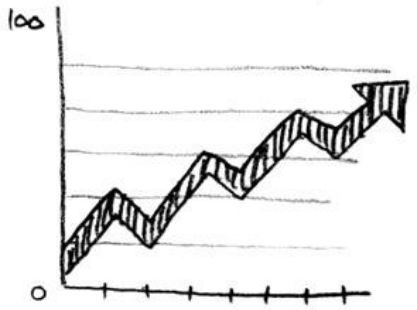
Association Measures:

Minimum Sensitivity

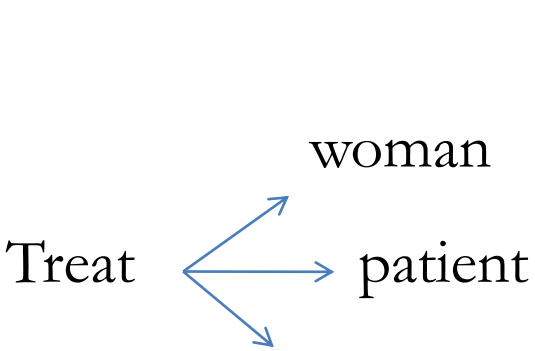


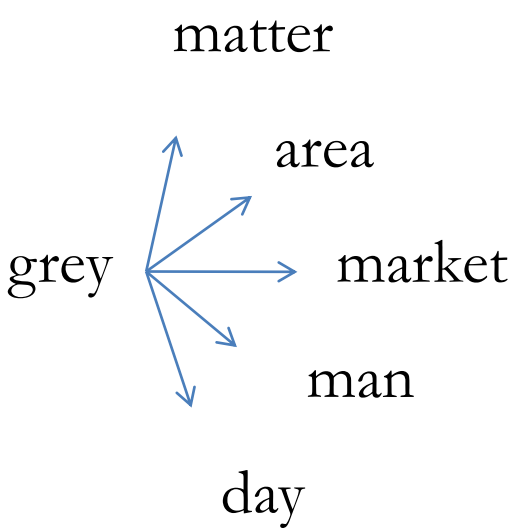
by Ihor Tytyk

Groningen 2011

Collostruction

Collocation:





Collostruction:

regret → **gerund** (I regret leaving school so young.)
→ **sentential** (I regret that I will be unable to attend.)
→ **infinitive** (I regret to inform you that.)

Collostruction Strength

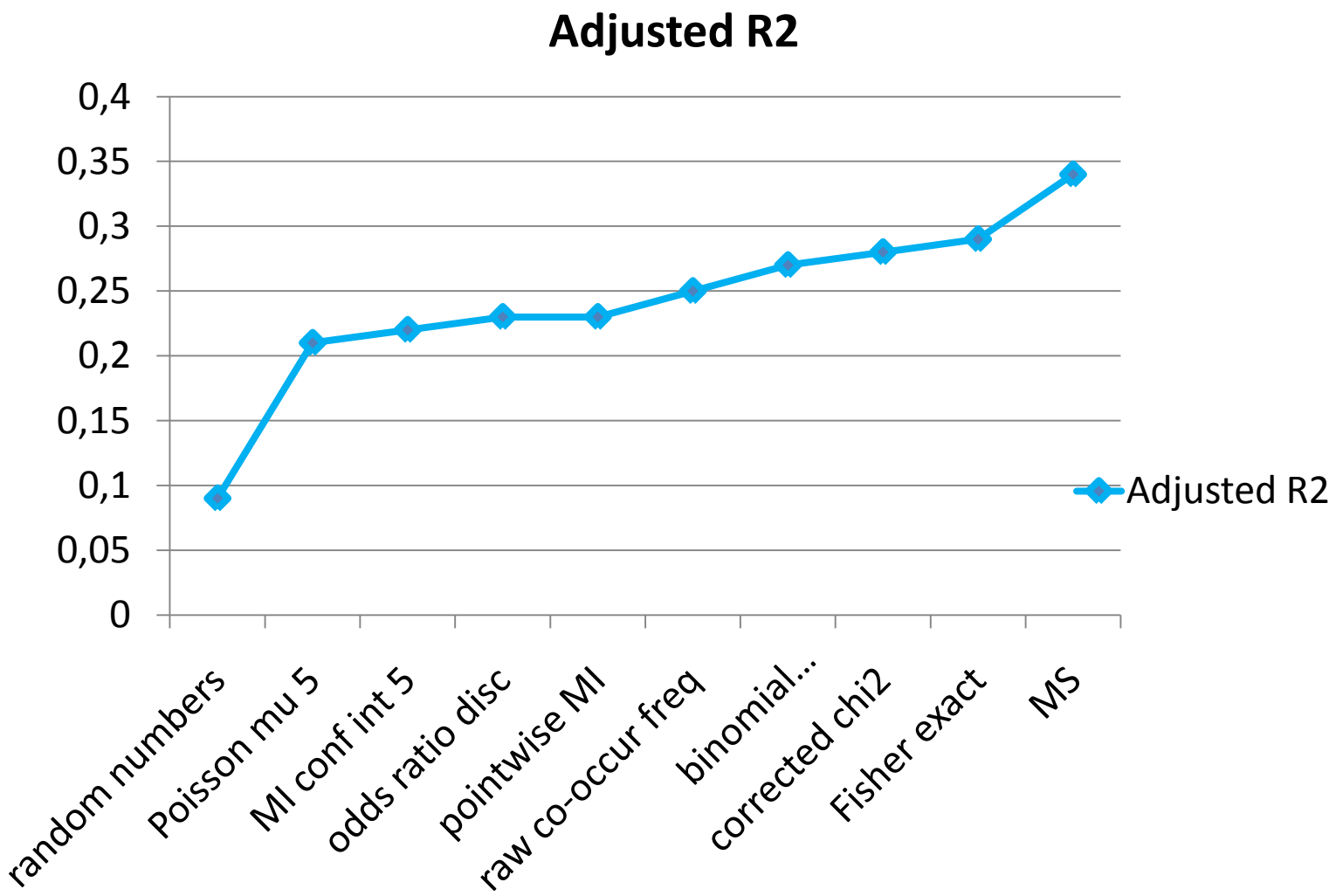
- Collostructional analysis *Stefan Th. Gries (University of California, Santa Barbara) and Anatol Stefanowitsch (University of Bremen)*
- Collostruction strength is the degree of attraction that a word C_j exhibits to a construction C_k

