

Webkorpora für die automatische Akquisition lexikalisch-semantischen Wissens

Webkorpora in Computerlinguistik und Sprachforschung

PD Dr. Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung (IMS)
Universität Stuttgart

28. September 2012

Institut für Deutsche Sprache, Mannheim

Überblick

① Fragestellung

② Korpora

③ Semantisches Wissen

Assoziationen und Korpus-Kookkurrenz
Kompositionalität von Nomen-Komposita

④ Zusammenfassung

⑤ Referenzen

Fragestellung

- Können Korpora für den automatischen Erwerb lexikalisch-semantischen Wissens genutzt werden?
- Welche Art von semantischer Information kann aus Korpus-Daten akquiriert werden, und wo sind die Grenzen?

Fragestellung

- Können Korpora für den automatischen Erwerb lexikalisch-semantischen Wissens genutzt werden?
- Welche Art von semantischer Information kann aus Korpus-Daten akquiriert werden, und wo sind die Grenzen?

→ *inferential vs. referential abilities* (Marconi, 1997)

Fragestellung

- Können Korpora für den automatischen Erwerb lexikalisch-semantischen Wissens genutzt werden?
- Welche Art von semantischer Information kann aus Korpus-Daten akquiriert werden, und wo sind die Grenzen?

→ *inferential vs. referential abilities* (Marconi, 1997)

- Welche Arten von Korpus-Daten sind für die Akquisition welcher Art von semantischer Information relevant?
- Welche Rolle spielen Webkorpora bei der Akquisition semantischer Information?

Anwendungsbereich

- Webkorpora aus Anwendersicht
- Bereich: **automatische, Korpus-basierte Akquisition von lexikalisch-semantischem Wissen**
- Forschungsbereiche:
 - Assoziationen als Wegweiser für die Akquisition von lexikalisch-semantischem Wissen
 - semantische Relationen und Klassifikationen
 - Kompositionalität von Nomen-Komposita und Partikelverben
- relevante Eigenschaften von Korpora:
 - Domäne(n) des Korpus
 - Größe des Korpus
 - Aufbereitung des Korpus

Anwendungsbereich

- Webkorpora aus Anwendersicht
- Bereich: automatische, Korpus-basierte Akquisition von lexikalisch-semantischem Wissen
- Forschungsbereiche:
 - Assoziationen als Wegweiser für die Akquisition von lexikalisch-semantischem Wissen
 - semantische Relationen und Klassifikationen
 - Kompositionalität von Nomen-Komposita und Partikelverben
- relevante Eigenschaften von Korpora:
 - Domäne(n) des Korpus
 - Größe des Korpus
 - Aufbereitung des Korpus

Korpora im Vergleich

- **Zeitungskorpus:** **Huge German Corpus** (HGC);
200 Millionen Wörter aus *Frankfurter Rundschau*, *Stuttgarter Zeitung*, *VDI-Nachrichten*, *die tageszeitung*, *Gesetzestexte*, *Donaukurier*, *Computerzeitung*
- **Webkorpora:**
 - **deutsches Webkorpus:** *deWaC* (Baroni et al., 2009);
Original mit 1,7 Milliarden Wörtern; Stuttgarter Versionen:
 - **WebKo:** 1,5 Milliarden Wörter
 - **sdeWaC** (version 3): 880 Millionen Wörter
 - **deutsche Wikipedia:**
 - **Wiki 2006** (Michael Roth): 407 Artikel, 2,8 Millionen Wörtern
 - **Wiki 2011** (Lukas Michelbacher): 430 Millionen Wörter

Semantische Aufgaben im Korpus-Vergleich

- 1 Assoziationen als Wegweiser für die Akquisition von lexikalisch-semantischem Wissen:

Analyse von Eigenschaften von Assoziationen und Stimulus-Assoziations-Paaren gibt Hinweise zu salienten Korpus-Merkmalen für die semantische Modellierung.

- 2 Kompositionalität von Nomen-Komposita:

Distributionelle Modellierung von Nomen-Komposita und ihren Konstituenten sagt den Grad der Kompositionalität der Komposita vorher.

