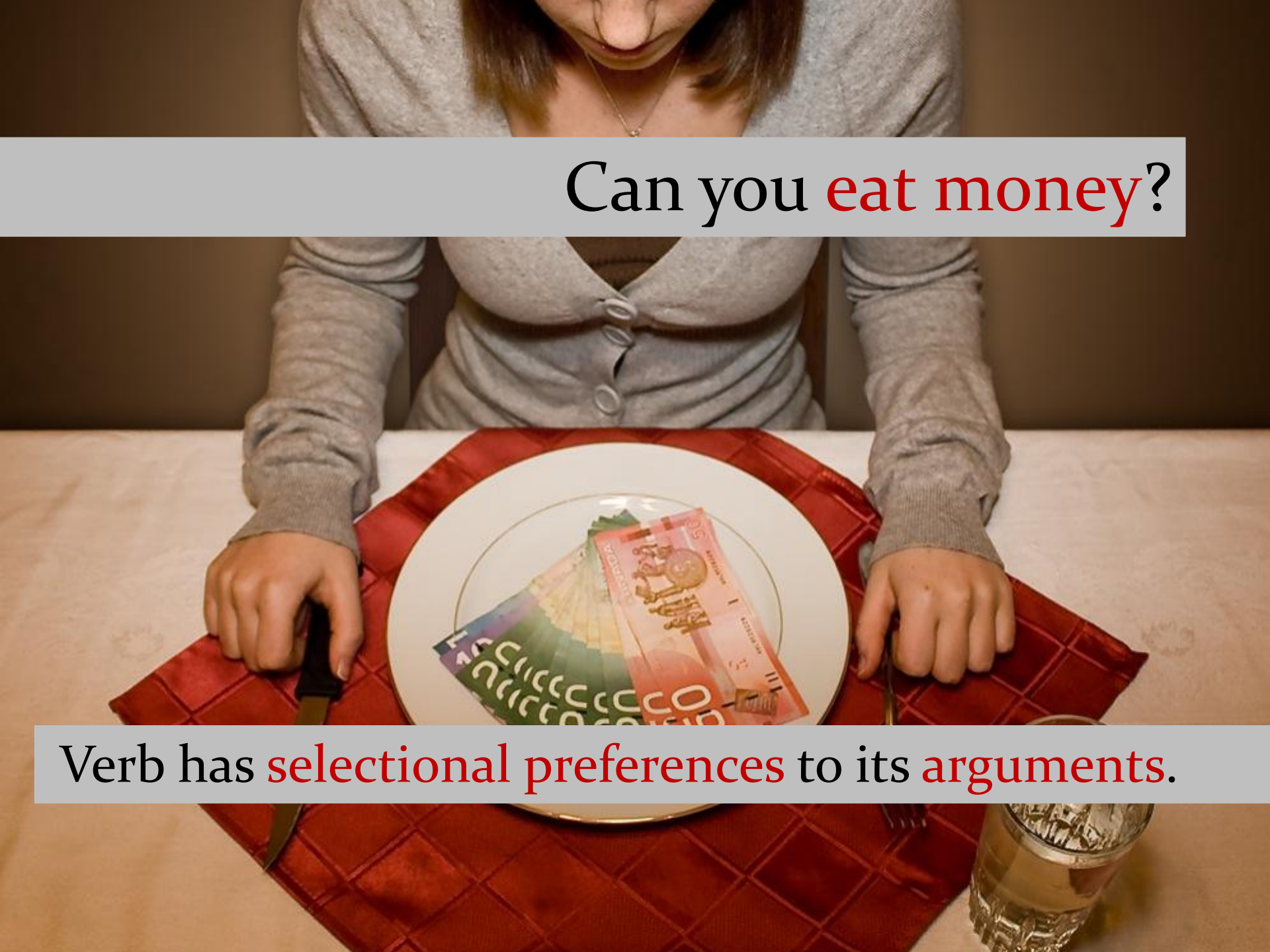
SOCIAL METAPHOR DETECTION VIA TOPICAL ANALYSIS

2013/10/13

Ting-Hao (Kenneth) Huang windx@cmu.edu
Language Technologies Institute, Carnegie Mellon University


Can you eat money?
Verb has selectional preferences to its arguments.

Could we capture **metaphors** in **social media** by selectional preference?

If **not**, why not?
Could **topic model** help ?

If **yes**, how?
Is it for **verb only** ?

