German Linking Elements

Presentation for the Course Morphology as part of the Graduiertenkolleg Ökonomie und Komplexität in der Sprache by the Lecturers: Anke Lüdeling and Penka Stateva

based on chapter 6 of Krott, Andrea (2001) Analogy in Morphology. The Selection of Linking Elements in Dutch Compounds
Dissertation at the Kath. University of Nijmegen, Ponsen & Looijen bv, Wageningen, p.143-147.

Student: Marije Michel
December 20th 2002

Outline

Introduction
Experimental support
the linking –s-
the linking –(e)n-
o linking element ø
Prediction with TiMBL
Interactive Activation Model
Discussion and Conclusion

linking elements

• insertion of linking elements between two nouns forming a compound e.g. state+s+man
• three different types of languages
  • fully predictable (Russian)
  • stored item by item in the lexicon (English)
  • partly predictable (Germanic languages) GERMAN and DUTCH

example: Dutch

• paradigmatic analogy
  • strongest predictor is the left constituent family = compounds that share left constituent with target compound
e.g. schaap+en+boul, schaap+en+tong → schaap+en+X
• smaller effect of right constituent family
e.g. schipper+s+trui, jongen+s+trui → X+trui
• other predictors are rime, semantics, class etc.
  ⇨ no strict rulebased procedure
German linking elements

- \(-s, -e, -n, -ens, -es, -er\) (Umlaut)
- 65% \(\emptyset\) (default) e.g. Buch+ø+laden
- 17% \(+s+\) e.g. Schiff+s+meldung
- 15% \(+e)n+\) e.g. Straße+n+arbeiter

very complex system
RULE or ANALOGY???

Two theories

RULE
Dressler, Libben, Stark, Pons & Jarema (2001)

10 ling. categories of left constituents
(gender, inflection, phonology)
schwa-final feminin nouns \(+n+\) X

paradigmatic

ANALOGY
Present paper: Krott (2001)

What are the effects of the left and right constituent family for \(+s+, +e)n+\) and \(+ø+\)?

What are the effects of rime, gender, inflection?

Experiment 1: linking \(+s+\)

- 3 sets of left constituents L1-3
- 3 sets of right constituents R1-3
  - L1 & R1 constituent family: bias \(+s\)
  - L2 & R 2 constituent family: neutral
  - L3 & R3 constituent family: bias \(-s\)
- 2 factorial paradigm
  - L1 __ R3 e.g. Teufel __ Fest
  - L3 __ R1 e.g. Auflauf __ Hilfe

Results: \(+s+\)

\[\begin{array}{ccc}
\text{negative} & \text{neutral} & \text{positive} \\
\text{right positive} & \text{right neutral} & \text{right negative}
\end{array}\]

Main effect for bias in left position
Experiment 2: linking +(e)n+

- 3 sets of left constituents L1-3
- 3 sets of right constituents R1-3
  - L1 & R1 constituent family: bias +(e)n
  - L2 & R 2 constituent family: neutral
  - L3 & R3 constituent family: bias -(e)n
- 2 factorial paradigm
  - L1__ R3 e.g. Zitrone __ Ball
  - L3__ R1 e.g. Bank __ König

Results: +(e)n+

Main effect of bias in left position

Experiment 3: no linking ø

- 3 sets of left constituents L1-3
- 3 sets of right constituents R1-3
  - L1 & R1 constituent family: bias +ø
  - L2 & R 2 constituent family: neutral
  - L3 & R3 constituent family: bias - ø
- 2 factorial paradigm
  - L1__ R3 e.g. Wand __ Sittich
  - L3__ R1 e.g. Suppe __ Analyse

Results: ø

Main effect of bias in left position
Overview results

• only the left constituent family seems to have an effect. The right constituent had no influence.
• even the default ø is (partially) analogical determined.
• tend to use overt linking elements

TiMBL

• an examplar-based machine-learning algorithm for the modeling of analogy (Daelemans, Zavrel, Van der Stoot, & Van den Bosch, 2000)
• combines similarity-based reasoning with extensive storage of exemplars in an instance database
• compares the target with exemplars in the instance base using user-specified features

Information Gain and Prediction

How much information do the separate characteristics provide in order to predict the correct linking element?

Method:
Let TiMBL it run over the constituents of the experiments and the compounds listed in CELEX and calculate its accuracy.

Prediction Accuracy

<table>
<thead>
<tr>
<th>features</th>
<th>celex</th>
<th>experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s</td>
<td>(e)n</td>
</tr>
<tr>
<td>Constituent 1</td>
<td>87,4</td>
<td>79,9</td>
</tr>
<tr>
<td>Constituent 1 and 2</td>
<td>86,0</td>
<td>79,3</td>
</tr>
<tr>
<td>Rime</td>
<td>79,0</td>
<td>80,0</td>
</tr>
<tr>
<td>Rime, gender, inflection</td>
<td>84,0</td>
<td>82,0</td>
</tr>
<tr>
<td>C1, rime, gender, inflection</td>
<td>91,0</td>
<td>79,3</td>
</tr>
</tbody>
</table>

Percentages of TiMBL prediction compared to reality (CELEX and experiments)
**Prediction Accuracy**

<table>
<thead>
<tr>
<th>Features</th>
<th>celex</th>
<th>experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rime</td>
<td>79,9</td>
<td>79,9</td>
</tr>
<tr>
<td>Constituent 1</td>
<td>87,4</td>
<td>79,9</td>
</tr>
<tr>
<td>Constituent 1 and 2</td>
<td>86,9</td>
<td>79,9</td>
</tr>
<tr>
<td>Rime, gender, inflection</td>
<td>84,0</td>
<td>62,0</td>
</tr>
<tr>
<td>C1, rime, gender, inflection</td>
<td>91,9</td>
<td>79,3</td>
</tr>
</tbody>
</table>

- left constituent is the best predictor
- sometimes other characteristics help!

**Discussion**

- strong paradigmatic effect of the left constituent family on the selection of linking elements in German
- bias for ø is more easily overruled than others (though default)
- TIMBL analysis confirmed that the left constituent is the strongest predictor
- In German: INTERPLAY of both ⇒ Rules & Analogy!

**Interactive Activation Model**

- Seife
  - Seife -n-Blase
  - Seife -n-Schaum
  - Seife -n-Kiste
  - Rose -n-Wasser
  - Seide -n-Papier
  - Kreide-Zeichnung

**Conclusion**

- Linking elements are independent units in the mental lexicon
- Strong effect of left constituent ⇒ tight connection between left constituent and linking element
<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>