UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO DI INGEGNERIA CHIMICA, GESTIONALE, MECCANICA ED INFORMATICA

COGNITIVE LINGUISTICS AS THE UNDERLYING FRAMEWORK FOR SEMANTIC APPLICATIONS: FROM SEMANTIC ANNOTATION TO COMPLETE NLIS

A case of study for a LMSs

Arianna Pipitone and Roberto Pirrone

ROMA – 15 | 06 | 2012

Focus

- The context of this work is the task of *automatically understanding, producing and answering Natural Language utterances/questions about a domain by a machine as if it were a human.*
  - Understanding natural language textual documents → Semantic Computing, Knowledge Acquisition, Semantic Annotation
  - Production → Talking about a domain, Ontology Verbalization
  - Understanding/production → Natural Language Interfaces with specific application: an intelligent Tutor for LMS

- This task has been a subject of research in:
  - Natural Language Processing
    - Natural Language Production
    - Natural Language Understanding
  - Knowledge Representation

Assumptions

The described methodology is helpful for intelligent support during interactions with users inside a LMS (as OLAT or Moodle).

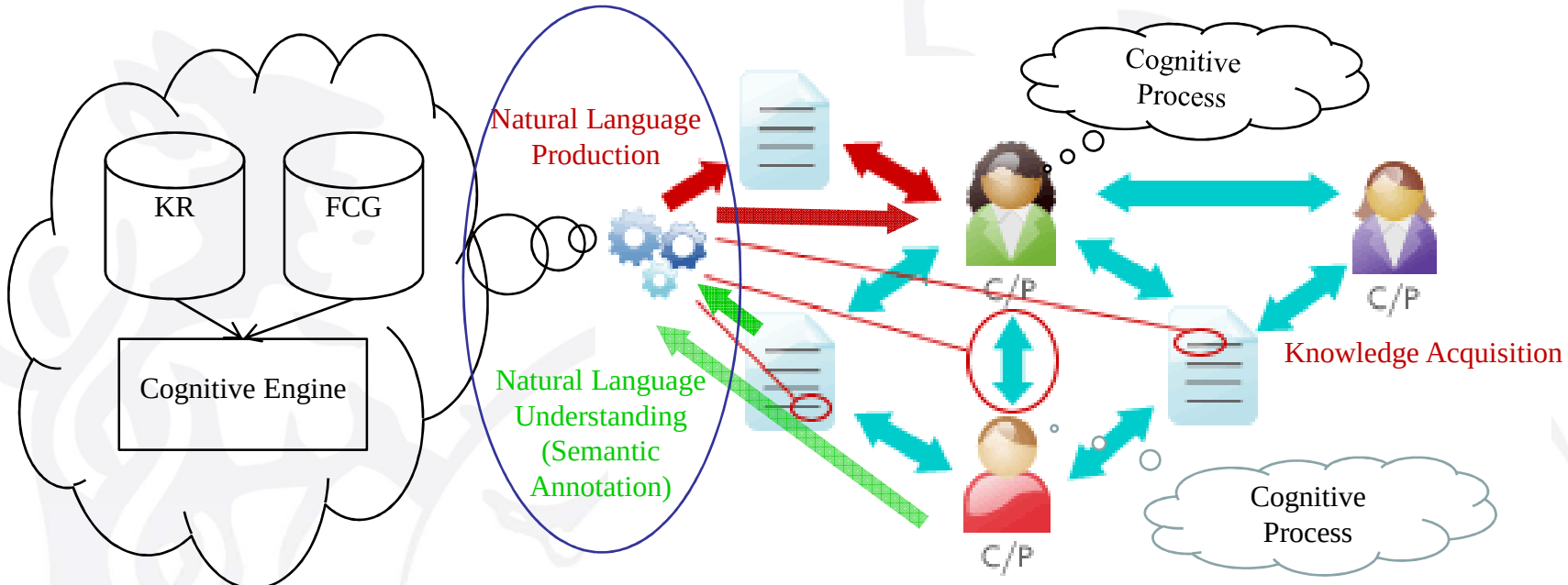
The theoretical background is described by followed assumptions:

- Ontology is the internal representation of the world (i.e. the domain) owned by the agent
- Knowledge may increase over time as the result of several understanding tasks
- Resources can be processed against different ontologies that is they’re analyzed by different agents at different times.
- Each agent acquires knowledge from the previous annotations, and annotates the resource with new tags.
- In this way the agent enriches its linguistic knowledge about the domain (the way of saying something).
- Linguistic abilities are triggered by parsing documents related to the domain itself, and this forms the basis of annotation.

Scenario

LMS

Many information on the LMS can be unstructured: these information become interoperable and understandable by computers structuring them according to Web standard, so machines can understand, use, share and reason about them, produce utterances and answer unskilled users queries



UNIVERSITÀ DEGLI STUDI DI PALERMO

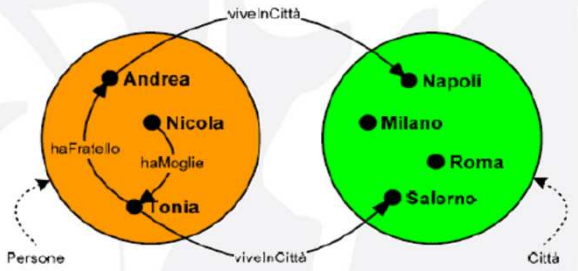
Knowledge Representation

- Many KR languages have been invented for standardization purposes including RDF and OWL.
- These languages encapsulate knowledge of the world through a set of concepts and relations between them. These are organised into triples which have the form

SUBJECT <predicate> OBJECT

that is, two concepts (subject and object) related by a predicate.

