Distinguishing Paradigmatic Semantic Relations: Human Ratings and Distributional Similarity

Sabine Schulte im Walde
schulte@ims.uni-stuttgart.de

Institut für Maschinelle Sprachverarbeitung

Motivation

Paradigmatic semantic relations

- Central in organisation of mental lexicon (Miller & Fellbaum, 1991; Murphy, 2003): *synonymy*, *antonymy*, *hypernymy*, *(co-)hyponymy*
- Provide a structure for the lexical concepts that words express.
- Natural relation structure differs across word classes:
  - hypernymy → noun lexicon; minor for verbs; unnatural for adjectives
  - antonymy → adjective lexicon
  - hypernymy, antonymy, synonymy, entailment → verb lexicon

Distributional vector space models

- Rely on the distributional hypothesis (Harris, 1954; Firth, 1957).
- Model meaning and "similarity" of target words (Turney & Pantel, 2010).
- Paradigmatic relations are difficult to distinguish:
  The boy/girl/person loves/hates the cat.

Perspectives

- Perspectives: cognitive semantics and distributional semantics
- Questions:
  - How do humans perceive and distinguish semantic relatedness?
  - To what extent are corpus-based approaches successful in the distinction?

Human Ratings

Target–response paradigmatic relation pairs

- Targets: Random choice of 99 WordNet targets per word class: nouns, verbs, adjectives (Scheible & Schulte im Walde, 2014), balanced for
  - frequency class (low; mid; high)
  - polysemy class (monosemous; two senses; >2 senses)
  - size of semantic class
- Experiments:
  - generation (5,745/8,910 pair types/tokens)
  - rating (1,684 pair types; scale: 0–5)

Experiment 1: Generation of relation pairs (examples)

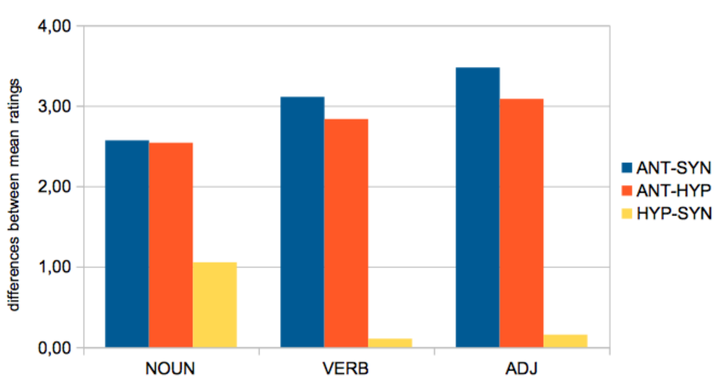
	ANT		SYN		HYP	
NOUN	*Bein/Arm* (leg/arm)	10	*Killer/Mörder* (killer)	8	*Ekel/Gefühl* (disgust/feeling)	7
	Zeit/Raum (time/space)	3	*Gerät/Apparat* (device)	3	*Arzt/Beruf* (doctor/profession)	5
VERB	*verbieten/erlauben* (forbid/allow)	10	*üben/trainieren* (practise)	6	*trampeln/gehen* (lumber/walk)	6
	setzen/stehen (sit/stand)	4	*setzen/platzieren* (place)	3	*wehen/bewegen* (wave/move)	3
ADJ	*dunkel/hell* (dark/light)	10	*mild/sanft* (smooth)	9	*grün/farbig* (green/colourful)	5
	heiter/trist (cheerful/sad)	2	*bekannt/vertraut* (familiar)	4	*heiter/hell* (bright/light)	1

Experiment 2: Rating of relation pairs (examples)

	Target Pair	Generation	ANT	SYN	HYP	Difference	
NOUN	*Arzt/Beruf* (doctor/profession)	HYP: 5	0.8	1.1	4.7	HYP–SYN	3.6
	Verhandlung/Gespräch (negotiation/conversation)	HYP: 4	0.6	2.8	4.0	HYP–SYN	1.2
VERB	*befehlen/gehorchen* (command/obey)	ANT: 6	4.4	0.3	0.1	ANT–SYN	4.1
	schmieren/streichen (grease/paint)	SYN: 4	0.9	2.2	3.3	SYN–HYP	-1.1
ADJ	*faul/fleißig* (lazy/diligent)	ANT: 8	5.0	0.5	0.0	ANT–SYN	4.5
	gewitzt/naiv (smart/naïve)	ANT: 3	3.0	0.3	0.4	ANT–SYN	2.7

