Grading

- Grading
  - Homework (~~10%~~ 20%)
  - ~~Midterm (40%)~~
  - Final (~~40%~~ 60%)
  - Project (~~10%~~ 20%)

- The final grade will be given on a curve

Final Exam

- TBD…
- Most likely:
  - in the 18-19th weeks

Project

- #1. Programming assignments ← the default option
  - 2 problems
  - To be done individually, no grouping
- #2. Define your own project
  - Use what you have learned in this course to solve a problem of your choice.
  - At most 4 people in a group
  - Expected workload: (#1 workload) × (group size)
  - Evaluation criteria
    - relevance to this course
    - substance, soundness, novelty
    - quality of the report and presentation
- Available at BB course menu → Project

Project

- Schedule
  - #1
    - To be released today
    - Due: 11:59pm, June 12, Sun
  - #2
    - Presentation on June 8, Wed
    - Report due: 11:59pm, June 10, Fri
  - Make your choice by 11:59pm, May 25, Wed
    - If option #2, enter your group members
    - BB course menu → Project → Project Registration

Project

- Plagiarism
  - All projects must be done by yourself
  - #1: Do not use any external code (other than those indicated in the problem description)
    - Plagiarism detection software will be used
  - #2: You may use external libraries, but must list them in your presentation and report
    - We will manually check your submitted code
- Plagiarism punishment
  - Zero point on the project
  - When one student copies from another student, both students are responsible

Lectures

- Week 15-17
  - No lectures
  - I will hold QA sessions online during 8:30-10am MW
    - Same Tencent Meeting ID
- June 8, Wed (week 17)
  - Presentation for project option #2

Final Review

What we have covered

2. Text normalization
3. Word representation
4. Text classification
5. Text clustering
6. Language modeling
7. Contextual word representation
8. Sequence labeling
9. Constituency parsing
10. Dependency parsing
11. Lexical semantics
12. Sentence semantics
13. Information extraction
14. Discourse analysis
15. Sequence to sequence

What we have covered -- grouped

- Basics
  - Text normalization
  - Word representation, Lexical semantics
  - Text classification
  - Text clustering
- Sequences
  - Language modeling
  - Pretrained LM & contextual word representation
  - Sequence labeling
  - Sequence to sequence
- Structures
  - Constituency parsing
  - Dependency parsing
  - Sentence semantics
  - Discourse analysis
- Applications
  - Information extraction

Text Normalization

- Word tokenization
  - Regular expression, BPE
- Word normalization
  - Lemmatization, stemming
- Sentence segmentation
- Methods: Rules + ML

Word Representation

- Sparse vector representations
  - Co-occurrence matrices
  - Weighting: tf-idf, PPMI
- Dense vector representations
  - Singular value decomposition, LSA
  - Word2vec: Skip-gram
- Evaluation

Lexical Semantics (Symbolic Word Representation)

- Word Senses
- WordNet
  - Organizing word senses according to their semantic relations
- Word Sense Disambiguation

Text Classification

- Rule-based methods
  - Regular expression
- Machine learning methods
  - Generative classifiers
    - Naive Bayes
  - Discriminative classifiers
    - Logistic regression
- Evaluation
  - Precision, recall, F-measure
  - Macro-/micro-averaging

Text Clustering

- Mixture of Gaussian
- (Unsupervised) Naive Bayes
- Topic models
  - pLSA, LDA
- Learning
  - Expectation-maximization

Language Modeling

- Compute the probability of a sentence
  - Chain rule: predicting the next word
  - Evaluation: perplexity
- n-gram LM
  - Probability of each word is conditioned on the preceding n-1 words.
- Recurrent neural networks
  - Probability of each word is conditioned on a hidden vector summarizing all the preceding words
- Transformers
  - Probability of each word is computed by attending to preceding words

Contextual word representations (& pretrained LM)

- ELMo
  - BiLSTM + LM
- BERT
  - Transformer + MLM (+NSP)
- Pretrain+finetune paradigm

