Natural Language for Communication

Chapter 23, Sec. 1-3, 3rd ed.
Chapter 22, Sec. 1-5, 2nd ed.
Outline

• Communication
• Grammar
• Syntactic and semantic analysis
• Problems
A Distinction

Distinguish:

- **Natural language processing**
  - Concerned with text classification, information extraction, etc.
  - E.g. spam detection, language translation (like Google translate).
  - Emphasis on statistical techniques

- **Natural language understanding (this part)**
  - Concerned with communication, deep understanding of language
  - E.g. verbal communication, story understanding
  - Emphasis on analysis and understanding
Communication

“Classical” view (pre-1953):
- Language consists of sentences that are true/false (cf. logic)

“Modern” view (post-1953):
- Language is a form of action

E.g.:
- Wittgenstein (1953) *Philosophical Investigations*
- Austin (1962) *How to Do Things with Words*
- Searle (1969) *Speech Acts*

Why?
“Classical” view (pre-1953):

- Language consists of sentences that are true/false (cf. logic)

“Modern” view (post-1953):

- Language is a form of action

E.g.:

- Wittgenstein (1953) *Philosophical Investigations*
- Austin (1962) *How to Do Things with Words*
- Searle (1969) *Speech Acts*

Why?

- *To change the actions of other agents*
Communication as action: Speech acts

Speech acts achieve the speaker’s goals:
- **Inform**: “There’s a pit in front of you”
- **Query**: “Can you see the gold?”
- **Command**: “Pick it up”
- **Promise**: “I’ll share the gold with you”
- **Acknowledge**: “OK”

Speech act planning requires knowledge of:
- Situation
- Semantic and syntactic conventions
- Hearer’s goals, knowledge base, and rationality
Stages in communication (informing)

**Intention**
S wants to inform H that \( P \)

**Generation**
S selects words \( W \) to express \( P \) in context \( C \)

**Synthesis**
S utters words \( W \)

**Perception**
H perceives \( W' \) in context \( C' \)

**Analysis**
H infers possible meanings \( P_1, \ldots, P_n \)

**Disambiguation**
H infers intended meaning \( P_i \)

**Incorporation**
H incorporates \( P_i \) into KB

How could this go wrong?

- Insincerity (S doesn’t believe \( P \))
- Speech wreck ignition failure
- I.e. It’s hard to wreck a nice beach!
- Ambiguous utterance
- Differing understanding of current context (\( C \neq C' \))
Stages in communication (informing)

*Intention*  
S wants to inform H that \( P \)

*Generation*  
S selects words \( W \) to express \( P \) in context \( C \)

*Synthesis*  
S utters words \( W \)

*Perception*  
H perceives \( W' \) in context \( C' \)

*Analysis*  
H infers possible meanings \( P_1, \ldots, P_n \)

*Disambiguation*  
H infers intended meaning \( P_i \)

*Incorporation*  
H incorporates \( P_i \) into KB

How could this go wrong?

• Insincerity (S doesn’t believe \( P \))
• Speech wreck ignition failure
  • I.e. It’s hard to wreck a nice beach!
• Ambiguous utterance
• Differing understanding of current context (\( C \neq C' \))
Stages in communication (informing)

*Intention*  
S wants to inform H that \( P \)

*Generation*  
S selects words \( W \) to express \( P \) in context \( C \)

*Synthesis*  
S utters words \( W \)

*Perception*  
H perceives \( W' \) in context \( C' \)

*Analysis*  
H infers possible meanings \( P_1, \ldots, P_n \)

*Disambiguation*  
H infers intended meaning \( P_i \)

*Incorporation*  
H incorporates \( P_i \) into KB

How could this go wrong?

- Insincerity (S doesn’t believe \( P \))
Stages in communication (informing)

**Intention**  
S wants to inform H that $P$

**Generation**  
S selects words $W$ to express $P$ in context $C$

**Synthesis**  
S utters words $W$

**Perception**  
H perceives $W'$ in context $C'$

**Analysis**  
H infers possible meanings $P_1, \ldots, P_n$

**Disambiguation**  
H infers intended meaning $P_i$

**Incorporation**  
H incorporates $P_i$ into KB

How could this go wrong?

- Insincerity (S doesn’t believe $P$)
- Speech wreck ignition failure
Stages in communication (informing)

*Intention*  
S wants to inform H that $P$

*Generation*  
S selects words $W$ to express $P$ in context $C$

*Synthesis*  
S utters words $W$

*Perception*  
H perceives $W'$ in context $C'$

*Analysis*  
H infers possible meanings $P_1, \ldots, P_n$

*Disambiguation*  
H infers intended meaning $P_i$

*Incorporation*  
H incorporates $P_i$ into KB

How could this go wrong?