- These triples together form *conceptual models* or *ontologies.*
- We think to structure LMS contents through topic categorization, realized by the recognition of relevant terms on the text (NER) and building the correspondent ontology.



We can use methodology by Cognitive Linguistics to model a tool for providing a NLI for a LMS; the model combines ontology representation with Construction Grammar (CxG)

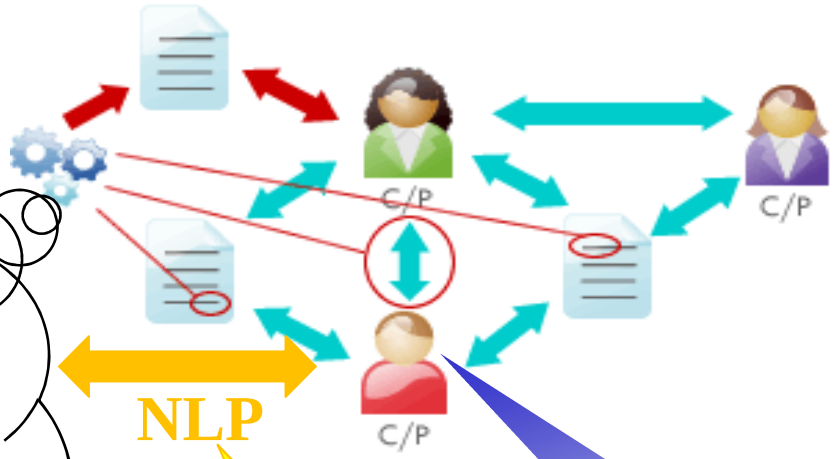
Mary works for University of Sheeld, which is located in Sheeld. Sheeld is placed in the United Kingdom.

NLU

MARY <is a> PERSON
MARY <work organization> UNIVERSITY
SHEFFIELD <is a> CITY
UNIVERSITY <is located in> CITY
UNITED KINGDOM <is a> COUNTRY

MARY <work organization> UNIVERSITY OF SHEELD
UNIVERSITY OF SHEELD <is located in> SHEELD
MARY <works at> SHEELD

NLP



Where does Mary work?

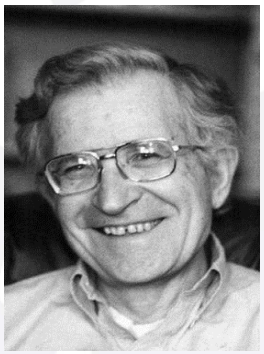
Mary is employed at University of Sheeld. Mary works at Sheeld.

Cognitive Linguistics (CL)

•CL is the study of the language in its cognitive function, where *cognitive* refers to the crucial role of intermediate informational structures in our encounters with the world.

•In the CL
- – no syntax without semantics
- – knowledge → the world (*through* the language)
- – mental schema → the language

•In the Generative Grammar
- – autonomy of syntax
- – knowledge → the knowledge *of* the language
- – mental schema → a method to learn the language



"colorless green ideas are meaningless"
Noam Chomsky (1957)

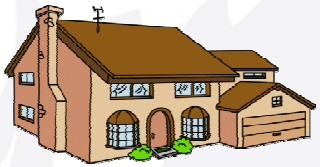
CxG in the OWL view

An OWL ontology is a net of constructions! We can argue that:

- *A RDF triple is the atomic construction for SVO statements*

- *In general, an OWL sub-graph can be put into correspondence to a set of constructions for describing complex statements*

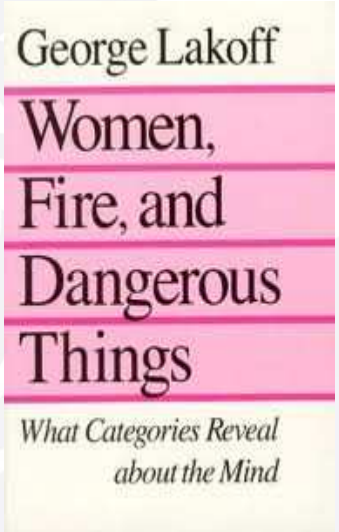
- *An OWL ontology conveys also semantic information, and can be used to support a (partial) Construction Grammar related to the domain.*



