1	1
p3		9	9		0	0		1	1		2	2		0	0		0	0
FULL	105	96	91	94	71	76	105	39	37	105	70	67	104	50	48	105	39	37
p1		55	52		38	4		16	15		44	42		27	26		19	18
p2		33	31		27	29		17	16		22	21		19	18		15	14
p3		8	8		6	6		6	6		4	4		4	4		5	5
WSD	105	97	92	94	61	65	105	39	37	105	68	65	104	46	44	105	32	30
p1		60	57		39	41		21	2		48	46		27	26		20	19
p2		31	3		18	19		13	12		16	15		15	14		6	6
p3		6	6		4	4		5	5		4	4		4	4		6	6

MEANING: WP8 User validation

CLIR

- Expansion with wordnet is only useful for synonymous queries in a monolingual setting
- Expansion with wordnet is always useful in cross-lingual setting
- Synonym selection is slightly better than concept selection (WSD based on SemNet and WordNet domains)
- Best approach: combining synonym-selection with concept selection

Classification

- Best results: using disambiguated classifiers and classifiers expanded with most frequent synonyms. Recall is up to 80% and precision is a bit lower than NO expansion. However, coverage is now 100%.

MEANING: WP8 User validation

EFE End-user validation

- 3 end-users (a, b, c)
- 3 systems (A=NP, B=FULL, C=WSD)
- 21 tasks organized in 3 test sets (1, 2, 3) of 7 tasks

	End-users		
Test sets	a	b	c
1	A	B	C
2	B	C	A
3	C	A	B

- No repetition of user, system or test-set

MEANING: WP8 User validation

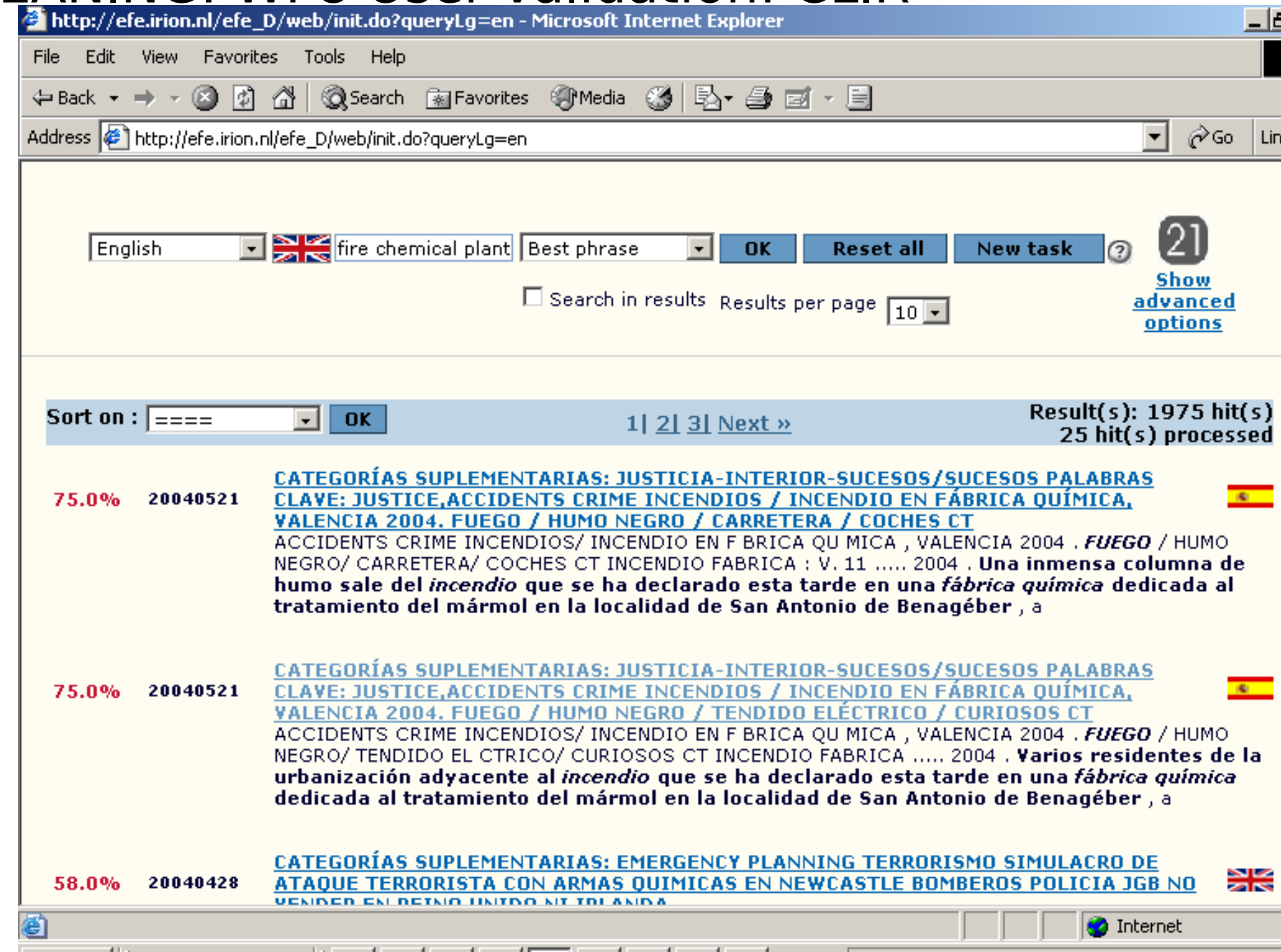
News Article 10

TOPIC = TERRORISMO
CONTEXT = Sigue la violencia en Colombia y especialmente en Medellín.
GOAL = Un entierro en Medellín.

