



# EVALITA 2011

## **Experiments with a Constraint-based Dependency Parser**

M. Grella, M. Nicola, D. Christen

**parsit**

<http://www.parsit.it>



# Outline

---

- Parsit at Evalita 2011
- Evaluation and results
- Description of the System
- Example
- How we came to a Constraint-based Parser
- Future Works



# Parsit at Evalita 2011

- Scientific Research plays a central role in **Parsit** activities
- **Parsit** technologies will be available for free for partners with researching purposes
- We focus on syntactic parsing of natural language
- The parser architecture used at **Evalita 2011** is one of a series of experiments we have been dedicating to in the past few months
- The parser was tested against the **Evalita 2009** dataset and the results encouraged us to participate to **Evalita 2011**



# RESULTS



# Results

Arc Accuracy (300 sentences for a total of 7401 tokens)

	LAS	UAS	LAS2
Parsit (M. Grella)	<b>91.23%</b>	<b>96.16%</b>	<b>93.87%</b>
#2	89.88%	93.73%	93.39%
#3	88.62%	92.85%	92.50%
#4	85.34%	91.47%	89.57%

The paper submitted includes some consideration on the difference between the LAS and UAS values

Correct Sentence Accuracy

CSA-L	CSA-U
22.33%	57.00%

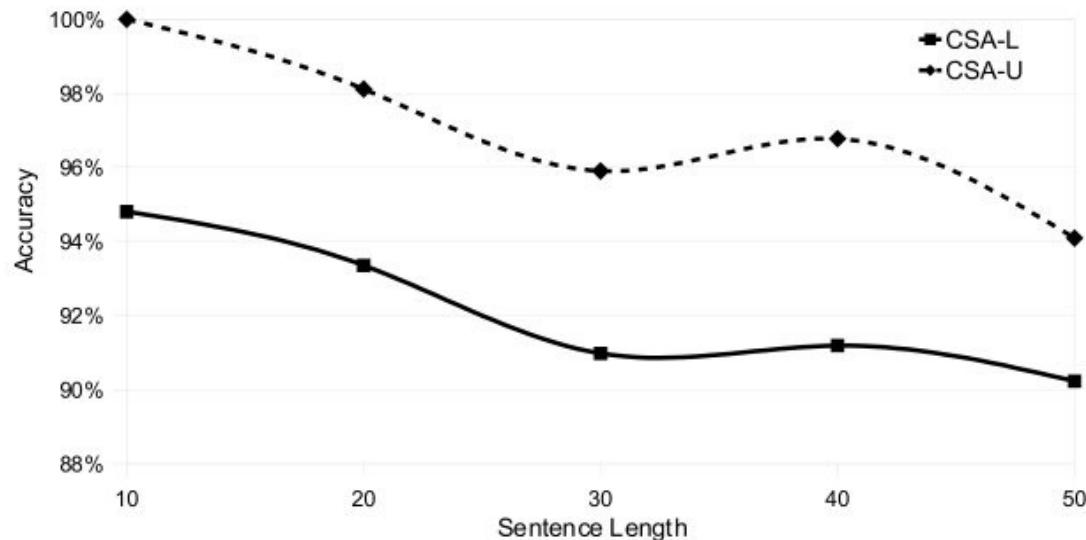
Average LAS and UAS per sentence

AVG-L	AVG-U
91.66%	96.63%



# Results

## Average Accuracy relative to sentence length



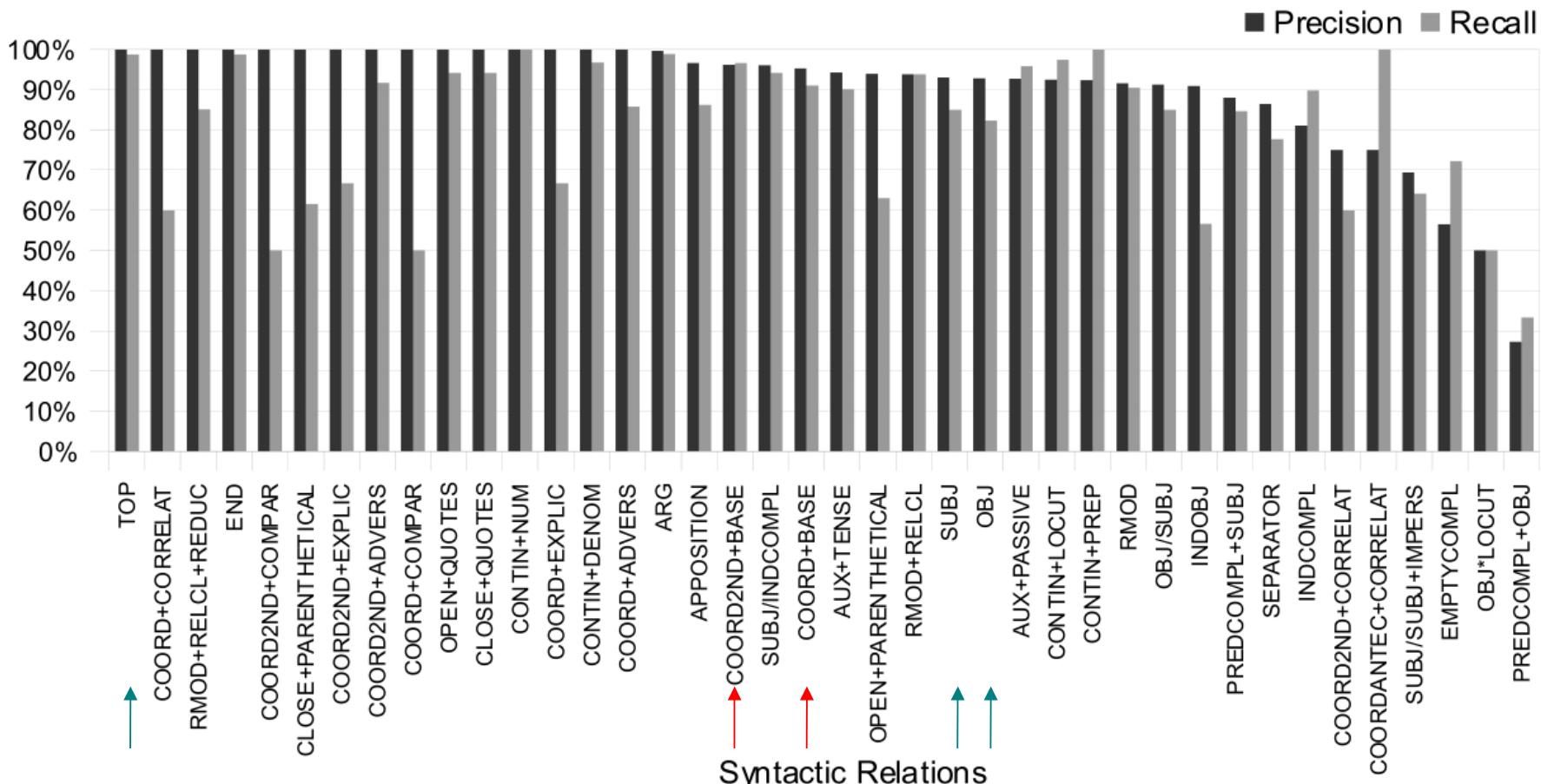
Group	Range	No. of sentences	Portion of dataset
1	1-10	15	5.00%
2	11-20	95	31.67%
3	21-30	108	36.00%
4	31-40	62	20.67%
5	41-50	20	6.67%

We considered meaningful groups 2, 3 only



# Results

Recall and Precision of every syntactic relation label attachment





# DESCRIPTION OF THE SYSTEM



# System - Goals

- Coherence (consistency) of the syntactic trees  
typical of top-down parsers
- High level of accuracy for label assignment
- Robustness  
typical of bottom-up parsers
- Capable of handling non-projectivity
- Morphological analysis without pre POS-tagger

Efficiency has not been considered to be relevant in this specific context.



