

Introduction to Computational Linguistics (WS 07/08)

Jan-Philipp Söhn

Seminar: Wednesday 16ct–17:45 at the SfS, Hörsaal 0.02

Regular seminar starts: Wednesday, 24th October 2006

Credits: 3 CP

Office Hours: on appointment

Office: Room 2.04 at Nauklerstr. 35, 2nd floor

Phone: 29-77153

Email: jp.soehn@uni-tuebingen.de

- Webpage: http://www.soehn.net/work/intro_cl.htm
- Midterm exam: December 5th
- Final exam: February 6th

Class Participation

Examination regulations of the Neuphilologische Fakultät require that students attend courses regularly. If students do not attend a course meeting on more than two occasions in one semester without proper excuse (e.g. doctor's note), the course instructor has to give them a failing grade. Please do not put me in a position to have to fail you for this reason. If you cannot come to class, please email me ahead of time, if at all possible.

You are expected to come on time. Being late without good reasons will count as not having attended a course meeting.

If you own a mobile phone and carry it with you, please turn it off before class.

Reading Assignments

Please read the assigned reading in advance of the class meeting for which it was assigned. I will presuppose that you have read the material when we discuss it in class.

Grading Policy

Your grade will be based on two components: midterm exam (50 %) and final exam (50 %).

Course Objective

This introductory course has five major goals:

- (Largely non-technical) introduction to the field of computational linguistics and its history.
- Survey of natural language processing applications.
- In-depth look at machine translation as a means to illustrate the major tasks for natural language processing
- Presentation of tools and resources needed for natural language processing applications.
- To give you credit for your work and to get you one step closer to your degree.

Time Table

Oct. 17	<i>Block Session (Saile)</i>	
Oct. 24	Start: General Introduction, History of CL	[Searle 1980], [Turing 1950]
Part I: Applications		
Oct. 31	Lexical Resources, Word Nets	[Miller et al. 1993], [Kunze and Lemnitzer 2002]
Nov. 7	Dialectometry	[Chambers and Trudgill, 1980]
Nov. 14	Foreign Language Learning	[Nerbonne 2003]
Nov. 21	Machine Translation I	[Hutchins and Somers 1992, Ch. 6], [Somers 2000]
Nov. 28	Machine Translation II	[Somers 1999]
Dec. 5	Midterm exam	
Part II: Methods & Techniques		
Dec. 12	Tokenization and Sentence Segmentation I	[Palmer 2000]
Dec. 19	Tokenization and Sentence Segmentation	[Palmer 2000]
Jan. 9	Regular Expressions, Finite State	[Karttunen 2003]
Jan. 16	Finite State Transducers	[Karttunen 2003]
Jan. 23	Morphological Analysis	[Trost 2003]
Jan. 30	Part of Speech Tagging	[Leech 1997]
Feb. 6	Final exam	
Feb. 13	Part of Speech Tagging contd., Resources in Computational Linguistics	

References

[Arnold et al. 1994]

Arnold, D., Balkan, R., Humphreys, R. L., Meijer, S., and Sadler, L. (eds) 1994. Machine Translation. An Introductory Guide. Manchester/Oxford: Blackwells-NCC
<http://www.essex.ac.uk/linguistics/clmt/MTbook/>

[Chambers and Trudgill 1980]

Chambers, J.K. and Trudgill, Peter. Dialectology, University Press Cambridge, 1980.
 (Chapter 1 + 2)

[Dale et al. 2000]

Dale, R., Moisl, H., and Somers, H. (eds) 2000. Handbook of Natural Language Processing. New York, Basel: Marcel Dekker, Inc.

[Garside et al. 1997]

Garside, R., Leech, G., and A., McEnery (eds) 1997. Corpus Annotation. Linguistic Information from Computer Text Corpora. Addison Wesley Longman Ltd.

[Hutchins and Somers 1992]

Hutchins, W. J. and Somers, H. L. (eds) 1992. An Introduction to Machine Translation. San Diego: Academic Press.

[Karttunen 2003]

Karttunen, L. 2003. Finite state technology. In [Mitkov, 2003], 339–358.

[Kunze and Lemnitzer 2002]

Lemnitzer, L. & C. Kunze (2002): "Adapting GermaNet for the Web". In: Proceedings of the First Global Wordnet Conference, Central Institute of Indian Languages, Mysore, India, 21.-25.01.2002, pp. 174-181.

[Leech 1997]

Leech, G. 1997. Grammatical tagging. In [Garside et al., 1997], 19–33.

[Locke and Booth 1955]

Locke, W. N. and Booth, A. D. (eds) 1955. Machine Translation of Languages. Cambridge, Mass.: MIT Press.

[Miller et al. 1993]

5 papers on wordnet, Ch. 1 + 2: Introduction to WordNet: An On-line Lexical Database / Nouns in WordNet: A Lexical Inheritance System
<http://wordnet.princeton.edu/5papers.pdf>

[Mitkov 2003]

Mitkov, R. (ed) 2003. The Oxford Handbook of Computational Linguistics. Oxford University Press.

[Nerbonne 2003]

Nerbonne, John. 2003. Computer-Assisted Language Learning and Natural Language Processing. In [Mitkov, 2003], 670-698.

[Palmer 2000]

Palmer, D. D. 2000. Tokenization and sentence segmentation. In [Dale et al., 2000], 11–36.

[Searle, 1980]

Searle, J. 1980. Minds, brains, and programs. Behavioral and Brain Sciences, 3:417–458.

[Slocum 1988]

Slocum, J. (ed) 1988. Machine Translation Systems. Cambridge University Press.

[Somers 1999]

Somers, H. 1999. Example-based Machine Translation. Machine Translation 14: 113–157. Kluwer Academic Publishers.

[Somers 2000]

Somers, H. 2000. Machine translation. In [Dale et al., 2000], 329–346.

[Trost 2003]

Trost, H. 2003. Morphology. In [Mitkov, 2003], 25–47.

[Turing 1950]

Turing, A. M. 1950. Computing machinery and intelligence. Mind, 59.