

Introduction

CS 685, Spring 2021

Advanced Topics in Natural Language Processing

<http://brenocon.com/cs685>

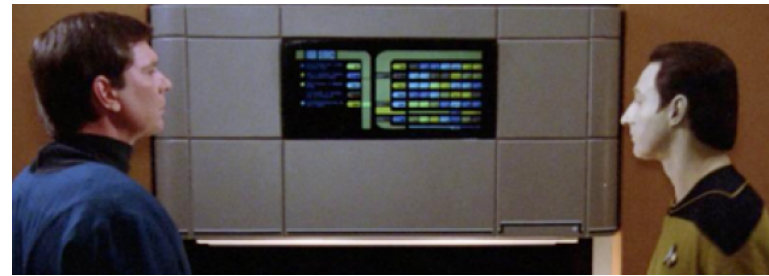
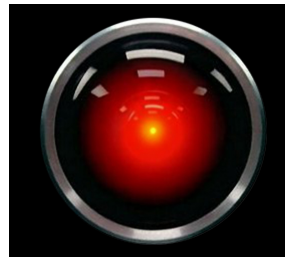
https://people.cs.umass.edu/~brenocon/cs685_s21/

Brendan O'Connor

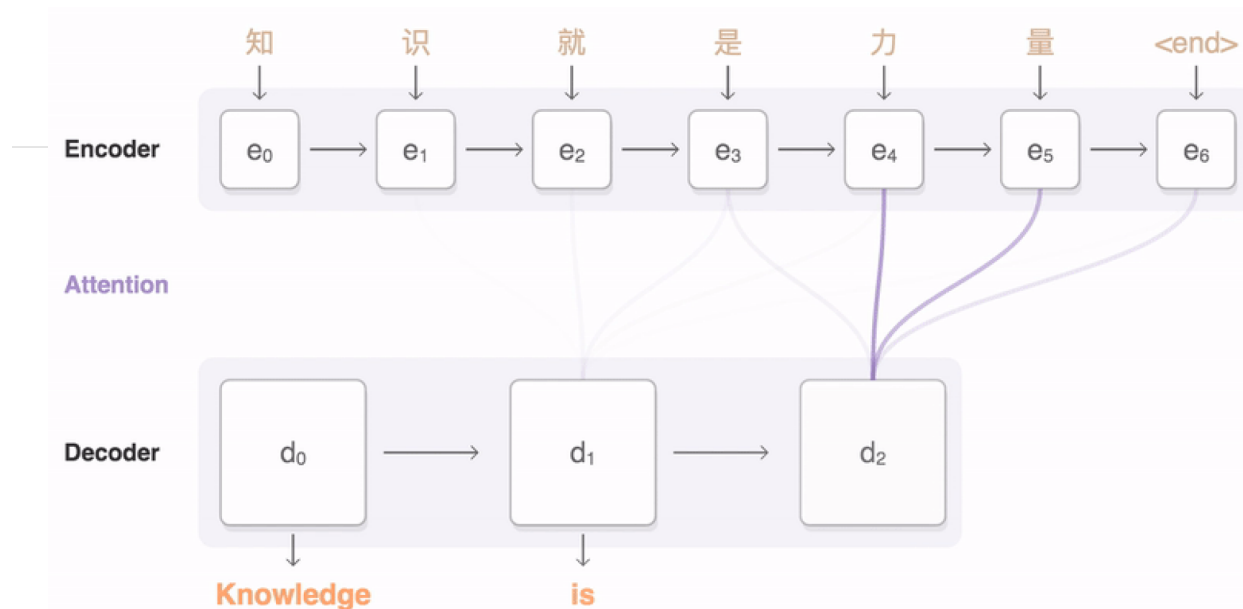
College of Information and Computer Sciences

University of Massachusetts Amherst

Computation + Language




Google Research Blog



TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

Entity Extraction



- Have technology (thanks to R6) – for English, Arabic and Chinese
- Allow queries like:
- Show me all the word documents with references to IAEO
- Show me all documents that reference Osama Bin Laden

- **Natural Language Processing:** computational {analysis, generation} of language
- **Computational Linguistics** - computational methods for the science of language
- Language is uniquely human
- Language is ambiguous, hard, and embedded in real-world context and knowledge
 - Is NLP is impossible?

