VerbOzean

Automated web-based learning of semantic relations between German verbs

Michaela Regneri 27.11.06
Motivation

- many approaches aiming to learn semantic relations automatically
- omnipresent: the sparse data problem
- using the web means using the biggest corpus available
- VerbOcean (Chklovski & Pantel 2004) uses the web as corpus
- VerbOzean modifies the approach for German, making use of the web’s multilingualism
Outline

- System overview
- VerbOcean
- VerbOzean
- Discussion
System overview

<table>
<thead>
<tr>
<th>similarity</th>
<th>enablement</th>
<th>antonymy</th>
<th>strength</th>
<th>happens-before</th>
</tr>
</thead>
</table>

verb pairs

- talk – speak
- cook – eat

patterns

to verb 1 and then verb 2

to talk and then speak

3 hits
System overview

verb pairs
- talk – speak
- cook – eat

patterns
- to verb1 and then verb2
- to cook and then eat

instantiated patterns

similarity enablement antonymy strength happens-before

210 hits
Outline

- System overview
- VerbOcean
- VerbOzean
- Discussion
VerbOcean (in a nutshell)
Chklovski & Pantel 2004

- extracting highly associated verb pairs
- creating lexico-syntactic patterns
- testing for a semantic relation
- evaluation
Extracting verb pairs

… automatically with DIRT (Lin & Pantel 2001)

- comparing environment in dependency trees

  ![Dependency Tree](image)

  - here: restriction on S-V-O constructions
  - high overlap of X and Y slot fillers → root verbs of the trees become a pair
  - 29165 pairs extracted
Lexico-syntactic patterns

Some examples:

- **antonymy:**
  - either X or Y
  - either Xs or Ys
  - either Xed or Yed
  - either Xing or Ying
  - whether to X or Y

- **enablement:**
  - Xed * by Ying the
  - to X * by Ying the

- **happens-before:**
  - to X and then Y
  - Xed and then Yed
  - to X and later Y
  - Xed and later Yed

- **similarity:**
  - X ie Y
  - to X and Y

- **strength:**
  - X even Y
  - not only Xed but Yed
Testing for a semantic relation

- frequent verbs $\rightarrow$ many Google hits
- **pointwise MI** as association measure:

$$PMI(v_1, pattern, v_2) = \frac{p(v_1, pattern, v_2)}{p(v_1)*p(pattern)*p(v_2)}$$

- a verb pair $v_1$-$v_2$ belongs to a relation, if

$$\sum_{pattern} PMI(v_1, pattern, v_2) > \text{Cut-Off}$$

- the cut-off is determined manually
Testing for a semantic relation (cont.)

- how to estimate the probabilities:

\[ p(v) \approx \frac{\text{hits ("to v") } * C}{N} \]

\[ p(\text{pattern}) \approx \frac{\text{hits}_{\text{est}}(\text{pattern})}{N} \]

\[ p(v_1, \text{pattern}, v_2) \approx \frac{\text{hits}(v_1, \text{pattern}, v_2)}{N} \]

\[ N = \text{estimated number of words indexed in Google} \]

\[ C = \text{"correction factor"} \]

\[ \text{hits} = \text{Google hits} \]
Evaluation

- evaluation with two human annotators
- judges were asked whether or not the assigned relation was acceptable

- accuracy: 65.5% (= average of both annotations)
- Kappa: 0.78 (unknown origin…)
Outline

- System overview
- VerbOcean
- VerbOzean
- Discussion
VerbOzean

- VerbOcean architecture adapted to German verbs
- restriction on happens-before
- main challenges:
  - extraction of verb pairs (no DIRT for German)
  - treatment of German morphology (simple encoding in patterns is not appropriate)
  - estimating frequencies verbs and patterns as precisely as possible
Verb pair extraction

- data from experiment with human verb associations (Schulte im Walde & Melinger 2005)
- participants got a list of verbs and wrote down their associations
- all the stimulus-association pairs with verbs as association collected
- for VerbOzean, we used pairs whose association was assigned more than once
- 4824 verb pairs, 1286 distinct verb types
Patterns & Morphology

- surface patterns do not encode morphology
- with each pattern, some morphology constraints are associated
- hand-crafted patterns are refined with corpus search (frequent patterns with high precision are preferred)
- corpus: DeWac (Baroni & Kilgarriff 2006)
Patterns & Morphology (cont.)

