Take care of your **morphs** with **mSUD** annotation format!

### Joint Annotation of Morphology and Syntax in Dependency Treebanks

Bruno Guillaume, Kim Gerdes, Kirian Guiller, Sylvain Kahane, Yixuan Li

Word level annotation may be difficult to apply
- Agglutinative languages (Turkish)
- Polysynthetic languages (Yupik)
- Languages written without spaces (Chinese)
- Languages with an oral tradition (Beja)

We propose **mSUD**
- Morph-level annotation framework
- Convertible to word-based format
- Easier inclusion of IGT-based source data

In **mSUD**
- Two types of dependency
  - regular (e.g. subj)
  - at the morphological (e.g. subj/m)
- Features **TokenType** (values: DerAff, InflAff, Root)
- Features for word-level **upos**: DerPos for derivational affixes
  - CpdPos for compounds

Three categories of **subword annotations**
- Derivation
- Composition
- Inflection

**Implementation**

- mSUD
- SUD
- mUD
- UD

In release 2.14, three treebanks are in **mSUD**
- mSUD_Beja-NSC
- mSUD_Chinese-Beginner
- mSUD_Chinese-PatentChar

Other treebanks are built in mSUD (IGT based)
- Gbay. Ye’kwana. Tuwari

---

**Derivation**

- Derivational affix as the head
  - it controls the distribution of the combination

**Composition**

- conj/m: Two roots from the same syntactic and semantic class
- nod/m: Modifier-head relation between two roots
- comp/m: predicate-complement relations

**Inflection**

- Inflectional affixes which control the distribution of the word
  - govern the root (e.g. TAM, affixes and case markers)

**Examples**

**Turkish inflectional groups** (Çöltekin, 2016)
- Mavi:
  - m: upernoun
  - mod: upernoun
- arabadıkiler: upernoun
- uyuyorlar: upernoun

**Yupik Polysynthetic example** (Park et al., 2021)
- Mavgi arabadıkilen uyuyorlar
  - Blue car,LOC-DAT.PL sleep.NGRCV.1P
  - ‘The ones in the blue car are sleeping.’

- Mangteghaghilállangilláayuq refill
  - house-big-to make-to want.to /IND.INTR-1PL
  - ‘We want to make a big house.’