Lexical Information for Coreferent Bridging

Yannick Versley
Sonderforschungsbereich 441
Universität Tübingen
E-mail: versley@sfs.uni-tuebingen.de

ESSLLI 2008 Workshop on Distributional Lexical Semantics

Overview

- Results for (concrete) noun and verb datasets
- Use linguistic intuition and lots of data
  - POS-based patterns
  - *no* parsing
  - *no* unsupervised learning of patterns
  - Sources: UK-WaC (Ferraresi, 2007), Google 1T n-gram
  - Window-based approach with SVD (cf. Rapp, 2003) for comparison

Background Assumptions

- Properties
  - can be used as features to construct a semantic space
  - can be used for taxonomy construction
    $\Rightarrow$ Formal Concept Analysis
  - different sets of (relevant) properties yield different taxonomies/clusterings
- Relational and Functional Concepts; Events
  - yield additional dimensions of meaning
  - make representation more complicated
  - verbs are worst:
    "X killed Y" entails "Y died"
    BUT: "to kill" is an action, "to die" is involuntary

Categorizing Verbs (1)

WordNet (Fellbaum, 1990)

- shallow structure
- multiple entailment relations:
  Temporal Inclusion (sleep-snore, buy-pay)
  Backward Presupposition (succeed-try) ...

FrameNet

- frames are perspective-independent
  ("buy" and "sell" both a *Commercial Transaction*)
- frame fillers $\Rightarrow$ roles

Categorizing Verbs (2)

EVCA (Levin, 1993)/VerbNet (Kipper et al., 2000)

- Defines verb classes based on alternations
  Peter broke the glass / The glass broke
- Linguistically motivated but not language independent
  cf. Sauerland (1994)

Association-based model (e.g. Schulte im Walde, 2008)

- corresponds best to GermaNet
- nouns predominant

Non-Concept approaches to clustering Verbs

- Schulte im Walde and Brew (2002), Gordon and Swanson (2007)
  - Subcategorization Frames
  - Fillers
  - Paths in parse trees
- Stevenson and Merlo (1999), Joanis et al. (2008)
  - Passivization
  - Tense (predominant POS tag)
  - Slot overlap
  - Subject Animacy

Patterns for nouns

- Syntactic (Hindle, 1990; Grefenstette, 1992; Lin, 1998)
  - Grammatical Roles (Subject, Object)
  - Modifying Adjectives
  - compounds (e.g. *soup bowl*, *kitchen knife*)
  - coordination
    note: coordinated terms as features rather than direct evidence
- Ontological Relations
  - subclass
    (Xs and other Ys; Ys such as Xs; see Hearst 1992)
  - part_of
    (the Y's X; the Y of X; e.g. Berland and Charniak 1999)

Patterns for verbs

- Syntactic
  - Grammatical Roles (Subject, Object)
  - Modifying Adverbs
  - Coordination
- Verb-Verb relation patterns (cf. Chklovski and Pantel 2004)
  - Strength: *X* even *Y*
  - Enablement: *X*ed by *Y*ing
  - Sequence: *X* then *Y*

Implementation (1)

Hand-coded pattern extraction:

- faster for one-shot use (but does not use indices)
- slowest part is subsequent `sort | uniq -c`

Home-grown sparse matrix toolkit

- *never* use full matrices
- scales to very large feature space

The devil is in the detail

- always take last NN(PS)
  (*furniture such as dog [kennels]*)
- match verb tenses

Implementation (2)

Generating Feature Vectors

- use posPMI weighting (locally, per relation)
- divide each component by its L_p norm
  ($p = 2$: map each vector to same length; $p = 1.5$ works best)

$\Rightarrow$ Best combination outperforms single patterns

Use Google 1T n-grams
for easily identifiable patterns (Xs and other Ys, . . .)

- no POS tagging
- no L/R context

Single Relations

Nouns:

- noun compounds: E=0.172 P=0.841
  (reverse: E=0.218 P=0.795)
- subjects: E=0.209 P=0.818
- objects: E=0.244 P=0.750
- coordination(and): E=0.241 P=0.750
- possession:
  UK-WaC: E=0.291 P=0.750
  Google 1T n-gram: E=0.211 P=0.818

Verbs:

- adverbs: E=0.342 P=0.622
- subjects: E=0.398 P=0.556
- objects: E=0.441 P=0.511
- then: E=0.424 P=0.533 (reverse: E=0.348 P=0.600)

Single Relations

Nouns:

