Verb polysemy and frequency effects in thematic fit modeling

Clayton Greenberg, Vera Demberg, and Asad Sayeed
Saarland University / M^{2}CI Cluster of Excellence

June 4, 2015

McRae et al. (1998) thematic fit

1. The cop arrested…
2. The crook arrested…

McRae et al. (1998) thematic fit

1. The cop arrested the crook.
2. The crook arrested by the cop confessed.

McRae et al. (1998) procedure

- On a scale from 1 (very uncommon) to 7 (very common), how common is it for a
  - snake
  - nurse
  - monster
  - baby
  - cat

  to **frighten** someone/something?

- How common is it for a
  - snake
  - nurse
  - monster
  - baby
  - cat

  to **be frightened by** someone/something?

Thematic fit datasets

Verb	Noun	Semantic role	Score
advise	doctor	agent	6.8
advise	doctor	patient	4.0
confuse	baby	agent	3.7
confuse	baby	patient	6.0
eat	lunch	agent	1.1
eat	lunch	patient	6.9
kill	lion	agent	2.7
kill	lion	patient	4.9
kill	man	agent	3.4
kill	man	patient	5.4

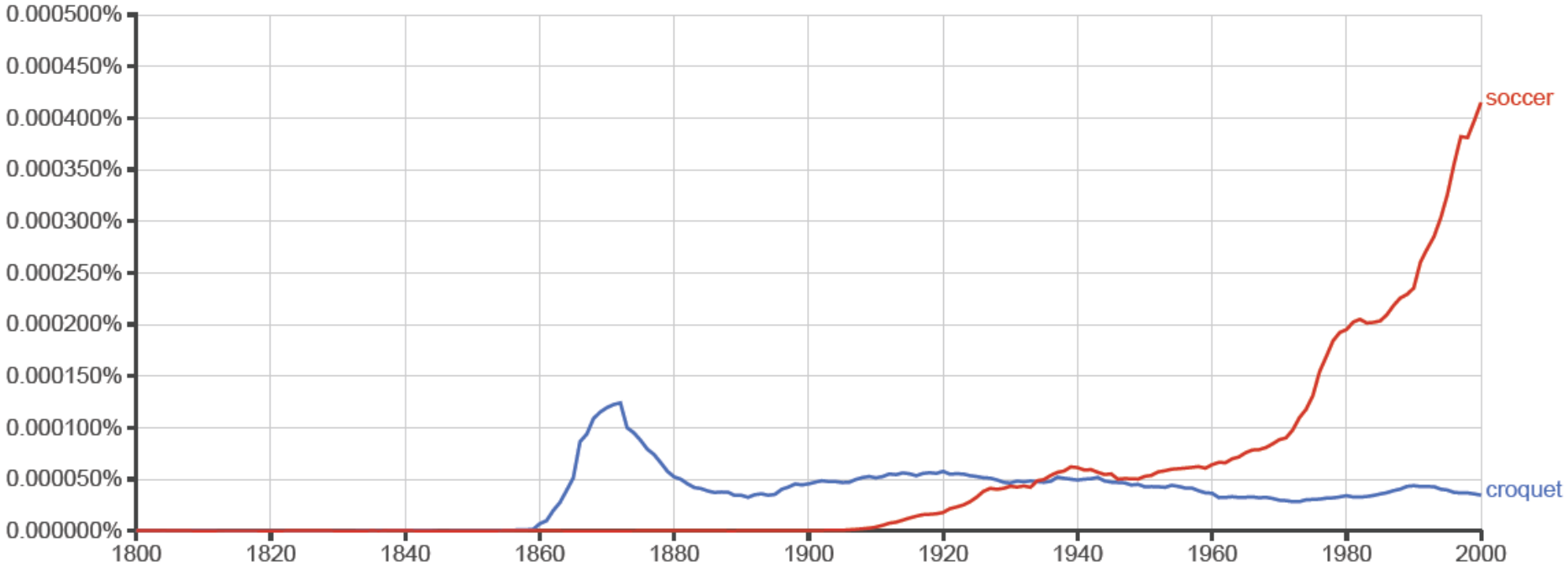
Judgements from Padó (2007)

Challenges to judgement well-formedness

Alice played croquet / soccer / piano / cheese in the garden

Role-filler frequency

How common is it for croquet/soccer to be played?

