One Tree is not Enough
Cross-lingual Accumulative Structure Transfer
for Semantic Indeterminacy

Patrick Ziering♠ Lonneke van der Plas♦

♠Institute for Natural Language Processing, University of Stuttgart, Germany
♦Institute of Linguistics, University of Malta, Malta

7th September 2015
RANLP 2015
Introduction
Parsing Complex Nominals

- Revealing the internal structure of noun compounds
- Important for natural language understanding

natural language understanding

natural language

understanding

natural language

language understanding

natural language

language understanding

natural language

understanding
Semantic Indeterminacy

Sometimes: virtually no difference in meaning between structures

- PP-attachment [Hindle and Rooth, 1993]
  \[\Rightarrow They \textit{mined the roads along the coast}\]
Semantic Indeterminacy

Sometimes: virtually no difference in meaning between structures

- PP-attachment [Hindle and Rooth, 1993]

  \[ \Rightarrow \text{They mined the roads along the coast} \]

- Parsing of 4NCs (e.g., \text{oil price increase strategy}):
Semantic Indeterminacy

Sometimes: virtually no difference in meaning between structures

- PP-attachment [Hindle and Rooth, 1993]
  \[ \Rightarrow \text{They mined the roads along the coast} \]
- Parsing of 4NCs (e.g., \textit{oil price increase strategy}):
Semantic Indeterminacy

- Established phenomenon in previous work
- Semantic indeterminacy is most often discarded in syntactic analysis
  - Vadas [2009]
  - Lauer [1995]
Semantic Indeterminacy

- Established phenomenon in previous work
- Semantic indeterminacy is most often discarded in syntactic analysis
  - Vadas [2009]
  - Lauer [1995]
- Incorporating semantic indeterminacy is important for NLP
  → All possible antecedents in anaphora resolution:
    - oil price, price increase, oil price increase, 
    - price increase strategy, . . .
Cross-lingual Structure Transfer (CST)

- Behaghel’s [1909] First Law: *Elements which belong close together intellectually will also be placed close together*

- Spelling variations for noun compound translations

- **human rights violations**
  \[\Rightarrow\] German: *Verletzungen der Menschenrechte*
Cross-lingual Structure Transfer (CST)

- Behaghel’s [1909] First Law: *Elements which belong close together intellectually will also be placed close together*

- Spelling variations for noun compound translations
  - **human** rights violations
    - ⇒ **Verletzungen der Menschenrechte**
  - ⇒ **human** and **rights** belong closer together:
    ```plaintext
    human rights violations
    /          \\
    human       rights
    /   \       |
    violations
    ```

Ziering and Van der Plas
One Tree is not Enough - Cross-lingual Accumulative Structure Transfer for Semantic Indeterminacy
Evidence for Semantic Indeterminacy in Parallel Corpora

Noun compound translations lead to structure variations

- *tobacco advertising ban*
  - German: *Werbeverbot für Tabakerzeugnisse* ⇒ RIGHT
  - Danish: *forbuddet mod tobaksreklamer* ⇒ LEFT

These variations are visible in particular across languages

A monolingual perspective suffers from conventional language use

<table>
<thead>
<tr>
<th>Language</th>
<th>Bigram</th>
<th>freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>🇬🇧</td>
<td><em>tobacco advertising ban</em></td>
<td>205</td>
</tr>
<tr>
<td>🇬🇧</td>
<td><em>advertising ban</em></td>
<td>84</td>
</tr>
<tr>
<td>🇩🇪</td>
<td><em>Tabakwerbung</em></td>
<td>31</td>
</tr>
<tr>
<td>🇩🇪</td>
<td><em>Werbeverbot</em></td>
<td>96</td>
</tr>
</tbody>
</table>
Language-Isolated Deterministic Structure Transfer

Bottom-up Parser

Word Alignment

Aligned Word Distance

[Ziering and Van der Plas, 2015]
Language-Isolated Deterministic Structure Transfer

