

# Die Repräsentation und Auflösung von ambigen Wortbedeutungen in der Computerlinguistik

## Wiederholung

PD Dr. Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung  
Universität Stuttgart

26. März 2010

- Untersuchungsgegenstand: lexikalische Bedeutung
- Bedeutung von lexikalischen Elementen und ihren semantischen Relationen
- Fokus im Seminar: (einzelne) Wörter
- Automatische Sprachverarbeitung: Computer-basierte Akquisition und Anwendung von lexikalischer Bedeutung
- Bereich: Daten-intensive lexikalische Semantik

- **Ambiguität**: mehrere Bedeutungen eines lexikalischen Elements
- Mangel (↔ Unklarheit) vs. Stilelement (↔ Ironie, Witz)
- **Desambiguierung**: Auflösung einer Ambiguität

- Erkennen und Auflösen einer Ambiguität
- Anzahl und Auswahl der Bedeutungen:  
Bezug auf Bedeutungsinventar
- Faktoren mit Einfluss auf die Wortbedeutung:
  - Kontext, Kollokationen, Syntagma
  - Domäne
  - Diskurs
  - Prototypikalität
  - Weltwissen

- **Lexikon**: Bestandteil einer linguistischen Theorie, die sich mit den Eigenschaften von Lexemen befaßt (Engelberg & Lemnitzer, 2004)
- Ressourcen zu lexikalischem Wissen:
  - **manuelle Ressourcen**, z.B. Wörterbücher, Thesauri, Taxonomien, Ontologien, Klassifikationen etc.
  - **automatische Erstellung von Ressourcen** auf der Basis von (annotierten) Korpus-Daten

- **Korpus**: (große) Sammlung von Äußerungen
- Empirischer Ansatz zur Linguistik
- **Annotation**: linguistische Beschreibung der Korpusdaten

↔ Ressource zu lexikalischem Wissen

- Typen und Token
- absolute Häufigkeit
- relative Häufigkeit
- Rangplatz aufgrund von Häufigkeit
- Häufigkeiten von Sequenzen

# Thematische Übersicht

- 1 Wortbedeutung und Ambiguität, Bedeutungsinventare
- 2 Ressourcen zur Wortbedeutung
- 3 Korpora und Korpus-Annotation
- 4 Korpus-Statistik
  
- 5 Distributionelle Semantik
- 6 Klassifikation als Repräsentationsform von Wortbedeutung
- 7 Automatische Ansätze zur Desambiguierung
- 8 Evaluierung
  
- 9 Übungen: Wortverteilungen und Konkordanzen; Annotation von Wortbedeutungen

# Termine

Fr	15.01.2010	9.00–13.00	C5.2/1.28
		14.00–17.00	C5.2/1.28
Sa	16.01.2010	9.00–13.00	C5.3/2.06
		14.00–15.30	Übung, CoLi-Pool
<hr/>			
Fr	26.03.2010	9.00–13.00	C5.3/2.06
		14.00–17.00	C5.3/2.06
Sa	27.03.2010	9.00–13.00	C5.3/2.06
		14.00–15.30	Übung, CoLi-Pool
<hr/>			
	???		Klausur