Knowledge Representation

Models to represent Linguistic Knowledge

- We will use certain *models to represent the required linguistic knowledge.*
- **State Machines** -- FSAs, FSTs, HMMs, ATNs, RTNs
- **Formal Rule Systems** -- Context Free Grammars, Unification Grammars, Probabilistic CFGs.
- Other common representational formalisms:
  – first order predicate logic
  – conceptual dependency graphs
  – semantic networks
  – Frame-based representations

Difficulty in NLP understanding

- arises from:
- Natural language is extremely **rich** in form and structure:
  – How to represent meaning,
  – Which structures map to which meaning structures.
- Natural language is **very ambiguous**. One input can mean many different things. Ambiguity can be at different levels.
  – Phonics Level: different meaning for the same sound
  - Lexical (word level) ambiguity -- different meanings of words
  – Syntactic ambiguity -- different ways to parse the sentence
  – Interpreting partial information -- how to interpret pronouns
  – Contextual information -- context of the sentence may affect the meaning of that sentence.

Speech Recognition - Complications

- No simple mapping between sounds and words
  – Variance in pronunciation due to gender, dialect, …
    - Restriction to handle just one speaker
  – Same sound corresponding to diff. words
    - e.g. bear, bare
  – Finding gaps between words
    - "how to recognize speech"
    - "how to wreck a nice beach"
  – Noise

Syntactic Analysis: Complications

- Rules of syntax (grammar) specify the possible organization of words in sentences and allows us to determine sentence's structure(s)
  – "John saw Mary with a telescope"
    - John saw (Mary with a telescope)
    - John (saw Mary with a telescope)
  – "fruit flies like a banana"
- Parsing: given a sentence and a grammar
  – Checks that the sentence is correct according with the grammar and if so returns a **parse tree** representing the structure of the sentence

Semantic Analysis – Complications Ambiguous Example

- Some interpretations of : I made her duck

Semantic Analysis – Complications Ambiguous Example

- Some interpretations of : I made her duck
  1. I cooked *duck* for her.
  2. I cooked *duck* belonging to her.
  3. I created a toy duck which she owns.
  4. I caused her to quickly lower her head or body.
  5. I used magic and turned her into a *duck*.
- duck – morphologically and syntactically ambiguous: noun or verb.
- her – syntactically ambiguous: for her/ to her/ her
- make – semantically ambiguous: cook or create.
- make – syntactically ambiguous:
- Transitive – takes a direct object. => 2
- Di-transitive – takes two objects. => 5
- Takes a direct object and a verb. => 4

Grammar types

- Regular: nonterminal ----> terminal[nonterminal]

  S ----> aS
- Context-free: nonterminal ----> anything

  S ----> aSb
- Context-sensitive: more nonterminals on right-hand side

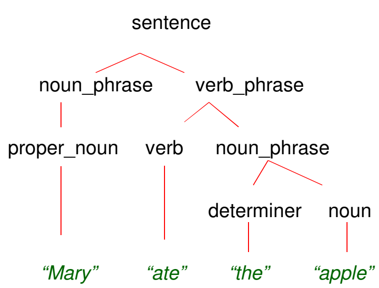
  ASB ----> AAaBB
- Recursively enumerable: no constraints

  Natural languages probably is considered (dealt with as) Context sensitive

Syntactic Analysis - Grammar

- **sentence -> noun_phrase, verb_phrase**
- **noun_phrase -> proper_noun**
- **noun_phrase -> determiner, noun**
- **verb_phrase -> verb, noun_phrase**
- **proper_noun -> [mary]**
- **noun -> [apple]**
- **verb -> [ate]**
- **determiner -> [the]**

Syntactic Analysis - Parsing



Syntactic Analysis : Complications

- Number (singular vs. plural) and gender
  - **sentence-> Noun_phrase(n),verb_phrase(n)**
  - **proper_noun(s) -> [mary]**
  - **noun(p) -> [apples]**
- Adjective
  - **noun_phrase-> determiner,adjectives,noun**
  - adjectives-> adjective, adjectives
  - adjective->[ferocious]
- Adverbs, …
- Having to parse syntactically incorrect sentences

Context-free grammars

We will look at the simplest Context-Free Grammars,

```
sentence     → noun_phrase  verb_phrase

noun_phrase  → proper_name
noun_phrase  → article  noun

verb_phrase  → verb
verb_phrase  → verb  noun_phrase
verb_phrase  → verb  noun_phrase
                     prep_phrase
verb_phrase  → verb  prep_phrase

prep_phrase  → preposition  noun_phrase
```

Context-free grammars

The still-undefined syntactic units are *preterminals*. They correspond to parts of speech. We can define them by adding lexical productions to the grammar:

```
article → the  |  a  |  an
noun → pizza  |  bus  |  boys  |  ...
preposition → to  |  on  |  ...
proper_name → Jim  |  Dan  |  ...
verb → ate  |  yawns  |  ...
```

This is not practical on a large scale. Normally, we have a lexicon (dictionary) stored in a database, that can be interfaced with the grammar.