- noun compounds: E=0.172 P=0.841
  (reverse: E=0.218 P=0.795)
- subjects: E=0.209 P=0.818
- objects: E=0.244 P=0.750
- coordination(and): E=0.241 P=0.750
- possession:
  UK-WaC: E=0.291 P=0.750
  Google 1T n-gram: E=0.211 P=0.818

Verbs:

- adverbs: E=0.342 P=0.622
- subjects: E=0.398 P=0.556
- objects: E=0.441 P=0.511
- then: E=0.424 P=0.533 (reverse: E=0.348 P=0.600)

Single Relations

Nouns:

- noun compounds: E=0.172 P=0.841
  (reverse: E=0.218 P=0.795)
- subjects: E=0.209 P=0.818
- objects: E=0.244 P=0.750
- coordination(and): E=0.241 P=0.750
- possession:
  UK-WaC: E=0.291 P=0.750
  Google 1T n-gram: E=0.211 P=0.818

Verbs:

- adverbs: E=0.342 P=0.622
- subjects: E=0.398 P=0.556
- objects: E=0.441 P=0.511
- then: E=0.424 P=0.533 (reverse: E=0.348 P=0.600)

Single Relations

Nouns:

- noun compounds: E=0.172 P=0.841
  (reverse: E=0.218 P=0.795)
- subjects: E=0.209 P=0.818
- objects: E=0.244 P=0.750
- coordination(and): E=0.241 P=0.750
- possession:
  UK-WaC: E=0.291 P=0.750
  Google 1T n-gram: E=0.211 P=0.818

Verbs:

- adverbs: E=0.342 P=0.622
- subjects: E=0.398 P=0.556
- objects: E=0.441 P=0.511
- then: E=0.424 P=0.533 (reverse: E=0.348 P=0.600)

Single Relations

Nouns:

- noun compounds: E=0.172 P=0.841
  (reverse: E=0.218 P=0.795)
- subjects: E=0.209 P=0.818
- objects: E=0.244 P=0.750
- coordination(and): E=0.241 P=0.750
- possession:
  UK-WaC: E=0.291 P=0.750
  Google 1T n-gram: E=0.211 P=0.818

Verbs:

- adverbs: E=0.342 P=0.622
- subjects: E=0.398 P=0.556
- objects: E=0.441 P=0.511
- then: E=0.424 P=0.533 (reverse: E=0.348 P=0.600)

Window-based Baseline

Use Google 1T n-grams (10^{12} words)

- huge (Lots More Data!!!)
- no POS tagging

Initial results

- Straight Window-based approach slightly worse than best single pattern
  noun compounds: E=0.172; 1W window: E=0.177
  adverbs: E=0.342; 1W window: E=0.376

Window-based Baseline

Use Google 1T n-grams (10^{12} words)

- huge (Lots More Data!!!)
- no POS tagging

Initial results

- Straight Window-based approach slightly worse than best single pattern
  noun compounds: E=0.172; 1W window: E=0.177
  adverbs: E=0.342; 1W window: E=0.376

$\Rightarrow$ try SVD with various weighting methods
(posPMI, Log, LogEnt)
use $\approx$2000 most frequent non-ambiguous nouns/verbs

SVD results

Results get actually worse!
What happened?

SVD results

Results get actually worse!

What happened?

v0: λ=56595		v1:λ=2043.5	
fundraise	*0.0000	ensure	*-9999.99
exhilarate	*Reserved	determine	Verzeichnis
socialize	*Advertise	process	-99
pend	*Cart	identify	-999
v2: λ=2028.7		**v3:λ=1760.5**	
f-ck	a-s	configure	src
suck	p-ssy	filter	header
*amend	*pursuant	*accuse	*father
*comply	*Agreement	*murder	*whom

SVD results

Results get actually worse!

What happened?

v0: λ=56595		v1:λ=2043.5	
fundraise	*0.0000	ensure	*-9999.99
exhilarate	*Reserved	determine	Verzeichnis
socialize	*Advertise	process	-99
pend	*Cart	identify	-999
v2: λ=2028.7		**v3:λ=1760.5**	
f-ck	a-s	configure	src
suck	p-ssy	filter	header
*amend	*pursuant	*accuse	*father
*comply	*Agreement	*murder	*whom

Yay! Web Genres!!!