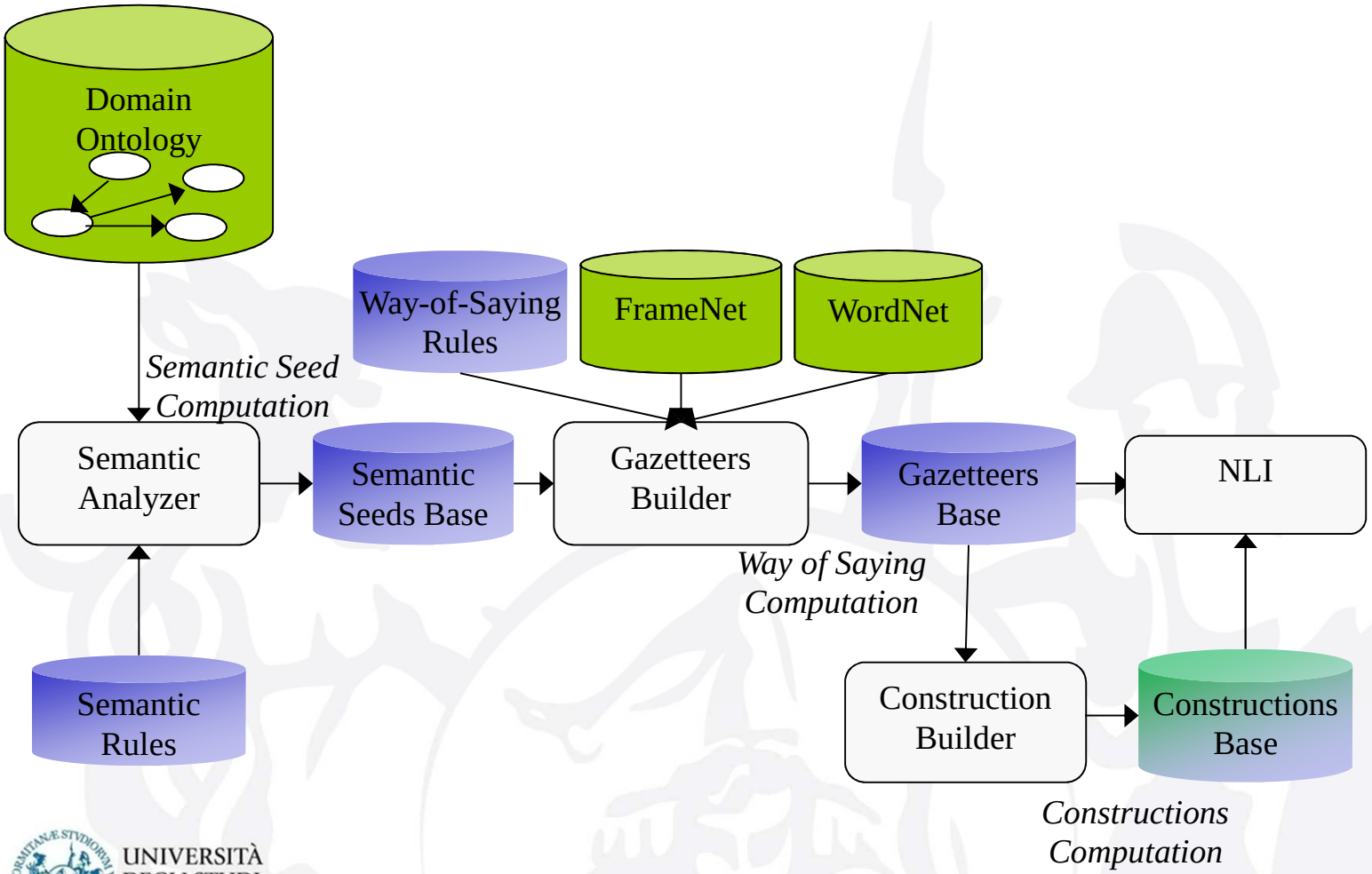
Construction Grammar (CxG)

- CxG is model in CL:
  - The atomic unit is a construction (form – meaning couple);
  - Consists of taxonomies of families of constructions;
  - Incorporates the cognitive and interactional foundations of language.

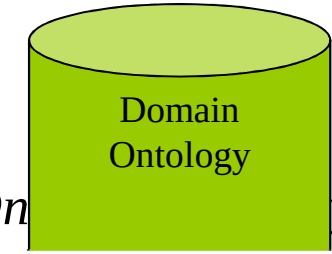
Charles Fillmore
Paul Kay
George Lakoff
Adele Goldberg



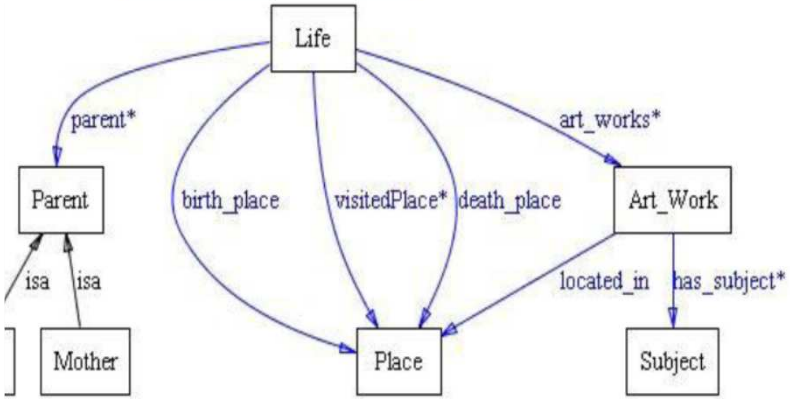
Architecture of the NLI prototype



Semantic Analyzer



On y



Semantic Seeds extraction rules: We represent the semantic seeds using First Order Logic (FOL) predicates. At first, a suitable axiomatization has been defined to model the three components of a SVO triple, and the triple itself:

(S ?s) asserts the existence of S components;

(O ?o) asserts the existence of O components;