- Insincerity ($S$ doesn’t believe $P$)
- Speech wreck ignition failure
  - I.e. It’s hard to wreck a nice beach!
### Stages in communication (informing)

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intention</strong></td>
<td>S wants to inform H that $P$</td>
</tr>
<tr>
<td><strong>Generation</strong></td>
<td>S selects words $W$ to express $P$ in context $C$</td>
</tr>
<tr>
<td><strong>Synthesis</strong></td>
<td>S utters words $W$</td>
</tr>
<tr>
<td><strong>Perception</strong></td>
<td>H perceives $W'$ in context $C'$</td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td>H infers possible meanings $P_1, \ldots, P_n$</td>
</tr>
<tr>
<td><strong>Disambiguation</strong></td>
<td>H infers intended meaning $P_i$</td>
</tr>
<tr>
<td><strong>Incorporation</strong></td>
<td>H incorporates $P_i$ into KB</td>
</tr>
</tbody>
</table>

How could this go wrong?

- Insincerity ($S$ doesn’t believe $P$)
- Speech wreck ignition failure
  - I.e. It’s hard to wreck a nice beach!
- Ambiguous utterance
Stages in communication (informing)

*Intention*  
S wants to inform H that $P$

*Generation*  
S selects words $W$ to express $P$ in context $C$

*Synthesis*  
S utters words $W$

*Perception*  
H perceives $W'$ in context $C'$

*Analysis*  
H infers possible meanings $P_1, \ldots, P_n$

*Disambiguation*  
H infers intended meaning $P_i$

*Incorporation*  
H incorporates $P_i$ into KB

How could this go wrong?

- Insincerity ($S$ doesn’t believe $P$)
- Speech wreck ignition failure
  - I.e. It’s hard to wreck a nice beach!
- Ambiguous utterance
- Differing understanding of current context ($C \neq C'$)
Vervet monkeys, antelopes etc. use isolated symbols

- restricted set of communicable propositions
- no *generative capacity*

Humans are unique among animals in many ways, with the use of *language* being particularly notable.

Consider the role of language in the Turing test or the Winograd Challenge.
Fundamentals of language: Grammar

- **Grammar** specifies the compositional structure of complex messages (Chomsky (1957): *Syntactic Structures*)
Fundamentals of language: Grammar

• *Grammar* specifies the compositional structure of complex messages (Chomsky (1957): *Syntactic Structures*)

• A *formal language* is a set of *strings* of *terminal symbols*

• Each string in the language can be analyzed/generated by the grammar
Fundamentals of language: Grammar

- **Grammar** specifies the compositional structure of complex messages (Chomsky (1957): *Syntactic Structures*)
- A **formal language** is a set of **strings** of **terminal symbols**
- Each string in the language can be analyzed/generated by the grammar
- The grammar is a set of **rules** that specifies a language
- Most grammar formalisms are based on the idea of a **phrase structure**
  - I.e. strings are composed of substrings called **phrases**.
  - E.g. noun phrases, verb phrases
Fundamentals of language: Grammar

- **Grammar** specifies the compositional structure of complex messages (Chomsky (1957): *Syntactic Structures*)
- A *formal language* is a set of *strings* of *terminal symbols*
- Each string in the language can be analyzed/generated by the grammar
- The grammar is a set of *rules* that specifies a language
- Most grammar formalisms are based on the idea of a *phrase structure*
  - I.e. strings are composed of substrings called *phrases*.
  - E.g. noun phrases, verb phrases
- Category names (like *NP*, *VP*) are called *nonterminals*; the words in a language are the *terminal* symbols.
  - $S \rightarrow NP \ VP$
  - $Article \rightarrow \textbf{the} \mid \textbf{a} \mid \textbf{an} \mid \ldots$
Fundamentals of language

- Both formal and natural languages associate a meaning or *semantics* to each valid string.

- For natural language the *pragmatics* of a string is also important.
Fundamentals of language

- Both formal and natural languages associate a meaning or *semantics* to each valid string.
- For natural language the *pragmatics* of a string is also important.
- “Standard” NLU is a 2-step process:
  - Given some grammar, *parse* a sentence into constituent parts to give a *parse tree*.
  - Given a parse tree, do a *semantic analysis* to determine the meaning.
Grammar types

**Regular:** nonterminal $\rightarrow$ **terminal**[nonterminal]

$$S \rightarrow aS$$
$$S \rightarrow \Lambda$$

**Context-free:** nonterminal $\rightarrow$ anything

$$S \rightarrow aSb$$

**Context-sensitive:** more nonterminals on right-hand side

$$ASB \rightarrow AAdBB$$

**Recursively enumerable:** no constraints; equivalent to Turing machines wrt expressiveness

Natural languages are “close to” context-free, parsable in real time!
Consider defining a fragment of English for agents to communicate in the wumpus world.

