

Exploring the Integration of WordNet and FrameNet

Egoitz Laparra Montse Cuadros German Rigau

Euskal Herriko Unibertsitatea (EHU/UPV)
IXA NLP group

February 2010

Outline

- 1 Introduction
- 2 SSI-algorithms
- 3 WordFrameNet
- 4 Future Work

Outline

- 1 Introduction
- 2 SSI-algorithms
- 3 WordFrameNet
- 4 Future Work

FrameNet I

- Lexical resource based on Frame Semantics (Fillmore, 1976)
- Each Frame corresponds to a particular event.

EDUCATION_TEACHING

- Lexical Units: A set of words that evokes the event.
student.n, teacher.n, learn.v, study.v
- Frame Elements: The typical participants (roles) of the event. STUDENT, SUBJECT, TEACHER

[Bernard Lansky] <i>STUDENT</i> <i>studied</i> [the piano] <i>SUBJECT</i> [with Peter Wallfisch] <i>TEACHER</i> .

FrameNet II

- Different senses for a word are represented by assigning different frames.
 - study.v
 - Def: Acquire knowledge on
 - Frame: EDUCATION_TEACHING
 - study.v
 - Def: Look at closely in order to observe or read.
 - Frame: SCRUTINY

FrameNet III

FrameNet 1.3

- 825 frames
- Around 10,000 lexical-units
 - Only 722 frames with a least one LU asociated
 - Only 9,360 LUs recognized by WN corresponding to only 708 frames
- The limited coverage is one of the main problems of this resource
 - WN 3.0 has more than 206,000 senses

Goal: Mapping LUs to WordNet synsets

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, teacher.n ...

WordNet

synset: 10694258n teacher#n#1, instructor#n#1

Goal: Mapping LUs to WordNet synsets

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, **teacher.n** ...

WordNet

synset: 10694258n **teacher#n#1**, instructor#n#1

Goal: Mapping LUs to WordNet synsets

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, **teacher.n** ...

WordNet

synset: 10694258n **teacher#n#1**, instructor#n#1

- We use a Knowledge based WSD algorithm
- We use as context all LUs of the same Frame

Motivation

By doing this integration process we expect:

- To extend the coverage of FrameNet
- To enrich WordNet with frame semantic information
- To extend FrameNet to other languages

Motivation: Extend the coverage of FrameNet

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, teacher.n ...

Motivation: Extend the coverage of FrameNet

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, **teacher.n** ...

Motivation: Extend the coverage of FrameNet

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, **teacher.n** ...

WordNet

synset: 10694258n **teacher#n#1**, instructor#n#1

Motivation: Extend the coverage of FrameNet

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, **teacher.n** ...

WordNet

synset: 10694258n **teacher#n#1**, **instructor#n#1**

Motivation: Extend the coverage of FrameNet

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, **teacher.n**, **instructor.n** ...

WordNet

synset: 10694258n **teacher#n#1**, **instructor#n#1**

Motivation: Enrich WordNet with Frame information

WordNet

synset: 10694258n teacher#n#1, instructor#n#1

synset: 10668698n student#n#1, pupil#n#1

Motivation: Enrich WordNet with Frame information

WordNet

synset: 10694258n teacher#n#1, instructor#n#1

No direct relation!!

synset: 10668698n student#n#1, pupil#n#1

Motivation: Enrich WordNet with Frame information

WordNet

synset: 10694258n teacher#n#1, instructor#n#1

No direct relation!!

synset: 10668698n student#n#1, pupil#n#1

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, student.n, pupil.n, teacher.n,...

Motivation: Enrich WordNet with Frame information

WordNet

synset: 10694258n teacher#n#1, instructor#n#1

No direct relation!!

synset: 10668698n student#n#1, pupil#n#1

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, student.n, pupil.n, teacher.n...

Motivation: Enrich WordNet with Frame information

WordNet

synset: 10694258n teacher#n#1, instructor#n#1

New Education-teaching relation

synset: 10668698n student#n#1, pupil#n#1

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, student.n, pupil.n, teacher.n...

Motivation: Extend FrameNet to other languages

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, student.n, pupil.n, teacher.n ...

Motivation: Extend FrameNet to other languages

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, [student.n](#), [pupil.n](#), teacher.n ...

