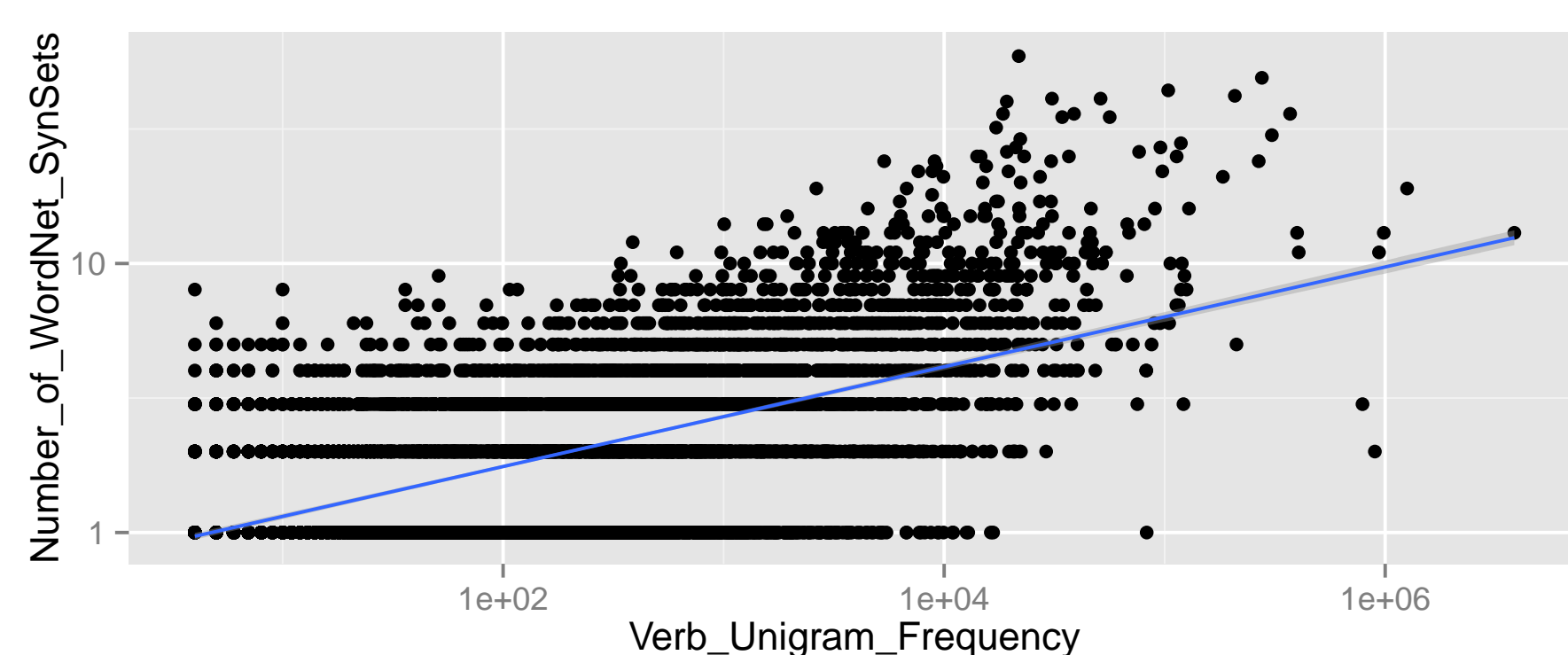


Introduction

Human judgements of *thematic fit* (e.g. how plausible {“croquet”, “soccer”, “harpichord”, “cheese”} are as patients of “play”) are likely sensitive to:

1) verb polysemy



Hypothesis: high polysemy → low thematic fit. Avoid confound: the most polysemous verbs are the most frequent.

2) sense frequency

play_1: participate in games or sport.
“We played hockey all afternoon.”
play_7: perform music on
(a musical instrument).
“He plays the flute.”

WordNet orders SynSets based on their frequencies. Hypothesis: more frequent senses get higher scores.

3) role-filler (noun) frequency



Soccer is most frequent, **croquet** is declining, and **harpichord** is oldest. Hypothesis: does not affect thematic fit.

Project goal:

Obtain human judgements of thematic fit that characterize the effects of verb polysemy, sense frequency, & role-filler frequency.