Need to define:

- the *lexicon*, or set of allowable words, and
- the *grammar*, which will combine words into phrases, and ultimately into *sentences*
# Wumpus lexicon

<table>
<thead>
<tr>
<th>Category</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>stench, breeze, glitter, nothing, wumpus, pit, pits, gold, east, ...</td>
</tr>
<tr>
<td>Verb</td>
<td>is, see, smell, shoot, feel, stinks, go, grab, carry, kill, turn, ...</td>
</tr>
<tr>
<td>Adjective</td>
<td>right, left, east, south, back, smelly, ...</td>
</tr>
<tr>
<td>Adverb</td>
<td>here, there, nearby, ahead, right, left, east, south, back, ...</td>
</tr>
<tr>
<td>Pronoun</td>
<td>me, you, I, it, ...</td>
</tr>
<tr>
<td>Name</td>
<td>John, Mary, Boston, SFU, ...</td>
</tr>
<tr>
<td>Article</td>
<td>the, a, an, ...</td>
</tr>
<tr>
<td>Prep</td>
<td>to, in, on, near, ...</td>
</tr>
<tr>
<td>Conjunction</td>
<td>and, or, but, ...</td>
</tr>
<tr>
<td>Digit</td>
<td>0, 1, 2, 3, 4, 5, 6, 7, 8, 9</td>
</tr>
</tbody>
</table>
Wumpus grammar

\[
S \rightarrow NP \ VP \quad \text{I + feel a breeze}
\]
\[
\quad \mid S \ Conj \ S \quad \text{I feel a breeze + and + I smell a wumpus}
\]
\[
NP \rightarrow \text{Pronoun} \quad \text{I}
\]
\[
\quad \mid \text{Noun} \quad \text{pits}
\]
\[
\quad \mid \text{Article Noun} \quad \text{the + wumpus}
\]
\[
\quad \mid \text{Digit Digit} \quad 3 \ 4
\]
\[
\quad \mid NP \ PP \quad \text{the wumpus + to the east}
\]
\[
\quad \mid NP \ RelCl \quad \text{the wumpus + that is smelly}
\]
\[
VP \rightarrow \text{Verb} \quad \text{stinks}
\]
\[
\quad \mid VP \ NP \quad \text{feel + a breeze}
\]
\[
\quad \mid VP \ Adjective \quad \text{is + smelly}
\]
\[
\quad \mid VP \ PP \quad \text{turn + to the east}
\]
\[
\quad \mid VP \ Adverb \quad \text{go + ahead}
\]
\[
PP \rightarrow \text{Prep} \ NP \quad \text{to + the east}
\]
\[
RelCl \rightarrow \text{that} \ VP \quad \text{that + is smelly}
\]
Examples

The grammar generates good English sentences, such as:

- John is in a pit
- The wumpus that stinks is in 2 2
- Mary is in Boston and John stinks
Examples

The grammar generates good English sentences, such as:

- John is in a pit
- The wumpus that stinks is in 2 2
- Mary is in Boston and John stinks

However it overgenerates, e.g.:

- Me go SFU
- John smell
- I smell pit gold wumpus east ahead
Syntactic analysis: Parsing

**Parsing** is the process of

- analysing a string of words to
- uncover its *phrase structure*,
- according to the rules of the *grammar*

Can generate one or more *parse trees* (assuming a grammatical sentence).

(Aside: A* search has been used in parsing algorithms to find the “best” parse and sometimes probabilities are attached to rules to give a “preferred” parse.)
Parse trees

Exhibit the grammatical structure of a sentence

I shoot the wumpus
Parse trees

Exhibit the grammatical structure of a sentence

I shoot the wumpus

Pronoun  Verb  Article  Noun

I  shoot  the  wumpus
Parse trees

Exhibit the grammatical structure of a sentence

I shoot the wumpus

Pronoun Verb Article Noun

NP VP NP
Parse trees

Exhibit the grammatical structure of a sentence

I shoot the wumpus

NP | VP | NP
---|----|---
Pronoun | Verb | Article Noun
I | shoot | the wumpus
Parse trees

Exhibit the grammatical structure of a sentence

I shoot the wumpus

NP VP NP

VP

S

I

Pronoun

the

Article

shoot

Verb

wumpus

Noun
Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” $\neq$ “John hit Mary”
Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” $\neq$ “John hit Mary”

