



Distributionelle Modellierung von semantischen Beziehungen: Potential und Grenzen

PD Dr. Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung (IMS)
Universität Stuttgart

Antrittsvorlesung

9. Juli 2014



Daten-intensive Lexikalische Semantik

Computerlinguistik,

Theoretische Linguistik, Kognitive Linguistik, Künstliche Intelligenz

Interdisziplinärer Ansatz



Daten-intensive Lexikalische Semantik

Computerlinguistik,

Theoretische Linguistik, Kognitive Linguistik, Künstliche Intelligenz

Interdisziplinärer Ansatz

Semantik

Bedeutungslehre



Daten-intensive Lexikalische Semantik

Computerlinguistik,

Theoretische Linguistik, Kognitive Linguistik, Künstliche Intelligenz

Interdisziplinärer Ansatz

Semantik

Bedeutungslehre

Lexikalische Semantik

Bedeutungen & Bedeutungsbeziehungen von (komplexen) Wörtern



Daten-intensive Lexikalische Semantik

Computerlinguistik,

Theoretische Linguistik, Kognitive Linguistik, Künstliche Intelligenz

Interdisziplinärer Ansatz

Semantik

Bedeutungslehre

Lexikalische Semantik

Bedeutungen & Bedeutungsbeziehungen von (komplexen) Wörtern

Daten-intensive Lexikalische Semantik

... mit Hilfe großer Datenmengen und Statistiken



Semantische Beziehungen



Semantische Beziehungen

Semantische Beziehungen/Semantic Relatedness:

Two words are semantically related if they have any kind of semantic relation. (Budanitsky and Hirst, 2006)



Semantische Beziehungen

Semantische Beziehungen/Semantic Relatedness:

Two words are semantically related if they have any kind of semantic relation. (Budanitsky and Hirst, 2006)

Phänomene:



Semantische Beziehungen

Semantische Beziehungen/Semantic Relatedness:

Two words are semantically related if they have any kind of semantic relation. (Budanitsky and Hirst, 2006)

Phänomene:

- **Paradigmatische semantische Relationen**
(Deutsch; Englisch; Italienisch)



Semantische Beziehungen

Semantische Beziehungen/Semantic Relatedness:

Two words are semantically related if they have any kind of semantic relation. (Budanitsky and Hirst, 2006)

Phänomene:

- **Paradigmatische semantische Relationen**
(Deutsch; Englisch; Italienisch)
- **Kompositionalität von deutschen Mehrwortausdrücken**
(Nomen-Komposita; Partikelverben)



Semantische Beziehungen

Semantische Beziehungen/Semantic Relatedness:

Two words are semantically related if they have any kind of semantic relation. (Budanitsky and Hirst, 2006)

Phänomene:

- **Paradigmatische semantische Relationen**
(Deutsch; Englisch; Italienisch)
- **Kompositionalität von deutschen Mehrwortausdrücken**
(Nomen-Komposita; Partikelverben)
- **Polysemie von deutschen Präpositionen**
(sowie Verben, Adjektiven, Nomen)



Semantische Beziehungen im Raum

Komposita-Beispiele:



Semantische Beziehungen im Raum

Komposita-Beispiele:





Semantische Beziehungen im Raum

Komposita-Beispiele:





