Treebank Translation for Cross-Lingual Parser Induction

Jörg Tiedemann\textsuperscript{1} Željko Agić\textsuperscript{2} Joakim Nivre\textsuperscript{1}

\textsuperscript{1}Department of Linguistics and Philology, Uppsala University
\textsuperscript{2}Department of Linguistics, University of Potsdam

CoNLL 2014, 2014-06-27
Motivation

The World is not English only many languages on the Web; most are under-resourced

Websites
- English: 57%
- Chinese: 5%
- Spanish: 5%
- Japanese: 5%
- Russian: 5%
- German: 7%
- Portuguese: 2%
- Italian: 2%
- French: 4%
- Others: 10%

Internet Users
- English: 27%
- Chinese: 25%
- Spanish: 8%
- Japanese: 5%
- Arabic: 3%
- German: 4%
- Portuguese: 4%
- French: 3%
- Russian: 3%
- Korean: 2%
- Others: 17%

> 2 billion Internet users
> 12 billion indexed web pages

Sources: W3Techs.com, Internet World Stats, WorldWideWebSize.com
Motivation

There are languages out there that require processing, but lack the required resources (Bender, 2011; Bender, 2013).

- most of World languages under-resourced (META-NET LWP, 2012)
- uniform language processing
  - lack of resources
  - *balkanization* – the one-scheme-per-language rule

- we focus on dependency parsing
- Is there a dependency treebank for… Croatian? Slovene?
Approaches

- annotation projection
- model transfer

- unsupervised
  - not addressed here
  - performance generally below previous two
Annotation projection

- take a parallel corpus
- word-align it
- parse it for syntactic dependencies
- project the annotation via alignment

- some variations
  - one side of parallel corpus is a treebank (rare)
  - word alignments are manual (rare)
  - usually relies on automatic word alignment and dependency parsing
    (Yarowsky et al., 2001; Hwa et al., 2005)

✓ language-specific features
× noise from parsing, alignment, projection
Model transfer

- train model on source language treebank
- rely on common features
- apply model on target language

- approaches
  - delexicalization (Zeman & Resnik, 2008; McDonald et al., 2013)
  - data point selection (Søgaard, 2011)
  - multi-source transfer (McDonald et al., 2011)
  - cross-lingual word clusters (Täckström et al., 2012)

✓ no resources required for target, no alignment and projection noise
✗ poor feature model
Treebank translation

- train a source-target SMT system
- translate source treebank into target language
- project annotations
- train dependency parser on synthetic treebank
- do parsing
Treebank translation

- differs from annotation projection
  - ✓ no source parsing noise
  - ✓ word alignment not separated, better for synthetic data
- and from model transfer
  - ✓ lexicalization
  - ✓ allows full feature set in target language
  - ✓ no assumptions on language universals
- potential issues
  - ✗ annotation projection noise still remains
  - ✗ quality of SMT
Setup

- treebanks
  - Google Universal Treebanks 1.0 (McDonald et al., 2013)
  - Universal POS (Petrov et al., 2012)
  - (adapted) Stanford Dependencies
  - excluded Korean as outlier: 5 languages
  - reliable cross-lingual dependency parsing assessment
  - existing train-dev-test split

- parsing
  - MaltParser (Nivre et al., 2007)
  - MaltOptimizer chooses optimal configuration (Ballesteros & Nivre, 2012)

- translation
  - Moses (Koehn et al., 2007), Europarl (Koehn, 2005)
Translation

- three scenarios
  - dictionary lookup
    - replace each word by default translation
    - no reordering
  - word-to-word
    - single-word translation table
    - distance-based reordering
    - 5-gram language model
  - phrase-based
    - standard phrase-based SMT model
- effects on non-projectivity
- projection requirements
Projection

- trivial for dictionary lookup
- same for word-to-word translation, non-projectivity occurs
Projection

- projection for phrase-based models
- multi-word alignments (m:n)
- labels must be projected as well
- one solution: dummy nodes (Hwa et al., 2005)

- our approach
  - use SMT phrase membership and phrase alignment information
  - use tree attachment heuristics
Projection

Input: source tree $S$, target sentence $T$, word alignment $A$, phrase segmentation $P$  
Output: syntactic heads head[], word attributes attr[]

