Computational Linguistics

Clayton Greenberg
Stefan Thater

FR 4.7 Allgemeine Linguistik (Computerlinguistik)
Universität des Saarlandes
Summer 2015

This Lecture

- **Lectures:**
  - Tuesday, 14:15 – 15:45
- **Exercise sessions:**
  - Thursday, 14:15 – 15:15
- **Lecturers:**
  - Stefan Thater (1st half)
  - Clayton Greenberg (2nd half)

Course website

- Lecture slides and exercise sheets can be obtained from the course website:
  - **http://www.coli.uni-saarland/courses/CL/2015**
- 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 theoreticall 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

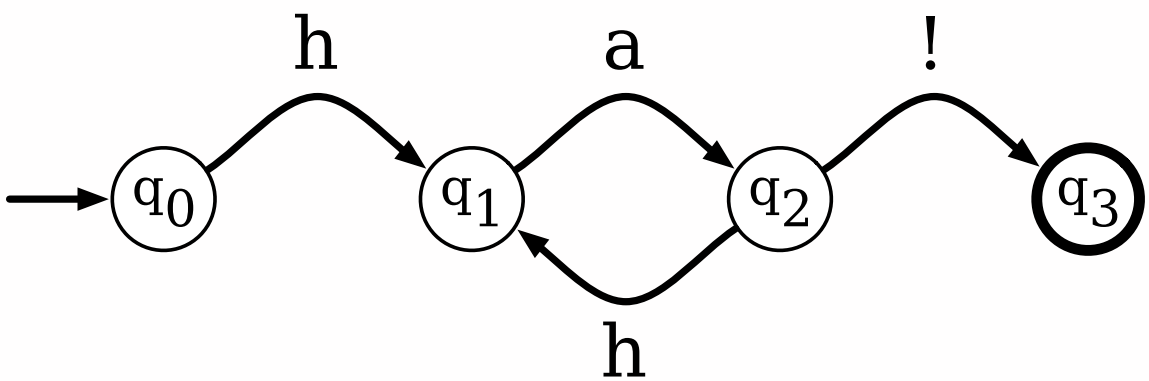
- The final exam will take place
  - Tuesday, 2015-07-28 (to be confirmed)
- Registration deadline: **July 15th**

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)

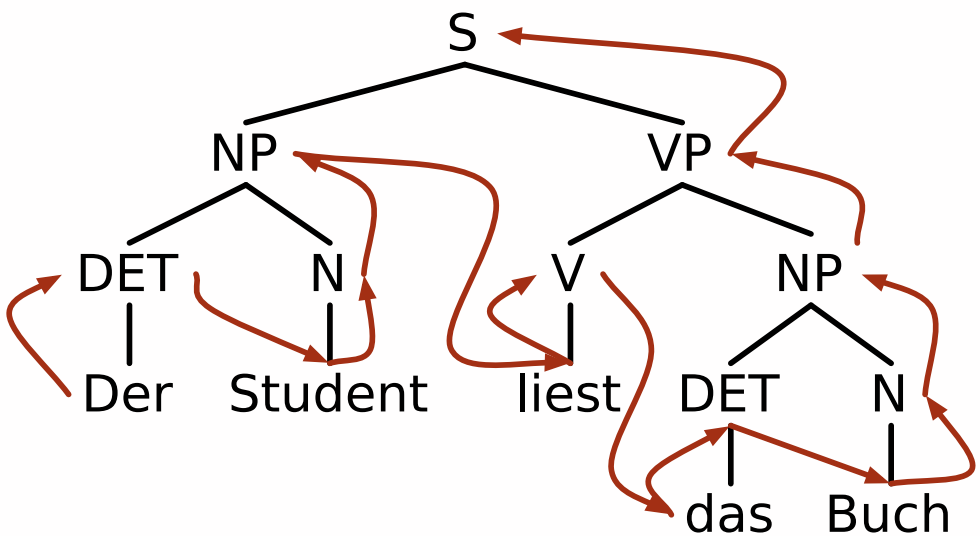
- Deterministic & non-deterministic finite state automata
- Algorithms for recognizing strings
- Algorithm for determinzing non-deterministic automata



ha!
haha!
hahaha!
hahahaha!
. . .