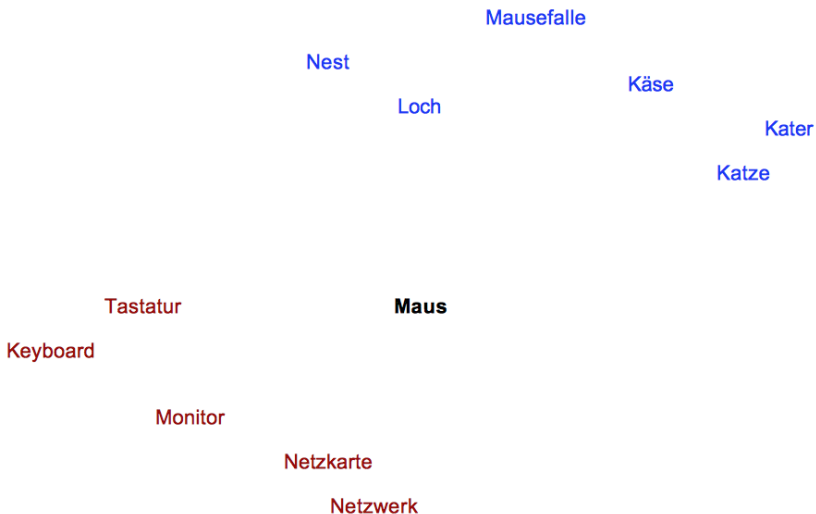
Ambiguität



Ambiguität im Raum – Klassisches Beispiel



Ambiguität im Raum – Klassisches Beispiel





Ambiguität im Raum – Kompositum



Ambiguität im Raum – Kompositum





Distributionelle Modelle



Distributionelle Modelle – Motivation

- Welches Wort verbirgt sich hier? Ergänzen Sie ein Zeichen.

AP EL



Distributionelle Modelle – Motivation

- Welches Wort verbirgt sich hier? Ergänzen Sie ein Zeichen.

AP EL → APFEL



Distributionelle Modelle – Motivation

- Welches Wort verbirgt sich hier? Ergänzen Sie ein Zeichen.

AP EL → APFEL

MÜT E



Distributionelle Modelle – Motivation

- Welches Wort verbirgt sich hier? Ergänzen Sie ein Zeichen.

AP EL → APFEL

MÜT E → MÜTZE



Distributionelle Modelle – Motivation

- Welches Wort verbirgt sich hier? Ergänzen Sie ein Zeichen.

AP EL → APFEL

MÜT E → MÜTZE

- Welches Wort verbirgt sich hier? Ergänzen Sie ein Wort.

Oma einen Kuchen



Distributionelle Modelle – Motivation

- Welches Wort verbirgt sich hier? Ergänzen Sie ein Zeichen.

AP EL → APFEL

MÜT E → MÜTZE

- Welches Wort verbirgt sich hier? Ergänzen Sie ein Wort.

Oma einen Kuchen → Oma backt einen Kuchen



Distributionelle Modelle – Hypothese

Distributionelle Hypothese/**Distributional Hypothesis**:

Each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts. (Harris, 1954)

You shall know a word by the company it keeps. (Firth, 1957)



Distributionelle Modelle – Hypothese

Distributionelle Hypothese/Distributional Hypothesis:

Each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts. (Harris, 1954)

You shall know a word by the company it keeps. (Firth, 1957)

Distributionelle Quelle:

- **Korpus-Kookkurrenz:**
Beschreibung von linguistischen Einheiten
- **Lexikalische Beschreibung** (idealisierte Vorstellung):
Summe von Kontext-Beschreibungen



Potential und Grenzen

- Quelle von distributioneller Information: [Korpora](#)



Potential und Grenzen

- Quelle von distributioneller Information: [Korpora](#)
- Arten von Korpus-Kookkurrenz:
 - Wörter in einem festen Wortfenster, im Satz, Paragraphen, Dokument
 - Syntax-basierte Kookkurrenz, ...



