### Supplementary Material: Appendices

#### A Details of UD Treebanks

The statistics of the Universal Dependency treebanks we used are summarized in Table 1.

<table>
<thead>
<tr>
<th>Language</th>
<th>Lang. Family</th>
<th>Treebank</th>
<th>#Sent.</th>
<th>#Token (w/o punct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic (ar)</td>
<td>Afro-Asiatic</td>
<td>PADT</td>
<td>train</td>
<td>6075</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>909</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>680</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>223881 (206041)</td>
</tr>
<tr>
<td>Bulgarian (bg)</td>
<td>IE.Slavic</td>
<td>BTB</td>
<td>train</td>
<td>8907</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>1115</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>1116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>124336 (106813)</td>
</tr>
<tr>
<td>Catalan (ca)</td>
<td>IE.Romance</td>
<td>AnCora</td>
<td>train</td>
<td>13123</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>1709</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>1846</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>417587 (371981)</td>
</tr>
<tr>
<td>Chinese (zh)</td>
<td>Sino-Tibetan</td>
<td>GSD</td>
<td>train</td>
<td>3997</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>986088 (84988)</td>
</tr>
<tr>
<td>Croatian (hr)</td>
<td>IE.Slavic</td>
<td>SET</td>
<td>train</td>
<td>6983</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>849</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>1057</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>154035 (135206)</td>
</tr>
<tr>
<td>Czech (cs)</td>
<td>IE.Slavic</td>
<td>PDT,CAC, CLTT,FicTree</td>
<td>train</td>
<td>102993</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>11311</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>12203</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1806230 (1542805)</td>
</tr>
<tr>
<td>Danish (da)</td>
<td>IE.Germanic</td>
<td>DDT</td>
<td>train</td>
<td>4383</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>564</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>565</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80378 (69219)</td>
</tr>
<tr>
<td>Dutch (nl)</td>
<td>IE.Germanic</td>
<td>Alpino, LassySmall</td>
<td>train</td>
<td>18038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>1394</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>1472</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>261180 (228902)</td>
</tr>
<tr>
<td>English (en)</td>
<td>IE.Germanic</td>
<td>EWT</td>
<td>train</td>
<td>12543</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>2077</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>204585 (180303)</td>
</tr>
<tr>
<td>Estonian (et)</td>
<td>Uralic</td>
<td>EDT</td>
<td>train</td>
<td>20827</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>2633</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>2737</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>287859 (240496)</td>
</tr>
<tr>
<td>Finnish (fi)</td>
<td>Uralic</td>
<td>TDT</td>
<td>train</td>
<td>12277</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>1364</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>1555</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>162621 (138524)</td>
</tr>
<tr>
<td>French (fr)</td>
<td>IE.Romance</td>
<td>GSD</td>
<td>train</td>
<td>14554</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>1478</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>416</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>356638 (316780)</td>
</tr>
<tr>
<td>German (de)</td>
<td>IE.Germanic</td>
<td>GSD</td>
<td>train</td>
<td>13814</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>799</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>977</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>263804 (229338)</td>
</tr>
<tr>
<td>Hebrew (he)</td>
<td>Afro-Asiatic</td>
<td>HTB</td>
<td>train</td>
<td>5241</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>484</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>491</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>137680 (123722)</td>
</tr>
<tr>
<td>Hindi (hi)</td>
<td>IE.Indic</td>
<td>HDTB</td>
<td>train</td>
<td>13304</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>1659</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>1684</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>281057 (262389)</td>
</tr>
<tr>
<td>Indonesian (id)</td>
<td>Austronesian</td>
<td>GSD</td>
<td>train</td>
<td>4477</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>559</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>557</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97531 (82617)</td>
</tr>
<tr>
<td>Italian (it)</td>
<td>IE.Romance</td>
<td>ISDT</td>
<td>train</td>
<td>13121</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>564</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>482</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>276019 (2446532)</td>
</tr>
<tr>
<td>Japanese (ja)</td>
<td>Japanese</td>
<td>GSD</td>
<td>train</td>
<td>7164</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>511</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>557</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>161900 (144045)</td>
</tr>
<tr>
<td>Korean (ko)</td>
<td>Korean</td>
<td>GSD, Kaist</td>
<td>train</td>
<td>27410</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>3016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>3276</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>353133 (312481)</td>
</tr>
<tr>
<td>Latin (la)</td>
<td>IE.Latin</td>
<td>PROIEL</td>
<td>train</td>
<td>15907</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>1234</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td>1260</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>171928 (171928)</td>
</tr>
<tr>
<td>Latvian (lv)</td>
<td>IE.Baltic</td>
<td>LVTB</td>
<td>train</td>
<td>5424</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dev</td>
<td>1051</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80666 (66270)</td>
</tr>
</tbody>
</table>
Table 1: Statistics of the UD Treebanks we used. For language family, “IE” is the abbreviation for Indo-European. “(w/o) punct” means the numbers of the tokens excluding “PUNCT” and “SYM”.

