Empirical NLP in a Cognitive Perspective: The Case of Word Meaning

Marco Baroni

CIMeC (University of Trento)

November 22, 2006

Outline

Introduction

Empirical NLP
- Supervised learning
- Unsupervised learning

Word Meaning
- Introduction
- Meaning as context
- Similarity in word space
- Dimensionality reduction

A Few Experiments
- Introduction
- Experimental setting
- Results

Current Directions

The "empirical turn" in NLP

- Early (but not earliest) NLP-work focused on development of (expert) hand-tuned systems that privileged "depth" over "width"
- Problems (of accuracy, efficiency and cost) to scale up to real life tasks
- From mid-eighties, "empirical" turn towards inductive methods
- Statistical generalizations extracted from *corpora*, i.e., large data-bases of text/transcribed speech produced in natural settings
- Famously, work at the IBM Watson Labs in speech recognition and machine translation
- Also: British lexicography, in particular COBUILD group from mid-seventies

Why?

- Classic sentences (computational) linguists focused on:

  `Every farmer that owns a donkey beats it.`

  `The idea that the idea suffices suffices.`

  `Who did you think that Mary said bought the present for John?`
- Real text:

  `Branch Prediction Analysis is a recent attack vector against RSA public-key cryptography`

  `Simon P. Chappell's review`

  `(Linux.com and Slashdot are both part of OSTG.)`

Corpus typology and some famous corpora

- Balanced, representative, ‘reference’ corpora: Brown/LOB (1M tokens), COBUILD (10M, ...), BNC (100M)
- Opportunistic: WSJ, la Repubblica-SSLMIT, Gigaword (1B)
- Web-derived corpora (WaCky project: 1.65B tokens of German, 1.9B tokens of Italian)
- Parallel: Hansard, OPUS, EuroParl
- Specialized, comparable, diachronic...
- More data is better data?
- According to Anonymous (p.c.), we hear/read/speak/write about 40M words per year, i.e., a 1B word corpus corresponds to more than 20 years of linguistic experience

Empirical NLP and human language competence

- Creation of large scale resources useful for investigations of human language (in particular: annotated corpora and lexical databases with frequency information)
- Computer seen as statistics-driven agent that “learns” from its environment: can it teach us something about human learning?
- Convergence with probabilistic models of cognition (see, e.g., *Trends in Cognitive Sciences* July 2006 issue)
- Emphasis on technology and practical applications often leads to “non-cognitively-plausible” choices

Outline

Introduction

Empirical NLP
- Supervised learning
- Unsupervised learning

Word Meaning
- Introduction
- Meaning as context
- Similarity in word space
- Dimensionality reduction

A Few Experiments
- Introduction
- Experimental setting
- Results

Current Directions

The typical setting in empirical NLP

- Large amount of input data (*training corpus*)
- Program extracts statistical generalizations from training corpus
- “Trained” program used for classification task on new input

Supervised learning

- Training corpus is (manually) annotated with target classification
- System learns model that predicts target classification from context features
- Model is used to predict classification from context features in unannotated text

Part-of-Speech Tagging

- Training corpus:

  `The/ART dog/NOUN barks/VERB`
- Features:
  - Tag of word, word to left, to right
  - Last N characters of target word
  - ...
- Machine-learning method: rule extraction, Markov models, maximum entropy...
- Text to be tagged:

  `The wug barks`

Supervised learning: A success story

- Current best accuracy in tagging English words: over 97% (Toutanova et al. 2003)
- Accuracy of 95% and above routinely achieved in Italian and other European languages
- Many other successful applications, ranging from text categorization (also by style!) to verbal Aktionsart semantics acquisition

Supervised learning: Problems

- "Resource bottleneck": you need plenty of training data
- Not a plausible model of how humans acquire language

Unsupervised learning

- Training corpus does not contain manually annotated examples of the information we want to extract (although it might contain *some* annotation: e.g., POS tags)
- Mostly used in lexical extraction studies, where task is to extract ranked lists of words or other units, or to cluster such units
- Examples:
  - Collocation/mwe/term extraction (Evert 2004)
  - Morpheme induction (Goldsmith 2001, Baroni 2003)
  - Bilingual lexicon induction (Tiedemann 2003)
  - Semantic similarity models (Sahlgren 2006 and references there)

