A Universal Dependencies Treebank for Gujarati

Mayank Jobanputra*, Maitrey Mehta*, Çağrı Çöltekin

Introduction

- **GujTB**: The first UD treebank in the Indo-Aryan language of Gujarati spoken by over 56M speakers
- 187 sentences with 100 sentences doubly-annotated
- **Diverse Genres** spanning Samanantar (news), UD Cairo (short), and Gujarati translations of the French novella “Le Petit Prince” (fiction), and grammar book examples (grammar)

<table>
<thead>
<tr>
<th>Genre</th>
<th>Sentences</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>news</td>
<td>93</td>
<td>1159</td>
</tr>
<tr>
<td>short</td>
<td>20</td>
<td>178</td>
</tr>
<tr>
<td>fiction</td>
<td>40</td>
<td>331</td>
</tr>
<tr>
<td>grammar</td>
<td>31</td>
<td>217</td>
</tr>
</tbody>
</table>

Data statistics in GujTB

- We discuss some interesting cases encountered during the annotation process

Non-projectivity

- Gujarati allows non-projectivity
- Partial free word order give rise to over-lapping dependency edges

Absence of clausal subjects

- A co-referential pronominal ‘te’ always present
- A dative nominal construction with experiencer semantics assigned as a ‘nsubj’ by prior work

Challenging Construction

Multiple Possible Annotations
1. Assigning ‘det:predet’ to ‘e’ and ‘je’ with ‘pustaka’ as their head
2. A change in order between ‘je’ and ‘pustaka’, where ‘je’ would act as a subordinating conjunction

The Case for Determiners

- According to Gujarati grammars, demonstrative pronouns like એ, તે, પેલું behave differently when attached to a nominal vs. when used independently
- Gujarati grammars do not discuss determiners as such; yet, the use of nominal-attached pronouns is closer to the UD definition of determiners
- Independently occurring cases treated as pronouns

Head-final conjunctions

- Gujarati carries case inflections and post-positional attachments on the final conjunct
- ‘jîtyâ’ has a plural inflection requiring it to agree with the head of the coordinated structure

Links and Contact

Email: mayank@lst.uni-saarland.de
maitrey@cs.utah.edu
ccoltekin@sfs.uni-tuebingen.de