<table>
<thead>
<tr>
<th>Pattern</th>
<th>constraints</th>
<th>count (DeWac)</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1 und schließlich v2</td>
<td>v1</td>
<td></td>
</tr>
<tr>
<td>v1 und später v2</td>
<td></td>
<td>814</td>
</tr>
<tr>
<td>zuerst v1 und dann v2</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td>v1 und anschließend v2</td>
<td></td>
<td>1576</td>
</tr>
<tr>
<td>v1 und bereits v2</td>
<td></td>
<td>76</td>
</tr>
<tr>
<td>v1 und bald v2</td>
<td>V1: PPP, v2:PPP</td>
<td>39</td>
</tr>
<tr>
<td>v1 und sofort v2</td>
<td></td>
<td>241</td>
</tr>
<tr>
<td>v1 und dann v2</td>
<td></td>
<td>1866</td>
</tr>
</tbody>
</table>
Gathering Google hits

verb pair -> morphology

restricts

pattern -> inflected forms

queries

Google

hits
Gathering Google hits

beginnen - enden

v1 || v2

v1 und später v2

beginnen und später enden
beginnt und später endet ...

19
+ 47
+ ...

Google™
Testing for “happens-before”

- starting point: PMI like in VerbOcean:

\[ PMI(v1, \text{pattern}, v2) = \frac{p(v1, \text{pattern}, v2)}{p(v1) \cdot p(\text{pattern}) \cdot p(v2)} \]

- we don’t consider constants like the web’s size

- our association measure S out of absolute numbers:

\[ S(v1, \text{pattern}, v2) = \frac{\text{hits}(v1, \text{pattern}, v2)}{\text{count}(v1) \cdot \text{count}(\text{pattern}) \cdot \text{count}(v2)} \]
Testing for “happens-before“ (cont.)

- a verb pair $v_1$-$v_2$ is labelled as happens-before, if

$$\sum_{pattern} S(v_1, pattern, v_2) > \text{Cut} - \text{Off}$$

- the cut-off is determined manually by examining a training set of 30 verb pairs

- asymmetry: label a pair $v_1$-$v_2$ only as h-b. if

$$\frac{\sum_{pattern} S(v_1, pattern, v_2)}{\sum_{pattern} S(v_2, pattern, v_1)} > 5$$

*(inherited from VerbOcean, cut-off of 5 refined)*
### Results – Top 10 for happens-before

<table>
<thead>
<tr>
<th>German Verb</th>
<th>English Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>verloben</td>
<td>affiance</td>
</tr>
<tr>
<td>verheiraten</td>
<td>marry</td>
</tr>
<tr>
<td>einfrieren</td>
<td>freeze</td>
</tr>
<tr>
<td>auftauen</td>
<td>thaw</td>
</tr>
<tr>
<td>verdauen</td>
<td>digest</td>
</tr>
<tr>
<td>ausscheiden</td>
<td>excrete</td>
</tr>
<tr>
<td>waschen</td>
<td>wash</td>
</tr>
<tr>
<td>trocknen</td>
<td>dry</td>
</tr>
<tr>
<td>erhitzen</td>
<td>heat</td>
</tr>
<tr>
<td>abkühlen</td>
<td>cool down</td>
</tr>
<tr>
<td>kauen</td>
<td>chomp</td>
</tr>
<tr>
<td>schlucken</td>
<td>swallow</td>
</tr>
<tr>
<td>pachten</td>
<td>rent</td>
</tr>
<tr>
<td>kaufen</td>
<td>buy</td>
</tr>
<tr>
<td>trennen</td>
<td>separate</td>
</tr>
<tr>
<td>scheiden</td>
<td>divorce</td>
</tr>
<tr>
<td>kneten</td>
<td>knead</td>
</tr>
<tr>
<td>essen</td>
<td>eat</td>
</tr>
<tr>
<td>schwimmen</td>
<td>swim</td>
</tr>
<tr>
<td>paddeln</td>
<td>canoe</td>
</tr>
</tbody>
</table>
Results – a composed frame

cut
schneiden
kochen
mix
mischen
formen
kneten
knead

cook

bake

ess

eat
Results - Evaluation

- evaluation set of 60 verb pairs:
  - 30 labelled as happens-before
  - 15 with hits, but under cut-off
  - 15 without any Google hits
- 5 annotators, ask to judge whether or not the verb pairs are typical for happens-before
- inter-annotator agreement of $K=0.54$
- precision of 50% (gold standard = majority decision)
Results - Evaluation (cont.)

- precision: annotator judgements ranking from 40% to 63.3% (VerbOcean: 67.6)
- ‘stricter’ annotation guideline than VerbOcean (“typical” vs. “acceptable”)
- Kappa: 0.33 with gold standard (worst inter-annotator-agreement: 0.35)
- recall: estimation of 5% – still a problem (which recent approaches try to solve)
Outline

- System overview
- VerbOcean
- VerbOzean
- Discussion
Discussion

- Proper definition of „happens-before“?
- PMI as association measure?
- More fine-grained treatment of German morphology?
- The web as corpus?
- How should we evaluate such a task?