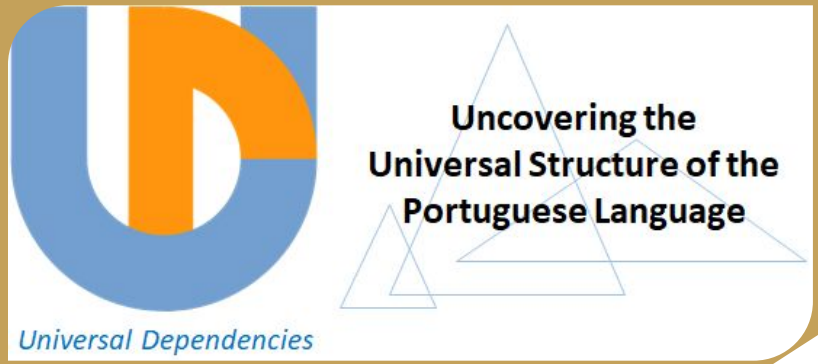
Shallow Parsing of Portuguese Annotated Texts under Universal Dependencies

Guilherme Martiniano de Oliveira, Paulo Berlanga Neto and Evandro Eduardo Seron Ruiz

Conference

Workshop



Introduction - Shallow Parsing (Parcial Parsing)

- Chunking: splitting a sentence into smaller parts of one or more tokens;
- **NP (Noun Phrase)**, VP (Verbal Phrase) and others applications;
- Commonly used: Named Entity Recognition, Sentiment Analysis and Information Retrieval;

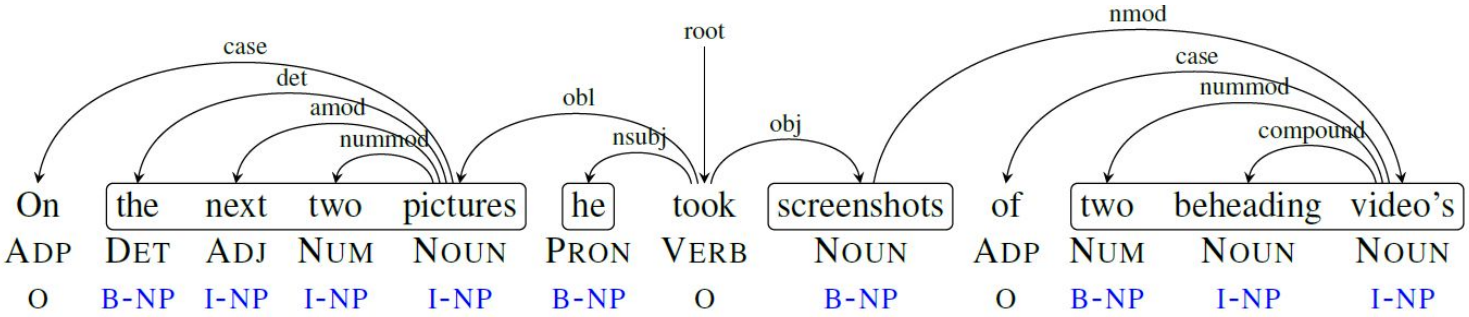


Image source: Investigating NP-Chunking with Universal Dependencies for English (Lacroix, 2018)

