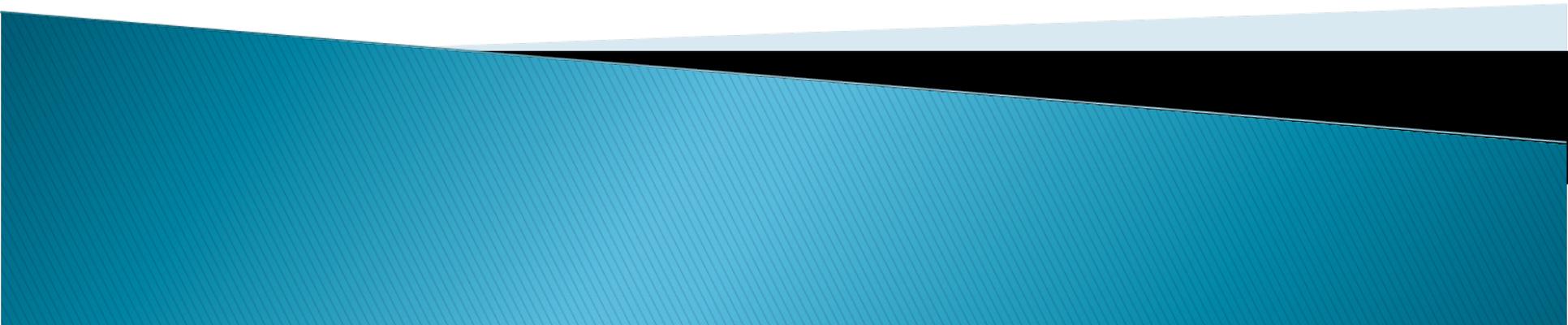


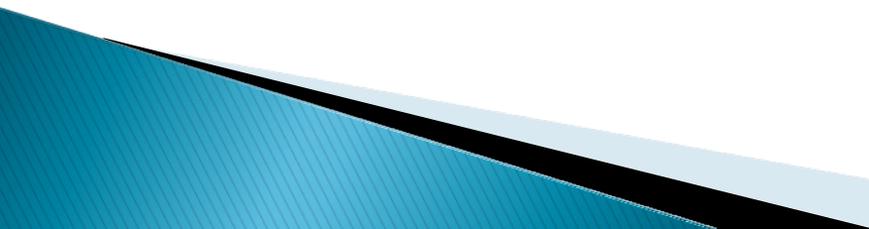
Midterm Review

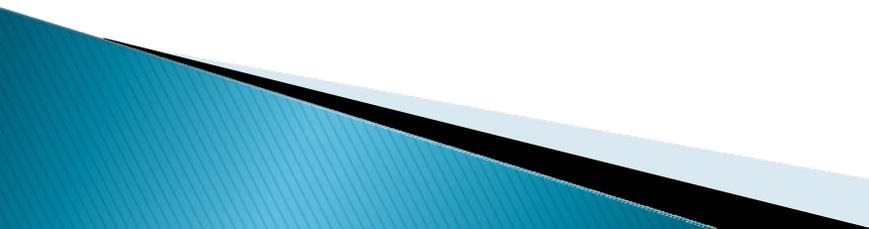
CS4705

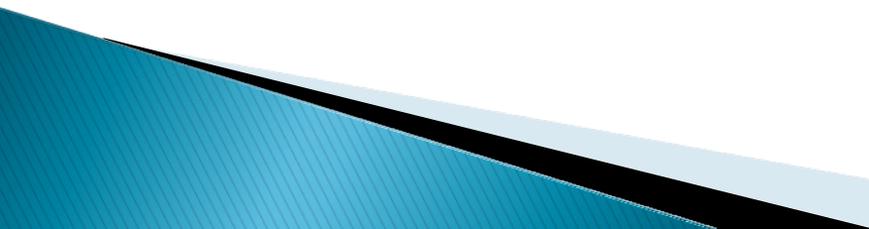
Natural Language Processing

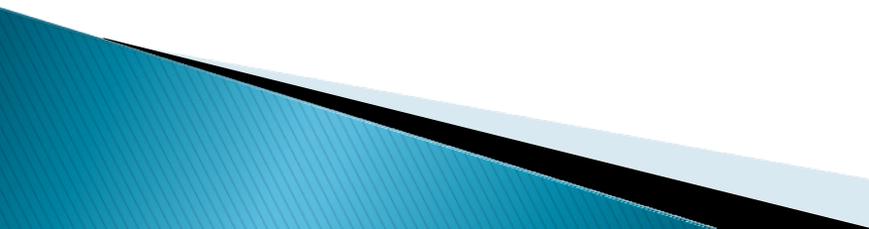


Midterm Review

- ▶ Regular Expressions
 - ▶ Finite State Automata
 - Determinism v. non-determinism
 - (Weighted) Finite State Transducers
 - ▶ Morphology
 - Word Classes and POS
 - Inflectional v. Derivational
 - Affixation, infixation, concatenation
 - Morphotactics
 - Different languages, different morphologies
 - Evidence from human performance
- 

- ▶ Noisy channel model
 - Bayesian inference
 - ▶ N-grams
 - Markov assumption
 - Chain Rule
 - Language Modeling
 - Simple, Adaptive, Class-based (syntax-based)
 - Smoothing
 - Add-one, Witten-Bell, Good-Turing
 - Back-off models
 - ▶ Creating and using ngram LMs
 - Corpora
 - Maximum Likelihood Estimation
- 

- ▶ Testing and Training
 - How to choose a corpus
 - How to divide
 - ▶ Part-of-Speech Tagging
 - Hand Written Rules v. Statistical v. Hybrid
 - Brill Tagging
 - HMMs
 - ▶ Syntax
 - Parse Trees
 - Constituent Structure vs. Dependency Structure
 - What is a good parse tree?
 - Types of Ambiguity
- 

- ▶ Context Free Grammars
 - Top-down v. Bottom-up Derivations
 - Early Algorithm
 - Normal Forms (CNF)
 - Modifying the grammar
 - ▶ Probabilistic Parsing
 - Derivational Probability
 - Computing probabilities for a rule
 - Choosing a rule probabilistically
 - Lexicalized head driven grammars
- 

▶ Semantics

- Where it fits
- Thematic roles
- First Order Predicate Calculus as a representation
- Compositional Semantics
- Word relations
- Word sense disambiguation
 - Naïve Bayes
 - Decision rules
 - Lesk simplified and corpus
 - Selectional restrictions