Natural Language Processing
(overview)

Emily M. Bender
ebender@u.washington.edu
CSE 415
February 13, 2006
Lecture outline

- What is NLP?
- Applications & Approaches
- Resources used in NLP
- NLP subtasks; Ambiguity
- Evaluation in NLP
- Precision grammar engineering
- NLP around UW
What is NLP?

- Processing language by computers
- Distinct from speech processing
- Not necessarily linguistically motivated
Applications (1/3)

- Linguistic research
- Grammar checking/spell checking
- Computer assisted language learning (CALL)
- Assistive & augmentative communication (AAC)
Lecture outline

- What is NLP?
- Applications & Approaches
- Resources used in NLP
- NLP subtasks; Ambiguity
- Evaluation in NLP
- Precision grammar engineering
- NLP around UW
Applications (2/3)

- Machine translation, machine assisted translation
- Information retrieval
- Information extraction
  -- Monolingual & multilingual
Applications (3/3)

- **HCI**
  - Natural language database access
  - UI navigation
  - Automated customer service
  - Games
- Other?
Approaches

- Knowledge engineering
- Machine learning
- Hybrid
Lecture outline

- What is NLP?
- Applications & Approaches
- Resources used in NLP
- NLP subtasks; Ambiguity
- Evaluation in NLP
- Precision grammar engineering
- NLP around UW
Resources

- Dictionaries (monolingual, bilingual)
- Corpora
- Annotated corpora
  - Tagged corpora (POS, word sense, ...)
- Treebanks
- Aligned bilingual/multilingual corpora
Useful for...

- Supervised learning
- Gold standard/evaluation
- Unsupervised/semi-supervised learning of the next layer of linguistic structure
- Linguistic hypothesis testing
Sources of Resources

- LDC: Linguistic Data Consortium
- ELDA: Evaluations and Language resources Distribution Agency
- Rosetta: All Languages Archive
Lecture outline

- What is NLP?
- Applications & Approaches
- Resources used in NLP
- NLP subtasks; Ambiguity
- Evaluation in NLP
- Precision grammar engineering
- NLP around UW
NLP subtasks (1/3)

- Language identification
- Part of speech tagging
- Word sense disambiguation
- Named entity recognition
- NP/other phrase detection
NLP subtasks (2/3)

- Stemming/morphological analysis
- Segmentation (documents to sentences, sentences to words)
- Sentence, phrase, word alignment (of bitext)
NLP subtasks (3/3)

- Parsing (string to tree; string to semantics)
- Generation (semantics to string)
- Reference resolution
- Speech act recognition
- Dialogue planning
- Others?
Ambiguity

- Natural language wasn’t designed to be processed by computers.
- Ambiguity (local and global) at every level of structure
- Potentially want to return multiple analyses
- ... while also being able to rank them
Ambiguity examples

- **Word boundary:**
  
  *Dungeon of Spit*

- **Part of speech:**
  
  *read, record, talk*

- **Morphological analysis:**
  
  *kayaking, singing, sing, anything, walks, unwrappable*
More ambiguity examples

- **Syntax:**
  - *Kim is our local unicode expert.*
  - *Have that report on my desk by Friday.*

- **Semantics:**
  - *Every cat chased some dog.*

- **Speech act:**
  - *Can you pass the salt?*
Still more ambiguity examples

- **Reference resolution:**
  
  The police denied the protesters a permit because they feared/advocated violence.

- **String realization:**
  
  Kim gave the dog a bone.
  Kim gave a bone to the dog.

- **Addressee recognition**
Lecture outline

- What is NLP?
- Applications & Approaches
- Resources used in NLP
- NLP subtasks; Ambiguity
- Evaluation in NLP
- Precision grammar engineering
- NLP around UW
**Evaluation**

- **Requires:**
  - *Test set with gold standard answers for comparison*
  - *Metric(s) of comparison*
  - *Baseline strategy to compare against*
  - *All three of these can be non-obvious in NLP*
Evaluation

- **Validation**: Does my system behave the way I think it behaves?
- **Regression testing**: What did I break today?
- **Experimental results**: How does my system compare to other systems?
Humans are expensive

- Evaluation processes should be automated wherever possible:
  - Speed
  - Cost
  - Integration into development cycle
Easy case: POS tagging

