

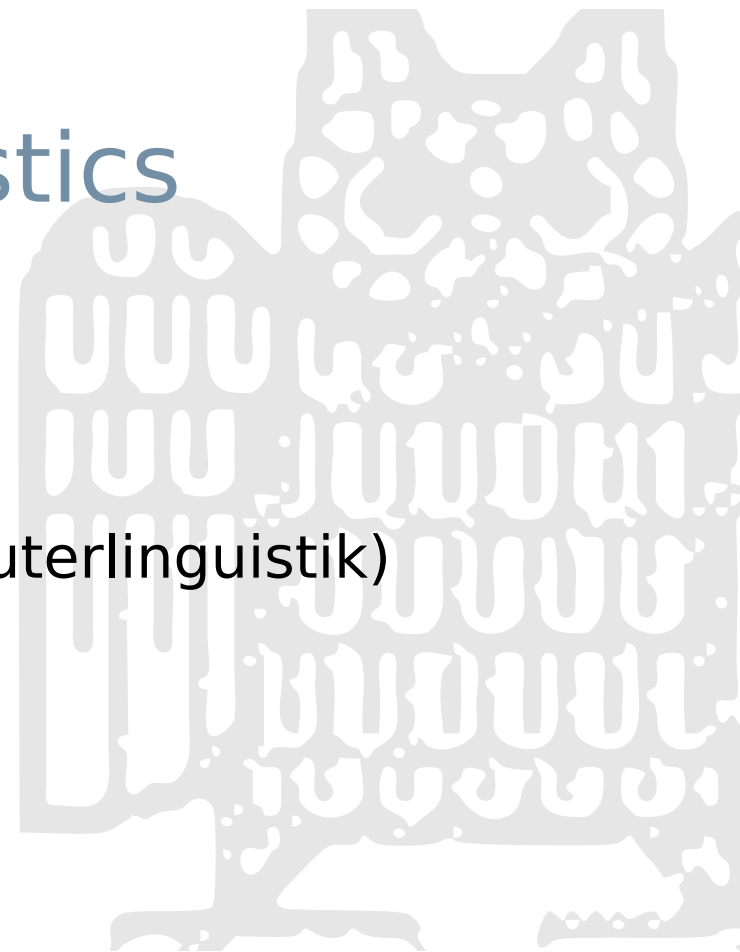
Computational Linguistics

Dietrich Klakow & Stefan Thater

FR 4.7 Allgemeine Linguistik (Computerlinguistik)

Universität des Saarlandes

Summer 2014



This Lecture

- **Lectures:**
 - Tuesday, 14:15 – 15:45
- **Exercise sessions:**
 - Thursday, 14:15 – 15:15
- **Lecturers:**
 - Stefan Thater (Week 1–7)
 - Dietrich Klakow (Week 8–12)

Course website

- Lecture slides and exercise sheets can be obtained from the course website:
 - <http://www.coli.uni-saarland/courses/CL/2014>
- Please subscribe to the follow mailing list:
 - <http://ml.coli.uni-saarland.de/cgi-bin/mailman/listinfo/compling>

Exercise sheets

- Assignments will be given after every lecture
 - You will (usually) have 1 week to submit your solutions
 - Exercise sheets will usually consists of theoretical questions as well as programming tasks
 - You have to get at least 50% of the points to be admitted to the final exam
- Programming tasks:
 - we recommend (and support) Python
 - consult with us if you prefer another language

Final Exam

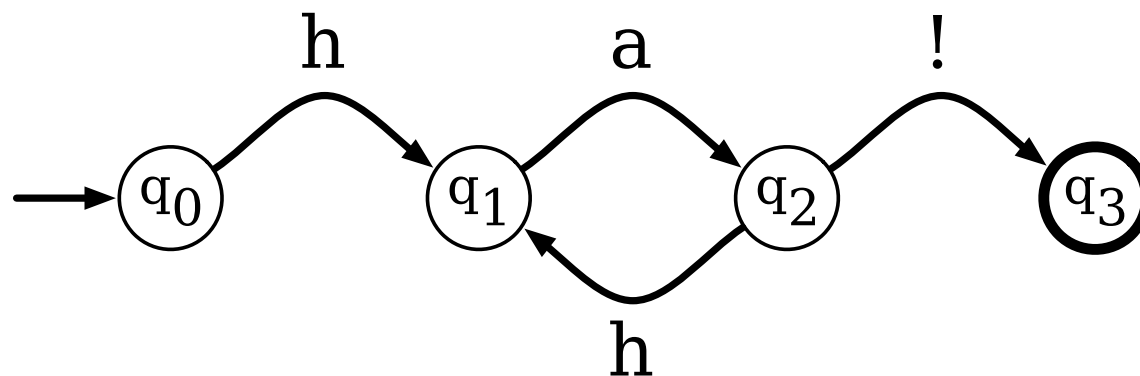
- The final exam will take place
 - Thursday, 2014-07-17 (to be confirmed)
- Registration deadline: **July 7th**

Textbooks

- D. Jurafsky and J. H. Martin: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall, 2009.
- C. Manning and H. Schütze: Foundations of Statistical Natural Language Processing, MIT Press, 1999.