Bottom-up Parser

Word Alignment

Aligned Word Distance

[Ziering and Van der Plas, 2015]
Language-Isolated Deterministic Structure Transfer

[Ziering and Van der Plas, 2015]
Language-Isolated Deterministic Structure Transfer

[Ziering and Van der Plas, 2015]
Language-Isolated Deterministic Structure Transfer

Bottom-up Parser

Word Alignment

Aligned Word Distance

[Ziering and Van der Plas, 2015]
[Ziering and Van der Plas, 2015]
Language-Isolated Deterministic Structure Transfer

Bottom-up Parser

Word Alignment

Aligned Word Distance

[Ziering and Van der Plas, 2015]
Language-Isolated Deterministic Structure Transfer

Limitations of LIDST:

(A) Deterministic output for each individual language
   → No output, if overlapping constituents have the same AWD

(B) Deterministic output for language ensemble using majority vote
   → Cannot handle semantic indeterminacy
Tree Accumulative Methods

Ziering and Van der Plas

One Tree is not Enough - Cross-lingual Accumulative Structure Transfer for Semantic Indeterminacy
Tree Accumulative Methods

Motivation:

(A) Translations with which overlapping constituents have the same AWD can still provide partial results ⇒ Non-deterministic classification criterion

(B) Semantic indeterminacy can only be captured non-deterministically ⇒ Cross-lingual tree accumulation
For a given language $l$ and $k$-partite noun compound $kNC$:

1. We create all binary trees.
We annotate all tree nodes $N_i$ with the AWD of their children

$$N_i.AWD = \begin{cases} 
\text{leaf}(N_i) & \mapsto 0 \\
\text{else} & \mapsto AWD(N_i.L, N_i.R)
\end{cases}$$

- **air traffic control centres**
  
  $\rightarrow$ **centres de contrôle du trafic aérien**

---

Ziering and Van der Plas

One Tree is not Enough - Cross-lingual Accumulative Structure Transfer for Semantic Indeterminacy
Tree validation:
A tree is valid, if the AWD annotation is monotonically decreasing top down.

⇒ Only valid trees are returned
Full Tree Accumulative Structure Transfer (FAST)

- Tree accumulation of all valid trees from all languages
- Ranking of all trees by frequency

<table>
<thead>
<tr>
<th>Rank</th>
<th>Structure</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[ air traffic ] [ control centres ]</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>[ [ air traffic ] control ] centres</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>[ air [ traffic control ] ] centres</td>
<td>10</td>
</tr>
</tbody>
</table>
Sometimes an invalid full tree still contains a valid subtree

- *church development aid projects*

  \[ \text{progetti ecclesiastici di aiuti allo sviluppo} \]

  church development aid projects \(X_{ft}\)

  \[ \text{AWD} = 1 \]

  \[ \begin{align*}
    \text{church} & \quad \text{development aid projects} \checkmark_{st} \\
    \text{AWD} & = 0 \quad \text{AWD} = 3
  \end{align*} \]

  \[ \begin{align*}
    \text{development aid} & \quad \text{projects} \\
    \text{AWD} & = 2 \quad \text{AWD} = 0
  \end{align*} \]

  \[ \begin{align*}
    \text{development} & \quad \text{aid} \\
    \text{AWD} & = 0 \quad \text{AWD} = 0
  \end{align*} \]
Subtree Accumulative Structure Transfer (SAST)

For all languages $l \in L$:

1. Creation of all binary trees
2. Tree annotation with AWD
3. Subtree validation and accumulation

Subtrees are assigned a subtree score ($sts$):

$$sts(st) = \frac{freq(st.\text{valid})}{|L| \cdot C_{\Delta}}$$

All full trees are assigned a full tree score ($fts$):

$$fts(ft) = \prod_{st \in ft} sts(st)$$

Full tree ranking according to $fts$
Difference between FAST and SAST

Given the language ensemble \{\text{\ding{123}}, \text{\ding{124}}, \text{\ding{125}}, \text{\ding{126}}\}:

