

# Syllabus for L645: Advanced Natural Language Processing

Sandra Kuebler

[skuebler@indiana.edu](mailto:skuebler@indiana.edu)

Memorial Hall 402

Office hours: Mon., Wed., 1:30—2:30 p.m.

## Course description

In recent years, statistical methods have become the standard in the field of Natural Language Processing (NLP). This course gives an introduction to statistical models and machine learning paradigms in NLP. Such methods are helpful for the following goals: reaching wide coverage, reducing ambiguity, automatic learning, increasing robustness, etc.

In this course, we will cover basic notions in statistics, focused on the concepts needed for NLP. Then we will discuss (Hidden) Markov Models, exemplified by an approach to POS tagging. The following sessions will be dedicated to probabilistic approaches to parsing. Here, we will start with an introduction to parsing in general, and then focus on probabilistic context-free grammars. In the last part of the course, we will cover statistical alignment methods and their use in statistical machine translation.

## Policies

**Attendance:** Your attendance is advised. I will not base any part of your grade directly on attendance, but I expect you to communicate directly with me if you are unable to attend class. You are responsible for the material covered in all classes.

**Grading:** Grading for the course is on a % scale, with the following ratio:

Assignments	40 %
Presentation:	20 %
Project Documentation:	40 %

**Reading:** I will expect you to do a significant amount of textbook reading in this course. The chapters are to be read **before** the class period in which they will be discussed. I do not expect you to understand everything (with the exception of the presenter) but I do expect you to be able to ask questions.

## Additional Reading (available on Oncourse):

JS: Jurafsky and Martin (2008) Speech and language Processing. 2<sup>nd</sup> ed.

KS: Krenn and Samuelsson (1997) The Linguist's Guide to Statistics – Don't Panic!

## Course Plan

	<b>Topic</b>	<b>Reading</b>
03 Sep	Organizational meeting	
08 Sep	Probability Theory	MS ch. 2.1, KS 1.1 – 1.4
10 Sep	Probability Theory	
15 Sep	Information Theory	MS ch. 2.2, KS 2.2
17 Sep	Finite-State Automata	
22 Sep	Markov Chains, Markov Models	MS 9.1, KS ch. 2.1.1 – 2.1.3
24 Sep	ngram POS Tagging	
29 Sep	Practical POS tagging	
01 Oct	Smoothing	JM ch. 4.5 (electronic)
06 Oct	Hidden Markov Models	MS ch. 9.2 – 9.3, KS 2.1.4
08 Oct	Calculating P(O)	MS ch. 9.3.1, KS 2.1.5
13 Oct	Finding the Optimal State Sequence	MS ch. 9.3.2, KS 2.1.6
15 Oct	Parameter Estimation	MS ch. 9.3.3, KS 2.1.7
20 Oct	Parameter Estimation	
22 Oct	CYK parsing	JM ch. 13 – 13.4.1
27 Oct	Treebanks	
29 Oct	Practical Parsing I	
03 Nov	Probabilistic PCFGs	MS ch. 11.1 – 11.3.3
05 Nov	Parsing with PCFGs	JM ch. 14.2, 14.4 – 14.5
10 Nov	Parsing with PCFGs	
12 Nov	Practical Parsing II	
17 Nov	Estimating PCFGs	MS ch. 11.3.4 – 11.4
19 Nov	Estimating PCFGs	
24 Nov	Project discussion	
<b>26 Nov</b>	<b>Thanksgiving</b>	
01 Dec	Beyond PCFGs	
03 Dec	Beyond PCFGs	
08 Dec		
10 Dec		
<b>15 Dec, 4pm</b>	<b>Term papers due</b>	