

Transitions all the Way Down: From Characters to Full Document Annotation in One System

Working Groups 1 & 3

Yuval Pinter



Miryam de Lhoneux

The logo for KU Leuven, consisting of a dark blue rectangular box with the text "KU LEUVEN" in white, bold, sans-serif capital letters.

- We propose a **transition-based** parsing scheme that annotates a document from the **character** level, through **morphology** and **syntax** to **discourse**, all in **one system**
- Our system fits a variety of phenomena in morphology and syntax, such as **null syntactic words** and **multi-word tokens**, tokenizing a document as it parses it
- We leverage multilingual resources: **Universal Dependencies** and **MorphyNet**
- Our decision and representation components are all **standard DL nets** implemented in PyTorch