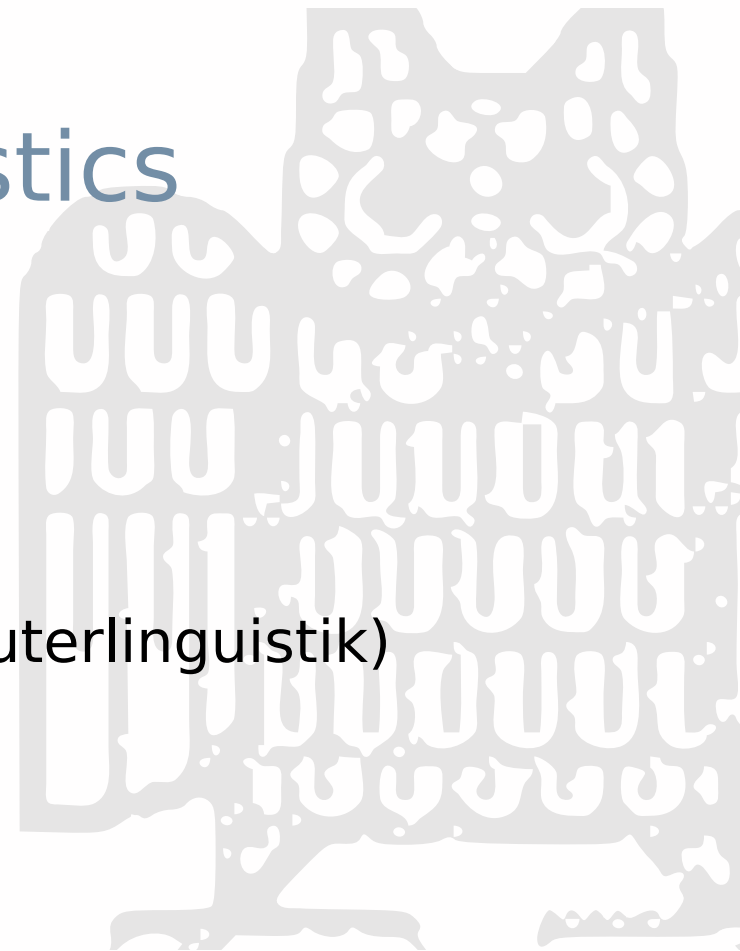


# Computational Linguistics

Clayton Greenberg  
Stefan Thater

FR 4.7 Allgemeine Linguistik (Computerlinguistik)  
Universität des Saarlandes  
Summer 2016



# This Lecture

- **Lectures:**

- Tuesday, 8:30 – 10:00

- **Exercise sessions:**

- Thursday, 8:30 – 10:00

- **Lecturers:**

- Stefan Thater (1<sup>st</sup> half)
- Clayton Greenberg (2<sup>nd</sup> half)

# Course website

- Lecture slides and exercise sheets can be obtained from the course website:
  - <http://www.coli.uni-saarland/courses/CL/2016>
- 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

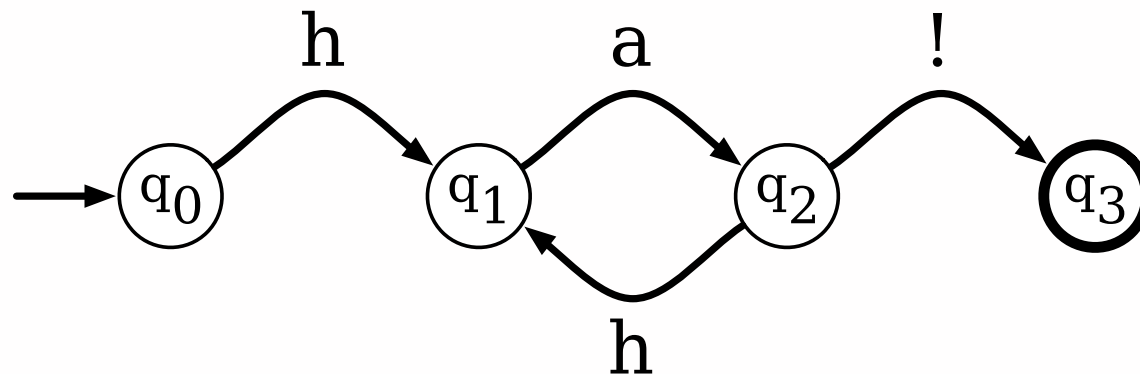
- The final exam will take place
  - Tuesday, 2016-08-02 (to be confirmed)
- Registration deadline: **July 11th**

