EVALITA 2011: FLAIT
Frame Labeling over Italian Texts

Roberto Basili, Alessandro Lenci, Diego Decao, Giulia Venturi, Alessandro Moschitti
(UniTorV, UniPi, ILC/SSA,UniTN)

Evalita 2011, Roma 24th January 2012

OUTLINE

- Motivation and Definitions
- Basic resources
- Task Set-up
- Participant systems
- Results
- Conclusions

FRAME SEMANTICS
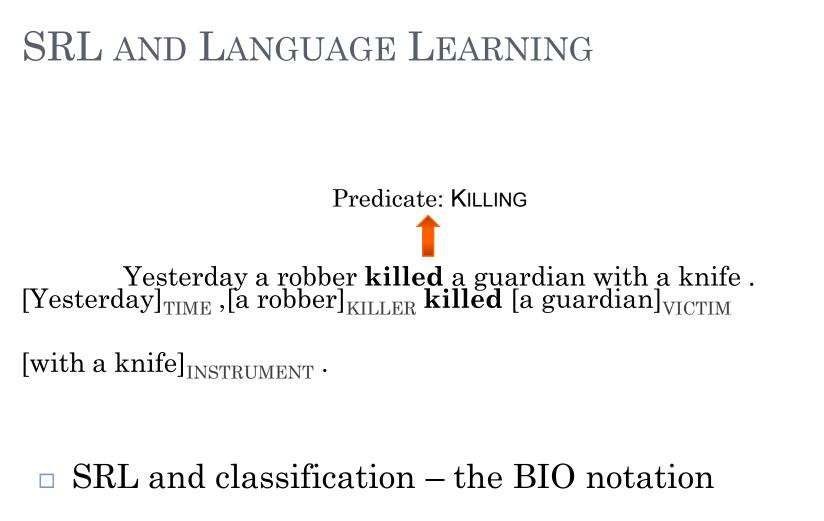
An example: the KILLING frame

Frame: KILLING
A KILLER or CAUSE causes the death of the VICTIM.

Frame Elements

CAUSE	**The rockslide** <u>killed</u> nearly half of the climbers.
INSTRUMENT	It's difficult to <u>suicide</u> **with only a pocketknife**.
KILLER	**John** <u>drowned</u> Martha.
MEANS	The flood <u>exterminated</u> the rats **by cutting off access to food**.
VICTIM	John <u>drowned</u> **Martha**.

Predicates

annihilate.v, annihilation.n, asphyxiate.v,assassin.n, assassinate.v, assassina-tion.n, behead.v, beheading.n, blood-bath.n, butcher.v, butchery.n, carnage.n, crucifixion.n, crucify.v, deadly.a, decapitate.v, decapitation.n, destroy.v, dis-patch.v, drown.v, eliminate.v, euthanasia.n, euthanize.v, . . .

FLAIT: THE SRL TASK

- **Frame Prediction (FP)**. Detecting the correct frame of a sentence given the presence of a possibly ambiguous lexical unit.

- **Semantic Role Labeling**. Annotate an entire sentence given a lexical unit with its frame information.
  - **Argument** or **Boundary Detection (BD)**. Locate all the semantic arguments of a frame realized in a sentence given a lexical unit.
  - **Argument Classification (AC)**. Assign the proper role, i.e. Frame Element, to every detected or given argument.

SRL and Language Learning

Predicate: KILLING

Yesterday a robber **killed** a guardian with a knife .

[Yesterday]$_{TIME}$,[a robber]$_{KILLER}$ **killed** [a guardian]$_{VICTIM}$

[with a knife]$_{INSTRUMENT}$.