# System - Approach

## Experiments with a Constraint-based Dependency Parser

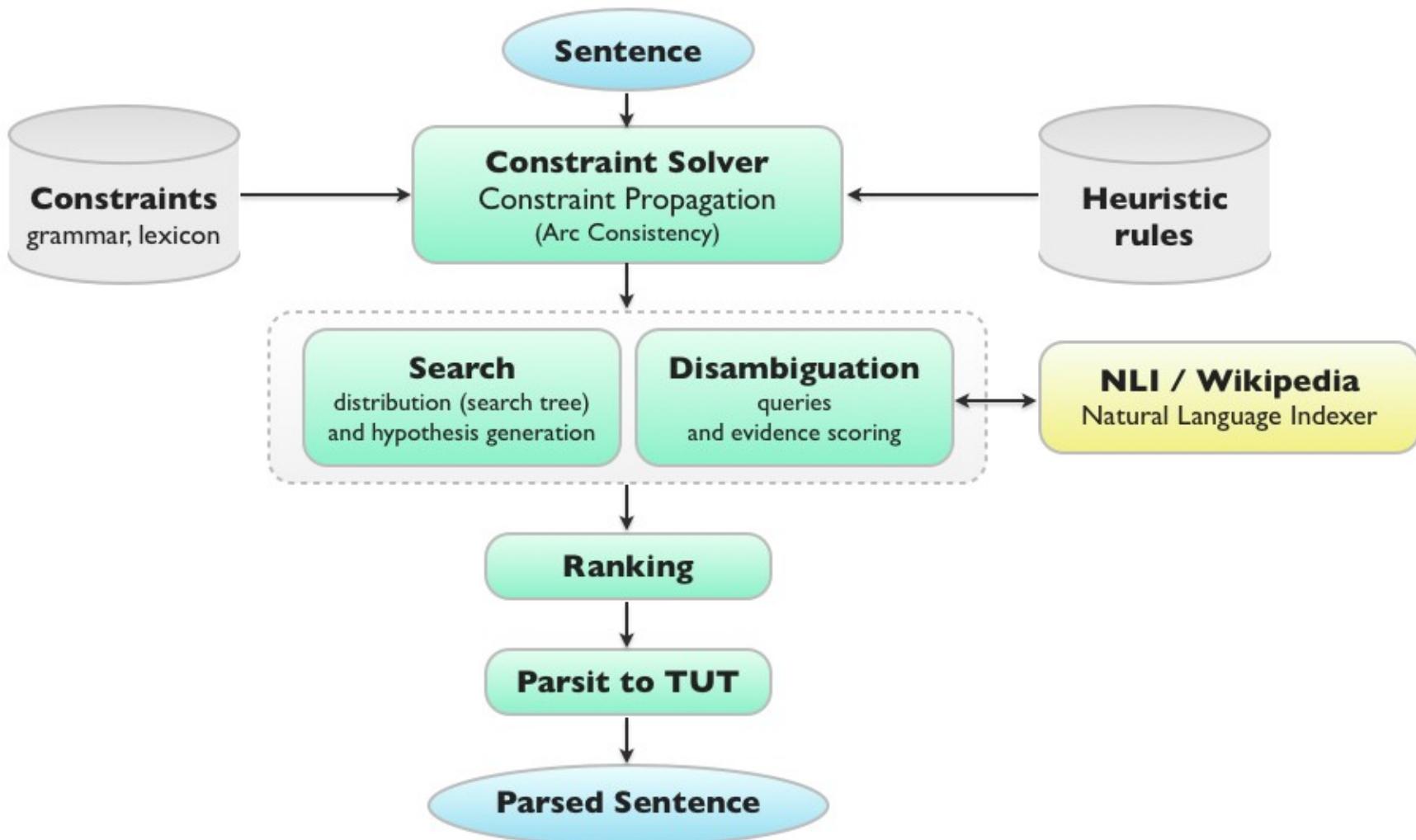
View of the parsing process as a finite configurations problem that can be formulated as a constraint satisfaction problem (CSP)

Dependency parsing problem is reduced to the problem of finding a dependency graph for a sentence that satisfies all the constraints defined by the grammar

cf. Maruyama, Harper and Helzerman, Menzel, Schröder, Duchier, Debusmann



# System - Architecture





# System - Constraint Solver

The Constraint Solver has been implemented in Ada 2005 using Gecode constraint programming library which provides constraint-solving algorithms with state-of-the-art performances.