QUERY = entierro medellín
TEXT = sepelio medellín
RESULT = **FH_1205173 20040524**
RESULT = **FH_1205172 20040524**

<entierro #35, sepelio #14, enterramiento #7> = <funeral>

MEANING: WP8 User validation: CLIR



MEANING: WP8 User validation: CLIR

Highlighted search result - Irion Technologies B.V. - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://efe.irion.nl/efe_D/web/highlightPage.do?seq=1 Go Links

CATEGORÍAS SUPLEMENTARIAS : JUSTICIA-INTERIOR-SUCESOS/ SUCESOS PALABRAS CLAVE : JUSTICE , ACCIDENTS CRIME INCENDIOS/ INCENDIO EN FÁBRICA QUÍMICA , VALENCIA 2004 . FUEGO/ HUMO NEGRO/ CARRETERA/ COCHES CT

Meta Original Close

Fecha	Categorias
20040521	TRI:JUSTICIA-INTERIOR-SUCESOS,SUCESOS SOC:SOCIEDAD-SALUD,SALUD 03027000000000 Incendios

INCENDIO FABRICA : V. 11 . Valencia , 21/ 05/ 2004 . Una inmensa columna de humo sale del incendio que se ha declarado esta tarde en una fábrica química dedicada al tratamiento del mármol en la localidad de San Antonio de Benagéber , a 15 kilómetros al norte de Valencia. EFE/ Kai Försterling .

This is the right picture | This is the wrong picture | Not sure about this picture

Internet

MEANING: WP8 User validation

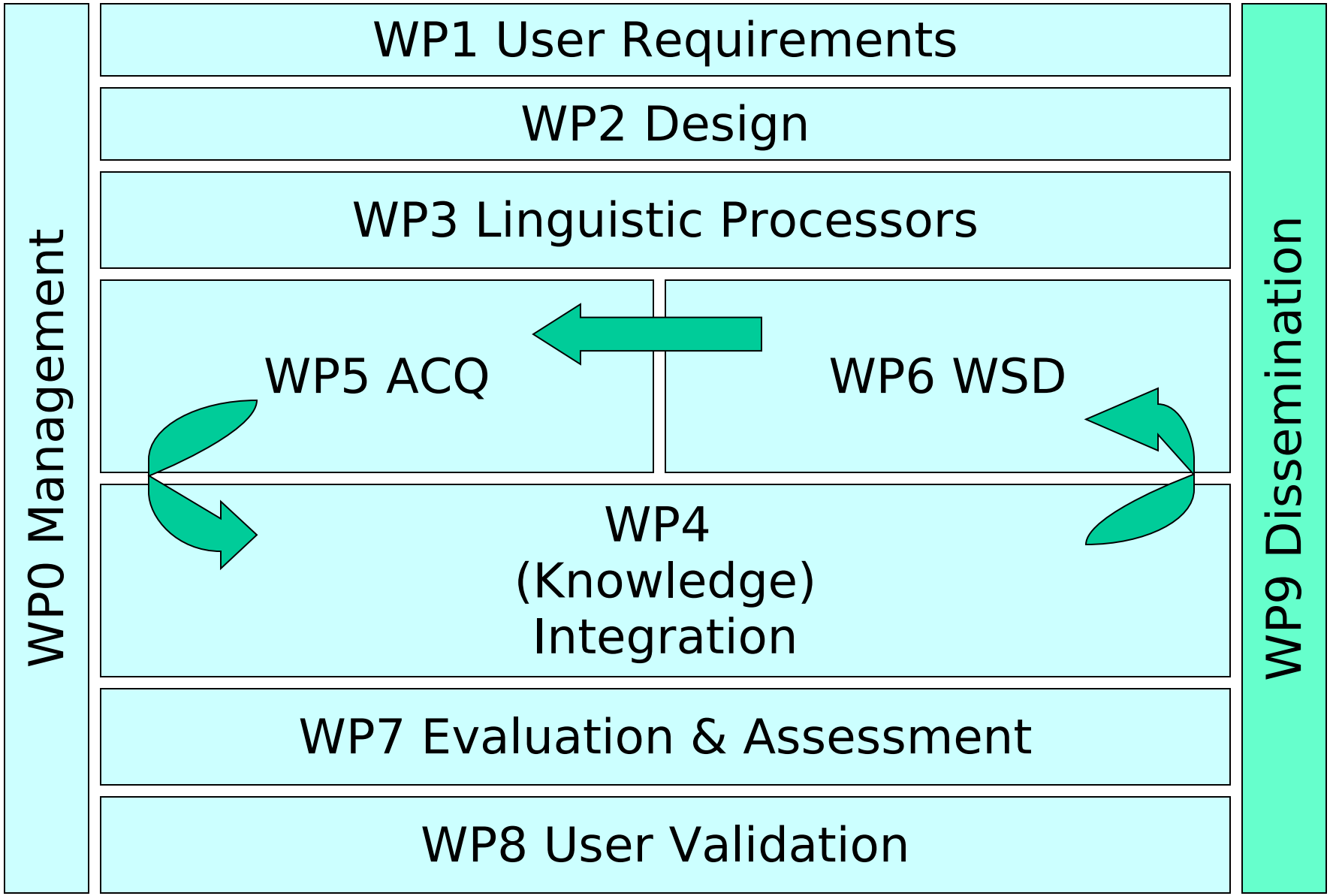
tester-all:	NP	FULL	WSD
SEARCH:	110	64	56
HIGHLIGHT:	105	55	60
DISAP:	57	28	27
CONFIRMED:	20	19	24
UNDEC:	3	6	1
TOTAL:	295	172	168

With MEANING:

- Half of the searching effort!
- More confirmed photographs
- Half of false possitives (highlight)
- Half of disaproved
- Less undecided

- Better IR and CLIR with the MCR and WSD ...
- ... unsupervised WSD for Spanish (or for any language ...)