□ SRL and classification – the BIO notation

Yesterday/B , /X a/B robber/O killed/X a/B guardian/O with/B a/I knife/O ./X

FLaIt: Open issues

- In SRL, language learning systems usually rely on a variety of linguistic observations and on a number of design choices such as
  - Lexical information (i.e. open or closed predicate dictionaries)
  - Morphological and grammatical features
  - Full or shallow grammatical parsing
  - Syntactic structures
- Challenges
  - Sources of *linguistic information*
  - Effective *feature engineering*
  - Manual or automatic *feature selection*
  - *Learning rates*

FLaIt: Basic Resources

- Framenet
  - Reference Model: Version 2.0
  - A source dictionary of about 12,000 Lexical Units referring to a set of about 1100 predicates, i.e. frames, triggering about 2-7 roles (i.e. frame elements) each
- Training Set for Italian. Two manually annotated sentence collections
  - ILC (A. Lenci, G. Venturi)
  - FBK (S. Tonelli, E. Pianta)
  - **1255** sentences, **38** frames, **282** different LU, **2836** arguments

FLaIt: Basic Resources

- Test Set:
  - Newly annotated data set
  - Source: Automatically annotated material from aligned sentences of Europarl
    - SRL over the English semtences plus cross-linguistic semantic transfer (Basili et al., 2009)
- Manual annotation at UniPi & ILC for validation by G. Venturi, C. Cinquesanti

- Size: **318** Sentences, **38** frames, **105** LUs (**+30**), **560** arguments
- 88 ambiguos sentences (10 LUs), FP baseline: ~68%

FLaIT: Data Formats (*.sem)

Example of the semantic annotation format.

1	Rilevata	V	–	–	–	–
2	la	RD	–	–	–	–
3	presenza	S	Presence	Target	–	–
4	di	E	–	Entity	–	–
5	gas	S	–	Entity	–	–
6	in	E	–	Location	–	–
7	uno	PI	–	Location	–	–
8	dei	EA	–	Location	–	–
9	tubi	S	–	Location	–	–
10	trasparenti	A	–	Location	–	–
11	che	PR	–	Location	–	–
12	compongono	V	–	Location	–	–
13	l'	RD	–	Location	–	–
14	opera	S	–	Location	–	–
15	,	FF	–	–	–	–
16	i	RD	–	–	–	–
17	guardiani	S	–	–	–	–
18	hanno	VA	–	–	–	–
19	fatto	V	–	–	–	–
20	scattare	V	Process_start	–	Target	–
21	uno	RI	–	–	Event	–
22	speciale	A	–	–	Event	–
23	piano	S	–	–	Event	–
24	d'	E	–	–	Event	–
25	emergenza	S	–	–	Event	–

FLAIT: DATA FORMATS (*.SYNT)

Example of the syntactic annotation format

1	Rilevata	rilevare	V	0	ROOT
2	la	il	RD	3	det
3	presenza	presenza	S	20	subj
4	di	di	E	3	comp
5	gas	gas	S	4	prep
6	in	in	E	20	comp
7	uno	uno	PI	6	prep
8	dei	di	EA	7	comp
9	tubi	tubo	S	8	prep
10	trasparenti	trasparente	A	9	mod
11	che	che	PR	12	subj
12	compongono	comporre	V	9	mod_rel
13	l'	il	RD	14	det
14	opera	opera	S	12	obj
15	,	,	FF	14	con
16	i	il	RD	17	det
17	guardiani	guardiano	S	20	subj
18	hanno	avere	VA	19	aux

FLAIT: "GOLD" DATA FORMAT

1	Rilevata	V	–	–	–	–
2	la	RD	–	–	–	–
3	presenza	S	Presence	Target	–	–
4	di	E	–	B	–	–
5	gas	S	–	O	–	–
6	in	E	–	B	–	–
7	uno	PI	–	I	–	–
8	dei	EA	–	I	–	–
9	tubi	S	–	I	–	–
10	trasparenti	A	–	I	–	–
11	che	PR	–	I	–	–
12	compongono	V	–	I	–	–
13	l'	RD	–	I	–	–
14	opera	S	–	O	–	–
15	,	FF	–	–	–	–
16	i	RD	–	–	–	–
17	guardiani	S	–	–	–	–
18	hanno	VA	–	–	–	–
19	fatto	V	–	–	–	–
20	scattare	V	Process_start	–	Target	–

