Using Associations to identify Salient Features for Data-intensive Lexical Semantic Tasks

PD Dr. Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung (IMS)
Universität Stuttgart

October 11, 2012

Institute of Psychology, University of Leuven

Overview

Background

- Background: Computational Linguistics
- Line of Research:
  - Data-intensive distributional lexical semantics
  - Focuses:
    - lexical acquisition
    - semantic classes and semantic relatedness
    - compositionality
    - (particle) verbs
    - evaluation
  - Interdisciplinary research: theory, cognition, computation

Motivation

- Goal: explore the potential and the limits of (text-based) distributional approaches to lexical semantics
- Tool: distributional models / vector space models (describe & compare by corpus-derived features)
- Role of associations:
  - Associations are used in gold standards for lexical semantics.
  - Associations help identifying salient semantic features.
- Basis:
  - *co-occurrence hypothesis*: associations $\leftrightarrow$ corpus co-occurrence
  - *distributional hypothesis*: corpus co-occurrence $\leftrightarrow$ meaning

Procedure

1. (Standard) Analyses of association norms:
   - part-of-speech analysis of associate responses
   - window co-occurrence of stimulus–associate types
   - syntax-semantic functions of associates with respect to stimuli
   - semantic relations between stimuli and associates
   - *etc.*
2. Apply associate information as gold standard (if appropriate).
3. Exploit associate knowledge with respect to semantic task.

Overview of Association Norms

- Associations to German verbs, collected in 2004 (Schulte im Walde et al., 2008):
  - 330 verbs including 36 particle verbs
  - 44–54 participants per stimulus
  - 38,769/79,480 stimulus–association types/tokens
- Associations to German particle verbs collected in 2004 (Schulte im Walde, 2005):
  - 100 verbs including 76 particle verbs
  - 32–35 participants per stimulus
  - 10,009/17,442 stimulus–association types/tokens
- Associations to German nouns, collected in 2003/2004 (Melinger and Weber, 2006):
  - 409 nouns referring to picturable objects
  - 50 participants per stimulus ($\times$ 2 modes)
  - 30,845/116,714 stimulus–association types/tokens

Overview of Association Norms

- Associations to German noun compounds collected in 2010–2012:
  - web experiment with 996 compounds+constituents for 442 noun compounds (Schulte im Walde et al., 2012):
    - 10–36 participants per stimulus
    - 28,238/47,249 stimulus–association types/tokens
  - AMT experiment with 571 compounds+constituents for 246 noun-noun compounds (unpublished):
    - 2–120 (in general: 30) participants per stimulus
    - 26,415/59,444 stimulus–association types/tokens
  - web data + AMT data contains a total of 47,523/106,693 stimulus–association types/tokens

Using the Association Norms in Distributional Semantics

Norms	Goal
Verbs Nouns	Identify salient features for distributional models
Verbs	Ditto; for semantic verb classification
Particle verbs	Gold standard to interpret distributional nearest neighbours
Compounds	Explore distributional factors of semantic relatedness between compounds and their constituents

Distributional Semantic Tasks

1. Semantic verb classification
2. Compositionality of Noun Compounds

Semantic Verb Classification: Motivation

- Resource-intensive vs. automatic methods
- Manual example classifications:
  - Levin classes based on syntax-semantics alternation behaviour (Levin, 1993)
  - *WordNet* based on synonymy (Fellbaum, 1998)
  - *FrameNet* based on situation agreement (Fillmore et al., 2003)
- Automatic example classifications:

  Merlo & Stevenson (2001); Korhonen et al. (2003); Schulte im Walde (2003; 2006); Joanis et al. (2008)
- Basis: distributional hypothesis
- Task: automatic, corpus-based semantic verb classification

Semantic Verb Classification relying on Associations

- Assumption: semantically related verbs have common associations → they are assigned to common classes
- Method: agglomerative hierarchical clustering with German verbs and associations as features
- Standard setup:
  - similarity measure: *skew divergence*
  - merging criterion: *Ward's method* (sum-of-squares)
- Validation as a gold standard:

  pair-wise comparison against *GermaNet* and *FrameNet*

  → F-score of 62.69% for GermaNet (upper bound: 82.35%)