MEANING: Workplan



MEANING: WP9 Exploitation and dissemination

- IXA, UPV/EHU

- Journals, conferences (First year: 41 published papers)
- Cooperation
  - SWAP – EDAMOK
  - ESPERONTO
  - BALKANET
- SENSEVAL-3
  - Coordinating several tasks: Basque, Catalan, Italian, Spanish
- During spring 2004:
  - First release of the MCR!
  - MEANING user group!
- Two workshops
  - First year: San Sebastián (Basque country)
  - Third year: Trento (Italy)

MEANING: WP9 First workshop

- Donostia / San Sebastian – April 10-12 2003
- Proceedings on the Web
- 8 invited speakers to give feedback (4 euro, 4 american)
  - Walter Daelemans (WSD, ML)
  - Fernando Gomez (Acquisition, semantic interpretation)
  - Julio Gonzalo (WSD, CLIR)
  - Anna Korhonen (Acquisition)
  - Dekang Lin (Acquisition)
  - Alexande Maedche (Acquisition, Semantic WEB)
  - Rada Mihalcea (WSD)
  - David Yarowsky (WSD)

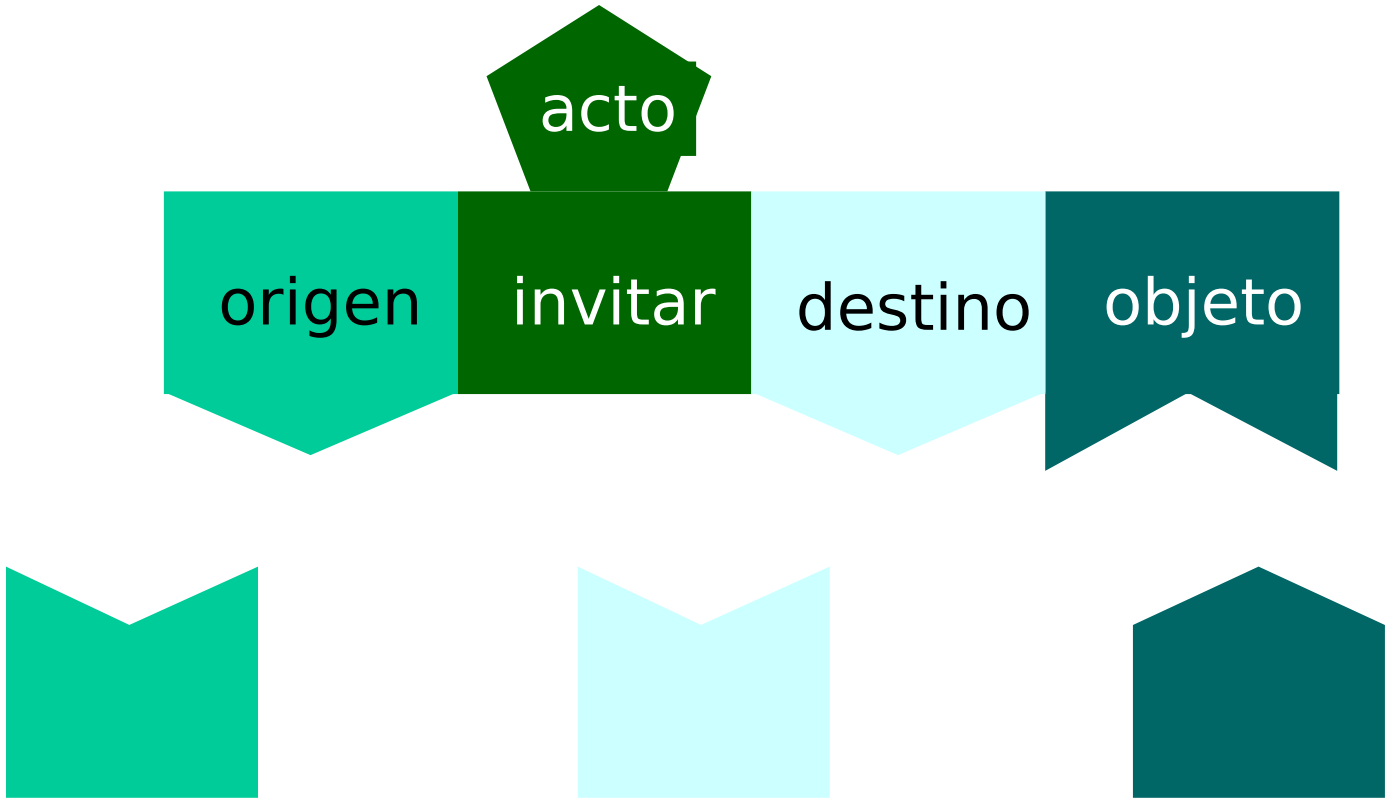
MEANING: WP9 Second workshop

- Trento – February 3-4 2005
- Proceedings on the Web
- 6 invited speakers
  - Paola Velardi (Acquisition, WSD)
  - Anna Korhonen (Acquisition)
  - Christiane Fellbaum (Knowledge Representation)
  - Mona Diab (Acquisition, WSD)
  - Dekai Wu (WSD, ML)
  - Eduard Hovy (Acquisition, WSD, applications)
- 2 panels
  - Martha Palmer, Shuly Wintner, Nicola Guarino
  - Oliviero Stock, Paul Buitelaar, Ido Dagan
- 8 papers, 5 + 3 posters
- 80 participants

MEANING: Conclusions and Results

- The good news:
  - MEANING works!
  - A Tool Set that using the semantic knowledge of MCR will obtain automatically from the web large collections of examples for each particular word sense.
  - A Tool Set for enriching the MCR using the knowledge acquired automatically from the Web.
  - A Tool Set for selecting accurately the senses of the open-class words for the languages involved in the project.
  - Multilingual Central Repository to maintain compatibility between wordnets of different languages and versions, past and new.
  - The results of MEANING will be **public** and **free**.