FLaIt: the Task Set-up

- Three Different Runs
- **First Run**.
  - Only sentences and (potential) predicate words. No frame information made available
  - Task: **FP, BD, AC**
- **Second Run**.
  - Frame information (of a predicate word) made available
  - Task: **BD, AC**
- **Third Run**.
  - Boundary information made available.
  - Task: **AC** only

PARTICIPANT TEAMS AND SYSTEMS

- Organisations:
  - CELI srl (1 system, 2 setup)
  - University of Roma Tor Vergata, SAG group (2 systems 3 set-ups)
- Systems:
  - RTV_SVM_SPTK systems
    - Use structured learning techniques, e.g. SVM-HMM
    - Exploits state-of-the-art tree kernels, in cooperation with the University of Trento (A. Moschitti)
  - RTV_SVM_Geom system
    - Hybrid generative and geometrical modeling
  - CELI_NT/WT systems
    - Use rules
    - Use lexical parameters

Results: Frame Prediction

Systems	CELI_NT	CELI_WT	TV_SVM-SPTK	TV_SVM-HMM
Gold Frame Total	318	318	318	318
Frame Correct	207	207	257	250
Frame Untagged	38	38	0	0
Frame Precision	73.93%	73.93%	80.82%	78.62%
Frame Recall	65.09%	65.09%	80.82%	78.62%
Frame F1	69.23%	69.23%	80.82%	78.62%

RESULTS: BOUNDARY DETECTION

First Run					
Systems	CELI_NT	CELI_WT	TV_SVM-SPTK	TV_SVM-HMM	
Gold Arg. Size	560	560	560	560	
Gold Arg. Token Size	3492	3492	3492	3492	
Sys. Arg. Size	255	332	609	568	
Sys. Arg. Token Size	1165	1477	3592	3962	
Correct Bound.	117	135	406	288	
Correct Tk. Bound.	945	1162	2945	2695	
BD Prec.	45.88%	40.66%	**66.67%**	50.70%	
BD Rec.	20.89%	24.11%	**72.50%**	51.43%	
BD F1	28.71%	30.27%	**69.46%**	51.06%	
BD Token Prec.	81.12%	78.67%	**81.99%**	68.02%	
BD Token Rec	27.06%	33.28%	**84.34%**	77.18%	
BD Token F1	40.58%	46.77%	**83.15%**	72.31%	

Second Run					
Systems	CELI_NT	CELI_WT	TV_SVM-SPTK	TV_SVM-HMM	RTV_SVM_Geom
Sys. Arg. Size	263	349	609	565	494
Sys. Arg. Token Size	1150	1487	3592	3930	3569
Correct Bound.	124	148	406	282	357
Correct Token Bound.	949	1193	2945	2678	2969
BD Prec.	47.15%	42.41%	66.67%	49.91%	**72.27%**
BD Rec.	22.14%	26.43%	**72.50%**	50.36%	63.75%
BD F1	30.13%	32.56%	**69.46%**	50.13%	67.74%
BD Token Prec.	82.52%	80.23%	81.99%	68.14%	**83.19%**
BD Token Rec.	27.18%	34.16%	84.34%	76.69%	**85.02%**
BD Token F1	40.89%	47.92%	83.15%	72.16%	**84.10%**

RESULTS: BOUNDARY DETECTION (2)

			Second Run		
Systems	CELI_NT	CELI_WT	TV_SVM-SPTK	TV_SVM-HMM	RTV_SVM_Geom
Sys. Arg. Size	263	349	609	565	494
Sys. Arg. Token Size	1150	1487	3592	3930	3569
Correct Bound.	124	148	406	282	357
Correct Token Bound.	949	1193	2945	2678	2969
BD Prec.	47.15%	42.41%	66.67%	49.91%	**72.27%**
BD Rec.	22.14%	26.43%	**72.50%**	50.36%	63.75%
BD F1	30.13%	32.56%	**69.46%**	50.13%	67.74%
BD Token Prec.	82.52%	80.23%	81.99%	68.14%	**83.19%**
BD Token Rec.	27.18%	34.16%	84.34%	76.69%	**85.02%**
BD Token F1	40.89%	47.92%	83.15%	72.16%	**84.10%**