1  treeSize = max_distance_to_root(S) ;
2  attr = [] ;
3  head = [] ;
4  for $t \in T$ do
5    if is_unaligned_trg($t$, A) then
6      for $t' \in \text{in_trg_phrase}(t, P)$ do
7        [$s_x, ..., s_y] = \text{aligned_to}(t') ;$
8        $\hat{s} = \text{find_highest}([s_x, ..., s_y], S) ;$
9        $\hat{t} = \text{find_aligned}(\hat{s}, S, T, A) ;$
10       attr[t] = DUMMY ;
11       head[t] = $\hat{t}$ ;
12    end
13  else
14    [$s_x, ..., s_y] = \text{aligned_to}(t) ;$
15    s = \text{find_highest}([s_x, ..., s_y], S) ;
16    attr[t] = attr(s) ;
17    $\hat{s} = \text{head_of}(s, S) ;$
18    $\hat{t} = \text{find_aligned}(\hat{s}, S, T, A) ;$
19    if $\hat{t} == t$ then
20      [$s_x, ..., s_y] = \text{in_src_phrase}(s, P) ;$
21      $s^* = \text{find_highest}([s_x, ..., s_y], S) ;$
22      $\hat{s} = \text{head_of}(s^*, S) ;$
23      $\hat{t} = \text{find_aligned}(\hat{s}, S, T, A) ;$
24      head[t] = $\hat{t}$ ;
25    end
26  end
27  end

function: find_aligned:

Input: node $s$, source tree $S$ with root ROOT, target sentence $T$, word alignment $A$  
Output: node $t^*$

1  if $s == \text{ROOT}$ then
2    return ROOT ;
3  end
4  while is_unaligned_src($s$, A) do
5    $s = \text{head_of}(s, S) ;$
6    if $s == \text{ROOT}$ then
7      return ROOT ;
8    end
9  end
10  $p = 0 ;$
11  $t^* = \text{undef} ;$
12  for $t' \in \text{aligned}(s, A)$ do
13    if position($t', T) > p$ then
14      $t^* = t'$ ;
15      $p = \text{position}(t', T) ;$
16    end
17  end
18  return $t^*$ ;

use phrase segmentation  
walk up the tree if unaligned  
attach to highest node  
heuristics for multiple targets: take right-most
Projection
Results
Baseline

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>72.13</td>
<td>87.50</td>
<td>78.54</td>
<td>77.51</td>
<td>81.28</td>
</tr>
<tr>
<td>en</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>es</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sv</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delexicalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>62.71</td>
<td>43.20</td>
<td>46.09</td>
<td>46.09</td>
<td>50.64</td>
</tr>
<tr>
<td>en</td>
<td>46.62</td>
<td>77.66</td>
<td>55.65</td>
<td>56.46</td>
<td>57.68</td>
</tr>
<tr>
<td>es</td>
<td>44.03</td>
<td>46.73</td>
<td>68.21</td>
<td>57.91</td>
<td>53.82</td>
</tr>
<tr>
<td>fr</td>
<td>43.91</td>
<td>46.75</td>
<td>59.65</td>
<td>67.51</td>
<td>52.01</td>
</tr>
<tr>
<td>sv</td>
<td>50.69</td>
<td>49.13</td>
<td>53.62</td>
<td>51.97</td>
<td>70.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>McDonald et al. (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>64.84</td>
<td>47.09</td>
<td>48.14</td>
<td>49.59</td>
<td>53.57</td>
</tr>
<tr>
<td>en</td>
<td>48.11</td>
<td>78.54</td>
<td>56.86</td>
<td>58.20</td>
<td>57.04</td>
</tr>
<tr>
<td>es</td>
<td>45.52</td>
<td>47.87</td>
<td>70.29</td>
<td>63.65</td>
<td>53.09</td>
</tr>
<tr>
<td>fr</td>
<td>45.96</td>
<td>47.41</td>
<td>62.56</td>
<td>73.37</td>
<td>52.25</td>
</tr>
<tr>
<td>sv</td>
<td>52.19</td>
<td>49.71</td>
<td>54.72</td>
<td>54.96</td>
<td>70.90</td>
</tr>
</tbody>
</table>
## Results