Potential und Grenzen

- Quelle von distributioneller Information: [Korpora](#)
- Arten von Korpus-Kookkurrenz:
 - Wörter in einem festen Wortfenster, im Satz, Paragraphen, Dokument
 - Syntax-basierte Kookkurrenz, ...
- Darstellung von Kookkurrenz: Matrizen/Vektoren



Potential und Grenzen

- Quelle von distributioneller Information: **Korpora**
- Arten von Korpus-Kookkurrenz:
 - Wörter in einem festen Wortfenster, im Satz, Paragraphen, Dokument
 - Syntax-basierte Kookkurrenz, ...
- Darstellung von Kookkurrenz: Matrizen/Vektoren
- Problem: Korpus enthält kein **Weltwissen**;
inferential (i.e., how to use language distributionally) vs. referential (incorporating world knowledge) abilities
(Marconi, 1997), z.B. **Ananas-gelb**; **auftauen-Wasser**



Modelle von Semantischen Beziehungen

- **Distributionelle Information**
 - Potential und Grenzen
 - Erweiterungen und Alternativen
- **Hervortretende distributionelle Merkmale/**
Salient distributional features
 - Standard-Merkmale
 - Phänomen-bezogene Merkmale
- **Ambiguität in Vektorräumen**
 - Vektorräume fassen Wortbedeutungen zusammen
 - Definition von Vektor-Regionen
 - Identifikation von ambigen Objekten
 - Charakterisierung von Bedeutungs-bezogenen Merkmalen



Modelle von Semantischen Beziehungen

1. [Test Case]
Enthält distributionelle Information wirklich Semantik?
→ semantische Assoziationen in Korpora
2. [Show Case]
Wie wendet man ein distributionelles Modell für semantische Beziehungen an?
→ Kompositionalitätsgrad von deutschen Nomen-Komposita
3. [Application Case]
Kann man distributionelle semantische Information in eine Anwendung integrieren?
→ Subkategorisierung in maschineller Übersetzung



Motivation: Semantik in Korpora

BLA Schlitten BLA BLA BLA BLA
BLA BLA BLA Schnee BLA BLA
BLA BLA weiß BLA BLA BLA
BLA BLA BLA BLA BLA Winter



Motivation

- Enthält distributionelle Information semantische Information?



Motivation

- Enthält distributionelle Information semantische Information?
- *Test case*: Assoziationen stehen in einer semantischen Beziehung zu ihren Stimuli.
(Nelson et al., (1997; 2000); McNamara (2005); i.a.)



Motivation

- Enthält distributionelle Information semantische Information?
- *Test case*: Assoziationen stehen in einer semantischen Beziehung zu ihren Stimuli.
(Nelson et al., (1997; 2000); McNamara (2005); i.a.)
- Assoziationen von *Schnee*



Motivation

- Enthält distributionelle Information semantische Information?
- *Test case*: Assoziationen stehen in einer semantischen Beziehung zu ihren Stimuli.
(Nelson et al., (1997; 2000); McNamara (2005); i.a.)
- Assoziationen von *Schnee* → *weiß*, *Winter*, *Schlitten*, ...



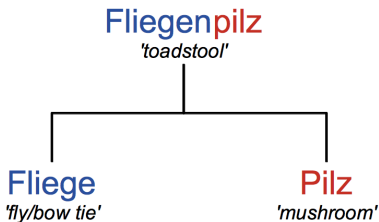
Motivation

- Enthält distributionelle Information semantische Information?
- *Test case*: Assoziationen stehen in einer semantischen Beziehung zu ihren Stimuli.
(Nelson et al., (1997; 2000); McNamara (2005); i.a.)
- Assoziationen von *Schnee* → *weiß*, *Winter*, *Schlitten*, ...
- **Distributionelle Information und semantische Beziehungen**:
Proportion von Stimulus-Assoziationspaaren in einem Korpus-Kontext-Fenster der Größe 5
(Schulte im Walde & Müller, 2013)

Deutsche Normen	Größe	Korpora			
		Zeitung	Wiki-de	SdeWaC	deWaC
		200	430	880	1.500
Verben	38.769	54%	50%	68%	73%
Nomen	30.845	53%	56%	72%	77%
Nomen-Komposita	14.326	23%	25%	42%	49%