For a sentence of length n, there are finitely many possible trees involving just n nodes. Out of this large number, we must select those that are grammatical.

Dependency Graph is represented using an Adjacency Matrix.

	ROOT	il	gatto	insegue	il	topo	in	giardino
il	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...		TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...					
gatto	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...		TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...				
insegue	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...		TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...
il	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...		TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...
topo	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...		TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...				
in	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...		TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...					
giardino	TOP, COORD-BASE, SUBJ, OBJ, RHOQ, ARQ, ...							

	ROOT	il	gatto	insegue	il	topo	in	giardino
il			DET					
gatto				SUBJ				
insegue	TOP							
il					DET			
topo						OBJ		
in							CONN	
giardino							RMOD	RMOD

The constraints propagation technique is applied to restrict possible analyses.



# System - Format

TUT Format	Parsit Format
ART → NOUN	ART ← NOUN
PREP → NOUN	PREP ← NOUN
CONJ-SUBORD → VERB	CONJ-SUBORD ← VERB



# System - Constraints

The Italian grammar has been expressed for the first time as a set of constraints

Wide coverage lexicon of subcategorization for nouns, adjectives, verbs and adverbs: each lexical entry contains information on category, agreement and valency (i. e. what types of complements are required by the word).

The grammar has been then fine-tuned on the TUT Treebank thanks to our automatic tools written in Python.

```
(deprel DET
  (governor NOUN)
  (dependent ART)
  (< (dependent ID) (governor ID))
  (morpho-agree dependent governor))
```

```
(deprel SUBJ
  (governor VERB)
  (dependent SUBJ)
  (complex-morpho-agree dependent governor))
```



# System - Search

- The Matrix is now consistent but, with the exception of simple cases, still shows ambiguous dependencies
- Simple search algorithm: spawns sequentially all possible non-ambiguous combinations of dependencies from the Matrix
- All incoherent parse trees are rejected (e.g. in the dependency graph each node has only one incoming edge, there are no cycles and there is precisely one root)



# System - Disambiguation

The score of each ambiguous relation in a given solution is computed by counting the number of occurrences of that relation in the Italian version of **Wikipedia** previously analyzed by **TULE parser** [Leonardo Lesmo]

Parsed sentences have been enriched with **Parsit Intra-Paragraph Anaphora Resolution (IPAR)** and indexed by our **Natural Language Indexer (NLI)**

**NLI** is a component of the high performance **Natural Language Retrieval Architecture (NLRA)** designed and developed by **Parsit** which allows for complex queries on parse trees data through a particular pattern matching technique

Our indexed version of the Italian Wikipedia can be explored at:

<http://www.parsit.it/evalita2011>



# System - Query

## Chi ha inventato il telefono?



```
{"multidep" :  
  [  
    {"L": "inventare"},  
    {"L" : "telefono", "D": "OBJ"},  
    {"D" : "SUBJ"}  
  ]  
}
```



# EXAMPLE



# Example

il gatto insegue il topo in giardino

the cat chases the mouse in the garden



# Example - Propagation

## Initial configuration – Start propagation



# Example - Propagation

## Determiner - Arc Constraints



# Example - Propagation

	ROOT	il	gatto	insegue	il	topo	in	giardino
il			<b>DET</b>					
gatto	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...			TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
insegue	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...			TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
il						<b>DET</b>		
topo	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...			TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
in	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
giardino	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	

## Determiner - Arc Constraints



# Example - Propagation

	ROOT	il	gatto	insegue	il	topo	in	giardino
il			<b>DET</b>					
gatto	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...			TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
insegue	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...			TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
il						<b>DET</b>		
topo	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...			TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
in	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
giardino	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	

## Connective - Arc Constraints



# Example - Propagation

	ROOT	il	gatto	insegue	il	topo	in	giardino
il			DET					
gatto	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...			TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
insegue	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...			TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
il						DET		
topo	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...				TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...
in								CONN
giardino	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...	TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		TOP, COORD+BASE, SUBJ, OBJ, RMOD, ARG, ...		

