Final Review

Linguistics 445/515

Autumn 2008

The final exam will be Monday, December 15, 8-10am.

1 Topics to be covered

- Writers' Aids
- Machine Translation (MT)
- Dialogue Systems
- Computer-Aided Language Learning (CALL)
- Social context of language technology use

2 Format of the exam

1. Matching: 10 terms [see below]
2. Multiple choice: 5 questions
3. "Calculations": 3-7 questions
    - Minimum edit distance
    - Bigram array (positional and non-positional)
    - Confusion matrix
    - Noisy Channel Model
        - Conditional probabilities
        - Bayes' Law
    - Bigram/Trigram real-word spell checkers (using confusion sets)
    - Phrase structure rules & syntactic trees
    - Venn Diagrams
    - Bag of words method
    - Statistical MT
    - Phrase-based translation
    - Dialogue mark-up: adjacency pairs, speech acts
    - Pronoun resolution
    - Finite-state automata

- Concordancing
- Mal-rules

4. Short answer: 5-10 questions

   - Types and causes of spelling errors
   - Context-sensitive spelling correction for web queries
   - Context-free grammars
   - Syntactic trees
   - Transformer systems
   - Direct transfer systems
   - Interlinguas
   - MT maching learning: text alignment, bag of words, expectation-maximization algorithm
   - MT evaluation and uses
   - What makes MT difficult
   - Basic facts about dialogue
   - Conversational (Gricean) maxims
   - ELIZA/chatterbot architecture
   - Parsing natural dialogue
   - Dialogue system architecture
   - Second Language Acquisition (SLA)
   - Using NLP in CALL
   - Authentic-text ICALL
   - Parser-based ICALL, including parsing ill-formed input
   - CALL exercise types (and motivation for them)
   - Feedback types and uses
   - Social context of language technology use

3 Terms to know

3.1 Writers' Aids

- productivity
- inflection
- tokenization
- detection
- correction
- homophone
- word recognition
- interactive spelling checkers
- automatic spelling correctors
- phonetic errors
- run-on errors
- split errors
- nonword error detection
- isolated-word error correction
- context-dependent word correction
- dictionary lookup
- dictionary construction
- insertion
- deletion
- substitution
- transposition
- single-error misspelling
- multi-error misspelling
- acyclic

- array
- positional bigram array
- nonpositional bigram array
- domain-specificity
- minimum edit distance
- dynamic programming
- transition probabilities
- confusion probabilities
- syntax
- linear order
- constituents
- syntactic tree
- category
- lexical category
- morphology
- function word
- open class
- closed class
- phrase structure rule
- generative
- structurally ambiguous
- hierarchical
- recursive
- infinite
- context-free grammar
- pushdown automaton
- stack
- parsing
- local syntactic errors
- long-distance syntactic errors

3.2 Machine Translation

- translation
- machine translation
- (lexical) ambiguity
- structural ambiguity
- source language
- target language
- hyponym
- hypernym
- synonym
- interlingua
- translation triangle
- dictionary
- lexical gap
- light verb
- idiom
- reversibility
- transformer architecture
- direct transfer
- transfer component
- comparative grammar
- robustness
- text alignment
- sentence alignment
- word alignment
- machine learning
- training data
- bag of words
- underlying representation
- intelligibility
- accuracy
- error analysis
- test suite

3.3 Dialogue Systems

- Turing Test
- ELIZA (chatterbots)
- implementations
- pattern recognition
- canned phrases
- case-based reasoning
- discourse
- turn-taking
- adjacency pairs
- utterances
- common ground
- backchannels
- maxims
- cooperative principle

- performative verbs
- direct speech acts
- indirect speech acts
- DAMSL
- forward-looking functions
- backward-looking functions
- discourse purpose
- subdialogues
- modularity
- dialogue manager
- state transition network
- finite-state automata
- frame-based systems
- templates
- semantic grammars
- confirmation
- repair
- reference
- indefinite noun phrases
- definite noun phrases
- pronouns
- coreference
- pronoun resolution
- selectional restrictions
- human-computer interaction
- natural dialogues
- Wizard of Oz dialogues

3.4 Computer-Aided Languge Learning

- first language acquisition
- stages of learning
- second language acquisition
- language awareness
- frame-based systems
- concordance
- lemmatization (lemma)
- morphological generation
- mal-rules
- error typology
- student model
- feedback
- explicit correction
- recast
- clarification request
- metalinguistic feedback
- elicitation
- repetition
- authentic text ICALL
- exercise generation
- parser-based CALL
- demand-driven architecture

3.5 Social Context of (Language) Technology Use

- Deskilling
- Upskilling
- Artificial intelligence