# Course Review

Kevin Duh
Intro to NLP, Fall 2019

# Congrats!

- We’re at the end of the course. You’ve learned a lot. Pat yourself on the back for your achievement!

# Final Exam Logistics

- Time: Thursday, December 19, 9:15AM - 10:30AM (1.5hr)
- Place: Shaffer 301
- Cheatsheet allowed: 1 page front/back A4/Letter-size
- Exam is comprehensive:
  - Covers all 11 modules of the course
  - Does not cover guest lectures

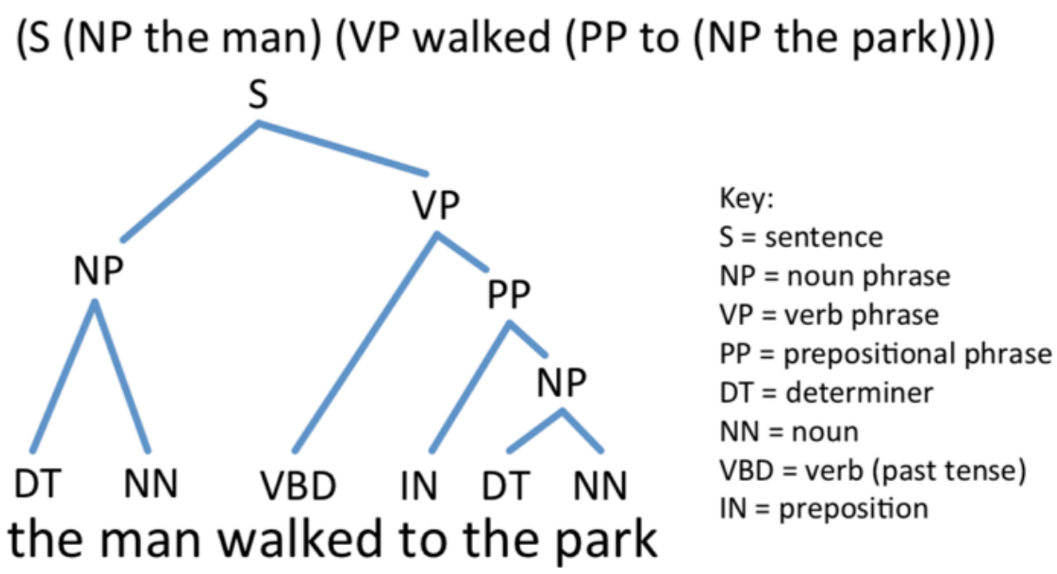
# Study Tips

- Review the slides and lecture notes
- Make sure you have mastered the assignments
  - e.g. can you now do the assignments all by yourself (without collaborators) if necessary?
- Make sure you have mastered the midterm exam
- Attend TA review session: **12:15 to 1:45 on Tuesday, December 17**

# Check your understanding…

# 1. Grammar Writing

- Context Free Grammar rules for English
    - What are the common rules? Can you derive some reasonable CFG rules based on sentence observations?
    - How do PCFG rules make things more flexible?
- Understand basic part-of-speech tags. Be able to tag yourself



# 2. Language Modeling

- Probability basics:
  - Explain joint probability, conditional probability, Bayes Rule, & entropy in mathematical terms
  - Perplexity — be able to implement LM evaluation
  - N-grams

$$\begin{aligned} p(\vec{w}) &= p(w_n \mid w_{n-1}, w_{n-2}, \ldots, w_1) \times p(w_{n-1} \mid w_{n-2}, \ldots, w_1) \\ &\times p(w_{n-2} \mid w_{n-3}, \ldots, w_1) \times p(w_{n-3} \mid w_{n-4}, \ldots, w_1) \\ &\times p(w_{n-4} \mid w_{n-5}, \ldots, w_1) \times \ldots \times p(w_2 \mid w_1) \times p(w_1) \end{aligned}$$

