Natural Language Processing

Review of course topics

Regular expressions

- Basic regular expression patterns
- Disjunction, grouping, and precedence
- Advanced operators

Finite-State Automata

- Relationship to regular expressions
- Formal definition
- Recognition/Generation
- DFA
- NFA
- Relationship between DFA and NFA

Words and Transducers

- Morphology
- Morphological parsing
  - Parsing vs. generation
- Finite-state transducers
  - Lexical level
  - Intermediate level
  - Surface level

N-Grams

- Unigrams, bigrams, trigrams
- Types vs. tokens
- Lemma vs. stem
- Language modeling
  - Conditional probability
  - Chain rule
  - Markov assumption
  - Maximum Likelihood Estimate (MLE)

N-Grams

- Evaluation
  - Training set, test set
- Unknown words
- Zero counts => smoothing

Text processing tools

- grep
- sed
- sort
- uniq
- paste
- head, tail
- awk
- JFlex

Part-of-Speech (PoS) tagging

- Different PoS
- Purpose of PoS tagging
- PoS tagset
- Ambiguity
- Different PoS tagging methods
  - Rule-based
  - Data-driven

Part-of-Speech (PoS) tagging

- Hidden Markov Model (HMM) tagging
  - Transition/contextual probabilities
  - Emission/observation probabilities
- Evaluation
  - Gold standard
  - Training
  - Testing
  - Cross-validation

Python and NLTK

- Format of a program / syntax
- Data types
- Control structures
- Data structures
- Functions
- Regular expressions
- Basics of NLTK

Formal grammars

- Context-free grammars
  - Terminals
  - Non-terminal
  - Rules
  - Derivations
- Constituents
- Treebanks
- Dependency grammars

Parsing

- Parsing and CFG
- Parse trees
- Top-down vs. bottom-up parsing
- CKY parsing
- Earley parsing
- Ambiguity in parsing
- Probabilistic parsing
- Partial/shallow parsing

Meaning Representation

- Semantic Analysis
- Meaning Representation Language
- Meaning of linguistic input vs. state of world → Linking the two
- Requirements of a useful meaning representation
  - Verifiable, unambiguous, canonical, inference, variables, expressive

Model-Theoretic Semantics

- Model
- Mapping expressions in meaning rep. language to elements of model
  - Non-logical vocabulary
  - Logical vocabulary
  - Denotation
  - Domain of a model
  - Objects, properties and relations
- Determining truth of complex expressions
  - Meaning of logical operators

First-Order Logic

- The kind of world that can be represented
  - Objects, properties, relations among objects
- Basic elements of FOL
  - Constants, functions, predicates, variables, quantifiers
- Semantics of FOL
- Syntax-driven semantic analysis
  - Semantic attachments as augmentation to context-free gammars

Discourse

- Discourse Analysis
- Purpose of language
  - Transmitting information vs. managing social relationships
- Discourse Function vs. Discourse Device
- Coherence vs. cohesion (lexical cohesion vs. non-lexical cohesion)
- Discourse Structure and Segmentation
  - Coherence relation, Rhetorical Structure Theory

Discourse

- Discourse Context
  - Cognitive, Situational and Textual
- Discourse Model, Discourse Entity
- Referring Expression, Anaphora, Referent, Coreferring,
- Reference resolution
  - Preferred potential referents
- Information Status, Givenness, Accessability

Conversation

- Turn-taking
  - Transition-relevant places
- Adjacency Pairs
- Speech Acts
  - Locutionary, Illocutionary , Perlocutionary
  - Assertives, Directives, Commissives, Delcaratives, Expressive

Grounding

- Acting collectively, cooperatively
- Principle of closure
- Grounding methods
  - Continued attention, next contribution, acknowledgement, demonstration, display
- Use of nonverbal cues

Conversation Structure

- Overall organization
- E.g. conversational openings
  1. Enter conversation (summon-response)
  2. Identify speakers
  3. Establish joint willingness to converse
  4. Raise first topic

Implicature

- Drawing inferences
  - Based on assumptions about speaker's behavior
- Gricean Maxims
  - Truth, Information, Relevance, Orderly

Speech Recognition

- Noisy Channel Model
- Simple Architecture
  - Feature extraction
  - Acoustic model + lexicon
  - Language model
  - Decoding/search

Nonverbal Behavior

- Evidence of communicative purpose
  - Social, form, synchrony and context
- Nonverbal behavior as discourse devices
- Studying the correlation of device and function
- Some results for: Gaze, gesture, posture
- Computer applications of this research
  - Speech Understanding, Animated characters, Computer Mediated Communication

Dialog Systems

- Basic Dialogue System Architecture
- Natural Language Understanding
  - frame-and-slot filling
  - semantic grammars vs. semantic HMMs
- Natural Language Generation
  - Template-based
  - Full natural language generator
- Text-to-Speech
  - Parametric control vs. concatenation

Dialog Manager

- Typical approaches
  - Finite-state, frame-based, information-state based (don't have to describe), plan-based (don't have to describe)
- System-initiative finite-state
- Mixed-initiative frame-based
- System Grounding
  - Explicit vs. implicit
  - Rejection and reprompting