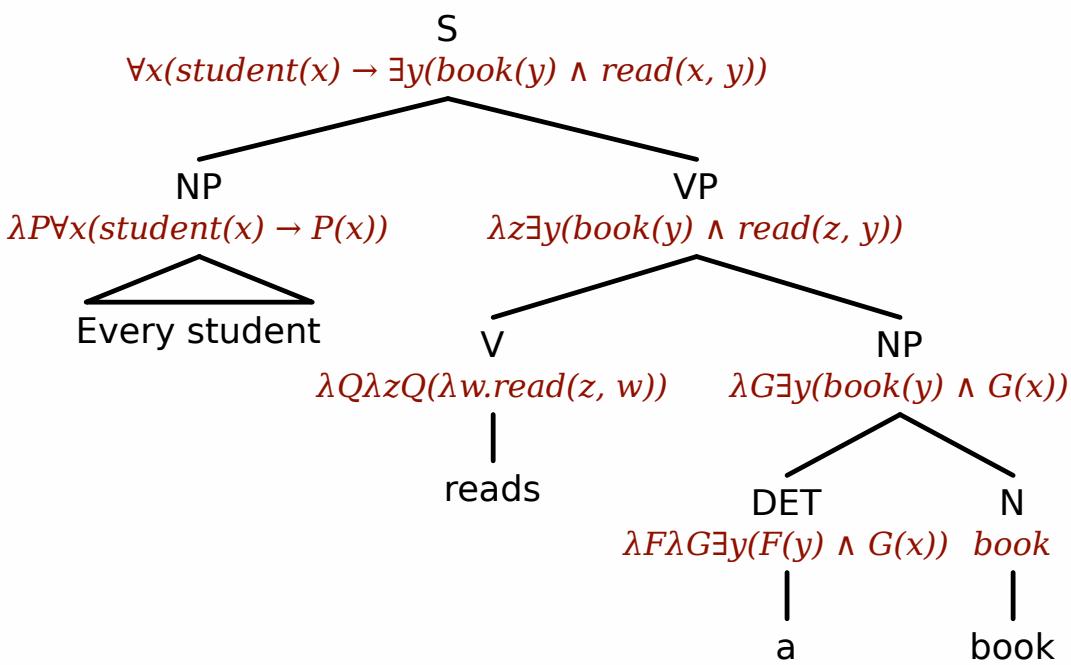
Grammars & Parsing (Thater)

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



Computational Semantics (Thater)

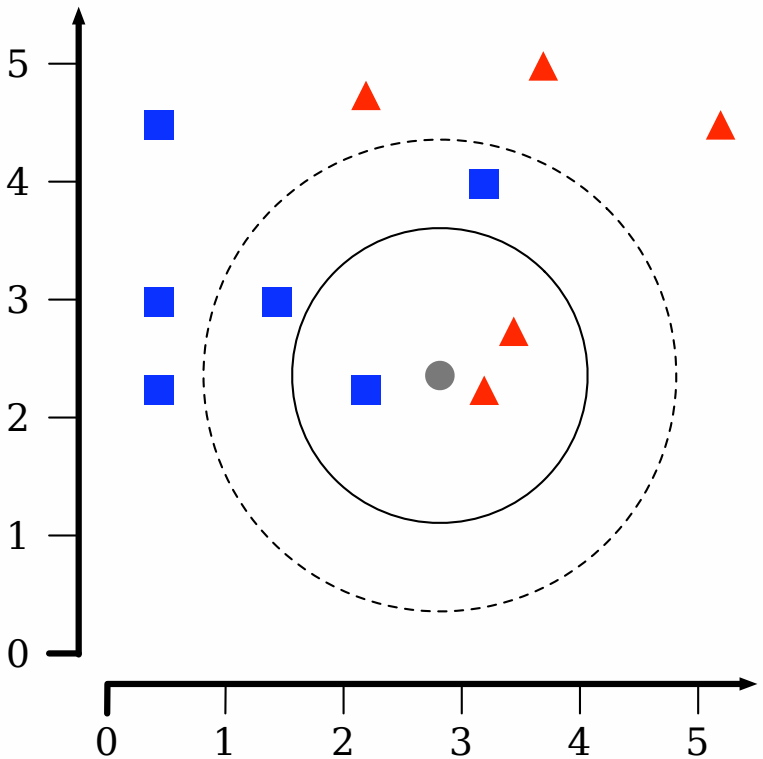
- Semantics construction
- Scope underspecification
- Dominance graphs & Tree Automata



Clustering (Greenberg)

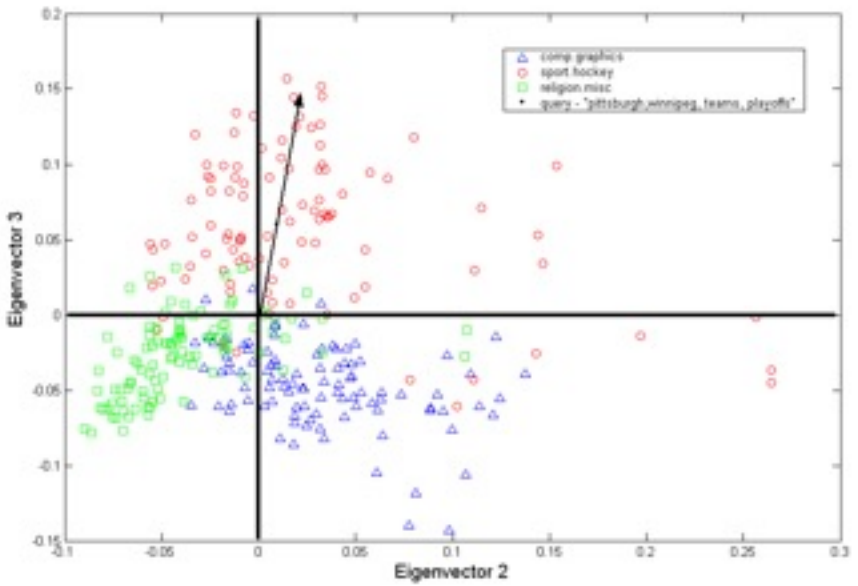
- Brown algorithm
- Exchange algorithm
- k-nearest neighbors algorithm
- k-means algorithm
- Expectation-maximization

