

Natural Language

References	2
Natural Language Ambiguity	3
Machine Translation	3
Types of Ambiguities in Natural Language	4
More Types of Ambiguities	5
Levels of Natural Language	6
Levels of Natural Language	6
More Levels	7
More Levels	8
Communication	9
The Modern View	9
Speech Acts	10
Stages in communication (informing)	11
Grammar	12
Grammar Definitions	12
Wumpus Lexicon	13
Wumpus Grammar	14
Wumpus Grammar	15
Grammaticality Judgements	16
Parsing	17
Syntax in Natural Language Processing	18
Context-free parsing	19
Augmented Grammars	20
Logical Grammars	20
Augmentation	21
Problems	22

Real Language	22
Ambiguity	23
Indexicality and Anaphora	24
Metonymy and Metaphor	25
Noncompositionality	26

References

The initial sections on natural language ambiguity and levels of natural language processing were taken (I think) from Terry Winograd, "Computer software for working with language," *Scientific American*, September 1984, pp. 230-245. I no longer have this paper.

The other sections were shamelessly taken from <http://aima.eecs.berkeley.edu/slides-tex/> with some modifications.

CS 5233 Artificial Intelligence

Natural Language – 2

Natural Language Ambiguity

3

Machine Translation

- Example English to Spanish translation
 - Did you see a white cow?
 - ¿Viste una vaca blanca?
- Can translation succeed by word or phrase substitution plus some reordering? Problems arise because natural language is ambiguous.
- Here is an infamous machine translation from English to Russian and back to English.

The spirit is willing, but the flesh is weak.

The liquor is holding out all right, but the meat has spoiled.

CS 5233 Artificial Intelligence

Natural Language – 3

Types of Ambiguities in Natural Language

- Lexical Ambiguity
 - Stay away from the bank.
- Structural Ambiguity
 - He saw that gasoline can explode.
 - I saw the man on the hill with a telescope.
 - The chickens are ready to eat.
- Semantic Ambiguity
 - David wants to marry a Norwegian.

CS 5233 Artificial Intelligence

Natural Language – 4

More Types of Ambiguities

- Referential Ambiguity
 - When a bright moon ends a dark day, a brighter one will follow.
 - She dropped the plate on the table and broke it.
- Pragmatic Ambiguity
 - Don't you know what day it is?
- Multiple Ambiguity
 - Time flies like an arrow.

CS 5233 Artificial Intelligence

Natural Language – 5

Levels of Natural Language

6

Levels of Natural Language

- Phonetics: speech sounds, e.g.,
sounds of “k”, “i”, and “t” in “kite”.
- Phonology: organization of speech sounds, e.g.,
different “k” sounds in “kite” vs. “coat”.
different “t” and “p” sounds in “top” vs. “pot”.
- Morphology: construction of words, e.g.,
use of “-s” to form plurals,
use of “-ed” to form past tense of verbs.

CS 5233 Artificial Intelligence

Natural Language – 6

More Levels

- Syntax: combination of words into phrases and sentences, e.g.,
Flying airplanes is dangerous.
Flying airplanes are dangerous.
- Prosody: rhythm and intonation of language, e.g., in English, questions usually end with increasing pitch.
- Semantics: meaning of language, e.g.,
The pig is in the pen.
The ink is in the pen.

CS 5233 Artificial Intelligence

Natural Language – 7

More Levels

- Pragmatics: effect of language on the speaker and listener, e.g.,
Can you pass the salt?
Do you know the time?
I swear to tell the truth ...
- World Knowledge: Knowledge of the physical world, social interactions, etc., e.g.,
The porridge is ready to eat.
There's a man in the room with a green hat on.

CS 5233 Artificial Intelligence

Natural Language – 8

Communication

9

The Modern View

- “Classical” view (pre-1953):
Language consists of sentences that are true/false (cf. logic).
- “Modern” view (post-1953):
Language is a form of action.
- Why?
Language is used to affect the actions of other agents.

CS 5233 Artificial Intelligence

Natural Language – 9

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

CS 5233 Artificial Intelligence

Natural Language – 10

Stages in communication (informing)

- *Intention*. S wants to inform H that P
- *Generation*. S selects words W to express P
- *Synthesis*. S utters words W
- *Perception*. H perceives W'
- *Analysis*. H infers possible meanings P_1, \dots, 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
- Ambiguous utterance
- Differing understanding of current situation

CS 5233 Artificial Intelligence

Natural Language – 11

Grammar

12

Grammar Definitions

- *Grammar* specifies the structure of messages.
- 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 *rewrite rules*, e.g.,

$$S \rightarrow NP VP$$
$$Article \rightarrow \text{the} \mid \text{a} \mid \text{an} \mid \dots$$

Here S is the *sentence* symbol, NP , VP , and *Article* are *nonterminals*

CS 5233 Artificial Intelligence

Natural Language – 12

Wumpus Lexicon

Noun → stench | breeze | glitter | nothing
| wumpus | pit | pits | gold | east | ...

