1. Introduction

- Problem: UD is believed to be suboptimal for parsing (de Marneffe et al., 2014).
- Solution: Create a parsing representation.
- Focus of the study: verb groups

![Figure 1: MS verb group: the auxiliary is the head](image)

![Figure 2: UD/PDT verb group: the main verb is the head](image)

(Nilsson et al., 2006, 2007; Schwartz et al., 2012)

2.1 Transformation Algorithm: UD to MS

![Figure 4: UD sentence with a VG](image)

Algorithm:
1. Find main verb and collect auxiliaries set
2. Head of main verb becomes head of outermost auxiliary
3. Make a chain from outermost auxiliary to main verb

![Figure 5: Intermediate representation](image)

2.2 Back Transformation: MS to UD

1. Find main verb and collect auxiliaries set
2. Attach auxiliaries to main verb
3. Attach auxiliaries to main verb

We obtain 100% back transformation accuracy on all but 4 treebanks.

2.4 Pipeline

![Figure 6: MS representation](image)

2.5 Software

- Parser: MaltParser (Nivre et al., 2006) with default settings and UD (coarse) POS tags.
- Transformation algorithms: released as part of oDETTE (DEpendency Treebank Transformation and Evaluation).
  https://github.com/mdelhoneux/oDETTE

2.6 Data

<table>
<thead>
<tr>
<th>Language</th>
<th>UD</th>
<th>PDT</th>
<th>SDT</th>
<th>τ</th>
<th>oo</th>
<th>do</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>700</td>
<td>300</td>
<td>200</td>
<td>100</td>
<td>150</td>
<td>500</td>
</tr>
<tr>
<td>French</td>
<td>600</td>
<td>500</td>
<td>400</td>
<td>350</td>
<td>300</td>
<td>500</td>
</tr>
<tr>
<td>German</td>
<td>500</td>
<td>400</td>
<td>300</td>
<td>250</td>
<td>200</td>
<td>300</td>
</tr>
<tr>
<td>Greek</td>
<td>400</td>
<td>300</td>
<td>200</td>
<td>150</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>Italian</td>
<td>300</td>
<td>200</td>
<td>100</td>
<td>50</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Polish</td>
<td>200</td>
<td>150</td>
<td>50</td>
<td>25</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>Portuguese</td>
<td>100</td>
<td>50</td>
<td>25</td>
<td>10</td>
<td>5</td>
<td>25</td>
</tr>
</tbody>
</table>

![Table 2: Summary of Table 7](image)

3.3 Role of POS tags ambiguity

Were improvements in PDT and SDT the result of POS disambiguation?

We modify POS tags to create 3 treebanks:
- σ₀: original treebank
- σ₁: disambiguated treebank
- σ₂: ambiguous treebank

![Table 3: 2 main verb group POS tags in SDT](image)

3.4 Predicted vs gold POS tags

Can UD benefit from the transformation when using predicted POS tags?

It seems not.

![Table 4: LAS on Predicted POS tags](image)

4. Conclusion

- Verb groups should stay as in UD.
- Gains from transforming from PDT style to MS style in previous studies were probably obtained because the approach helped disambiguate POS tags.
- Future work
  - Looking at other parsing models.
  - More in-depth error analysis.
  - Looking at other representations (e.g., PPs).

References