Predicate Matrix: extending SemLink through WordNet mappings

Maddalen López de Lacalle,
Egoitz Laparra and German Rigau
IXA group, EHU/UPV
http://ixa.si.ehu.es

LREC 2014, Reykyavik, Island
28/05/2014

Introduction

- **Predicate models** are core resources in most advanced NLP tasks (QA, TE, IE...)
- **Natural Language Understanding** capabilities require a large and precise amount of semantic **knowledge at the predicate argument level**
  - Shallow semantic parsing and explicit and implicit semantic role labeling *(Erk and Pado, 2004), (Shi and Mihalcea, 2005), (Giuglea and Moschitti, 2006), (Laparra and Rigau, 2013)*

Introduction

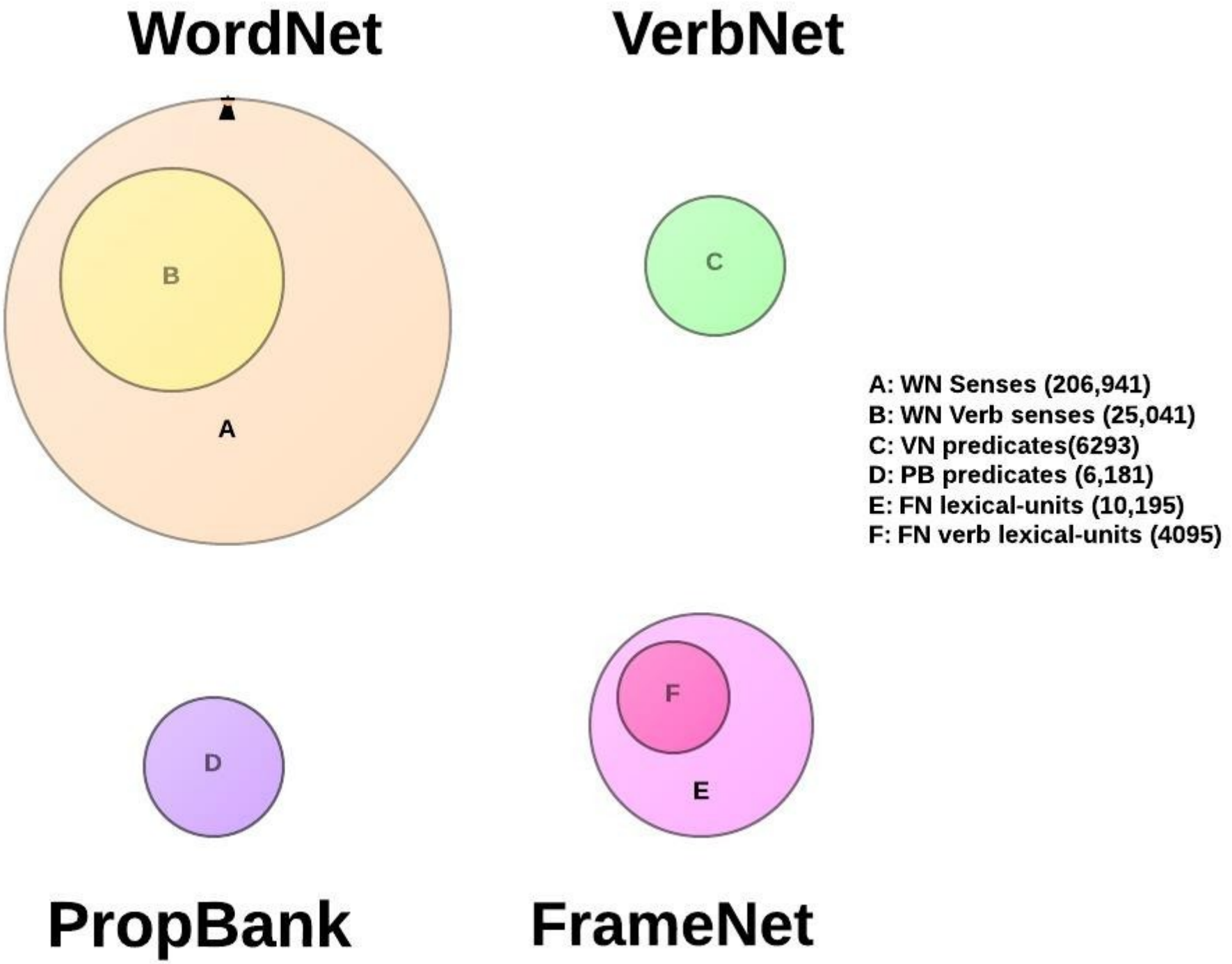
- **Building** large and rich enough predicate models takes a great deal of **expensive manual effort**
- **Coverage** of currently available predicate-argument resources is still **far from being complete** *(Burchardt et al., 2005),(Shen and Lapata, 2007)*
- **Same effort** should be invested for each **different language** *(Subirats and Petruck, 2003)*