  → F-score of 34.68% for FrameNet (upper bound: 60.31%)

Semantic Verb Classification relying on Associations

Example classes and verbs	Strongest association features
bedauern 'regret', *heulen* 'cry', *jammern* 'moan', *klagen* 'complain, moan, sue', *verzweifeln* 'become desperate', *weinen* 'cry'	*Trauer* 'mourning', *weinen* 'cry', *traurig* 'sad', *Tränen* 'tears', *jammern* 'moan', *Angst* 'fear', *Mitleid* 'pity', *Schmerz* 'pain', *etc.*
abnehmen 'lose weight', *abspecken* 'lose weight', *zunehmen* 'gain weight'	*Diät* 'diet', *Gewicht* 'weight', *dick* 'fat', *abnehmen* 'lose weight', *Waage* 'scale', *Essen* 'food', *essen* 'eat', *Sport* 'sports', *dünn* 'thin', *Fett* 'fat', *etc.*

Corpus-based Semantic Verb Classification

- Goal: compare corpus-based features in standard verb clustering with association-based and manual gold standards
- Method: agglomerative hierarchical clustering with German verbs and corpus-based features (details as above)
- Features: 20-window co-occurrence and dependency-based corpus features from 200-million word newspaper corpus
- Gold standards:
    - association-based clustering: 100 clusters with 330 verbs
    - GermaNet: hard random selection of 100 synsets, 233 verbs
    - FrameNet: hard version of all 77 classes with 406 verbs

Corpus-based Semantic Verb Classification

	grammar relations						
	$\langle \underline{NP_n} \rangle$	$\langle \underline{NP_n}, NP_a \rangle$	$\langle NP_n, \underline{NP_a} \rangle$	NP	PP	NP&PP	ADV
Assoc	35.90	37.18	39.25	39.14	37.97	**41.28**	38.53
GN	**58.01**	53.37	51.90	53.10	54.21	51.77	51.82
FN	29.46	30.13	32.74	34.16	28.72	33.91	**35.24**

	co-occurrence: window-20				
	all	ADJ	ADV	N	V
Assoc	**39.33**	37.31	36.89	**39.33**	38.84
GN	51.53	50.88	47.79	**52.86**	49.12
FN	32.01	31.08	31.00	**34.24**	31.75

Semantic Verb Classification: Summary

- Associations provide the knowledge we need for automatic semantic classification.

  → Modelling the (syntax-)semantic relatedness between stimuli and associations can guide us towards salient features.

- Caveats:
  - There are significant differences in accuracy: feature type ↔ gold standard type.
  - The association-based clustering is modelled worst.

- Conclusions:
  - Association-based clustering represents one gold standard semantic classification among others.
  - We need to model association knowledge beyond standard feature types.

Noun-Noun Compositionality: Motivation

- Interest: semantic relatedness between noun-noun compounds and their nominal constituents
- Examples:
  - *Blockflöte* *'flute'* / *Block* *'block; fragment; pad'* / *Flöte* *'flute'*
  - *Fliegenpilz* *'fly agaric'* / *Fliege* *'fly; bow tie'* / *Pilz* *'mushroom'*
  - *Schlittenhund* *'sledge dog'* / *Schlitten* *'sledge'* / *Hund* *'dog'*
- Basis: distributional hypothesis
- Task: automatic prediction of the degree of compositionality of the compounds with respect to the constituents

Noun-Noun Compositionality: Data

- 246 depictable German noun-noun compounds
  (von der Heide and Borgwaldt, 2009)
- Transparent vs. opaque compounds
- Compositionality judgements:
  - for compounds and each constituent on a scale 1–7
  - 35 participants for each compound–constituent pair
  - gold standard: mean values of judgements
- Associations for compounds and constituents
  (85,049 tokens over 34,560 types for 571 stimuli)