<table>
<thead>
<tr>
<th>Language</th>
<th>Family</th>
<th>Language Family</th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
<th>(w/o) punct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norwegian (no)</td>
<td>IE.Germanic</td>
<td>Bokmaal, Nynorsk</td>
<td>29870</td>
<td>4300</td>
<td>3450</td>
<td>489217(432597)</td>
</tr>
<tr>
<td>Polish (pl)</td>
<td>IE.Slavic</td>
<td>LFG, SZ</td>
<td>19874</td>
<td>2772</td>
<td>2827</td>
<td>167251(136504)</td>
</tr>
<tr>
<td>Portuguese (pt)</td>
<td>IE.Romance</td>
<td>Bosque, GSD</td>
<td>17993</td>
<td>1770</td>
<td>1681</td>
<td>462494(400343)</td>
</tr>
<tr>
<td>Romanian (ro)</td>
<td>IE.Romance</td>
<td>RRT</td>
<td>8043</td>
<td>752</td>
<td>729</td>
<td>185513(161429)</td>
</tr>
<tr>
<td>Russian (ru)</td>
<td>IE.Slavic</td>
<td>SynTagRus</td>
<td>48814</td>
<td>6584</td>
<td>6491</td>
<td>870674(71467)</td>
</tr>
<tr>
<td>Slovak (sk)</td>
<td>IE.Slavic</td>
<td>SNK</td>
<td>8483</td>
<td>1060</td>
<td>1061</td>
<td>805575(65042)</td>
</tr>
<tr>
<td>Slovenian (sl)</td>
<td>IE.Slavic</td>
<td>SSI, SST</td>
<td>8556</td>
<td>734</td>
<td>1898</td>
<td>132003(116730)</td>
</tr>
<tr>
<td>Spanish (es)</td>
<td>IE.Romance</td>
<td>GSD, AnCora</td>
<td>28492</td>
<td>3054</td>
<td>2147</td>
<td>827053(730062)</td>
</tr>
<tr>
<td>Swedish (sv)</td>
<td>IE.Germanic</td>
<td>Talbanken</td>
<td>4303</td>
<td>504</td>
<td>1219</td>
<td>666455(59268)</td>
</tr>
<tr>
<td>Ukrainian (uk)</td>
<td>IE.Slavic</td>
<td>IU</td>
<td>4313</td>
<td>577</td>
<td>783</td>
<td>750980(60976)</td>
</tr>
</tbody>
</table>

Table 2: Hyper-parameters in our experiments.

Table 2 summarizes the hyper-parameters that we used in our experiments. Most of them are similar to those in (Dozat and Manning, 2017) and (Ma et al., 2018).

<table>
<thead>
<tr>
<th>Layer</th>
<th>Hyper-Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>Word dimension</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>POS dimension</td>
<td>50</td>
</tr>
<tr>
<td>RNN</td>
<td>encoder layer</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>encoder size</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>arc MLP size</td>
<td>312</td>
</tr>
<tr>
<td></td>
<td>label MLP size</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>Dropout</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td></td>
<td>learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>batch size</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>$d_{model}$</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>$d_{ff}$</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>$d_{mlp}$</td>
<td>512</td>
</tr>
<tr>
<td></td>
<td>label MLP size</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>Dropout</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td></td>
<td>learning rate</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>batch size</td>
<td>80</td>
</tr>
</tbody>
</table>

