EVALITA 2011
Parsing Task

Cristina Bosco and Alessandro Mazzei

Università degli Studi di Torino

Outline

- Introduction

- Evaluation:
  - dataset
  - formats
  - metrics
  - results

- Discussion

- Conclusions

Introduction

- The parsing task is the activity of assigning a syntactic structure to a given set of PoS tagged sentences

- A large set of sentences is given for tuning and training (development set)

- The evaluation is based on a smaller set of new sentences (test set)

Introduction

- The parsing task is organized in two tracks, i.e. dependency and constituency

- The datasets are the same for both tracks, but annotated in different format

- The evaluation procedures and metrics are different for each track

Evaluation

Datasets

Development set from the Turin University Treebank:

- 3,542 sentences (102,150 tokens)

- 3 text genres (1,983 sentences from legal texts; 1,100 from newspaper; 459 from Wikipedia)

Test set:

- 300 sentences (7,836 tokens)

- 3 text genres (150 legal; 75 newspaper; 75 Wikipedia)

Evaluation

Formats

For Dependency:

 data are in TUT-CoNLL format, obtained by native TUT deleting null elements and reducing the amount of relations

For Constituency:

data are in TUT-Penn format

Evaluation

TUT in CoNLL (Dependency Format)

- pure dependency trees labeled with 72 different relations

- only projective trees

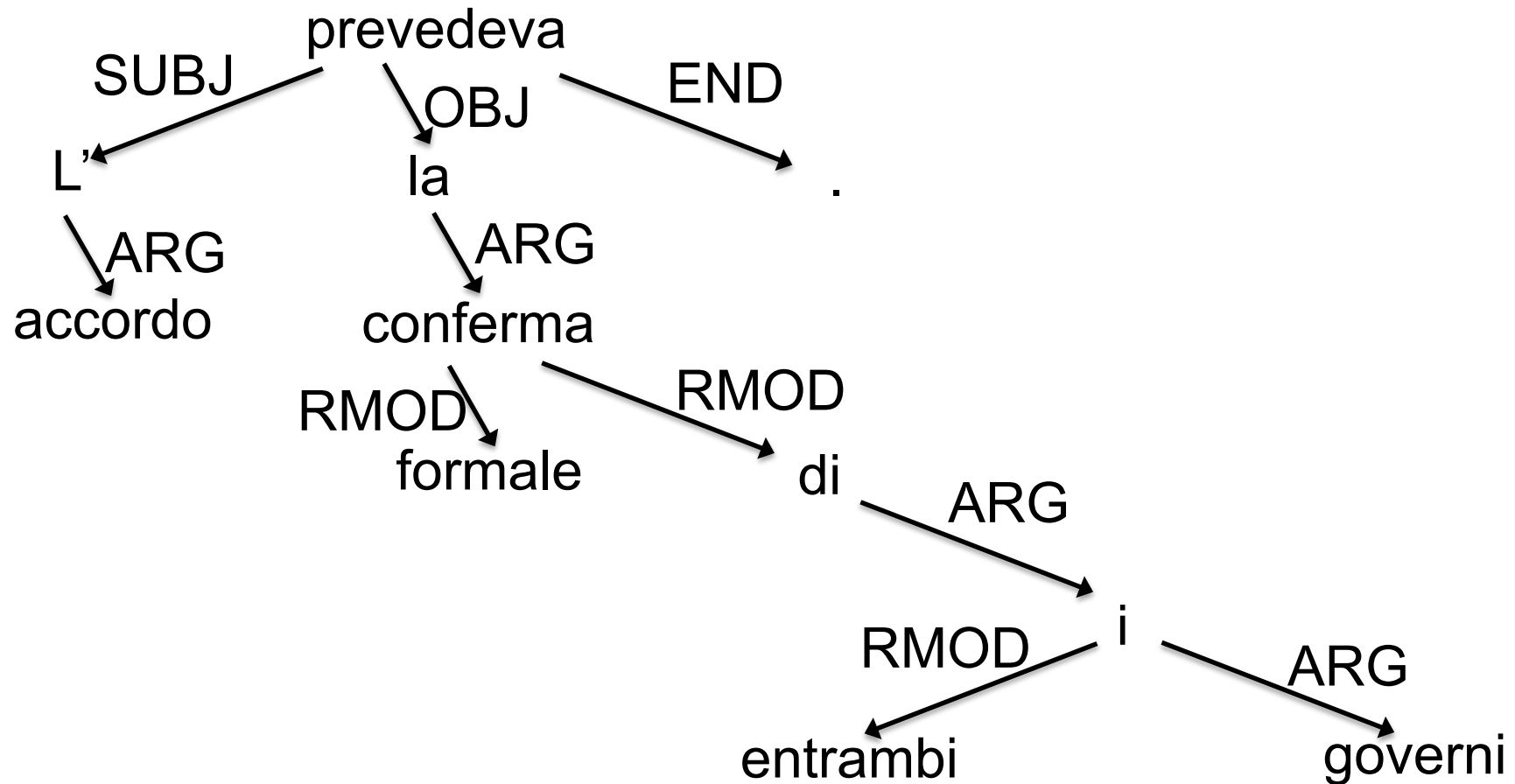- data filling eight columns of CoNLL's format