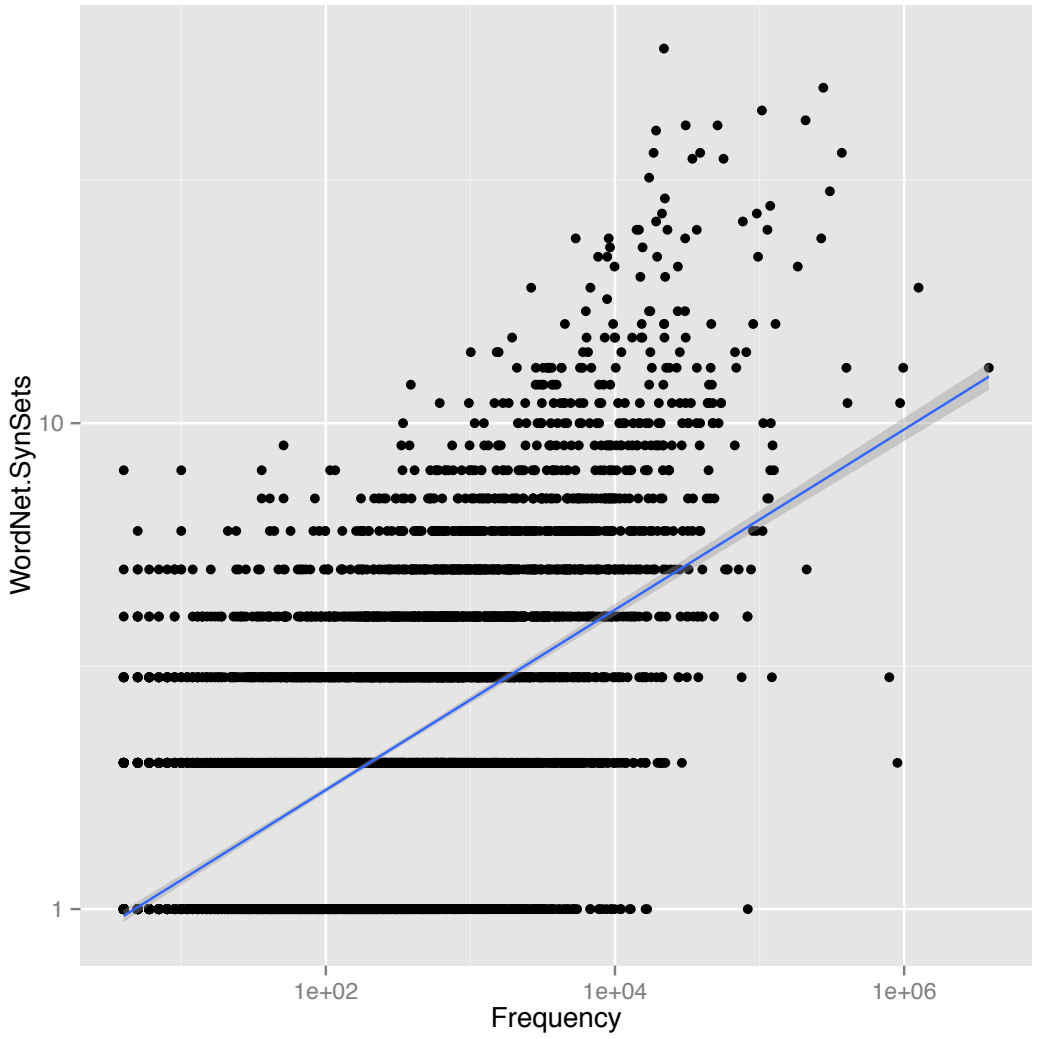


Google ngram (Michel et al., 2010) comparison of “croquet” and “soccer”

Polysemy

How common is it for soccer/the piano to be played?

Right: polysemy versus frequency of the most frequent verbs in COCA. Corpus obtained from Davies (2008).



Sense frequency

WordNet (Fellbaum, 1998) orders SynSets
based on their frequencies

play_1: participate in games or sport. "We played hockey all afternoon"; "play cards"; "Pele played for the Brazilian teams in many important matches"

play_7: perform music on (a musical instrument). "He plays the flute"; "Can you play on this old recorder?"