Assoziationen und Assoziationsnormen

- **Assoziationen:** Wörter in spontaner Reaktion auf einen Stimulus

Assoziationen und Assoziationsnormen

- **Assoziationen**: Wörter in spontaner Reaktion auf einen Stimulus
Beispiel: *rennen*

Assoziationen und Assoziationsnormen

- **Assoziationen:** Wörter in spontaner Reaktion auf einen Stimulus
Beispiel: *rennen* → *Marathon, schnell, bewegen, Wald*

Assoziationen und Assoziationsnormen

- **Assoziationen:** Wörter in spontaner Reaktion auf einen Stimulus
Beispiel: *rennen* → *Marathon, schnell, bewegen, Wald*
- **Assoziationsnormen:**
Quantifizierung über alle Assoziationen zu einem Stimulus

Assoziationen und Assoziationsnormen

- **Assoziationen:** Wörter in spontaner Reaktion auf einen Stimulus

Beispiel: *rennen* → *Marathon, schnell, bewegen, Wald*

- **Assoziationsnormen:**

Quantifizierung über alle Assoziationen zu einem Stimulus

<i>klagen</i>	
<i>Gericht</i>	19
<i>jammern</i>	18
<i>weinen</i>	13
<i>Anwalt</i>	11
<i>Richter</i>	9
<i>Klage</i>	7
<i>Leid</i>	6
<i>Trauer</i>	6
<i>Klagemauer</i>	5
<i>laut</i>	5

<i>Schloss</i>	
<i>Schlüssel</i>	51
<i>Tür</i>	15
<i>Prinzessin</i>	8
<i>Burg</i>	8
<i>sicher</i>	7
<i>Fahrrad</i>	7
<i>schließen</i>	7
<i>Keller</i>	7
<i>König</i>	7
<i>Turm</i>	6

Assoziationen als semantische Ressource

- Assoziationen reflektieren typische **linguistische und konzeptuelle Merkmale** des Stimulus.
- Grundlage für psycholinguistische Untersuchungen zum **semantischen Gedächtnis** (Repräsentation und Zugriff).
- Grundlage für korpus-linguistische Untersuchungen zu **Bedeutungen und Bedeutungsbeziehungen**.
- **Kookkurrenz-Hypothese:**
Assoziationen ↔ Korpus-Kookkurrenz
- **Distributionelle Hypothese:**
Korpus-Kookkurrenz ↔ Bedeutung

Assoziationen und Korpus-Kookkurrenz

- Welche Art von semantischer Information kann aus Korpus-Daten akquiriert werden, und wo sind die Grenzen?
- Werkzeug: **Assoziationsnormen** (Stimulus-Assoziations-Paare)
- Untersuchungen anhand von Kookkurrenz-Fenstern: Stimuli und Assoziationen . . .
 - in Fenstern verschiedener Korpora
 - in Fenstern verschiedener Korpus-Größen

→ Abdeckung von Bedeutungsaspekten im Korpus-Vergleich

Assoziationen und Korpus-Kookkurrenz: Korpus-Vergleich

Korpus-Kookkurrenz in HGC, deWaC (WebKo, sdeWaC) und Wikipedia

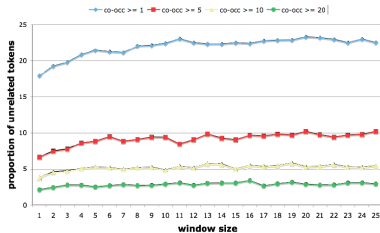
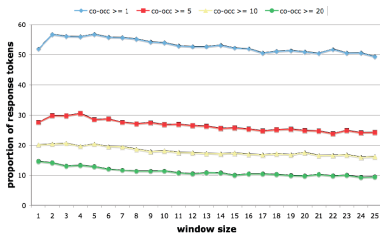
Frage: Wie groß ist der prozentuale Anteil der Stimulus-Assoziations-Typen, die im entsprechenden Korpus in einem Fenster von x Wörtern (links und rechts) vorkommen?

