You find a POS-tagged corpus of text in an unknown language. Can you parse this?

- **VERB DET NOUN ADJ DET NOUN**

Let’s extract interesting features of the whole corpus ("surface cues to structure").
- Our universal parser sees these corpus features, along with the input sentence.
- The universal parser is trained end-to-end on diverse languages, with supervision from treebanks.
- Supervision can include treebanks for thousands of additional synthetic training languages. This helps.
- Our best method improved UAS and LAS on held-out test languages by an average of 5.6 percentage points over past work.

## Surface Cues to Structure

<table>
<thead>
<tr>
<th>Feature</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>POS</td>
<td>NOUN</td>
</tr>
<tr>
<td>Adj</td>
<td>DET</td>
</tr>
<tr>
<td>Det</td>
<td>NOUN</td>
</tr>
<tr>
<td>Verb</td>
<td>DET</td>
</tr>
</tbody>
</table>

## Neural features

- Hand engineered features
  - How often do NOUNs tend to appear shortly before or after VERBs?
  - How often do ADJs tend to appear shortly before or after NOUNs?
  - How often do ADPs tend to appear shortly before or after NOUNs?

## The Typology Component

- More than 50,000 synthetic languages
- Resemble real languages, but not found on Earth
- Each has a corpus of dependency parses
- In the Universal Dependencies format
- Vertices are words labeled with POS tags
- Edges are labeled syntactic relationships
- Provide train/dev/test splits, alignments, tools

## Galactic Treebanks (Wang & Eisner 2016)

- How often do NOUNs appear shortly before or after VERBs?
- How often do ADJs appear shortly before or after NOUNs?
- How often do ADPs appear shortly before or after NOUNs?

## Results (each bar stretches from labeled to unlabeled score)

- **Attachment Scores**
  - [Graph showing attachment scores]
- **Precision, Recall, F1**
  - [Graph showing precision, recall, F1 scores]