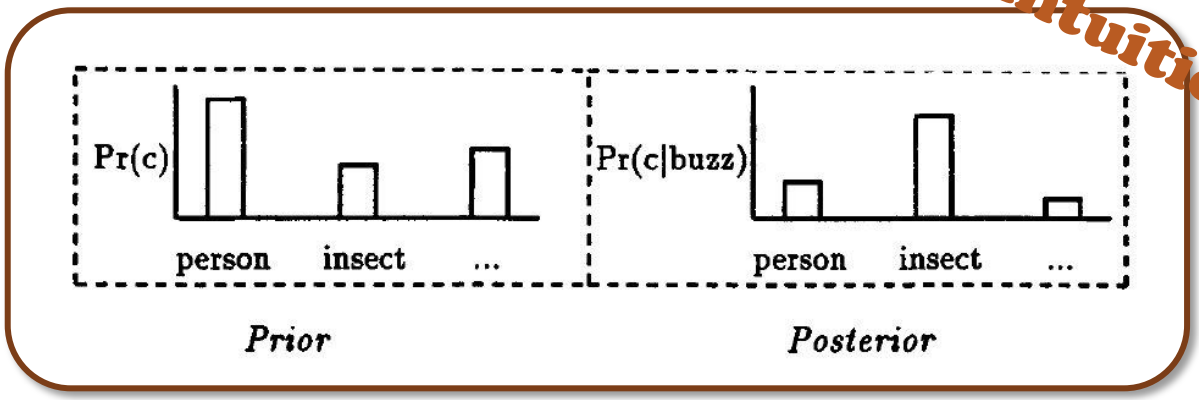
Outline

- Selectional Preference
- 3-Step Framework
  1. Pre-processing
  2. Modeling & Detection
  3. Post-processing
- Topical Analysis
- Experiment & Result
- Conclusion

Selectional Preference

- Selectional Association (SA) (Resnik, 1997)

Intuition!



$$A_R(p,c) = \frac{1}{S_R(p)} \Pr(c|p) \log \frac{\Pr(c|p)}{\Pr(c)}$$

$$S_R(p,c) = D(\Pr(c|p)||\Pr(c))$$

$$= \sum_c \Pr(c|p) \log \frac{\Pr(c|p)}{\Pr(c)}$$

p: predicate
c: noun class

3-step Framework

Pre-processing -
Word Extraction &
Noun Clustering

Modeling & Detection -
SA Outlier Detection

Post-processing -
SA Strength Filter

Step 1: Pre-processing (1)

- Word Extraction
  - Why?
    - Parsing & POS tagging is hard on noisy data
  - How?
    - Using lemma form
    - Set minimal term frequency
    - Set minimal "POS rate"
      - Proportion of occurrence of certain POS
    - Predicates should be more strict than the nouns
      - Noun: TF > 5, POS rate >= 0.7
      - Verb & Adj: TF > 50, POS rate >= 0.8

Step 1: Pre-processing (2)

- Semantic Noun Clustering

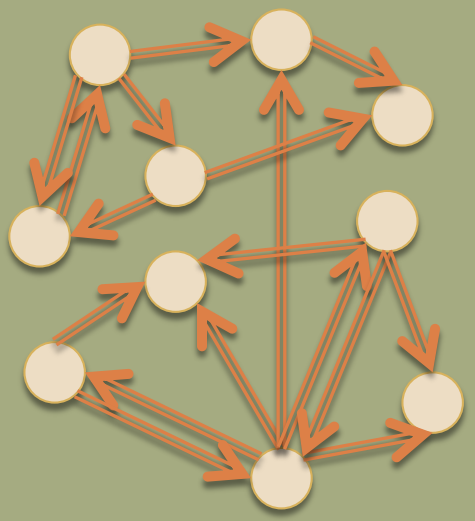
Top 100 Similar Nouns

money:

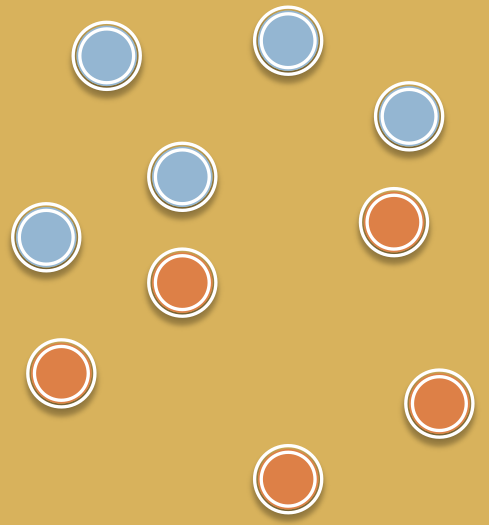
1. funds
2. cash
3. profits
4. millions
5. monies
6. dollars
7. royalties

...