## Connective - Arc Constraints



# Example - Propagation

	ROOT	il	gatto	insegue	il	topo	in	giardino
il			DET					
gatto				SUBJ, OBJ				
insegue	TOP							
il						DET		
topo				SUBJ, OBJ				
in								CONN
giardino				RMOD		RMOD		

Matrix is consistent



# Example - Propagation

	ROOT	il	gatto	insegue	il	topo	in	giardino
il				DET				
gatto				SUBJ, OBJ				
insegue	TOP							
il						DET		
topo				SUBJ, OBJ				
in							CONN	
giardino				RMOD		RMOD		

## Heuristic Rules



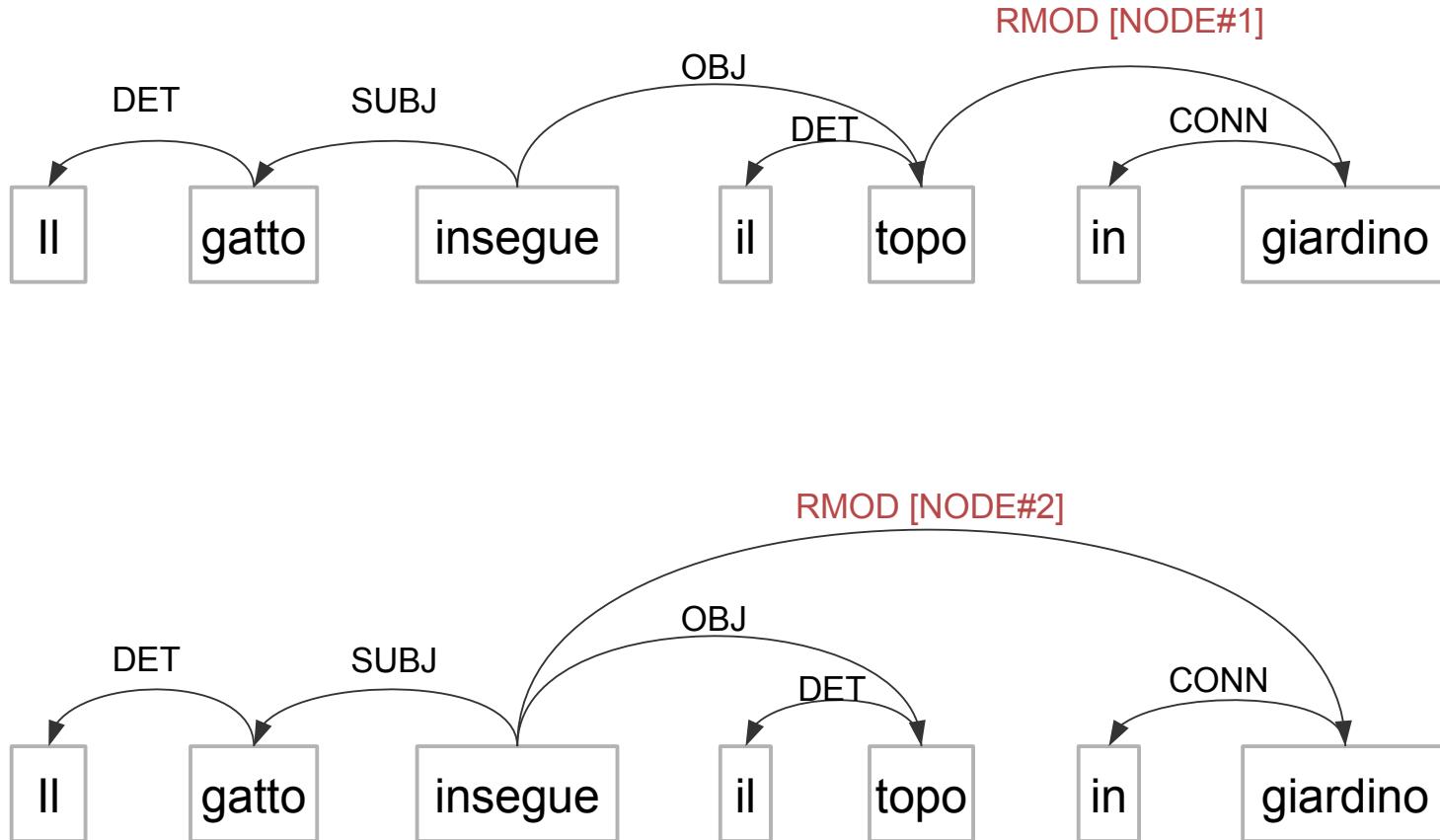
# Example - Propagation

	ROOT	il	gatto	insegue	il	topo	in	giardino
il				DET				
gatto				SUBJ				
insegue	TOP							
il						DET		
topo				OBJ				
in							CONN	
giardino				RMOD		RMOD		

