

Computational Linguistics

Dietrich Klakow & Stefan Thater

FR 4.7 Allgemeine Linguistik (Computerlinguistik)
Universität des Saarlandes
Summer 2012

This Lecture

- **Lectures:**
 - Thursday, 14:15 - 15:45
- **Exercise sessions:**
 - Monday, 10:15 - 11:45 (ST)
 - Monday, 16:15 - 17:45 (DK)
- **Lecturers:**
 - Stefan Thater (Week 1-8)
 - Dietrich Klakow (Week 9-12)

2

Course website

- Lecture slides and exercise sheets can be obtained from the course website:
 - <http://www.coli.uni-saarland/courses/CL/2012>
- Please subscribe to the follow mailing list:
 - <http://ml.coli.uni-saarland.de/cgi-bin/mailman/listinfo/compling>

3

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

4

Final Exam

- The final exam will take place Monday, July 24
 - Registration deadline: July 10
 - Date to be confirmed!

5

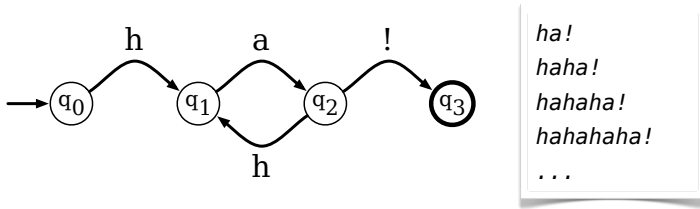
Textbooks

- 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.

6

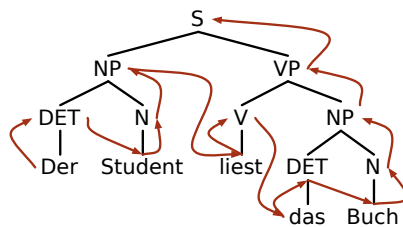
Lecture 1 (ST) Finite State Automata

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



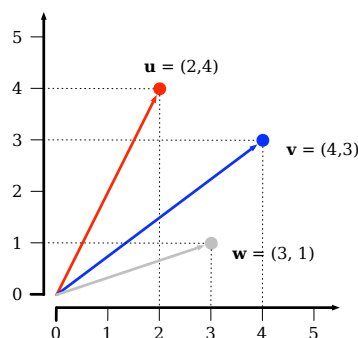
Lectures 2, 3, 4 (ST) Grammars & Parsing

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



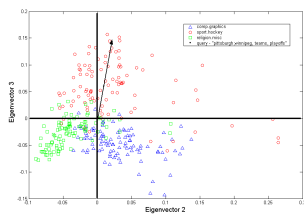
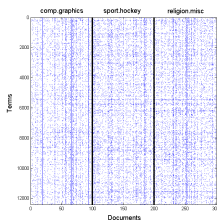
Lecture 5 (DK) Distributional Semantics

- Distributional Semantics
- Vector space models
- Similarity measures



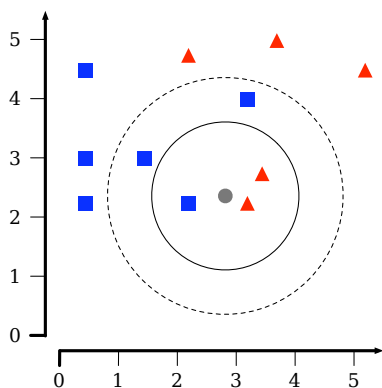
Lecture 6 (DK) Matrix Factorization

- Latent semantic analysis (LSA)

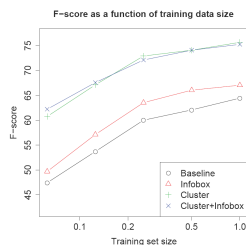


- Non-negative matrix factorization (NMF)

Lecture 7 (DK) Clustering



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



Lecture 8 (ST) Word Sense Disambiguation

- 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:

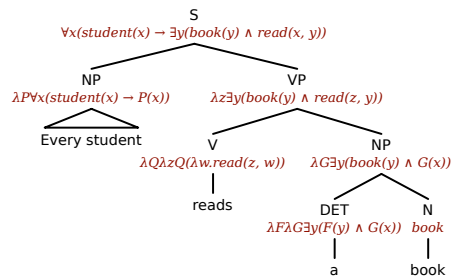
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

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**, **zharabang**, **double-decker**, **jitney**, **motorbus**, **motorcoach**, **omnibus**, **passenger vehicle** (a vehicle carrying many passengers; used for public transport) "he always rode the bus to work"

Lecture 9 (ST) Computational Semantics

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



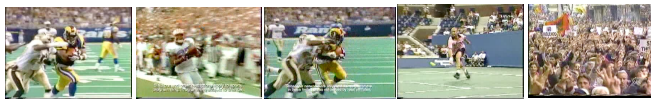
13

Lecture 10 (DK) Combining Visual & Textual Data

Weather News



Football



Horse



Alternative:
caption generation

14