Weighted Directed Graph for Nouns

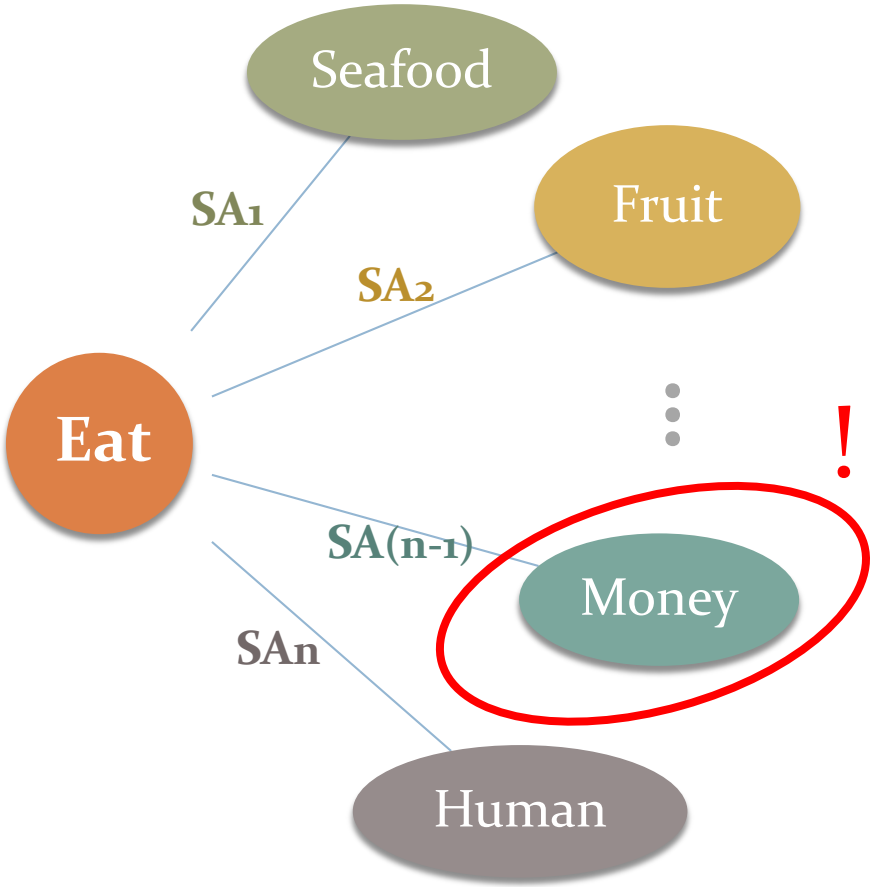


Spectral Clustering



Step 2: Modeling & Detection

- Selectional Association



Another Candidate
Semantic Outlier Word Detection

- "Semantic Coherence" outlier (Inkpen et al., 2005)
- Based on pair-wise word semanic similarity
- Very High False Positive
  - The influences of "general words"
  - Semantic similarity is not reliable

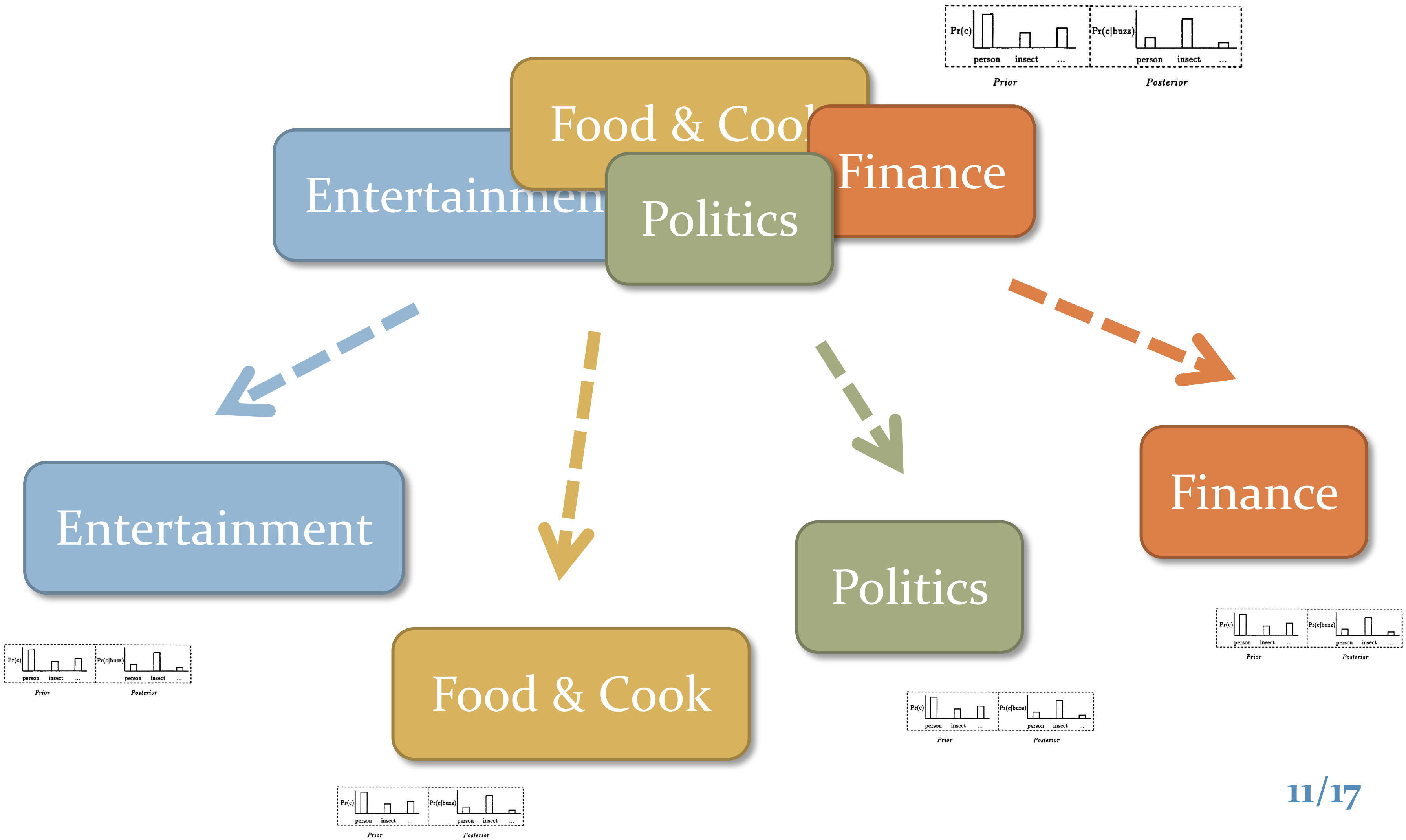
Fail...