MEANING: Semantic Interpretation

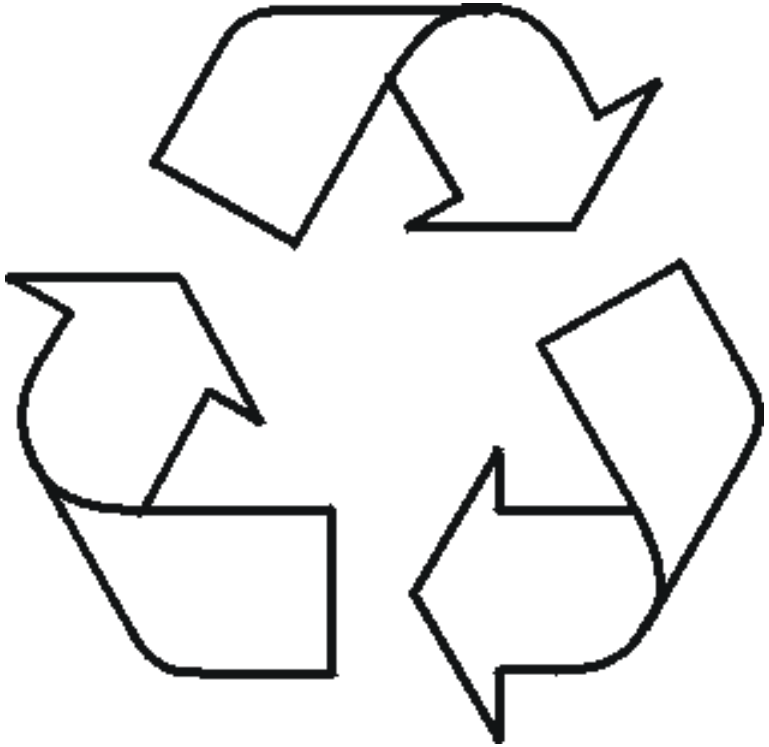


MEANING as a framework

- The bad news:
  - MEANING will focus only on the most promising research lines
  - MEANING has a large amount of work to do!
  - MEANING has only one more cycle!
- MEANING can be also seen as a **common framework** to acquire and port knowledge (information/data?) across languages, resources and tools useful for many large-scale Semantic Processing tasks
- Your collaborations and contributions are welcome!

MEANING as a framework

- Don’t waste your effort!
- MEANING can recycle your resources!



MEANING: Results

- MEANING works!
  - 3 cycles of ACQ+WSD+PORT
  - Acquiring better knowledge => better WSD
  - High precision WSD => acquiring better knowledge
  - Concept-based (with MCR and WSD) CLIR

KNOW

Developing large-scale multilingual technologies for language understanding

TIN2006-15049-C03-01

German Rigau i Claramunt

KNOW: Setting

- Introduction
- Current Content of the MCR
- Automatic Selection of Base Level Concepts
- New Topic Signatures
- Reasoning

KNOW: Introduction

- From NLP to NLU
- Large-scale Semantic Processing dealing with concepts (senses) rather than words
- Two complementary OPEN problems:

- Acquisition bottleneck
  - Autonomous large-scale knowledge acquisition systems
- Ambiguity bottleneck
  - Highly accurate Semantic Processing systems

KNOW: Introduction

Dealing with the ACQ/Semantic Processing deadlock

- Dealing with large-scale knowledge acquisition
  - Need of texts automatically tagged with semantics
- Dealing with open-domain Semantic Processing
- Dealing with multilingualism
  - Need of compatibility across resources
- Dealing with Advanced Reasoning
- Dealing with Semantic and Reasoning IR

KNOW: Introduction

Dealing with the ACQ/SEM deadlock

- Addressing Acquisition and Semantic Processing simultaneously
  - three consecutive KNOW cycles
- Language is highly polysemous
  - but also highly redundant
- Multilingualism
  - maybe is part of the solution using EuroWordNet
- Reuse of incompatible large-scale resources
  - Mapping technology to connect already available data

MEANING: Current MCR Content

- ILI
  - WordNet1.6
  - EuroWordNet Base Concepts
  - EuroWordNet Top Ontology
  - Multiwordnet Domains
  - SUMO
- Local wordnets
  - Wordnets of five Languages
    - Basque, Catalan, English, Italian, Spanish
    - Seven WordNet versions (1.5, 1.6, 1.7, 1.7.1, 2.0, 2.1, 3.0)
  - eXtended WordNet
- Large collections of Semantic Preferences
  - Acquired from SemCor (179,942)
  - Acquired from BNC (295,422)
- Instances
  - Named Instances

Automatic selection of Base Level Concepts

- In EuroWordNet, the **Base Concept** are supposed to be the concepts that play the most important role in the various wordnets of different languages.
- This role was measured in terms of two main criteria:
  - A high position in the semantic hierarchy
  - Having many relations to other concepts
- Thus, the Lexicografic Files (or Supersenses) of WN could be considered the most basic set of BC

Automatic selection of Base Level Concepts

- **Basic Level Concepts** (Rosch, 1977) should not be confused with Base Concepts.
- BLC are the result of a compromise between two conflicting principles of characterization:
  - Represent as many concepts as possible
  - Represent as many features as possible
- As a result of this, BLC typically occur in the middle of hierarchies and less than the maximum number of relations.