It is some sort of glue between units, a verb and a complement.

- based on frequency
- can be measured

association strength – degree of collocativity – degree of attraction – collostructional strength

Methods comparison



R^2 – coefficient of determination

Contingency Table

	infinitive	gerund	
Verb v (e.g. regret)	O_{11}	O_{12}	R_1
All other verbs	O_{21}	O_{22}	R_2
	C_1	C_2	N

O_{11}, O_{12}, O_{21}, O_{22} – frequencies; C_1,C_2,R_1,R_2 – totals; N – sample size

In order to calculate MS two probabilities are needed:
P (verb | construction) and P **(construction | verb)**

$$S_{w1} = \frac{O_{11}}{C_1} = P(w1|w2) \qquad S_{w2} = \frac{O_{11}}{R_1} = P(w2|w1)$$

Minimum sensitivity

$$S_{w1} = \frac{O_{11}}{C_1} = P(w1|w2) \quad S_{w2} = \frac{O_{11}}{R_1} = P(w2|w1)$$

$$MS = min\left\{\frac{O_{11}}{R_1}; \frac{O_{11}}{C_1}\right\}$$

$$AM = MS$$

Experinmental data

Verb synonyms: ***to propose, to suggest***

Collostruction comparison

Complements:

Suggest

+ infinitive (*They suggest to go see a doctor.*)
+ to NP (*He suggested me to go to Europe.*)
+ gerund (*Tim suggests finding your life purpose.*)
+ *that*-clause (*It was suggested to me to do this upon arrival in Norway.)*
+ NP (*I suggest a solution for designing service agents.*)

Corpora / Tools

- Ukwac (UK Web archive)

A ukwac file with 17mil. words

- CWB (Open Corpus Workbench) is a collection of open-source tools for managing and querying large text corpora
- CQP was used for querying the corpus (collostructions extraction)
- With the help of UCS association measures were calculated