- **Gold standard**: A corpus with POS tags annotated (by humans)
- **Evaluation metric**: Precision (number of correct tags/total tags)
- **Baseline**: Random assignment of possible tags for each lexical item
- **Wrinkle**: Count performance on unambiguous items?
Harder case: Parsing

- How to create a gold standard?
- Sources of variation:
  - Genuine structural ambiguity (usually, but not always, resolved in context)
  - Different styles of representation/different linguistic theories
Examples

S
NP
Kim
VP
V
saw
NP
D
the
N
students
PP
P
with
NP
D
the
N
telescope
Examples

S
  NP
    Kim
  VP
    VP
      V
        saw
      NP
        D
          the
        N
          students
    PP
      P
        with
      NP
        D
          the
        N
          telescope
Examples

NP
  NOM
    Kim

VP
  V
    saw
  D
    the

NP
  NOM
    students

PP
  P
    with
  D
    the

NP
  NOM
    telescope
Examples

NP
Kim

VP
saw

PP
with

NP
the students

NP
the telescope
Harder case: Parsing

- In practice, the most common gold standard is the Penn Treebank
- 1 million words of hand-parsed Wall Street Journal text + 1 million words of hand-parsed Brown corpus
- More or less internally consistent; not consistent with any particular linguistic theory
Harder case: Parsing

- Evaluation metric?
  - How many sentences got exactly the gold standard tree
  - A more sophisticated solution is PARSEVAL (there are others)
PARSEVAL

- **Labeled precision:**
  \[
  \frac{\text{correct constituents in candidate parse}}{\text{constituents in candidate parse}}
  \]

- **Labeled recall:**
  \[
  \frac{\text{correct constituents in candidate parse}}{\text{constituents in gold standard parse}}
  \]

- **Crossing brackets:**
  \[
  (A (B C)) \lor ((A B) C)
  \]
Harder case: Parsing

- What would be a sensible baseline?
- Randomly choosing among all possible structures assigned by the grammar
- Comparison to other existing systems
Even harder case: MT

- (NB: Human evaluation is particularly expensive in this case.)
- What should be the gold standard?
- Are all things that differ from the gold standard necessarily wrong?
- More so than with parsing?
MT and BLEU


- A good translation will have a distribution of n-grams similar to other good translations
BLEU

- Modified n-gram precision

\[ \frac{\sum_{C \in \text{candidates}} \sum_{n-\text{gram} \in C} \text{Count}_{clip}(n - \text{gram})}{\sum_{C \in \text{candidates}} \sum_{n-\text{gram}' \in C'} \text{Count}_{clip}(n - \text{gram}')} \]

- Geometric mean of n-gram precisions for different N, plus a brevity penalty
Evaluation in NLP summary

- It’s always important
- The nature of the tasks makes it often hard to define a gold standard and evaluation metric
- Gold standards can also be expensive
Lecture outline

- What is NLP?
- Applications & Approaches
- Resources used in NLP
- NLP subtasks; Ambiguity
- Evaluation in NLP
- Precision grammar engineering
- NLP around UW
Natural language syntax & semantics

- Constituent structure
- Mapping of linear string to predicate-argument structure (word order, case, agreement)
- Long distance dependencies
  - What did Kim think Pat said Chris saw?
- Idioms, collocations
Formal/‘Generative’ Grammars

- Characterize a set of strings (phrases and sentences)
- These strings should correspond to those that native speakers find acceptable
- Assign one or more syntactic structures to each string
- Assign one or more semantic structures to each string
Formal/‘Generative’ Grammars

- No complete generative grammar has ever been written for any language
Precision Computational Grammars

- Knowledge engineering of formal grammars, for:
  - Parsing: assigning syntactic structure and semantic representation to strings
  - Generation: assigning surface strings to semantic representations
Hurdles

- **Efficient processing** (Oepen et al 2002)
- **Ambiguity resolution** (Baldridge & Osborn 2003, Toutanova et al 2005, Riezler et al 2002)
- **Domain portability** (Baldwin et al 2005)
- **Lexical acquisition** (Baldwin & Bond 2003, Baldwin 2005)
- **Extragrammatical/ungrammatical input** (Baldwin et al 2005)
- **Scaling to many languages**
The Grammar Matrix: Overview

