

Computational Lexical Semantics

Zusammenfassung

Timm Lichte & Younes Samih

HHU Düsseldorf

13. Sitzung, 19. Juli 2018



- 1 Worum ging's?
- 2 Dimensionen
- 3 Seminarüberblick
- 4 Fragestellungen und Herausforderungen
- 5 Seminarkritik

Worum ging's?

Aus der (modifizierten) Kursbeschreibung:

Das Lexikon enthält diejenigen minimalen Form-Bedeutung-Einheiten, die für die Bedeutungskomposition notwendig sind.

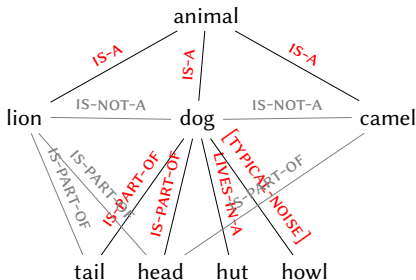
- 1 Welche Einheiten enthält das Lexikon?
⇒ Morpheme, **Worte** (Wortformen, Lexeme), Phrasen, ...
- 2 Was ist “die Bedeutung” einer lexikalischen Einheit?
⇒ Problem: Mehrdeutigkeit (Polysemie, Homonymie),
Abstraktheit (**Freiheit, der**), Vagheit (**rot**)
- 3 Wie repräsentieren wir (lexikalische) Bedeutung?
⇒ Paraphrasen, logische Formeln, Merkmalsstrukturen, Typenhierarchien, Ontologien, Vektoren, ...

In diesem Seminar

eine **Auswahl** (2 SWS!) computerlinguistischer “Antworten”

Repräsentation: kontextuell versus komponentiell

kontextuell: Die Bedeutung einer lexikalischen Einheit wird durch das Verhältnis zu anderen lexikalischen Einheiten definiert.



Menge der Relationen?

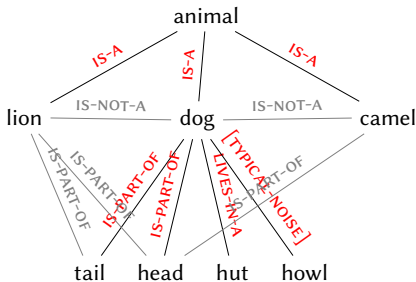
komponentiell: Die Bedeutung einer lexikalischen Einheit wird durch die Konfiguration semantischer Komponenten definiert.

- filly = [HORSE] [FEMALE] [YOUNG]
- boy = [HUMAN] [MALE] [YOUNG]
- kill = [CAUSE] [BECOME] [NOT] [ALIVE]
- chair = [OBJECT] [FURNITURE] [FOR SITTING]
[FOR ONE PERSON] [WITH BACK]

Menge der Komponenten/Atome?

Repräsentation: paradigmatische versus syntagmatische Relationen

paradigmatische Relationen: zwischen lexikalischen Einheiten in einer Menge von Sätzen (nicht direkt beobachtbar) \Rightarrow WSD



syntagmatische Relationen: zwischen lexikalischen Einheiten in einem Satz (direkt beobachtbar) \Rightarrow SRL

- (1) [Arg_0 He] [$ArgM-MOD$ would][$ArgM-NEG$ n't] accept [Arg_1 anything of value] [Arg_2 from those he was writing about].

- 1 Theorie: Lexikon & lexikalische Semantik, Verbklassen
- 2 Lexikalische Ressourcen: WordNet, FrameNet, VerbNet
- 3 Semantisch annotierte Korpora: SemCor, PropBank, OntoNotes
- 4 Anwendung: Word Sense Disambiguation (Word Similarity, Synonym Selection, ...)
 - Lesk-Algorithmus und WordNet (Banerjee & Pedersen 2002)
 - Lexical Choice bei Fast-Synonymen (Edmonds & Hirst 2002)
 - Evaluation von WordNet-basierten Ansätzen (Raganato, Camacho-Collados & Navigli 2017)
 - Neuronaler Ansatz (Popov 2018)
- 5 Anwendung: Semantic Role Labeling
 - Problem der Domänenadaption (Pradhan, Ward & Martin 2008)
 - Neuronaler Ansatz (Zhou & Xu 2015)
- 6 Verknüpfung mit distributionellen Vektoren
 - Retrofitting (Faruqui et al. 2015)
 - AutoExtend (Rothe & Schütze 2017)

