

## MWE Finder

- Enables search for flexible MWEs
- Based on an example in canonical form
- Embedded in GrETEL
- Application to support linguistic and lexicographic research into MWEs

## Multiword Expression (MWE)

- word combination with linguistic properties that cannot be predicted from the properties of the individual words or the way they have been combined by the rules of grammar

*de boeken neerleggen*

*the books down.put `stop business`*

*Te-r plaats-e*

*On-DAT location-DAT `on location`*

## Flexible MWEs

- components : multiple forms, need not be adjacent, multiple orders
- *Saab heeft gisteren **de boeken neergelegd**.*
- *Ik dacht dat Saab gisteren **de boeken wilde neerleggen**.*
- *Saab **legde de boeken neer**.*
- *Saab **legde gisteren de boeken neer**.*

## Canonical Form

- Defined for (Dutch) MWEs
- 11k MWEs in canonical form with various annotations
- Examples:
  - *iemand zal de boeken neerleggen*
  - *iemand zal de dans ontspringen*
  - *iemand zal Oeen \*+poging doen*

## Queries (Automatically generated)

- MWE query: the MWE
- Near-Miss query: the MWE with unexpected modifiers / determiners
- Major Lemma query: major lemmas

## Search Results

- *iemand zal de dans ontspringen*
- Treebank Mediargus 103M sentences
- MWE-Query: 1158 hits
- Near-Miss: 1271 hits
- Major Lemma: 1309 hits

## Analysis (off-line Analysis Component)

- Near-Miss minus MWE:
  - Different determiners of *dans*
  - Modification of *dans*
- Major lemma minus Near Miss:
  - Wrong parses
  - *Aan de dans ontspringen*

## Conclusions

- MWE Finder enables searching for flexible MWEs
- Enables efficient determination of MWE properties based on corpus data
- → improved canonical form:
  - *iemand zal Ode \*dans ontspringen*