Identifying Semantic Relations and Functional Properties of Human Verb Associations

Sabine Schulte im Walde & Alissa Melinger
Saarland University, Germany
(schulte, melinger)@coli.uni-sb.de

Introduction

• Goal: use human verb associations for an investigation of verb properties
• Basis: collection of semantic associates evoked by German verbs in a web experiment
• Semantic associate: concept spontaneously called to mind by a stimulus word
• Task: identification and quantification of relationships between stimulus and response

Semantic Verb Relations

1. Semantic verb relations: The notion of semantic verb relations is crucial for NLP tasks and applications: verb clustering, thesaurus extraction, word sense discrimination, summarisation, etc. Different applications incorporate different semantic verb relations, varying with respect to their demands. We suggest that an analysis of human verb-verb associations may identify the range of semantic relations which are crucial in NLP.

2. Functional noun properties: In data-intensive lexical semantics, words are commonly modelled by distributional vectors, and the relatedness of words is measured by vector similarity. The features in the distributional descriptions can be varied in nature. We assume that the noun associates in our verb experiment are related to conceptual roles of the respective verbs, and investigate their functional languages, to identify prominent roles and conditions for distributional verb descriptions.

Morpho-Syntactic Analysis:

• Based on an empirical, quantitative dictionary, each response was classified according to its part-of-speech (POS).
• For ambiguous responses such as überlegen (verb “think about” vs. adjective “superior”), the response frequency was split over the possible POS tags. 4% of all response types were ambiguous.
• Table 2 provides the overall distributions for response tokens over POS tags.
• The average pattern varies with respect to individual verbs, frequencies, semantic class.
• Correlation between POS and target verb frequency:
  • verb frequency ↑ → noun/verb responses, verb frequency ↓ → noun/adj responses
  • Relation between semantic class and POS?
• Verb-verb relations (such as temporal order, cause, consequence) are represented in a large proportion of the verb-verb pairs.

Syntax-Semantic Noun Functions

• Our lexicalised grammar model contains empirical distributions of verbs for subcategorisation frame types and nominal argument fillers (Schulte im Walde, 2003). The model parameters were estimated in an unsupervised training procedure, using 35 million words of a large German newspaper corpus from the 1990s.
• We looked up the linguistic relationships between target verbs and noun associates. For each noun response, we evaluated whether the response co-occurs with the target verb as an argument filler in a particular frame (ambiguity considered). For example, the associate Kuchen “cake” in response to backen “bake” appears as the direct object of the verb in transitive frames AND as intransitive subject.
• 28% of all noun associates were identified with a subcategorisation function of the target verbs. 11 frame-slot functions were evoked by the noun tokens with prob > 1%.

Conclusions

• 20,000 verb associates:
  • 37% of the verb-verb pairs are specified as paradigmatic GermanNet relations. Verb-verb pairs with no GermanNet relation provide an empirical basis for detecting missing links in the taxonomy.
  • Non-classical verb-verb relations (such as temporal order, cause, consequence) are represented in a large proportion of the verb-verb pairs.
  • They represent an excellent basis for defining an exhaustive set of non-classical relations.

• Window look-up to determine the distance between two associated verbs:
  • 200 million words newspaper corpus, co-occurrence windows 5/20/50 words left/right; co-occurrence rates of 85/90/97% for paradigmatic (GermanNet) relations, i.e. 37% of data; 67/74/99% for non-GermanNet relations, i.e. 63% of data.

Web Experiment

• 330 German verbs were selected.
  • Variety of semantic classes (loosely based on Levin, 1993).
  • Variety of verb frequencies (based on 35 million word corpus).
  • Random division into 6 presentation lists with 55 verbs each.
  • 299 native German speakers participated, between 44 and 54 for each data set.
  • Ellicition was conducted over the Internet.
  • Participants had 30 secs per task to provide as many responses as possible.
  • In total, we collected data for 16,445 trials with over 80,000 responses.
  • Each response was quantified by its frequency of occurrence for any individual target, cf. Table 1.

Semantic Verb Relations

The lexicical semantic taxonomy GermanNet organises words into sets of synonyms. The sets are interconnected by semantic relations like hypernymy, antonymy, etc.

We looked up the semantic relation between the target and its responses (polysynonymy considered). The semantic relations were quantified by the target-response frequency.

Figure 1 shows the distribution of relations between target verbs and verb responses. 54% of response tokens were not encoded as GermanNet semantic relation: the semantic relations were quantified by the target-response frequency.

Table 1. Most frequent responses for Allegen.

| Allegen | complan, moan, sus | Gericht | court | 19
|---------|---------------------|---------|------|---
| Gimmern | moan | 18
| Jekken | cry | 13
| Kiwens | lauern | 11
| Richter | judge | 9
| Klage | complaint | 7
| Leid | suffering | 5
| Trauer | mourning | 6
| Klagemauer | Wailing Wall | 5
| Welt | 5

Table 2. POS distributions over all verb and response tokens.

<table>
<thead>
<tr>
<th>POS</th>
<th>Freq</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>19,863</td>
<td>25%</td>
</tr>
<tr>
<td>N</td>
<td>48,905</td>
<td>62%</td>
</tr>
<tr>
<td>ADJ</td>
<td>8,510</td>
<td>11%</td>
</tr>
<tr>
<td>ADV</td>
<td>1,268</td>
<td>2%</td>
</tr>
</tbody>
</table>

Figure 1. Distribution over semantic relations.

Acknowledgements:

• This work was supported by the project: ‘Towards a GermanNet-Based Multilingual Dictionary’ (DFG, TRR 47, project A17.3) and the project: ‘Towards a GermanNet-Based Multilingual Dictionary’ (DFG, TRR 47, project A17.3).