English WordNet: 10668698n [student#n#1](#), [pupil#n#1](#)

Motivation: Extend FrameNet to other languages

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, **student.n**, **pupil.n**, teacher.n ...

English WordNet: 10668698n **student#n#1**, **pupil#n#1**

Spanish WordNet: 10668698n **estudiante#n#1**, **alumno#n#1**

Italian WordNet: 10668698n **studente#n#1**, **alunno#n#1**

Basque WordNet: 10668698n **ikasle#n#1**

Outline

- 1 Introduction
- 2 SSI-algorithms**
- 3 WordFrameNet
- 4 Future Work

SSI-Dijkstra(Cuadros & Rigau, 2008)

Based on Structural Semantic Interconnections algorithm (SSI)
(Navigli & Velardi, 2005)

- Uses a very large connected graph with 99,635 nodes (synsets) and 636,077 edges corresponding to direct relations between synsets gathered from WordNet 1.6 and eXtended WordNet.
- Computes Dijkstra algorithm for finding the shortest path distance between one node and the rest of nodes of the graph.

SSI-Dijkstra(Cuadros & Rigau, 2008)

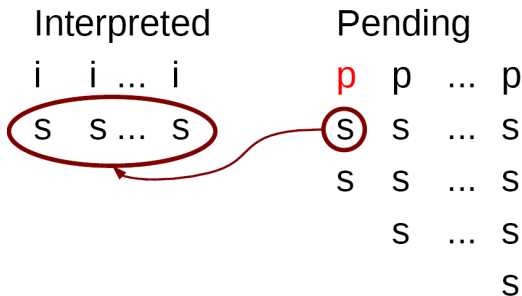
Interpreted

i i ... i
s s ... s

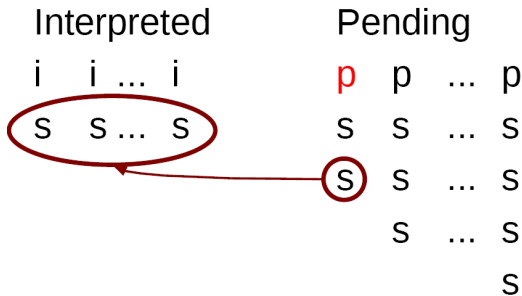
Pending

p p ... p
s s ... s
s s ... s
 s ... s
 s

SSI-Dijkstra(Cuadros & Rigau, 2008)



SSI-Dijkstra(Cuadros & Rigau, 2008)



SSI-Dijkstra(Cuadros & Rigau, 2008)

Interpreted

i i ... i
s s ... s

Pending

p p ... p
s s ... s
Ⓢ s ... s
 s ... s
 s

SSI-Dijkstra(Cuadros & Rigau, 2008)

Interpreted

i i ... i p
s s ... s **S**

Pending

p ... p
s ... s
s ... s
s ... s
s

FSI & ASI

- At least one monosemous word is required in the context
- Some Frames only have polysemous LUs
- We devised two different approaches to start-up the process

FSI

Interpreted

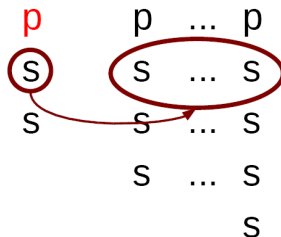
Pending

p	p	...	p
S	S	...	S
S	S	...	S
	S	...	S
			S

FSI

Interpreted

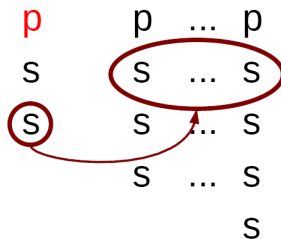
Pending



FSI

Interpreted

Pending



FSI

Interpreted

p

s

Pending

p ... p

s ... s

s ... s

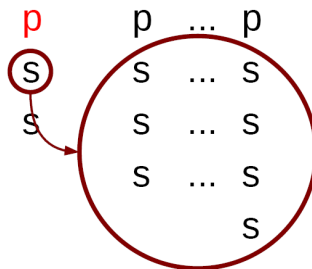
s ... s

s

ASI

Interpreted

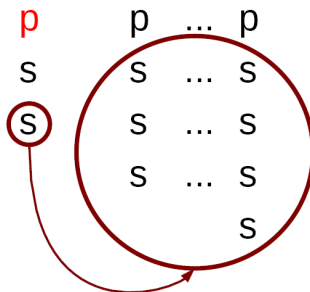
Pending



ASI

Interpreted

Pending



ASI

Interpreted

p

s

Pending

p ... p

s ... s

s ... s

s ... s

s

FSP & ASP

- The remaining words in P may contribute in the disambiguation process
- We devised two different approaches following the F and A strategies