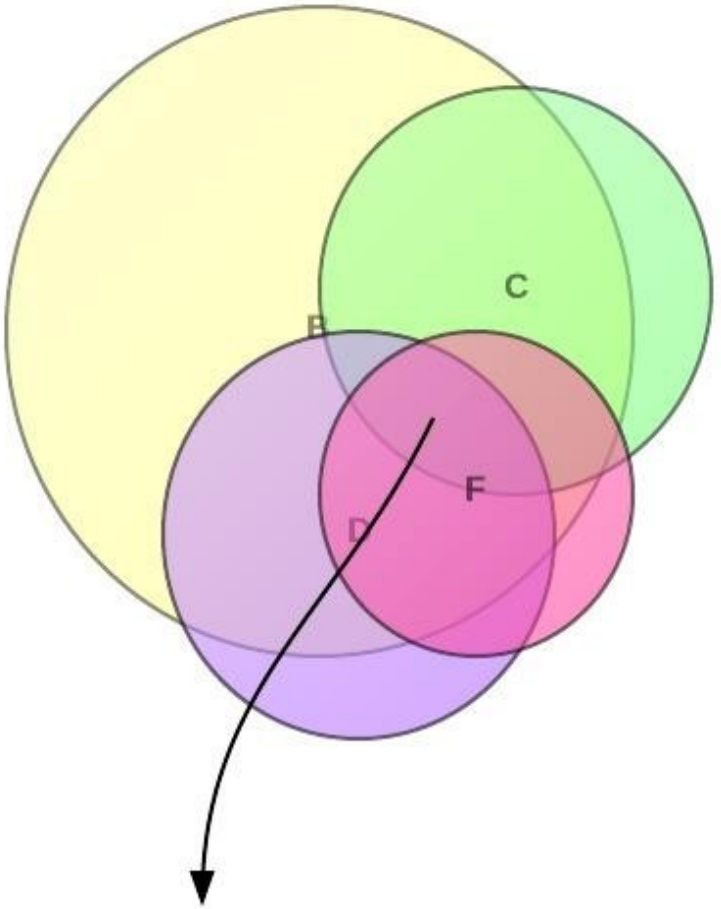
Introduction

- Most previous research efforts on the **integration of resources** targeted on **nouns and named entities**:
  - *YAGO (Suchanek et al., 2007), Freebase (Bollacker et al., 2008), DBPedia (Bizer et al., 2009), BabelNet (Navigli and Ponzetto, 2010) or UBY (Gurevych et al., 2012)*
- Works on integration of **predicate information**:
  - *(Shi andMihalcea, 2005), (Burchardt et al., 2005), (Johansson and Nugues, 2007), (Pennacchiotti et al., 2008), (Cao et al., 2008), (Tonelli and Pianta, 2009), (Laparra et al., 2010)*

Sources of Predicate Information

- **SemLink** (Palmer, 2009)
  - SemLink aim is to **connect together different predicate resources**.
  - SemLink provides **partial mappings** to:
    - VerbNet (Kipper, 2005)
    - PropBank (Palmer et al.,2005)
    - FrameNet (Baker et al., 1997)
    - WordNet (Fellbaum, 1998)
  - However, its coverage is still far from complete


WordNet
VerbNet
B
A
C
A: WN Senses (206,941)
B: WN Verb senses (25,041)
C: VN predicates(6293)
D: PB predicates (6,181)
E: FN lexical-units (10,195)
F: FN verb lexical-units (4095)
D
F
E
PropBank
FrameNet



B: WN Verb senses (25,041)
C: VN predicates(6,293)
D: PB predicates (6,181)
F: FN verb lexical-units (4,095)

SemLink < UNION (B,C,D,E)

SemLink Coverage

- **WN vs VN**: most of the WN senses (74%) are not aligned to VN
- **WN vs VN**: some VN lemmas not in WN
- **PB vs VN**: all lemmas of PB are contained in VN lexicon
- **PB vs VN**: about the half of the roles/arguments not mapped in both directions
- **FN vs VN**: only 16% of LUs aligned to a VN predicate
- **FN vs VN**: 88% of the FEs are not aligned to any VN thematic-role (and hence to PB)