    - What assumptions are made?
    - How to estimate probabilities & do smoothing

# 3. Text Classification

- Log-linear model

$$p(y \mid x) = \frac{1}{Z(x)} \exp(score(x, y)) = \frac{1}{Z(x)} \exp(\sum_{k=1}^{K} \theta_k f(x, y))$$

  - Explain how it's defined, i.e. the math
  - Explain why it's a powerful model

- Machine learning concepts: generalization, overfitting, underfitting, model expressiveness, bias-variance

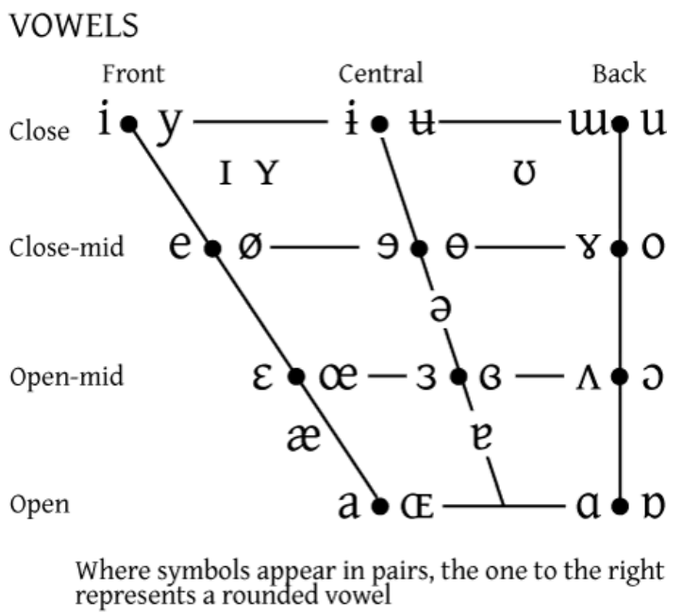
- How to setup experiments (train/dev/test split & procedures) when asked to deploy a classifier somewhere

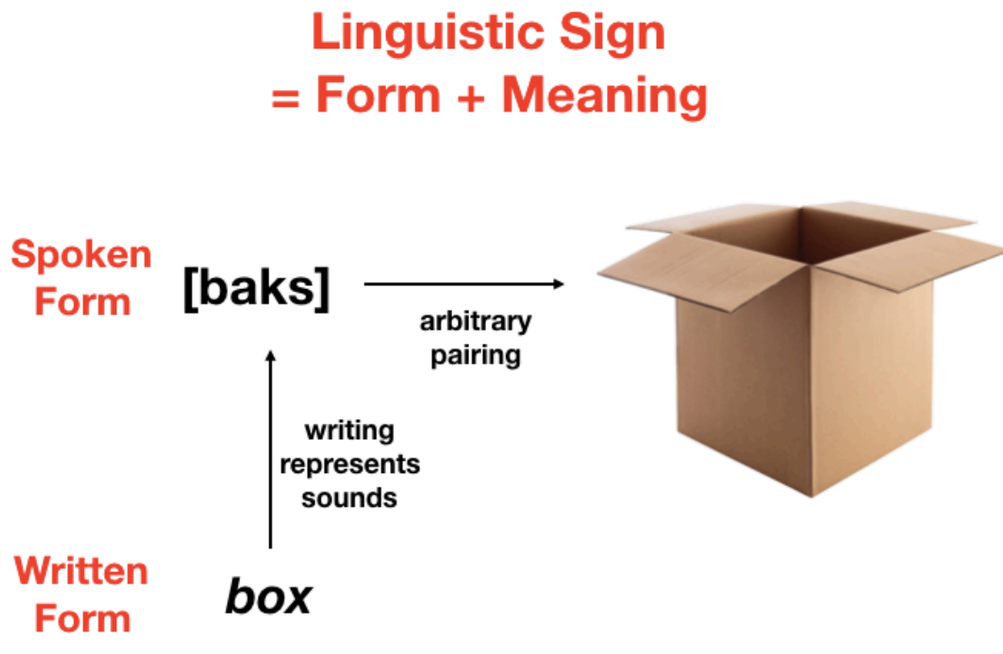
# 4. Linguistics 101

- Recognize that language is a spoken phonemon. i.e. Language != written text

- Phonetics: how to classify vowels & consonants?