Still multiple solutions



# Example - Search





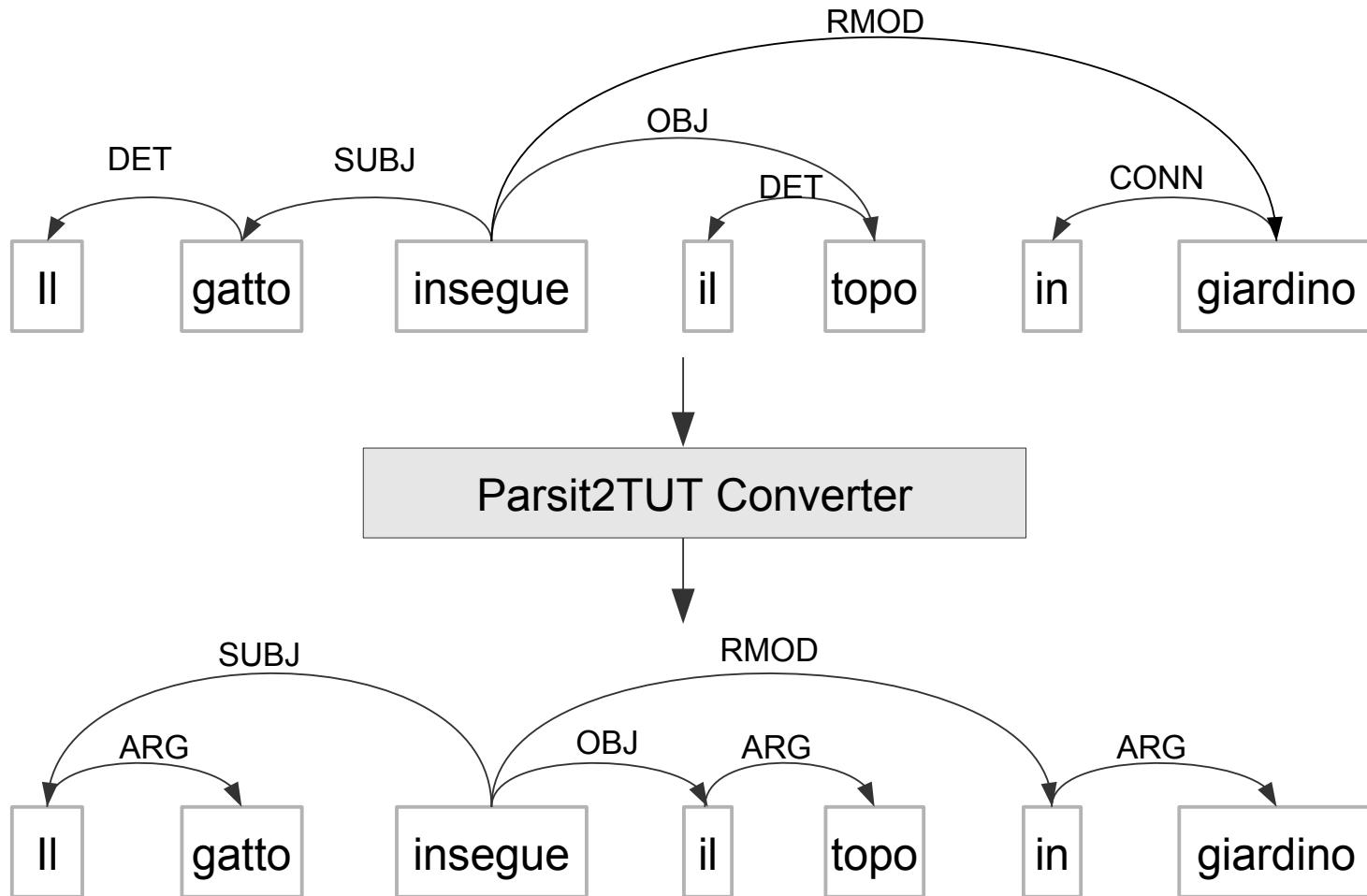
# Example - Disambiguation

QUERY NODE#1	SCORE
{"multidep": [{"L": "topo"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"P": "NOUN", "D": "ARG"}]}]}	%
{"multidep": [{"L": "topo"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"L": "giardino", "D": "ARG"}]}]}	%
{"multidep": [{"L": "topo"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"S": "LOC", "D": "ARG"}]}]}	%
{"multidep": [{"P": "NOUN"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"P": "NOUN", "D": "ARG"}]}]}	%
{"multidep": [{"P": "NOUN"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"L": "giardino", "D": "ARG"}]}]}	%
{"multidep": [{"P": "NOUN"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"S": "LOC", "D": "ARG"}]}]}	%
TOTAL	%

QUERY NODE#2	SCORE
{"multidep": [{"L": "inseguire"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"P": "NOUN", "D": "ARG"}]}]}	%
{"multidep": [{"L": "inseguire"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"L": "giardino", "D": "ARG"}]}]}	%
{"multidep": [{"L": "inseguire"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"S": "LOC", "D": "ARG"}]}]}	%
{"multidep": [{"P": "VERB"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"P": "NOUN", "D": "ARG"}]}]}	%
{"multidep": [{"P": "VERB"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"L": "giardino", "D": "ARG"}]}]}	%
{"multidep": [{"P": "VERB"}, {"multidep": [{"L": "in", "D": "RMOD"}, {"S": "LOC", "D": "ARG"}]}]}	%
TOTAL	%



# Example - Conversion





# HOW WE CAME TO A CONSTRAINT- BASED PARSER?



# Hybrid Transition-based

## Experiments with a Hybrid Transition-based Dependency Parser

- Multi-threaded Multilayer Perceptron (OpenMP has been used)