### Delexicalized models

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word-to-word</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>–</td>
<td>48.12 (4.92)</td>
<td>50.84 (4.75)</td>
<td>52.92 (6.83)</td>
<td>55.52 (4.88)</td>
</tr>
<tr>
<td>en</td>
<td>49.53 (2.91)</td>
<td>–</td>
<td>57.41 (1.76)</td>
<td><strong>58.53</strong> (2.07)</td>
<td><strong>57.82</strong> (0.14)</td>
</tr>
<tr>
<td>es</td>
<td>45.48 (1.45)</td>
<td>48.46 (1.73)</td>
<td>–</td>
<td>58.29 (0.38)</td>
<td>55.25 (1.43)</td>
</tr>
<tr>
<td>fr</td>
<td>46.59 (2.68)</td>
<td>47.88 (1.13)</td>
<td><strong>59.72</strong> (0.07)</td>
<td>–</td>
<td>52.31 (0.30)</td>
</tr>
<tr>
<td>sv</td>
<td><strong>52.16</strong> (1.47)</td>
<td><strong>49.14</strong> (0.01)</td>
<td>56.50 (2.88)</td>
<td>56.71 (4.74)</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phrase-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>–</td>
<td>45.43 (2.23)</td>
<td>47.26 (1.17)</td>
<td>49.14 (3.05)</td>
<td>53.37 (2.73)</td>
</tr>
<tr>
<td>en</td>
<td>49.16 (2.54)</td>
<td>–</td>
<td>57.12 (1.47)</td>
<td><strong>58.23</strong> (1.77)</td>
<td><strong>58.23</strong> (0.55)</td>
</tr>
<tr>
<td>es</td>
<td>46.75 (2.72)</td>
<td>46.82 (0.09)</td>
<td>–</td>
<td>58.22 (0.31)</td>
<td>54.14 (0.32)</td>
</tr>
<tr>
<td>fr</td>
<td>48.02 (4.11)</td>
<td><strong>49.06</strong> (2.31)</td>
<td><strong>60.23</strong> (0.58)</td>
<td>–</td>
<td>55.24 (3.23)</td>
</tr>
<tr>
<td>sv</td>
<td><strong>50.96</strong> (0.27)</td>
<td>46.12 (3.01)</td>
<td>55.95 (2.33)</td>
<td>54.71 (2.74)</td>
<td>–</td>
</tr>
</tbody>
</table>
## Results

### Lexicalized models

#### Lookup

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>de</td>
<td>–</td>
<td>48.63 (5.43)</td>
<td>52.66 (6.57)</td>
<td>52.06 (5.97)</td>
<td>58.78 (8.14)</td>
</tr>
<tr>
<td>en</td>
<td>48.59 (1.97)</td>
<td>–</td>
<td>57.79 (2.14)</td>
<td>57.80 (1.34)</td>
<td>62.21 (4.53)</td>
</tr>
<tr>
<td>es</td>
<td>47.36 (3.33)</td>
<td>49.13 (2.40)</td>
<td>–</td>
<td>62.24 (4.33)</td>
<td>57.50 (3.68)</td>
</tr>
<tr>
<td>fr</td>
<td>47.57 (3.66)</td>
<td><strong>54.06 (7.31)</strong></td>
<td><strong>66.31 (6.66)</strong></td>
<td>–</td>
<td>57.73 (5.72)</td>
</tr>
<tr>
<td>sv</td>
<td><strong>51.88 (1.19)</strong></td>
<td>48.84 (0.29)</td>
<td>54.74 (1.12)</td>
<td>52.95 (0.98)</td>
<td>–</td>
</tr>
</tbody>
</table>

#### Word-to-word

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>de</td>
<td>–</td>
<td>51.86 (3.74)</td>
<td>55.90 (5.06)</td>
<td>57.77 (4.85)</td>
<td>61.65 (6.13)</td>
</tr>
<tr>
<td>en</td>
<td><strong>53.80 (4.27)</strong></td>
<td>–</td>
<td>60.76 (3.35)</td>
<td>63.32 (4.79)</td>
<td><strong>62.93 (5.11)</strong></td>
</tr>
<tr>
<td>es</td>
<td>49.94 (4.46)</td>
<td>49.93 (1.47)</td>
<td>–</td>
<td><strong>65.60 (7.31)</strong></td>
<td>59.22 (3.97)</td>
</tr>
<tr>
<td>fr</td>
<td>52.07 (5.48)</td>
<td><strong>54.44 (6.56)</strong></td>
<td><strong>65.63 (5.91)</strong></td>
<td>–</td>
<td>57.67 (5.36)</td>
</tr>
<tr>
<td>sv</td>
<td>53.18 (1.02)</td>
<td>50.91 (1.77)</td>
<td>60.82 (4.32)</td>
<td>59.14 (2.43)</td>
<td>–</td>
</tr>
</tbody>
</table>

