University of Oslo

Who Did What to Whom?

[A Contrastive Study of Syntacto-Semantic Dependencies]

Angelina Ivanova, Stephan Oepen, Lilja Øvrelid, and Dan Flickinger (Stanford University)

The 6th Linguistic Annotation Workshop (The LAW VI), ACL 2012
Introduction

Dependency representations:

- Useful for diverse tasks
  - Machine Translation
  - Semantic Search
  - Ontology Learning
  - Sentiment Analysis
  - Question Answering

- Can be obtained automatically
Motivation

Variety of incompatible representation formats that challenges the task of parser evaluation.

Example: A similar technique is almost impossible to apply to other crops, such as cotton, soybeans and rice.
Goals

Theoretical

Commonalities and differences between a broad range of dependency formats!

Practical

Making LinGO Redwoods Treebank accessible for a broader range of users
Dependency formats overview

- **PEST corpus:**
  - Language: **English**
  - Two sets: **10 sentences** and **15 sentences** from Wall Street Journal
  - CoNLL Syntactic Dependencies (CD)
  - CoNLL PropBank Semantics (CP)
  - Stanford Basic Dependencies (SB)
  - Stanford Collapsed Dependencies (SD)
  - Enju Predicate-Argument Structures (EP)

- **Conversion from LinGO ERG:**
  - DELPH-IN Syntactic Derivation Tree (DT)
  - DELPH-IN MRS-derived Dependencies (DM)
## Summary of dependency formats

<table>
<thead>
<tr>
<th>Abbr.</th>
<th>Format</th>
<th>Head status</th>
<th>Is the structure an acyclic tree?</th>
<th>Are the tokens connected?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>CoNLL Syntactic Dependencies</td>
<td>Functional</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CP</td>
<td>CoNLL PropBank Semantics</td>
<td>Substantive</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SB</td>
<td>Stanford Basic Dependencies</td>
<td>Substantive</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SD</td>
<td>Stanford Collapsed Dependencies</td>
<td>Substantive</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EP</td>
<td>Enju Predicate-Argument Structures</td>
<td>Substantive</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>DT</td>
<td>DELPH-IN Syntactic Derivation Tree</td>
<td>Functional</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>DM</td>
<td>DELPH-IN MRS-derived Dependencies</td>
<td>Substantive</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
**Root choice**

*A similar technique is almost impossible to apply to other crops, such as cotton, soybeans and rice.*

<table>
<thead>
<tr>
<th>CD</th>
<th>CoNLL Syntactic Dependencies</th>
<th>is</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>CoNLL PropBank Semantics</td>
<td>-</td>
</tr>
<tr>
<td>SB</td>
<td>Stanford Basic Dependencies</td>
<td>impossible</td>
</tr>
<tr>
<td>SD</td>
<td>Stanford Collapsed Dependencies</td>
<td>impossible</td>
</tr>
<tr>
<td>EP</td>
<td>Enju Predicate-Argument Structures</td>
<td>is</td>
</tr>
<tr>
<td>DT</td>
<td>DELHP-IN Syntactic Derivation Tree</td>
<td>is</td>
</tr>
<tr>
<td>DM</td>
<td>DELPH-IN MRS-derived Dependencies</td>
<td>almost</td>
</tr>
</tbody>
</table>
Conjunction

CoNLL Syntactic Dependencies (CD)

CoNLL PropBank Semantics (CP)

Stanford Basic Dependencies (SB)

Stanford Collapsed Dependencies (SD)

Enju Predicate-Argument Structures (EP)

DELPH-IN Derivation Tree (DT)

DELPH-IN MRS-derived Dependencies (DM)
Infinitive

CoNLL Syntactic Dependencies (CD)
Enju Predicate-Argument Structures (EP)
DELPH-IN Syntactic Derivation Tree (DT)
Stanford Basic Dependencies (SB)
Stanford Collapsed Dependencies (SD)
CoNLL PropBank Semantics (CP)
DELPH-IN MRS-derived Dependencies (DM)
Article

CoNLL Syntactic Dependencies (CD)
Stanford Basic Dependencies (SB)
Stanford Collapsed Dependencies (SD)
DELPH-IN Syntactic Derivation Tree (DT)

Enju Predicate-Argument Structures (EP)
DELPH-IN MRS-derived Dependencies (DM)

CoNLL PropBank Semantics (CP)
Adjective

CoNLL Syntactic Dependencies (CD)
Stanford Basic Dependencies (SB)
Stanford Collapsed Dependencies (SD)
DELPH-IN Syntactic Derivation Tree (DT)