Fenstergröße	Korpus						
	HGC	Wiki	WebKo			sdeWaC	
			Teile	ext	int	int	
	200	430	200	1,500	1,500	880	
1	17	18	19	19	37	36	30
2	29	32	32	32	52	51	47
5	41	44	43	43	62	61	59
10	48	52	50	49	67	65	64
20	54	57	55	55	71	67	66

Assoziationsnorm: **Nomen-Komposita und ihre Konstituenten**; 571 Komposita und Konstituenten mit 34.560 Typen von Stimulus-Assoziations-Paaren

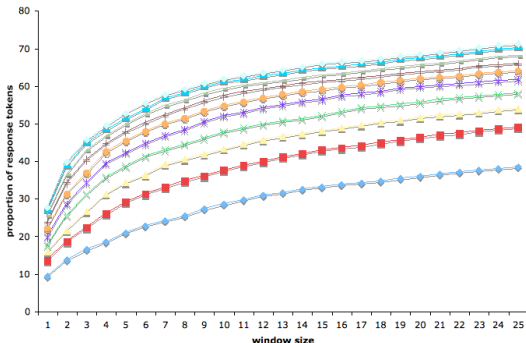
Korpus-Kookkurrenz von (nicht-)semantischen Paaren

Vergleich von Korpus-Kookkurrenz (HGC) für Verb-Assoziationen-Paare mit semantischen Beziehungen vs. Paare ohne semantische Beziehungen:



Assoziationen und Korpus-Kookkurrenz: Größen-Vergleich

Abdeckung von Verb-Assoziationen Korpus-Kookkurrenz in einem Fenster von 25 Wörtern (links und rechts) bei 10, 20, ..., 100% des HGC-Korpus:



Assoziationen: Korpus, Wörterbuch, Enzyklopädie

- Welche Art von semantischer Information kann aus Korpus-Daten akquiriert werden, und wo sind die Grenzen?
- Vergleich der Abdeckung von Stimulus-Assoziations-Paaren in Korpus (HGC) vs. Wörterbuch (WDG) vs. Enzyklopädie (Wikipedia)
- Wörterbuch der Deutschen Gegenwartssprache (WDG):
 - 130.000 Einträge mit 593.000 Wörtern; \emptyset 840 Wörter pro Eintrag
 - 7/0% der Paare fehlen in den Wörterbuch-Einträgen (Stimulus oder Assoziation oder beide).
 - 12/13% der Nomen/Verb-Assoziationen-Paare vorhanden.
- Wikipedia (2006):
 - 2.447 Artikel mit 2,9 Millionen Wörtern; \emptyset 1.164 Wörter pro Artikel
 - 2/56% der Paare fehlen in den Wikipedia-Einträgen (Stimulus oder Assoziation oder beide).
 - 26/6% der Nomen/Verb-Assoziationen-Paare vorhanden.

Assoziationen: Korpus, Wörterbuch, Enzyklopädie

- Token-Typ-Verhältnisse für Nomen/Verb-Assoziations-Paare:
Korpus: 1,2/1,2;
Wikipedia: 1,8/1,7;
Wörterbuch: 2,3/2,0

→ Abdeckung von starken Stimulus-Assoziations-Paaren:
Wörterbuch > Wikipedia > Korpus
- Introspektive Analyse zeigt, dass Wörterbuch und Enzyklopädie nicht offensichtlich mehr Weltwissen abdecken als Korpus-Kookkurrenz.

Assoziationen und Korpus-Kookkurrenz: Datenbeispiele

Frage: Wie oft kommen die Stimulus-Assoziations-Paare in einem Fenster von 20 Wörtern (links und rechts) in den entsprechenden Korpora vor?

Stimulus-Assoziations-Paare (mit Nenn-Häufigkeit)	Korpus						
	HGC	Wiki	WebKo			sdeWaC	
			Teile	ext	int	int	
	200	430	200	1,500	1,500	880	
<i>Blockflöte – Musik</i> (23)	26	73	6	6	90	43	22
<i>Fliegenpilz – giftig</i> (34)	0	9	5	9	40	14	9
<i>Obstkuchen – backen</i> (7)	1	1	1	1	17	12	5
<i>Schlittenhund – Winter</i> (10)	1	5	0	0	6	3	3
<i>Telefonzelle – gelb</i> (25)	16	5	3	6	17	14	6
<i>Wasserhahn – tropfen</i> (9)	14	8	21	11	120	89	55