Finite State Automata (Thater)

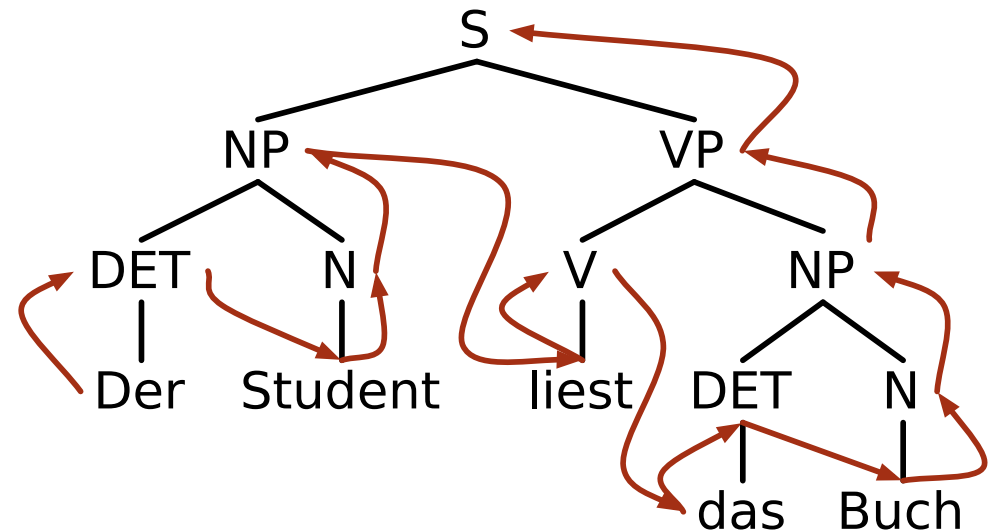
- Deterministic & non-deterministic finite state automata
- Algorithms for recognizing strings
- Algorithm for determinizing non-deterministic automata



ha!
haha!
hahaha!
hahahaha!
...

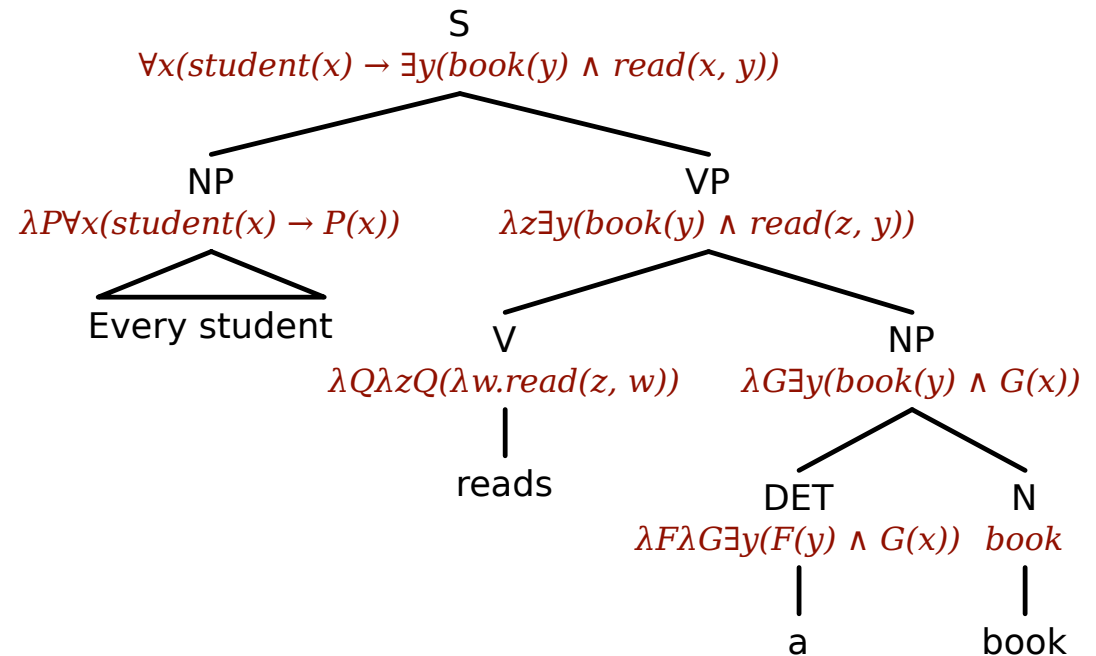
Grammars & Parsing (Thater)