- Motivation
- HPSG
- Semantic representations
- Cross-linguistic core
- Modules
Matrix: Motivation

- **English Resource Grammar:**
  - 140,000 lines of code (25,000 exclusive of lexicon)
  - ~3000 types
  - 16+ person-years of effort
- Much of that is useful in other languages
- Reduces the cost of developing new grammars
Matrix: Motivation

- Hypothesis testing (monolingual and cross-linguistic)
  - Interdependencies between analyses
  - Adequacy of analyses for naturally occurring text
Matrix: Motivation

- Promote consistent semantic representations
- Reuse downstream technology in NLU applications while changing languages
- Transfer-based (symbolic or stochastic MT)
HPSG

- Head-Driven Phrase Structure Grammar
  (Pollard & Sag 1994)

- Mildly-context sensitive (Joshi et al 1991)

- Typed feature-structures

- Declarative, order-independent, constraint-based formalism
An HPSG consists of

- A collection of feature-structure descriptions for phrase structure rules and lexical entries
- Organized into a type hierarchy, with supertypes encoding appropriate features and constraints inherited by subtypes
- All rules and entries contain both syntactic and semantic information
An HPSG is used

- By a parser to assign structures and semantic representations to strings
- By a generator to assign structures and strings to semantic representations
- Rules, entries, and structures are DAGs, with type name labeling the nodes
- Constraints on rules and entries are combined via unification
Example rule type

head-subj-phrase:
[ binary-headed-phrase &
head-compositional
SUBJ ⟨ ⟩
COMPS 1
HEAD-DTR
SUBJ ⟨ 2 ⟩
COMPS 1
NON-HEAD-DTR 2
]
Example rule type

head-final: [binary-headed-phrase &
HEAD-DTR 1
NON-HEAD-DTR 2
ARGS ⟨ 2, 1 ⟩]

subj-head: head-subj-phrase & head-final
Example parse

[HEAD  verb]
SUBJ  ⟨ ⟩
COMPS ⟨ ⟩

[HEAD  noun]
1  SPR  ⟨ ⟩
COMPS ⟨ ⟩

Kim

[HEAD  verb]
SUBJ  ⟨ 1 ⟩
COMPS ⟨ ⟩
danced
Semantic Representations

- Not going for an interlingua
- Not representing connection to world knowledge
- Not representing lexical semantics (the meaning of life is life’)
- Making explicit the relationships among parts of a sentence
Semantic Representations

- *Kim gave a book to Sandy*

- `give(e,x,y,z), name(x, 'Kim'), book(y), name(z, 'Sandy'), past(e)`
Semantic Representations

- Sandy was given a book by Kim.
- A book was given to Sandy by Kim.
- Kim continues to give books to Sandy.
- This is the book that Kim gave Sandy.
- Which book did Kim give Sandy?
- Which book do people often seem to forget that Pat knew Kim gave to Sandy?
- This book was difficult for Kim to give to Sandy.
Semantic representations

- **Languages may still differ:**
  - **Lexical predicates**
    - Japanese: kore, sore, are
  - **Grammaticized tense/aspect, discourse status**
  - **Ways of saying**
    - make a wish, center divider
Matrix Architecture

- Cross-linguistic core encoding language universals
- Set of mutually-compatible ‘modules’ encoding recurring, but non-universal patterns
- Rapid prototyping of precision grammars
- Ongoing development through Ling 567
Lecture outline

- What is NLP?
- Applications & Approaches
- Resources used in NLP
- NLP subtasks; Ambiguity
- Evaluation in NLP
- Precision grammar engineering
- NLP around UW
NLP around UW

- **Professional MA in Computational Linguistics**  
  [http://www.compling.washington.edu](http://www.compling.washington.edu)

- **Computational Linguistics Lab**  
  [http://depts.washington.edu/uwcl](http://depts.washington.edu/uwcl)

- **Turing Center**  

- **SSLI Lab**  
  [http://ssli.ee.washington.edu](http://ssli.ee.washington.edu)

- **MS/UW Symposium in Computational Linguistics**  

- **iSchool, Med School, ...**