Research question

How do

1. role-filler frequency
2. polysemy
3. sense frequency

affect thematic fit judgements?

Stimuli selection

McRae et al. (1998)

- Many purposes
- Verbs have “well-defined” roles
- Role-fillers selected to fit their roles well
- Animate role-fillers preferred
- 146 verbs
- 1,444 (F,R,V) triples

Padó (2007)

- One purpose
- Verbs are most frequent in Penn Treebank & FrameNet
- Role-fillers selected to have a wide range of fit ratings
- Fully mixed animacy
- 18 verbs
- 414 (F,R,V) triples

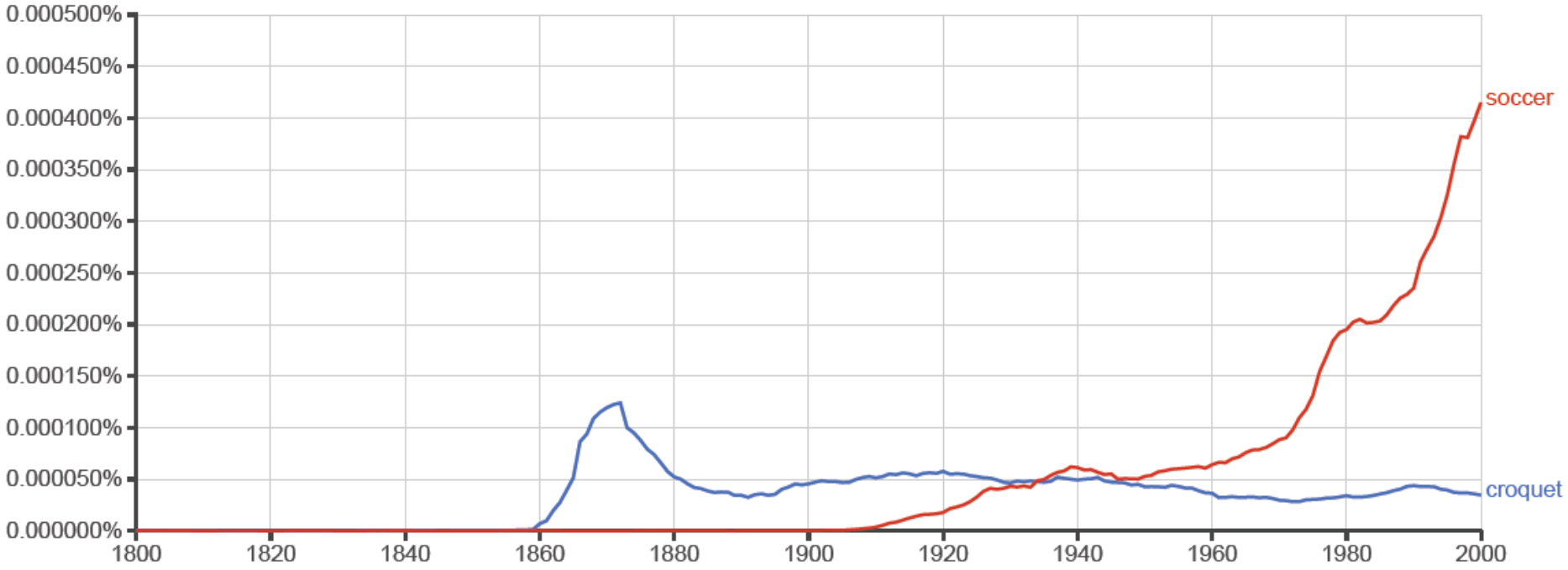
New formulation of the question

How common is it for croquet to be played?

Google ngram (Michel et al., 2010) comparison of “croquet” and “soccer”

New formulation of the question

Agreement scale: croquet is something that is played.



Google ngram (Michel et al., 2010) comparison of “croquet” and “soccer”

Verb selection

- Start with 500,000 most common word forms in COCA.
- Filter for verbs.
- Lemmatize using the WordNet lemmatizer in NLTK (Bird et al., 2009).
- Filter for only those that retrieve exactly one SynSet.
- Sort by frequency.
- Choose the first 48 that fit the paradigm (transitive, etc…).
- For each MONOSEMOUS verb, find a POLYSEMOUS verb (at least 2 salient senses, ~7 SynSets) with similar unigram frequency.

