**GOALS**

1. Improve parsing quality by pre-processing historical input data (PIPELINE)

2. Trace syntactic change in EN and DE scientific discourse (APPLICATION)

**PIPELINE**

1. Normalization of Historical Data
   - Replacing the formerly common virgule (slash) by the analogous comma: „Wann jemand etwas seinem Nächsten zum Besten aufrecht heraus gibt / so gering es auch ist / billig zu Dank soll angenommen werden.” → „Wann jemand etwas seinem Nächsten zum Besten aufrecht heraus gibt, so gering es auch ist, billig zu Dank soll angenommen werden.

2. Extraction of “good sentences” (GS)
   - Detection of non-sentential constructions (“bad sentences” – BS)
     - Sentences beginning in lower case and the preceding sentence (incomplete)
     - Sentences with less than 8 tokens (too short)
     - Sentences lacking a verb (verbless)
     - Foreign-language sentences (foreign)

3. UD-parsing
   - Parser: UDPipe 1.0, UD Models (2.5): GUM (EN), GSD (DE)
     - Input tokens: normalized word forms
     - Preservation of sentence splitting and tokenization

4. Evaluation
   - Sample 100 GS and BS (20 / 50 years period, e.g., 1650–1699)

   4.1 Parsability of a sentence: grammatically interpretable structure
     - Accepted structures: title-like noun phrases and dates
     - Excluded structures: sentences in other language, fragments without grammatical structure, i.e., equations, abbreviations

   4.2 Roots: number and accuracy of sentence roots

   4.3 Parsing accuracy: UD label and syntactic head
     - Overall parsing accuracy ~ 80% for both corpora
     - Significant improvement (0.23%) of accuracy for GS vs. BS
     - Stable accuracy of GS over time

**APPLICATION**

Case study and sanity check: Noun phrase development

1. Does the data reflect previous observations? – Yes!
   - MORE NP internal modification with phrasal features – in the UD-framework: nmod, appos, nummod, amod, det, compound, flat.
   - LESS clausal postmodification, i.e., finite and non-finite clausal modifiers (acl/acl:relcl)

2. Is NP densification a cross-linguistic development? – Yes!
   - Significant decrease of clausal features in both EN and DE.

3. If so, is change driven by the same NP modifiers? – Almost!
   - EN and DE develop towards similar proportions of amod and nummod.

**REFERENCES**


LREC 2022