- Context free grammars
- Bottom up parsing (shift-reduce)
- Chart parsing (CYK)
- Probabilistic grammars and parsing
- Dependency parsing



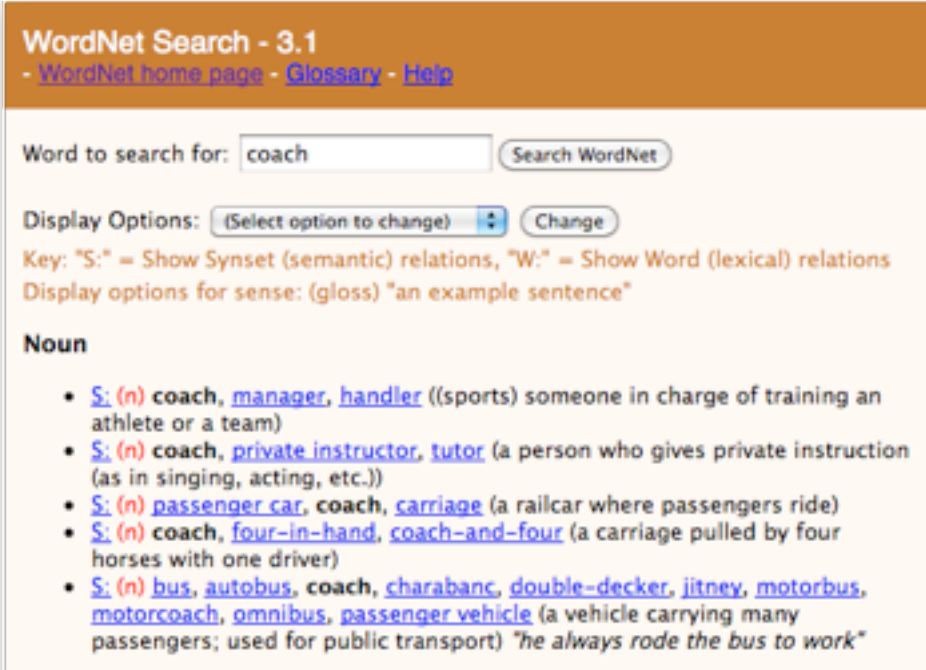
Computational Semantics (Thater)

- Semantics construction
- Scope underspecification
- Dominance graphs & Tree Automata



Word Sense Disambiguation (Thater)

- *Wladyslaw Stachurski was named Monday as Poland's new **coach** following the team's elimination from the European Nations Cup.*
- *The powerful bombs, set off by a timer, exploded inside a **coach** reserved for army and paramilitary personnel, starting a major blaze and derailing three adjoining cars.*



WordNet Search - 3.1
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

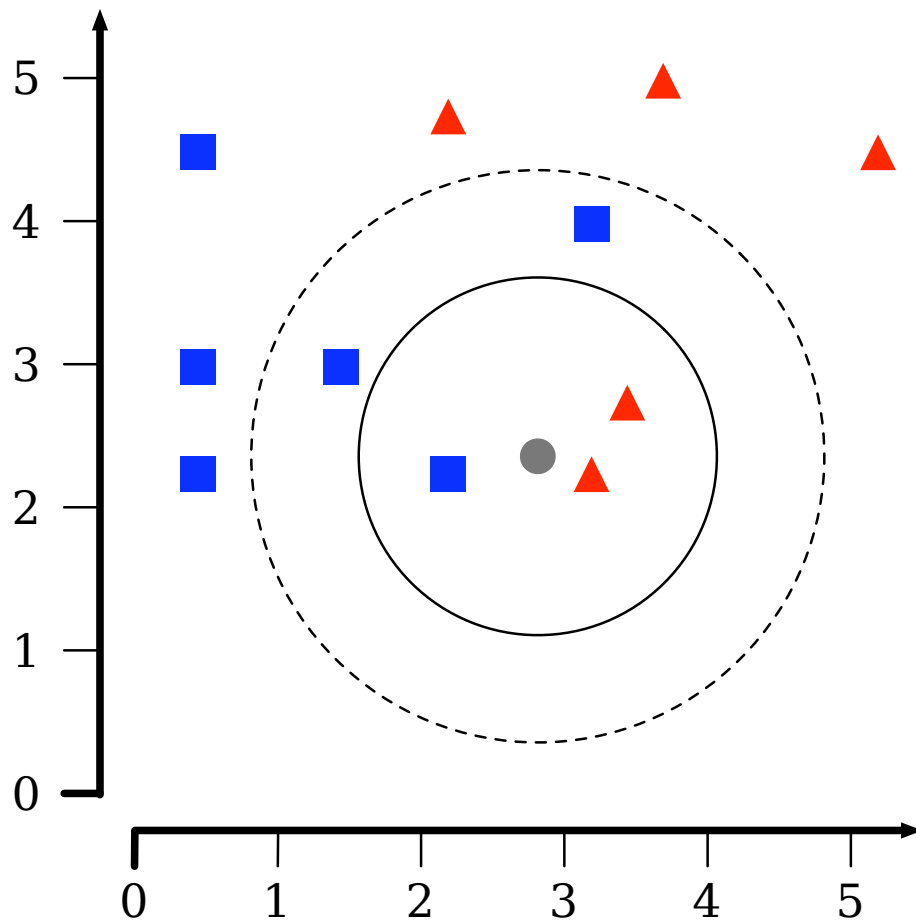
Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