B Hyper-Parameters

Table 2 summarizes the hyper-parameters that we used in our experiments. Most of them are similar to those in (Dozat and Manning, 2017) and (Ma et al., 2018).
C Details about augmented dependency types

<table>
<thead>
<tr>
<th>Type</th>
<th>Avg. Freq. (%)</th>
<th>#Lang.</th>
<th>Type</th>
<th>Avg. Freq. (%)</th>
<th>#Lang.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ADP, NOUN, case)</td>
<td>7.47</td>
<td>31</td>
<td>(PROPN, VERB, nsubj)</td>
<td>0.81</td>
<td>30</td>
</tr>
<tr>
<td>(PUNCT, VERB, punct)</td>
<td>6.91</td>
<td>30</td>
<td>(PRON, VERB, obj)</td>
<td>0.77</td>
<td>30</td>
</tr>
<tr>
<td>(NOUN, NOUN, nmod)</td>
<td>4.97</td>
<td>31</td>
<td>(NOUN, ROOT, root)</td>
<td>0.66</td>
<td>31</td>
</tr>
<tr>
<td>(ADJ, NOUN, amod)</td>
<td>4.92</td>
<td>31</td>
<td>(VERB, VERB, xcomp)</td>
<td>0.61</td>
<td>28</td>
</tr>
<tr>
<td>(DET, NOUN, det)</td>
<td>4.69</td>
<td>30</td>
<td>(VERB, VERB, ccomp)</td>
<td>0.60</td>
<td>30</td>
</tr>
<tr>
<td>(VERB, ROOT, root)</td>
<td>4.31</td>
<td>31</td>
<td>(ADP, PRON, case)</td>
<td>0.57</td>
<td>29</td>
</tr>
<tr>
<td>(NOUN, VERB, obl)</td>
<td>3.96</td>
<td>30</td>
<td>(AUX, NOUN, cop)</td>
<td>0.57</td>
<td>28</td>
</tr>
<tr>
<td>(NOUN, VERB, obj)</td>
<td>3.10</td>
<td>31</td>
<td>(ADV, ADJ, advmod)</td>
<td>0.54</td>
<td>29</td>
</tr>
<tr>
<td>(NOUN, VERB, nsubj)</td>
<td>2.89</td>
<td>31</td>
<td>(AUX, ADJ, cop)</td>
<td>0.50</td>
<td>27</td>
</tr>
<tr>
<td>(PUNCT, NOUN, punct)</td>
<td>2.75</td>
<td>30</td>
<td>(PROPN, VERB, obl)</td>
<td>0.48</td>
<td>29</td>
</tr>
<tr>
<td>(ADV, VERB, advmod)</td>
<td>2.43</td>
<td>31</td>
<td>(PRON, VERB, obl)</td>
<td>0.44</td>
<td>30</td>
</tr>
<tr>
<td>(AUX, VERB, aux)</td>
<td>2.29</td>
<td>28</td>
<td>(ADV, NOUN, advmod)</td>
<td>0.41</td>
<td>28</td>
</tr>
<tr>
<td>(PRON, VERB, nsubj)</td>
<td>1.53</td>
<td>30</td>
<td>(ADJ, ROOT, root)</td>
<td>0.39</td>
<td>29</td>
</tr>
<tr>
<td>(ADP, PROPN, case)</td>
<td>1.46</td>
<td>29</td>
<td>(PRON, NOUN, nmod)</td>
<td>0.39</td>
<td>22</td>
</tr>
<tr>
<td>(NOUN, NOUN, conj)</td>
<td>1.32</td>
<td>30</td>
<td>(NOUN, ADJ, obl)</td>
<td>0.37</td>
<td>25</td>
</tr>
<tr>
<td>(VERB, NOUN, acl)</td>
<td>1.31</td>
<td>31</td>
<td>(PROPN, PROPN, conj)</td>
<td>0.35</td>
<td>29</td>
</tr>
<tr>
<td>(Sconj, VERB, mark)</td>
<td>1.27</td>
<td>28</td>
<td>(NOUN, ADJ, nsubj)</td>
<td>0.35</td>
<td>30</td>
</tr>
<tr>
<td>(Cconj, VERB, cc)</td>
<td>1.18</td>
<td>30</td>
<td>(Cconj, ADJ, cc)</td>
<td>0.29</td>
<td>28</td>
</tr>
<tr>
<td>(PROPN, NOUN, nmod)</td>
<td>1.14</td>
<td>30</td>
<td>(PUNCT, NOUN, punct)</td>
<td>0.26</td>
<td>24</td>
</tr>
<tr>
<td>(Cconj, NOUN, cc)</td>
<td>1.13</td>
<td>30</td>
<td>(NOUN, NOUN, nsubj)</td>
<td>0.25</td>
<td>31</td>
</tr>
<tr>
<td>(NUM, NOUN, nummod)</td>
<td>1.11</td>
<td>31</td>
<td>(ADJ, ADJ, conj)</td>
<td>0.25</td>
<td>26</td>
</tr>
<tr>
<td>(PROPN, PROPN, flat)</td>
<td>1.09</td>
<td>26</td>
<td>(Cconj, PROPN, cc)</td>
<td>0.22</td>
<td>26</td>
</tr>
<tr>
<td>(VERB, VERB, conj)</td>
<td>1.05</td>
<td>30</td>
<td>(PRON, VERB, obj)</td>
<td>0.21</td>
<td>21</td>
</tr>
<tr>
<td>(PUNCT, PROPN, punct)</td>
<td>0.94</td>
<td>29</td>
<td>(ADV, ADV, advmod)</td>
<td>0.19</td>
<td>21</td>
</tr>
<tr>
<td>(VERB, VERB, advcl)</td>
<td>0.89</td>
<td>30</td>
<td>(NOUN, NOUN, appos)</td>
<td>0.18</td>
<td>23</td>
</tr>
<tr>
<td>(PUNCT, ADJ, punct)</td>
<td>0.89</td>
<td>30</td>
<td>(PROPN, VERB, obj)</td>
<td>0.17</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 3: Selected augmented dependency types sorted by their average frequencies. "#Lang." denotes in how many languages the specific type appears. Since the augmented dependency types can be in hundreds or larger than 1k, but mostly infrequent, we prune them according to average frequency and number of appearing languages. Our pruning criterion is “Freq > 0.1% and #Lang ≥ 20”.
D  Punctuation-included Evaluation on the test sets