Stimuli examples

Filler type	Frequency	whip (1686, 6 SynSets)	punish (2908, 1 SynSet)

To find a good patient-filler, query COCA for: `VERB [at*] [nn*]`

Stimuli examples

Filler type	Frequency	whip (1686, 6 SynSets)	punish (2908, 1 SynSet)
Good		horse (32384)	outlaw (1487)

Find a much higher or lower (~10x) frequency synonym.

Stimuli examples

Filler type	Frequency	whip (1686, 6 SynSets)	punish (2908, 1 SynSet)
Good	high	horse (32384)	criminal (9271)
	low	stallion (818)	outlaw (1487)

For POLYSEMOUS verbs, repeat for second sense.

Stimuli examples

Filler type	Frequency	whip (1686, 6 SynSets)	punish (2908, 1 SynSet)
Sense1	high	horse (32384)	criminal (9271)
	low	stallion (818)	outlaw (1487)
Sense2	high	cream (19727)	
	low	frosting (905)	

Randomly shuffle good patient-fillers to assign poor ones.

Stimuli examples

Filler type	Frequency	whip (1686, 6 SynSets)	punish (2908, 1 SynSet)
Sense1	high	horse (32384)	criminal (9271)
	low	stallion (818)	outlaw (1487)
Sense2	high	cream (19727)	
	low	frosting (905)	
Bad	high	party (118292)	criminal (9271)
	low	gathering (7025)	outlaw (1487)

Reshuffle all of the ones that are too good.

Stimuli examples

Filler type	Frequency	whip (1686, 6 SynSets)	punish (2908, 1 SynSet)
Sense1	high	horse (32384)	criminal (9271)
	low	stallion (818)	outlaw (1487)
Sense2	high	cream (19727)	
	low	frosting (905)	
Bad	high	party (118292)	baby (70498)
	low	gathering (7025)	fetus (2329)

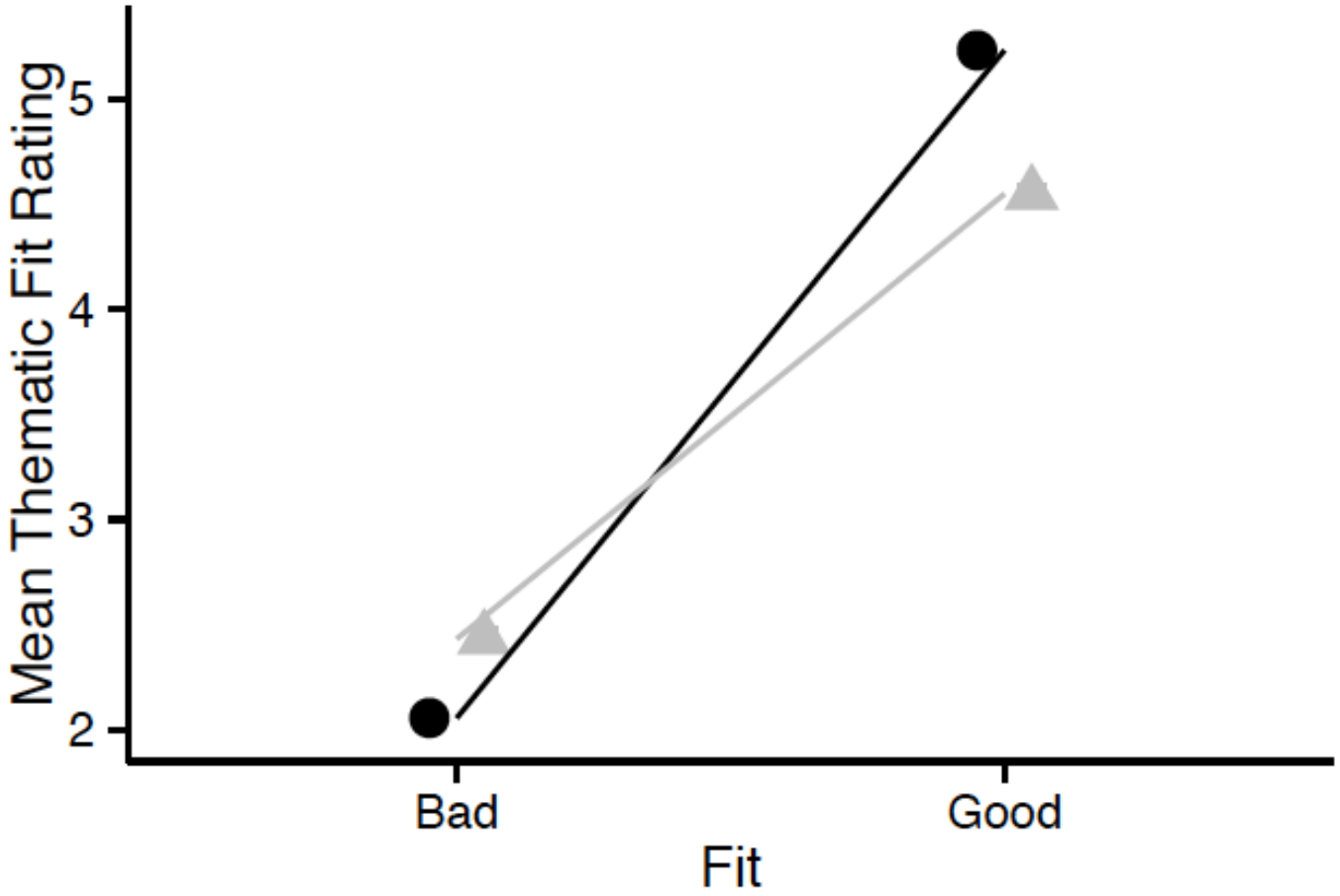
Filler items: the 240 most frequent triples from McRae et al. (1998)