Nomen-Komposita: Daten und Aufgabe

- 246 zweigliedrige Nomen-Nomen-Komposita
- Beispiele: *Blockflöte*, *Fliegenpilz*, *Schlittenhund*
- transparente vs. opake Komposita
- Kompositionalitätsbewertungen auf einer Skala 1 – 7;
Bewertungen in Bezug auf beide Konstituenten;
35 Teilnehmer pro Kompositum-Konstituenten-Paar
- Aufgabe: distributionelle Modellierung der Kompositionalität

Nomen-Komposita: Modellierung

- **Distributionelles Modell:** Vorhersage der Kompositionalität auf Basis von Korpus-Kookkurrenz
- **Distributionelle Hypothese:**
Korpus-Kookkurrenz \leftrightarrow Bedeutung
- Modell: Fenster-basierte Kookkurrenz
- Vorgehen:
 - 1 Merkmals-Vektoren für Komposita und Konstituenten erstellen.
 - 2 Kosinus-Distanz zwischen Vektoren von Komposita und Konstituenten berechnen.
 - 3 Korrelation zwischen Kosinus-Distanzen und Kompositionalitätsbewertungen berechnen.

Nomen-Komposita: Modellierung im Korpus-Vergleich

Frage: Wie stark ist die Korrelation zwischen Kosinus-Distanzen und Kompositionalitätsbewertungen (*Spearman Rank-Order Correlation, (r_s)*)?

Fenstergröße	Korpus						
	HGC	Wiki	WebKo				sdeWaC
			Teile	ext	int	int	
	200	430	200	1,500	1,500	880	
1	.2611	.2683	.1138	.1543	.2017	.2168	.3779
2	.3133	.3001	.1614	.1867	.2322	.2411	.4469
5	.2831	.3072	.2102	.2435	.3268	.3118	.4624
10	.2407	.3336	.2639	.2808	.3664	.3299	.4703
20	.2214	.3549	.2999	.3024	.4065	.3306	.4742

Nomen-Komposita: Modellierung im Korpus-Vergleich

Frage: Wie stark ist die Korrelation zwischen Kosinus-Distanzen und Kompositionalitätsbewertungen (*Spearman Rank-Order Correlation, (r_s)*)?

Fenstergröße	Korpus						
	HGC	Wiki	WebKo			sdeWaC	
			Teile	ext	int	int	
	200	430	200	1,500	1,500	880	
1	.2611	.2683	.1138	.1543	.2017	.2168	.3779
2	.3133	.3001	.1614	.1867	.2322	.2411	.4469
5	.2831	.3072	.2102	.2435	.3268	.3118	.4624
10	.2407	.3336	.2639	.2808	.3664	.3299	.4703
20	.2214	.3549	.2999	.3024	.4065	.3306	.4742

Nomen-Komposita: Modellierung im Korpus-Vergleich

Frage: Wie stark ist die Korrelation zwischen Kosinus-Distanzen und Kompositionalitätsbewertungen (*Spearman Rank-Order Correlation, (r_s)*)?

Fenstergröße	Korpus						
	HGC	Wiki	WebKo			sdeWaC	
			Teile	ext	int	int	
	200	430	200	1,500	1,500	880	
1	.2611	.2683	.1138	.1543	.2017	.2168	.3779
2	.3133	.3001	.1614	.1867	.2322	.2411	.4469
5	.2831	.3072	.2102	.2435	.3268	.3118	.4624
10	.2407	.3336	.2639	.2808	.3664	.3299	.4703
20	.2214	.3549	.2999	.3024	.4065	.3306	.4742

Nomen-Komposita: Modellierung im Korpus-Vergleich

Frage: Wie stark ist die Korrelation zwischen Kosinus-Distanzen und Kompositionalitätsbewertungen (*Spearman Rank-Order Correlation, (r_s)*)?