Unsupervised learning: pros and cons

- Pros:
  - No resource bottleneck
  - Cognitive angle more plausible: children do perform some kind of unsupervised learning
    - At the very least, success of unsupervised algorithm shows that cues are in input, independently of learning procedure
    - “Macro” simulations of learning over spans of years
- Cons:
  - Resources still needed for evaluation and training (e.g., unsupervised learning of verbal semantics from *parsed* corpus)
  - Performance generally lower than for supervised algorithms

Middle grounds

- “Weakly supervised” learning, bootstrapping, co-training. . . :
  1. Start with a small set of annotated examples
  2. Train model on current annotated set
  3. Use trained model to annotate more data
  4. Go back to 2., repeat as needed
- Famously used by Yarowsky (1995) for word sense disambiguation
- Start with small set of collocates of sense A and sense B; iteratively bootstrap larger set

Outline

Introduction

Empirical NLP
- Supervised learning
- Unsupervised learning

Word Meaning
- Introduction
- Meaning as context
- Similarity in word space
- Dimensionality reduction

A Few Experiments
- Introduction
- Experimental setting
- Results

Current Directions

Why?

- A promising area to build a bridge between NLP and cognitive (neuro-)science
- Large amount of work in NLP (and psychology) on word meaning and semantic similarity
- We don't know the "right answer": exploratory character of analysis
- NLP models have some flavor of cognitive plausibility:
  1. Prevalence of unsupervised algorithms
  2. Contextual view of meaning
  3. (Lexical) semantics as network of similarities in context-based multi-dimensional space
  4. (Incremental algorithms)

The Word Space Model

E.g., Sahlgren 2006

- Meaning of words defined by *set of contexts* in which word occurs
- Similarity of words represented as *geometric distance* among *context vectors*
- *Dimensionality reduction* of context vectors

Contextual view of meaning

	leash	walk	run	owner	pet
dog	3	5	2	5	3
cat	0	3	3	2	3
lion	0	3	2	0	1
light	0	0	0	0	0
bark	1	0	0	2	1
car	0	0	1	3	0

Contextual view meaning

Theoretical background

- "You should tell a word by the company it keeps" (Firth 1957)
- "[T]he semantic properties of a lexical item are fully reflected in appropriate aspects of the relations it contracts with actual and potential contexts [...] [T]here are good reasons for a principled limitation to linguistic contexts" (Cruse 1986)
- "Because the capsule was hermetically broamed, . . . " (McDonald and Ramscar 2001)

The contextual approach to meaning
Psycholinguistic evidence

- Relatively large literature (see refs. in Sahlgren 2006, and in particular work by Scott McDonald)
- For example, McDonald and Ramscar (2001) found that human similarity ratings on marginally familiar words and nonce forms are biased by collocations (extracted from corpus with statistical methods)
- there was a *balak* dispensing tea → *urn*
- boiled water in the *balak* → *kettle*

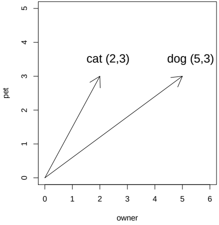
Which context?

- Documents/large textual spans
- All words in a narrow window
- Lemmatized content words in a narrow window
- Content words in specific syntactic constructions
- Larger contexts tend to capture "topical" similarity, narrower contexts "ontological" or "collocational" similarity
- More "linguistics" does not imply better results:
  - Data sparseness problems
  - A bad analysis might be more damaging than no analysis at all
- Context needs not be linguistic! Vectors could include, e.g., co-occurrence counts with sensory stimuli

Counting occurrence-in-context

- 0/1 vector reporting if word occurs in context or not (typical for document contexts)
- Frequencies of co-occurrence of target and context
- Weighted frequencies of co-occurrence, e.g.:
  - Giving more weight to nearer contexts
  - Giving more weight to unexpected co-occurrences

Similarity in word space



The cosine measure

- Cosine of angle between vectors: decreases with angle size, 0 for orthogonal vectors

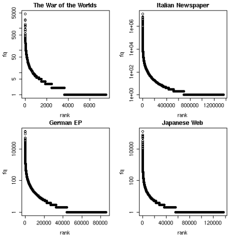
$$\cos(\overrightarrow{x}, \overrightarrow{y}) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

- Same as *correlation coefficient* (of normalized vectors)
- Intuition:
  - Numerator: will be larger if x_i tends to be high when y_i is also high
  - Denominator has normalizing function

Semantic similarity via the cosine measure

- Continuous notion of similarity
- Words will have "some" degree of similarity with a large set of other words
- Single measure of similarity, multiple ways to be similar

Dimensionality reduction: why?