Procedure

- Rewrite each verb in its past-participle form.
- Normalize each role-filler to singular with appropriate determiner.
- Choose either the +human or the –human template:
  - +human: __ is someone who is ___
  - –human: __ is something that is ___

- One survey
  - 6 POLYSEMOUS, 4 MONOSEMOUS, 5 fillers
  - Workers do not see a verb in more than one condition
  - Compensation: $0.15
  - 159 workers participated, 10 ratings per item.

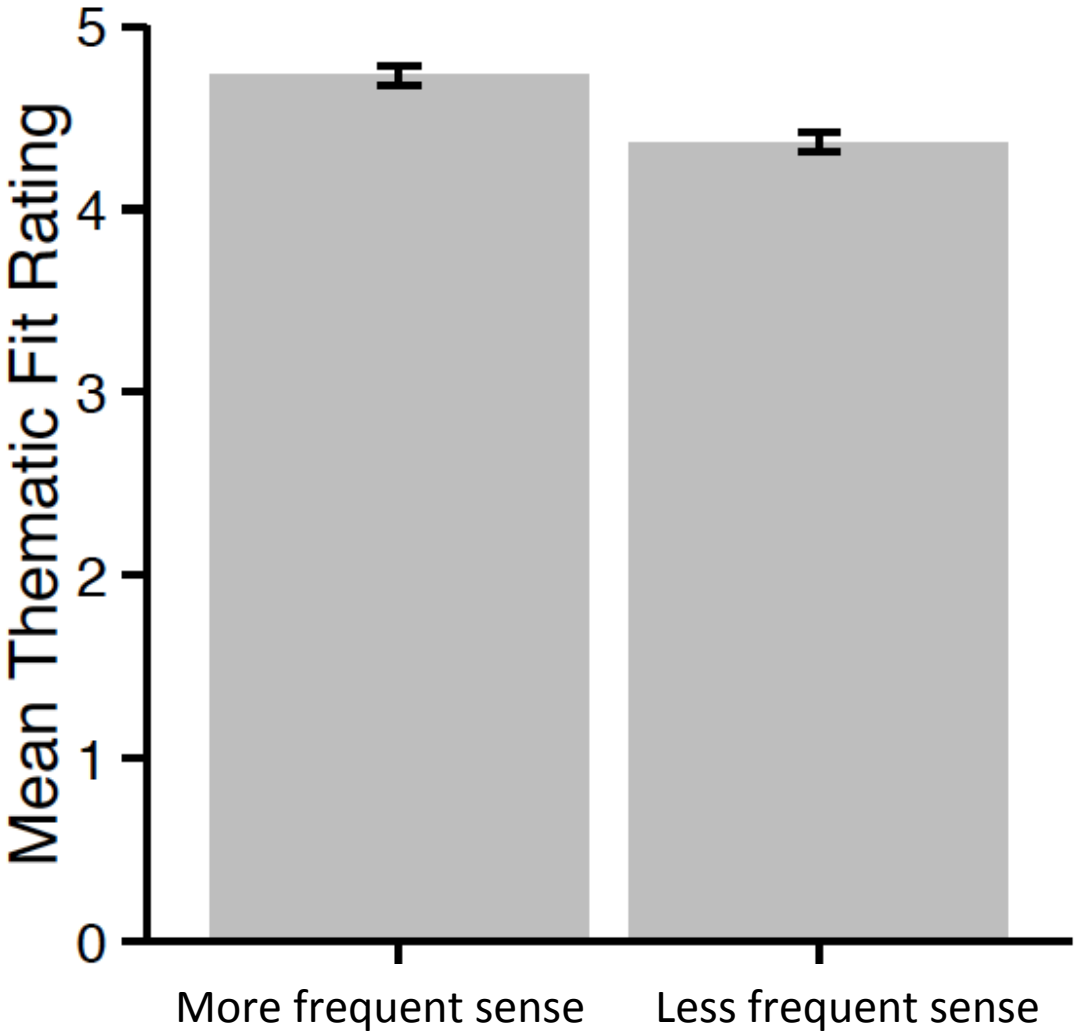
ANOVA results: polysemy-fit interaction



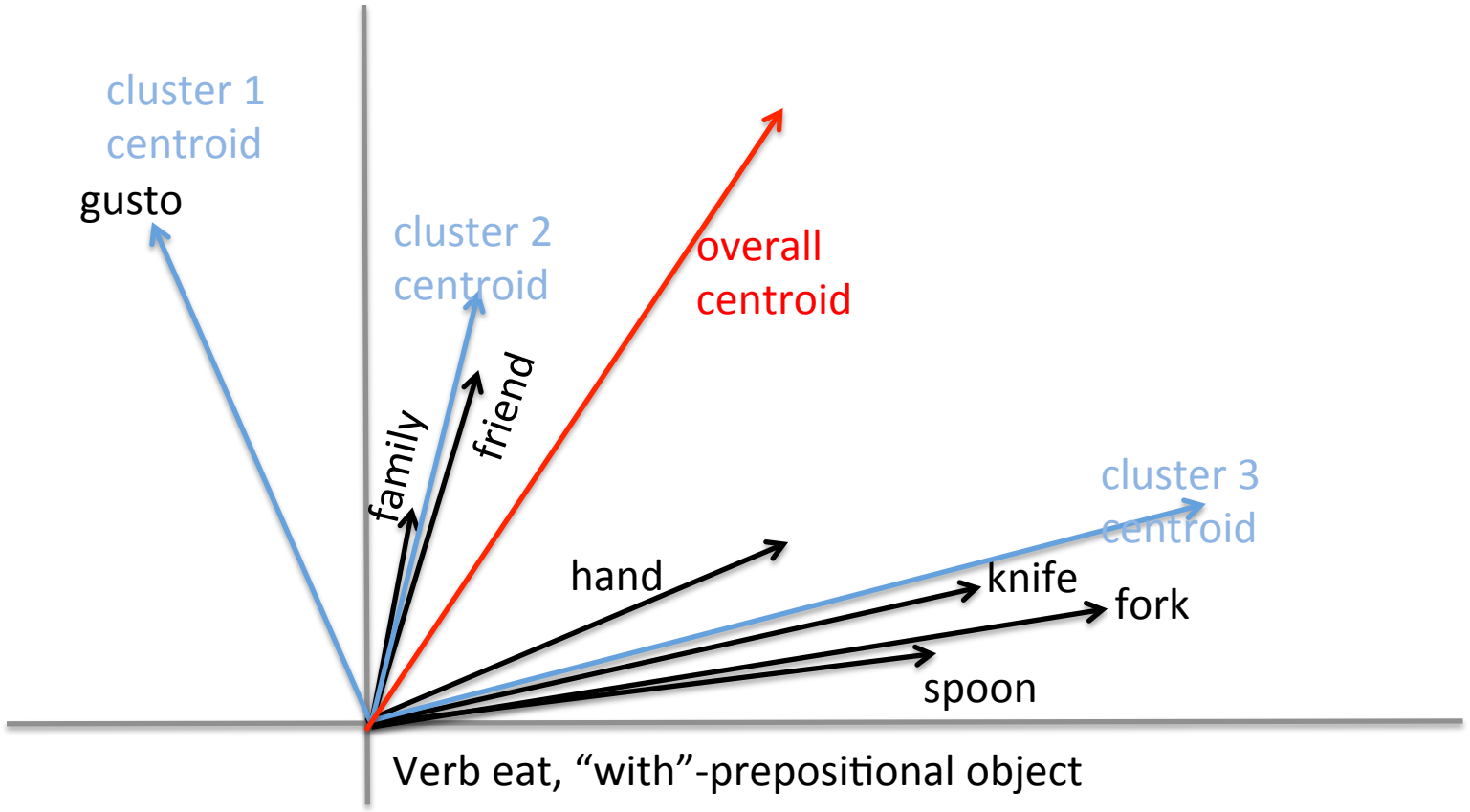
Follow-up ANOVAs

- Good: polysemy (***), frequency (**)
- Bad: polysemy (***), frequency ( )
- POLYSEMOUS: fit (***), frequency (.)
- MONOSEMOUS: fit (***), frequency (***)

Comparing senses



Greenberg, Sayeed, and Demberg (2015)



Overall modelling results

Method	Spearman's rho (TypeDM), range = [-1,1]
Centroid	0.53
OneBest	0.54
kClusters	0.55

Correlation between our experimental human judgements and automatic scores using LMIs from TypeDM, by prototype generation method.

Modelling results by verb type

Method	POLYSEMOUS	MONOSEMOUS
Centroid	0.41	0.66
OneBest	0.45	0.64
kClusters	0.43	0.67

