

# Final Exam

---

- ▶ Time
  - ▶ in class (8:15-9:45am) on May 30 (Thu)
  - ▶ 90 minutes
- ▶ Location
  - ▶ 教学中心 201 (this classroom)
  - ▶ Seat arrangement will be announced later
- ▶ Format
  - ▶ Similar to homework
  - ▶ Closed-book. You can bring an A4-size cheat sheet and nothing else.
  - ▶ TBA: 涂卡笔、计算器
- ▶ Grade
  - ▶ 60% of the total grade





# Final Review



# Disclaimer

---

- ▶ Topics covered in this review may not appear in the exam.
- ▶ Topics not covered in this review may appear in the exam.



# Text Normalization

---

- ▶ Word tokenization
  - ▶ Regular expression, BPE
- ▶ Word normalization
  - ▶ Lemmatization, stemming
- ▶ Sentence segmentation



# Text Representation

---

- ▶ Word representation
  - ▶ Sparse vector: word-word matrix, PPMI
  - ▶ Dense vector: word2vec (skip-gram)
- ▶ Document representation
  - ▶ Sparse vector: term-document matrix, TF-IDF
  - ▶ Dense vector: LSA, neural methods



# 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
- ▶ Evaluation



# 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 (LSTM, GRU, +Attention)
  - ▶ 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





# Sequence to Sequence

---

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



# Pretrained LM

---

- ▶ Pretraining: the new paradigm
- ▶ Encoder (auto-encoding)
  - ▶ ELMo: two LSTMs + LM
  - ▶ BERT: Transformer + MLM
- ▶ Encoder-Decoder
  - ▶ BART, T5
- ▶ Decoder (auto-regressive)
  - ▶ GPT: Transformer + LM + ...
  - ▶ LLM
- ▶ Utilizing PLMs
  - ▶ Finetuning, parameter efficient tuning
  - ▶ Prompting



# Sequence Labeling

---

- ▶ Many applications
- ▶ Hidden Markov model (HMM)
  - ▶ Inference: Viterbi, Forward, Backward
  - ▶ Learning: Maximum Likelihood Estimate, Expectation-Maximization / SGD
- ▶ Conditional random field (CRF)
  - ▶ Undirected discriminative models
  - ▶ Inference: Viterbi, Forward, Backward
  - ▶ Learning: conditional likelihood, margin-based loss, CRF-AE
- ▶ Neural models
  - ▶ Neural softmax, neural CRF



# Constituency Parsing

---

- ▶ Concepts, evaluation
- ▶ Span-based Parsing
  - ▶ Tree score = sum of constituent scores
  - ▶ Parsing: CYK
  - ▶ Learning: discriminative methods
- ▶ (Probabilistic) Context-Free Grammars
  - ▶ Tree score = product of rule probabilities
  - ▶ Parsing: CYK
  - ▶ Learning
    - ▶ Supervised: generative & discriminative methods
    - ▶ Unsupervised: EM with inside-outside algorithm / SGD
- ▶ Transition-based parsing
  - ▶ Tree score = product of action probabilities
  - ▶ Bottom-up parsing



# Dependency Parsing

---

- ▶ Concepts, evaluation
- ▶ Relation to constituency parsing
- ▶ Graph-based parsing
  - ▶ 1st-order: Eisner
  - ▶ Learning
    - ▶ Supervised: discriminative methods
    - ▶ Unsupervised: EM, CRF-AE
- ▶ Transition-based parsing
  - ▶ Arc-standard



# Semantics

---

- ▶ Vector vs. symbolic representation
- ▶ Lexical Semantics
  - ▶ Word Senses
  - ▶ WordNet: organizing word senses according to semantic relations
  - ▶ Word Sense Disambiguation



# Semantics

---

- ▶ Sentence Semantics
  - ▶ Formal Meaning Representation
    - ▶ Special/general-purpose representations (logic, semantic graph)
  - ▶ Syntax-Driven Semantic Parsing
    - ▶  $\lambda$ -Calculus, Semantic Attachments to CFG
  - ▶ Neural Semantic Parsing
    - ▶ Seq2seq, parsing to graph
  - ▶ Semantic Role Labeling
    - ▶ PropBank, FrameNet
    - ▶ Methods: sequence labeling, graph-based methods, seq2seq
  - ▶ Information Extraction
    - ▶ NER, Entity Linking, Relation Extraction, Event Extraction



# 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
    - ▶ Mention ranking







## Final Remarks



# Topics covered in this course...

---

## ▶ Basics

- ▶ Text normalization
- ▶ Text representation
- ▶ Text classification
- ▶ Text clustering

## ▶ Structures

- ▶ Constituency parsing
- ▶ Dependency parsing
- ▶ Semantics
- ▶ Discourse analysis

## ▶ Sequences

- ▶ Language modeling
- ▶ Sequence to sequence
- ▶ Pretrained language models
- ▶ Sequence labeling



# Topics not covered

---

- ▶ Information Retrieval
- ▶ Multilingual and Cross-Lingual NLP
- ▶ Multimodal NLP (language+X)
- ▶ Interpretability
- ▶ Biases & ethics in NLP
- ▶ Adversarial NLP
- ▶ Pragmatics
- ▶ ...