Dimensionality reduction: why?

- Type richness, rare occurrence of most types
- Vectors with millions of dimensions
- Most elements will be zero

(Reduced) Singular Value Decomposition

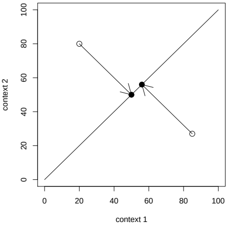
Aka Latent Semantic Analysis (Landauer et al. 1998)

- Linear algebra techniques to find “best” k-dimensional approximation to a m-dimensional matrix, where $k << m$ (*best* in a least-squares error sense)
- Context vectors of thousands (or millions) of dimensions are reduced to vectors of hundreds of dimensions
- The dimensions in the reduced space are those that account for larger proportion of variation in original data
- Cosine similarity measured in reduced space
- Cf. Principal Component Analysis, Factor Analysis
- For the math see, e.g., Manning and Schütze (1999)

Dimensionality reduction as generalization

- Contexts with similar co-occurrence patterns likely to be collapsed onto same dimension in reduced space
- Accounts for “synonymic contexts”
- E.g., occurring near *spaceman* or near *astronaut* should count as essentially the same thing
- Should we look for a semantics of the new dimensions?

Dimension reduction as generalization



Packages implementing SVD

- Infomap: `http://infomap-nlp.sourceforge.net/`
- SenseClusters: `http://senseclusters.sourceforge.net/`
- Good and bad of out-of-the-box solutions
- Efficient implementation not trivial

SVD takes the TOEFL

- TOEFL synonym match task
- Target: *levied*; Candidates: *imposed, believed, requested, correlated*

SVD takes the TOEFL (cont.)

- Performance on TOEFL synonym match task
- Average foreign test taker: 64.5%
- SVD trained on lemmatized BNC (Rapp 2003 quoted in Sahlgren 2006): 92.5%
- See Landauer et al. (1998) for more SVD-based simulations of aspects of human knowledge

Problems with SVD

- Inefficient:
  - You have to build the full matrix first
  - SVD on very large matrices is computationally expensive
- Implementation tricky
- (These factors matter even if you are not an engineer: you can't experiment much if you must rely on black-box solution that takes months to run each experiment)
- Non-incremental: first build full matrix, then reduce (bad as a model of human learning)

The low-cost alternative: Random Indexing

Sahlgren 2005

- Represent each context element with a (low-dimensional) index of randomly assigned 1, -1 and (mostly) 0:

```
pet     0  -1   0  0
owner   1   0   0  0
leash  -1   0  -1  0
```

- As you go through corpus, add random index corresponding to each context to target word contextual vector:

```
                         dog  0   0   0  0
dog is a pet        -->  dog  0  -1   0  0
owner of the dog    -->  dog  1  -1   0  0
dog on the leash    -->  dog  0  -1  -1  0
```

- Cosine similarity computed on resulting contextual vectors

Pros and cons

- Pros:
  - Very efficient: low dimensionality from the beginning to the end
  - Implementation trivial (assign random values to vector, sum vectors)
  - Incremental: at any stage, target vectors constitute low-dimensional semantic space
- Cons:
  - No latent semantic space effect: contexts are “squashed” randomly
  - Lower accuracy, at least on some tasks (Gorman and Curran 2006)

Outline

Introduction

Empirical NLP
- Supervised learning
- Unsupervised learning

Word Meaning
- Introduction
- Meaning as context
- Similarity in word space
- Dimensionality reduction

A Few Experiments
- Introduction
- Experimental setting
- Results

Current Directions

Typical evaluation

- Quantitative, often large scale
- Relies on existing lexical resource (e.g., WordNet, Moby thesaurus, TOEFL synonym list) as gold standard
- Problems:
  - As much an evaluation of lexical resource as of semantic similarity models (e.g., WordNet has *escargot* but not *steak* as neighbor of *sausage*)
  - Each resource will highlight specific aspects of semantic similarity (e.g., “ontological” vs. ”collocational” similarity): evaluation of suitability for a certain task
  - Differences among different classes of words will cancel out in large scale evaluation