This course

- “Advanced topics” in natural language processing
 - Assuming some machine learning background
 - Not assuming linguistics background or extensive NLP experience, but will be helpful
 - At graduate student level, designed to get you ready for NLP research
- General NLP topics
 - NLP-relevant ML essentials: linear, neural network, unsup. learning
- **"Syn-Sem-Soc"** version of the course, emphasizing:
 - Syntax (structure) in NLP
 - Semantics (meaning) in NLP
 - Social factors (context) in NLP
- Assignments emphasize research reading/writing

Other NLP at UMass

- NLP courses
 - Fall 2020 685: emphasis on deep learning and generation
 - This sem: examine DL's relationship to Sym-Sem-Soc
 - 490A/485/(old 585): substantial overlap with this course. This course could still be useful in addition, but not sure I'd recommend
 - Ling 509 + Ling 492B comp ling sequence (undergrad)
- Other related courses
 - CS 682 (Deep Learning), CS 688 (Graphical Models)
 - Ling 692C (Cognitive Modeling of Language)

People & logistics

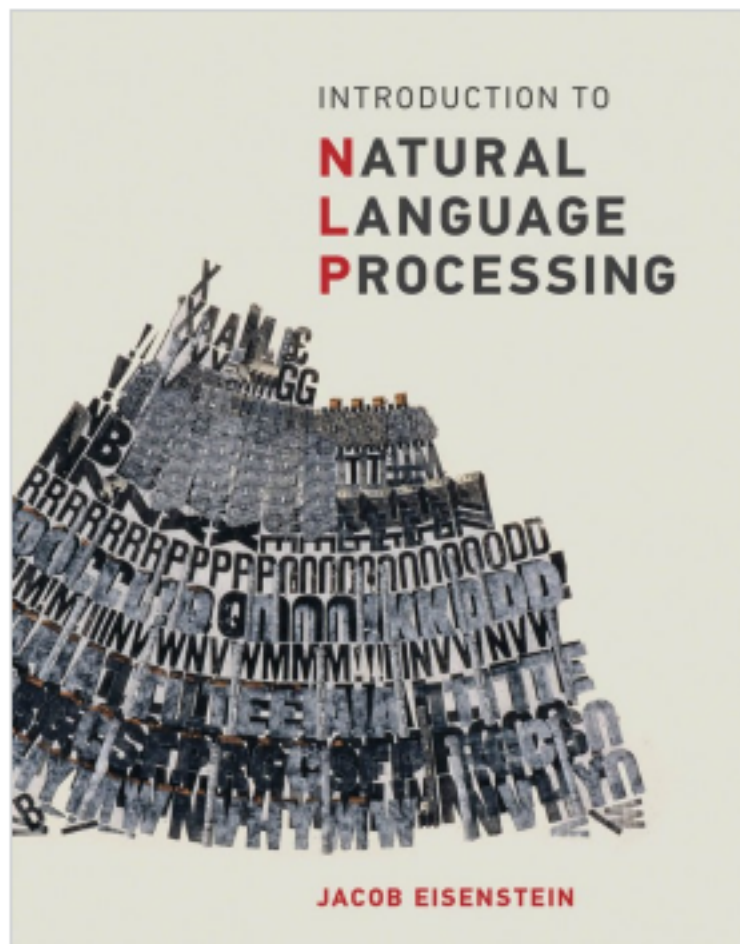
- TA:
 - Nader Akoury
- Public course webpage:
 - https://people.cs.umass.edu/~brenocon/cs685_s21/
- QUESTION: feelings about Discord?
 - or we can be boring and use piazza

Coursework / grading

- 40%: ~4 homework assignments (written, programming, analysis)
- 10%: Approx. weekly paper reviews or reading reactions (graded mostly for completion)
- 20%: Literature review (~midterm), done solo
- 30%: Final project, in small groups (2-3)
 - Project proposal
 - Progress report
 - Class presentations (late sem)
 - Final report (end of sem.)

