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*

[DET_i $NP1_i$-s [NP2]]	>	[DET_j [NP1-s] $NP2_j$]
head = NP1	>	head = NP2
descriptive genitive	>	adjective/degree modifier

Descriptive genitives – no head shift

- ***ett*** ***slags*** *bizarrt* *mönster* ORDAT
  a-NEUT kind-NEUT.GEN bizarre-NEUT pattern-NEUT
  'a kind of bizarre pattern'
- ***något*** ***slags*** *profet* GP09
  some-NEUT kind-NEUT.GEN prophet-COMM
  'a kind of prophet'

- ***en*** ***sorts*** *explosion* Press65
  a-COMM sort-COMM.GEN explosion-COMM
  'a sort of explosion'
- ***en*** ***sorts*** *skvaller* GP09
  a-COMM sort-COMM.GEN gossip-NEUT
  'a sort of gossip'

Adjectives – head shift

- ***en*** ***slags*** *vaggsång* Brem1830
  a-COMM kind-NEUT.GEN lullaby-COMM
  ‘a kind of lullaby’
- ***vilken*** ***slags*** *makt* GP09
  which-COMM kind-NEUT.GEN power-COMM
  ‘which kind of power’

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

	neuter	common	total
slags + sg. noun	n_{11}	n_{12}	n_{1+}
noun	n_{21}	n_{22}	n_{2+}

aesvr	neuter	common	total
slags + sg. noun	134	260	394
noun	28	72	100

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

Contingency tables – sorts

	neuter	common	total
slags + sg. noun	n_{11}	n_{12}	n_{1+}
noun	n_{21}	n_{22}	n_{2+}

aesvr	neuter	common	total
slags + sg. noun	11	78	394
noun	28	72	100

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

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:

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

- sorts:

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

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 θ = 1 might continue