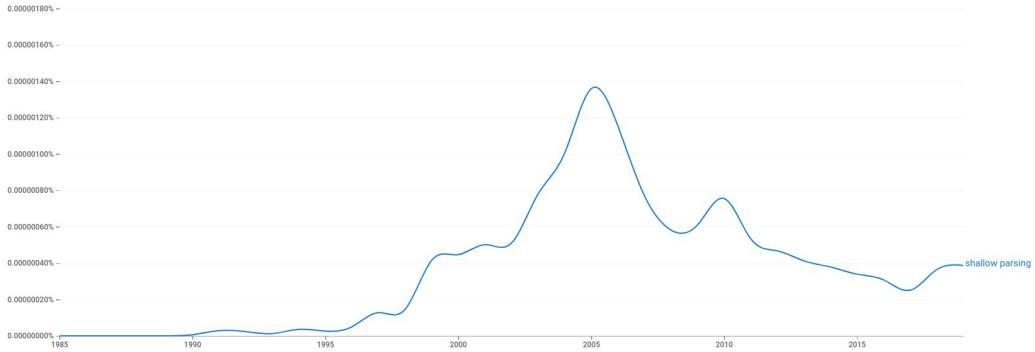
Problem

- Outdated studies on the field;
- New standard annotation guidelines, Universal Dependencies (UD);

Result

- Method to tag UD annotated texts;
- Metrics:

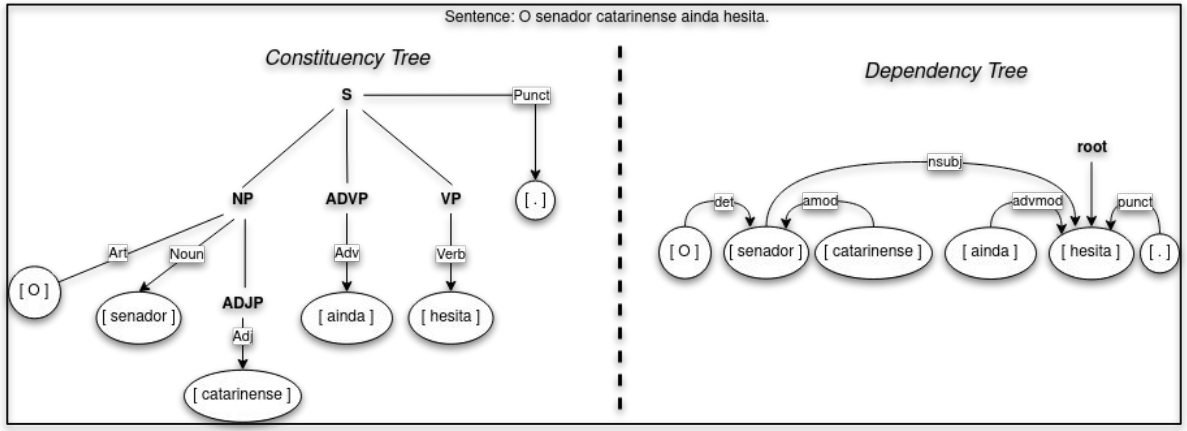
Precision	Recall	F-measure
84.8	85.3	85.1