Predicate Matrix

- **Predicate Matrix** is a new lexical resource resulting from the integration of multiple sources of predicate information: *FN,VN, PB and WN (and SemLink)*
- We expect to provide a more robust **interoperable lexicon** across *FN,VN, PB and WN*
- Discover and solve **inherent inconsistencies** among the integrated resources
- Extend the **coverage** of current predicate resources
- To extend predicate information to **languages other than English**

Predicate Matrix

- **Predicate Matrix** uses WN as a **central resource**.
- Each row (or line) in the matrix presents a (partial) **role alignment** of a particular WN word sense

Predicate Matrix

- First version 1.0 (GWC 2014)
  - SemLink +
  - Monosemous verbs from VN +
  - Synonyms from WN

- Second version 1.1 (LREC 2014)
  - SemLink +
  - Automatic mappings between predicates +
    - WN-VN and WN-FN (**new mappings!**)
  - Project VN roles to FN roles (**complete gaps!**) +
  - Synonyms from WN

Mapping approach

- Following (Laparra et al 2010):
  - We use **graph-based WSD algorithms**
    - WN as a graph (WN3.0 + glosses)
    - UKB (Agirre & Soroa 2009)
    - SSI-Dijkstra+ (Laparra et al. 2010)
  - We use as **context**:
    - all predicates of the same VN class
    - all LUs of the same FN frame

Mapping approach

- **FrameNet**
- *Frame*: Education teaching
- *LUs*: coach.v, cram.v, educate.v, educational.a, education.n, graduate.n, instruction.n, instruct.v, learn.v, lecturer.n, master.v, professor.n, protege.n, pupil.n, schoolmaster.n, schoolmistress.n, schoolteacher.n, school.v, student.n, study.v, teacher.n, teach.v, training.n, train.v, tutee.n, tutor.n, tutor.v

Mapping approach

- **FrameNet**
- *Frame*: Education teaching
- *LU*: coach.v, cram.v, educate.v, educational.a, education.n, graduate.n, instruction.n, **instruct.v**, **learn.v**, lecturer.n, master.v, professor.n, protege.n, pupil.n, schoolmaster.n, schoolmistress.n, schoolteacher.n, school.v, student.n, study.v, teacher.n, **teach.v**, training.n, train.v, tutee.n, tutor.n, tutor.v

- **WordNet**
- *Synset* 00829107-v impart skills or knowledge to: *I taught them French; He instructed me in building a boat*;
  - instruct%2:32:00 learn%2:32:00 teach%2:32:00

Mapping approach

- **Average lenght** of contexts:
  - VN: 23.30 verbs
  - FN 19.38 LUs

- As a **gold standard** we used:
  - 272 VN classes
  - 214 FN frames
  - with at least one WN sense manually assigned

Mapping approach

VerbNet	Method	LKB	P	R	F1
	UKB	WN	84.2	84.2	84.2
	UKB	WN+G	**85.3**	**85.3**	**85.3**
	SSID+	WN	83.8	83.5	83.7
	SSID+	WN+G	83.8	83.5	83.7
FrameNet	Method	LKB	P	R	F1
	UKB	WN	79.0	79.0	79.0
	UKB	WN+G	79.4	79.4	79.4
	SSID+	WN	82.5	81.3	81.9
	SSID+	WN+G	**82.9**	**81.8**	**82.4**

Table 1: Resuls of the disambiguation process.

Mapping Approach

- We already have **new** predicate aligments like

  vn:learn-14 learn%2:31:02 fn:Education_teaching

	VN-FN	VN-WN	FN-WN	VN-WN-FN
SemLink	3,285	7,620	4,342	5,168
New links	4,712	8,504	6,338	6,745

Table 2: New links added to the mapping between the lexicons.

New role aligments

- We already have new predicate aligments like

  vn:learn-14 learn%2:31:02 fn:Education_teaching

- What about their **roles**?
  - VN: *Agent, Topic, Source*
  - FN: *Teacher, Subject, Student, Means, Manner*
  - From SemLink we already have that
    - vn:Agent fn:Student
    - vn:Source fn:Teacher
    - vn:Topic fn:Subject

New role aligments

- What about *paddle.v*?
  - VN (spank-18.3): *Agent, Patient, Instrument, Location, Result*
  - FN (Corporal_punishment): *Agent, Evaluee, Reason, Instrument, Degree, Body part, etc.*
  - Which FE is aligned the VN role *Location*?

- We used **three simple methods** based on frequencies

New role aligments

- Example **method 1**: *paddle.v*
  - Which FE is aligned the VN role *Location*?
  - when the available FEs are *Agent, Evaluee, Reason, Instrument, Degree, Body part*, ...