Automatic selection of Base Level Concepts

freq.	#rel	synset
2338	18	00017954-n group 1,grouping 1
0	19	05962976-n social group 1
729	**37**	05997592-n organisation 2,organization 1
30	10	06002286-n establishment 2,institution 1
15	**12**	06023733-n faith 3,religion 2
62	5	06024357-n Christianity 2,church 1,Christian church 1
11	14	00001740-n entity 1,something 1
51	29	00009457-n object 1,physical ob ject 1
1	39	00011937-n artifact 1,artefact 1
68	63	03431817-n construction 3,structure 1
50	**79**	02347413-n building 1,edifice 1
0	11	03135441-n place of worship 1,house of prayer 1,house of God 1
59	**19**	02438778-n church 2,church building 1
25	20	00017487-n act 2,human action 1,human activity 1
611	**69**	00261466-n activity 1
2	5	00662816-n ceremony 3
0	**11**	00663517-n religious ceremony 1,religious ritual 1
243	7	00666638-n service 3,religious service 1,divine service 1
11	1	00666912-n church 3,church service 1

Automatic selection of Base Level Concepts

PoS	#BLC	Av. depth.
Noun	3,210	5.08
Verb	1,442	2.45

Table 3: BALKANET Base Concepts using WN2.0

PoS	#BLC	Av. depth.
Noun	793	4.93
Verb	742	1.36

Table 4: MEANING Base Concepts using WN1.6

	Senses	BLC	SuperSenses
Nouns	4.92	4.10	3.01
Verbs	11.00	8.67	1.03
Nouns + Verbs	7.66	6.16	3.47

Table 5: Polysemy degree over SensEval–3

MEANING
Automatic selection of Base Level Concepts

Threshold	Rel.	PoS	#BLC	Av. depth.
0	all	Noun	3,094	7.09
		Verb	1,256	3.32
	hypo	Noun	2490	7.09
		Verb	1041	3.31
10	all	Noun	971	6.20
		Verb	719	1.39
	hypo	Noun	993	6.23
		Verb	718	1.36
20	all	Noun	558	5.81
		Verb	673	1.25
	hypo	Noun	558	5.80
		Verb	672	1.21
50	all	Noun	253	5.21
		Verb	633	1.13
	hypo	Noun	248	5.21
		Verb	633	1.10

Table 2: Automatic Base Level Concepts for WN1.6

MEANING
Class-based WSD

An ancient stone [BLC:artifact.n#1, SS:noun.artifact, D=building] church [BLC20:building.n#1, SS:noun.artifact, D:building] stands [SS=verb.stative] amid the fields [BLC:*geographic_area.n#1*:physical_object.n#1, SS:*noun.location*:noun.object, D:*factotum*:geography], the sound [BLC:property.n#2, SS:noun.attribute, D:*factotum*:acoustics] of bells [BLC:device.n#1, SS:noun.artifact, D:*factotum*:acoustics)] cascading [SS:verb.motion] from its tower [BLC:construction.n#3, SS:noun.artifact, D:factotum], calling [SS:*verb.stative*:verb.communication] the faithful [SS:*group.n#1*:social_group.n#1, SS:noun.group, D:*person*:religion] to evensong [BLC:*time_of_day.n#1*:writing.n#2, SS:noun.communication, D:religion].

Table 1: Example text annotated automatically with several semantic class labels

MEANING
Class-based WSD

An ancient stone [BLC:artifact.n#1, SS:noun.artifact, D=building] church [BLC20:building.n#1, SS:noun.artifact, D:building] stands [SS=verb.stative] amid the fields [BLC:*geographic_area.n#1*:physical_object.n#1, SS:*noun.location*:noun.object, D:*factotum*:geography], the sound [BLC:property.n#2, SS:noun.attribute, D:*factotum*:acoustics] of bells [BLC:device.n#1, SS:noun.artifact, D:*factotum*:acoustics)] cascading [SS:verb.motion] from its tower [BLC:construction.n#3, SS:noun.artifact, D:factotum], calling [SS:*verb.stative*:verb.communication] the faithful [SS:*group.n#1*:social_group.n#1, SS:noun.group, D:*person*:religion] to evensong [BLC:*time_of_day.n#1*:writing.n#2, SS:noun.communication, D:religion].

Table 1: Example text annotated automatically with several semantic class labels

Classifier	Examples	# of examples
church.n#2 *(sense approach)*	church.n#2	58
building, edifice *(class approach)*	church.n#2	58
	building.n#1	48
	hotel.n#1	39
	hospital.n#1	20
	barn.n#1	17
	.....	.....
		TOTAL= 371 examples

Table 3: Number of examples in Semcor, for sense approach and for class approach

MEANING
Class-based WSD (on SE3)

Nouns		Verbs	
System	**F1**	**System**	**F1**
Sense → BLC20			
SVM-semBLC20	76.52	GAMBL-AW	63.56
base-SemCor	*76.29*	SVM-semSS	61.29
SVM-semBLC50	75.73	SVM-semSUMO	61.15
GAMBL-AW	74.77	SVM-semWND	60.88
kuaw	74.69	kuaw	60.66
LCCaw	74.44	SVM-semBLC50	60.60
UNTaw	74.40	SVM-semBLC20	59.92
SVM-semWND	74.24	R2D2	59.79
base-WordNet	*74.16*	UNTaw	59.73
SVM-semSS	73.82	Meaning-allwords	59.37
SVM-semSUMO	73.71	*base-SemCor*	*58.82*
Meaning-allwords	73.11	*base-WordNet*	*58.28*

Sense → SuperSense			
SVM-semBLC50	81.73	SVM-semWND	79.75
base-SemCor	*81.50*	GAMBL-AW	79.4
SVM-semBLC20	81.39	*base-SemCor*	*79.07*
SVM-semSUMO	81.05	*base-WordNet*	*78.25*
kuaw	79.89	Meaning-allwords	78.14
SVM-semWND	79.82	SVM-semSS	77.84
UNTaw	79.71	Meaning-simple	77.72
GAMBL-AW	79.62	SVM-semBLC20	77.70
upv-eaw2	79.27	SVM-semBLC50	77.70
upv-eaw	78.42	SVM-semSUMO	77.70
base-WordNet	*78.25*	kuaw	77.53
SVM-semSS	76.46	upv-eaw2	77.21