Enju Predicate-Argument Structures (EP)
DELPH-IN MRS-derived Dependencies (DM)

CoNLL PropBank Semantics (CP)
Preposition

CoNLL Syntactic Dependencies (CD)
Stanford Basic Dependencies (SB)

DELPH-IN Syntactic Derivation Tree (DT)

Enju Predicate-Argument Structures (EP)

DELPH-IN MRS-derived Dependencies (DM)

CoNLL PropBank Semantics (CP)
Stanford Collapsed Dependencies (SD)
Tough adjective

A similar technique is almost impossible to apply

The long-distance dependency is detected only in:

CoNLL PropBank Semantics (CP)
Enju Predicate-Argument Structures (EP)
DELPH-IN MRS-derived Dependencies (DM)
Pairwise Jaccard similarity on PEST

<table>
<thead>
<tr>
<th></th>
<th>CD</th>
<th>CP</th>
<th>SB</th>
<th>SD</th>
<th>EP</th>
<th>DT</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td></td>
<td>.171</td>
<td>.427</td>
<td>.248</td>
<td>.187</td>
<td>.488</td>
<td>.115</td>
</tr>
<tr>
<td>CP</td>
<td>.171</td>
<td></td>
<td>.171</td>
<td>.177</td>
<td>.122</td>
<td>.158</td>
<td>.173</td>
</tr>
<tr>
<td>SB</td>
<td>.427</td>
<td>.171</td>
<td></td>
<td>.541</td>
<td>.123</td>
<td>.319</td>
<td>.147</td>
</tr>
<tr>
<td>SD</td>
<td>.248</td>
<td>.177</td>
<td>.541</td>
<td></td>
<td>.14</td>
<td>.264</td>
<td>.144</td>
</tr>
<tr>
<td>DT</td>
<td>.488</td>
<td>.158</td>
<td>.319</td>
<td>.264</td>
<td>.192</td>
<td></td>
<td>.13</td>
</tr>
<tr>
<td>DM</td>
<td>.115</td>
<td>.173</td>
<td>.147</td>
<td>.144</td>
<td>.462</td>
<td>.13</td>
<td></td>
</tr>
</tbody>
</table>

DELPH-IN Syntactic Derivation Tree (DT) format is closer to CoNLL Syntactic Dependencies (CD)

DELPH-IN MRS-derived Dependencies (DM) are closer to Enju Predicate-Argument Structures (EP)
Goals

Theoretical

Commonalities and differences between a broad range of dependency formats!

Practical

Making LinGO Redwoods Treebank accessible for a broader range of users
The LinGO Redwoods Treebank

- **Language:** English
- **Size:** 45 000 utterances
- **Linguistic approach:** HPSG
- **Grammar:** LinGO ERG
- **Data:**
  - Verbmobil and e-commerce corpora
  - LOGON Norwegian-English MT corpus
  - English Wikipedia (from WeScience)
  - Brown corpus (SemCor)
  - other
DELPH-IN Syntactic Derivation Tree

Conversion to bilexical dependencies

A similar technique is
Special cases during the conversion

Special cases:

• Contracted negation: doesn’t does n’t

• Punctuation: bark. bark .

• Multiword expressions: such as such as

• Hyphenated words: end-state end-state
A similar technique is almost impossible to apply to other crops.
Elementary Dependency Structure

\[
\{ \\
\ldots \\
x_{33}: \text{cotton\_n\_1} \\
_5: \text{udef\_q(BV i_{38})} \\
x_{27}: \text{implicit\_conj(L-INDEX x_{33}, R-INDEX i_{38})} \\
_6: \text{udef\_q(BV x_{43})} \\
x_{43}: \text{soybeans\_nns\_u\_unknown} \\
i_{38}: \text{and\_c(L-INDEX x_{43}, R-INDEX x_{47})} \\
_7: \text{udef\_q(BV x_{47})} \\
x_{47}: \text{rice\_n\_1}\}
\]

cotton, soybeans and rice.
Conclusions

• Qualitative and quantitative comparison of various dependency formats

• Automatic mapping from HPSG representations to syntactic and semantic dependencies
Future work

• Release of the converted Redwoods treebanks and conversion software

• Modification of DELPH-IN MRS-derived Dependencies conversion into a dependency tree

• Training parsers on DELPH-IN Syntactic Derivation Tree and DELPH-IN MRS-derived Dependencies formats

• Experimentation in domain adaptation for parsing on Redwoods treebanks
Thank you for your attention!

Questions?