Noun-Noun Compositionality: Associations

- Assumption: transparent compounds have more associations in common with their constituents than opaque compounds
- Method: standard vector space model compares vectors of compounds with vectors of constituents
- Standard setup:
  - features: window co-occurrence
  - similarity measure: *cosine*
- Evaluation: *Spearman rank-order correlation coefficient* (r_S) for cosine values against mean compositionality judgements

Noun-Noun Compositionality: Predictions

r_S varying the vector space features (using the *sdeWaC* corpus):

Features		r_S both const	r_S const1	r_S const2
Baseline	association overlap	.5394	.5702	.5680
Vector space	associations	.5676	.5752	.6267
	window 20: *all*	.1918	.1958	.1190
	window 20: nouns	.4742	.4806	.4416
	window 20: verbs	.2773	.1883	.2432
	window 20: adjectives	.2261	.2136	.2000

Noun-Noun Compositionality: Predictions

r_S varying the corpus:

Features	Corpora				
	HGC	Wikipedia	WebKo		sdeWaC
	ext	ext	ext	int	int
	200	430	1,500	1,500	880
window 20: nouns	.2214	.3549	.4065	.3306	.4742

Noun-Noun Compositionality: Summary

- Confirmation: Transparent compounds have more associations in common with their constituents than opaque compounds.
- Associations provide the knowledge we need for predicting compositionality.

  → Modelling the overlap of associations to compounds and constituents can guide us towards salient features.
- Association properties: Different feature types provide complementary information for compositionality.

  → Can we distinguish between the contributions of the various parts-of-speeches?

  → Can we specify the compositionality with respect to modifier vs. head?

Conclusions

- Associations provide a lot of the knowledge we need for distributional lexical semantics.
- Questions:
  - how can we improve the automatic identification of stimulus–associate relationships?
  - what is the role of corpus domain(s)?
  - what is the role of corpus size?
  - how can we exploit world knowledge in association norms?

Colleagues

- Susanne Borgwaldt (Braunschweig/Erfurt)
- Ronny Jauch (Stuttgart)
- Alissa Melinger (Dundee)
- Stefan Müller (Stuttgart)

References: Web Corpora

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta.
The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora.
Language Resources and Evaluation, 43(3):209–226, 2009.

Marco Baroni and Adam Kilgarriff.
Large Linguistically-processed Web Corpora for Multiple Languages.
In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 2006.

Gertrud Faaß Ulrich Heid, and Helmut Schmid.
Design and Application of a Gold Standard for Morphological Analysis: SMOR in Validation.
In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 803–810, Valletta, Malta, 2010.

References: Distributional Hypothesis

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

Zellig Harris.
Distributional Structure.
In Jerold J. Katz, editor, *The Philosophy of Linguistics*, Oxford Readings in Philosophy, pages 26–47. Oxford University Press, 1968.

Diego Marconi.
Lexical Competence.
MIT Press, Cambridge, MA, 1997.

References: Distributional Hypothesis

Marco Baroni and Alessandro Lenci.
Distributional Memory: A General Framework for Corpus-based Semantics.
Computational Linguistics, 36(4):673–721, 2010.

Kenneth W. Church and Patrick Hanks.
Word Association Norms, Mutual Information, and Lexicography.
Computational Linguistics, 16(1):22–29, 1990.

Dekang Lin.
Extracting Collocations from Text Corpora.
In *Proceedings of the First Workshop on Computational Terminology*, 1998.

Kevin Lund and Curt Burgess.
Producing High-Dimensional Semantic Spaces from Lexical Co-Occurrence.
Behavior Research Methods, Instruments, and Computers, 28(2):203–208, 1996.

Hinrich Schütze.
Dimensions of Meaning.
In *Proceedings of Supercomputing*, pages 787–796, 1992.

Hinrich Schütze.
Automatic Word Sense Discrimination.
Computational Linguistics, 24(1):97–123, 1998.