- Writing systems: logographic, syllabary, alphabet, abugida, abjad



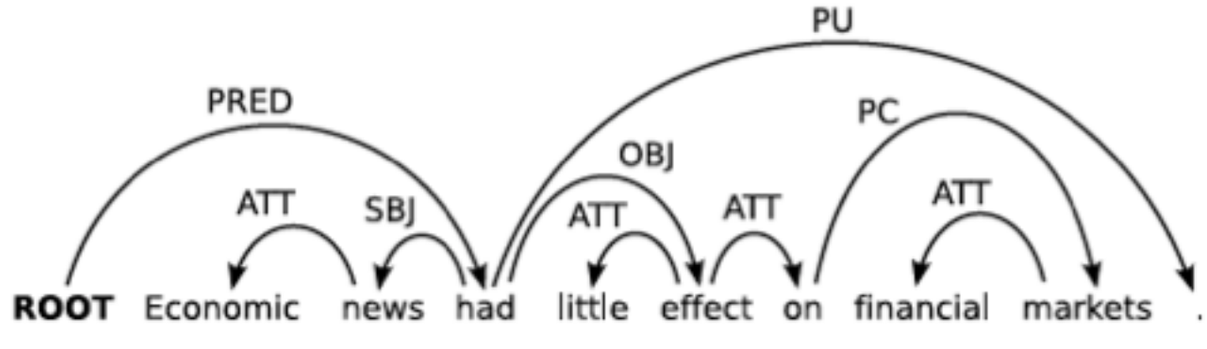



# 4. Linguistics 101

- Morphology
    - Identify morphemes, derivations, and inflections
    - Classify: analytic, agglutinative, fusional
- Syntax:
    - Know that languages may be SVO, SOV, etc. or free-word order. Recognize the diversity in expression: some things expressed in syntax in English might be expressed in morphology in another.
- Semantics vs. Pragmatics — what's the difference?

# 5. Parsing

- What is dynamic programming?
- Be able to implement CKY for constituency parsing
    - i.e. I give you a grammar and a sentence, you give me all the valid parse trees
- Be able to read and understand a dependency parse



# 6. Neural Networks

- Understand how to derive backpropagation equations (don’t need to do it for real, but understand it’s chain rule)

- Basic familiarlity with Word2Vec & Neural LM models

- Be able to figure out how to set parameters to a simple neuron, e.g. midterm question on fitting AND/OR operators

# 7. HMM

- What's it good for? Sequence labeling problems
- Mathematical form $P(O,Q) = P(O|Q)P(Q) = \prod_{t=1}^{T} P(o_t|q_t) \times \prod_{t=1}^{T} P(q_t|q_{t-1})$
- What are the 3 problems for an HMM
    - Be able to compute likelihood and decode by hand, if given model parameters and observation
    - Understand the basic mechanics of Baum-Welch
- Basically, master the homework assignment

# 8. Topic Models

- High level differences (advantage/disadvantages) between Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA)

- Generative story of LDA and how it maps to graphical model. Be clear about what's 'latent', 'observed', 'hyperparameter' variables in the model