Fenstergröße	Korpus						
	HGC	Wiki	WebKo			sdeWaC	
			Teile	ext	int		int
	200	430	200	1,500	1,500	880	
1	.2611	.2683	.1138	.1543	.2017	.2168	.3779
2	.3133	.3001	.1614	.1867	.2322	.2411	.4469
5	.2831	.3072	.2102	.2435	.3268	.3118	.4624
10	.2407	.3336	.2639	.2808	.3664	.3299	.4703
20	.2214	.3549	.2999	.3024	.4065	.3306	.4742

Nomen-Komposita: Modellierung im Korpus-Vergleich

Frage: Wie stark ist die Korrelation zwischen Kosinus-Distanzen und Kompositionalitätsbewertungen (*Spearman Rank-Order Correlation, (r_s)*)?

Fenstergröße	Korpus						
	HGC	Wiki	WebKo			sdeWaC	
			Teile	ext	int		int
	200	430	200	1,500	1,500	880	
1	.2611	.2683	.1138	.1543	.2017	.2168	.3779
2	.3133	.3001	.1614	.1867	.2322	.2411	.4469
5	.2831	.3072	.2102	.2435	.3268	.3118	.4624
10	.2407	.3336	.2639	.2808	.3664	.3299	.4703
20	.2214	.3549	.2999	.3024	.4065	.3306	.4742

Kompositionalität und Korpus-Kookkurrenz: Datenbeispiele

Vergleich von Kosinus-Distanzen und Kompositionalitätsbewertungen:

Kompositum	Modifikator	Kopf	Modifikator		Kopf	
			Kosinus	Wertung	Kosinus	Wertung
<i>Blockflöte</i>	<i>Block</i>	<i>Flöte</i>	.18	1.73	.70	6.57
<i>Feuerwerk</i>	<i>Feuer</i>	<i>Werk</i>	.41	4.20	.48	2.80
<i>Fliegenpilz</i>	<i>Fliege</i>	<i>Pilz</i>	.17	1.93	.25	6.55
<i>Obstkuchen</i>	<i>Obst</i>	<i>Kuchen</i>	.11	4.80	.40	5.93
<i>Schlittenhund</i>	<i>Schlitten</i>	<i>Hund</i>	.47	4.30	.60	5.33
<i>Telefonzelle</i>	<i>Telefon</i>	<i>Zelle</i>	.47	6.13	.20	3.23
<i>Wasserhahn</i>	<i>Wasser</i>	<i>Hahn</i>	.73	6.10	.43	2.27

Akquisition lexikalisch-semantischen Wissens

- Korpusdaten enthalten semantische Information, die auch mit einfachsten, Fenster-basierten Kookkurrenz-Verfahren akquiriert werden kann.
- Kein Korpus enthält alle gewünschten semantischen Informationen.
- Verschiedene Korpora sind komplementär in der Art der semantischen Information.
- Gut aufbereitete Webkorpora sind geeignet (und vielleicht sogar optimal) für die Akquisition von lexikalisch-semantischer Information.

Eigenschaften von Korpora

- **Domäne(n)** des Korpus: Domänenvielfalt gewünscht
- **Größe** des Korpus: wichtig, aber nicht alleine ausschlaggebend
- **Aufbereitung** des Korpus: sehr wichtig
- **Status von Webkorpora**: *Is more data always better?*
Die Größe und die Domänenvielfalt von Webkorpora sind wichtig, aber erst nach guter Aufbereitung wertvoll!

Kollegen

- Susanne Borgwaldt (Braunschweig/Erfurt)
- Ronny Jauch (Stuttgart)
- Alissa Melinger (Dundee)
- Stefan Müller (Stuttgart)
- Michael Roth (Saarbrücken → Heidelberg)

Referenzen: Webkorpora



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.

Referenzen: Distributionelle Hypothese



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.

Referenzen: Webkorpora



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*, Montreal, Canada, 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.

Referenzen: Assoziationsnormen



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.

Referenzen: Assoziationsnormen



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.

Referenzen: Assoziationsnormen



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.

Referenzen: Kookkurrenz-Hypothese



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.

Referenzen: Analysen von Assoziationsnormen



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.

Referenzen: Analysen von Assoziationsnormen



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.

Referenzen: Nomen-Komposita



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.

Referenzen: Kompositionalität von Komposita



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, France, 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.