**Werdy**: Recognition and Disambiguation of Verbs and Verb Phrases with Syntactic and Semantic Pruning

Luciano Del Corro, Rainer Gemulla, Gerhard Weikum

**Contributions**

1. **Recognition** of dictionary entries in text (single and multi-word)
2. Improve existing disambiguation methods via **syntactic and semantic pruning**
3. Construction of a **public corpus** of \(< verb\_sense, object\_sense >\) pairs (+17,000 collected)

**Overview**

1. Dependency parse
2. Identify clauses
3. Identify entries
4. Syntactic Pruning
5. Semantic Pruning
6. Disambiguate

**Find all dictionary entries in text**

**Challenge**: discontinuous entries, multiple senses

```
He takes my hand and a deep breath.
```

```
She must attend to this matter.
```

Clauses type: SVA(dverbial)

**Syntactic pruning**: if clause structure does not match

```
PRP nsubj VBZ aux prep to DT NN
```

5 possible senses (e.g. attend a class)
1 sense remaining (WordNet)
(take charge of or deal with)

```
PRP nsubj VBZ aux to DT NN
```

35 possible senses (e.g. play sports)
1 sense remaining (perform music on …)

**Semantic pruning**: if pairs not in our VOS repository

```
PRP nsubj VBZ aux to DT NN
```

Which senses of play have guitar as a possible object?
VOS repository: \(< play\_7, instrument\_6 >\)

For each verb sense collect all sense tagged nouns appearing in its WordNet gloss:

**play-1**: participate in \(< games\_1,3 >\) or \(< sport\_1 >\)
(games and sport are sense tagged)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Gold Entry</th>
<th>Pruning</th>
<th>MFS backoff</th>
<th>threshold depth</th>
<th>Verbs (clause heads)</th>
<th>F1</th>
<th>F1 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>SimpleLex (DR)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>6</td>
<td>72.28</td>
<td>75.27</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
<td>73.70</td>
<td>52.82</td>
<td>61.17</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
<td>74.33</td>
<td>72.57</td>
<td>73.44</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
<td>70.30</td>
<td>77.57</td>
<td>78.52</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
<td>69.85</td>
<td>50.54</td>
<td>58.65</td>
</tr>
</tbody>
</table>

**VOS repository**

For each verb sense collect all sense tagged nouns appearing in its WordNet gloss:

**play-1**: participate in \(< games\_1,3 >\) or \(< sport\_1 >\)

(games and sport are sense tagged)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Gold Entry</th>
<th>Pruning</th>
<th>MFS backoff</th>
<th>threshold depth</th>
<th>Verbs (clause heads)</th>
<th>F1</th>
<th>F1 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFS</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>76.61</td>
<td>74.62</td>
<td>75.60</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>73.67</td>
<td>71.92</td>
<td>72.79</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>1</td>
<td>77.75</td>
<td>76.24</td>
<td>76.99</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>1</td>
<td>71.24</td>
<td>69.48</td>
<td>71.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Gold Entry</th>
<th>Pruning</th>
<th>MFS backoff</th>
<th>threshold depth</th>
<th>Verbs (clause heads)</th>
<th>F1</th>
<th>F1 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>SimpleLex (DR)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>6</td>
<td>72.28</td>
<td>75.27</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
<td>73.70</td>
<td>52.82</td>
<td>61.17</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
<td>74.33</td>
<td>72.57</td>
<td>73.44</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
<td>70.30</td>
<td>77.57</td>
<td>78.52</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
<td>69.85</td>
<td>50.54</td>
<td>58.65</td>
</tr>
</tbody>
</table>

- L. Del Corro, R. Gemulla, G. Weikum Werdy: Recognition and Disambiguation of Verbs and Verb Phrases with Syntactic and Semantic Pruning, EMNLP 2014
- L. Del Corro, R. Gemulla, ClausIE: Clause-Based Open Information Extraction, WWW, 2013

Best previous (F1): 76.26  Werdy: 81.18

max planck institut informatik