Recognizing MWEs
- MWE same synt str. as literal
- MWE in a sentence
  - replace it in semantics or deep syntax
    - by a single node (opaque idioms)
    - By multiple nodes ((semi-transparent idioms)
  - Replacement: not here
- MWE identification: XPATH Query

Alpino
- Parser for Dutch
- Deals with some fixed MWEs
- Deals with some flexible MWEs that require special syntax
- [http://www.let.rug.nl/vannoord/alp/Alpino](http://www.let.rug.nl/vannoord/alp/Alpino)

PaQu
- Search interface to (Dutch) treebanks
- Enables you to
  - upload your own corpus
  - have it parsed by Alpino
  - and then to search in it
- [http://portal.clarin.nl/node/4182](http://portal.clarin.nl/node/4182)

Flexible MWEs
- MWE:
  - word combination with idiosyncratic linguistic properties
- Flexible:
  - No fixed word order, not necessarily contiguous, inflectional variation (in principle) possible on multiple words
- Examples
  - *Hij heeft gisteren de plaat gepoest*
  - *Hij poetste gisteren de plaat*
  - *...dat hij de plaat wilde poetsen*

Mini-Treebank
- 35 example sentences relevant for MWEs (mostly V+NP idioms)
- Go to PaQu
  - Login with your e-mail address
  - Click on the link in the mail sent to you
  - Select Corpus: PARSEME MWE Test
    - [List all sentences: //node[@cat="top"]](http://portal.clarin.nl/node/4182)
    - Download fully parsed corpus

Morpho-syntactic phenomena:
- Verb: Vfinal, V2, V1 (YNQ), VR
- NP: Topicalisation, Wh-questioning, scrambling, independent occurrence
- N: (internal) modification, relativisation, diminutives, plural, other dets
- V+NP: passive

Extensions/adaptations needed:
- Overacceptance (extension needed)
  - Modification of *plaat*
  - Scrambling of *de plaat*, Topicalisation of *de plaat*
  - Passivisation (idiosyncratic MWE property)
- Underacceptance
  - Relativisation of *bok, flater* (adaptation)
  - MWEs with cranberry words (extension)

Caveats
- Internal Alpino structure / analysis may be different
- Rules in the Alpino system may differ from XPATH queries (esp. wrt matching properties)
- Maybe sufficient for analysis (parsing), but not for production.