Der wichtigste Gegensatz:

- manuell (WordNet) vs. automatisch (distributionelle Vektoren)

Herausforderungen:

- Mehrdeutigkeit: Repräsentation und Auflösung
 - Mehrworteinheiten: “irreguläre” lexikalische Einheiten
 - distributionelle Semantik für Nicht-Worte: Komposition von distributionellen Vektoren
 - Multimodalität: Bild und Ton als Teil der lexikalischen Information
- ⇒ ganzheitliche semantische Repräsentation?

Wissen:

- Kennenlernen verschiedener computerlinguistischer Aspekte der lexikalischen Semantik

Fähigkeit:

- wissenschaftliche Texte lesen
- über wissenschaftliche Texte referieren
- über wissenschaftliche Texte diskutieren
- über wissenschaftliche Texte schreiben

X tun lernt man nur, indem man X tut.

Kritikpunkt: Textauswahl

- Zuviel/zuwenig Theorie?
- Zuviel/zuwenig Praxis (z.B. in Form von Programmieraufgaben)?
- Form und Anzahl der Referate?
- Was sollte bei einem **4-stündigem** Kurs hinzukommen oder verstärkt behandelt werden?
- Was haben Sie (nicht) gelernt?

Inhalt:

- Fragestellung → Beantwortung durch Literatur → eigener Standpunkt
- Aufgaben
 - Eingrenzung, Abgrenzung des Themas
 - Darstellung von Frage und Antwort
 - Bewertung, eigener Standpunkt, eigene Ideen?
 - wissenschaftlicher Stil: richtiges Zitieren, Quellenangaben, ...
- imaginiertes Publikum: Kommilitonen, die den Kurs vielleicht nicht besucht haben (Dozent)

Form:

- Bestandteile: Titelseite, Inhaltsverzeichnis, Einleitung, Hauptteil, Schluss, Literaturverzeichnis
- Umfang: ca. 15 Seiten
- Textprozessor: MS-Word, OO-Writer, \LaTeX
- Abgabe: PDF

Freischuss: spätestens drei Wochen vor Abgabe

- [1] Banerjee, Satanjeev & Ted Pedersen. 2002. An adapted Lesk algorithm for word sense disambiguation using WordNet. In Alexander Gelbukh (ed.), *Computational linguistics and intelligent text processing (CICLing 2002)* (Lecture Notes in Computer Science 2276), 136–145. Berlin: Springer.
- [2] Edmonds, Philip & Graeme Hirst. 2002. Near-synonymy and lexical choice. *Computational Linguistics* 28(2). 105–144.
- [3] Faruqi, Manaal, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy & Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1606–1615.
- [4] Popov, Alexander. 2018. Neural network models for word sense disambiguation: An overview. *Cybernetics and Information Technologies* 18(1).
- [5] Pradhan, Sameer S., Wayne Ward & James H. Martin. 2008. Towards robust semantic role labeling. *Computational Linguistics* 34(2). 289–310.
- [6] Raganato, Alessandro, Jose Camacho-Collados & Roberto Navigli. 2017. Word sense disambiguation: A unified evaluation framework and empirical comparison. In *Proceedings of the 15th conference of the European Chapter of the Association for Computational Linguistics: Volume 1, long papers*, 99–110.
- [7] Rothe, Sascha & Hinrich Schütze. 2017. AutoExtend: Combining word embeddings with semantic resources. *Computational Linguistics* 43(3). 593–617.

- [8] Zhou, Jie & Wei Xu. 2015. End-to-end learning of semantic role labeling using recurrent neural networks. In *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long papers)*, 1127–1137. Beijing, China: Association for Computational Linguistics.
<http://www.aclweb.org/anthology/P15-1109>.