<table>
<thead>
<tr>
<th>Language</th>
<th>SelfAtt-Graph</th>
<th>RNN-Graph</th>
<th>SelfAtt-Stack</th>
<th>RNN-Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>en</td>
<td>89.29/87.52</td>
<td>89.46/87.52</td>
<td>89.16/87.26</td>
<td>90.83/89.07</td>
</tr>
<tr>
<td>sv</td>
<td>70.9/72.69</td>
<td>79.94/72.99</td>
<td>79.24/72.24</td>
<td>81.44/73.98</td>
</tr>
<tr>
<td>fr</td>
<td>75.58/71.05</td>
<td>76.11/71.79</td>
<td>74.32/69.87</td>
<td>73.56/69.16</td>
</tr>
<tr>
<td>pt</td>
<td>73.07/65.30</td>
<td>72.82/66.38</td>
<td>71.61/63.96</td>
<td>71.21/63.76</td>
</tr>
<tr>
<td>da</td>
<td>74.03/66.52</td>
<td>74.99/67.67</td>
<td>73.76/66.15</td>
<td>75.81/67.76</td>
</tr>
<tr>
<td>es</td>
<td>70.98/63.84</td>
<td>71.50/64.40</td>
<td>69.54/62.44</td>
<td>69.73/62.37</td>
</tr>
<tr>
<td>it</td>
<td>78.19/73.77</td>
<td>78.63/74.31</td>
<td>76.52/72.11</td>
<td>78.29/73.84</td>
</tr>
<tr>
<td>hr</td>
<td>60.58/52.60</td>
<td>58.60/50.28</td>
<td>59.03/50.65</td>
<td>59.27/50.72</td>
</tr>
<tr>
<td>ca</td>
<td>70.47/62.37</td>
<td>70.96/62.85</td>
<td>68.91/60.87</td>
<td>68.79/60.45</td>
</tr>
<tr>
<td>pl</td>
<td>74.78/64.68</td>
<td>71.73/60.83</td>
<td>73.82/63.19</td>
<td>72.24/62.11</td>
</tr>
<tr>
<td>sl</td>
<td>57.57/51.16</td>
<td>56.32/50.25</td>
<td>54.58/48.18</td>
<td>57.31/50.81</td>
</tr>
<tr>
<td>nl</td>
<td>66.50/55.84</td>
<td>64.55/53.84</td>
<td>64.83/53.88</td>
<td>66.07/55.03</td>
</tr>
<tr>
<td>bg</td>
<td>66.92/59.59</td>
<td>66.45/59.54</td>
<td>66.05/58.59</td>
<td>68.10/61.01</td>
</tr>
<tr>
<td>ru</td>
<td>60.95/53.83</td>
<td>59.56/51.80</td>
<td>58.88/50.86</td>
<td>59.63/51.13</td>
</tr>
<tr>
<td>de</td>
<td>65.11/57.76</td>
<td>63.66/56.38</td>
<td>63.68/56.21</td>
<td>64.97/57.08</td>
</tr>
<tr>
<td>he</td>
<td>49.00/44.07</td>
<td>47.08/42.78</td>
<td>47.03/42.17</td>
<td>47.12/42.38</td>
</tr>
<tr>
<td>cs</td>
<td>66.53/49.52</td>
<td>66.95/49.66</td>
<td>64.50/47.72</td>
<td>65.98/48.46</td>
</tr>
<tr>
<td>ro</td>
<td>64.83/49.83</td>
<td>65.04/49.98</td>
<td>63.41/48.61</td>
<td>64.97/49.63</td>
</tr>
<tr>
<td>sk</td>
<td>63.50/45.88</td>
<td>63.08/45.45</td>
<td>61.74/44.12</td>
<td>62.15/44.57</td>
</tr>
<tr>
<td>id</td>
<td>40.46/25.52</td>
<td>39.54/24.74</td>
<td>38.37/23.55</td>
<td>39.26/24.25</td>
</tr>
<tr>
<td>lv</td>
<td>37.15/27.79</td>
<td>32.37/25.42</td>
<td>31.69/23.46</td>
<td>32.04/24.73</td>
</tr>
<tr>
<td>fi</td>
<td>33.96/17.99</td>
<td>33.08/16.96</td>
<td>31.68/16.04</td>
<td>32.81/16.17</td>
</tr>
<tr>
<td>et</td>
<td>36.90/28.52</td>
<td>30.94/23.55</td>
<td>32.65/24.92</td>
<td>26.80/19.49</td>
</tr>
<tr>
<td>zh*</td>
<td>27.83/21.25</td>
<td>18.39/12.59</td>
<td>20.33/13.56</td>
<td>15.01/9.75</td>
</tr>
<tr>
<td>Average</td>
<td>62.21/53.27</td>
<td>60.91/52.12</td>
<td>60.26/51.34</td>
<td>60.62/51.46</td>
</tr>
</tbody>
</table>

