

## **Statistical Language Processing (CL III): 6 ECTS**

**Instructors: Prof. Dr. Erhard Hinrichs**

**Sommersemester 2024;** meeting times: Tuesday, 16:15 – 17:45 and Thursday, 14:15 – 15:45

**Room 0.02, Wilhelmstr. 19**

**First class meeting: Tue, April 16, 16:15 – 17:45**

Successful completion of this course satisfies one of two course requirements for the ISCL-BA-08 module of the Bachelor of Arts Hauptfach curriculum in Computational Linguistics.

- a) Statistical Language Processing (CLIII) (4 semester credit hours, 6 ECTS)
- b) Tutorial Statistical Language Processing (CL III) (2 semester credit hours, 3 ECTS)

The **tutorial Statistical Language Processing (CL III)** will be taught this semester as well. There will be a separate Moodle page for that course.

### **Important Note:**

For B.A. students with Computational Linguistics as their Hauptfach, enrollment in this course is only possible for those who fulfill the following pre-requisites: successful completion of modules ISCL-BA-06: Symbolic Computational Linguistics: Parsing, and ISCL-BA-07: Advanced Programming

For B.A./B.Sc. students enrolled in programs other than Computational Linguistics and for M.A. students in Computational Linguistics: please consult with me as the course instructor (email: [erhard.hinrichs@uni-tuebingen.de](mailto:erhard.hinrichs@uni-tuebingen.de)) concerning pre-requisites for this course.

For M.A. students in Computational Linguistics, please consult with the M.A. Student Advisor Dr. Çağrı Çöltekin as to whether you can receive credit for this course.

### **Credit Points and Grading Policy**

The following requirements will be mandatory for all participants who want to receive course credit for this class:

**1. Active participation in class:**  $4\text{SWS} * 15 = 60$  hours (2 CP)

**Please note:** Missing more than two classes without valid excuses automatically results in failing the class. Attendance and participation will count for **20% of your final grade**.

**Reading Assignments :**  $4\text{h} * 15 = 60$  hours (2 CP)

Reading assignments need to be completed in advance of each lecture. Write down any open questions and raise these questions during class.

**2 . Final Exam:** 4h \* 15 = 60h (2 CP)

This class concludes with a written examination (Klausur). The final exam will count for **80% of your final grade**.

## **Time Schedule**

### **Week 1**

**Tue, April 16** Course Overview (30 min); Vector Semantics and Embeddings: Word Meaning

Reading: J&M, 6.1

**Thur, Apr 18** Vector Semantics and Embeddings: Vector Semantics; Words and Vectors, Cosine similarity

Reading: J&M, 6.2 – 6.4

### **Week 2**

**Tue, April 23** Vector Semantics and Embeddings: TF-IDF, PMI, Word2Vec

Reading: J&M, 6.5 – 6.8.1; Further Reading: Alamar, Jay. The Illustrated Word2Vec.

**Thur, April 25** Vector Semantics and Embeddings: Learning the Embeddings; Properties of Embeddings

Reading: J&M, 6.8.2 - 6.12

### **Week 3**

**Tue, April 31** Simple Neural Networks and Neural Language Models: Units in Neural Networks, The XOR problem

Reading: J&M, Chapter 7 – 7.2

**Thur, May 02** Feedforward Neural Networks

Reading: J&M, Chapter 7.3

### **Week 4**

**Tue, May 07** Applying feedforward networks to NLP tasks: Classification and Language Modelling

Reading: J&M, Chapter 7.4 - 7.5

**Thur, May 09** Holiday

## **Week 5**

**Tue, May 14** Training

Reading: J&M, Chapter 7.6 – 7.7

Videos:      What is Backpropagation really doing?:  
<https://www.youtube.com/watch?v=Ilg3gGewQ5U>

Backpropagation Calculus:  
<https://www.youtube.com/watch?v=tIeHLnjs5U8>

**Thur, May 16** Training

Reading: J&M, Chapter 7.6 – 7.7

## **Week 6**

**No classes (Pfingstferien)**

## **Week 7**

**Tue, May 28** English Word Classes, Part-of-Speech Tagging; Named Entity and Named Entity Tagging

Reading: J&M, Chapter 8.1 – 8.3

**Thur, May 30** Holiday

## **Week 8**

**Tue, June 04** HMM Tagging

Reading: J&M, Chapter 8.4

**Thur, June 06** Conditional Random Fields (CRF)

Reading: J&M, Chapter 8.5

## **Week 9**

**Tue, June 11** Recurrent Neural Networks (RNNs); RNNs as LMs; RNNs for other NLP tasks

Reading: J&M, Chapter 9.1; J&M, Chapter 9.2; J&M, Chapter 9.3

**Thur, June 13** Stacked and bi-directional RNN architecture;

Reading: J&M, Chapter 9.4

## **Week 10**

**Tue, June 18** Common RNN NLP Architectures, The Encoder-Decoder Model with RNNs

Reading: J&M, Chapter 9.6 – 9.7

**Thur, June 20** Attention

Reading: J&M, Chapter 9.8

## **Week 11**

**No classes**

## **Week 12**

**Tue, July 02** The Transformer: A Self-Attention Network

Reading: J&M, Chapter 10.1

**Thur, July 04** Multihead Attention; Transformer Blocks

Reading: J&M, Chapter 10.2.; Chapter 10.3.

## **Week 13**

**Tue, July 09** Chapter 10

**Thur, July 11** Chapter 11: BERT

## **Week 14**

**Tue, July 16** Chapter 11: BERT

**Thur, July 18** Prompting, In-Context Learning, and Instruct Tuning

## **Week 15**

**Tue, July 23** Prompting, In-Context Learning, and Instruct Tuning

**Thur, July 25** Review

**Final Exam: Tue, July 30 (mandatory date for all participants)**

## Textbook

Daniel Jurafsky and James H. Martin (2024). *Speech and Language Processing - An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Third Edition draft. February 03, 2024.

URL: <https://web.stanford.edu/~jurafsky/slp3/>

## Further Readings

Alammar, Jay. The Illustrated Word2Vec. <https://jalammar.github.io/illustrated-word2vec/>

Alammar, Jay. The Illustrated Transformer. <https://jalammar.github.io/illustrated-transformer/>

Alammar, Jay. Visualizing A Neural Machine Translation Model (Mechanics of Seq2seq Models With Attention). <https://jalammar.github.io/visualizing-neural-machine-translation-mechanics-of-seq2seq-models-with-attention/>

Karpathy, Andrej. The Unreasonable Effectiveness of Recurrent Neural Networks. <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>