Table 7: Results for sense to semantic classes on SensEval3

SSI-Dijkstra

- A version of SSI algorithm (Navigli & Velardi 2004)
  - Graph-based (i.e. MCR)
  - all-words
  - all-languages (connected to WN)
  - Good results on topically related terms
  - Applications:
    - WSD on Topic Signatures => KnowNets
    - WSD on FrameNet => FrameWordNet
    - WSD on Wikipedia => JumboWNs

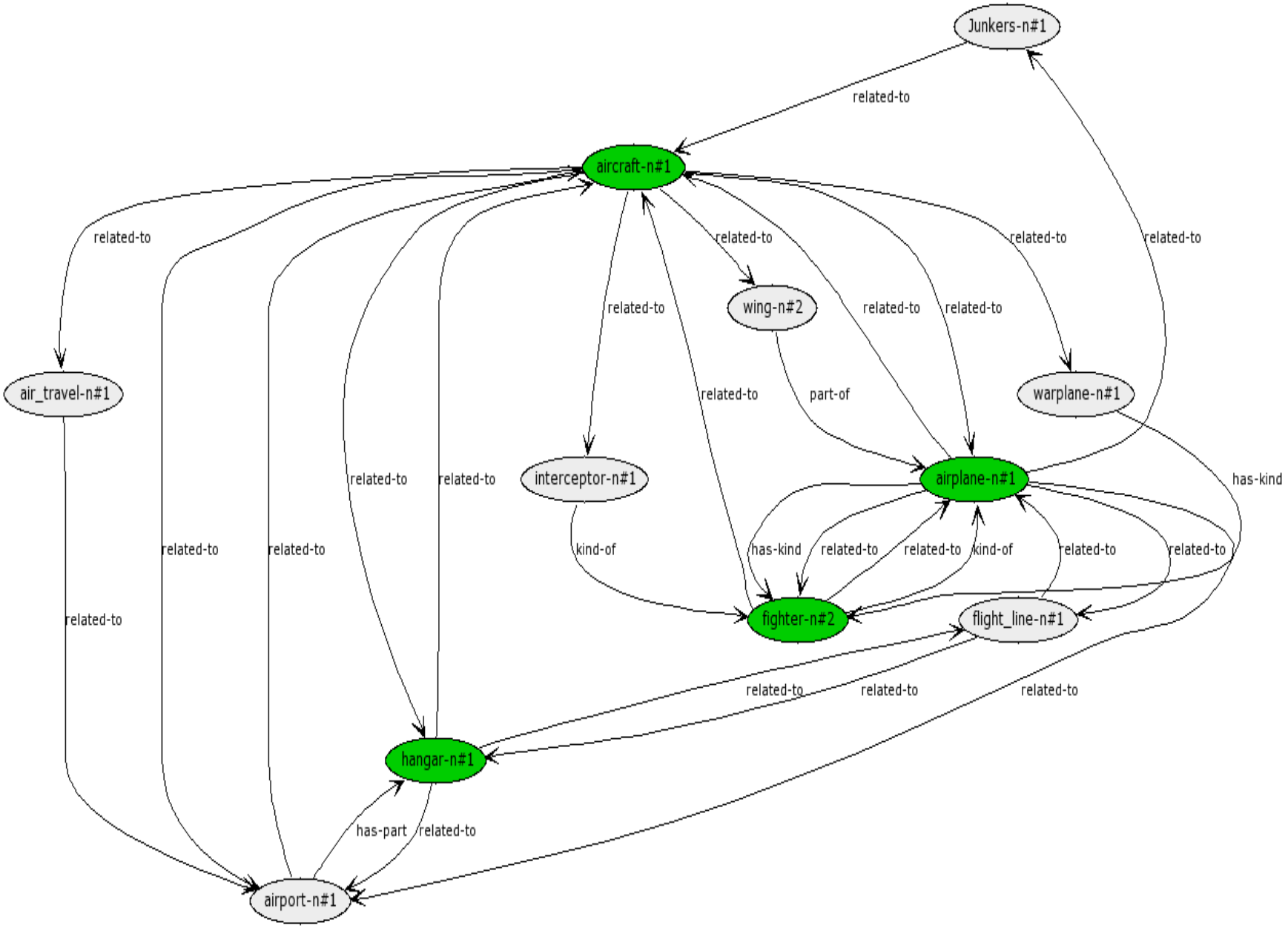
New Topic Signatures

- Use monosemous terms to obtain TS
  - Any corpora (BNC, WEB, Wikipedia, ...)
  - No query construction strategy!
  - TS for all monosemous words ...
- Disambiguate the words of the TS using SSI and the MCR
  - Disambiguated TS
  - Reversing the TSs to obtain TS for polysemous words!
- Millions of new relations!

New Topic Signatures

<airplane, aeroplane, plane>
aircraft (237.25)
fighter (94.71)
propeller (59.85)
hangar (52.09)
helicopter (51.04)
pilotless (50.19)
whirlybird (48.61)
...