- Oracle Parsit Strategy
- Online disambiguation (POS tagging during Syntactic Parsing)
- Multilayer Linguistic Supervision (Surface, Morphology, Syntax)

"Cade [la neve]"                   *Not Object!*  
"Mi piace [la prugna]"



# Hybrid Transition-based

```
S ← []
I ← [w1 , w2 , . . . , wn]
A ← []

while I != [] do
    X ← get_context(S, I, A)
    Y ← estimate_synactions(X, model)

    foreach y in Y do
        if is_permissible(y, S, I)
            if check_linguistic_supervision(y, S, I, A, rules)
                transition(y, S, I, A)
                break
```

*extremely simplified version!*





# FUTURE WORKS



# Future works

- Accuracy improvements
- Performance improvements
- Hybrid statistical transition-based and constraint-based



# Future works

```
S ← []
I ← [w1 , w2 , . . . , wn]
A ← []
```

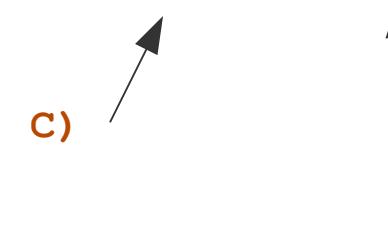
```
C ← make_arc_consistency(I) →
```

```
while I != [] do
    X ← get_context(S, I, A, C) →
    Y ← estimate_synactions(X, model)
```

```
foreach y in Y do
    if is_permissible(y, S, I)
        if is_consistent(y, S, I, A, C)
            transition(y, S, I, A)
            update(C, A)
    break
```

*extremely simplified version!*

	ROOT	il	gatto	insegue	il	topo	in	giardino
il			DET					
gatto				SUBJ				
insegue	TOP							
il						DET		
topo					OBJ			
in								CONN
giardino				RMOD		RMOD		





# EVALITA 2011

*Evaluation of NLP and Speech Tools for Italian*

# THANKS

Please send your questions, opinions and requests to [info@parsit.it](mailto:info@parsit.it)

**parsit**

<http://www.parsit.it>