Qualitative analysis of nearest neighbors

- Choose a variety of words across range of distributional and semantic properties
- Analyze *nearest neighbor set* of each word (set of words with highest cosine similarity to target word)
- In current pilot study, 10 nearest neighbors
- NB: absolute cosine value ignored!
- This is bad – consider, e.g., *svd-narrow* data:
  - Nearest neighbor of *red* is *yellow* with cosine > 0.79
  - Nearest neighbor of *minister* is *bin* with cosine < 0.64

Input

- BNC
- POS-tagged and lemmatized
- Only nouns, verbs and adjectives kept

```
This virus affects the body's defence system
so that it can not fight infection.

virus-nn affect-vv body-nn defence-nn system-nn
fight-vv infection-nn
```

SVD models

- SVD as implemented in Infomap
- 20,000 most frequent input words as targets (rows)
- 2,000 most frequent input words as contexts (raw columns)
- Reduction to 300 dimensions
- "Narrow" contexts: 2 words to left and right, and not across sentence boundaries
- "Wide: contexts: in theory, 50 words to left and right; because of a bug in my implementation, more like 30 words to left and right

RI models

- 20,000 most frequent input words as targets (rows)
- 10,000 most frequent input words as contexts (raw columns)
- Reduction to 1800 dimensions
- "Narrow" and "wide" contexts: same as for SVD models
- For nouns only, "relational" contexts based on regular expressions that look for syntactically meaningful verbal, adjectival and nominal collocates

Test items

item	*pos*	*freq*	*item*	*pos*	*freq*
good	adj	124,236	minister	noun	30,146
bad	adj	19,576	sausage	noun	964
red	adj	12,191	starfish	noun	102
mauve	adj	182	cost	noun	26,937
sing	verb	5,994	sorrow	noun	657
hesitate	verb	2,027	favouritism	noun	102
annotate	verb	94			

Neighbor classification

- Constructed *a posteriori*
- Interpretation of categories changes with test class
- Preliminary, plenty of arbitrary choices!

syn	(near) synonym
syn-scale	same side on graded scale
anto	antonym
anto-scale	opposite side on graded scale
hyp	co-hyponym or direct hyper-/hyponym
deriv	morphological variant
colloc	typical syntagmatic neighbor
sem	“family resemblance”
out	no close relation

good-aj

svd-narrow		*svd-wide*	
bad-aj	anto	bad-aj	anto
excellent-aj	syn-scale	decent-aj	syn-scale
improved-aj	syn-scale	excellent-aj	syn-scale
improved-vv	syn-scale	give-vv	out
improving-vv	syn-scale	honest-aj	syn-scale
perfect-aj	syn-scale	impressed-aj	syn-scale
poor-aj	anto-scale	improved-vv	syn-scale
satisfactory-aj	syn-scale	like-vv	syn-scale
superb-aj	syn-scale	nice-aj	syn-scale
wonderful-aj	syn-scale	well-aj	syn-scale

ri-narrow		*ri-wide*	
able-aj	syn-scale	expect-vv	out
come-vv	colloc	happen-vv	out
going-vv	out	keep-vv	out
like-vv	syn-scale	only-aj	out
marvellous-aj	syn-scale	put-vv	out
mean-vv	colloc	rest-nn	colloc
suppose-vv	out	start-vv	out
want-vv	syn-scale	taking-vv	out
wanted-vv	syn-scale	try-vv	out
wonderful-aj	syn-scale	whole-aj	syn-scale

bad-aj

svd-narrow		*svd-wide*	
appalling-aj	syn-scale	appalling-aj	syn-scale
awful-aj	syn-scale	awful-aj	syn-scale
blame-vv	syn-scale	blame-vv	syn-scale
dreadful-aj	syn-scale	dreadful-aj	syn-scale
freak-aj	syn-scale	good-aj	anto
good-aj	anto	nasty-aj	syn-scale
nasty-aj	syn-scale	terrible-aj	syn-scale
poor-aj	syn-scale	trouble-nn	colloc
terrible-aj	syn-scale	worrying-aj	syn-scale
worse-aj	syn-scale	worse-aj	syn-scale