New Topic Signatures



SSI-Dijkstra: KnowNet

word	offset	weight	gloss
flight#n	00195002n	0.017	a scheduled trip by plane between designated airports
travelling#n	00191846n	0	the act of going from one place to another
train#n	03528724n	0.012	a line of railway cars coupled together and drawn by a locomotive
passenger#n	07460409n	0	a person traveling in a vehicle (a boat or bus or car or plane or train etc) who is not operating it
station#n	03404271n	0.019	a building equipped with special equipment and personnel for a particular purpose
airport#n	02175180n	0	an airfield equipped with control tower and hangers as well as accommodations for passengers and cargo
ferry#n	02671945n	0.010	a boat that transports people or vehicles across a body of water and operates on a regular schedule
airfield#n	02171984n	0	a place where planes take off and land

Table 4. Sense disambiguated TS for airport#n#1 obtained from BNC using InfoMap and SSI-Dijkstra

SSI-Dijkstra: KnowNet

Source	#relations
Princeton WN3.0	235,402
Selectional Preferences from SemCor	203,546
eXtended WN	550,922
Co-occurring relations from SemCor	932,008
New KnowNet-5	436,997
New KnowNet-10	1,354,905
New KnowNet-15	2,731,273
New KnowNet-20	4,692,553

Table 1. Number of synset relations

SSI-Dijkstra: KnowNet

KB	P	R	F1	Av. Size
TRAIN	*65.1*	*65.1*	*65.1*	
TRAIN-MFS	*54.5*	*54.5*	*54.5*	
WN-MFS	*53.0*	*53.0*	*53.0*	
TSSEM	52.5	52.4	52.4	103
SEMCOR-MFS	*49.0*	*49.1*	*49.0*	
MCR^2	45.1	45.1	45.1	26,429
MCR	45.3	43.7	44.5	129
KnowNet-20	44.1	44.1	44.1	610
KnowNet-15	43.9	43.9	43.9	339
spSemCor	43.1	38.7	40.8	56
KnowNet-10	40.1	40.0	40.0	154
$(WN+XWN)^2$	38.5	38.0	38.3	5,730
WN+XWN	40.0	34.2	36.8	74
TSWEB	36.1	35.9	36.0	1,721
XWN	38.8	32.5	35.4	69
KnowNet-5	35.0	35.0	35.0	44
WN^3	35.0	34.7	34.8	503
WN^4	33.2	33.1	33.2	2,346
WN^2	33.1	27.5	30.0	105
spBNC	36.3	25.4	29.9	128
WN	44.9	18.4	26.1	14
RANDOM	*19.1*	*19.1*	*19.1*	

Table 7. P, R and F1 fine-grained results for the resources evaluated at Senseval-3, English Lexical Sample Task.

KB	P	R	F1	Av. Size
TRAIN	87.6	87.6	87.6	
TRAIN-MFS	81.2	79.6	80.4	
WN-MFS	66.2	59.9	62.9	
$(WN+XWN)^2$	54.9	51.1	52.9	5,153
TSWEB	54.8	47.8	51.0	700
KnowNet-20	49.5	46.1	47.7	561
KnowNet-15	47.0	43.5	45.2	308
XWN	50.1	39.8	44.4	96
KnowNet-10	44.0	39.8	41.8	139
WN+XWN	45.4	36.8	40.7	101
SEMCOR-MFS	42.4	38.4	40.3	
MCR	40.2	35.5	37.7	149
TSSEM	35.1	32.7	33.9	428
KnowNet-5	35.5	26.5	30.3	41
MCR^2	32.4	29.5	30.9	24,896
WN^3	29.3	26.3	27.7	584
RANDOM	27.4	27.4	27.4	
WN^2	25.9	27.4	26.6	72
spSemCor	31.4	23.0	26.5	51.0
WN^4	26.1	23.9	24.9	2,710
WN	36.8	16.1	22.4	13
spBNC	24.4	18.1	20.8	290

Table 8. P, R and F1 fine-grained results for the resources evaluated at SemEval-2007, English Lexical Sample Task.

SSI-Dijkstra: KnowNet

KB	P	R	F1	Av. S
TRAIN	*81.8*	*68.0*	*74.3*	962
MiniDir-MFS	*67.1*	*52.7*	*59.2*	
KnowNet-15	54.7	48.9	**51.6**	176
KnowNet-20	51.8	**49.6**	50.7	319
KnowNet-10	53.5	43.1	47.7	81
MCR	46.1	41.1	43.5	66
WN^2	56.0	29.0	42.5	51
$(WN+XWN)^2$	41.3	41.2	41.3	1,892
KnowNet-5	58.5	26.9	36.8	22
TSSEM	33.6	33.2	33.4	208
XWN	42.6	27.1	33.1	24
WN	**65.5**	13.6	22.5	8
RANDOM	**21.3**	**21.3**	**21.3**	

Cuadro 5: P, R and F1 fine-grained results for the resources evaluated individually on Spanish.

SSI-Dijkstra: KnowNet

KB	P	R	F1	Av. S
TRAIN	*81.8*	*68.0*	*74.3*	962
MiniDir-MFS	*67.1*	*52.7*	*59.2*	
KnowNet-15	54.7	48.9	**51.6**	176
KnowNet-20	51.8	**49.6**	50.7	319
KnowNet-10	53.5	43.1	47.7	81
MCR	46.1	41.1	43.5	66
WN^2	56.0	29.0	42.5	51
$(WN+XWN)^2$	41.3	41.2	41.3	1,892
KnowNet-5	58.5	26.9	36.8	22
TSSEM	33.6	33.2	33.4	208
XWN	42.6	27.1	33.1	24
WN	**65.5**	13.6	22.5	8
RANDOM	**21.3**	**21.3**	**21.3**	

Cuadro 5: P, R and F1 fine-grained results for the resources evaluated individually on Spanish.