Step 3: Post-processing

- SA Strength Filtering (Shutova, et al., 2010)
  - SA Strength
    - Strong (e.g., filmmake)
    - Weak (e.g., "light verb", put, take, …)
  - Predicates with weak selectional preference barely "violates" their own preference.

$$S_R(p,c) = D(\Pr(c|p)||\Pr(c))$$
$$= \sum_c \Pr(c|p) \log \frac{\Pr(c|p)}{\Pr(c)}$$

Topical Analysis



Data

- Online breast cancer support community
- All the public posts from Oct 2001 to Jan 2011.
- 90,242 unique users who posted 1,562,459 messages belonging to 68,158 discussion threads. (Wang, *et al.*, 2012; Wen, *et al.*, 2013)

Experiment Setting

Pre-processing -
- Stanford NLP/Parser
- 55k nouns, 3k adjs, and 1.8k verbs

Modeling & Detection -
- 3 deps: ***nsubj, dobj, amod***
- Observe negative pairs

Post-processing -
- Follow (Shutova, et al., 2010)

Topical Model: JGibbLDA, 20 topics (k = 20)

Result

- Most outliers are **NOT** metaphors
  - Parsing Error
    "…yearly breast MRI…": amod(breast, yearly)
  - Non-metaphor
    "…cancer cells float around in my blood…": dobj(float, cancer)
  - Metonymy
    "If John win tomorrow night, …": dobj(win, tomorrow)
- Only **very few** metaphors are identified
  - "…keep my head occupied …": nsubj(occupy, head)
  - "… my belly has overtaken the boobs …": nsubj(overtake, belly)
- Topic model does **NOT** help much

Could we capture **metaphors** in **social media** by selectional preference?

Maybe not by fully-automatic approaches.

If **not**, why not? Could **topic model** help ?

Good parsing is challenging on social media.
Outliers of SA are not always metaphors.
Topic modeling does not help much.

If **yes**, how? Is it for **verb only** ?

Maybe seed-expansion method works better.
No, it could also work for amod dependency.

Thanks!

- **Acknowledgement**
  - Zi Yang, Prof. Teruko Mitamura, Prof. Eric Nyberg for academic supports, and Yi-Chia Wang , Dong Nguyen for data collection.
  - Supported by the Intelligence Advanced Re-search Projects Activity (IARPA) via Department of Defense US Army Research Laboratory contract number W911NF-12-C-0020.

- **Main References**

  - Resnik, P. (1997). Selectional preference and sense disambiguation. In Proceedings of the ACL'97 SIGLEX Workshop.
  - Shutova, E.; Sun, L.; and Korhonen, A. (2010). Metaphor identification using verb and noun clustering. ACL'10.
  - Wang, Y.-C., Kraut, R. E., & Levine, J. M. (2012). To Stay or Leave? The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups. CSCW'2012.

License