Why is Syntax Important?

• Statistical *bag of words* approaches work well for a number of NLP applications:
  – Information Retrieval (IR)
  – Text Categorization
• Bag of words approaches, however, exhibit asymptotic behavior
• For other applications, bag of word approaches don’t work well at all

Syntax is Important

• To improve the performance, NLP applications need to consider the structure of the text:
  – Bob hit John.
  – John hit Bob.
  – Bob was hit by John.
  – Q: Who hit John?

Some Background

• Two notions of grammar equivalency
  – Weak
  – Strong
• Lexical grammars
Complexity of Natural Languages

• The Chomsky hierarchy of languages
  - Regular languages
  - Context-free languages
  - Context-sensitive languages
  - Recursively enumerable languages

Where do Natural Languages (NLs) fit?
  - It is universally agreed that NLs are not regular
  - Many theorists believe that NLs are not Context-Free
    • Evidence: cross constructs in German have the form $wuy^nxy^mz^n p$
  - NLs are believed to be Mildly Context Sensitive

Mildly Context Sensitive Languages (MCSLs)

• A formalism introduced to deal with NL
• Defined by the following properties:
  - CFL are properly contained in MCSL
  - Languages in MCSL can be parsed in polynomial time
  - MCSG can only capture certain dependencies (e.g. nested dependencies)
  - Languages in MCSL have the linear growth property

Syntactic Theories in Linguistics and Psychology

• Classical linguistics
• New directions in syntax
The Reigning Metaphor

Language as LEGO

Classical Linguistics

- Syntax described by CFG
- Semantic and other considerations expressed in terms of constraints
- Theory of movement and traces
- Innate grammar hypothesis

Classical Syntax in NLP

- Penn Treebank
- Parsers have been used successfully in NLP:
  - Stochastic and symbolic parsers; partial parsers
  - Parsers used in many NLP applications
    - Question Answering, Coreference Resolution, Machine Translation, etc
- Grammars typically hand-crafted

Classical Syntax in NLP Cont’d

- Attempts to learn the structure empirically from data:
  - Stolcke & Omohundro (1994) attempted to induce the structure of probabilistic grammars
  - Clark (2001) - Unsupervised induction of Stochastic Context-Free Grammars

(Slide taken from Shimon Edelman’s presentation)
Classical Linguistics - Problems

- No comprehensive transformational generative grammar for any language
- Imperfections of performance and competence
- Lack of empirical support for traces and movement
- Limited defacto language productivity

Example

You already know what it's hittin for
Ma I got whatever outside and you know what I'm sittin on
50/50 venture with them S dots kickin off
Armadale poppin now, only bring a brotha more
Only thing missin is a Missus
You ain't even gotta do the dishes, got two dishwashers
Got one chef, one maid, all I need is a partner
to play spades with the cards up, ALL TRUST
Who else you gon' run with, the truth is us
Only dudes movin units - Em, Pimp Juice and us
.. it's the Roc in here!
(Jay-Z - Excuse me miss)

New Approaches to Syntax

- Cognitive Grammar
- Construction Grammars
- Tree-Adjoining Grammars

Cognitive Grammar

- Langacker, Taylor
- Linguistic cognition simply is cognition; it is an inextricable phenomenon of overall human cognition.
- Therefore, patterns of cognition observed by psychologists, neurologists, and the like should be observed in language.
Cognitive Grammar Cont’d

• “Grammar is not a distinct level of linguistic representation, but reduces instead to the structuring and symbolization of conceptual content...”

• “Lexicon, morphology, and syntax form a continuum of symbolic units, divided only arbitrarily into separate "components"---it is ultimately as pointless to analyze grammatical units without reference to their semantic value as it is to write a dictionary which omits the meanings of its lexical units. “

Construction Grammar

• Constraint based system; syntactic and semantic information represented within single feature structure (attribute value matrix)

• Jackendoff (2002): “[...] we must explicitly deny that conceptual structures [...] mean anything. Rather, we want to say that they are meaning: they do exactly the things meaning is supposed to do, such as support inference and judgment”

Construction Grammar Cont’d

• Constructions
  – stored pairings of form and function

• Elements vary in complexity and in the degree in which they are specified

<table>
<thead>
<tr>
<th>Morpheme</th>
<th>e.g., antia, -ing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>e.g., freezing, amercia, and</td>
</tr>
<tr>
<td>Complex word</td>
<td>e.g., Thoroughbred, show-in</td>
</tr>
<tr>
<td>Idiom (filled)</td>
<td>e.g., losing great gain</td>
</tr>
<tr>
<td>Idiom (partially filled)</td>
<td>e.g., 'dog someone's memory</td>
</tr>
</tbody>
</table>

Covariational Conditional construction [10]
Form: The Xer the Yer
(e.g., The more you think about it, the less you understand)
Meaning: linked independent and dependent variables; see text.

Ditransitive (double object) construction
Form: Subj [V Obj1 Obj2]
(e.g., He gave her a Coke, He baked her a muffin.)
Meaning: transfer (intended or actual); see text.

Passive
Form: Subj aux VPpp (PP by)
(e.g., The armadillo was hit by a car)
Discourse function: to make undergoer topical and/or actor non-topical.

Tree-Adjoining Grammars (TAGs)

• Joshi, Schabes, Vijay-Shanker, Wier, Rambow...

• Work on TAGs motivated by linguistics considerations

• TAGs, however, of interest in formal languages and automata
  – Mildly context sensitive
  – Can be parsed in polynomial time by embedded push-down automata
  – Derive from the lexicalization of CFG
TAG elements

- Initial tree
- Auxiliary tree

LTAG composition operations

- Adjoining

LTAG composition operations

- Substitution

LTAG linguistic relevance

- Extended domain of locality
- Factoring recursion from the domain of locality
  - Dependencies such as agreement and subcategorization stated over elementary structures of TAGs
  - Long-distance behavior follows from adjoining
Other approaches

- Stochastic Tree Substitution Grammars (STSGs)
  - Bod (1998); Scha, Bod, and Sima'an (1999)
- Split and Merge Pattern Learning
  - Wolff (1988)
- ADIOS
  - Shimon Edelman et al

New Approaches - Connections to NLP

- Information Extraction – Extraction Patterns
  - <victim> was murdered
- Question answering
- A challenge:
  - Develop methods to automatically identify semantic units from (un/weakly) supervised training data
  - Take advantage of the vast amount of unsupervised data (untagged text) available