Table 4: Evaluations with punctuation included (average UAS%/LAS% over 5 runs) on the test sets. The patterns are similar to the punctuation-excluded evaluations in the main content. (Languages are sorted by the word-ordering distance to English, '*' refers to results of delexicalized models.)
E Results on the original training sets

<table>
<thead>
<tr>
<th>Language</th>
<th>SelfAtt-Graph</th>
<th>RNN-Graph</th>
<th>SelfAtt-Stack</th>
<th>RNN-Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>en°</td>
<td>90.35/88.40</td>
<td>90.44/88.31</td>
<td>90.18/88.06</td>
<td>91.82/89.89</td>
</tr>
<tr>
<td>no</td>
<td>80.72/72.45</td>
<td>80.59/72.41</td>
<td>80.06/71.60</td>
<td>81.46/72.75</td>
</tr>
<tr>
<td>sv</td>
<td>80.07/71.91</td>
<td>80.42/72.39</td>
<td>79.45/71.28</td>
<td>80.87/72.25</td>
</tr>
<tr>
<td>fr</td>
<td>79.31/74.73</td>
<td>79.99/75.52</td>
<td>78.62/74.02</td>
<td>76.84/72.22</td>
</tr>
<tr>
<td>pl</td>
<td>77.06/69.33</td>
<td>77.33/69.91</td>
<td>75.84/68.22</td>
<td>75.39/67.75</td>
</tr>
<tr>
<td>da</td>
<td>75.75/67.12</td>
<td>75.95/67.41</td>
<td>75.18/66.55</td>
<td>76.98/67.50</td>
</tr>
<tr>
<td>es</td>
<td>73.91/66.48</td>
<td>74.39/67.03</td>
<td>72.84/65.38</td>
<td>72.46/64.78</td>
</tr>
<tr>
<td>it</td>
<td>80.37/75.48</td>
<td>80.89/75.99</td>
<td>79.15/74.17</td>
<td>79.05/73.91</td>
</tr>
<tr>
<td>hr</td>
<td>61.57/52.40</td>
<td>59.74/50.37</td>
<td>59.94/50.43</td>
<td>60.44/50.68</td>
</tr>
<tr>
<td>ca</td>
<td>74.40/65.73</td>
<td>74.94/66.21</td>
<td>73.01/64.42</td>
<td>72.75/63.68</td>
</tr>
<tr>
<td>pl</td>
<td>75.32/63.26</td>
<td>73.12/59.76</td>
<td>74.28/61.46</td>
<td>73.21/61.02</td>
</tr>
<tr>
<td>uk</td>
<td>65.70/57.48</td>
<td>64.77/56.40</td>
<td>64.10/55.83</td>
<td>65.82/57.13</td>
</tr>
<tr>
<td>sl</td>
<td>69.13/58.92</td>
<td>67.35/56.87</td>
<td>67.74/57.08</td>
<td>68.95/58.26</td>
</tr>
<tr>
<td>nl</td>
<td>68.98/60.00</td>
<td>68.37/59.52</td>
<td>68.22/59.02</td>