Stimuli selection

McRae et al. (1997) and others obtained human judgements on a scale from 1 (lowest fit) to 7 (highest fit).

Their question: How common is it for croquet to be played?
Our question: Croquet is something that is played. Agree?

Verb selection

MONOSEMOUS: frequent in COCA, 1 SynSet in WordNet
POLYSEMOUS: matched for COCA freq., many SynSets

Role-filler type

Good for more frequent sense

Frequency

high
low

whip

horse
stallion

punish

criminal
outlaw

Good for less frequent sense
(POLYSEMOUS only)

high
low

cream
frosting

-
-

Bad for all senses
(from shuffle of good nouns)

high
low

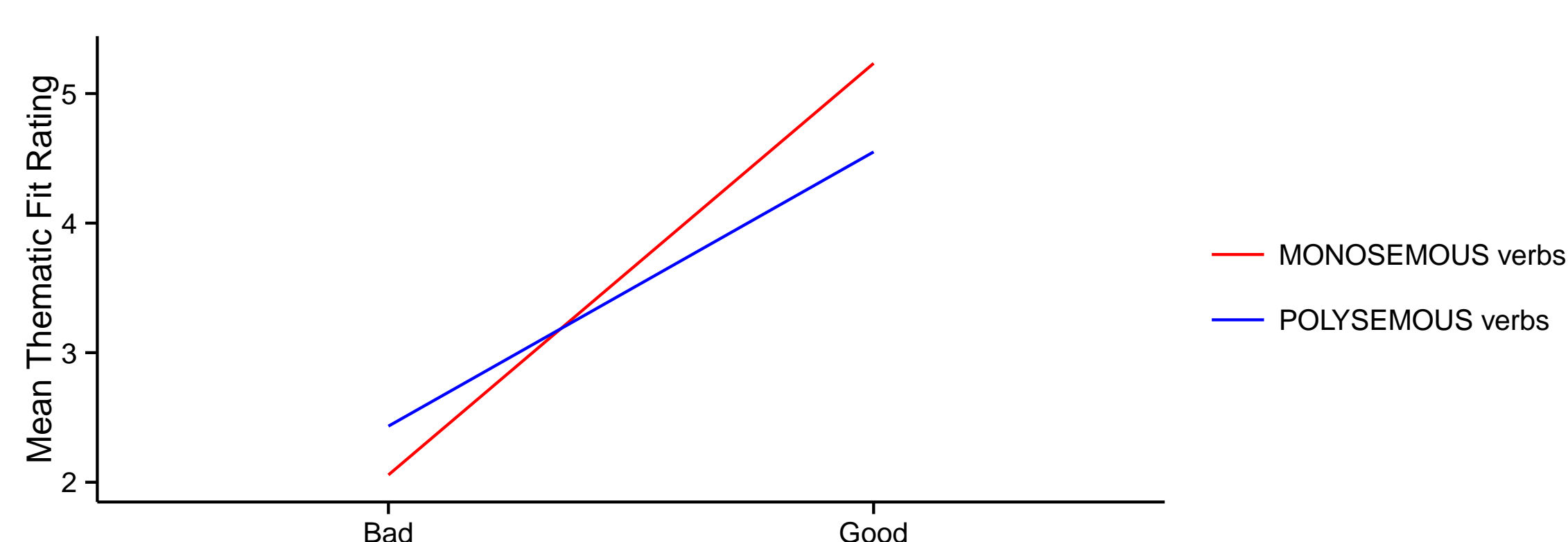
party
gathering

baby
fetus

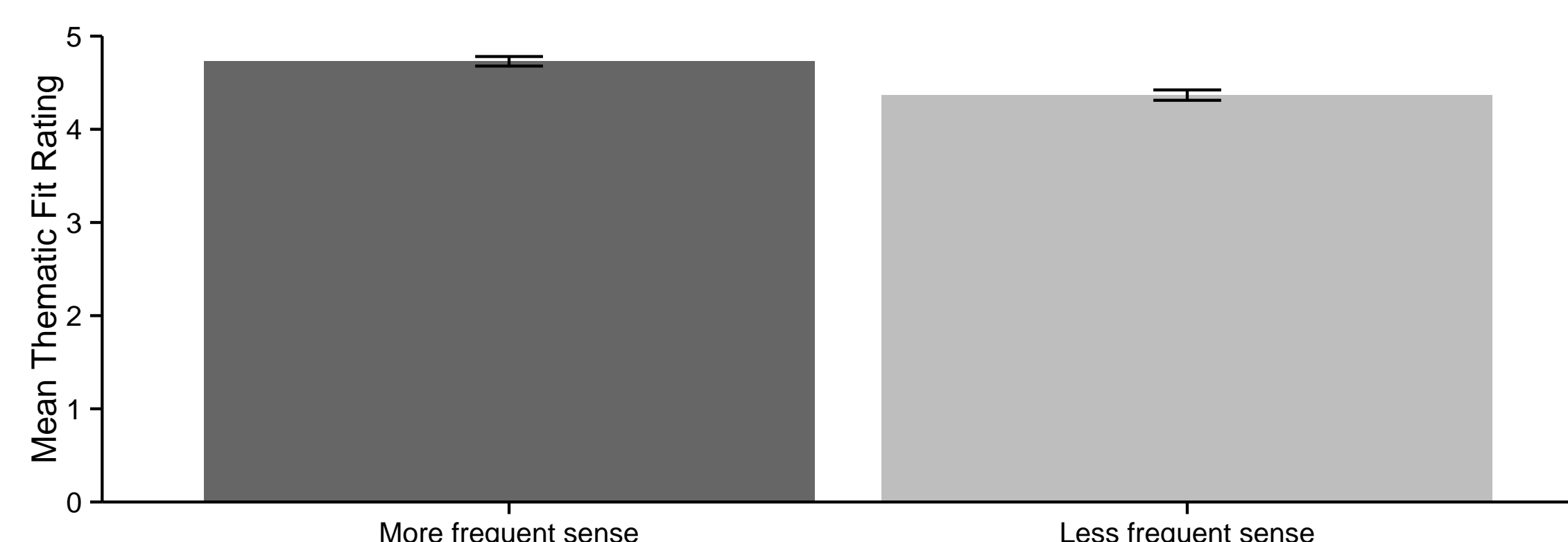
Role-filler selection

- ① To find a good patient-filler, query COCA for: VERB [at*] [nn*]. Example: “whip the horse”
- ② Find a much higher or lower frequency synonym. Example: “horse” (32384) → “stallion” (818)
- ③ For POLYSEMOUS verbs, repeat 1 and 2 for second sense. Example: “whip” → “cream” (19727) → “frosting” (905)

Analysis of human judgements



For POLYSEMOUS verbs, bad patient-fillers were not as bad and good patient-fillers were not as good (**).



Participants rated good patient-fillers for the more frequent sense higher than those for the less frequent sense (**).