# Topics not covered in detail

---

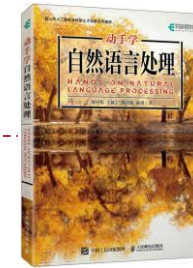
- ▶ Machine Translation
- ▶ Question answering
- ▶ Dialog
- ▶ Sentiment Analysis
- ▶ Summarization
- ▶ ...



# Where to learn more...

---

- ▶ Text books: SLP3, INLP, 动手学NLP, ...
- ▶ 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 (CL)
    - ▶ Transactions of the Association for Computational Linguistics (TACL)



# Doing NLP research at SIST... (for undergraduates)

---

- ▶ My research group

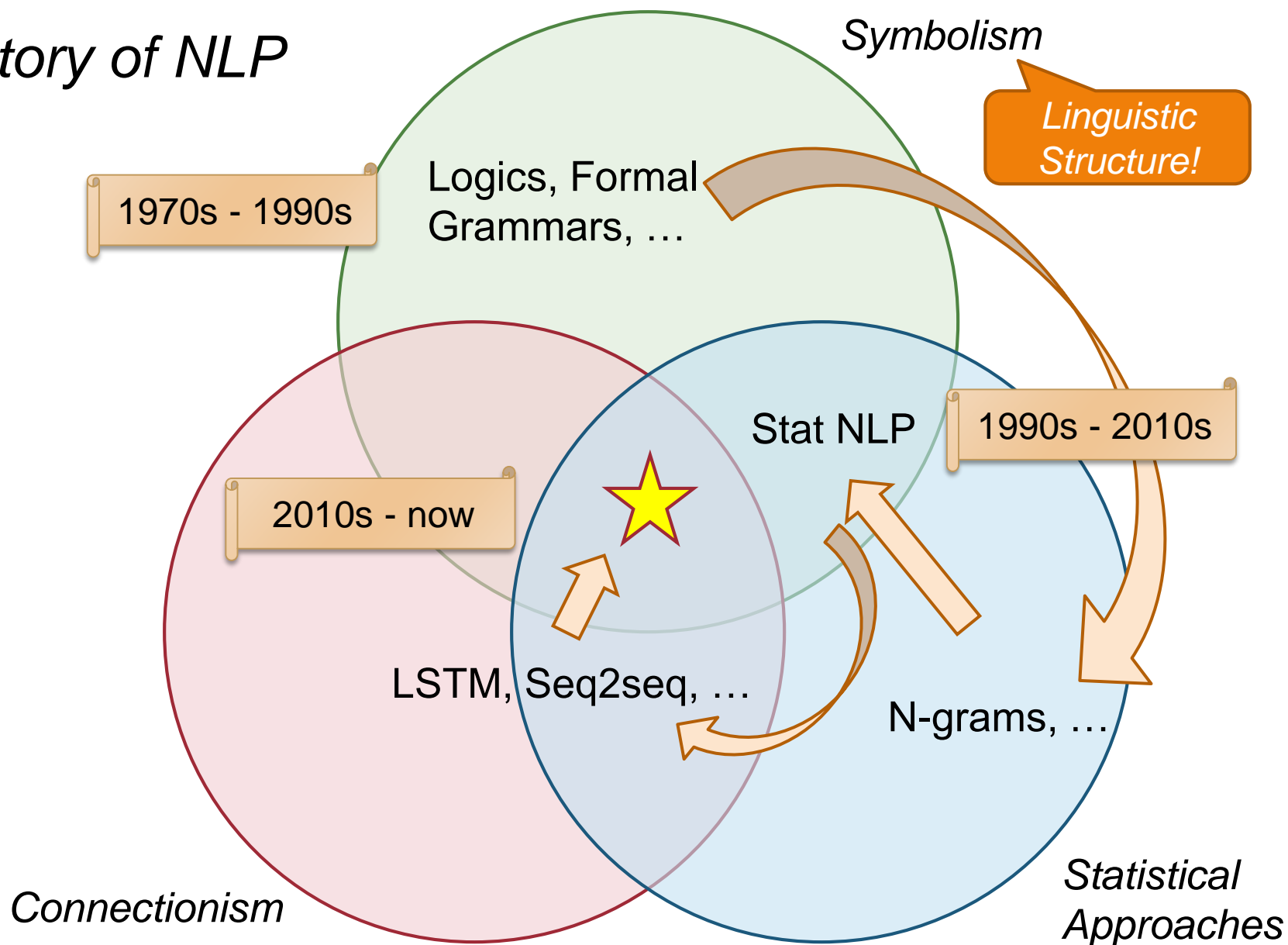
- ▶ Combining symbolic, probabilistic and neural approaches to the representation, learning and utilization of linguistic structures
- ▶ Integrating linguistic structures into transformers and LLMs

- ▶ Other groups

- ▶ CV+language (He, Yang, etc.)
- ▶ NLP safety (Wang, etc.)



# History of NLP



# The past 1+ year

---

- ▶ The revolution of LLM!
  - ▶ Huge impact not only in NLP, but also spreading to other fields
- ▶ Near future
  - ▶ Rise of general purpose NLP: one model for all tasks
  - ▶ Improvements & innovations in foundation models
- ▶ Long term
  - ▶ Integrating symbolic & statistic NLP into neural NLP
  - ▶ Incorporating linguistic structures in foundation models
  - ▶ Towards better performance, efficiency, theoretical foundation, interpretability, ...





*That's all!*  
*Good luck in your project and final exam!*

CS274A Spring 2024