Thematic-Role	FrameElement	Freq.
Location	Area	285
Location	Goal	228
Location	Path	99
Location	Sound_source	73
Location	Ground	54
Location	Source	49
Location	Location	23
Location	Body_part	21

Table 4: Frequency of frame-elements mapped to the thematic-role *Location*.

New role aligments

lemma	VN-class	Thematic-Role	FN-frame	FrameElement
sit	spatial_configuration-47.6	Location	Placing	Area
spew	substance_emission-43.4	Location	Excreting	Goal
move	roll-51.3.1	Location	Change_position_on_a_scale	Path
paddle	spank-18.3	Location	Corporal_punishment	Body_part

Table 5: Examples of new frame-elements mapped to the thematic-role *Location*.

New role aligments

- Example **method 2**: *feel.v*
  - VN (see-30.1): *Experiencer, Stimulus*
  - FN (Seeking): *Cognizer_agent, Sought_entity, Ground*
  - We use the example patterns from VN and FN

Source	Class/Frame	Pattern				Freq.
VerbNet	see-30.1	Experiencer	v	Stimulus		100%
FrameNet	Seeking	Cognizer_agent	v	Sought_entity	Ground	80%
		Sought_entity	v	Ground	Cognizer_agent	20%

Table 6: Frequency of role patterns in VN class *see-30.1* and frame *Seeking*.

New role aligments

- **Method 3**:
  - Same as method 1 after applying method 2
  - Now, new role aligments have been acquired
  - To solve new cases not solved by method 1

New role aligments

Method	New	Total	P	R	F
SemLink	-	6,201	-	-	-
Method 1	6,686	12,887	**88.2**	**88.2**	**88.2**
Method 2	1,088	13,975	76.0	48.6	59.3
Method 3	1,193	15,168	80.6	80.6	80.6

Table 3: Number of new role alignments and performances of the different methods.

Adding WN synonyms

- Example:
  - We already have predicate aligments like
    - vn:leave-51.2-1 desert%2:31:00 fn:Departing
  - In WN desert%2:31:00 has tree **synonyms**
    - abandon%2:31:00, forsake%2:31:00, desolate%2:31:00
  - We **port** the predicate information to the rest of word senses

Predicate Matrix 1.1

	lexicon				roles
	VN-FN	VN-WN	FN-WN	VN-WN-FN	
SemLink	3,285	7,620	4,342	5,168	6,201
New links	9,952	14,900	11,391	12,267	31,889

Table 9: Number of new lexicon and role alignments after the extension to WordNet synonyms.

Predicate Matrix 1.1

- The aligment to **PropBank** also improves

	lexicon			roles	
	PB-VN	PB-WN	PB-FN	PB-VN	PB-FN
SemLink	5,292	6,174	2,462	12,164	1,461
New-Links	11,990	12,249	2,901	23,498	1,923

Table 10: Number of new lexicon and role alignments for PropBank.

Predicate Matrix 1.1

LF	WN senses	not in VN (%)	not in FN (%)	LF name
29	1,130	549 (48.58)	794 (70.27)	body
30	4,171	2,561 (61.40)	3,393 (81.35)	change
31	1,404	828 (58.97)	1,053 (75.00)	cognition
32	3,120	1,723 (55.22)	2,161 (69.26)	communication
33	733	518 (70.67)	595 (81.17)	competition
34	476	266 (55.88)	363 (76.26)	consumption
35	3,698	1,833 (49.57)	2,716 (73.45)	contact
36	1,151	718 (62.38)	898 (78.02)	creation
37	763	228 (29.88)	491 (64.35)	emotion
38	2,491	1,257 (50.46)	1,731 (69.49)	motion
39	820	372 (45.37)	548 (66.83)	perception
40	1,431	834 (58.28)	1,116 (77.99)	possession
41	2,202	1,372 (62.31)	1,647 (74.80)	social
42	1,409	881 (62.53)	1,084 (76.93)	stative
43	146	57 (39.04)	80 (54.79)	weather

Table 11: WordNet verbal senses not covered by VerbNet classes and FrameNet frames in the Predicate Matrix v1.1

Conclusions and Future work

- We extend **SemLink** by automatic means
- This is an **ongoing work** towards a more **complete**, **robust** and **interoperable** verbal lexicon
- http://adimen.si.ehu.es:/web/PredicateMatrix
- Plans:
  - Extending PM through WN hierarchies
  - Acquiring new aligments induced from corpora
  - Extending PM to nominal predicates
  - Extending to other languages
  - Exploiting the PM

Predicate Matrix: extending SemLink through WordNet mappings

Maddalen López de Lacalle,
Egoitz Laparra and German Rigau
IXA group, EHU/UPV
http://ixa.si.ehu.es

LREC 2014, Reykyavik, Island
28/05/2014