Action Categorisation in Multimodal Instructions
van der Sluis, Ielka; Vergeer, Renate; Redeker, Gisela

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 16-12-2019
Action Categorisation in Multimodal Instructions

ABSTRACT

We present an explorative study for the (semi-)automatic categorisation of actions in Dutch multimodal first aid instructions, where the actions needed to successfully execute the procedure in question are presented verbally and in pictures. We start with the categorisation of verbalised actions and expect that this will later facilitate the identification of those actions in the pictures, which is known to be hard. Comparisons of verb and picture representations showed that the verbal and visual representations will allow us to determine the effectiveness of picture-text combinations and will eventually support the automatic generation of multimodal documents. We used Natural Language Processing tools to identify and categorise 2,388 verbs in a corpus of 78 multimodal instructions (MIs). We show that the main action structure of an instruction can be retrieved through verb identification using the Alpino parser followed by a manual selection operation. The selected main action verbs were subsequently generalised and categorised with the use of Cornetto, a lexical resource that combines a Dutch Wordnet and a Dutch Reference Lexicon. Results show that these tools are useful but also have limitations which make human intervention essential to guide an accurate categorisation of actions in multimodal instructions.

Keywords: instructions, actions, verbs, categorisation, task structure.

RESEARCH QUESTION

How can the constituent actions in Dutch first aid procedures be identified and categorised by (semi-)automatic natural language analysis, so that the resulting action categories can be used to identify picture-text relations in multimodal instructions?

CORPUS

78 multimodal instructions (MIs) from the annotated PAT corpus1 Published in two editions of Het Oranje Kruis Boekje (2011*, 2016) Overlap in terms of the first aid tasks:
• 25 tasks appear in both editions (yielding 50 MIs),
• and 28 appear only in one edition.

Het Oranje Kruis2 is a Dutch organisation that provides learning materials for first aid certification trainings.

1 http://www.rug.nl/zkt/ pat

http://www.hetoranjecruis.nl/

EXAMPLES

How to place a victim in the recovery position.

Renate Vergeer  r.j.vergeer@rug.nl
Ielka van der Sluis            i.f.van.der.sluis@rug.nl
Gisela Redeker             g.redeker@rug.nl

Figure 1: Mijlja Placing a victim in the recovery position (Het Oranje Kruis Boekje, 2011).

Figure 2: Part of Mijlja Placing a victim in the recovery position (Het Oranje Kruis Boekje, 2006).

METHOD

1. Selection of the verbs from MIs using Alpino (Van Noord, 2006)
2. Manual exclusion of modalised actions, negated actions, conditional actions, and elements
3. Selection of the verbs from MIs using Alpino (Van Noord, 2006)
4. Abstraction from hyperonyms using Cornetto 2.0 synset XML and other sources
5. Further analysis using the Cornetto Demo of 21 hyperonyms with a frequency > 1%

RESULTS

Table 1: Features and words used to identify modality verbs, negated actions, conditional actions and warnings in the parsed MI texts and the numbers and percentages of excluded verbs.

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>WORDS</th>
<th>EXCLUDED VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modal Verbs</td>
<td>221</td>
<td>(9.3%)</td>
</tr>
<tr>
<td>Explicit Negation</td>
<td>297</td>
<td>(12.4%)</td>
</tr>
<tr>
<td>Integrated Negation</td>
<td>96</td>
<td>(4.0%)</td>
</tr>
<tr>
<td>Other Negation</td>
<td>174</td>
<td>(7.3%)</td>
</tr>
</tbody>
</table>

Table 2: Hyperonym sources.

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>UNIQUE VERBS</th>
<th>TOTAL NUMBER OF VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornetto 2.0 DB</td>
<td>127</td>
<td>938 (66.5%)</td>
</tr>
<tr>
<td>Cornetto Demo</td>
<td>118</td>
<td>365 (25.3%)</td>
</tr>
<tr>
<td>Van Dale Dictionary</td>
<td>22</td>
<td>88 (6.8%)</td>
</tr>
<tr>
<td>Prefix stripping</td>
<td>12</td>
<td>25 (1.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>282</td>
<td>1,411 (100%)</td>
</tr>
</tbody>
</table>

Table 3: Eight hyperonym categories with their synsets, frequencies and corpus examples

The table shows the eight hyperonym categories with their synsets, frequencies and corpus examples.

DISCUSSION & FUTURE WORK

• Improvement of existing tools to automatically analyse the Dutch MI dataset:
  - Alpino does not recognise all verbs + does not identify negations, conditions and warnings.
  - Cornetto provides hyperonyms for only two thirds of the MI lemma’s.
• Categorisation of the verbs in the PAT MI corpus to eventually automatically recognise actions in text and pictures.
• Comparison of matching textual and pictorial presentations using
  - more detailed analysis of aspect, modality, and adverbial specifications of manner
  - more fine-grained visual analysis identifying posture, gaze, and the positions of persons.
• Conducting user studies to test the effects of possible text-picture pairings.