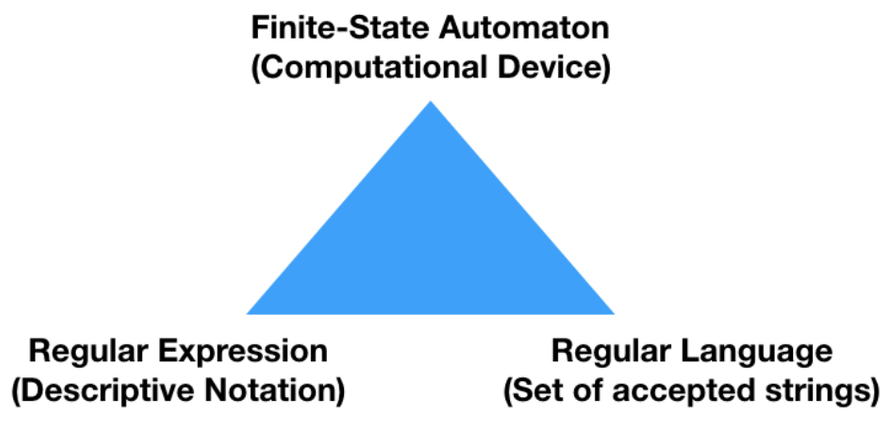
# 9. FST

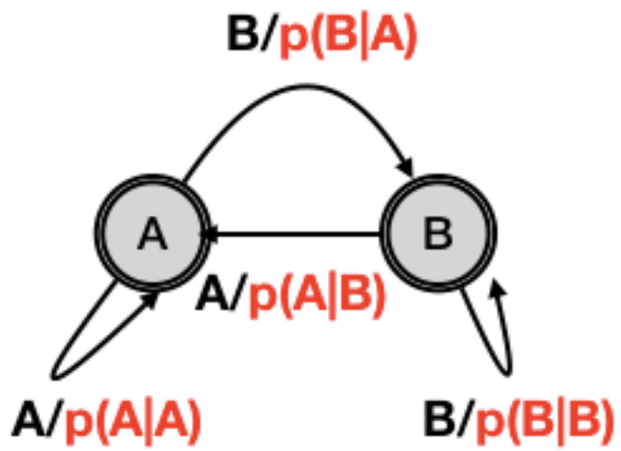
- Understand relationship between FSA, RegEX, set of strings

- If given a weighted FST, be able to explain what it's doing; be able to generate output based on some input.

- Be able to do the above under different semirings.

- Understand the purpose of various FST operations, e.g. composition



# 10. Semantics

- Recognize that semantics is hard, that different “kinds” of semantics work for different purposes.
- Distributional Semantics: what is it? some examples?
- Word sense:
  - Be able to think of multiple senses when given a word
  - Describe relations between senses, e.g. hypernym
  - What’s a synset? How is WordNet organized?

# 10. Semantics

- Semantic Role Labeling: what is the task and how does this relate to semantics?
    - Identify different thematic roles in a sentence
- PropBank and FrameNet:
    - How do these approaches differ?
    - If given an annotation, explain in words what it means

[Arg0 The shop] increased [Arg1 the price] [Arg3 today]

# 11. Structured Prediction

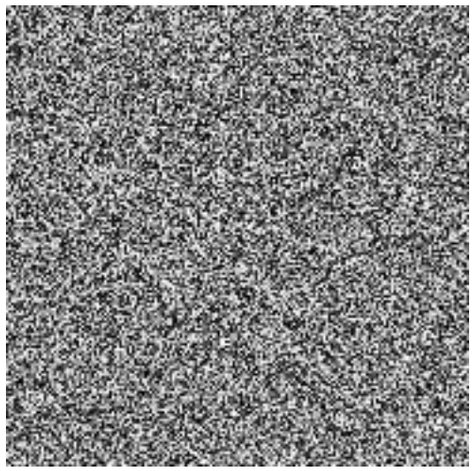
- What problems can be called structured prediction?
- Generative vs. Discriminative.
    - Explain the difference in terms of math. Explain the benefits of each. Understand how bad features affect generative models.
- Local vs. Global. Explain the difference. Explain the label bias problem.
- CRF - how does it relate to log-linear models?
- Structured Perceptron - understand the pseudocode
- When to use CRF, Structured Perceptron, or Structured SVM?