# 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)

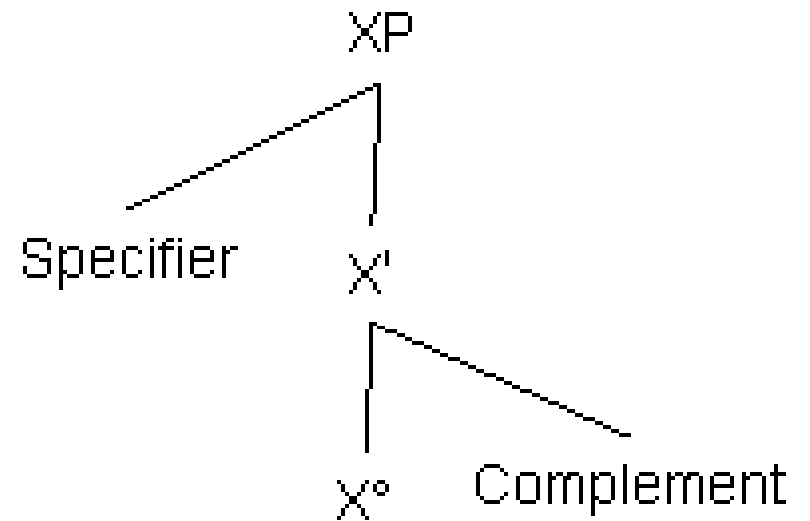
- Deterministic & non-deterministic finite state automata
- Algorithms for recognizing strings
- Algorithm for determinizing non-deterministic automata



*ha!*  
*haha!*  
*hahaha!*  
*hahahaha!*  
*...*

# Linguistic Essentials (Greenberg)

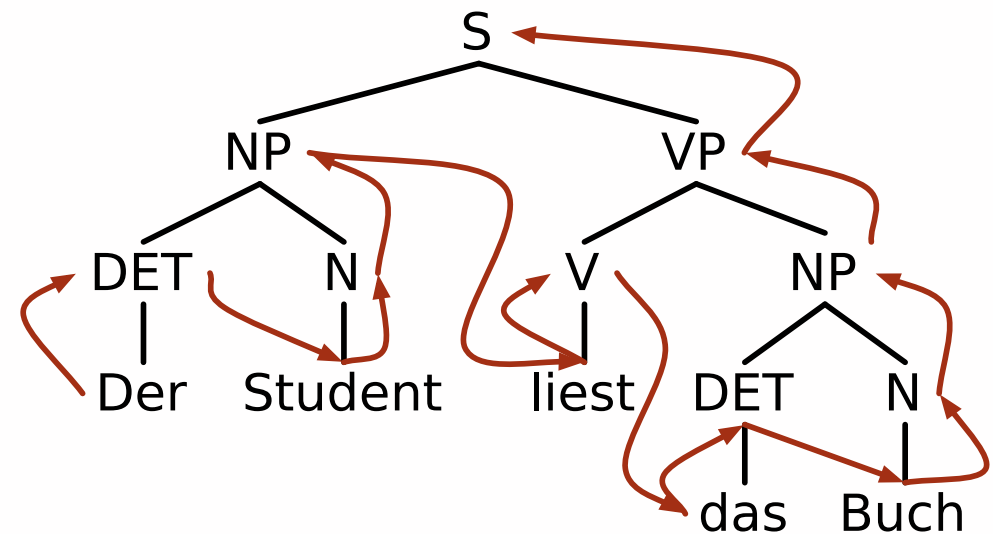
- A few ways to cluster words
- Penn Treebank tagset
- Brown Tagset
- Features of nominals
- Features of verbals
- Survival syntax
- Bridge to formal grammars





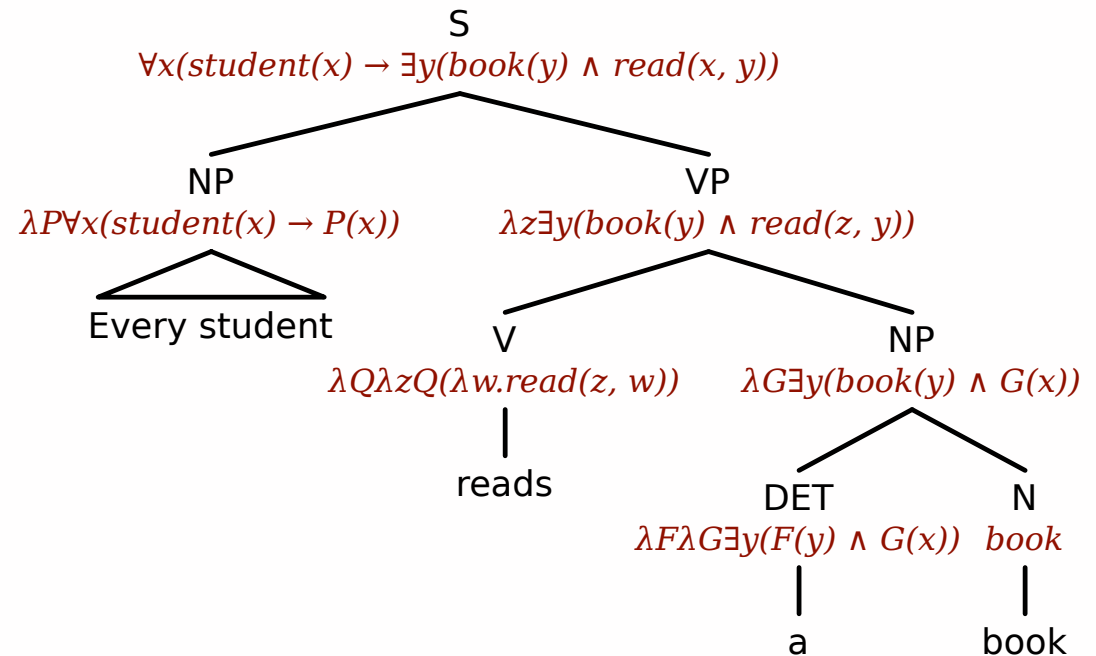
# Grammars & Parsing (Thater)

