CSE440: Natural Language Processing II

Dr. Farig Sadeque
Associate Professor
Department of Computer Science and Engineering
BRAC University

Lecture 1: Introduction

Class Expectation

What can we expect?

- Some linguistics knowledge
- A whole lot of algorithms. This is an **algorithms** course
- A bunch of programming

What this course is not:

- This is not a linguistics course. We will learn whatever linguistics we require during the course
- This is not a machine learning/neural network course. There will be a refresher within the first couple of weeks, but for deeper understanding, please take the respective courses.

Class Expectations

- Several semesters of programming
- Primary language: Python
- Linguistics experience helpful, not required
- Mathematical experience helpful, not required
    - As long as you can understand some maths notation, you will do well

Course Structure

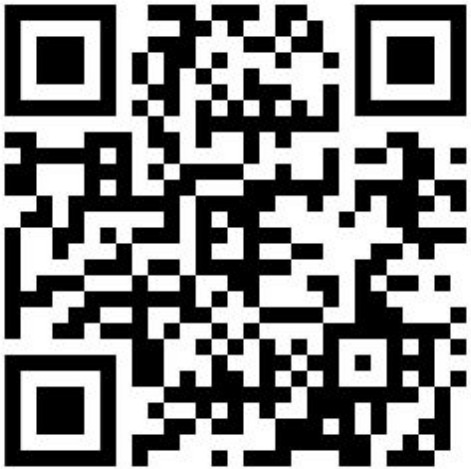
- Attendance: 0%
- Lab: 25%
- Quiz (best 3 out of 4): 15%
- Midterm: 30%
- Final 30%

Course Classroom

- Classroom will contain all notifications, contents (books, lectures, recordings) etc.
- https://classroom.google.com/c/Nzg2MzYyNzQyNDg0?cjc=ailtmmov
- Join the class using this QR code:

Consultation

- You can visit me during consultation hours: I prefer if you book a slot
- Appointment link: https://calendar.app.google/LXSrnyDF9a7B9sXq6



Course Plan

Linguistics essentials
- Sentence segmentation
- Tokenization
- Lemmatization/Stemming
- Parts-of-Speech tagging
- Named Entity Recognition
- Parsing
- Coreference Resolution

Machine Learning Essentials Review
- Probability Review
- Naive Bayes, Logistic regression
- Splits, metrics, statistical significance
- Essential ML maths refresher

Text Representation
- Representation basics
- Word embeddings
- Contextual embeddings

Sequence tagging
- Sequence tagging basics
- Markov Models
- Deep Learning Architectures: Recurrent Neural Network

Translation
- Probabilistic Translation
- Seq2seq model
- Attention mechanism
- Translation issues

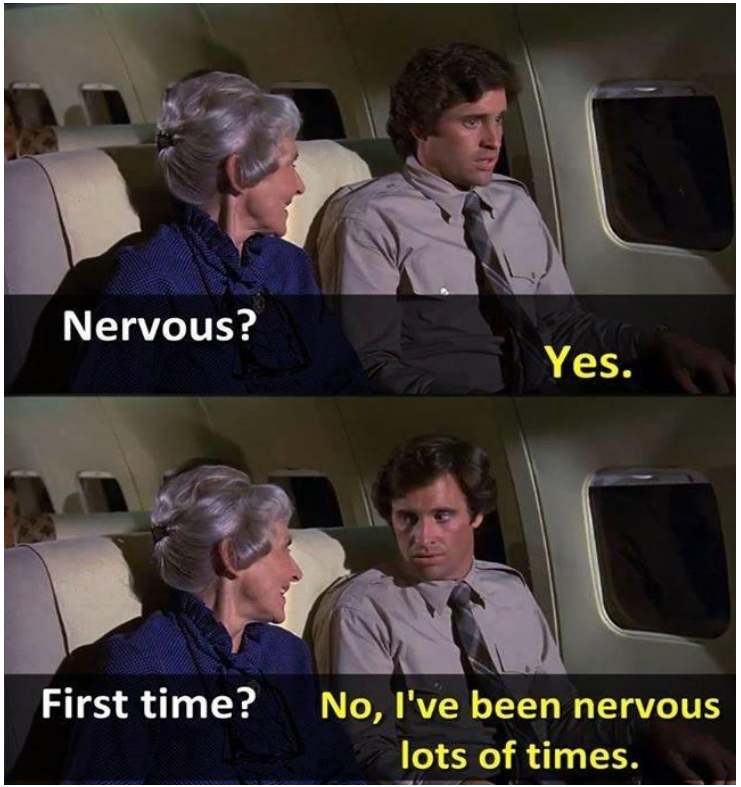
**Parsing
- Parsing Basics
- Constituency Grammar
- Constituency Parsing
- Dependency Parsing

**Coreference

**Text Generation: Encoder-Decoder Algorithm

**Question Answering

Why is NLP Hard?



Ambiguity

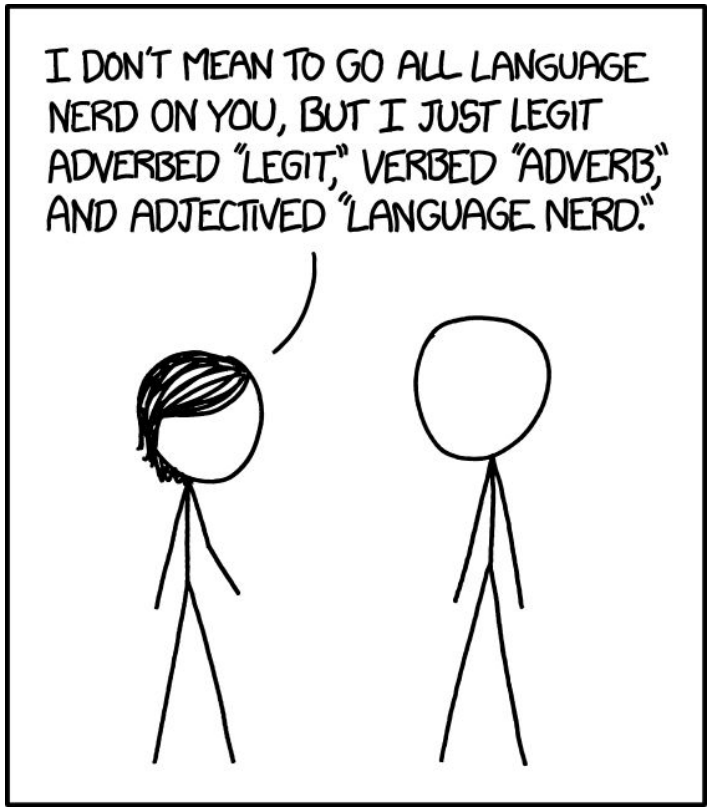
- Phonetics: I scream? Ice cream?
- Morphology: Union-ized? Un-ionized?
- Syntax: Squad helps dog bite victim.
    - Squad helps a dog to bite a victim?
    - Squad helps a dog-bite victim?
- Semantics: Ball: an orb, or a dance?
- “High-end” nonsense: Colorful green ideas sleep furiously.
- Discourse: see that photo again

Variability

He bought it

- He purchased it
- He acquired it
- It was bought by him (and all other synonyms with passive voice)
- It was sold to him
- …….

Language Change



Language Change

- English beats up other languages in dark alleys, then rifles through their pockets for loose grammar and spare vocabulary
- Example: We eat beef, but we raise cows.
- Fun video: https://www.youtube.com/watch?v=Jl3K63Rbygw