Nomen-Nomen-Komposita





Komposita-Daten

- **Komposition:**
 - 244 konkrete, abbildbare deutsche Nomen-Nomen-Komposita; Teilmenge der Komposita: von der Heide & Borgwaldt (2009)
 - Komposita, Modifikatoren und Köpfe sind Nomen.
 - Vier Kompositionalitätsklassen (O=opak; T=transparent): O+O, T+T, O+T, T+O



Komposita-Daten

- **Komposition:**
 - 244 konkrete, abbildbare deutsche Nomen-Nomen-Komposita; Teilmenge der Komposita: von der Heide & Borgwaldt (2009)
 - Komposita, Modifikatoren und Köpfe sind Nomen.
 - Vier Kompositionalitätsklassen (O=opak; T=transparent): O+O, T+T, O+T, T+O

- **Menschliche Bewertungen:**

Komposita			Durchschnittsbewertungen		
Kompositum	wörtliche Bedeutungen		als Ganzes	Modifikator	Kopf
<i>Ahornblatt</i> 'maple leaf'	maple	leaf	6.03 ± 1.49	5.64 ± 1.63	5.71 ± 1.70
<i>Postbote</i> 'post man'	mail	messenger	6.33 ± 0.96	5.87 ± 1.55	5.10 ± 1.99
<i>Seezunge</i> 'sole'	sea	tongue	1.85 ± 1.28	3.57 ± 2.42	3.27 ± 2.32
<i>Windlicht</i> 'storm lamp'	wind	light	3.52 ± 2.08	3.07 ± 2.12	4.27 ± 2.36
<i>Löwenzahn</i> 'dandelion'	lion	tooth	1.66 ± 1.54	2.10 ± 1.84	2.23 ± 1.92
<i>Maulwurf</i> 'mole'	mouth	throw	1.58 ± 1.43	2.21 ± 1.68	2.76 ± 2.10
<i>Fliegenpilz</i> 'toadstool'	fly/bow tie	mushroom	2.00 ± 1.20	1.93 ± 1.28	6.55 ± 0.63
<i>Flohmarkt</i> 'flea market'	flea	market	2.31 ± 1.65	1.50 ± 1.22	6.03 ± 1.50
<i>Feuerzeug</i> 'lighter'	fire	stuff	4.58 ± 1.75	5.87 ± 1.01	1.90 ± 1.03
<i>Fleischwolf</i> 'meat chopper'	meat	wolf	1.70 ± 1.05	6.00 ± 1.44	1.90 ± 1.42



Modelle

1. **Distributionelle Modelle** mit textueller, Korpus-basierter Kookkurrenz (Schulte im Walde et al., 2013):
 - **Aufgabe**: Vorhersage des Kompositionalitätsgrads
 - **Teilaufgabe 1**: Vergleich von Fenster-basierten und Syntax-basierten Merkmalen
 - **Teilaufgabe 2**: Vergleich der semantischen Beiträge von Modifikatoren vs. Köpfen



Modelle

1. **Distributionelle Modelle** mit textueller, Korpus-basierter Kookkurrenz (Schulte im Walde et al., 2013):
 - **Aufgabe:** Vorhersage des Kompositionalitätsgrads
 - **Teilaufgabe 1:** Vergleich von Fenster-basierten und Syntax-basierten Merkmalen
 - **Teilaufgabe 2:** Vergleich der semantischen Beiträge von Modifikatoren vs. Köpfen
2. **Multimodales LDA-Modell** mit
 - **textuellen Daten** (Kookkurrenz)
 - **experimentellen Daten** (Assoziations- und Merkmalsnormen)
 - **visuellen Daten** (Bildern)

Roller & Schulte im Walde (2013)

 - **Aufgabe:** Vorhersage des Kompositionalitätsgrads



Distributionelle Modelle

- **Ziel:** Vektorraum-Modelle nutzen saliente distributionelle Merkmale, um den Kompositionalitätsgrad der Komposita vorherzusagen.
- **Korpora:** deutsche Webkorpora (1.500/880 Millionen Wörter)
- **Merkmalswerte:** *local mutual information* (Evert, 2005) der Kookkurrenz-Frequenzen (Ziel-Nomen und Merkmale)
- **Semantische Beziehung:** Kosinus \sim Kompositionalitätsgrad
- **Evaluation:** Kosinus gegen menschliche Bewertungen; Spearman Rank-Order Correlation Coefficient ρ (Siegel & Castellan, 1988)