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



# Computational Semantics (Thater)

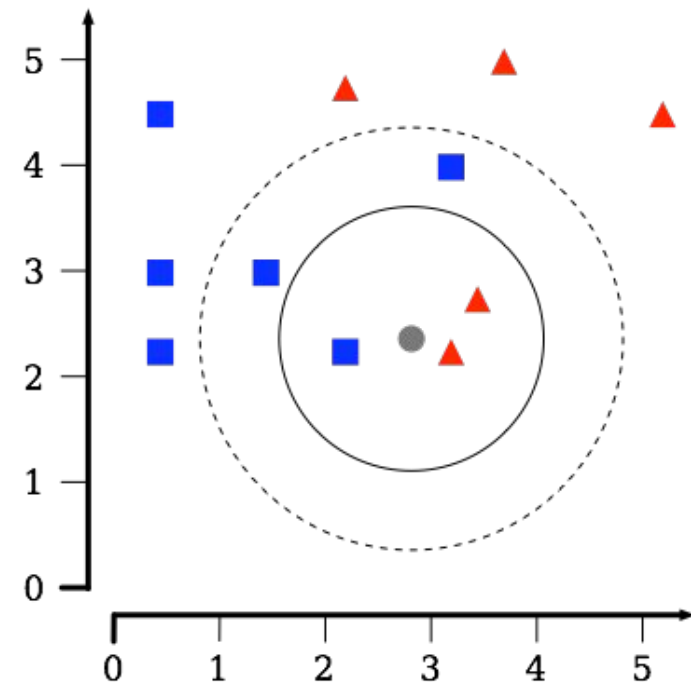
- Semantics construction
- Scope underspecification
- Dominance graphs & Tree Automata



# Clustering (Greenberg)

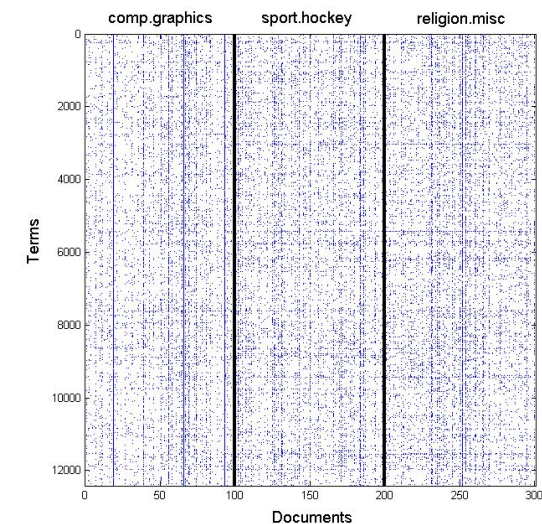
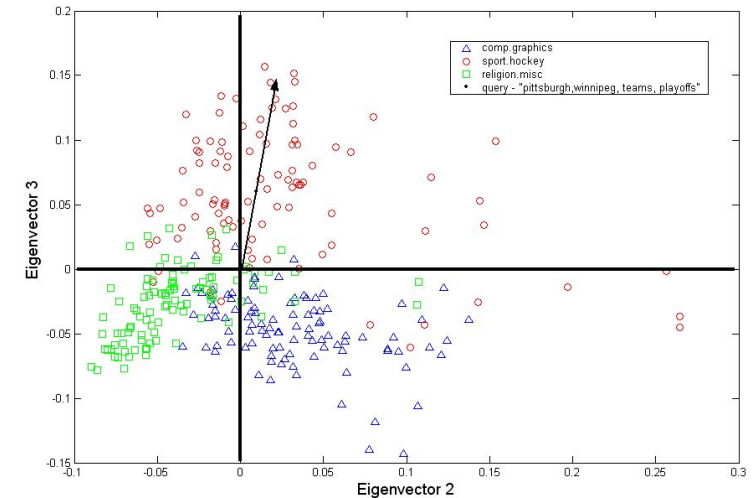
- Brown Algorithm
- Exchange Algorithm
- K-nearest neighbors Algorithm
- K-means Algorithm
- Variance Ratio Criterion

