Association Strength and Lexical Choice
Swedish slags and sorts

Methodology & Statistics
Caroline Morris
June 5th 2012
Research questions

• Has head shift become more frequent in the period 1830-2009?

• Is *slags* used more with neuter nouns and *sorts* more with common gender nouns?
Gender in Swedish NPs

- COMMON (former masculine and feminine) and NEUTER
- 72% of all nouns are of common gender
- nouns, adjectives, pronouns and articles (both definite and indefinite) agree in gender (in singular NPs)
Head shift in slags and sorts

\[ \text{DET}_i \text{ NP1}_i \text{-s} \ [\text{NP2}] \] > \ [\text{DET}_j \ [\text{NP1-}\text{s}] \ \text{NP2}_j]\]

head = NP1 > head = NP2

descriptive genitive > adjective/degree modifier
Descriptive genitives – no head shift

- **ett slags bizzarrt mönster**  
  a-NEUT kind-NEUT.GEN bizarre-NEUT pattern-NEUT  
  ‘a kind of bizarre pattern’

- **något slags profet**  
  some-NEUT kind-NEUT.GEN prophet-COMM  
  ‘a kind of prophet’

- **en sorts explosion**  
  a-COMM sort-COMM.GEN explosion-COMM  
  ‘a sort of explosion’

- **en sorts skvaller**  
  a-COMM sort-COMM.GEN gossip-NEUT  
  ‘a sort of gossip’
Adjectives – head shift

- **en slags vaggsång** by Brem1830
  - a-COMM kind-NEUT.GEN lullaby-COMM
  - ‘a kind of lullaby’

- **vilken slags makt** by GP09
  - which-COMM kind-NEUT.GEN power-COMM
  - ‘which kind of power’

- **ett sorts pris** by Hufv1999
  - a-NEUT sort-COMM.GEN prize-NEUT
  - ‘a sort of prize’

- **något sorts bakverk** by GP09
  - some-NEUT sort-COMM.GEN pastry-NEUT
  - ‘some sort of pastry’
Odds ratio

• association strength for categorical data
• tests whether variables are independent
• 2x2 contingency tables, one for each period
Contingency tables - slags

<table>
<thead>
<tr>
<th></th>
<th>neuter</th>
<th>common</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>slags + sg. noun</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
<td>$n_{1+}$</td>
</tr>
<tr>
<td>noun</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
<td>$n_{2+}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>neuter</th>
<th>common</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>slags + sg. noun</td>
<td>134</td>
<td>260</td>
<td>394</td>
</tr>
<tr>
<td>noun</td>
<td>28</td>
<td>72</td>
<td>100</td>
</tr>
</tbody>
</table>

$$\hat{\theta} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$
### Contingency tables – sorts

<table>
<thead>
<tr>
<th></th>
<th>neuter</th>
<th>common</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>slags + sg. noun</td>
<td>(n_{11})</td>
<td>(n_{12})</td>
<td>(n_{1+})</td>
</tr>
<tr>
<td>noun</td>
<td>(n_{21})</td>
<td>(n_{22})</td>
<td>(n_{2+})</td>
</tr>
<tr>
<td>slags + sg. noun</td>
<td>11</td>
<td>78</td>
<td>394</td>
</tr>
<tr>
<td>noun</td>
<td>28</td>
<td>72</td>
<td>100</td>
</tr>
</tbody>
</table>

\[
\hat{\theta} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}
\]
Odds ratio

- if $\theta = 1$ there is no association: the variables are independent
- if $\theta > 1$ the odds are higher in row 1
- if $\theta < 1$ the odds are higher in row 2
Hypothesis

• if *slags* is indeed more often combined with neuter nouns then the odds will be higher in row 1
• if *sorts* is indeed more often combined with common gender nouns then the odds will be higher in row 2
Results (lexical choice)

- *slags* 1830 – 1976: \( \theta > 1 \) (+/- 1,3)
- *slags* 1999, 2009: \( \theta = 1 \)
- *sorts* in all periods: \( \theta < 1 \) (+/- 0,5)

-> there is an association between the choice for *slags* (neuter) and *sorts* (common) and the gender of the noun that they are combined with
# Results (head shift)

## slags:

<table>
<thead>
<tr>
<th>sub-corpus</th>
<th>period</th>
<th>headshift</th>
<th>no headshift</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Older Sw. novels</td>
<td>1800s</td>
<td>31</td>
<td>229</td>
<td>0.135</td>
</tr>
<tr>
<td>Strindberg</td>
<td>1861-1912</td>
<td>45</td>
<td>166</td>
<td>0.271</td>
</tr>
<tr>
<td>ORDAT (periodical)</td>
<td>1925–1958</td>
<td>2</td>
<td>41</td>
<td>0.049</td>
</tr>
<tr>
<td>Press65 and Press 76</td>
<td>1965, 1976</td>
<td>36</td>
<td>134</td>
<td>0.269</td>
</tr>
<tr>
<td>Hufvudstadsbladet</td>
<td>1999</td>
<td>74</td>
<td>173</td>
<td>0.428</td>
</tr>
<tr>
<td>GP09</td>
<td>2009</td>
<td>83</td>
<td>198</td>
<td>0.419</td>
</tr>
</tbody>
</table>

## sorts:

<table>
<thead>
<tr>
<th>sub-corpus</th>
<th>period</th>
<th>headshift</th>
<th>no headshift</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Older Sw. novels</td>
<td>1800s</td>
<td>3</td>
<td>8</td>
<td>0.375</td>
</tr>
<tr>
<td>Strindberg</td>
<td>1861-1912</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>ORDAT (periodical)</td>
<td>1925–1958</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Press65 and Press 76</td>
<td>1965, 1976</td>
<td>2</td>
<td>9</td>
<td>0.222</td>
</tr>
<tr>
<td>Hufvudstadsbladet</td>
<td>1999</td>
<td>7</td>
<td>16</td>
<td>0.438</td>
</tr>
<tr>
<td>GP09</td>
<td>2009</td>
<td>16</td>
<td>47</td>
<td>0.340</td>
</tr>
</tbody>
</table>
Discussion

• head shift with *slags* has become more frequent since 1830, no clear development for *sorts*

• association has become weaker for *slags*, but *sorts* is still mainly used with common gender nouns
  – less likely to change from common > neuter
  – *sorts* is less frequent, absolute numbers are lower
  – *sorts* is more recent, slight increase towards $\theta = 1$ might continue