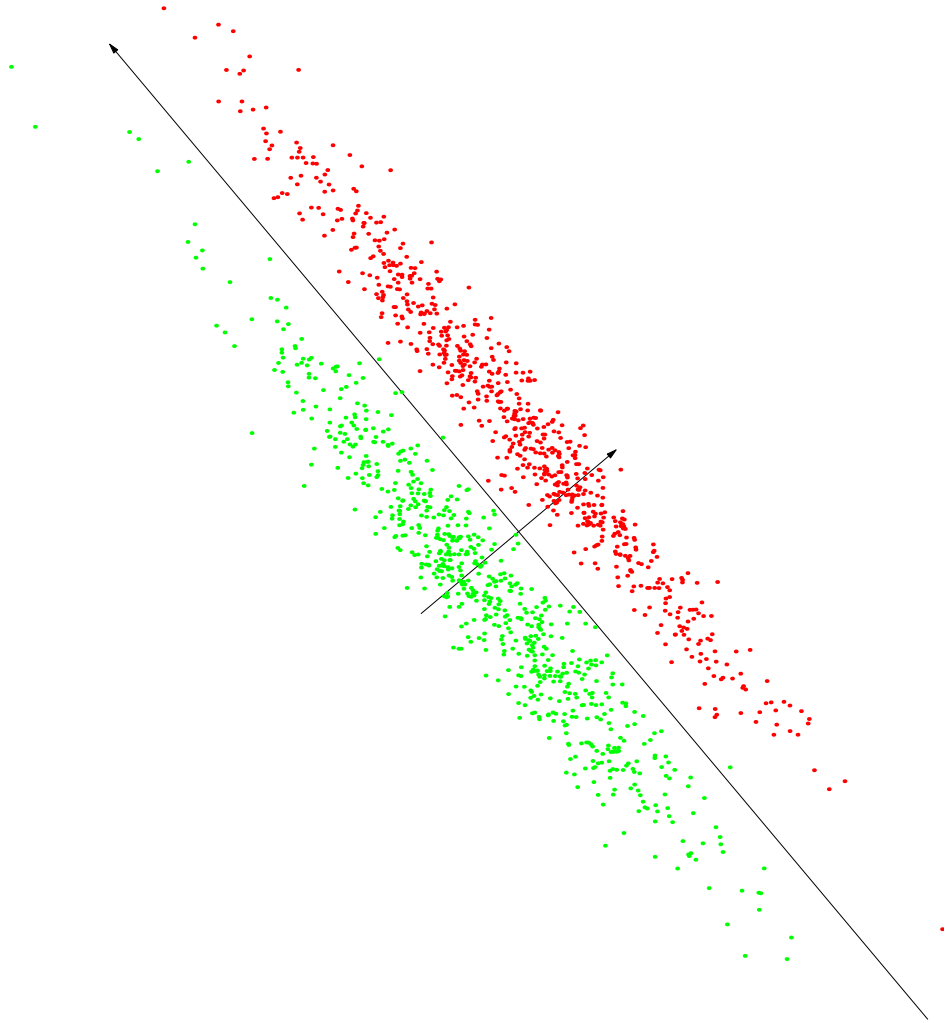
But we want lexical semantics!?