Context-free grammars

```
sentence →
noun_phrase verb_phrase →
proper_name verb_phrase →
Jim verb_phrase →
Jim verb noun_phrase prep_phrase →
Jim ate noun_phrase prep_phrase →
Jim ate article noun prep_phrase →
Jim ate a noun prep_phrase →
Jim ate a pizza prep_phrase →
Jim ate a pizza preposition noun_phrase →
Jim ate a pizza on noun_phrase →
Jim ate a pizza on article noun →
Jim ate a pizza on the noun →
Jim ate a pizza on the bus
```

Context-free grammars

```
sentence     → noun_phr_sg  verb_phr_sg
sentence     → noun_phr_pl  verb_phr_pl
noun_phr_sg  → art_sg  noun_sg
noun_phr_sg  → proper_name_sg
noun_phr_pl  → art_pl  noun_pl
noun_phr_pl  → proper_name_pl
art_sg       → the  |  a  |  an
art_pl       → the
noun_sg      → pizza  |  bus  |  ...
noun_pl      → boys  |  ...
```

Context-free grammars

Other examples of sentences generated by this grammar:

```
Jim ate a pizza
Dan yawns on the bus
```

These *wrong* data will also be recognized:

```
Jim ate an pizza
Jim yawns a pizza
Jim ate to the bus
the boys yawns
the bus yawns
```

Direction of parsing

In practice, parsing is never "pure".

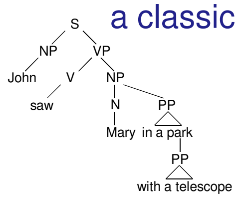
➢Top-down, enriched: check data early to discard wrong hypotheses (somewhat like recursive-descent parsing in compiler construction).

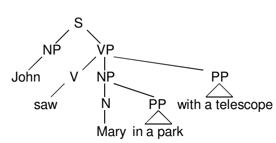
➢Bottom-up, enriched: use productions, suggested by data, to limit choices

➢A popular bottom-up analysis method: chart parsing.

➢Popular top-down analysis methods:
  transition networks (used with Lisp),
  logic grammars (used with Prolog).

Syntactic ambiguity
a classic example





Semantic Analysis

- Generates (partial) meaning/representation of the sentence from its syntactic structure(s)
- Compositional semantics: meaning of the sentence from the meaning of its parts:
  - Sentence: A tall man married Mary
  - Representation: ∃x man(x) & tall(x) & married(x,mary)
- Grammar + Semantics
  - `Sentence (Smeaning)-> noun_phrase(NPmeaning),verb_phrase(VPmeaning), combine(NPmeaning,VPmeaning,Smeaning)`

15

Understanding Natural Language

From Luger

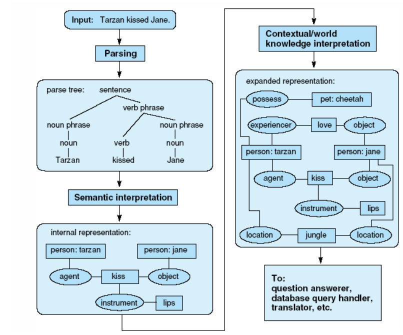
15.1 Deconstructing Language: A Symbolic Analysis

15.2 Syntax

15.3 Syntax and Knowledge with ATN Parsers

Stages of Language Analysis

Fig 15.2 Stages in producing an internal representation of a sentence.



Syntax Example.

Consider the following context free grammar

1. Sentence→ Noun_phrase Verb_phrase
2. Noun_phrase → Noun
3. Noun_phrase → Article Noun
4. Verb_phrase →Verb
5. Verb_phrase →Verb Noun_phrase
6. Article → a
7. Article → the
8. Noun → man
9. Noun → dog
10. Verb → likes
11. Verb → bites

The derivation of a sentence

A Top-Down derivation:

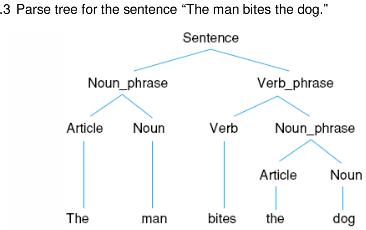
String	Rule
Sentence	1
Noun_phrase Verb_phrase	3
Article Noun Verb_phrase	7
the Noun Verb_phrase	8
the man Verb_phrase	5
the man Verb Noun_phrase	11
the man bites Noun_phrase	3
the man bites Article Noun	7
the man bites the Noun	9
the man bites the dog	

1. Sentence→ Noun_phrase Verb_phrase
2. Noun_phrase → Noun
3. Noun_phrase → Article Noun
4. Verb_phrase →Verb
5. Verb_phrase →Verb Noun_phrase
6. Article → a
7. Article → the
8. Noun → man
9. Noun → dog
10. Verb → likes
11. Verb → bites

Do it Bottom-Up

The parse Tree

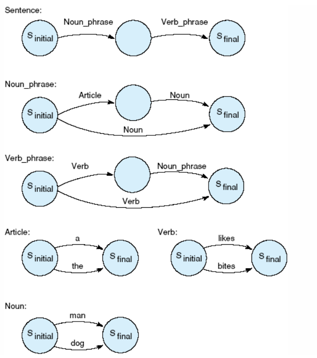
Fig 15.3 Parse tree for the sentence "The man bites the dog."