<td>69.16/60.11</td>
</tr>
<tr>
<td>bg</td>
<td>80.25/68.88</td>
<td>78.39/67.03</td>
<td>79.19/66.76</td>
<td>79.66/68.22</td>
</tr>
<tr>
<td>ru</td>
<td>60.50/51.35</td>
<td>59.55/50.17</td>
<td>59.01/49.71</td>
<td>60.71/51.57</td>
</tr>
<tr>
<td>de</td>
<td>67.23/58.27</td>
<td>66.64/57.48</td>
<td>66.10/56.89</td>
<td>65.88/56.63</td>
</tr>
<tr>
<td>he</td>
<td>58.32/49.80</td>
<td>57.75/49.07</td>
<td>56.36/47.62</td>
<td>58.79/43.83</td>
</tr>
<tr>
<td>cs</td>
<td>63.04/53.92</td>
<td>61.75/52.91</td>
<td>61.11/51.91</td>
<td>62.21/52.48</td>
</tr>
<tr>
<td>ro</td>
<td>65.31/54.22</td>
<td>63.17/52.16</td>
<td>63.03/51.95</td>
<td>61.78/50.52</td>
</tr>
<tr>
<td>sk</td>
<td>76.07/62.75</td>
<td>74.67/61.15</td>
<td>75.93/61.97</td>
<td>75.37/60.94</td>
</tr>
<tr>
<td>id</td>
<td>47.92/41.93</td>
<td>45.07/39.91</td>
<td>46.23/40.16</td>
<td>45.62/39.67</td>
</tr>
<tr>
<td>lv</td>
<td>71.69/50.43</td>
<td>72.48/50.85</td>
<td>70.24/48.97</td>
<td>71.60/49.56</td>
</tr>
<tr>
<td>fi</td>
<td>64.64/46.21</td>
<td>64.63/46.22</td>
<td>63.07/44.82</td>
<td>64.74/46.09</td>
</tr>
<tr>
<td>et</td>
<td>66.63/45.58</td>
<td>65.78/45.01</td>
<td>64.94/44.04</td>
<td>65.06/44.33</td>
</tr>
<tr>
<td>zh*</td>
<td>41.05/23.85</td>
<td>40.11/23.02</td>
<td>39.49/22.68</td>
<td>39.89/22.49</td>
</tr>
<tr>
<td>ar</td>
<td>38.74/28.24</td>
<td>33.66/25.44</td>
<td>34.25/24.69</td>
<td>33.31/24.86</td>
</tr>
<tr>
<td>la</td>
<td>49.04/35.48</td>
<td>47.12/34.36</td>
<td>46.78/33.56</td>
<td>45.26/31.97</td>
</tr>
<tr>
<td>ko</td>
<td>34.62/15.14</td>
<td>33.91/14.16</td>
<td>32.70/13.77</td>
<td>32.95/13.14</td>
</tr>
<tr>
<td>hi</td>
<td>36.01/27.24</td>
<td>29.59/21.75</td>
<td>32.02/23.79</td>
<td>26.37/18.56</td>
</tr>
<tr>
<td>ja*</td>
<td>28.19/21.74</td>
<td>18.23/12.68</td>
<td>20.53/13.78</td>
<td>15.21/10.37</td>
</tr>
<tr>
<td>Average</td>
<td>64.57/54.14</td>
<td>63.25/52.94</td>
<td>62.88/52.44</td>
<td>62.88/52.16</td>
</tr>
</tbody>
</table>