Calculation

	that-clause	X - construction	
Suggest	7541	14930	22471
Verb	248438		
	255979		

$$MS = min\left\{\frac{7541}{255979}; \frac{7541}{22471}\right\} =$$

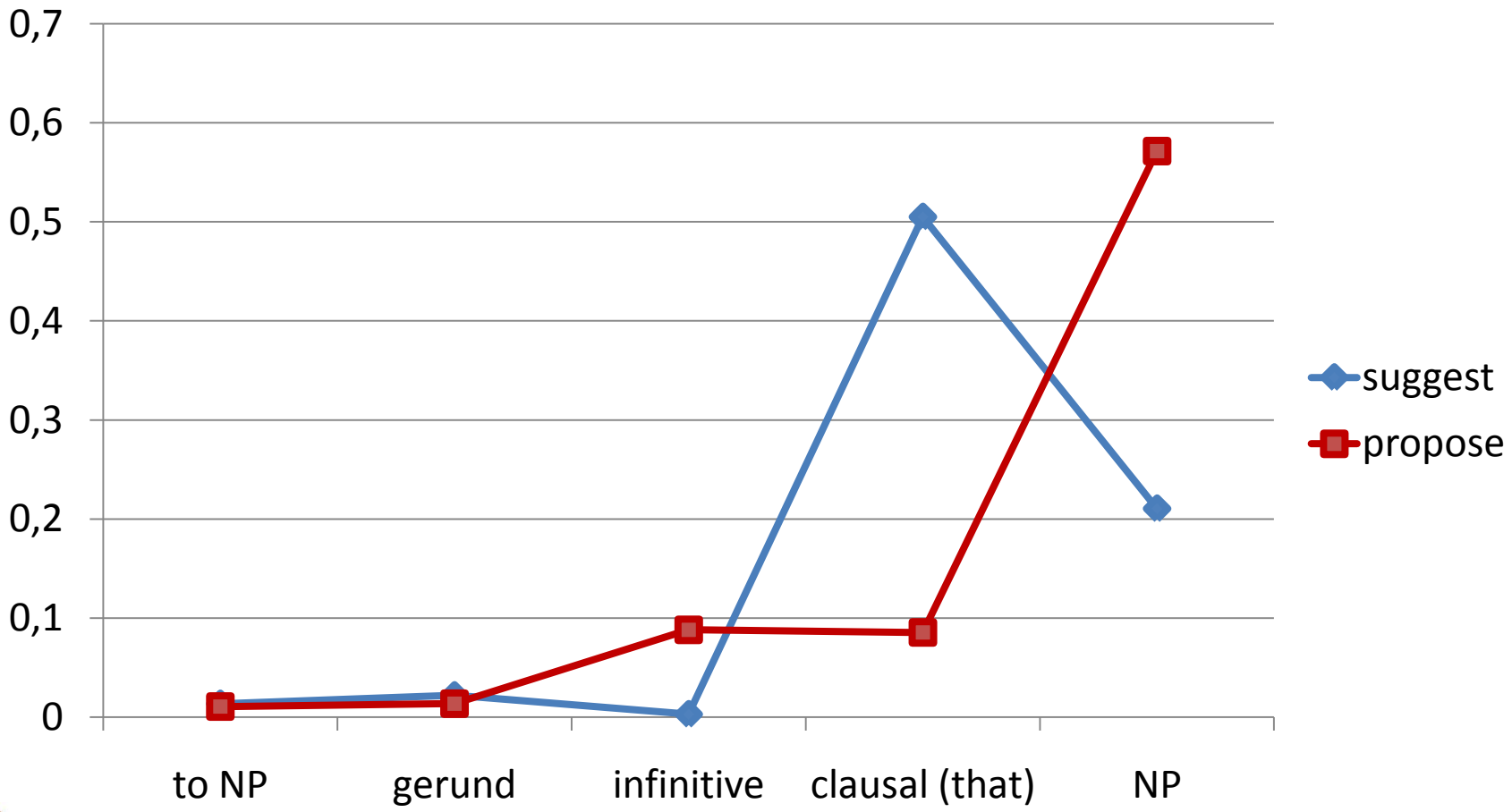
$$= min\{0.0294594; 0.335588\} = 0.02945$$