The derivation of a sentence

- One difficulty, that can add huge complexity to the parsing problem is: Determining which of several applicable rules should be used at any step of the derivation.
- If a wrong choice is made, the parser may fail to recognize a legal sentence.
- The problem of selecting the correct rule at any stage of the parse is handled either by
  →allowing the parser to set backtrack pointers and return if an incorrect choice was made
  → *using* look-ahead to check the input string for features that will help determine the proper rule to apply.

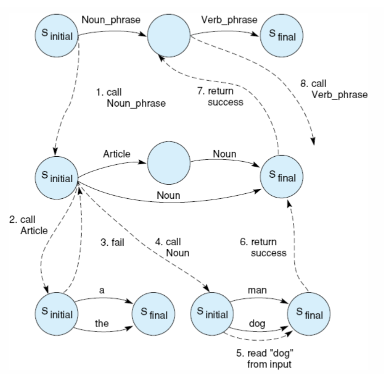
Fig 15.5 Transition network definition of a simple English grammar.



Parsing a sentence

1. The parser begins with the sentence network and tries to move along the arc labeled noun_phrase. To do so, it retrieves the network for noun_phrase.
2. In the noun-phrase network, the parser first tries the transition marked article. This causes it to branch to the network for article.
3. It fails to find a path to the finish node of the article network because the first word of the sentence, "Dog," matches neither of the arc labels. The parser fails and backtracks to the nounphrass network.
4. The parser attempts to follow the arc labeled noun in the noun_phrase network and branches to the network for noun.
5. The parser successfully crosses the arc labeled "dog," because this corresponds to the first word of the input stream.
6. The noun network returns success. This allows the arc labeled noun in the noun_phrase network to be crossed to the final state.
7. The noun_phrase network returns success to the top-level network, allowing the transition of the arc labeled noun-phrase.
8. A sequence of similar steps is followed to parse the verbphrase portion of the sentence.

Fig 15.6 Trace of a transition network parse of the sentence “Dog bites.”



Parsing a sentence

Adding these rules to the above grammar
Noun → men
Noun → dogs
Verb → bites
Verb → like

Produce sentences like
**) The dogs like the men*
**) A men bites a dogs !!!*

Chomsky in 1965 proposed the hierarchy of grammars.
A context free grammar doesnot enforce the number…

To get rid of phrases like above, we use context sensitive grammar

Context Sensitive Grammar

sentence → noun_phrase verbphrase
noun_phrase → article number noun
nounphrase → number noun
number → singular
number → plural
article singular → a singular
article singular → the singular
article plural → some plural
article plural → the plural
singular noun → dog singular
singular noun → man singular
plural noun → men plural
plural noun → dogs plural
singular verb_phrase → singular verb
plural verb_phrase → plural verb
singular verb → bites
singular verb → likes
plural verb → bite
plural verb → like

In this grammar, the nonterminals singular and plural offer constraints to determine when different article, noun, and verb-.-phrase rules can be applied, ensuring number agreement

Context Sensitive Grammar

Similarly, we can use context-sensitive grammars to perform checks for semantic agreement. For example, we could disallow sentences such as "Man bites dog" by adding
a nonterminal, act_of_biting, to the grammar. This could be checked in the rules to prevent any sentence involving "bites" from having "man" as its subject.
Though context-sensitive grammars can define language structures that cannot be captured using context-free grammars, they have a number of disadvantages

Context Sensitive Grammar

1. Context-sensitive grammars increase drastically the number of rules and nonterminals in the grammar.
2. They obscure the phrase structure of the language that is so clearly represented in the context-free rules.
3. By attempting to handle more complicated checks for agreement and semantic consistency in the grammar itself, they lose many of the benefits of separating the
syntactic and semantic components of language.
4. Context-sensitive grammars do not address the problem of building a semantic representation of the meaning of the text. A parser that simply accepts or rejects sentences is not sufficient

NLP Topics

- **Phonology** – concerns how words are related to the sounds that realize them.
- **Morphology** – concerns how words are constructed from more basic meaning
- **Syntax** – concerns how can be put together to form correct sentences – Grammar
- **Semantics** – concerns what words mean and how these meaning combine in sentences to form sentence meaning.
- **Pragmatics** – concerns how sentences are used in different situations
- **Discourse** – concerns how the immediately preceding sentences affect the interpretation of the next sentence. “context”
- **World Knowledge** – includes general knowledge about the world.

Example of knowledge base:

John went to the diner to eat lunch. He ordered a hamburger. But John wasn't very hungry so he didn't finish it. John told the waiter that he wanted a doggy bag. John gave the waiter a tip. John then went to the hardware store and home.