Semantics:
Semantic Grammar
& Information Extraction

CMSC 35100

Natural Language Processing

May 13, 2003

Potential Projects

- Develop a morphological transducer for a portion of a language of your choice
- Improve the output of a word-for-word translation procedure – say, by improved translation selection, fluency smoothing or other technique
- Implement a keyword spotting system for a small vocabulary
- Build a speech recognizer for a restricted domain, e.g. stock quotes
- Improve the durational modeling of a base ASR system
- Incorporate phrase boundary pause modeling in ASR
- Build a commercial/music/report detector for audio
- Describe and critique a set of at least three contemporary approaches to machine translation, information retrieval, or other area
- Implement a finite-state style information extraction approach for a small domain of your choice

Roadmap

- Domain & Syntax-driven Semantics
  - Semantic Grammar
    - Example
    - Pros & Cons
  - Information Extraction
    - Template filling
      - Cascaded grammar automata
    - Goals & Limitations

Approach II: Semantic Grammars

- Issue:
  - Grammatical overkill
    - Constituents with little (no) contribution to meaning
    - Constituents so general that semantics are vacuous
  - Mismatch of locality
    - Components scattered around tree
- Solution: Semantic Grammars
  - Developed for dialogue systems
    - Tied to domain
    - Exclude unnecessary elements

Semantic Grammar Example

- What do I have on Thursday?
  - CalQ -> What Aux UserP have {on} DateP
    - Cal action:=find; CalOwner:= head UserP; Date:=head DateP;
  - UserP-> Pron
    - Head:=Head Pron
  - Pron-> I
    - Head:= USER
  - DateP -> Dayof Week
    - Head:= sem DayofWeek

Semantic Grammar Pros & Cons

- Useful with ellipsis & anaphora
  - Restrict input by semantic class: e.g. DataP
- Issues:
  - Limited reuse
    - Tied to application domain
  - Simple rules may overgenerate

Information Extraction

- Requirements:
  - Information specified within simple, fixed template
    - Semantic analysis = Slot filling
  - Small proportion of text relevant
- Example: Message Understanding (MUC)
  - Joint venture info, terrorist reports
  - Template:
    - Relationship, Entities, Joint Venture Co, Activity, Amount
      - » Activity: Company, Product, Start date
    - Possible hierarchical template structure

Cascaded Automata & IE

- Typical approach:
  - Cascade of automata build up representation
    - Match syntax & produce semantics
    - Phases:
      - Multi-word unit recognizer: expressions, names, numbers
        - » E.g. San Francisco Symphony Orchestra
      - Basic phrases: noun groups, verb groups,
      - Complex phrases: conjunctions
      - Semantic patterns: e.g. joint venture co
        - » Often domain specific
      - Merging: Combine references to same entity

Information Extraction Issues

- Domain specificity:
  - Only extract small subset of available info
  - Rules tuned to vocabulary, grammar, domain
    - Some systems use learning components
      - Generalize constructions, word use

Lexical Semantics & Word Sense Disambiguation

CMSC 35100

Natural Language Processing

May 13, 2003

Roadmap

- Lexical Semantics
  - Motivation: Word sense disambiguation
  - Meaning at the word level
  - Issues
    - Ambiguity
    - Meaning
    - Meaning structure
      - Relations to other words
      - Subword meaning composition
  - WordNet: Lexical ontology

What is a plant?

There are more kinds of plants and animals in the rainforests than anywhere else on Earth. Over half of the millions of known species of plants and animals live in the rainforest. Many are found nowhere else. There are even plants and animals in the rainforest that we have not yet discovered.

The Paulus company was founded in 1938. Since those days the product range has been the subject of constant expansions and is brought up continuously to correspond with the state of the art. We’re engineering, manufacturing, and commissioning world-wide ready-to-run plants packed with our comprehensive know-how.

Lexical Semantics

- Focus on word meanings:
  - Relations of meaning among words
    - Similarities & differences of meaning in sim context
  - Internal meaning structure of words
    - Basic internal units combine for meaning
- Lexeme: meaning entry in lexicon
  - Orthographic form, phonological form, sense

Sources of Confusion

- Homonymy:
  - Words have same form but different meanings
    - Generally same POS, but unrelated meaning
    - E.g. bank (side of river) vs bank (financial institution)
      - » Bank1 vs bank2
    - Homophones: same phonology, diff't orthographic form
      - » E.g. two, to, too
    - Homographs: Same orthography, diff't phonology
- Why?
  - Problem for applications: TTS, ASR transcription, IR

Sources of Confusion II

- Polysemy
  - Multiple RELATED senses
    - E.g. bank: money, organ, blood,…
  - Big issue in lexicography
    - # of senses, relations among senses, differentiation
    - E.g. serve breakfast, serve Philadelphia, serve time

Relations between Words

- Synonymy:
  - "same meaning": substitutability?
  - Issues:
    - Polysemy – same as some sense
    - Shades of meaning – other associations:
      - Price/fare
    - Collocational constraints: e.g. babbling brook
    - Register: social factors: e.g. politeness, formality
- Hyponomy:
  - Isa relations:
    - More General (hypernym) vs more specific (hyponym)
      - E.g. dog vs golden retriever
  - Organize as ontology/taxonomy

WordNet Taxonomy

- Manually constructed lexical database
  - 3 Tree-structured hierarchies
    - Nouns, verbs, adjective+adverb
    - Entries: synonym set, gloss, example use
- Relations between entries:
  - Synonymy: in synset
  - Hypo(per)nym: Isa tree
- Heavily used resource

Word-internal Structure

- Thematic roles:
  - Characterize verbs by their arguments
    - E.g. transport: agent, theme, source, destination
      - They transported grain from the fields to the silo.
    - Deep structure: passive / active: same roles
- Thematic hierarchy
  - E.g. agent > theme > source, dest
    - Provide default surface positions
  - Tie to semantics (e.g. Levin): Interlinguas
    - Cluster verb meanings by set of syntactic alternations
    - Limitations: only NP,PP: other arguments predicates less

Selectional Restrictions

- Semantic constraints on filling of roles
  - E.g. Bill ate chicken
    - Eat: Agent: animate; Theme: Edible
  - Associate with sense
    - Most commonly of verb/event; possibly adj, noun…
- Specifying constraints:
  - Add a term to semantics, e.g. Isa(x,Ediblething)
  - Tie to position in WordNet
    - All hyponyms inherit

Primitive Decompositions

– Jackendoff(1990), Dorr(1999), McCawley (1968)

- Word meaning constructed from primitives
  - Fixed small set of basic primitives
    - E.g. cause, go, become,
    - kill=cause X to become Y
  - Augment with open-ended “manner”
    - Y = not alive
    - E.g. walk vs run
- Fixed primitives/Infinite descriptors