ri-narrow		*ri-wide*	
happen-vv	out	happen-vv	out
idea-nn	colloc	keep-vv	out
matter-vv	out	put-vv	out
mean-vv	colloc	start-vv	out
reckon-vv	out	supposed-aj	out
suppose-vv	out	trouble-nn	sem
terrible-aj	syn-scale	try-vv	out
think-vv	colloc	well-aj	anto-scale
thought-vv	colloc	worse-aj	syn-scale
worse-aj	syn-scale	wrong-aj	syn-scale

red-aj

svd-narrow		*svd-wide*	
blue-aj	hyp	blue-aj	hyp
brown-aj	hyp	brown-aj	hyp
grey-aj	hyp	grey-aj	hyp
pink-aj	hyp	orange-aj	hyp
purple-aj	hyp	pale-aj	hyp
red-nn	deriv	pink-aj	hyp
scarlet-aj	hyp	purple-aj	hyp
shiny-aj	hyp	red-nn	deriv
white-aj	hyp	white-aj	hyp
yellow-aj	hyp	yellow-aj	hyp

ri-narrow		*ri-wide*	
black-aj	hyp	black-aj	hyp
bright-aj	hyp	blue-aj	hyp
colour-aj	hyp	bright-aj	hyp
dark-aj	hyp	green-aj	hyp
dark-nn	hyp	grey-aj	hyp
grey-aj	hyp	sight-nn	sem
palea-aj	hyp	thin-aj	out
pink-aj	hyp	tiny-aj	out
red-nn	deriv	white-aj	hyp
yellow-aj	hyp	yellow-aj	hyp

mauve-aj

svd-narrow		svd-wide	
blue-aj	hyp	flower-nn	colloc
emerald-aj	hyp	foliage-nn	colloc
flowered-aj	hyp	hyacinth-nn	colloc
pink-aj	hyp	pink-aj	hyp
purple-aj	hyp	pink-nn	sem
red-aj	hyp	purple-aj	hyp
red-nn	hyp	scarlet-aj	hyp
satin-nn	colloc	variegated-aj	hyp
yellow-aj	hyp	yellow-aj	hyp
yellow-nn	sem	yellow-nn	sem

ri-narrow		ri-wide	
beige-aj	colloc	brown-aj	hyp
green-nn	sem	flower-nn	colloc
pink-aj	hyp	pink-aj	hyp
pink-nn	sem	pink-nn	sem
purple-aj	hyp	purple-aj	hyp
red-nn	sem	red-nn	sem
shade-nn	colloc	rose-nn	sem
violet-nn	sem	shade-nn	colloc
yellow-aj	hyp	violet-nn	sem
yellow-nn	sem	yellow-aj	hyp

sing-vv

svd-narrow		svd-wide	
bagpipes-nn	sem	chorus-nn	colloc
carol-nn	colloc	hymn-nn	colloc
dance-vv	hyp	music-nn	sem
duet-nn	colloc	musical-aj	sem
flute-nn	sem	musician-nn	sem
music-nn	sem	sing-nn	deriv
pop-aj	sem	singer-nn	deriv
sing-nn	deriv	song-nn	deriv
song-nn	deriv	tune-nn	colloc
trumpet-nn	sem	vocal-aj	colloc

ri-narrow		ri-wide	
chant-nn	colloc	dance-vv	hyp
cry-vv	hyp	dancing-vv	hyp
dance-vv	hyp	folk-nn	colloc
listen-vv	hyp	listen-vv	hyp
lyric-nn	colloc	listening-vv	hyp
music-nn	sem	sing-nn	deriv
shout-vv	hyp	song-nn	deriv
shouting-vv	hyp	sound-vv	hyp
sing-nn	deriv	tune-nn	colloc
sound-nn	sem	wonderful-aj	sem

hesitate-vv

svd-narrow		svd-wide	
dare-vv	hyp	afraid-aj	sem
grin-vv	hyp	frown-vv	hyp
hesitancy-nn	deriv	glance-nn	sem
luden-nn	out	glance-vv	hyp
moment-nn	sem	mutter-vv	hyp
nod-vv	hyp	nod-vv	hyp
pause-vv	syn	pause-vv	syn
puzzle-aj	sem	reply-vv	hyp
shrug-vv	hyp	shrug-vv	hyp
smile-vv	hyp	smile-vv	hyp