Distinction between relation pairs

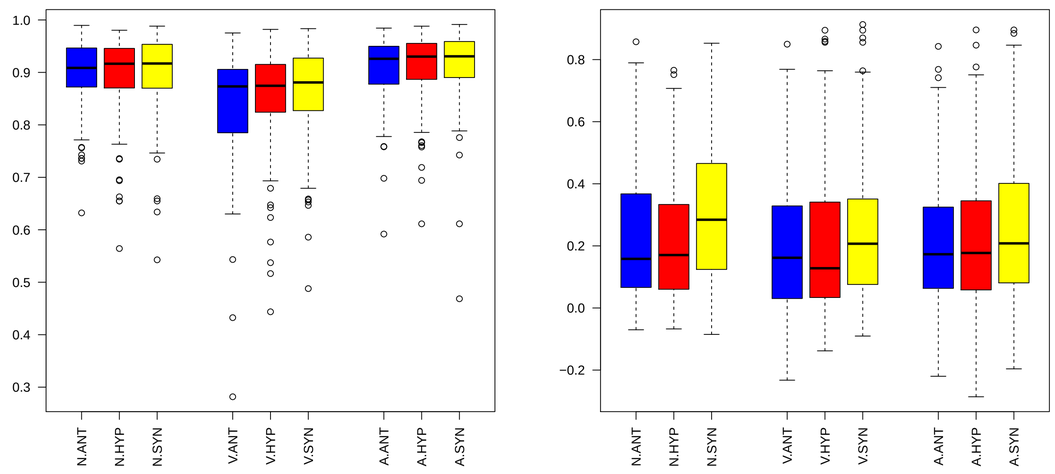
How well do experiment participants distinguish between paradigmatic relations?
→ differences in mean ratings across relation pairs



Distributional Models

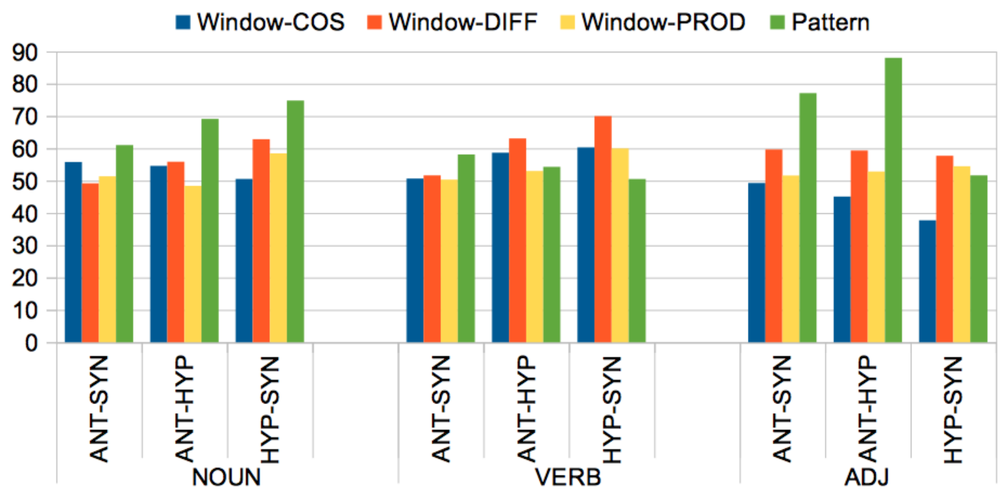
Cosine similarity

- Paradigmatic semantic relation pairs are expected to be close in word space.
- Vector space: 20-word co-occurrence in web corpus *DECOW14AX*, weighted by *frequency* vs. *local mutual information* (Evert, 2005; Schäfer & Bildhauer, 2012)
- Distributional similarity: cosine of vector angle



Automatic classification

- Series of classification experiments
- Features: window co-occurrence vs. lexico-syntactic patterns
- Vector representations of relation word pairs:
  - Window-COS: cosine scores between word pairs
  - Window-DIFF: difference vector for word pair
  - Window-PROD: vector product for word pair
  - Pattern: linear word sequences between related words
- Corpus: *SdeWaC* (Faaß & Eckart, 2013)
- Nearest-centroid classifier (also known as *Rocchio classifier*)
  - Use training pairs to initiate relation class centroids.
  - Assign test pairs to nearest class centroid.
  - Evaluation: 5-fold cross-validation, precision values.



Results:

- Salience of feature types depends on the word class.
- Pattern information outperforms window information (nouns + adjectives).
- Automatic classification is best for natural relations (nouns + adjectives).
- Verbs are different.

References

Stefan Evert. *The Statistics of Word Co-Occurrences: Word Pairs and Collocations*. PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 2005.

Gertrud Faaß & Kerstin Eckart. SdeWaC – A Corpus of Parsable Sentences from the Web. In *Proceedings of GSCL*, pages 61–68, Darmstadt, Germany, 2013.

John R. Firth. *Papers in Linguistics 1934-51*. Longmans, London, UK, 1957.

Zellig Harris. Distributional structure. *Word*, 10(23):146–162, 1954.

George A. Miller & Christiane Fellbaum. Semantic Networks of English. *Cognition*, 41:197–229, 1991.

M. Lynne Murphy. *Semantic Relations and the Lexicon*. Cambridge University Press, 2003.

Roland Schäfer & Felix Bildhauer. Building Large Corpora from the Web Using a New Efficient Tool Chain. In *Proceedings of LREC*, pages 486–493, 2012.

Silke Scheible & Sabine Schulte im Walde. A Database of Paradigmatic Semantic Relation Pairs for German Nouns, Verbs and Adjectives. In *Proceedings of the COLING Workshop on Lexical and Grammatical Resources for Language Processing*, pages 111–119, 2014.

Peter D. Turney & Patrick Pantel. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37:141–188, 2010.