

# Natural Language Processing

Magdalena Wolska   Benno Stein   Matthias Hagen   Henning Wachsmuth   Martin Potthast

# Contents

I. Introduction

II. Corpus Linguistics

III. Words

IV. Syntax

V. Semantics

VI. Pragmatics

VII. Speech

VIII. NLP Architectures

IX. NLP Applications

# Objectives

- ❑ understand the basics of language processing
- ❑ acquire background on selected elements of language processing theory
- ❑ implement and apply NLP algorithms in practice
- ❑ comprehend practical considerations, enabling comparative evaluation
- ❑ being able to educate oneself

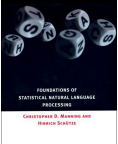
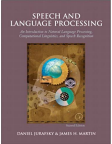
# Related Fields

1. Linguistics [paradigms, models]
2. Statistics
3. Psychology
4. Machine Learning [methods, algorithms]
5. Data Mining
6. Information Retrieval
7. Knowledge Processing
8. Text-to-Speech Systems [applications]
9. Dialog Systems
10. Conversational Systems

# Literature

## Natural Language Processing:

- ❑ D. Jurafsky, J. H. Martin.  
*Speech and Language Processing*  
Prentice Hall 2008.
- ❑ C. D. Manning, H. Schütze.  
*Foundations of Statistical Natural Language Processing*  
MIT Press 1999.



# Literature

## Top-tier natural language processing conferences:

- ❑ **AACL.**  
Conference of the Asian Chapter of the ACL.
- ❑ **ACL.** (the only A\* NLP conference)  
Annual Meeting of the Association for Computational Linguistics.
- ❑ **COLING.** (linguistic focus)  
International Conference on Computational Linguistics.
- ❑ **CoNLL.**  
International Conference on Natural Language Learning.
- ❑ **EACL.**  
Conference of the European Chapter of the Association for Computational Linguistics.
- ❑ **EMNLP.** (empirical focus, probably the second best NLP conference)  
Conference on Empirical Methods in Natural Language Processing.
- ❑ **NAACL.** Conference of the North-American Chapter of the ACL.

# Literature

## Other relevant natural language processing conferences:

- ❑ CICLing.  
International Conference on Computational Linguistics and Intelligent Text Processing.
- ❑ IJCNLP.  
International Joint Conference on Natural Language Processing.
- ❑ INLG. (generation focus)  
International Conference on Natural Language Generation.

## Conferences from related fields:

- ❑ SIGIR, ECIR, ICTIR, AIRS, CHIIR, TREC, CLEF
- ❑ CIKM, WSDM, WWW, SPIRE

## Top-tier Journals:

- ❑ ACM TOIT. [toit.acm.org](http://toit.acm.org)
- ❑ Computational Linguistics. [www.mitpressjournals.org/loi/coli](http://www.mitpressjournals.org/loi/coli)
- ❑ TACL. Transactions of the ACL. [transacl.org](http://transacl.org)

# Software

## Industry (Closed Source):

- ❑ ExplosionAI  
*prodi.gy: An Annotation Tool Powered by Active Learning*  
[prodi.gy](https://prodi.gy)

## Industry (Open Source):

- ❑ ExplosionAI  
*spaCy: Industrial-Strength Natural Language Processing*  
[spacy.io](https://spacy.io)
- ❑ Hiroki Nakayama  
*Doccano: Text annotation for Human*  
[doccano.herokuapp.com](https://doccano.herokuapp.com)

## Research:

- ❑ NLTK Project  
*Natural Language Toolkit*  
[nltk.org](https://nltk.org)
- ❑ Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton and Christopher D. Manning  
*Stanza - A Python NLP Library for Many Human Languages*  
[stanfordnlp.github.io/stanza/](https://stanfordnlp.github.io/stanza/)