Fixing domain sensibility: Whitening

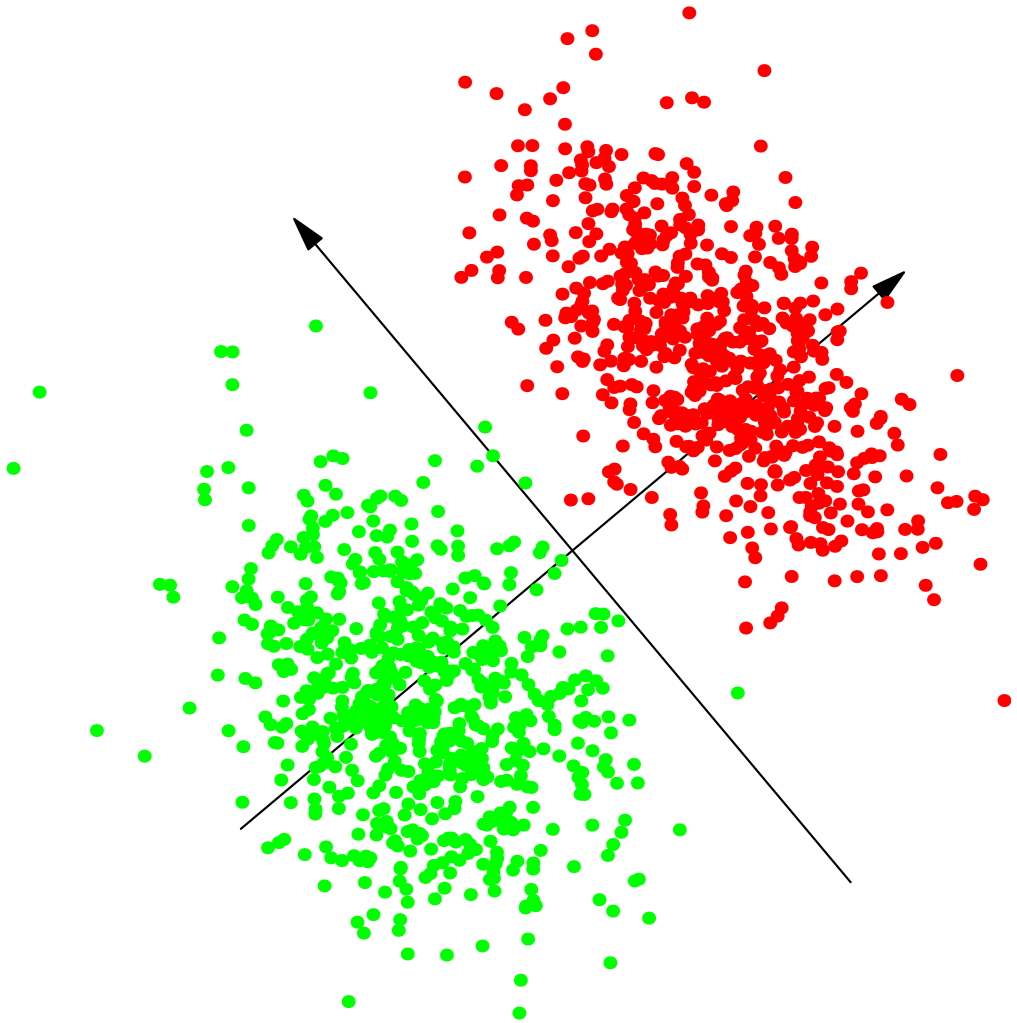
Decorrelation/Whitening = Set all singular values to 1

- standard "tool" from PCA literature
- evens out influence of dominating singular vectors
- k-means also works better with globular clusters