Table 5: Results (average UAS%/LAS% over 5 runs, excluding punctuation) on the original training sets. (Languages are sorted by the word-ordering distance to English, ‘°’ refers to results of delexicalized models, ‘en°’ means that for English we use results on the test set since models are trained with the English training set.)

F Results on Google Universal Dependency Treebanks v2.0

We also ran our models on Google Universal Dependency Treebanks v2.0 (McDonald et al., 2013), which is an older dataset that was used by (Guo et al., 2015). The results show that our models perform better consistently.

<table>
<thead>
<tr>
<th>Language</th>
<th>SelfAtt-Graph</th>
<th>RNN-Graph</th>
<th>SelfAtt-Stack</th>
<th>RNN-Stack</th>
<th>(Guo et al., 2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>65.03/55.03</td>
<td>64.60/54.57</td>
<td>63.63/54.40</td>
<td>65.51/55.82</td>
<td>60.35/51.54</td>
</tr>
<tr>
<td>French</td>
<td>74.45/63.28</td>
<td>76.75/65.20</td>
<td>73.63/62.76</td>
<td>75.13/64.44</td>
<td>72.93/63.12</td>
</tr>
<tr>
<td>Spanish</td>
<td>72.00/61.50</td>
<td>73.99/63.46</td>
<td>71.73/61.42</td>
<td>74.13/64.00</td>
<td>71.90/62.28</td>
</tr>
</tbody>
</table>

Table 6: Comparisons (UAS%/LAS%) on Google Universal Dependency Treebanks v2.0.
G Results on specific dependency types for Czech

In table 7, we show results of Czech on some dependency types with evaluation breakdowns on dependency directions. We select Czech mainly for two reasons: (1) It has the largest dataset; (2) Czech is famous for relatively flexible word order. Generally, we can see that models that are more flexible on word ordering perform better. Interestingly, for objective and subjective types, we can see that LAS scores for all models are quite low even when the correct heads are predicted. The reason might be that even the relative-positional self-attention encoder can capture some positional information which further reveals word ordering information in some way.

### Table 7: Evaluation breakdowns (UAS%/LAS%) on dependency directions for Czech on some specific dependency types.

<table>
<thead>
<tr>
<th>Direction</th>
<th>Percentage</th>
<th>SelfAtt-Graph</th>
<th>RNN-Graph</th>
<th>SelfAtt-Stack</th>
<th>RNN-Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>mod-first</td>
<td>99.99%</td>
<td>75.34/75.34</td>
<td>74.62/74.61</td>
<td>74.46/74.43</td>
<td>74.17/74.08</td>
</tr>
<tr>
<td>head-first</td>
<td>0.01%</td>
<td>-</td>
<td>75.34/75.34</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>all</td>
<td>100.00%</td>
<td>75.33/75.33</td>
<td>74.61/74.61</td>
<td>74.45/74.43</td>
<td>74.17/74.07</td>
</tr>
</tbody>
</table>