#### Phrase-based

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>de</td>
<td>–</td>
<td>50.89 (5.46)</td>
<td>52.54 (5.28)</td>
<td>54.99 (5.85)</td>
<td>59.46 (6.09)</td>
</tr>
<tr>
<td>en</td>
<td><strong>53.71 (4.55)</strong></td>
<td>–</td>
<td>60.70 (3.58)</td>
<td>62.89 (4.66)</td>
<td><strong>64.01 (5.78)</strong></td>
</tr>
<tr>
<td>es</td>
<td>49.59 (2.84)</td>
<td>48.35 (1.53)</td>
<td>–</td>
<td><strong>64.88 (6.66)</strong></td>
<td>58.99 (4.85)</td>
</tr>
<tr>
<td>fr</td>
<td>51.83 (3.81)</td>
<td><strong>53.81 (4.75)</strong></td>
<td><strong>65.55 (5.32)</strong></td>
<td>–</td>
<td>59.01 (3.77)</td>
</tr>
<tr>
<td>sv</td>
<td>53.22 (2.26)</td>
<td>49.06 (2.94)</td>
<td>58.41 (2.46)</td>
<td>58.04 (3.33)</td>
<td>–</td>
</tr>
</tbody>
</table>
Conclusions

▶ substantial improvements
  ▶ delexicalized up to +6.38 LAS
  ▶ lexicalized up to +7.31 LAS
▶ phrase-based projection fails to deliver
  ▶ quality of SMT
  ▶ unreliable POS mappings, link ambiguity
  ▶ no tree constraints
▶ overall results very positive
  ▶ lexical features
  ▶ reordering
  ▶ per-language parser optimization
▶ future work
  ▶ better translation
  ▶ better projection (Tiedemann, 2014)
  ▶ multi-synthetic-source transfer using n-best lists
  ▶ closely related languages (Agić et al., 2012)
Thank you for your attention. 😊
## Non-projectivity

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>en</th>
<th>es</th>
<th>fr</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>14.0</td>
<td>0.00</td>
<td>7.90</td>
<td>13.3</td>
<td>4.20</td>
</tr>
<tr>
<td><strong>Word-to-word</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>–</td>
<td>49.1</td>
<td>62.6</td>
<td>52.8</td>
<td>60.4</td>
</tr>
<tr>
<td>en</td>
<td>43.3</td>
<td>–</td>
<td>27.6</td>
<td>34.8</td>
<td>0.00</td>
</tr>
<tr>
<td>es</td>
<td>54.9</td>
<td>25.1</td>
<td>–</td>
<td>12.3</td>
<td>18.3</td>
</tr>
<tr>
<td>fr</td>
<td>68.2</td>
<td>39.6</td>
<td>32.8</td>
<td>–</td>
<td>57.8</td>
</tr>
<tr>
<td>sv</td>
<td>34.1</td>
<td>5.20</td>
<td>21.6</td>
<td>33.7</td>
<td>–</td>
</tr>
<tr>
<td><strong>Phrase-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>–</td>
<td>51.5</td>
<td>57.3</td>
<td>58.8</td>
<td>46.8</td>
</tr>
<tr>
<td>en</td>
<td>49.3</td>
<td>–</td>
<td>50.3</td>
<td>61.7</td>
<td>14.6</td>
</tr>
<tr>
<td>es</td>
<td>65.9</td>
<td>66.7</td>
<td>–</td>
<td>62.8</td>
<td>49.0</td>
</tr>
<tr>
<td>fr</td>
<td>58.0</td>
<td>53.7</td>
<td>44.7</td>
<td>–</td>
<td>38.2</td>
</tr>
<tr>
<td>sv</td>
<td>43.9</td>
<td>43.6</td>
<td>49.6</td>
<td>57.1</td>
<td>–</td>
</tr>
</tbody>
</table>
Link ambiguity

Tous ses produits sont de qualité et d'une fraicheur exemplaires.

All his products are high-quality and a cold mullet copies.

ADP NOUN CONJ ADP DET NOUN ADJ

DEP DET NOUN VERB NOUN ADP CONJ DET NOUN NOUN ADJ