References: Association Norms

Susanne Borgwaldt, Catherine-Marie Longtin, Rachel Kemps, and Gary Libben.
Semantic Transparency Ratings and Associations to English Compounds.
Unpublished raw data, 2005.

Simon de Deyne and Gert Storms.
Word associations: Norms for 1,424 dutch words in a continuous task.
Behavior Research Methods, 40(1):198–205, 2008.

Simon de Deyne and Gert Storms.
Word Association Study, Ongoing.
URL: www.smallworldofwords.com/.

Annamaria Guida.
The Representation of Verb Meaning within Lexical Semantic Memory: Evidence from Word Associations.
Master's thesis, Universit degli studi di Pisa, 2007.

George R. Kiss, Christine Armstrong, Robert Milroy, and James Piper.
An Associative Thesaurus of English and its Computer Analysis.
In *The Computer and Literary Studies*. Edinburgh University Press, 1973.

References: Association Norms

Alissa Melinger and Andrea Weber.
Database of Noun Associations for German, 2006.
URL: www.coli.uni-saarland.de/projects/nag/.

Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber.
The University of South Florida Word Association, Rhyme, and Word Fragment Norms, 1998.

D.S. Palermo and James J. Jenkins.
Word Association Norms: Grade School through College.
University of Minnesota Press, Minneapolis, 1964.

Sabine Schulte im Walde.
Exploring Features to Identify Semantic Nearest Neighbours: A Case Study on German Particle Verbs.
In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 608–614, Borovets, Bulgaria, 2005.

Sabine Schulte im Walde, Susanne Borgwaldt, and Ronny Jauch.
Association Norms of German Noun Compounds.
In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 632–639, Istanbul, Turkey, 2012.

References: Association Norms

Sabine Schulte im Walde, Alissa Melinger, Michael Roth, and Andrea Weber.
An Empirical Characterisation of Response Types in German Association Norms.
Research on Language and Computation, 6(2):205–238, 2008.
DOI 10.1007/s11168-008-9048-4.

Claudia von der Heide and Susanne Borgwaldt.
Assoziationen zu Unter-, Basis- und Oberbegriffen. Eine explorative Studie.
In *Proceedings of the 9th Norddeutsches Linguistisches Kolloquium*, pages 51–74, 2009.

References: Analyses of Association Norms

Herbert H. Clark.
Word Associations and Linguistic Theory.
In John Lyons, editor, *New Horizon in Linguistics*, chapter 15, pages 271–286. Penguin, 1971.

Simon de Deyne and Gert Storms.
Word associations: Network and semantic properties.
Behavior Research Methods, 40(1):213–231, 2008.

Christiane Fellbaum and Roger Chaffin.
Some Principles of the Organization of Verbs in the Mental Lexicon.
In *Proceedings of the 12th Annual Conference of the Cognitive Science Society of America*, 1990.

Annamaria Guida.
The Representation of Verb Meaning within Lexical Semantic Memory: Evidence from Word Associations.
Master's thesis, Universit degli studi di Pisa, 2007.

Hans Jürgen Heringer.
The Verb and its Semantic Power: Association as the Basis for Valence.
Journal of Semantics, 4:79–99, 1986.

References: Analyses of Association Norms

Reinhard Rapp.
The Computation of Word Associations: Comparing Syntagmatic and Paradigmatic Approaches.
In *Proceedings of the 19th International Conference on Computational Linguistics*, Taipei, Taiwan, 2002.

Michael Roth and Sabine Schulte im Walde.
Corpus Co-Occurrence, Dictionary and Wikipedia Entries as Resources for Semantic Relatedness Information.
In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pages 1852–1859, Marrakech, Morocco, 2008.

Sabine Schulte im Walde and Alissa Melinger.
An In-Depth Look into the Co-Occurrence Distribution of Semantic Associates.
Italian Journal of Linguistics. Alessandro Lenci (guest editor): From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science", 20(1):89–128, 2008.

