En Garde Project
The redesign of a Dutch Electronic Lexicon of Multiword Expressions

Nicole Grégoire

CLARIN-NL/UiL OTS, Utrecht

August 4, 2010
Convert DUELME into LMF format to create a curated DUELME resource fully compliant with the standards supported by CLARIN.

Create a mapping between DUELME-specific data categories and (possibly newly created) ISOCAT data categories to ensure semantic interoperability of the curated resource with other resources and tools.

March 1, 2010 - August 31, 2010
1. DUELME
2. DUELME → DUELME LMF
Dutch Electronic Lexicon of Multiword Expressions

- Identification and Representation of Multiword Expressions (IRME) project (2005-2007)
- Creating a lexicon that is highly theory- and implementation-independent, and which can be used in various NLP systems.
- Total number of 5,232 Multiword Expressions (MWEs).
Multiword Expressions

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.

Examples: *miss the boat, throw someone to the lions*
Design

- Use in various NLP-systems
- Equivalence Class Method (Odijk, 2003)
- MySQL database with two tables: mwes and patterns.
- DUELME-GUI
MWEs that have the same structure require the same treatment in NLP.

- EC1: miss the boat, kick the bucket
- EC2: spill the beans
Parameterized equivalence classes

- EC1: miss the boat, kick the bucket, spill the beans
- Representation: kick the bucket[sg], spill the bean[pl]
Dutch Electronic Lexicon of Multiword Expressions

Lexicon

DuELME (Dutch Electronic Lexicon of Multiword Expressions) is one of the results of the project Identification and Representation of Multiword Expressions (IRME) and contains lexical descriptions of over 5,000 multiword expressions (MWEs). The lexical descriptions boast to be highly theory- and implementation-neutral. The lexicon is primarily intended for the use in various NLP systems.

Project

The IRME project has been carried out within the STEVIN programme which is funded by the Dutch and Flemish Governments http://taalunieversum.org/stevin.

GUI

The GUI can be used to create new entries, search on predefined queries and adapt existing entries. See the documentation section for more information.
### MWE pattern description

**Pattern Name**: ec142

**POS**: npx p d n v

**Pattern**:

```
[ VP [ .obj1:NP (var) ] [ .hd:PP [ .hd:P (1) ] [ .ppcomp:NP [ .det:D (2) ] [ .hd:N (3) ] ] ] [ .hd:V ]
```

**Mapping**: 4 5 6 7

**MWE**: onder de duim houden

**Example**: hij heeft iemand onder de duim gehouden

**Description**:

Expressions headed by a verb, taking (1) a variable direct object, and (2) a locative/directional complement headed by a fixed preposition taking a complement consisting of a fixed determiner and an unmodifiable noun.

**Diagram**:

```
  VP
   /\  
  /   \ 
obj1:NP  ld:PP  hd:V
   |     |     |
  (var) (1) (2) (4)
```

```
  hd:P
   |  
  (1)  
  det:D  
  (2)
```

```
  hd:N
   |  
  (3)
```
### MWE description

| **EXPRESSION** | throw to the lions |
| **CL** | throw to the lion[pl][count] |
| **LISTA** | n.a. |
| **LISTB** | n.a. |
| **POS** | ec74 - npx p d n v - onder de duim houden - hij heeft iemand onder de duim gehouden |
| **PATTERN1** | he threw her to the lions |
| **EXAMPLE1** | - n.a. - |
| **EXAMPLE2** | n.a. |
| **EXAMPLE3** | n.a. |
| **MODIFIER** | - |
| **SUBJECT** | - |
| **OBJECT** | [human] |
| **RPRON** | n.a. |
| **CONJUGATION** | hebben |
| **POLARITY** | - |
| **COMMENTS** | |
DUELME → DUELME LMF
Requirements:

1. Maintain the necessary ingredients to describe MWEs and MWE patterns, and
2. maintain the ECM.
3. Where possible make use of existing LMF extensions.
ISO 24613

- NLP multiword expression patterns extension (Annex M)
- Morphology extension (Annex A)
- NLP syntax extension (Annex E)
- NLP semantics extension (Annex G)
NLP multiword expression patterns extension
MWE Pattern
List of Components

- **Lexical Entry**
  - 0..1 to **List Of Components**
  - 1 to **Lemma**
- **List Of Components**
  - 0..* to **Component**
- **Lemma**
  - 1 to **Lexical Entry**
- **Component**
  - 2..* {ordered} to **List Of Components**
Variables (DUELME LMF)
Lexical Entry and other classes
Expressions headed by a verb, taking (1) a variable direct object, and (2) a locative/directional complement headed by a fixed preposition taking a complement consisting of a fixed determinant and an unmodifiable noun.
### DUELME-LMF example (MWE description)

<table>
<thead>
<tr>
<th>EXPRESSION</th>
<th>throw to the lions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>throw to the lion[pl][count]</td>
</tr>
<tr>
<td>LISTA</td>
<td>n.a.</td>
</tr>
<tr>
<td>LISTB</td>
<td>n.a.</td>
</tr>
<tr>
<td>POS</td>
<td>Narrow potential patterns</td>
</tr>
<tr>
<td>PATTERN1</td>
<td>ec74 - npx p d n v - onder de duim houden - hij heeft iemand onder de duim gehouden</td>
</tr>
<tr>
<td>EXAMPLE1</td>
<td>he threw her to the lions</td>
</tr>
<tr>
<td>PATTERN2</td>
<td>n.a.</td>
</tr>
<tr>
<td>EXAMPLE2</td>
<td>n.a.</td>
</tr>
<tr>
<td>PATTERN3</td>
<td>n.a.</td>
</tr>
<tr>
<td>EXAMPLE3</td>
<td>n.a.</td>
</tr>
<tr>
<td>MODIFIER</td>
<td>-</td>
</tr>
<tr>
<td>SUBJECT</td>
<td>-</td>
</tr>
<tr>
<td>OBJECT</td>
<td>[human]</td>
</tr>
<tr>
<td>RPRON</td>
<td>n.a.</td>
</tr>
<tr>
<td>CONJUGATION</td>
<td>hebben</td>
</tr>
<tr>
<td>POLARITY</td>
<td>-</td>
</tr>
<tr>
<td>COMMENTS</td>
<td></td>
</tr>
</tbody>
</table>

Nicole Grégoire
En Garde
DUELME-LMF example (MWE description)
Thank you!

www.clarin.nl