FSP

Interpreted

i i ... i

s s ... s

Pending

p p ... p

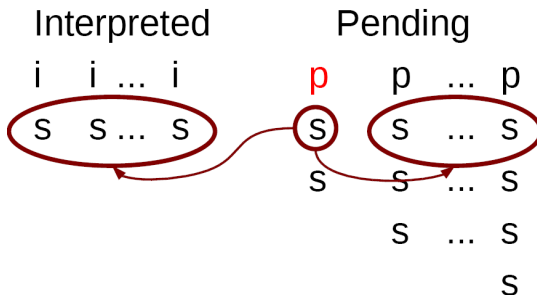
s s ... s

s s ... s

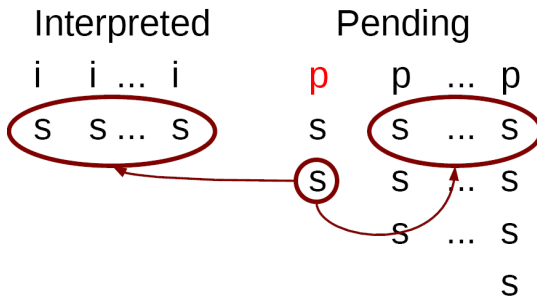
s s ... s

s

FSP



FSP



FSP

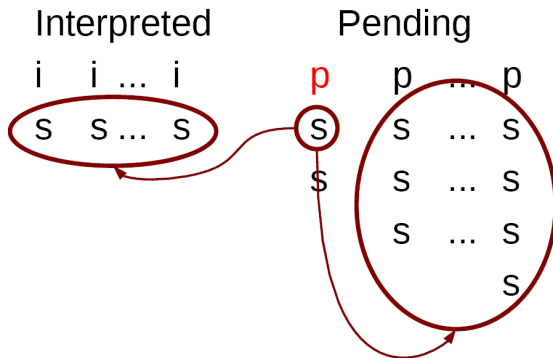
Interpreted

i i ... i p
s s ... s **S**

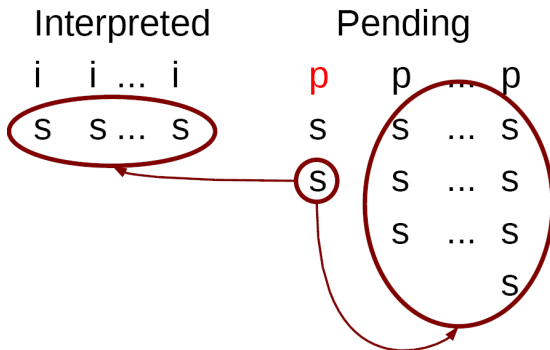
Pending

p ... p
s ... s
s ... s
s ... s
s

ASP



ASP



ASP

Interpreted

i i ... i p
s s ... s **S**

Pending

p ... p
s ... s
s ... s
s ... s
s

Evaluation I

- We have evaluated the different versions of the algorithm using a dataset provided by Sara Tonelli (FBK, Trento)
 - 372 LUs corresponding to 372 different frames from FrameNet 1.3
 - Manually annotated with WN 1.6 synsets
 - 9 frames with only one LU \Rightarrow No context words
 - Our approach produces no results

Evaluation II

Results of the different SSI algorithms on the GS dataset

	nouns			verbs			adjectives		
	P	R	F	P	R	F	P	R	F
wn-mfs	0.75	0.75	0.75	0.64	0.64	0.64	0.80	0.80	0.80
SSI-Dijkstra	0.84	0.65	0.73	0.70	0.56	0.62	0.90	0.82	0.86
FSI	0.80	0.77	0.79	0.66	0.65	0.65	0.89	0.89	0.89
ASI	0.80	0.77	0.79	0.67	0.65	0.66	0.89	0.89	0.89
FSP	0.75	0.73	0.74	0.71	0.69	0.70	0.79	0.79	0.79
ASP	0.72	0.69	0.70	0.68	0.66	0.67	0.75	0.75	0.75