Graph source: Google Books Ngram Viewer http://books.google.com/ngrams

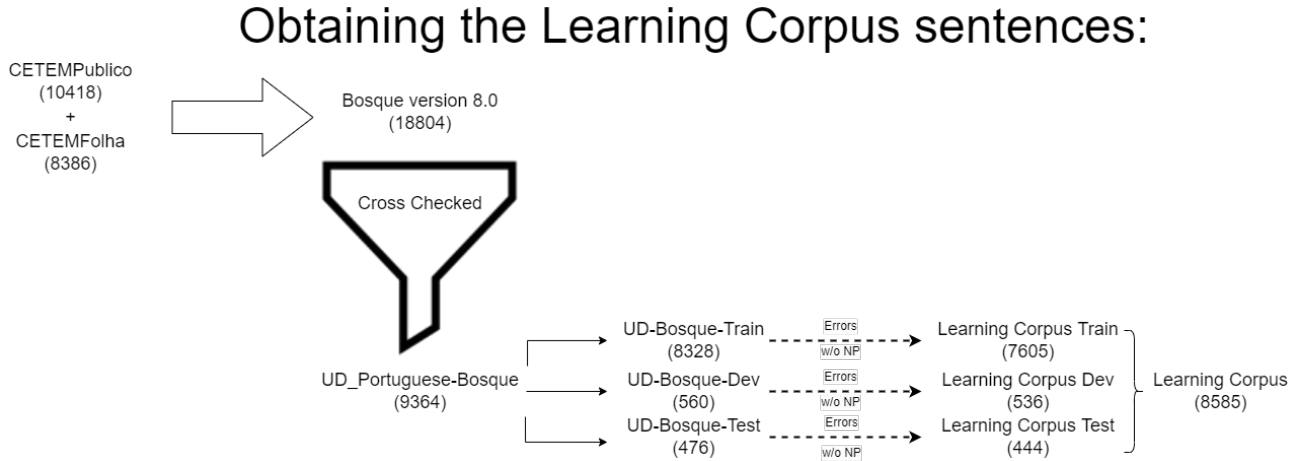
Objectives

- Multi-genre text shallow parser;
- Constituency Tree → Dependency Tree;
- Baseline algorithm;



Methodology - Data

- Corpora Bosque (version 8.0): tagged noun phrases chunks → validation;
- UD_Portuguese-Bosque: UD tagged tokens (without chunk tags) → input;
- Divided into: *Learning Corpora* (88.6%) and *Test Corpora* (11.4%);
- Cross checked to match existence;



Methodology - Classification

- Deep Bi-Long-Short Term Memory (Bi-LSTM) Recurrent Neural Network (RNN) → Multi-task Learning (MTL) model;

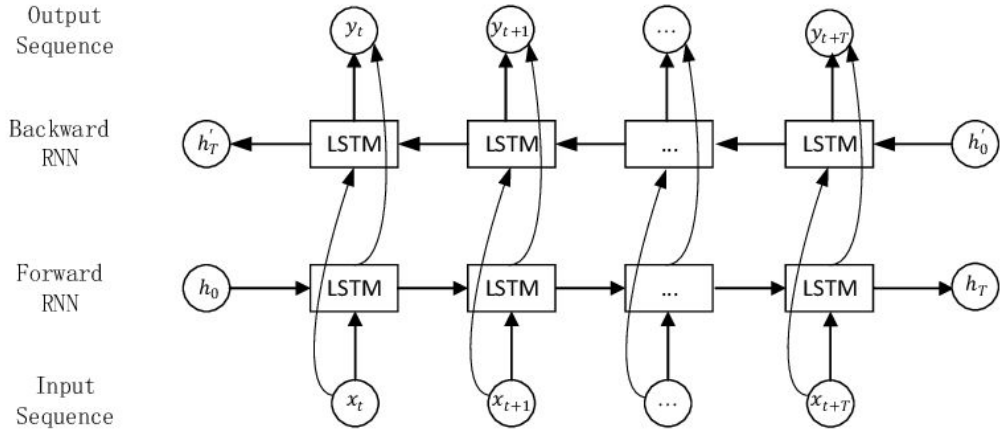


Image source: Multi-time scale wind speed prediction based on WT-bi-LSTM - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Bi-LSTM-neural-network-structure-deployed-in-time-direction_fig2_339679582 [accessed 19 Mar, 2022]

Conclusion & Final Remarks

- Method proposed has potential;
- Capability to improve further;

Precision	Recall	F-measure
84.8	85.3	85.1

- Preliminary results, underachieved in comparison to Ophélie's results;
- Corpora credibility, algorithm's parameters augmentation and sentence analysis (longer/shorter);

1. . . . o governador do Rio e o **Presidente**[PROPN] da **República**[PROPN] chamaram o Exército.
2. . . . o **presidente**[NOUN] da **República**[NOUN] abriu uma fresta . . .
3. No caso de impedimento de o **presidente**[NOUN] da **República**[PROPN] . . .

Acknowledgement

- This work was carried out at the Center for Artificial Intelligence (C4AI-USP), with support by the São Paulo Research Foundation (FAPESP grant#2019/07665-4) and by the IBM Corporation.

- **E-mail: {guizera11, pauloberlanga, evandro}@usp.br**

Center for Artificial Intelligence