(NOUN, NOUN, nmod): (mod-first% in English is 4.72%)

<table>
<thead>
<tr>
<th>Direction</th>
<th>Percentage</th>
<th>SelfAtt-Graph</th>
<th>RNN-Graph</th>
<th>SelfAtt-Stack</th>
<th>RNN-Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>mod-first</td>
<td>99.99%</td>
<td>88.93/88.92</td>
<td>89.42/89.41</td>
<td>85.39/85.21</td>
<td>87.26/86.37</td>
</tr>
<tr>
<td>head-first</td>
<td>0.01%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>all</td>
<td>100.00%</td>
<td>85.63/85.29</td>
<td>85.72/85.41</td>
<td>81.85/81.14</td>
<td>83.98/81.71</td>
</tr>
</tbody>
</table>

(NOUN, VERB, obl): (mod-first% in English is 9.62%)

<table>
<thead>
<tr>
<th>Direction</th>
<th>Percentage</th>
<th>SelfAtt-Graph</th>
<th>RNN-Graph</th>
<th>SelfAtt-Stack</th>
<th>RNN-Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>mod-first</td>
<td>37.80%</td>
<td>48.84/40.33</td>
<td>46.39/38.49</td>
<td>48.75/41.08</td>
<td>50.16/41.64</td>
</tr>
<tr>
<td>head-first</td>
<td>62.20%</td>
<td>62.81/55.97</td>
<td>60.38/53.41</td>
<td>62.22/55.37</td>
<td>61.73/55.32</td>
</tr>
<tr>
<td>all</td>
<td>100.00%</td>
<td>57.53/50.06</td>
<td>55.09/47.77</td>
<td>57.13/49.97</td>
<td>57.36/50.15</td>
</tr>
</tbody>
</table>

(NOUN, VERB, nsubj): (mod-first% in English is 85.07%)

<table>
<thead>
<tr>
<th>Direction</th>
<th>Percentage</th>
<th>SelfAtt-Graph</th>
<th>RNN-Graph</th>
<th>SelfAtt-Stack</th>
<th>RNN-Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>mod-first</td>
<td>20.65%</td>
<td>61.42/58.33</td>
<td>58.12/54.51</td>
<td>60.88/58.24</td>
<td>60.34/58.98</td>
</tr>
<tr>
<td>head-first</td>
<td>79.35%</td>
<td>73.18/65.24</td>
<td>71.30/62.28</td>
<td>72.12/63.81</td>
<td>72.76/64.65</td>
</tr>
<tr>
<td>all</td>
<td>100.00%</td>
<td>62.47/56.65</td>
<td>60.03/54.09</td>
<td>61.48/53.25</td>
<td>60.38/54.32</td>
</tr>
</tbody>
</table>

(ADV, VERB, advcl): (mod-first% in English is 31.02%)

<table>
<thead>
<tr>
<th>Direction</th>
<th>Percentage</th>
<th>SelfAtt-Graph</th>
<th>RNN-Graph</th>
<th>SelfAtt-Stack</th>
<th>RNN-Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>mod-first</td>
<td>41.75%</td>
<td>65.79/65.28</td>
<td>65.02/64.33</td>
<td>65.33/64.55</td>
<td>61.93/60.53</td>
</tr>
<tr>
<td>head-first</td>
<td>58.25%</td>
<td>51.53/49.06</td>
<td>51.54/48.52</td>
<td>51.54/48.52</td>
<td>56.07/48.52</td>
</tr>
<tr>
<td>all</td>
<td>100.00%</td>
<td>81.53/80.86</td>
<td>80.04/79.17</td>
<td>80.29/79.05</td>
<td>79.26/76.80</td>
</tr>
</tbody>
</table>

Table 7: Evaluation breakdowns (UAS%/LAS%) on dependency directions for Czech on some specific dependency types. “mod-first” means the dependency edges whose modifier is before head, “head-first” means the opposite, and “all” indicates both “mod-first” and “head-first”. “–” replaces results that are unstable because of rare appearance (below 1%).
References


