

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

- Single form (cf. Lemma)
- DUCAME >11k canonical forms (Dutch)
- 'headword' for entry with properties
- Some properties as annotations
- MWE-Finder
 - Input canonical form
 - Applies principles to automatically derive queries

Guiding Principles

- Well-formed expression
- Head: lemma if inflectable
- Free arguments: indef pronouns
- Bound arguments: 3p pronouns
- Head: modifiable, inflectable (if lemma)
- Non-head: not inflectable or modifiable
- Deviations: annotations
- Only MWE properties

Generalisation principles

- Properties
 - Pos. Rel, lemma
 - Subcat (S)
 - Interpretable (I)
 - Purely Grammatical (G)
- Pos, rel, lemma, subcat: fixed
- Freely Inflectable: I and G can vary
- Other: G can vary

Annotations

notation	interpretation
*word	word is modifiable/determinable
+word	word is inflectable
=word	word must occur as given
!word	word is not modifiable / determinable
0word	word is not part the MWE
<text>	freely replaceable argument
dd:[word]	word must be a definite determiner

Examples

Iemand zal de dans ontspringen

Iemand zal iemands hart breken

Iemand zal zijn geduld verliezen

*Iemand zal Oeen *+poging doen*

De schellen zullen iemand van de ogen vallen

Summary

- Canonical form for MWEs
- Largely language-independent
- >11k lexicon (DUCAME) for Dutch
- Actually used in MWE-Finder

Future Work

- Support verb constructions
- Negative polarity