ri-narrow		ri-wide	
hesitancy-nn	deriv	afraid-aj	sem
let-vv	out	frown-vv	hyp
moment-nn	sem	moment-nn	sem
pause-vv	syn	nod-vv	hyp
reply-vv	hyp	pause-vv	syn
smile-vv	hyp	sigh-vv	hyp
stand-vv	out	silence-nn	sem
step-vv	out	smile-vv	hyp
stop-vv	syn	wait-vv	syn
thought-vv	hyp	wonder-vv	hyp

annotate-vv

svd-narrow		svd-wide	
log-vv	syn	annotation-nn	deriv
motif-nn	sem	clipboard-nn	sem
note-nn	deriv	drawing-vv	hyp
playback-nn	sem	paste-vv	hyp
pleading-nn	out	pen-nn	colloc
printing-vv	hyp	printer-nn	sem
reference-vv	hyp	reprint-nn	hyp
rewrite-vv	hyp	revision-nn	sem
widget-nn	out	workbook-nn	colloc
workbook-nn	colloc	worksheet-nn	colloc

ri-narrow		ri-wide	
ascii-nn	sem	copy-vv	hyp
copy-vv	hyp	edit-vv	hyp
copying-vv	hyp	format-nn	sem
delete-vv	hyp	illustration-nn	colloc
edit-vv	hyp	manual-nn	colloc
format-nn	sem	print-nn	hyp
handwritten-aj	colloc	print-vv	hyp
print-vv	hyp	printed-aj	sem
printed-vv	hyp	printed-vv	hyp
record-vv	hyp	text-nn	colloc

minister-nn

svd-narrow		*svd-wide*	
affair-nn	sem	affair-nn	sem
bin-vv	out	bin-vv	out
deputy-nn	hyp	cabinet-nn	sem
minister-vv	deriv	deputy-nn	hyp
ministry-nn	deriv	minister-vv	deriv
minster-nn	syn	ministerial-aj	deriv
mover-nn	out	prime-aj	colloc
office-nn	sem	reshuffle-nn	colloc
reshuffle-nn	colloc	reshuffle-vv	colloc
retd-aj	colloc	resignation-nn	sem

ri-narrow		*ri-wide*		*ri-rel*	
bin-vv	out	cabinet-nn	colloc	fodder-nn	out
deputy-nn	hyp	deputy-nn	hyp	lending-nn	out
list-vv	out	leader-nn	hyp	meridian-nn	out
malaysian-aj	sem	ministerial-aj	deriv	minster-nn	syn
ministerial-aj	deriv	official-nn	hyp	mover-nn	out
minster-nn	syn	opposition-nn	sem	purveyor-nn	out
mover-nn	out	parliament-nn	sem	slot-nn	sem
reappoint-vv	colloc	president-nn	hyp	suspect-nn	out
suspect-nn	out	resign-vv	colloc	target-nn	out
then-aj	colloc	resignation-nn	colloc	warden-nn	sem

sausage-nn

svd-narrow		*svd-wide*	
bacon-nn	hyp	bake-aj	colloc
bake-aj	colloc	cheese-nn	hyp
beef-nn	hyp	chicken-nn	hyp
cheese-nn	hyp	cook-aj	colloc
chicken-nn	hyp	ham-nn	hyp
meat-nn	hyp	pork-nn	hyp
mince-nn	sem	pudding-nn	hyp
pye-nn	hyp	pye-nn	hyp
salad-nn	hyp	roast-aj	colloc
steak-nn	hyp	steak-nn	hyp

ri-narrow		*ri-wide*		*ri-rel*	
bacon-nn	hyp	bacon-nn	hyp	breakfast-nn	sem
cheese-nn	hyp	bread-nn	hyp	cake-nn	hyp
chicken-nn	hyp	cheese-nn	hyp	chicken-nn	hyp
chop-nn	hyp	chicken-nn	hyp	meal-nn	sem
fry-aj	colloc	ham-nn	hyp	meat-nn	hyp
ham-nn	hyp	potato-nn	hyp	pizza-nn	hyp
meat-nn	hyp	pudding-nn	hyp	pye-nn	hyp
omelette-nn	hyp	pye-nn	hyp	sandwich-nn	hyp
pye-nn	hyp	sandwich-nn	hyp	steak-nn	hyp
sandwich-nn	hyp	soup-nn	hyp	supper-nn	sem