Sequence Labeling

- Hidden Markov model (HMM)
  - Inference: Viterbi, Forward, Backward
  - Learning: Maximum Likelihood Estimate, Expectation-Maximization / SGD
- Conditional random filed (CRF)
  - Label bias problem
  - Inference: Viterbi, Forward, Backward
  - Learning: conditional likelihood, margin-based loss, CRF-AE
- Neural models
  - Neural softmax, neural CRF

Sequence to Sequence

- Many applications
  - MT, paraphrase, summarization, ...
- Methods: encoder-decoder
  - Recurrent neural network (with attention)
  - Transformer: cross-attention
- Learning
  - Maximizing conditional likelihood on a parallel corpus
- Decoding
  - Greedy, beam-search
- Extensions
  - Pointer Net / Copy Mechanism
  - Seq2Set, X2Seq, Null2Seq

Constituency Parsing

- Context-Free Grammars
  - Terminals, nonterminals, start symbol, production rules
  - Probabilistic Grammars: each rule has a probability
- Parsing
  - CYK algorithm on CNF
- Learning
  - Supervised: generative & discriminative methods
  - Unsupervised: inside-outside algorithm
- Beyond CFG
  - RG, MCSG (TAG)
  - Span-based, transition-based, sequence labeling, top-down splitting

Dependency Parsing

- Dependency Parsing
  - Projectivity
  - Relation to constituency parsing
- Graph-based parsing
  - 1st-order: Eisner, MST
  - Learning
    - Supervised: discriminative methods
    - Unsupervised: EM, CRF-AE
- Transition-based parsing
  - Arc-standard
  - Learning: from configuration to transition

Sentence Semantics

- Vector vs. symbolic representation of sentences
- Formal Meaning Representation
  - Special-purpose representations
  - General-purpose representations: formal logic, semantic graphs
- Syntax-Driven Semantic Parsing
  - λ-Calculus, Semantic Attachments to CFG
- Neural Semantic Parsing
  - Seq2seq, parsing to graph, ...
- Semantic Role Labeling
  - PropBank, FrameNet
  - Methods: sequence labeling, graph-based methods

Discourse Analysis

- A discourse is a coherent structured group of sentences.
  - Text spans are connected with coherence relations.
  - These relations form a hierarchical structure.
  - Discourse parsing: EDU segmentation + RST parsing

- Coreference Resolution
  - Mention Detection
  - Mention Clustering
    - Binary classification vs. ranking

Information Extraction

- Subtasks
  - Named entity recognition
  - Relation extraction
  - Event extraction
  - ...
- Methods
  - Sequence labeling
  - Span/arc classification
  - Joint extraction
  - ...

Final Remarks

Topics not covered in this course...

- Text generation
- Question answering
- Dialog
- Multilingual NLP
- Multimodal NLP (language+X)
- Interpretability
- Biases in NLP
- Adversarial NLP
- …