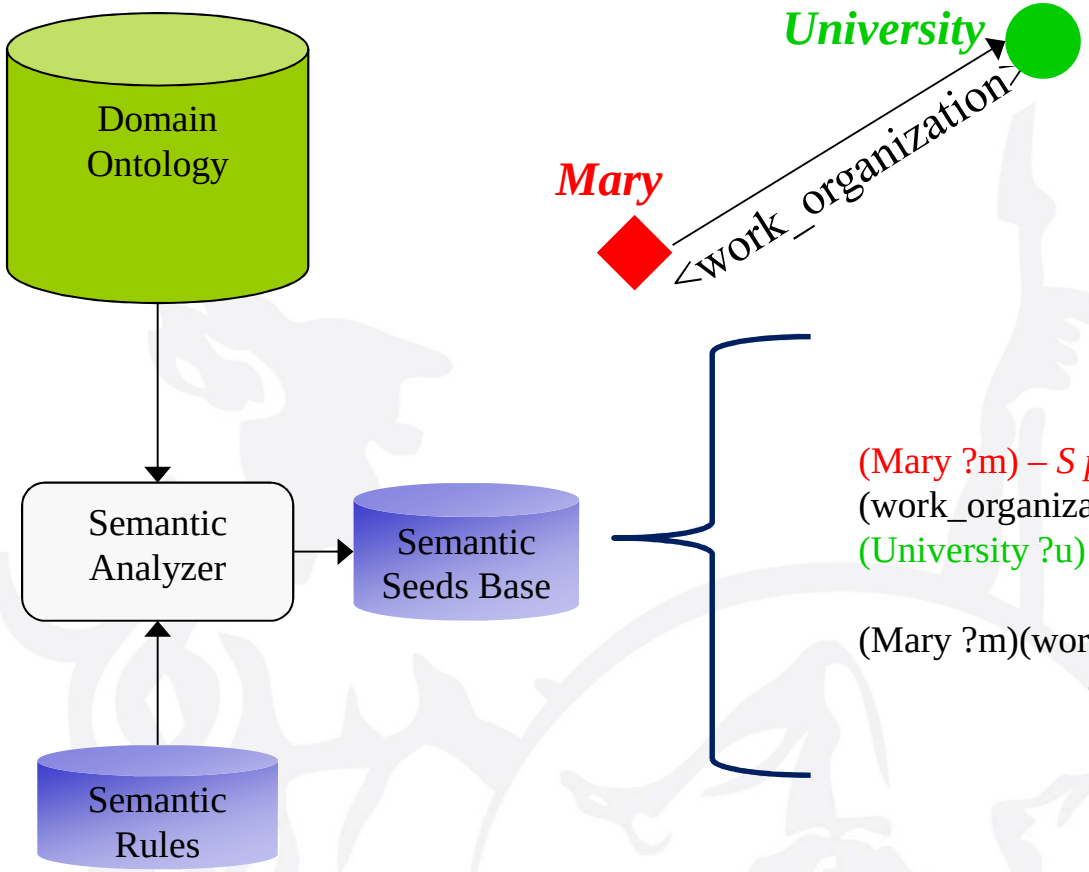
(P ?s ?p ?o) asserts the existence of V components;

(S ?s)(P ?s ?p ?o)(O ?o) asserts a SVO triple

UNIVERSITÀ DEGLI STUDI DI PALERMO

A. Pipitone, R. Pirrone

Semantic Rules



(Mary ?m) – *S predicate*
(work_organization ?m ?p ?u) – *P predicate*
(University ?u) – *O predicate*