- SAST ranks the left tree higher by exploiting the valid subtree derived from the Italian translation.
Experiments
## Dataset

- **Noun Compound Database**  
  [Ziering and Van der Plas, 2014]
- **OPUS Europarl corpus**  
  [Tiedemann, 2012]
- **10 languages out of 3 families:**
  - **Germanic**
    - English
    - Danish
    - Dutch
    - German
    - Swedish
  - **Romance**
    - French
    - Italian
    - Portuguese
    - Spanish
  - **Hellenic**
    - Greek

### Extraction of 3NCs and 4NCs by PoS patterns

- 24,848 3NC tokens (16,565 types)
- 1468 4NC tokens (1257 types)

---

Ziering and Van der Plas  
One Tree is not Enough - Cross-lingual Accumulative Structure Transfer for Semantic Indeterminacy
Gold Standard

- **3NC test set** [Ziering and Van der Plas, 2015]
  - 278 LEFT- or RIGHT-branched 3NC tokens
  - 120 cases of semantic indeterminacy

- **4NC test set**
  - 50 4NC tokens
  - Annotation guidelines of Vadas and Curran [2007]
  - Two trained annotators
    - Single tree (1, . . . , 5)
    - Semantic Indeterminacy \([i; \ldots; j]\)
  - Single trees \(\cup\) Semantic Indeterminacy \(\rightarrow\) 33 4NC tokens
Structure Retrieval

- How well does the system ranking fit to the set of gold trees?

- R-Precision

\[
R\text{-Prec}(kNC) = \frac{|\text{top-R(sys trees)} \cap \text{gold trees}|}{|\text{top-R(sys trees)}|}
\]

→ Mean R-Precision (MRP) as macro average

[Buckley and Voorhees, 2000]
Models in Comparison

**LIDST**  Language-Isolated Deterministic Structure Transfer

**LINDST**  Language-Isolated Non-Deterministic Structure Transfer
- Expansion of LIDST
- Frequency ranking instead of majority vote

**CHANCE**  Creates a random tree ranking

**FREQ**  Creates a tree ranking according to structure pattern frequencies
## Results

- **MRP on test set of 3NCs and 4NCs**

<table>
<thead>
<tr>
<th>System</th>
<th>MRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>93.7%</td>
</tr>
<tr>
<td>SAST</td>
<td>94.0%</td>
</tr>
<tr>
<td>LIDST</td>
<td>92.6%</td>
</tr>
<tr>
<td>LINDST</td>
<td>92.0%</td>
</tr>
<tr>
<td>FREQ</td>
<td>84.6%</td>
</tr>
<tr>
<td>CHANCE</td>
<td>62.5%</td>
</tr>
</tbody>
</table>

All CST systems outperform the baselines. FAST and SAST outperform LIDST and LINDST, but differences are small.
MRP on test set of 3NCs and 4NCs

<table>
<thead>
<tr>
<th>System</th>
<th>MRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>93.7%</td>
</tr>
<tr>
<td>SAST</td>
<td>94.0%</td>
</tr>
<tr>
<td>LIDST</td>
<td>92.6%</td>
</tr>
<tr>
<td>LINDST</td>
<td>92.0%</td>
</tr>
<tr>
<td>FREQ</td>
<td>84.6%</td>
</tr>
<tr>
<td>CHANCE</td>
<td>62.5%</td>
</tr>
</tbody>
</table>

