

LASSY

Large Scale Syntactic Annotation of Written Dutch



Who:

- Groningen: *Gertjan van Noord, Gosse Bouma*
- Leuven: *Ineke Schuurman, Frank van Eynde*

Who:

- Groningen: *Gertjan van Noord, Gosse Bouma*
- Leuven: *Ineke Schuurman, Frank van Eynde*
- User group with partners from industry and academia

What:

- 1,000,000 word treebank (manually corrected)
 - ★ POS-tags
 - ★ lemma's
 - ★ CGN/D-Coi dependency structures
- 500,000,000 word treebank (automatically constructed)
- Tools for corpus exploitation
 - ★ compressed XML / XML database technology
 - ★ XPATH, Xquery
 - ★ extensions for matching and capturing
- Case Studies
 - ★ Lexical Acquisition
 - ★ Corpus Linguistics
 - ★ Knowledge Acquisition