**INTRODUCTION**

**Goal:** Automatic conversion of PerDT (a none-UD Persian Dependency treebank with 29K sentences) to its UD version.

**OUR APPROACH**

Taking following steps:
- Token unification
- POS mapping
- Syntactical changes to main corpus
- Dependency mapping

**TOKEN UNIFICATION**

Separating multiword infections or attached clitics:

**POS MAPPING**

- NER tagging for Proper Nouns
- Correcting some POS tags & mapping

**SYNKTACTICAL CHANGES**

- Reversing order of verbal conjunctions
- Systematic & manual corrections to PerDT:
  - Lemma: 0.7% correction
  - Dependency head: 5.6% correction
  - Dependency label: 3.8% correction

**DEPENDENCY MAPPING**

Applying rules with specific order:

**EXPERIMENTS**

**Supervised parser**

Training UDpipe V.2 with fastText embeddings on both our data & the only previous Persian UD corpus (Seraji)

**Results:**

- Trained & tested on our corpus: **85.2** (LAS)
- Trained & tested on Seraji: **79.4** (LAS)
- Trained on ours/tested on Seraji: **61** (LAS)
- Trained on Seraji/tested on ours: **62.6** (LAS)

**Delexicalized model transfer**

Using delexicalized parser transfer to test consistency of our corpus & Seraji with UD project:
- Sampling the same number of tokens as Seraji from our corpus.
- Delexicalizing both of them
- Learning Yara Parser with 15 epochs
- Evaluating on delexicaled test set of the Universal English Web Treebank

**Results:**

<table>
<thead>
<tr>
<th>Training Data</th>
<th>UAS</th>
<th>LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uppsala Universal Treebank (Seraji)</td>
<td>45.37</td>
<td>36.42</td>
</tr>
<tr>
<td>Our corpus</td>
<td>47.31</td>
<td>38.59</td>
</tr>
</tbody>
</table>