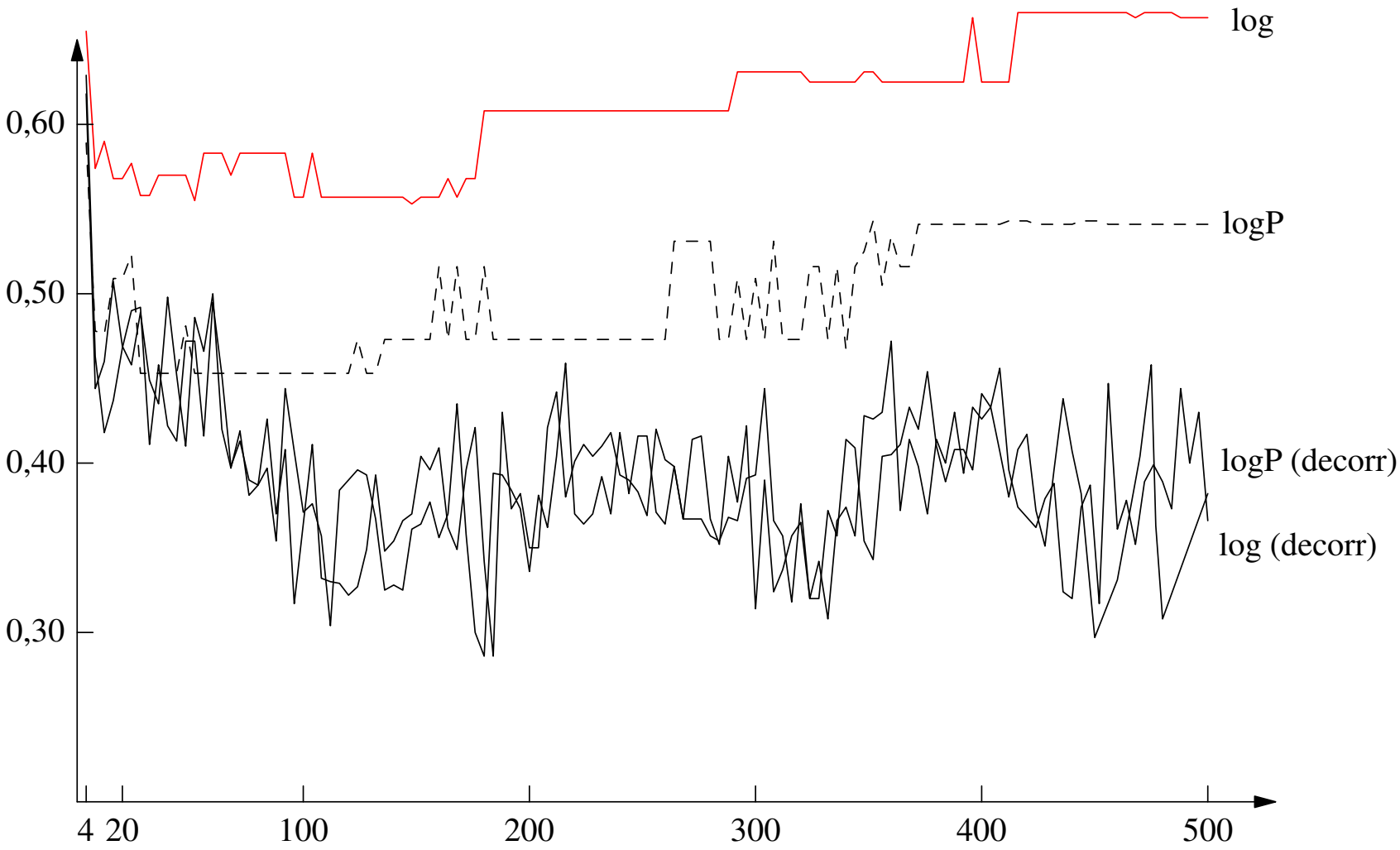
Whitening: Illustration



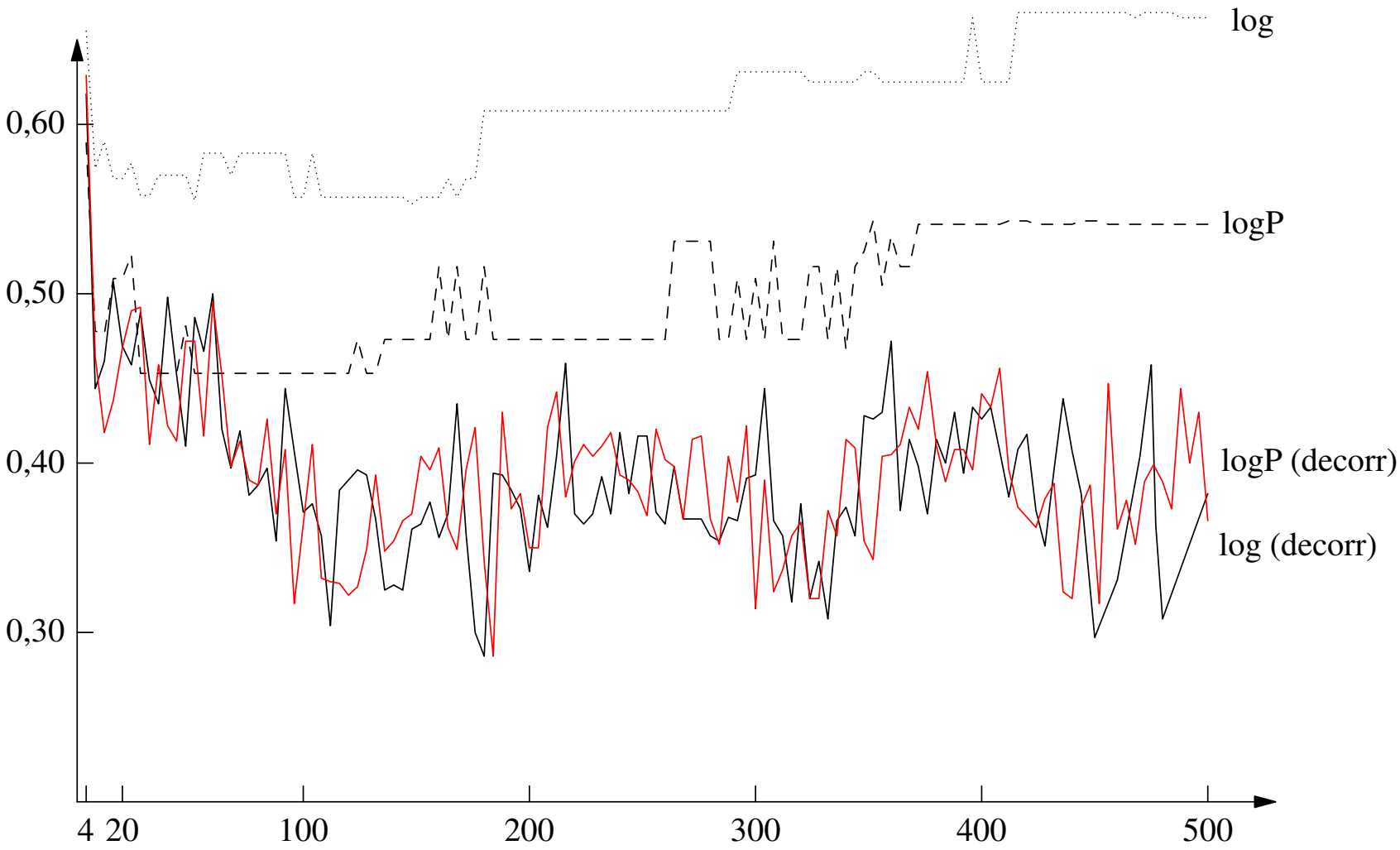
Whitening: Illustration



Influence of Decorrelation



Influence of Decorrelation



Results Summary

- Good noun clusters (subject, adjectives, and-other)
  - "chicken" as a vegetable
  - tools (chisel/scissors) vs. other (kettle/telephone) artifacts
- OK verb clusters (and, then^{-1})
  (caveat: need well-defined criteria)
  - "fall" in break/destroy/die cluster
  - "smile" in feel/look/notice cluster
  - move,forget with arrive/enter/leave
  - check/evaluate with request/suggest, repair
  - know/remember with read/speak/talk, listen

Further work

- Try multiple datasets
- Are results comparable across languages?
  - Homonymy/Polysemy
  - Surface distance corresponds to different things
  - Many shallow patterns only work for configurational languages (English, French, Italian) but not for free(r)-word order languages (German, Dutch, Arabic, . . .)
    $\Rightarrow$ need fast (and reasonably good) parsing
- Influence of clustering method
- Use (semi-)supervised training
  (Baroni and Lenci, 2008; Snow et al., 2005)

Thanks for listening!

THE END

Baroni, M. and Lenci, A. (2008). Concepts and word spaces. *Italian Journal of Linguistics*, to appear.

Berland, M. and Charniak, E. (1999). Finding parts in very large corpora. In *Proceedings of ACL-1999*.

Chklovski, T. and Pantel, P. (2004). VerbOcean: Mining the web for fine-grained semantic verb relations. In *Proc. EMNLP 2004*.

Fellbaum, C. (1990). English verbs as a semantic net. *International Journal of Lexicography*, 3(4):278–301.

Ferraresi, A. (2007). Building a very large corpus of English obtained by Web crawling: ukWaC. Master's thesis, Università di Bologna.