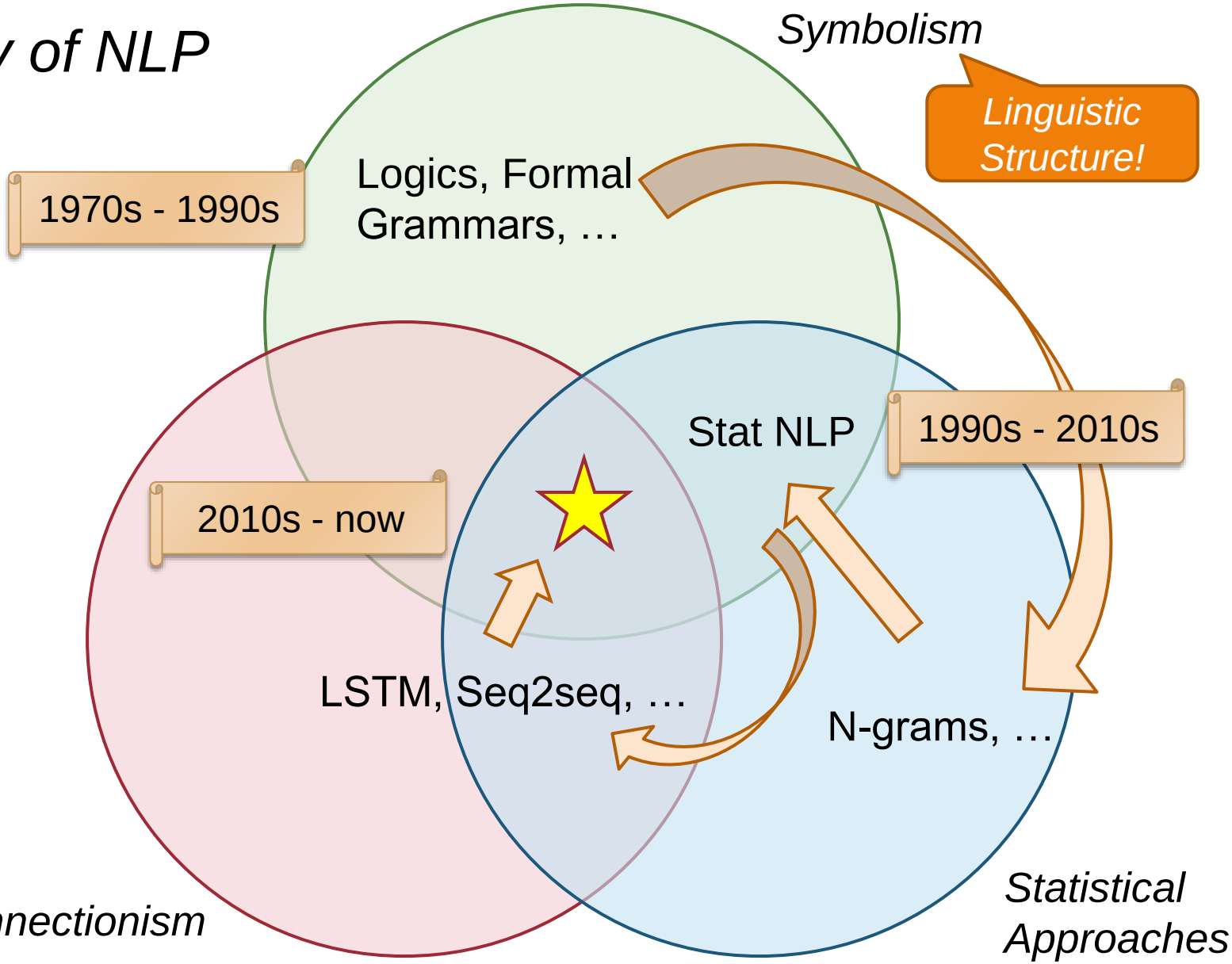
Where to learn more…

- Text books: SLP3, INLP, …
- Online lectures: Stanford CS224n, …
- Research papers
  - Conferences
    - ACL: Meeting of the Association for Computational Linguistics
    - EMNLP: Conference on Empirical Methods in Natural Language Processing
    - NAACL: Conference of the North American Chapter of the Association for Computational Linguistics
    - COLING, EACL, AACL, CoNLL, SemEval, …
    - AI/ML conferences
  - Journals
    - Computational linguistics
    - Transactions of the Association for Computational Linguistics (TACL)

Doing research at SIST... (for undergraduates)

- My research group
  - Focus
    - Linguistic structures: representation, inference, learning
  - Methodology
    - A combination of symbolic, statistical, and neural approaches
    - Learning under low resources
  - Applications
    - Mainly NLP, but also: CV, knowledge representation & reasoning, probabilistic modeling, …
  - Undergraduate interns
    - Talk to me if GPA≥3.5 for courses in math, CS basics, and AI
- Other groups
  - CV+language

History of NLP



That's all!

Good luck in your project and final exam!

CS274A Spring 2022