Readings

- Papers provided as PDFs
- Also readings from the Eisenstein textbook.
[Online draft PDF will be fine]



From Adaptive Computation and Machine Learning series

Introduction to Natural Language Processing

By Jacob Eisenstein

A survey of computational methods for understanding, generating, and manipulating human language, which offers a synthesis of classical representations and algorithms with contemporary machine learning techniques.

HW0

- Not a “real” homework
- Sending out later today, due next Monday
 - A few math review questions
 - Writing: why are you interested in NLP?
- Submit PDF via Gradescope

Levels of linguistic structure

Communication, Beliefs, Actions...

Discourse

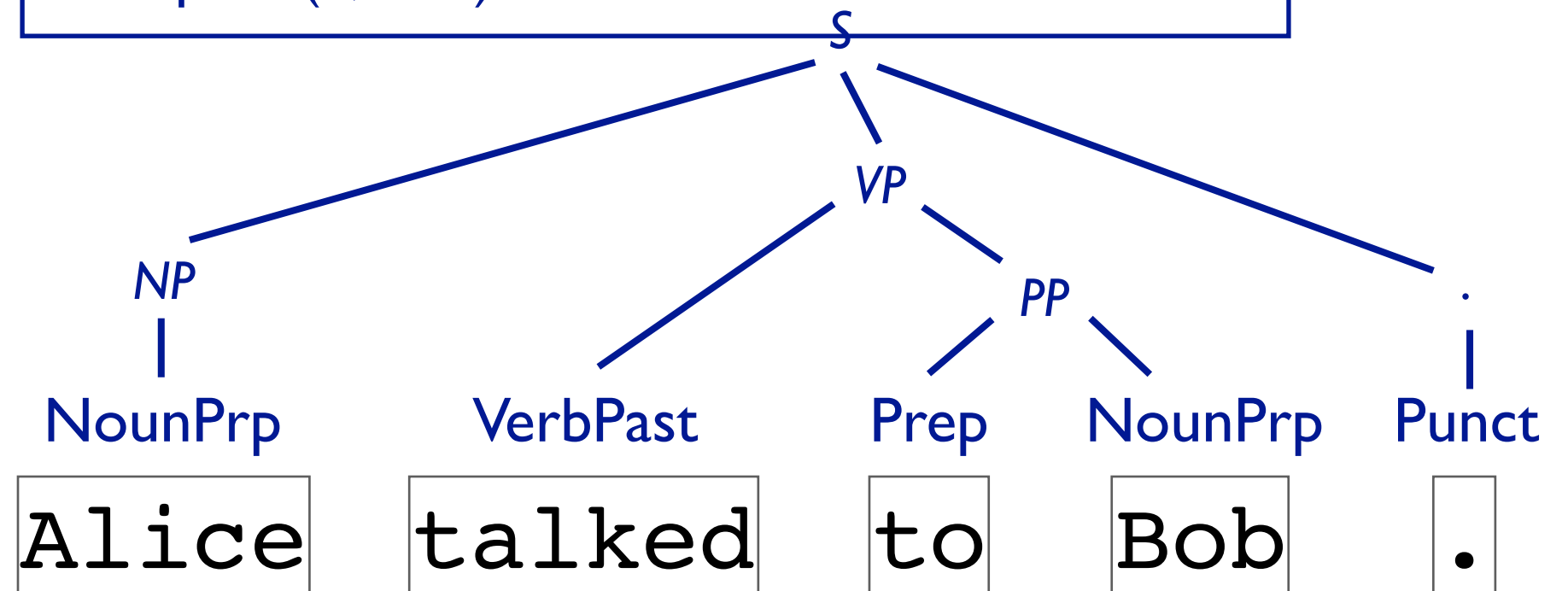
Alice saw Mary too. She talked to her.

Semantics

SpeakingEvent(e)
Agent(e, Alice)
Recipient(e, Bob)

SpeakerContext(s)
TemporalBefore(e, s)

Syntax



Words

Alice

talked

to

Bob

.

Morphology

talk -ed

Characters

Alice talked to Bob.

demos!
(allennlp.org)

python demo!
(colab.research.google.com)