Of course, things can get complicated:

“And since I was not informed—as a matter of fact, since I did not know that there were excess funds until we, ourselves, in that checkup after the whole thing blew up, and that was, if you’ll remember, that was the incident in which the attorney general came to me and told me that he had seen a memo that indicated that there were no more funds.”
Context-free parsing

- Bottom-up parsing works by replacing any substring that matches the RHS of a rule with the rule’s LHS
- Efficient algorithms $O(n^3)$ for context-free
  - run at several thousand words/sec for real grammars
- Context-free parsing $\equiv$ Boolean matrix multiplication

⚠️ Unlikely to find faster practical algorithms
Augmented Grammars

- Unfortunately, the assumption that a language is context free makes notions like
  - *case* (e.g. subjective, objective case) and
  - *number* agreement
  awkward.
  
  Also, difficult to connect syntax to semantics
Augmented Grammars

- Unfortunately, the assumption that a language is context free makes notions like case (e.g. subjective, objective case) and number agreement awkward.
- Also, difficult to connect syntax to semantics
- One (poor) solution: Add more rules
Augmented Grammars

• Unfortunately, the assumption that a language is context free makes notions like
  • case (e.g. subjective, objective case) and
  • number agreement
awkward.

  Also, difficult to connect syntax to semantics

• One (poor) solution: Add more rules

• Solution: Augment the grammar by allowing parameters on nonterminal symbols.

• This results in a formalism called a definite clause grammar (or DCG).
Logical grammars: DCGs

Idea: express grammar rules as logic

\[ X \rightarrow YZ \quad \text{becomes} \quad Y(s_1) \land Z(s_2) \Rightarrow X(Append(s_1, s_2)) \]
\[ X \rightarrow \text{word} \quad \text{becomes} \quad X([\text{"word"}]) \]
\[ X \rightarrow Y \mid Z \quad \text{becomes} \quad Y(s) \Rightarrow X(s) \quad Z(s) \Rightarrow X(s) \]

- Rules are “turned around” to make logical implications, or definite clauses
- Can then add arguments to enforce case, number, etc.
- Here, \( X(s) \) means that string \( s \) can be interpreted as an \( X \)

DCGs allow us to talk about parsing as logical inference.
Logical grammars

- Big benefit: Can *augment* the nonterminal symbols with additional parameters.
- For example:

\[
NP(s_1) \land Number(s_1, n) \land VP(s_2) \land Number(s_2, n) \\
\Rightarrow S(Append(s_1, s_2))
\]
Semantic Interpretation

To add semantics to a grammar, need a “target” representational formalism, such as FOL

- E.g. go from “John loves Mary” to $Loves(John, Mary)$. 

Lambda calculus ≈ formalism for defining and applying functions. (Aside: Lisp and Scheme are based on the lambda calculus)

Idea: For “John loves Mary”:
- The NP “John” should have as semantic interpretation the logical term $John$
- The sentence as a whole should have interpretation $Loves(John, Mary)$.
- Tricky part: the VP “loves Mary”. Intuitively, “loves Mary” can be regarded as a 1-place predicate, true of those who love Mary.
Semantic Interpretation

To add semantics to a grammar, need a “target” representational formalism, such as FOL

- E.g. go from “John loves Mary” to $Loves(John, Mary)$.
- Can do this by annotating a parse tree, and using it to build a logical formula via the *lambda calculus*.
- Lambda calculus $\approx$ formalism for defining and applying functions.
  (Aside: Lisp and Scheme are based on the lambda calculus)
Semantic Interpretation

To add semantics to a grammar, need a “target” representational formalism, such as FOL

- E.g. go from “John loves Mary” to Loves(John, Mary).
- Can do this by annotating a parse tree, and using it to build a logical formula via the lambda calculus.
- Lambda calculus $\approx$ formalism for defining and applying functions.
  (Aside: Lisp and Scheme are based on the lambda calculus)
- Idea: For “John loves Mary”:
  - The NP “John” should have as semantic interpretation the logical term John
  - The sentence as a whole should have interpretation Loves(John, Mary).
  - Tricky part: the VP “loves Mary”. Intuitively, “loves Mary” can be regarded as a 1-place predicate, true of those who love Mary.
Semantic Interpretation

Roughly, “John loves Mary” is parsed and interpreted as follows:

- “John” denotes the term *John*, which is a NP
Semantic Interpretation

Roughly, “John loves Mary” is parsed and interpreted as follows:

- “John” denotes the term *John*, which is a NP
- “Loves” denotes a 2-place predicate *Loves*, which is a Verb
Semantic Interpretation

Roughly, “John loves Mary” is parsed and interpreted as follows:

- “John” denotes the term *John*, which is a NP
- “Loves” denotes a 2-place predicate *Loves*, which is a Verb
- “Mary” denotes the term *Mary*, which is a NP
Semantic Interpretation

Roughly, “John loves Mary” is parsed and interpreted as follows:

- “John” denotes the term *John*, which is a NP
- “Loves” denotes a 2-place predicate *Loves*, which is a Verb
- “Mary” denotes the term *Mary*, which is a NP
- “Loves” and Mary combine to give a VP with interpretation the 2-place predicate *Loves*(x, *Mary*)
Semantic Interpretation

Roughly, “John loves Mary” is parsed and interpreted as follows:

- “John” denotes the term *John*, which is a NP
- “Loves” denotes a 2-place predicate *Loves*, which is a Verb
- “Mary” denotes the term *Mary*, which is a NP
- “Loves” and Mary combine to give a VP with interpretation the 2-place predicate *Loves*(x, *Mary*)
- The VP combines with the NP *John* to give sentence with interpretation *Loves*(John, Mary).
Real language

Real human languages provide many problems for NLP:

- ambiguity
- anaphora
- indexicality
- vagueness
- discourse structure
- metonymy
- metaphor
- noncompositionality
Ambiguity

Bystander helps dog bite victim
Ambiguity

Bystander helps dog bite victim
Helicopter powered by human flies
Ambiguity

Bystander helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans
Ambiguity

Bystander helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans
I ate spaghetti with meatballs
Ambiguity

Bystander helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans
I ate spaghetti with meatballs
   salad
Ambiguity

Bystander helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans
I ate spaghetti with meatballs
   salad
   enjoyment
Ambiguity

Bystander helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans

I ate spaghetti with meatballs
salad
enjoyment
a fork
Bystander helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans

I ate spaghetti with meatballs
  salad
  enjoyment
  a fork
  a friend
Bystander helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans

I ate spaghetti with meatballs
  salad
  enjoyment
  a fork
  a friend

- Ambiguity can be lexical (polysemy), syntactic, semantic, referential
Anaphora

Using pronouns to refer back to entities already introduced in the text

- After Mary proposed to John, *they* found a preacher and got married.
Anaphora

Using pronouns to refer back to entities already introduced in the text

- After Mary proposed to John, they found a preacher and got married.
- For the honeymoon, they went to Hawaii

(Recall the Winograd Challenge from the first lecture!)
Anaphora

Using pronouns to refer back to entities already introduced in the text

- After Mary proposed to John, they found a preacher and got married.
- For the honeymoon, they went to Hawaii
- Mary saw a ring through the window and asked John for it
Anaphora

Using pronouns to refer back to entities already introduced in the text

- After Mary proposed to John, *they* found a preacher and got married.
- For the honeymoon, *they* went to Hawaii
- Mary saw a ring through the window and asked John for *it*
- Mary threw a rock at the window and broke *it*
Anaphora

Using pronouns to refer back to entities already introduced in the text

- After Mary proposed to John, they found a preacher and got married.
- For the honeymoon, they went to Hawaii
- Mary saw a ring through the window and asked John for it
- Mary threw a rock at the window and broke it

(Recall the Winograd Challenge from the first lecture!)
Indexical sentences refer to utterance situation (place, time, S/H, etc.)

- *I am over here*
- *Why did you do that?*
Metonymy

Using one noun phrase to stand for another

- I’ve read *Shakespeare*
- *Chrysler* announced record profits
- The *ham sandwich* on Table 4 wants another beer
Metaphor

“Non-literal” usage of words and phrases, often systematic:

- I’ve tried killing the process but it won’t die. Its parent keeps it alive.
Noncompositionality

basketball shoes
Noncompositionality

basketball shoes
baby shoes
Noncompositionality

basketball shoes
baby shoes
alligator shoes
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
red herring
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
red herring
small moon
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
red herring
small moon
large molecule
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
red herring
small moon
large molecule
mere child
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
red herring
small moon
large molecule
mere child
alleged murderer
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
red herring
small moon
large molecule
mere child
alleged murderer
real leather
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
red herring
small moon
large molecule
mere child
alleged murderer
real leather
artificial grass
Summary

- NLU is one of the oldest and most important areas of AI.
- Unlike other areas of AI, it requires an investigation of human behaviour and cognition.
- Formal language theory and phrase structure grammars are useful tools for dealing with some aspects of NLU.
- By augmenting a grammar, problems such as subject-verb agreement can be handled.
  - Definite clause grammars allow such augmentations.
- Ambiguity is a very important problem, and relies on world knowledge and the current situation.