In a nutshell

• We have seen representations, datasets, models, and algorithms for computationally reasoning about textual language.

  ‣ Persistent challenges: Zipf’s Law, ambiguity & flexibility, variation, context

• Core NLP tasks (judgments about the language itself): tokenization, POS tagging, syntactic parsing (constituency, dependency), word sense disambiguation, word similarity, semantic role labeling, coreference resolution

• NLP applications (solve some practical problem involving/using language): spam classification, language/author identification, sentiment analysis, named entity recognition, question answering, machine translation

• Which of these are generally easy, and which are hard?
Language complexity and diversity

- **Ambiguity** and **flexibility** of expression often best addressed with corpora & statistics
  
  - Treebanks and statistical parsing

- Grammatical forms help convey meaning, but the relationship is complicated, motivating **semantic** representations
  
  - proposed by linguists, or
  
  - induced from data

- Typological variation: Languages vary extensively in **phonology**, **morphology**, and **syntax**
Methods useful for more than one task

- annotation, crowdsourcing
- rule-based/finite-state methods, e.g. regular expressions
- classification (naïve Bayes, perceptron)
- language modeling (n-gram or neural)
- grammars & parsing
- sequence modeling (HMMs, structured perceptron)
- structured prediction—dynamic programming (Viterbi, CKY)
Models & Learning

• Because language is so complex, most NLP tasks benefit from statistical learning.

• In this course, mostly **supervised learning** with *labeled* data. Exceptions:
  
  ‣ **unsupervised learning:** the EM algorithm (e.g. for word alignment, topic models)
  
  ‣ language models, distributional similarity/embeddings: supervised learning, but no extra labels necessary—the context is the supervision

• In NLP research, a tension between building a lot of linguistic insights into models vs. learning almost purely from the data.
  
  ‣ Current research on neural networks tries to bypass hand-designed features/intermediate representations as much as possible.
  
  ‣ We still don’t quite know how to capture “deep” understanding.
Generative and discriminative models

- Assign probability to language AND hidden variable? Or just score hidden variable GIVEN language?

- Independence assumptions: how useful/harmful are they?
  - “all models are wrong, but some are useful”
  - bag-of-words; Markov models
  - combining statistics from different sources, e.g. Noisy Channel Model

- Avoiding overfitting (smoothing, regularization)

- Evaluation: gold standard? sometimes difficult
Dynamic Programming Algorithms

• Allow us to search a combinatorial (exponential) space efficiently by reusing partial results.
Dynamic Programming Algorithms

• Allow us to search a combinatorial (exponential) space efficiently by reusing partial results.

• In a sentence of length $N$, what is the asymptotic runtime complexity of:

  ‣ Word edit distance, where the other sentence has length $M$? $O(M \cdot N)$

  ‣ Viterbi (in a first-order HMM), with $L$ possible labels?
Dynamic Programming Algorithms

• Allow us to search a combinatorial (exponential) space efficiently by reusing partial results.

• In a sentence of length $N$, what is the asymptotic runtime complexity of:

  ▶ **Viterbi** (in a first-order HMM), with $L$ possible labels? $O(NL^2)$

  ▶ **CKY**, with a grammar of size $G$?
Dynamic Programming Algorithms

• Allow us to search a combinatorial (exponential) space efficiently by reusing partial results.

• In a sentence of length $N$, what is the asymptotic runtime complexity of:

  ‣ **Viterbi** (in a first-order HMM), with $L$ possible labels? $O(NL^2)$

  ‣ **CKY**, with a grammar of size $G$? $O(N^3G)$
Applications

• Sentiment analysis, machine translation

• Your projects!

• Now that you know the tools in the toolbox, you can

BUILD ALL THE THINGS
The Final Exam

• Thursday 5/10, 4:00-6:00

• Largely similar in style to the midterm & quizzes, but with content covering the entire course.

• …and more short answer questions. For each major concept or technique, be prepared to define it, explain its relevance to NLP, discuss its strengths and weaknesses, and compare to alternatives.

  ‣ E.g.: “Why is smoothing used? For a model covered in class, describe two methods for smoothing and their pros/cons.”

• Study guide will be posted.

• Review session: Wednesday 1:00–2:00, ICC 462
Other Adminstrivia

- Projects due midnight tomorrow!

- Peer evaluations for the final project (watch for an announcement after tomorrow; **we need these to determine your grade**)

- No more office hours (unless you contact us)

- Related courses next semester include Advanced Semantic Representation (COSC/LING-672) and Dialogue Systems (COSC-483/LING-463)

- TA & course evaluations
  [https://eval.georgetown.edu/](https://eval.georgetown.edu/)