Multimodales LDA-Modell

- Erweiterung des **Latent-Dirichlet-Allocation**-Modells mit ursprünglich zwei-dimensionalen *Topics* (Andrews et al., 2009)

- **Multimodale Merkmale:**
 - **Textuelle Modalität:** Wort-Dokument-*Topics* (WebKo)
 - **Kognitive Modalität:**
 1. Assoziationsnormen
 2. Merkmalsnormen
 - **Visuelle Modalität:** *BilderNetle*, Nomen-*ImageNet*-Abbildungen für Komposita und Konstituenten
 1. SURF (Cluster): wählt wichtige Punkte eines Bilds aus
 2. GIST (Cluster): berechnet hoch-dimensionalen Vektor für ein Bild



Ergebnisse

- Nomen repräsentieren die salientesten Merkmale: $\rho = .6497$
- Fenster-basierte Merkmale > Syntax-basierte Merkmale
- Unterschiedliche saliente Merkmale für Kompositionalitätsgrad von Kompositum–Modifikator vs. Kompositum–Kopf:
 - kleine Fenster:
 - Kompositum–Kopf > Kompositum–Modifikator
 - syntaktische Merkmale:
 - Kompositum–Kopf > Kompositum–Modifikator
- Einfluss der Modifikator-Bedeutung auf Kompositum ist stärker als Einfluss der Kopf-Bedeutung
- Hybrides LDA-Modell (Konkatenation von textuellen Merkmalen, Assoziationsnormen, SURF-Merkmalen und GIST-Clustern) übertrifft textuelles Modell und verschiedene 2- und 3-dimensionale LDA-Modelle



Ergebnisse

Information	Kompositum–Konstituente			als Ganzes über Funktion			
	beide	Mod.	Kopf	Mod.	Kopf	Add	Mult
Baseline und Bewertungen							
Baseline (zufällig)	-.0672	.0496	.0527	.0959	.1019	.1168	.1079
Obere Grenze (Bewertungen)				.6002	.1385	.7687	.7829
Experimentelle Daten							
Assoziationsnormen	.6227	.6128	.6547	.4470	.0264	.3297	.4462
Distributionelle Modelle							
Fenster (NN, Größe=20)	.6497	.5698	.5745	.4388	.1001	.3058	.4509
Syntax (ADJ, mod)	.3455	.2224	.3502				
LDA-Modelle							
unimodal (nur Text)	.2040						
bimodal (Text+AN)	.3280						
3D (Text+FN+SURF)	.2860						
hybrid (Text+AN+SURF+GC)	.4060						



Distributionelle Information für Maschinelle Übersetzung (SMÜ)

BLA sledge BLA BLA BLA BLA
BLA BLA BLA snow BLA BLA
BLA BLA white BLA BLA BLA
BLA BLA BLA BLA BLA winter



BLA Schlitten BLA BLA BLA BLA
BLA BLA BLA Schnee BLA BLA
BLA BLA weiß BLA BLA BLA
BLA BLA BLA BLA BLA Winter



Distributionelle Information für SMÜ

- Phrasenbasiertes statistisches Übersetzungssystem
- Problem: Übersetzung von Komposita Deutsch–Englisch
- Aufgabe: Spaltung von Komposita in Konstituenten?



