# VENSES
# A Linguistically-Based System for Semantic Evaluation

Rodolfo Delmonte
Department of Language Sciences
Università Ca' Foscari - Venezia (Italy)
website http://project.cgm.unive.it


Southampton - April, 12th

# OUTLINE

- System Architecture
- Quantitative Evaluation Module
- Linguistic Evaluation Module
- Results
- Conclusions

# COMPLETE SYSTEM ARCHITECTURE

- GETARUNS : Text Understanding and Answer Generation
- LOW LEVEL MODULE : Tokenizer, Tagger, Parser, Interpretation for Grammatical Relations and Semantic Roles Assignment, Quantifier Raising, Pronominal Binding
- HIGH LEVEL MODULE : Discourse Model, Centering-like Algorithm, Temporal Interpretation, Logical Form, Discourse Structure

# Discourse Model in Situation Semantics

- Collection of Facts
- Entities, Sits and Infons
- Semantic Types for Entities: Individuals, Sets, Classes

# Situation Semantics Representations

- Infon (Sit)
  - Index
  - Relation (Property)
  - List of Arguments with Semantic Roles
  - Polarity - 0 negative; 1 positive
  - Pair Indices for SpatioTemporal Locations

# From Discourse Models to AHDSs

- Clinton's new book is not big seller here.
  Clinton's book is a big seller.

```
entity(ind,id1,0,facts([
fact(infon18, inst_of, [ind:id1, class:place], 1, univ, univ),
fact(infon19, isa, [ind:id1, class:here], 1, univ, id1),
fact(infon41, isa, [arg:id1, arg:here], 1, tes(id_78f_1), id1),
fact(infon42, in, [arg:id5, locative:id1], 1, tes(id_78f_1), id1),
fact(infon5, main_sloc, here, 1, univ, _)])).


entity(class,id2,7,facts([
fact(infon7, 'Clinton-s_', [ind:id2], 1, univ, id1),
fact(infon8, seller, [nil:id2], 0, univ, id1),
fact(infon10, inst_of, [ind:id2, class:thing], 1, univ, univ),
fact(infon11, isa, [ind:id2, class:book], 1, univ, id1),
fact(infon34, new, [arg:id2], 1, univ, id1),
fact(infon53, isa, [arg:id2, arg:book], 1, id7, id1),
fact(infon54, 'Clinton-s_', [arg:id2], 1, id7, id1),
fact(infon55, seller, [nil:id2], 1, id7, id1)])).
```

```
be(adj-locat, here).
seller(ncmod, big).
new(ncmod, book).
book(ncmod-specif, 'Clinton-s_').
be(xcomp-prop, seller).
be(subj-theme_bound, book).
be(neg, not).


seller(ncmod, big).
book(ncmod-specif, 'Clinton-s_').
be(xcomp-prop, seller).
be(subj-theme_bound, book).
```

# Semantic Evaluator

- Input : DAGs turned into Augmented Head-Dependent Structures
  - Implicit Relations
  - Negation, Modals, Progressive Mood
  - Semantic Roles
  - Antecedents for Pronominal Expressions

# Semantic Evaluator

- From Integrity Constraint Violations to
- Semantic Equivalence
- Scoring Functions and Thresholds
- Bad Penalties for Propositional Level Interpretation

# Quantitative Evaluation

- Heads & Dependants
- Grammatical Relations & Semantic Roles
- Propositional Level Attributes
  - Weighted Score
  - Thresholds

# Quantitative Evaluation

- Heads & Dependants
  - Same Linguistic Description
  - Similar Linguistic Description
    - Single Word
    - Multiword

# Quantitative Evaluation

- Close Synonyms (WN Synsets)
- Belong to the same Semantic Field
- Morphologically Derivable
- Very Short Edit Distance
- Share Congruent Number of Inherent Features

# Quantitative Evaluation

- Grammatical Relations & Semantic Roles
  - Same GRs & SRs
  - Compatible GRs & SRs
    - Coupled to Equivalent AHDS

# Quantitative Evaluation

- Propositional Level Attributes
  - Negation
  - Modality
  - Opacity
    - Yes/NO Penalty

# Linguistic Evaluation

- Same Concept
  Reformulation, Paraphrase
  - WN Definition
  - Predicate-Argument Manipulation
  - Predicate-Adjunct Manipulation

# Linguistic Evaluation

- Same Best Predicates
- Core Arguments vs. Adjuncts
  - Non Conflicting Semantic Roles
  - Non Conflicting Grammatical Relations
  - Consistent Propositional Attributes

# System Evaluation

| | CWS | Accuracy | Precision | RECALL | F-Measure |
|---|---|---|---|---|---|
| Develop | 64.59 | 60.32 | 62.50 | 51.24 | 56.31 |
| Test | 62.06 | 59.75 | 61.75 | 51.25 | 56.01 |

# System Evaluation2

| cws | CD | QA | PP | IE | IR | RC | MT |
|---|---|---|---|---|---|---|---|
| Dev | 74.16 | 63.6 | 62.5 | 61.2 | 67.9 | 55.3 | 63.6 |
| Test | 70.52 | 49.6 | 84.71 | 61.5 | 72.1 | 57.3 | 50.9 |

# *Conclusions & Future Work*

- **Intrinsic Limitations in Deep Parse Approach**
  - **10% Error parsing + tagging**
  - **10% Error lack semantic resources**

# *Conclusions & Future Work*

- **Improve on semantic resources**
- **Apply the SE to the Complete System**
- **Attemp the Integrity Violation Approach**

THANKS FOR YOUR ATTENTION