Evaluation

TUT in CoNLL (Dependency Format)

1	L'	IL	ART	ART	DEF|M|SING	3	SUBJ
2	accordo	ACCORDO	NOUN	NOUN	COMMON|M|SING	1	ARG
3	prevedeva	PREVEDERE	VERB	VERB	MAIN|IND|IMPERF|TRANS|3|SING	0	TOP
4	la	IL	ART	ART	DEF|F|SING	3	OBJ
5	conferma	CONFERMA	NOUN	NOUN	COMMON|F|SING	4	ARG
6	formale	FORMALE	ADJ	ADJ	QUALIF|ALLVAL|SING	5	RMOD
7	di	DI	PREP	PREP	MONO	5	RMOD
8	entrambi	ENTRAMBI	PREDET	PREDET	M|PL	9	RMOD
9	i	IL	ART	ART	DEF|M|PL	7	ARG
10	governi	GOVERNO	NOUN	NOUN	COMMON|M|PL	9	ARG
11	.	#\.	PUNCT	PUNCT	_	3	END

Evaluation

TUT in CoNLL (Dependency Format)

prevedeva

SUBJ   OBJ   END

L'   la   .

ARG   ARG

accordo   conferma

RMOD   RMOD

formale   di

ARG

RMOD   i   ARG

entrambi   governi

Evaluation

TUT–Penn (Constituency Format)

- trees with the same structure of the English Penn Treebank format

- richer PoS tagset and richer inventory of functional relations with respect to Penn Treebank, in order to better describe Italian

Evaluation

TUT-Penn (Constituency Format)

```
(S

    (NP-SBJ (ART~DE L') (NOU~CS accordo))

    (VP (VMA~IM prevedeva)

        (NP

            (NP (ART~DE la) (NOU~CS conferma) (ADJ~QU formale))

            (PP (PREP di)

                (NP (PRDT entrambi) (ART~DE i) (NOU~CP governi)))))

    (. .))
```

Evaluation

Metrics

For Dependency:

Labeled Attachment Score (LAS) = percentage of tokens with correct head and dependency type

Unlabeled Attachment Score (UAS) = percentage of tokens with correct head

For Constituency:

Labelled Precision (LP), Labelled Recall (LR), F-score

Evaluation

Results

Dependency

LAS	UAS	
91.23	96.16	Parsit_Grella
89.88	93.73	UniPisa_Attardi
88.62	92.85	FBK_Lavelli
85.34	91.47	UniTorino_Lesmo

Constituency

LP	LR	F-score	
82.94	82.97	82.96	FBK_Lavelli

Discussion

Results positively compares with previous both for dependency and constituency; but scores for dependency remain higher.

Less participants than in previous editions.

Difference of results on different genres confirms as in previous editions that legal texts are less hard to parse.

Conclusions

The results of the parsing task are very good, but there is a lot of work to do in the future.

With respect to parsing, it should be deeply investigated the use of different kinds of knowledge (e.g. null elements, punctuation, semantics) in the same or in other formats.

With respect to Evalita, we hope for a larger participation in current and new tasks in the future, for the assessment of our results in a wider community of researchers.