All CST systems outperform the baselines
## Results

- **MRP on test set of 3NCs and 4NCs**

<table>
<thead>
<tr>
<th>System</th>
<th>MRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>93.7%</td>
</tr>
<tr>
<td>SAST</td>
<td>94.0%</td>
</tr>
<tr>
<td>LIDST</td>
<td>92.6%</td>
</tr>
<tr>
<td>LINDST</td>
<td>92.0%</td>
</tr>
<tr>
<td>FREQ</td>
<td>84.6%</td>
</tr>
<tr>
<td>CHANCE</td>
<td>62.5%</td>
</tr>
</tbody>
</table>

- All CST systems outperform the baselines
- **FAST and SAST outperform LIDST and LINDST, but differences are small**
Results

Evaluation on test set of 4NCs

<table>
<thead>
<tr>
<th>System</th>
<th>MRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>70.0%</td>
</tr>
<tr>
<td>SAST</td>
<td>69.5%</td>
</tr>
<tr>
<td>LIDST</td>
<td>54.5%</td>
</tr>
<tr>
<td>LINDST</td>
<td>62.9%</td>
</tr>
<tr>
<td>FREQ</td>
<td>60.1%</td>
</tr>
<tr>
<td>CHANCE</td>
<td>32.0%</td>
</tr>
</tbody>
</table>

FAST and SAST significantly outperform LI(N)DST in MRP.
### Results

- **Evaluation on test set of 4NCs**

<table>
<thead>
<tr>
<th>System</th>
<th>MRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>70.0%</td>
</tr>
<tr>
<td>SAST</td>
<td>69.5%</td>
</tr>
<tr>
<td>LIDST</td>
<td>54.5%‡</td>
</tr>
<tr>
<td>LINDST</td>
<td>62.9%‡</td>
</tr>
<tr>
<td>FREQ</td>
<td>60.1%</td>
</tr>
<tr>
<td>CHANCE</td>
<td>32.0%</td>
</tr>
</tbody>
</table>

- **FAST and SAST significantly outperform LI(N)DST in MRP**
Discussion
Accumulative CST also means a benefit in a deterministic take
→ partial evidence from several languages can be combined

energy efficiency action plan
→ plan de acción de eficiencia energética
Accumulative CST also means a benefit in a deterministic take partial evidence from several languages can be combined

- energy efficiency action plan
  → Aktionsplan zur Effizienz von Energie
Discussion

- Accumulative CST also means a benefit in a deterministic take → partial evidence from several languages can be combined

- Both languages fail to provide a single deterministic output

- In contrast:
  the cross-lingual tree accumulation provides the correct tree:
  
  energy efficiency action plan

  energy  efficiency

  action   plan
Ziering and Van der Plas

One Tree is not Enough - Cross-lingual Accumulative Structure Transfer for Semantic Indeterminacy

Conclusion
Conclusion

- Two models of accumulative cross-lingual structure transfer
  1. FAST
  2. SAST

  → Outperform previous CST approach significantly on 4NCs

- Non-deterministic CST is suitable for semantic indeterminacy

- Cross-lingual tree accumulation combines partial results in a deterministic take
Ziering and Van der Plas

One Tree is not Enough - Cross-lingual Accumulative Structure Transfer for Semantic Indeterminacy
# Full Results Table on 4NC Test Set

<table>
<thead>
<tr>
<th>System</th>
<th>MRP</th>
<th>MP@1</th>
<th>MR@1</th>
<th>MP@2</th>
<th>MR@2</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td><strong>70.0%</strong></td>
<td><strong>72.7%</strong></td>
<td><strong>47.5%</strong></td>
<td><strong>60.6%</strong></td>
<td><strong>74.2%</strong></td>
</tr>
<tr>
<td>SAST</td>
<td>69.5%</td>
<td>69.7%</td>
<td>44.4%</td>
<td><strong>63.6%</strong></td>
<td><strong>78.8%</strong></td>
</tr>
<tr>
<td>LIDST</td>
<td>54.5%‡</td>
<td>69.7%</td>
<td>44.4%</td>
<td>47.0%‡</td>
<td>59.1%‡</td>
</tr>
<tr>
<td>LINDST</td>
<td>62.9%†</td>
<td>69.7%</td>
<td>44.4%</td>
<td>54.5%†</td>
<td>66.7%†</td>
</tr>
<tr>
<td>FREQ</td>
<td>60.1%</td>
<td>63.6%</td>
<td>38.4%</td>
<td><strong>56.1%</strong></td>
<td>65.2%</td>
</tr>
<tr>
<td>CHANCE</td>
<td>32.0%</td>
<td>39.4%</td>
<td>23.7%</td>
<td>33.3%</td>
<td>42.4%</td>
</tr>
</tbody>
</table>