RESULTS: ARGUMENT CLASSIFICATION

First Run					
Systems	CELL_NT	CELL_WT	TV_SVM-SPTK	TV_SVM-HMM	
Gold Arg. Size	560	560	560	560	
Gold Arg. Token Size	3492	3492	3492	3492	
Sys. Arg. Size	255	332	609	568	
Sys. Arg. Token Size	1165	1477	3592	3962	
Correct Arg.	83	91	295	188	
Correct Token Arg.	558	731	2248	1853	
AC Prec.	32.55%	27.41%	**48.44%**	33.10%	
AC Rec.	14.82%	16.25%	**52.68%**	33.57%	
AC F1	20.37%	20.40%	**50.47%**	33.33%	
AC Token Prec.	47.90%	49.49%	**62.58%**	46.77%	
AC Token Rec	15.98%	20.93%	**64.38%**	53.06%	
AC Token F1	23.96%	29.42%	**63.47%**	49.72%	

Second Run					
Systems	CELI_NT	CELI_WT	TV_SVM-SPTK	TV_SVM-HMM	RTV_SVM_Geom
Sys. Arg. Size	263	349	609	565	494
Sys. Arg. Token Size	1150	1487	3592	3930	3569
Correct Arg.	95	109	312	212	256
Correct Token Arg.	716	960	2479	2147	2198
AC Prec.	36.12%	31.23%	51.23%	37.52%	**51.82%**
AC Rec.	16.96%	19.46%	**55.71%**	37.86%	45.71%
AC F1	23.09%	23.98%	**53.38%**	37.69%	48.58%
AC Token Prec.	62.26%	64.56%	**69.01%**	54.63%	61.59%
AC Token Rec.	20.50%	27.49%	**70.99%**	61.48%	62.94%
AC Token F1	30.85%	38.56%	**69.99%**	57.86%	62.26%
Third Run					
Systems	CELI_NT	CELI_WT	TV_SVM-SPTK	TV_SVM-HMM	RTV_SVM_Geom
Sys. Arg. Size	247	300	560	549	543
Sys. Arg. Token Size	1657	2160	3492	3481	3475
Correct Arg.	181	225	394	366	363
Correct Token Arg.	1269	1798	2736	2705	2489
AC Prec.	73.28%	**75.00%**	70.36%	66.67%	66.85%
AC Rec.	32.32%	40.18%	**70.36%**	65.36%	64.82%
AC F1	44.86%	52.33%	**70.36%**	66.01%	65.82%
AC Token Prec.	76.58%	**83.24%**	78.35%	77.71%	71.63%
AC Token Rec.	36.34%	51.49%	**78.35%**	77.46%	71.28%
AC Token F1	49.29%	63.62%	**78.35%**	77.59%	71.45%

General Remarks

- The Frame detection task is accurately carried out by most systems
  - Small number of frames
  - Closed world assumption
  - Baseline (F1 68.39%)
- Systems are, by design, different in recall for AC and BD
- The tree kernel-based system, in particular SPTK, are outperforming the other systems in BD and AC
- A significant performance increment is observed between the second and third run, as the quality of parsing is critical (except for the SVM-HMM)

CONCLUSIONS

- FLaIt is a rather important achievement given the status of the resources before the first launch of this task in Evalita

- Chances to make new resources available

- FLaIt has been a joint effort between the Organizers and:

  - iFrame, a joint research network on the development of a Framenet-like resource for Italian
    http://sag.art.uniroma2.it/iframe/doku.php?id=start

  - PARLI, the PRIN 2008 project for the development of resources for Italian
    http://parli.di.unito.it/project_en.html