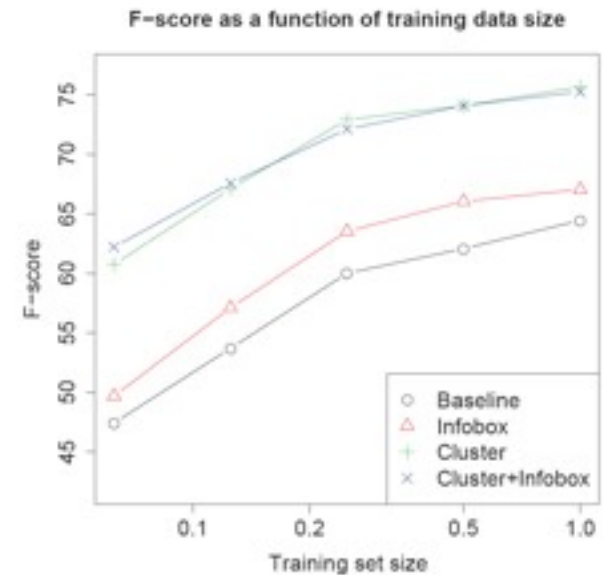
Noun

- **S:** (n) [coach](#), [manager](#), [handler](#) ((sports) someone in charge of training an athlete or a team)
- **S:** (n) [coach](#), [private instructor](#), [tutor](#) (a person who gives private instruction (as in singing, acting, etc.))
- **S:** (n) [passenger car](#), [coach](#), [carriage](#) (a railcar where passengers ride)
- **S:** (n) [coach](#), [four-in-hand](#), [coach-and-four](#) (a carriage pulled by four horses with one driver)
- **S:** (n) [bus](#), [autobus](#), [coach](#), [charabanc](#), [double-decker](#), [jitney](#), [motorbus](#), [motorcoach](#), [omnibus](#), [passenger vehicle](#) (a vehicle carrying many passengers; used for public transport) "he always rode the bus to work"

Clustering (Klakow)

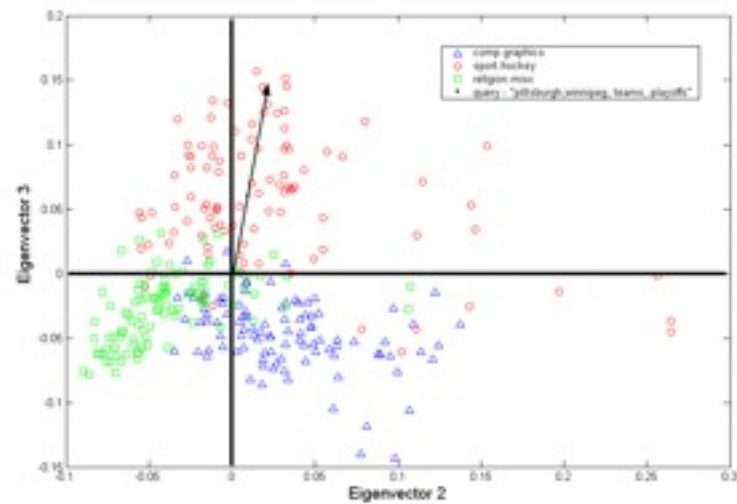
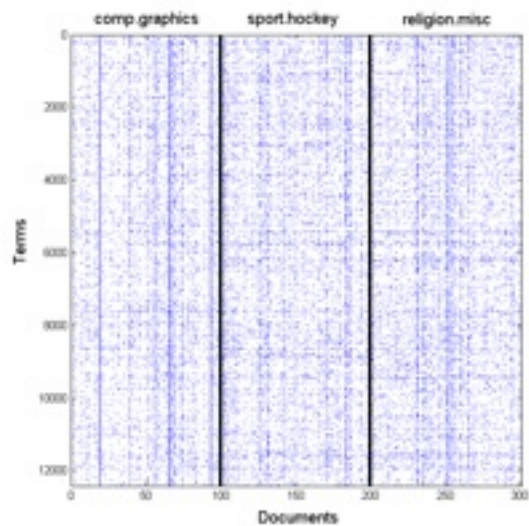