Gordon, A. S. and Swanson, R. (2007). Generalizing semantic role annotations across syntactically similar verbs. In *ACL 2007*.

Grefenstette, G. (1992). Sextant: Exploring unexplored contexts for semantic extraction from syntactic analysis. In *ACL Student Session 1992*.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proc. of the 14th International Conference on Computational Linguistics (COLING 92)*.

Hindle, D. (1990). Noun classification from predicate argument structures. In *Proceedings of the 28th annual meeting of the Association for Computational Linguistics*.

Joanis, E., Stevenson, S., and James, D. (2008). A general feature space for automatic verb classification. *Natural Language Engineering*, 14(3):337–367.

Kipper, K., Dang, H. T., and Palmer, M. (2000). Class-based construction of a verb lexicon. In *AAAI-2000*.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.

Lin, D. (1998). Automatic retrieval and clustering of similar words. In *Proc. CoLing/ACL 1998*.

Rapp, R. (2003). Word sense discovery based on sense descriptor dissimilarity. In *Proc. Ninth Machine Translation Summit*.

Sauerland, U. (1994). German diathesis and verb morphology. In *Verb Classes and Alternations in Bangla, German, English and Korean*, number 1517 in A.I. Memo. Massachusetts Institute of Technology.

Schulte im Walde, S. (2008). Human associations and the choice of features for semantic verb classification. *Research on Language and Computation*, to appear.

Schulte im Walde, S. and Brew, C. (2002). Inducing german semantic verb classes from purely syntactic subcategorization information. In *Proc. ACL 2002*.

Snow, R., Jurafsky, D., and Ng, A. Y. (2005). Learning syntactic patterns for automatic hypernym discovery. In *NIPS 2005*.

Stevenson, S. and Merlo, P. (1999). Automatic verb classification using grammatical features. In *Proc. EACL 1999*.