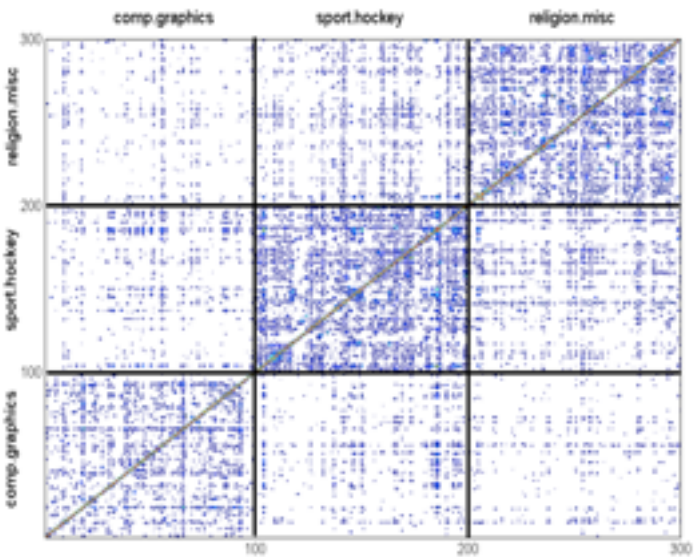
Cluster	Example members
1	Groß, Rau, Müller, Zimmermann, Frei, Becker, Möllemann, Schmidt
2	Düsseldorf, Berlin, München, Köln, Stuttgart, Hannover, Hamburg
3	nahmen, macht, zeigt, gleichen, bringt, biete, machte, sorgt, enthält



Topic spaces (Greenberg)

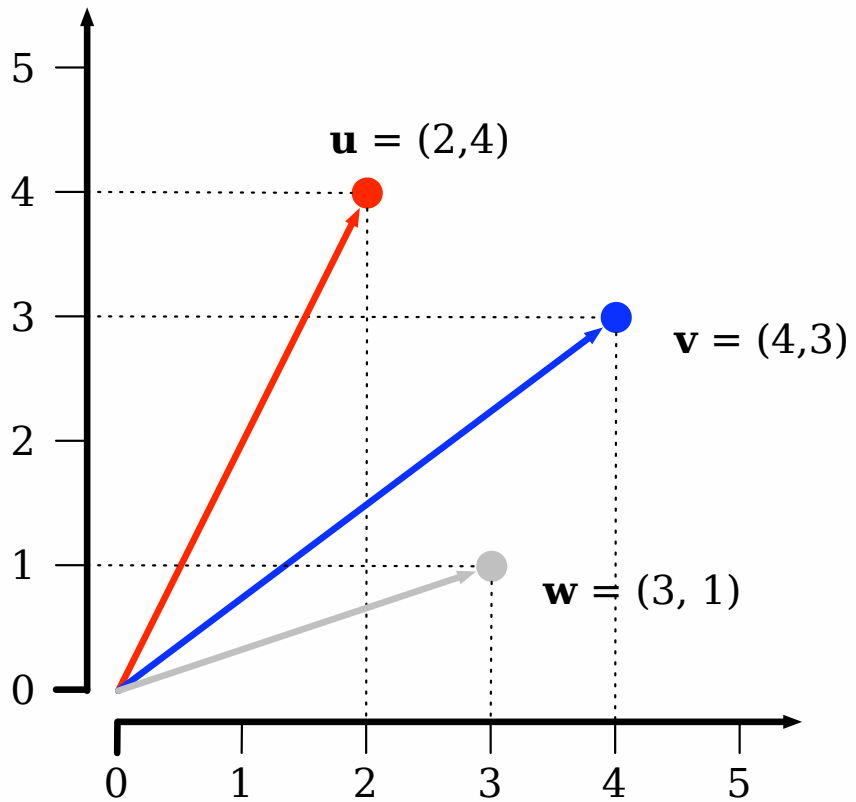
- Latent semantic analysis (LSA)
- Probabilistic latent semantic analysis (PLSA)
- Principal component analysis (PCA)
- Singular value decomposition (SVD)
- Non-negative matrix factorization (NMF)
- Latent Dirichlet allocation (LDA)





Lexical spaces (Greenberg)

- Distributional Semantics
- Vector space models
- Word embeddings
- Similarity measures
- Evaluation tasks
- Compositionality?



woman - man + king = queen

Students' choice lecture (Greenberg)

- Image retrieval
- Caption generation/image tagging
- Applications of semantic spaces
- Applications of information theory
- Other?