Cluster	Example members
1	Groß, Rau, Müller, Zimmermann, Frei, Becker, Möllemann, Schmidt
2	Düsseldorf, Berlin, München, Köln, Stuttgart, Hannover, Hamburg
3	nahmen, macht, zeigt, gleichen, bringt, biete, machte, sorgt, enthält



Matrix Factorization (Klakov)

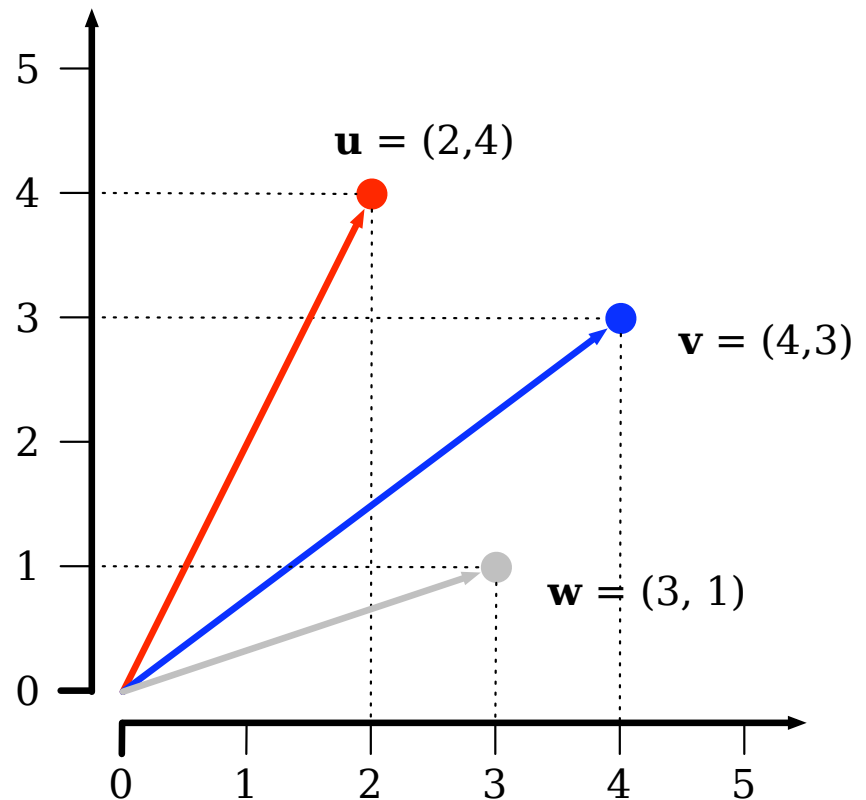
- Latent semantic analysis (LSA)



- Non-negative matrix factorization (NMF)

Distributional Semantics (Klakov)

- Distributional Semantics
- Vector space models
- Similarity measures



Combining Visual & Textual Data (Klakow)

Weather News

1



2



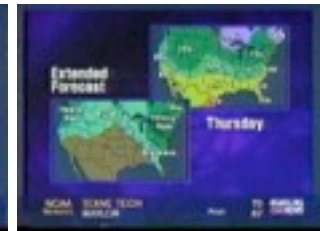
3



4



5



Football



Horse



Combining Visual & Textual Data (Klakow)

Weather News

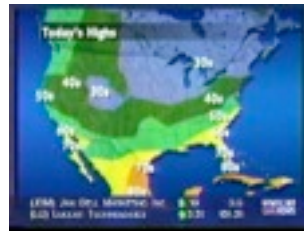
1



2



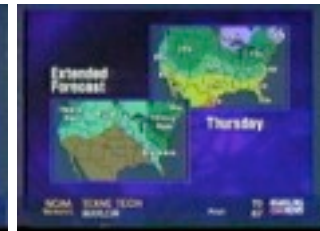
3



4



5



Football



Horse



Alternative:
caption generation