(Mary ?m)(work_organization ?m ?p ?u)(University ?u)
- SPO conjunction -

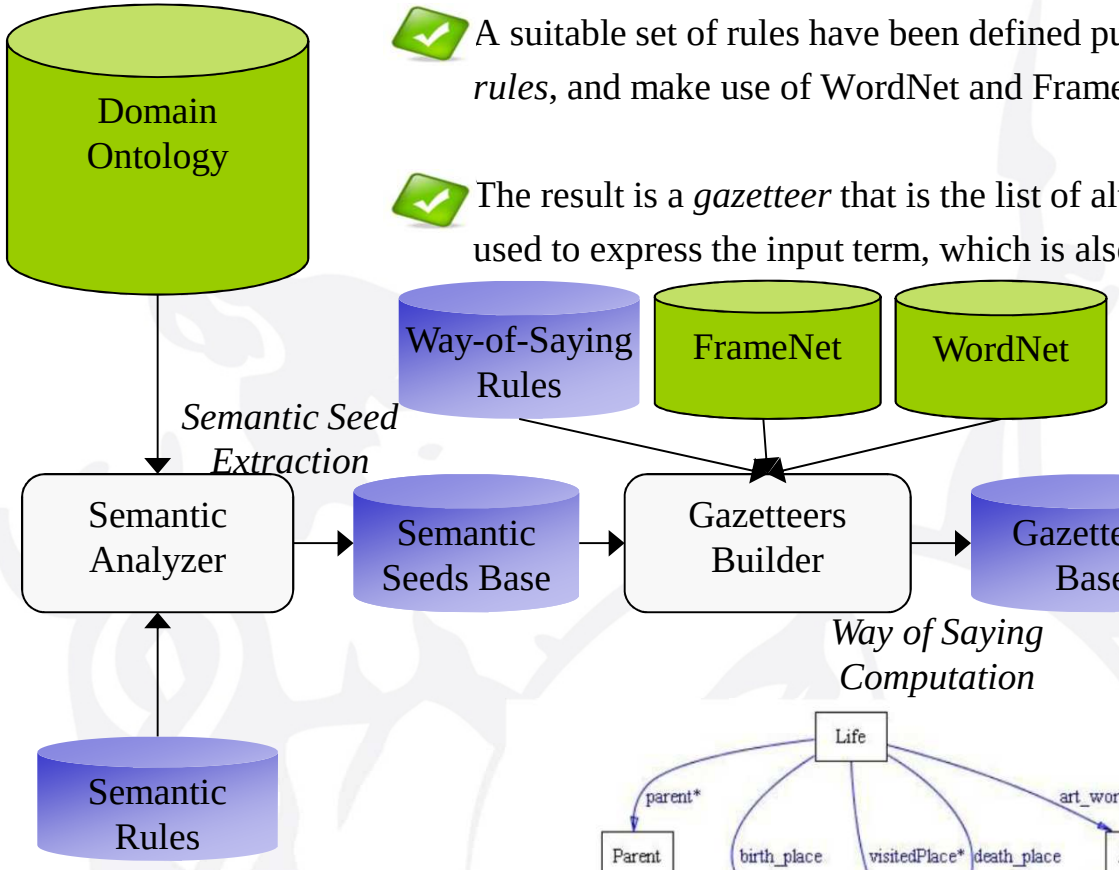
UNIVERSITÀ DEGLI STUDI DI PALERMO

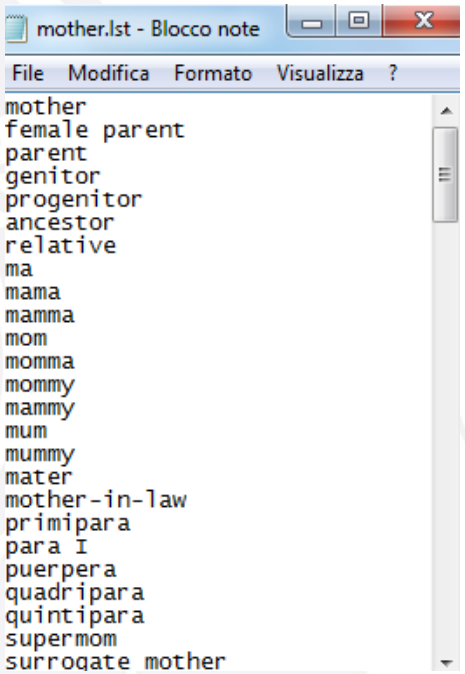
Gazetteers Builder

Using the ontology as the only source for the lexicon leads to very poor annotation due to the limited syntax that can be inferred from the model.

A suitable set of rules have been defined purposely; they're called *way-of-saying rules*, and make use of WordNet and FrameNet as their information sources.

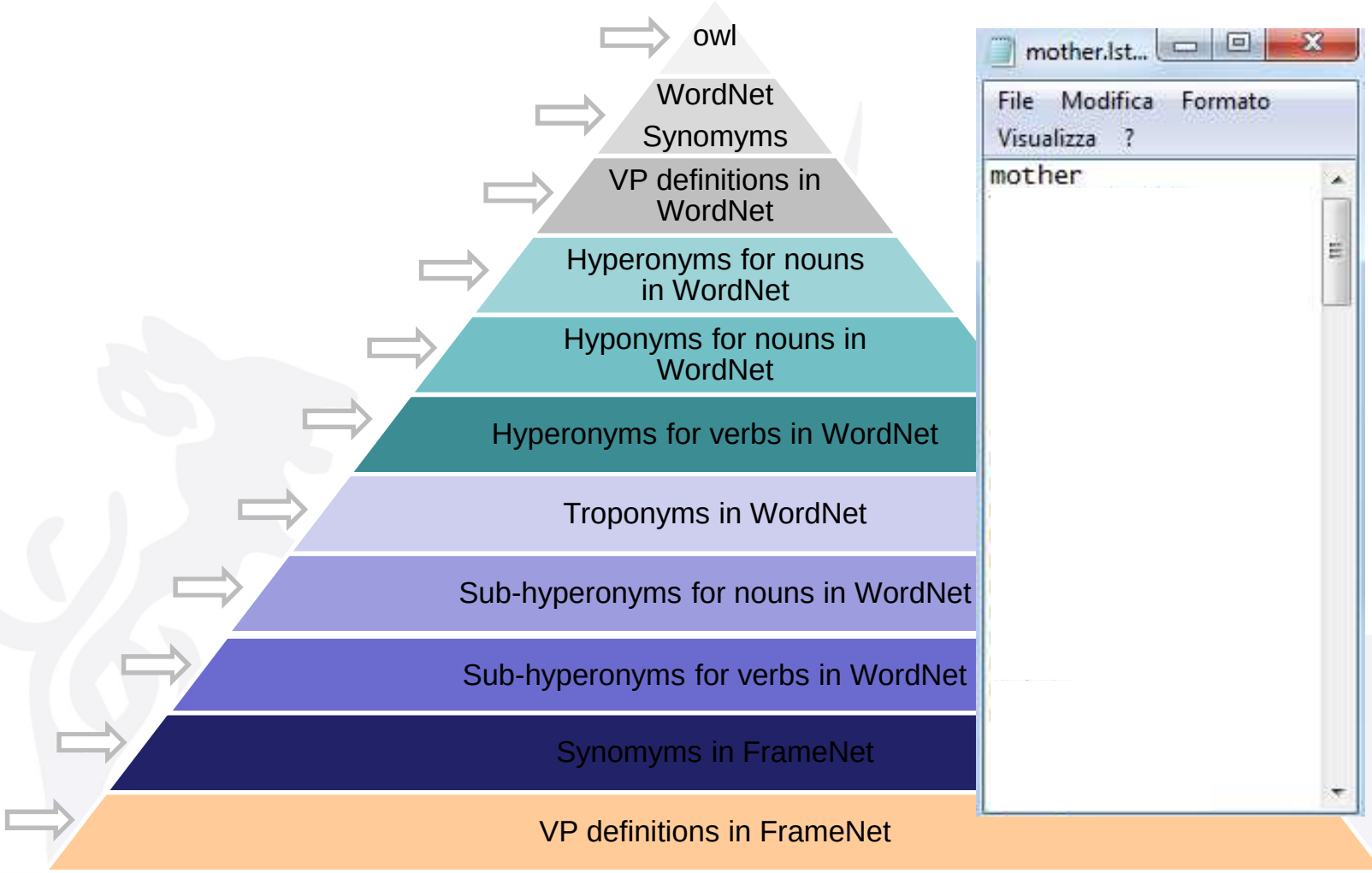
The result is a *gazetteer* that is the list of alternative words and/or ways of saying used to express the input term, which is also included in the list.





