Redefining Syntactic and Morphological Tasks for Typologically Diverse Languages

Omer Goldman, Leonie Weissweiler and Reut Tsarfaty

Motivation

NLP for polysynthetic languages!

Let’s try Yupik:
Mangteghangllaghyugtukut
English: We want to make a house

Word Segmentation

Inconsistent
  • across languages
  • across treebanks

Hebrew: האבתי
  • HTB: האבתי - הא
  • IAHLT: האבתי

Our Solution

Content words/morphemes → “syntactic” nodes
Function words/morphemes → “morphological” features

More parallel trees

go
you
will
not
student
because
you
were
my

ID | Form | Lemma | POS | FEATS | HEAD | DEP | MS-FEATS
--- | --- | --- | --- | --- | --- | --- | ---
1  | you  | you   | PRON| Nom;2;Sg | nsubj | Nom;2;Sg |
2  | will  | will  | AUX | Fin    | aux  |       |       |
3  | not   | not   | PART | Neg    | advmod|       |       |
4  | go    | go    | VERB | Inf    | root | Fin;Ind;Fut;Neg|
5  | because | because | SCONJ | - | mark |       |       |
6  | you   | you   | PRON| Nom;2;Sg | nsubj | Nom;2;Sg |
7  | were  | be    | AUX | Fin;Ind;Past;2;Sg | cop |       |       |
8  | my    | my    | PRON| Gen;1;Sg | nmod:poss | Gen;1;Sg |
9  | student | student | NOUN | Sg | advcl:because | Sg;Ind;Past |