- FAST and SAST significantly outperform LI(N)DST in MRP

**MP/R@k:** Macro average of Precision/Recall at $k$
### Full Results Table on 4NC Test Set

<table>
<thead>
<tr>
<th>System</th>
<th>MRP</th>
<th>MP@1</th>
<th>MR@1</th>
<th>MP@2</th>
<th>MR@2</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>70.0%</td>
<td>72.7%</td>
<td>47.5%</td>
<td>60.6%</td>
<td>74.2%</td>
</tr>
<tr>
<td>SAST</td>
<td>69.5%</td>
<td>69.7%</td>
<td>44.4%</td>
<td>63.6%</td>
<td>78.8%</td>
</tr>
<tr>
<td>LIDST</td>
<td>54.5%†</td>
<td>69.7%</td>
<td>44.4%</td>
<td>47.0%†</td>
<td>59.1%†</td>
</tr>
<tr>
<td>LINDST</td>
<td>62.9%†</td>
<td>69.7%</td>
<td>44.4%</td>
<td>54.5%†</td>
<td>66.7% †</td>
</tr>
<tr>
<td>FREQ</td>
<td>60.1%</td>
<td>63.6%</td>
<td>38.4%</td>
<td>56.1%</td>
<td>65.2%</td>
</tr>
<tr>
<td>CHANCE</td>
<td>32.0%</td>
<td>39.4%</td>
<td>23.7%</td>
<td>33.3 %</td>
<td>42.4 %</td>
</tr>
</tbody>
</table>

- FAST and SAST significantly outperform LI(N)DST in MRP
- No difference between the CST systems in MP/R@1

**MP/R@k**: Macro average of Precision/Recall at $k$
### Full Results Table on 4NC Test Set

<table>
<thead>
<tr>
<th>System</th>
<th>MRP</th>
<th>MP@1</th>
<th>MR@1</th>
<th>MP@2</th>
<th>MR@2</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>70.0%</td>
<td>72.7%</td>
<td>47.5%</td>
<td>60.6%</td>
<td>74.2%</td>
</tr>
<tr>
<td>SAST</td>
<td>69.5%</td>
<td>69.7%</td>
<td>44.4%</td>
<td>63.6%</td>
<td>78.8%</td>
</tr>
<tr>
<td>LIDST</td>
<td>54.5%†</td>
<td>69.7%</td>
<td>44.4%</td>
<td>47.0%†</td>
<td>59.1%†</td>
</tr>
<tr>
<td>LINDST</td>
<td>62.9%†</td>
<td>69.7%</td>
<td>44.4%</td>
<td>54.5%†</td>
<td>66.7%†</td>
</tr>
<tr>
<td>FREQ</td>
<td>60.1%</td>
<td>63.6%</td>
<td>38.4%</td>
<td>56.1%</td>
<td>65.2%</td>
</tr>
<tr>
<td>CHANCE</td>
<td>32.0%</td>
<td>39.4%</td>
<td>23.7%</td>
<td>33.3%</td>
<td>42.4%</td>
</tr>
</tbody>
</table>

- FAST and SAST significantly outperform LI(N)DST in MRP
- No difference between the CST systems in MP/R@1
- FAST and SAST significantly outperform LIDST in MP/R@2

MP/R@k: Macro average of Precision/Recall at k