Noun frequency had a small effect on ratings of good patient-fillers (**), but not on ratings of bad patient-fillers ().
Noun frequency affected ratings on POLYSEMOUS verbs (.) less than ratings on MONOSEMOUS verbs (**).

Modeling results

	POLYSEMOUS	MONOSEMOUS	FILLERS	ALL
<i>Centroid</i>	0.405	0.655	0.313	0.464
<i>OneBest</i>	0.447	0.641	0.223	0.452
<i>kClusters</i>	0.432	0.669	0.304	0.479

- Spearman’s ρ between human judgements and Greenberg, Sayeed, & Demberg (NAACL, 2015) estimates.
- These confirm that using multiple prototypes (*OneBest*, *kClusters*) is more helpful for POLYSEMOUS verbs than MONOSEMOUS verbs, and that clustering (*kClusters*) is best for mixed datasets.

Conclusions

- We generated the first dataset of thematic fit judgements that systematically manipulates verb polysemy and role-filler frequency, avoiding confounds with verb unigram frequency.
- Modeling results show: multiple prototypes per verb-role help with polysemy, but sense frequency should still be addressed.
- These effects help characterize the nature of linguistic and conceptual material activated by verbs and their arguments.
- See details in our paper in the CMCL 2015 proceedings: <http://aclweb.org/anthology/W/W15/W15-1106.pdf>
- Distributional Memory models and datasets are available at: <http://rollen.mmci.uni-saarland.de>