SSI-Dijkstra+ I

- SSI-Dijkstra+
 - Combines ASI and FSP algorithms
 - Applies ASI approach to nouns and adjectives
 - Applies FSP approach to verbs

SSI-Dijkstra+ II

Results of the different SSI algorithms on the GS dataset

	all		
	P	R	F
wn-mfs	0.69	0.69	0.69
SSI-Dijkstra	0.78	0.63	0.69
FSI	0.74	0.73	0.73
ASI	0,75	0,73	0,74
FSP	0.73	0.72	0.72
ASP	0.70	0.69	0.69
SSI-Dijkstra+	0.76	0.74	0.75

- To our knowledge, SSI-Dijkstra+ presents the best results in this dataset.

Outline

- 1 Introduction
- 2 SSI-algorithms
- 3 WordFrameNet**
- 4 Future Work

WordFrameNet¹

Frame: EDUCATION_TEACHING		
train.v	00407541-v	prepare for a future task or career
instruct.v	00562446-v	impart skills or knowledge to
educational.a	02716766-a	relating to the process of education
tutee.n	07654181-n	learns from a tutor
schoolteacher.n	07551404-n	a teacher in a school below the college level
educate.v	00407541-v	prepare for a future task or career
study.v	00405251-v	be a student of a certain subject
instruction.n	00567704-n	activities that impart knowledge
teacher.n	07632177-n	a person whose occupation is teaching
student.n	07617015-n	a learner who is enrolled in an educational institution
schoolmistress.n	07550942-n	a woman schoolteacher
tutor.v	00562981-v	be a tutor to someone; give individual instruction
lecturer.n	07367816-n	someone who lectures professionally
...		

Table: LUs corresponding to EDUCATION_TEACHING frame

¹<http://adimen.si.ehu.es/WordFrameNet>

WordFrameNet

- Extend the coverage of FrameNet
- Enrich WordNet with frame semantic information
- Extend FrameNet to other languages

Extend the coverage of FrameNet I

Extend the coverage of FrameNet

Example:

educate.v : 00407541-v prepare for a future task or career

- 00407541-v -> develop.v
- 00407541-v -> prepare.v
- 00407541-v -> educate.v

Original FrameNet has 9,328 LUs corresponding to 6,565 synsets

WordFrameNet has 20,587 LUs

Extend the coverage of FrameNet II

- Transport to the disambiguated LUs the knowledge currently available from other semantic resources associated to WordNet
 - SUMO(Niles & Pease,2001), WordNet Domains(Magnini & Cavaglia,2000), etc
- Now the LU corresponding to *student.n* can also have associated the SUMO label *SocialRole* and its corresponding logical axioms, and the WordNet Domains *school* and *university*.

Enrich WordNet with frame information

WordNet

synset: 10694258n teacher#n#1, instructor#n#1

New Education-teaching relation

synset: 10668698n student#n#1, pupil#n#1

FrameNet

Frame: Education_teaching

LU: coach.v, educate.v, teach.v, study.v, education.n, educational.a, student.n, pupil.n, teacher.n...

Enrich WordNet with frame information

- 124,718 new semantic relations between the original 6,565 synsets.
- 121,813 of these relations that connect synsets of the same frame do not appear in WordNet

Extend FrameNet to other languages

Spanish LUs inferred for EDUCATION_TEACHING frame

train.v	00407541-v	adistrar.v amastrar.v
instruct.v	00562446-v	enseñar.v instruir.v
student.n	07617015-n	alumna.n alumno.n estudiante.n

- Current version of the Spanish FrameNet (Subirats & Petruck, 2003) consists of 308 frames with 1,047 LUs
- WordFrameNet for Spanish has 14,106 LUs

Outline

- 1 Introduction
- 2 SSI-algorithms
- 3 WordFrameNet
- 4 Future Work**

Future work I

- Improve the results of the desambiguation
 - Use the definitions of the LUs to enrich the context of each frame
- Complete the integration
 - Disambiguate the Frame Elements using a similar approach

Future work II

Thanks!