SSI-Dijkstra: FrameWordNet

Lexical Unit	synset	#senses	Gloss
education.n	00567704-n	2	"activities that impart knowledge"
teacher.n	07632177-n	2	"a person whose occupation is teaching"
instruct.v	00562446-v	3	"impart skills or knowledge"
study.v	00410381-v	6	"be a student; follow a course of study; be enrolled at an institute of learning"
student.n	07617015-n	2	"a learner who is enrolled in an educational institution"
pupil.n	07617015-n	3	"a learner who is enrolled in an educational institution"

Table 1: Partial result of the WSD process of the LUs of the frame EDUCATION_TEACHING

	nouns			verbs			adjectives			all		
	P	R	F	P	R	F	P	R	F	P	R	F
VM	0.00	0.00	0.00	0.93	0.66	0.77	0.00	0.00	0.00	0.93	0.34	0.50
wn-mfs	0.75	0.75	0.75	0.64	0.64	0.64	0.80	0.80	0.80	0.69	0.69	0.69
ukb	0.70	0.69	0.70	0.68	0.68	0.68	0.84	0.84	0.84	0.71	0.71	0.71
SSI-Dijkstra	**0.84**	0.65	0.73	0.70	0.56	0.62	**0.90**	0.82	0.86	**0.78**	0.63	0.69
FSI	0.80	**0.77**	**0.79**	0.66	0.65	0.65	0.89	**0.89**	**0.89**	0.74	0.73	0.73
ASI	0,80	**0,77**	**0,79**	0,67	0,65	0,66	0,89	**0,89**	**0,89**	0,75	0,73	0,74
FSP	0.75	0.73	0.74	**0.71**	**0.69**	**0.70**	0.79	0.79	0.79	0.73	0.72	0.72
ASP	0.72	0.69	0.70	0.68	0.66	0.67	0.75	0.75	0.75	0.70	0.69	0.69
SSI-Dijkstra+	0.79	0.77	0.78	0.70	0.68	0.69	0.89	**0.89**	**0.89**	0.76	**0.74**	**0.75**

Table 3: Results of the different SSI algorithms on the *GS* dataset

SSI-Dijkstra: FrameWordNet

	nouns			verbs			adjectives			all		
	P	R	F	P	R	F	P	R	F	P	R	F
VM	0.00	0.00	0.00	0.93	0.80	0.86	0.00	0.00	0.00	0.93	0.36	0.52
wn-mfs	0.76	0.76	0.76	0.61	0.61	0.61	0.76	0.76	0.76	0.69	0.69	0.69
ukb	0.81	0.81	0.81	0.62	0.62	0.62	0.82	0.82	0.82	0.72	0.72	0.72
SSI-Dijkstra	**0,86**	0,78	0,82	0,66	0,63	0,64	**0,88**	0,85	0,87	**0,77**	0,72	0,75
FSI	0,85	**0,85**	**0,85**	0,64	0,64	0,64	**0,88**	**0,88**	**0,88**	0,76	0,76	0,76
ASI	0,85	**0,85**	**0,85**	0,64	0,64	0,64	**0,88**	**0,88**	**0,88**	0,76	0,76	0,76
FSP	0,81	0,81	0,81	**0,67**	**0,67**	**0,67**	0,74	0,74	0,74	0,73	0,73	0,73
ASP	0,76	0,76	0,76	0,63	0,63	0,63	0,71	0,71	0,71	0,69	0,69	0,69
SSI-Dijkstra+	0.85	**0.85**	**0.85**	**0.67**	**0.67**	**0.67**	**0.88**	**0.88**	**0.88**	**0.77**	**0.77**	**0.77**

Table 4: Results of the different SSI algorithms on frames having at least 10 LUs

	P	R	F
mfs-wn	0.67	0.67	0.67
ukb	0.76	0.76	0.76
SSI-Dijkstra	**0.79**	0.74	0.76
FSI	0.78	0.78	0.78
ASI	0.78	0.77	0.78
FSP	0.72	0.71	0.71
ASP	0.70	0.70	0.70
SSI-Dijkstra+	**0.79**	**0.79**	**0.79**

Table 5: Results using FrameNet–WordNet Verbal mapping from (Shi and Mihalcea, 2005) as gold standard

SSI-Dijkstra: JumboWN (or WikiWN)

[rigau@adimen MCRGraphDistances]$ ./SSI.v4.pl
Reading Graph from file ...
Monosemous: Pablo_Picasso|n 1
Polysemous: sculpture|n 2
Monosemous: Spanish_people|n 1
Polysemous: painter|n 3
Polysemous: art|n 4
Monosemous: cubism|n 1
sculpture|n 00599509-n 0.3333 3
painter|n 07453414-n 0.5208 4
art|n 00598038-n 0.5 5
Interpretation: cubism n 06252762-n 0 0 an artistic movement in France beginning in 1907 that ...
Interpretation: painter n 07453414-n 0.5208 4 an artist who paints
Interpretation: Pablo_Picasso n 07747579-n 0 0 1881-1973
Interpretation: sculpture n 00599509-n 0.3333 3 making figures or designs in three dimensions
Interpretation: Spanish_people n 07039306-n 0 0 the people of Spain
Interpretation: art n 00598038-n 0.5 5 the creation of beautiful or significant things

SSI-Dijkstra: JumboWN (or WikiWN)

enJumboWN				WN
<Picasso, Pablo_Picasso, ...>			=	07747579-n
<Picasso, Pablo_Picasso, ...>	->	<cubism, ...>	=	06252762-n
<Picasso, Pablo_Picasso, ...>	->	<painter, ...>	=	07453414-n
<Picasso, Pablo_Picasso, ...>	->	<sculpture, ...>	=	00599509-n
<Picasso, Pablo_Picasso, ...>	->	<Spanish_people, ...>	=	07039306-n
<Picasso, Pablo_Picasso, ...>	->	<art, ...>	=	00598038-n
<Picasso, Pablo_picasso, ...>	->	<George_Braque, ...>	=	?

...

enJumboWN		esJumboWN		WN
<Picasso, Pablo_Picasso, ...>	=	<Picasso, Pablo_Picasso, ...>	=	07747579-n
<cubism, ...>	=	<cubismo, ...>	=	06252762-n
<painter, ...>	=	<pintor, ...>	=	07453414-n

...

WordNet Extensions

German Rigau i Claramunt

german.rigau@ehu.es

IXA group

Departamento de Lenguajes y Sistemas Informáticos

UPV/EHU