UNIVERSITÀ DEGLI STUDI DI PALERMO

Way-of-Saying Rules

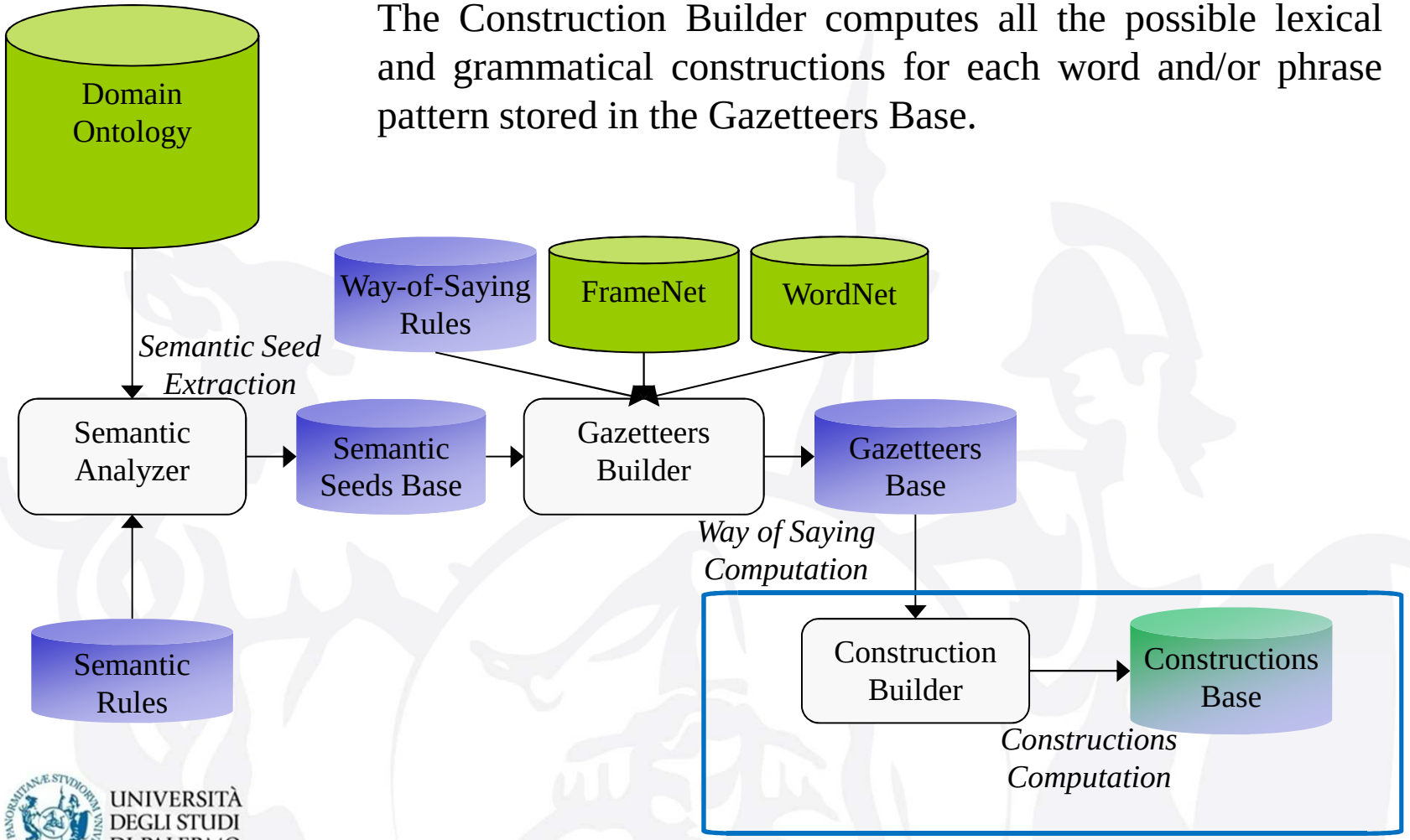


UNIVERSITÀ DEGLI STUDI DI PALERMO

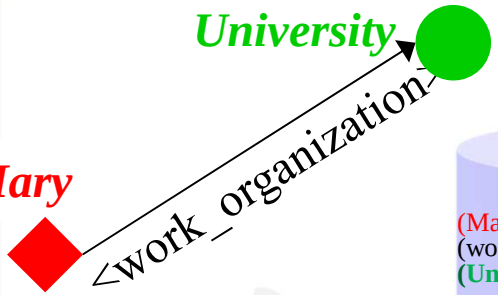
A. Pipitone, R. Pirrone

Construction Builder

The Construction Builder computes all the possible lexical and grammatical constructions for each word and/or phrase pattern stored in the Gazetteers Base.