Sabine Schulte im Walde, Alissa Melinger, Michael Roth, and Andrea Weber.
An Empirical Characterisation of Response Types in German Association Norms.
Research on Language and Computation, 6(2):205–238, 2008.
DOI 10.1007/s11168-008-9048-4.

References: Co-Occurrence Hypothesis

Christiane Fellbaum.
Co-Occurrence and Antonymy.
Lexicography, 8(4):281–303, 1995.

George Miller.
The Organization of Lexical Memory: Are Word Associations sufficient?
In George A. Talland and Nancy C. Waugh, editors, *The Pathology of Memory*, pages 223–237. Academic Press, New York, 1969.

Donald P. Spence and Kimberly C. Owens.
Lexical Co-Occurrence and Association Strength.
Journal of Psycholinguistic Research, 19:317–330, 1990.

References: Semantic Verb Classification (manual)

Katrin Erk, Andrea Kowalski, and Manfred Pinkal.
A Corpus Resource for Lexical Semantics.
In *Proceedings of the 5th International Workshop on Computational Semantics*, Tilburg, The Netherlands, 2003.

Christiane Fellbaum, editor.
WordNet – An Electronic Lexical Database.
Language, Speech, and Communication. MIT Press, Cambridge, MA, 1998.

Charles J. Fillmore, Christopher R. Johnson, and Miriam R.L. Petruck.
Background to FrameNet.
International Journal of Lexicography, 16:235–250, 2003.

Claudia Kunze.
Extension and Use of GermaNet, a Lexical-Semantic Database.
In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, pages 999–1002, Athens, Greece, 2000.

Beth Levin.
English Verb Classes and Alternations.
The University of Chicago Press, 1993.

References: Semantic Verb Classification (automatic)

Vasileios Hatzivassiloglou and Kathleen R. McKeown.
Towards the Automatic Identification of Adjectival Scales: Clustering Adjectives According to Meaning.
In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pages 172–182, Columbus, OH, 1993.

Eric Joanis, Suzanne Stevenson, and David James.
A General Feature Space for Automatic Verb Classification.
Natural Language Engineering, 14(3):337–367, 2008.

Anna Korhonen, Yuval Krymolowski, and Zvika Marx.
Clustering Polysemic Subcategorization Frame Distributions Semantically.
In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 64–71, Sapporo, Japan, 2003.

Paola Merlo and Suzanne Stevenson.
Automatic Verb Classification Based on Statistical Distributions of Argument Structure.
Computational Linguistics, 27(3):373–408, 2001.

References: Semantic Verb Classification (automatic)

Sabine Schulte im Walde.
Experiments on the Automatic Induction of German Semantic Verb Classes.
PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 2003.
Published as AIMS Report 9(2).

Sabine Schulte im Walde.
Experiments on the Automatic Induction of German Semantic Verb Classes.
Computational Linguistics, 32(2):159–194, 2006.

References: Noun Compounds

- MWE Community.
  Multi-Word Expressions Web, Ongoing.
  URL: multiword.sourceforge.net.

- Wolfgang Fleischer and Irmhild Barz.
  Wortbildung der deutschen Gegenwartssprache.
  de Gruyter, 2012.

- Rochelle Lieber and Pavol Stekauer, editors.
  The Oxford Handbook of Compounding.
  Oxford University Press, 2009.

References: Compositionality

Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do, and Chung chieh Shan.
Entailment above the Word Level in Distributional Semantics.
In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, Francs, 2012.

Jeff Mitchell and Mirella Lapata.
Composition in Distributional Models of Semantics.
Cognitive Science, 34:1388–1429, 2010.

Siva Reddy, Ioannis P. Klapaftis, Diana McCarthy, and Suresh Manandhar.
Dynamic and Static Prototype Vectors for Semantic Composition.
In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 705–713, Chiang Mai, Thailand, 2011.

Siva Reddy, Diana McCarthy, and Suresh Manandhar.
An Empirical Study on Compositionality in Compound Nouns.
In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 210–218, Chiang Mai, Thailand, 2011.