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

Southampton - April, 12th
From Discourse Models to AHDSs

Clinton’s new book is not a big seller here.

Clinton’s book is a big seller.
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 & Dependents
- Grammatical Relations & Semantic Roles
- Propositional Level Attributes
- Weighted Score
- Thresholds
Quantitative Evaluation

- Heads & Dependents
  - 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

<table>
<thead>
<tr>
<th></th>
<th>CWS</th>
<th>Accuracy</th>
<th>Precision</th>
<th>RECALL</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop</td>
<td>64.59</td>
<td>60.32</td>
<td>62.50</td>
<td>51.24</td>
<td>56.31</td>
</tr>
<tr>
<td>Test</td>
<td>62.06</td>
<td>59.75</td>
<td>61.75</td>
<td>51.25</td>
<td>56.01</td>
</tr>
</tbody>
</table>

Southampton - April, 12th
## System Evaluation 2

<table>
<thead>
<tr>
<th></th>
<th>cws</th>
<th>CD</th>
<th>QA</th>
<th>PP</th>
<th>IE</th>
<th>IR</th>
<th>RC</th>
<th>MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dev</td>
<td>74.16</td>
<td>63.6</td>
<td>62.5</td>
<td>61.2</td>
<td>67.9</td>
<td></td>
<td>55.3</td>
<td>63.6</td>
</tr>
<tr>
<td>Test</td>
<td>70.52</td>
<td>49.6</td>
<td>84.71</td>
<td>61.5</td>
<td>72.1</td>
<td>57.3</td>
<td></td>
<td>50.9</td>
</tr>
</tbody>
</table>
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
- Attempt the Integrity Violation Approach
THANKS FOR YOUR ATTENTION