

Converting a standard representation of MWEs to the Alpino lexicon

Nicole Grégoire

Uil-OTS
University of Utrecht

CLIN 17



Please read the following text

Steeds meer koperdieven slaan hun slag langs het spoor. Dat blijkt uit de tellingen die sinds begin 2006 worden gehouden, meldt de Spoorwegpolitie.

De diefstallen leiden ook steeds vaker tot vertragingen op het spoor. De eerste negen maanden gebeurde dat al 91 keer. Ook vandaag was het weer raak: op twee locaties was gestolen koper aanleiding voor treinenuitval.

Het rood-gele metaal is bij criminelen sterk in trek geraakt als gevolg van hoge koperprijzen. Spoorbeheerder ProRail denkt na over maatregelen ter voorkoming van de diefstal.

www.rtl.nl



8 sentences, 9 multiword expressions

Steeds meer koperdieven **slaan hun slag** langs het spoor. Dat **blijkt uit** de tellingen die sinds begin 2006 worden gehouden, meldt de Spoorwegpolitie.

De diefstallen **leiden** ook steeds vaker **tot** vertragingen op het spoor. De eerste negen maanden gebeurde dat al 91 keer. Ook vandaag **was het weer raak**: op twee locaties was gestolen koper **aanleiding voor** treinenuitval.

Het rood-gele metaal is bij criminelen sterk **in trek** geraakt **als gevolg van** hoge koperprijzen. Spoorbeheerder ProRail **denkt na over** maatregelen **ter voorkoming van** de diefstal.

www.rtl.nl



- 1 Multiword Expressions
- 2 The MWE lexicon
- 3 A mapping to Alpino



Definition

In this research Multiword Expressions (MWEs) are defined as a combination of words that has linguistic properties not predictable from the individual components or the normal way they are combined.



Linguistic properties

- Lexical: *de plaat poetsen, zware/*sterke shag*
- Orthographic: *Yahoo!*
- Phonological: *over de rooie/*rode (gaan/zijn/raken)*
- Morphological: *ten gevolge van*
- Syntactic: *in opdracht van, bijvoeglijk(*e) naamwoord*
- Semantic: *de plaat poetsen, een bok schieten*
- Pragmatic: *dames en heren*



Creating the lexicon

What ingredients must be part of an MWE description if

- 1 we want to cover as many different combinations as possible, and
- 2 the design must be implementable in any NLP system.



Creating the lexicon

Covering as many combinations as possible means making a distinction between:

- *de plaat poetsen* and *een blunder maken/begaan/*doen*



Ingredients of an MWE description

MWE Lexicon for Dutch

[Lexicon](#) | [Documentation](#) | [About IRME](#) | [Sitemap](#)

[Advanced search](#) | **New MWE** | [New from File](#) | [New pattern](#) | [View patterns](#) | [View MWEs](#)

EXPRESSION	<input type="text" value="de plaat poetsen"/>	
PATTERN	<input type="text" value="ec1 - de plaat poetsen"/>	<input type="button" value="Show details"/>
POS	<input type="text"/>	<input type="button" value="Narrow potential patterns"/> <input type="button" value="Show potential patterns"/>
CL	<input type="text" value="de plaat poetsen"/>	
	<input type="button" value="Parameters"/>	
VL	<input type="text" value="-"/>	
XL	<input type="text" value="-"/>	
SUBJECT	<input type="text" value="-"/>	
EXAMPLE	<input type="text" value="hij heeft de plaat gepoetst"/>	
COMMENTS	<input type="text" value=" [.VP [.obj1:NP [.det:D (1)] [.hd:N (2)]] [.hd:V (3)]]"/>	
VP MWE	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="button" value="Add Entry"/>	



Een blunder maken/begaan

- Expression: blunder
- Pattern: EC3 – [.VP [.obj1:NP [.hd:N (1)]] [.hd:V (list)]]
- CL: blunder
- VL: maken begaan
- Example: hij heeft een blunder gemaakt



MWEs in Alpino

fixed and **part_fixed** patterns in the verbs lexicon (verbs.pl)

- steken fixed([[de,draak],pc(met)],imp_passive)
- brengen
part_fixed(uit,[[acc(verslag),pc(over)]],imp_passive)



Pattern

Standard representation \Rightarrow Alpino representation

[.VP [.obj1:NP [.det:D (1)] [.hd:N (2)]] [.hd:V (3)]]



- (3) fixed([[(1), (2)]], norm_passive)
- (verb_3) part_fixed((prt_3), [[(1), (2)]], norm_passive)



De plaat poetsen

- Source Pattern: [.VP [.obj1:NP [.det:D (1)] [.hd:N (2)] [.hd:V (3)]]
- Target pattern: (3) fixed([[(1), (2)]], norm_passive)
- mapping: 3 4 5
- example sentence: *hij heeft de plaat gepoetst*
- CL: de plaat poetsen

⇒ poetsen fixed([[de,plaat]], norm_passive)



Blunder maken/begaan

- Source Pattern: [.VP [.obj1:NP [.hd:N (1)]] [.hd:V (list)]]
- Target pattern: (2) fixed([acc((1))],norm_passive)
- mapping: 4 5
- example sentence: *hij heeft een blunder gemaakt*
- VL: maken begaan

⇒ maken fixed([acc(blunder)],norm_passive)

⇒ begaan fixed([acc(blunder)],norm_passive)



End of project: 31st of May 2007

- Filling the lexicon up to 5,000 expressions.
- Validation of the lexicon (by CST Copenhagen).
- Converting the standard to Alpino.
- Converting the standard to Rosetta?



Thank You

IRME website

<http://www-uilots.let.uu.nl/irme>

Financed by NTU STEVIN

<http://www.taaluniversum.org/stevin>

