Project

- Programming assignments
  - BB course menu → Project
  - To be done individually, no grouping
- Schedule
  - Due: 11:59pm, May 25, Thu

Project

- Plagiarism
  - The project must be done by yourself
  - Do not use any external code (other than those indicated in the problem description)
  - Plagiarism detection software will be used
- Plagiarism punishment
  - Zero point on the project
  - When one student copies from another student, both students are responsible

Final Exam

- Time
  - in class (10:15-11:45am) on May 9 (Tue)
  - 90 minutes
- Location
  - 教学中心 201
  - Seat arrangement will be announced later
- Format
  - Similar to midterm
  - Closed-book. You can bring an A4-size cheat sheet + a calculator and nothing else.
  - 建议带涂卡笔
- Grade
  - 35% 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.

Constituency Parsing

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

Dependency Parsing

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

Lexical Semantics

- (Symbolic word representation)
- Word Senses
- WordNet
  - Organizing word senses according to their semantic relations
- Word Sense Disambiguation

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, seq2seq

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
  - Constituency/dependency parsing
  - Joint extraction
  - Decoding based
  - ...

Final Remarks

Topics covered in this course...

- Basics
  - Text normalization
  - Text representation
  - Text classification
  - Text clustering
- Sequences
  - Language modeling
  - Pretrained language models
  - Sequence labeling
  - Seq2seq
- Structures
  - Constituency parsing
  - Dependency parsing
  - Semantics
  - Discourse analysis
- Applications
  - Information extraction

Topics not covered in this course...

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

Where to learn more...

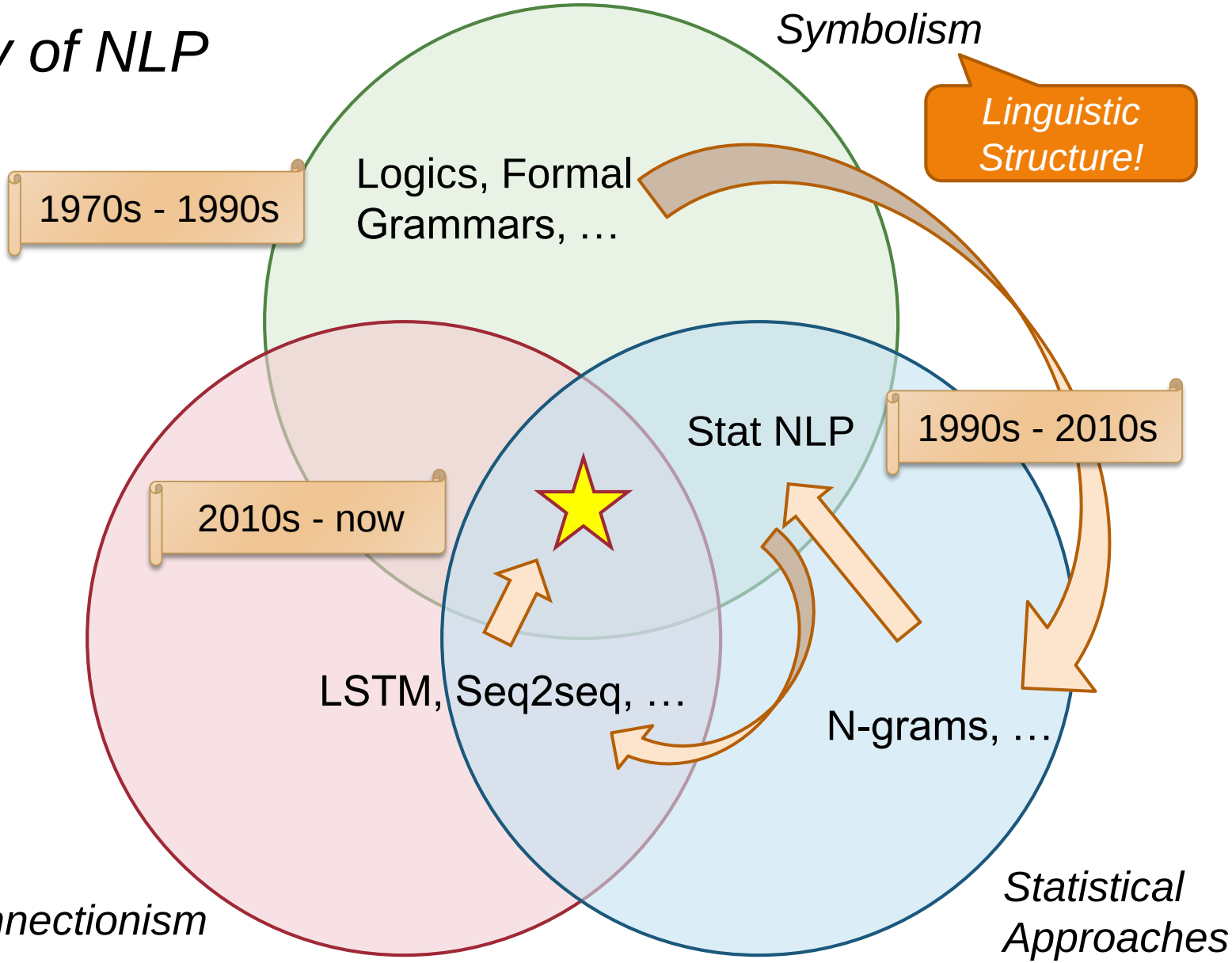
We will post new chapters in the following months.

- 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
  - Focus
    - Linguistic structures: representation, inference, learning
  - Methodology
    - A combination of symbolic, statistical, and neural approaches
    - Integration with PLMs
  - Applications
    - Mostly NLP, but also: CV, KR, probabilistic modeling, …
- Other groups
  - CV+language (He, Yang, etc.)

History of NLP



Now?

- The revolution of LLM!
  - Very fast development over the past 0.5 year
  - Huge impact not only in NLP, but also spreading to other fields
- Trends
  - Rising of general purpose LLMs
  - Demising of many intermediate tasks?
    - But some techniques may still be useful in LLMs
- Future
  - Lots of exciting new developments in the following years
  - The dawn of the next industrial revolution?

That's all!

Good luck in your project and final exam!

CS274A Spring 2023