starfish-nn

svd-narrow		*svd-wide*	
crab-nn	hyp	anemone-nn	hyp
diurnal-aj	sem	brachiopod-nn	hyp
dorsal-aj	sem	coral-nn	hyp
dragonfly-nn	hyp	crustacean-nn	hyp
extinction-nn	sem	invertebrate-aj	sem
fauna-nn	sem	invertebrate-nn	hyp
flail-vv	out	mollusc-nn	hyp
invertebrate-nn	hyp	silurian-aj	sem
shallow-nn	sem	squid-nn	hyp
spine-nn	sem	trilobite-nn	hyp

ri-narrow		*ri-wide*		*ri-rel*	
anemone-nn	hyp	anemone-nn	hyp	ancestor-nn	out
jewel-nn	hyp	coral-nn	hyp	buzzard-nn	sem
kayak-nn	sem	crab-nn	hyp	cavity-nn	out
lobster-nn	hyp	crustacean-nn	hyp	coat-nn	out
prince-nn	sem	invertebrate-aj	hyp	denominator-nn	out
prosecution-nn	out	invertebrate-nn	hyp	flagstone-nn	out
slug-nn	hyp	mollusc-nn	hyp	martini-nn	out
thorn-nn	colloc	reef-nn	sem	plaster-nn	out
triple-aj	colloc	seaweed-nn	hyp	seaweed-nn	hyp
urchin-nn	hyp	squid-nn	hyp	shrapnel-nn	out

cost-nn

svd-narrow		*svd-wide*	
bpd-nn	sem	cost-vv	deriv
cost-vv	deriv	expense-nn	syn
disbursement-nn	syn	expensive-aj	sem
expense-nn	syn	incur-vv	colloc
fee-nn	syn	marginal-aj	colloc
incur-vv	colloc	monopolist-nn	sem
marginal-aj	colloc	overheads-nn	syn
outlay-nn	syn	quantify-vv	colloc
overheads-nn	syn	realizable-aj	colloc
running-nn	colloc	sink-aj	colloc

ri-narrow		*ri-wide*		*ri-rel*	
capacity-nn	hyp	additional-aj	colloc	concentration-nn	sem
cost-vv	deriv	benefit-vv	colloc	expense-nn	syn
demand-nn	hyp	current-aj	colloc	frequency-nn	hyp
expected-vv	colloc	financial-aj	colloc	outgoings-nn	syn
expense-nn	syn	incentive-nn	hyp	overheads-nn	syn
level-nn	hyp	increase-vv	colloc	price-nn	syn
price-nn	syn	operating-vv	colloc	risk-nn	hyp
salary-nn	hyp	reduce-vv	colloc	salary-nn	hyp
value-nn	hyp	substantial-aj	colloc	tax-nn	hyp
volume-nn	hyp	total-aj	colloc	volume-nn	hyp

sorrow-nn

svd-narrow		*svd-wide*	
anguish-nn	syn	anguish-nn	syn
despair-nn	syn	despair-nn	syn
fear-nn	hyp	grief-nn	syn
grief-nn	syn	grieve-vv	sem
joy-nn	anto	joy-nn	anto
loneliness-nn	hyp	love-aj	hyp
longing-nn	hyp	mourn-vv	sem
misery-nn	syn	sadness-nn	syn
sadness-nn	syn	tragedy-nn	sem
torment-nn	syn	tragic-aj	colloc

ri-narrow		*ri-wide*		*ri-rel*	
anguish-nn	syn	affection-nn	hyp	admiration-nn	hyp
despair-nn	syn	despair-nn	syn	affection-nn	hyp
disappointment-nn	hyp	grief-nn	syn	anger-nn	hyp
excitement-nn	hyp	happiness-nn	anto	disappointment-nn	hyp
fear-nn	hyp	heart-nn	sem	dismay-nn	hyp
frustration-nn	hyp	joy-nn	anto	fear-nn	hyp
happiness-nn	anto	love-nn	hyp	joy-nn	anto
joy-nn	anto	sad-aj	sem	regret-nn	hyp
misery-nn	syn	sadness-nn	syn	sadness-nn	syn
sadness-nn	syn	soul-nn	sem	sympathy-nn	hyp