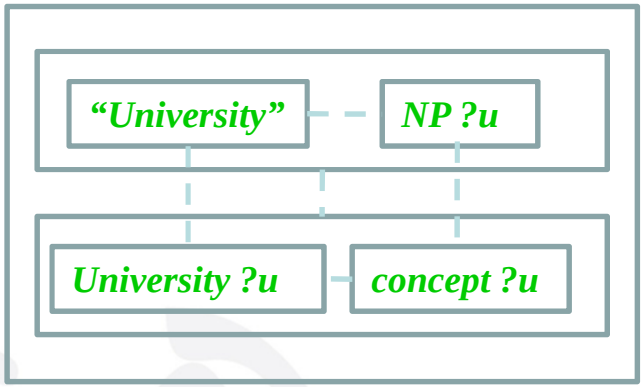
Construction Builder: lexical constructions

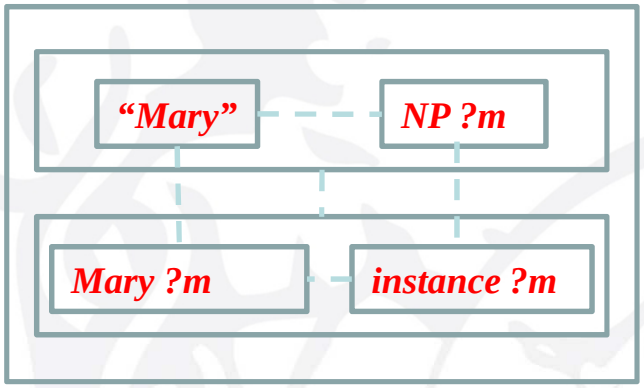


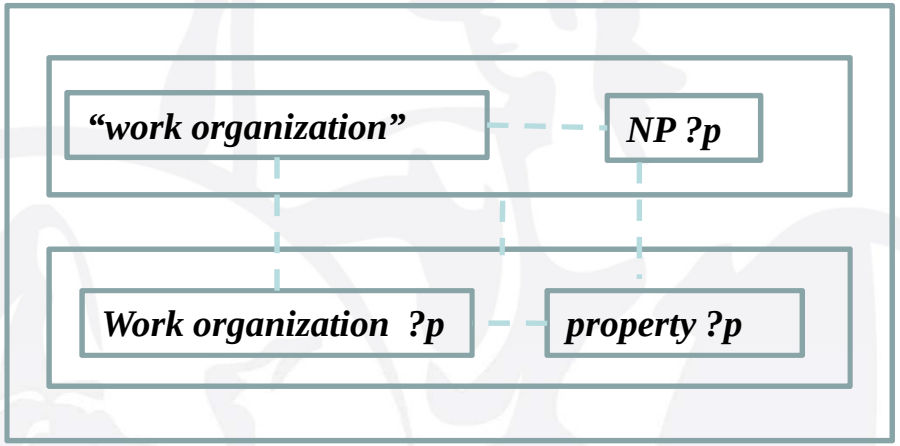
Semantic Seeds Base

(Mary?m)
(work_organization ?m ?p ?u)
(University ?u)

(Mary ?m)(work_organization ?m ?p ?u) (University ?u)