# Final Remarks

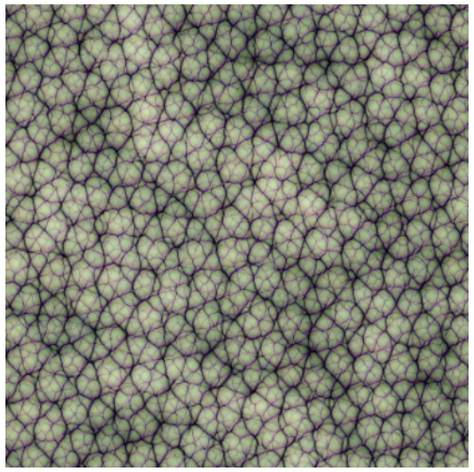
# Two Major Themes

- Language has structure
  - There are patterns in what we say; this can be exploited this for more efficient learning and inference
- Language processing involves ambiguity resolution
  - There is ambiguity in what we say; this has to be resolved, e.g. by probabilistic models

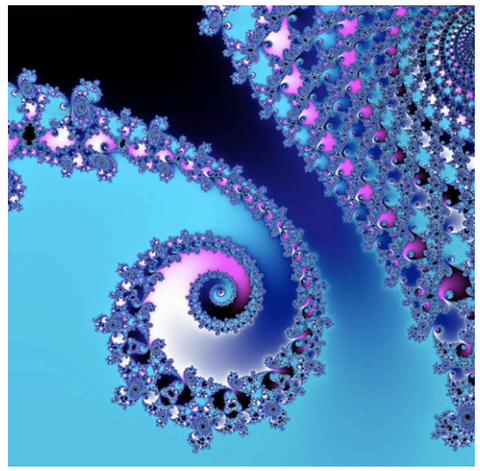
# What is structure?



No apparent structure



Some structure



Some structure

**Structure = there is some pattern, not just randomness**

# What is structure?

**sdfpfkgpowkpork**
**opvsdkofaewpewmd**
**fdfadffpkbwkr**

**No apparent structure**

**abc abc abc**
**efg efg efg**
**xyz xyz xyz**

**Some structure**

# How do you describe this image?

**There are infinite set of sentences:**

**a cute dog**
**a very cute dog**
**super cute puppy**
**adorable puppy looking at me**
**....**

**But not all are likely:**
**dog cute a**
**dog cut a very**
**puppy cute super**
**me at puppy looking adorable at**
**....**

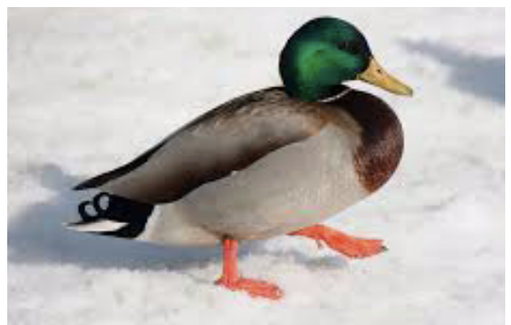
# Language is full of ambiguity

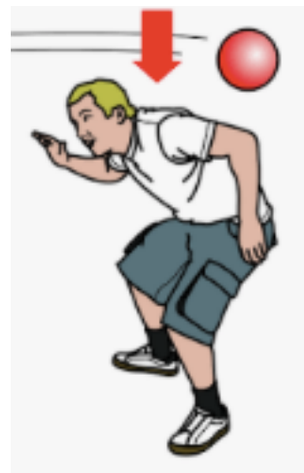
- “I made her duck”

**From: Jurafsky and Martin, Speech and Language Processing, 2nd ed.**

# To resolve ambiguity, we'll be exploring different probabilistic models

- Counting statistics
  - e.g. how many times "duck" means





- Linear (or log-linear) models
  - e.g. extract features: "I", "made", "her", "duck" and combine with weights
- Neural network models
  - can be viewed as a logical extension of linear models

# Questions? Comments?
# Thanks for a good semester!

감사합니다 Natick

Grazie Danke Ευχαριστίες Dalu obrigado

Thank You Köszönöm

Tack

Спасибо Dank Gracias

谢谢 Merci Seé

ありがとう