Correlation between our experimental human judgements and automatic scores using LMIs from TypeDM, by prototype generation method and verb type.

The MONOSEMOUS verb “obey”

1. injunction
2. will
3. wish
4. limit
5. equation
6. master
7. law, rule, commandment, principle, regulation, teaching, convention
8. voice, word
9. order, command, instruction, call, summons

The POLYSEMOUS verb “observe”

1. day
2. silence
3. difference, change
4. object, star, bird
5. effect, phenomenon, pattern, behaviour, practice, behavior, reaction, movement, trend
6. rule, custom, law, condition

Conclusions and future work

- Our dataset is available at: http://rollen.mmci.uni-saarland.de/
- It is the first thematic fit dataset to vary polysemy of verbs and frequency of role-fillers systematically.
- We found that polysemy makes good role-fillers not as good and bad role-fillers not as bad.
- The good role-fillers of a more frequent sense get higher ratings.
- We verified the trends in Greenberg, Sayeed, and Demberg (2015).
- Clustering prototypes navigates a trade-off between addressing polysemy and smoothing out noise.
- The next step is a model that successfully integrates sense frequencies.

Thank you!

UNIVERSITÄT
DES
SAARLANDES

Data from this project available at http://rollen.mmci.uni-saarland.de/

DFG

References

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

Davies, M. (2008). The corpus of contemporary american english: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

Fellbaum, C. (1998). *WordNet: an electronic lexical database*. Wiley Online Library.

Greenberg, C., Sayeed, A., and Demberg, V. (2015). Improving unsupervised vectorspace thematic fit evaluation via role-filler prototype clustering. In *Proceedings of the 2015 conference of the North American chapter of the Association for Computational Linguistics - Human Language Technologies*, Denver, USA.

McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3):283-312.

Michel, J., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Brockman, W., The Google Books Team, Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M.A., and Aiden, E.L. (2010) Quantitative Analysis of Culture Using Millions of Digitized Books. *Science.*

Padó, U. (2007). The integration of syntax and semantic plausibility in a wide-coverage model of human sentence processing. PhD thesis, Saarland University.