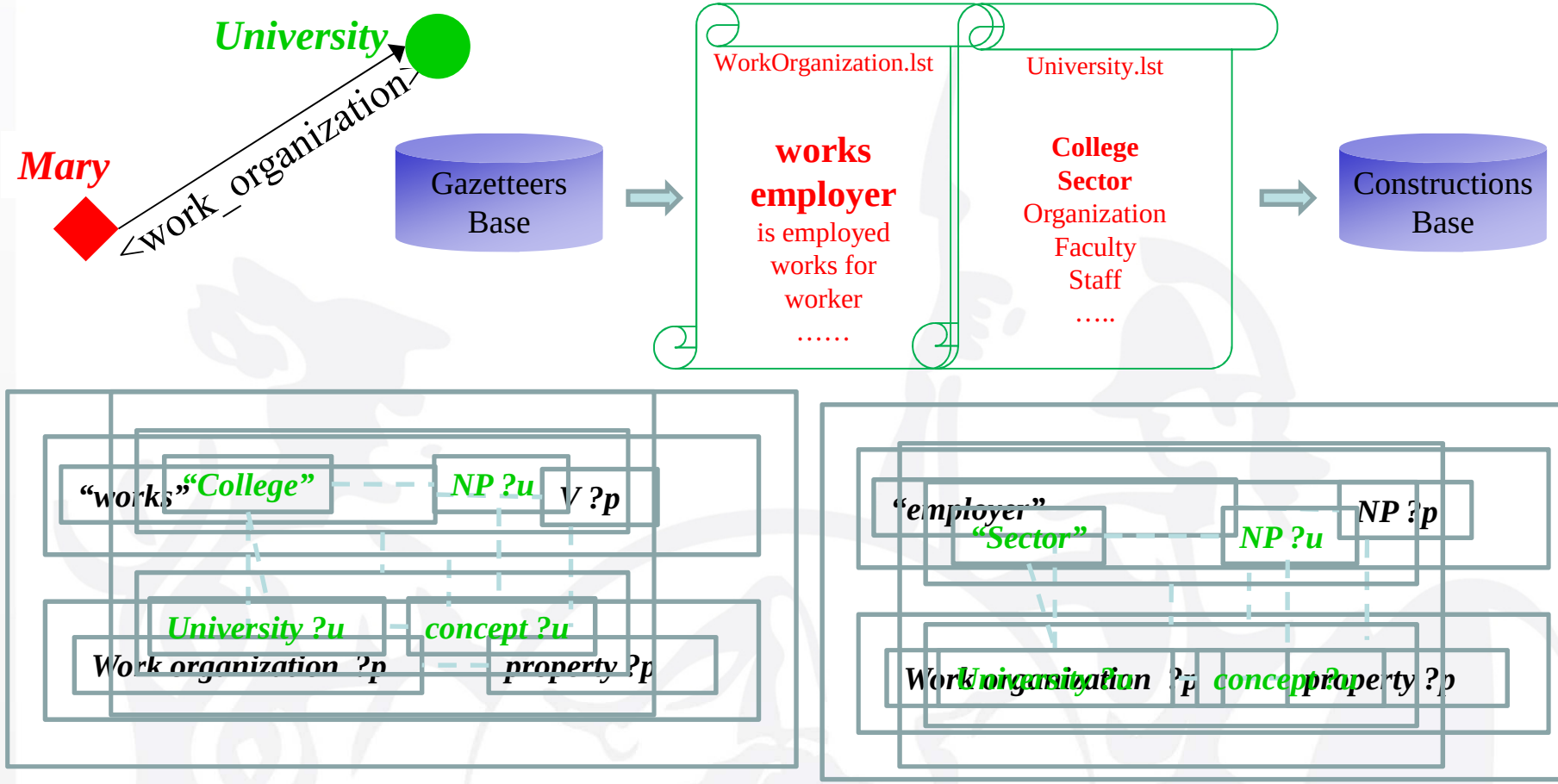






UNIVERSITÀ DEGLI STUDI DI PALERMO

Construction Builder: synonymous lexical constructions



Grammatical constructions are all the possible *SPO* conjunctions Induced by the ontology structure

UNIVERSITÀ DEGLI STUDI DI PALERMO

A. Pipitone, R. Pirrone

Context disambiguation: a possible approach

- Disambiguation occurs using semantic seeds.

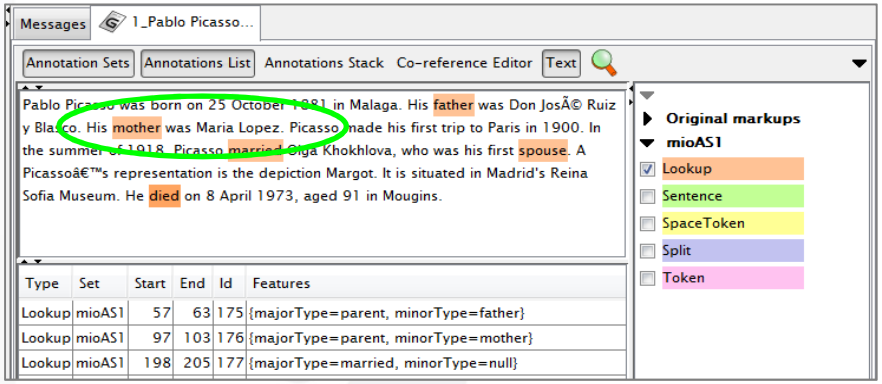
- Phrase pattern: "*born in*"
  - Means *birth_place* if the seed has the form <Person> <born> <Place>
  - Means *birth_date* if the seed has the form <Person> <born> <Year>

  - Direct use of the class names in case of object properties

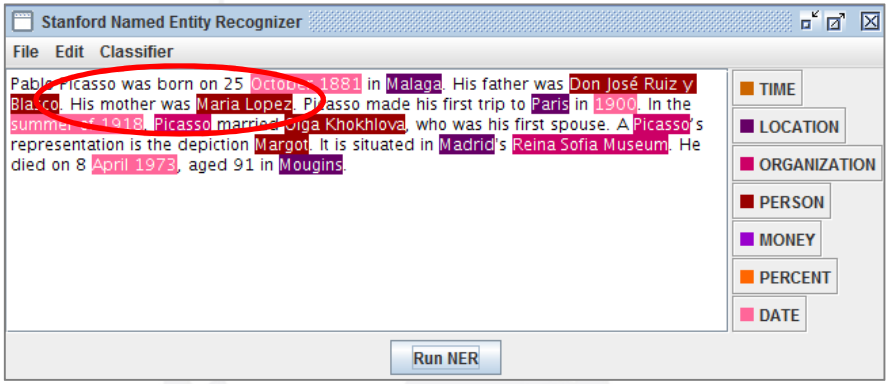
  - Use of NER categories against the property range in case of datatype properties

Context disambiguation: a possible approach

GATE Annotation



Stanford NER Annotation



His ***mother*** *as Maria Lopez.*

↓

Property: *mother*

↓

Range: *Parent*

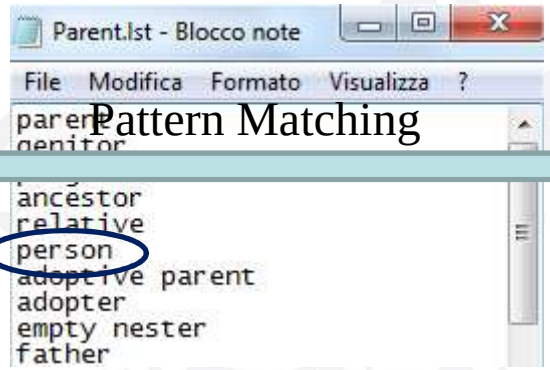
His mother was M ***Maria Lopez.***

↓

Text: *Maria Lopez*

↓

Label: *Person*



Pattern Matching

UNIVERSITÀ DEGLI STUDI DI PALERMO

Information: - -

A. Pipitone, R. Pirrone

Current work

- Applications to IR in Enterprise Information Systems
  - Collaboration with Consorzio Operativo – Gruppo MPS
- Development of Intelligent Tutoring Systems
  - European Grant I-TUTOR (LLP action KA3-ICT Multilateral Projects)