LING242 2024 projects

difficulties: technical (***)/theoretical (***)/ fun (***), *** is the highest

1. */*/* constituency parsing with and without null elements, including tree edit – removing null elements (English – given, Chinese, Korean): implementation + readme

2. ∅/***/** generalization of tree structures (many languages): almost no implementation + report

3. */*/* treebank tree binarization (English, Chinese, Korean): implementation + readme

4. */**/* revise penn treebanks by correcting punctuation mark nodes (English, Chinese, Korean): implementation + readme

5. */∅/* GEC dataset for seq2seq format: implementation + readme

6. */***/*** Linearization dependency parsing (any languages in UD): implementation + readme

7. */**/*** UniMorph+UD (any languages in UD): implementation + report

8. */**/*** UD+MWE typology (any languages in UD) + visualization: implementation + report

9. ***/**/∅ phrase2dep (English – given, Chinese, Korean): implementation + readme

10. ***/***/∅ dep2phrase (English, Chinese, Korean): implementation + readme