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



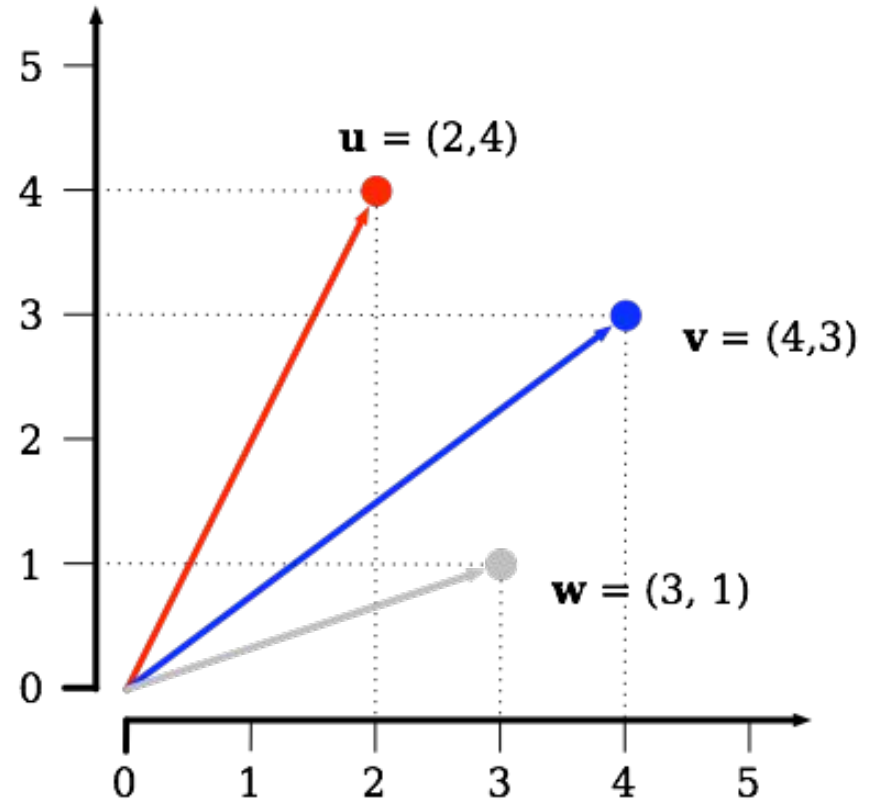
# Topic Spaces (Greenberg)

- 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)



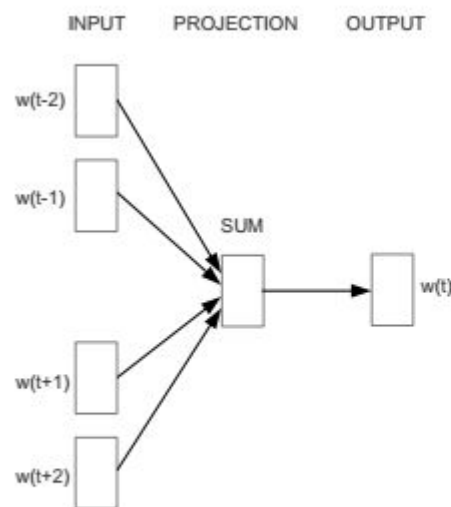
# Lexical Spaces (Greenberg)

- Distributional Semantics
- Vector space models
- Similarity measures
- Evaluation tasks
- Compositionality

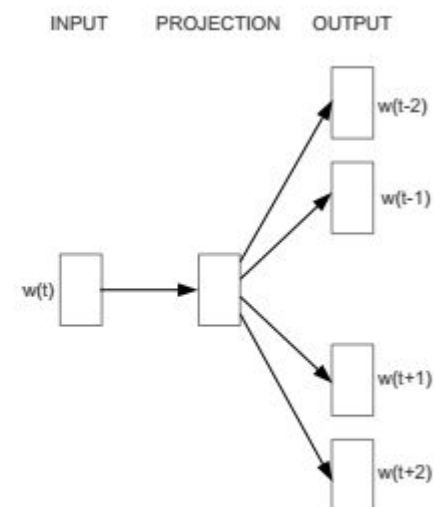


# Continuous Representations (Greenberg)

- Computational advantages?
- Why does Word2Vec work?
- Metaparameter tuning
- Comparison of neural and count based methods
- Applications



CBOW



Skip-gram