Distributionelle Information für SMÜ

- Phrasenbasiertes statistisches Übersetzungssystem
- Problem: Übersetzung von Komposita Deutsch–Englisch
- Aufgabe: Spaltung von Komposita in Konstituenten?
 - *Holzzaun* → *wooden fence*



Distributionelle Information für SMÜ

- Phrasenbasiertes statistisches Übersetzungssystem
- Problem: Übersetzung von Komposita Deutsch–Englisch
- Aufgabe: Spaltung von Komposita in Konstituenten?
 - *Holzzaun* → *wooden fence*
 - *Jägerzaun* → *lattice fence*



Distributionelle Information für SMÜ

- Phrasenbasiertes statistisches Übersetzungssystem
- Problem: Übersetzung von Komposita Deutsch–Englisch
- Aufgabe: Spaltung von Komposita in Konstituenten?
 - *Holzzaun* → *wooden fence*
 - *Jägerzaun* → *lattice fence*
 - *Jägerhut* → *lattice hat* !!!



Distributionelle Information für SMÜ

- Phrasenbasiertes statistisches Übersetzungssystem
- Problem: Übersetzung von Komposita Deutsch–Englisch
- Aufgabe: Spaltung von Komposita in Konstituenten?
 - *Holzzaun* → *wooden fence*
 - *Jägerzaun* → *lattice fence*
 - *Jägerhut* → *lattice hat* !!!
- Modell: **Spaltung nur bei transparenten Komposita**;
Vorhersage des Kompositionalitätsgrads (Weller et al., 2014)
- Evaluation: marginale Effekte aber qualitativ interessant;
die meisten Komposita sind (semi-)transparent; Ansatz hilft bei
niedrigfrequenten aber kaum bei opaken Komposita



Distributionelle Information für SMÜ

- Hierarchisches statistisches Übersetzungssystem
- Übersetzung in zwei Schritten: (i) Erstellen des Übersetzungssystems mit Stämmen; (ii) Flexion der Übersetzung
- Beispiel für **Kasus-Fehler** bei Übersetzung Englisch–Deutsch:

Eingabe		[why] ₁ [the government] ₂ [ordered] ₃ [the ongoing military actions] ₄
Ausgabe	Stämme Flexion	[warum] ₁ [d Regierung] ₂ [d anhaltend militärisch Aktion] ₄ [angeordnet] ₃ [warum] ₁ [die Regierung] ₂ [der anhaltenden militärischen Aktionen] ₄ [angeordnet] ₃

- **Integration von Subkategorisierungsinformation** (Weller et al., 2014):
 - syntaktische Subkategorisierung der Quell-Sprache
 - externe Wissensbasis mit quantitativer Dependenz-Information zu Subkategorisierungsrahmen der Ziel-Sprache
- Evaluation: positive Wirkung auf die Übersetzungsqualität



Zusammenfassung

- **Distributionelle Information**
 - erfasst semantische Assoziationen,
 - unterscheidet zwischen paradigmatischen Relationen,
 - sagt den Kompositionalitätsgrad von Nomen-Komposita und Partikelverben vorher,
 - (identifiziert polyseme Objekte im Vektorraum) und
 - ist nützlich für maschinelle Übersetzung.
- **Saliente distributionelle Merkmale:**
 - Standard-Merkmale sind ein erster Schritt, aber . . .
 - für linguistische Modelle brauchen wir Phänomen-bezogene Merkmale.
- **Ambiguität in Vektorräumen:**
 - Co-Disambiguierung ist ein einfacher aber effektiver Ansatz für paarweise Disambiguierung.
 - Wo im Vektorraum befinden sich polyseme Objekte?
 - Wie erreichen wir Phänomen-übergreifende Disambiguierung?



SemRel-Team (2011–2014)

- Dr. Stefan Bott
- Max Kisselew
- Maximilian Köper
- Gabriella Lapesa (SFB/MGK)
- Stefan Müller
- Cornelius Putzler
- Stephen Roller (SFB/MGK)
- Dr. Michael Roth
- Enrico Santus (Erasmus Placement)
- Dr. Silke Scheible
- Sylvia Springorum (SFB/MGK)
- Marion Weller
- Moritz Wittmann