Summary

	svd narrow	*svd wide*	*ri narrow*	*ri wide*	*ri rel*
syn	26.9%	20.8%	14.6%	8.5%	11.7%
hyp	31.5%	31.5%	41.5%	40.8%	35.0%
deriv	6.1%	6.1%	3.8%	2.3%	0.0%
colloc	7.7%	17.7%	13.1%	15.4%	0.0%
sem	16.1%	18.5%	10.8%	15.4%	20.0%
out	11.5%	5.4%	16.1%	17.7%	33.3%

favouritism-nn

svd-narrow		*svd-wide*	
bully-vv	sem	brazilian-nn	out
courage-nn	out	bridle-nn	out
drawing-nn	out	detest-vv	sem
figure-nn	out	filly-nn	out
histogram-nn	out	mule-nn	out
lose-aj	out	outspoken-aj	out
onshore-aj	out	prejudiced-aj	sem
rebellious-aj	out	prim-aj	out
textile-nn	out	refute-vv	sem
vested-aj	out	scornful-aj	sem

ri-narrow		*ri-wide*		*ri-rel*	
diagram-nn	out	accepted-vv	out	aptitude-nn	sem
fig-nn	out	challenge-nn	out	diagram-nn	out
figure-nn	out	confidence-nn	sem	heterogeneity-nn	sem
inclination-nn	sem	continue-vv	out	histogram-nn	out
insensitiveness-nn	hyp	effort-nn	out	histology-nn	out
map-nn	out	failed-vv	out	inclination-nn	sem
mercy-nn	hyp	insist-vv	out	poll-nn	out
photograph-nn	out	loyalty-nn	hyp	resilience-nn	sem
sign-nn	out	move-nn	out	sign-nn	out
table-nn	out	succeed-vv	sem	willingness-nn	sem

General remarks

- Nearest neighbors close to a linguist's idea of semantic vicinity
- Different words inhabit different neighborhoods (but the algorithms adjust quite well)
- Some words have a more sensible semantic neighborhood than others
- Window size effect not huge
- Not clear if SVD is worth the effort (improve RI with more parameter tuning?)
- Bad linguistics definitely not worth the effort
- Low frequency is a problem

Outline

Introduction

Empirical NLP
- Supervised learning
- Unsupervised learning

Word Meaning
- Introduction
- Meaning as context
- Similarity in word space
- Dimensionality reduction

A Few Experiments
- Introduction
- Experimental setting
- Results

Current Directions

Current directions
The obvious things

- Larger, more systematic test set
- Sampling wider range of neighbors, taking cosine values into account
- More granular classification
- Parameter hell!

Current directions
Looking forward

- Once semantic characteristics of word space model(s) are better understood
- look for non-trivial (psycho-/neuro-)linguistic predictions made by model
- (Crosses with ongoing work on compound processing and generation)

Some references

A. Cruse (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
J. Gorman and J. Curran (2006). Scaling distributional similarity to large corpora. *Proceedings of ACL 2006*, 361-368.
T. Landauer, P. Foltz, and D. Laham (1998). An introduction to Latent Semantic Analysis. *Discourse Processes* **25**, 259-284.
C. Manning and H. Schütze (1999). *Foundations of statistical natural language processing*. MIT Press, Cambridge MA.
S. McDonald and M. Ramscar (2001). Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 611-616.
M. Sahlgren (2005). An introduction to Random Indexing. *TKE 2005*.
M. Sahlgren (2006). *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces.* PhD thesis, Stockholm University.

Some more references

M. Baroni (2003). Distribution-driven morpheme discovery: A computational/experimental study. *Yearbook of Morphology 2003*, 213-248.

S. Evert (2004). *The statistics of word cooccurrences: Word pairs and collocations*. PhD thesis, University of Stuttgart.

J. Firth (1957). A synopsis of linguistic theory 1930-55. In *Studies in Linguistic Analysis*, Oxford: The Philological Society, 1-32.

J. Goldsmith (2001). Unsupervised learning of the morphology of a natural language. *Computational Linguistics* **27**, 153-189.

J. Tiedemann (2003). *Recycling translations: Extraction of lexical data from parallel corpora and their application in natural language processing*. PhD thesis, University of Uppsala.

K. Toutanova, D. Klein, Ch. Manning and Y. Singer (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of HLT-NAACL 2003*, 252-259.

D. Yarowsky (1995). Unsupervised word sense disambiguation rivaling supervised methods. *Proceedings of ACL 1995*, 189-196.