Results

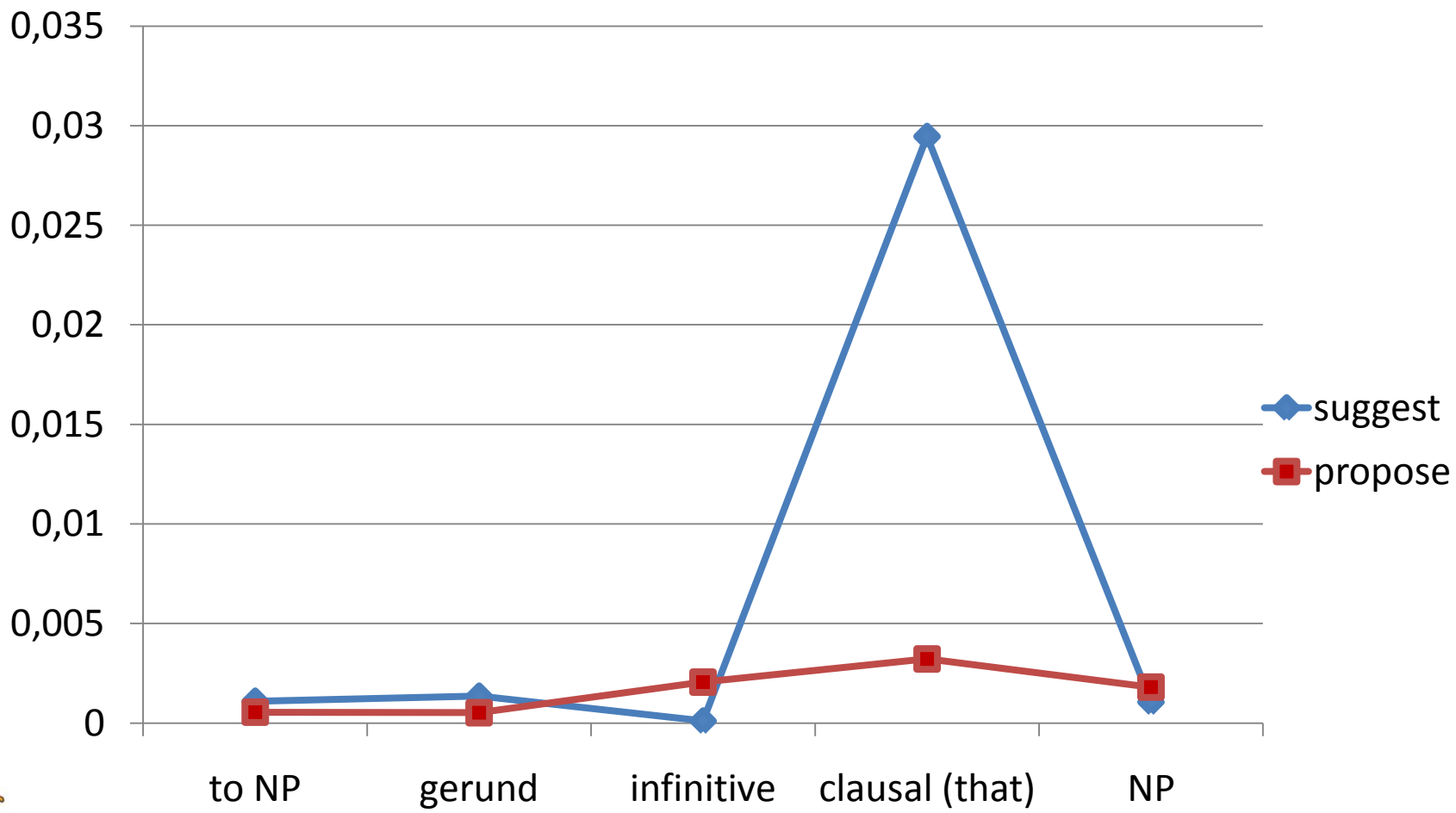
MS	to NP	gerund	infinitive	that-clause	NP
suggest	0,00110	0,00135	0,00010	0,02945	0,00105
propose	0,00054	0,00052	0,00205	0,00322	0,00180

MI					
suggest	1,69	1,78	0,68	3,13	1,67
propose	1,59	1,57	2,68	2,36	2,11

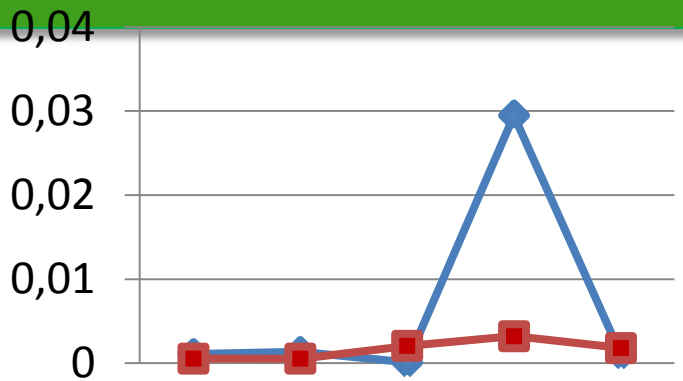
Complement taking probabilities

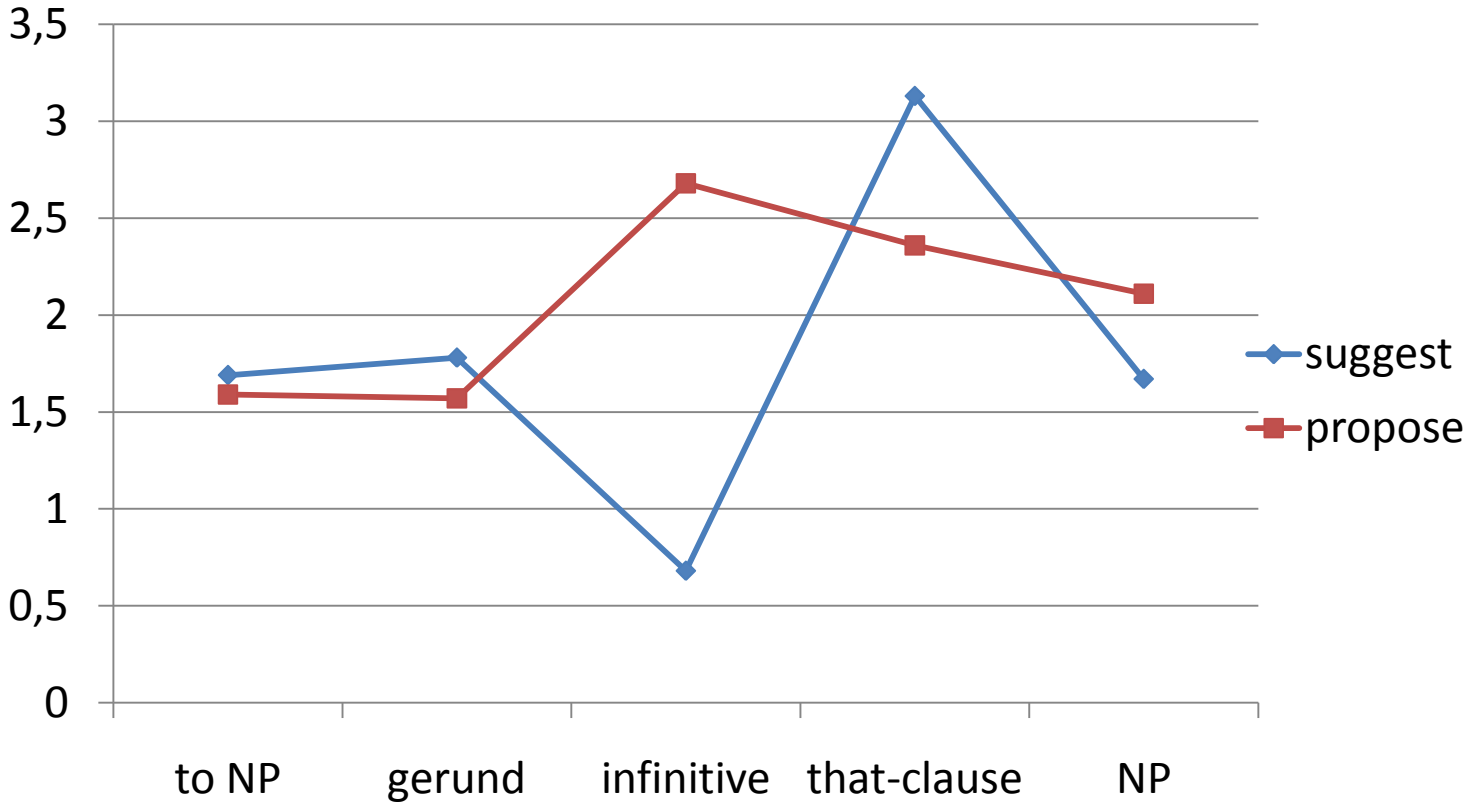


Minimum sensitivity



Mutual information





Method advantages

- free from underlying distributional assumptions
- computationally is less demanding
- less dependant on the sample size
- empirically is most adequate (acc. to D. Wiechmann)

Issues to consider

- More collostructions to analyze
- Some ‘gold standard’ is needed for more objective evaluation
- Construction-tag annotated corpora would be very useful

References

- Daniel Wiechmann (2008) *On the computation of collostructional strength: Testing measures of association as expressions of lexical bias.*
- Daniel Wiechmann (2008) *Some thoughts how to measure association strength.*