Verb → is | see | smell | shoot | feel | stinks
| go | grab | carry | kill | turn | ...

Adjective → right | left | east | south | back | smelly | ...

Adverb → here | there | nearby | ahead
| right | left | east | south | back | ...

Pronoun → me | you | I | it | ...

Name → John | Mary | Boston | UCB | PAJC | ...

Article → the | a | an | ...

Preposition → to | in | on | near | ...

Conjunction → and | or | but | ...

Digit → 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

CS 5233 Artificial Intelligence

Natural Language – 13

Wumpus Grammar

$S \rightarrow NP VP$ I + feel a breeze
 | S Conjunction S I feel a breeze + and
 + I smell a wumpus

$NP \rightarrow$ *Pronoun* I
 | *Noun* pits
 | *Article Noun* the + wumpus
 | *Digit Digit* 3 4
 | *NP PP* the wumpus + to the east
 | *NP RelClause* the wumpus
 + that is smelly

CS 5233 Artificial Intelligence

Natural Language – 14

Wumpus Grammar

$VP \rightarrow$ *Verb* stinks
 | $VP NP$ feel + a breeze
 | $VP Adjective$ is + smelly
 | $VP PP$ turn + to the east
 | $VP Adverb$ go + ahead

$PP \rightarrow$ *Preposition NP* to + the east

$RelClause \rightarrow$ *that VP* that + is smelly

CS 5233 Artificial Intelligence

Natural Language – 15

Grammaticality Judgements

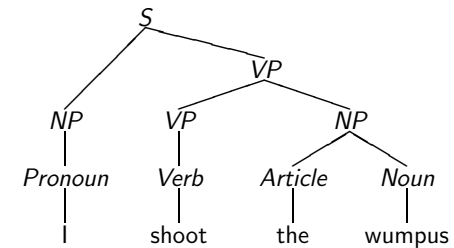
- Formal language L_1 may differ from natural language L_2
- Adjusting L_1 to agree with L_2 is a learning problem!
 - * the gold grab the wumpus
 - * I smell the wumpus the gold
 - I give the wumpus the gold
 - * I donate the wumpus the gold
- Real grammars are 10–500 pages, insufficient even for “proper” English.

CS 5233 Artificial Intelligence

Natural Language – 16

Parsing

- A parse tree exhibits the grammatical structure of a sentence.



CS 5233 Artificial Intelligence

Natural Language – 17

Syntax in Natural Language Processing

- Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “John hit Mary”

“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.”

CS 5233 Artificial Intelligence

Natural Language – 18

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 (e.g., chart parsing, Ch. 22) are $O(n^3)$ for context-free grammars and run at several thousand words/sec for real grammars.
- Context-free parsing \equiv Boolean matrix multiplication (Lee, 2002). This implies faster practical algorithms are unlikely.

CS 5233 Artificial Intelligence

Natural Language – 19

Augmented Grammars

20

Logical Grammars

- BNF notation for grammars makes it difficult:
 - to add “side conditions” (number agreement, etc.)
 - to connect syntax to semantics
- Idea: express grammar rules as logic.
 $X \rightarrow YZ$ becomes
 $Y(s_1) \wedge Z(s_2) \rightarrow X(\text{Append}(s_1, s_2))$
 $X \rightarrow \text{word}$ becomes $X([\text{“word”}])$
 $X \rightarrow Y | Z$ becomes
 $Y(s) \rightarrow X(s)$ and $Z(s) \rightarrow X(s)$
- $X(s)$ means s can be interpreted as an X .

CS 5233 Artificial Intelligence

Natural Language – 20

Augmentation

- Now it's easier to augment the rules:
 $NP(s_1) \wedge \text{Agent}(\text{Ref}(s_1)) \wedge VP(s_2)$
 $\rightarrow NP(\text{Append}(s_1, [\text{“who”}], s_2))$
 $NP(s_1) \wedge \text{Number}(s_1, n) \wedge VP(s_2)$
 $\wedge \text{Number}(s_2, n) \rightarrow S(\text{Append}(s_1, s_2))$
- Parsing is reduced to logical inference:
 $\text{ASK}(KB, S([\text{“I” “am” “a” “wumpus”}]))$
- Generation is a query with variables:
 $\text{ASK}(KB, S(x))$
- Extra arguments can be added for the parse trees, features, and semantics.

CS 5233 Artificial Intelligence

Natural Language – 21

Problems

22

Real Language

Real human languages provide many problems for NLP:

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

CS 5233 Artificial Intelligence

Natural Language – 22

Ambiguity

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

- Squad helps dog bite victim.
- Helicopter powered by human flies.
- American pushes bottle up Germans.
- I ate spaghetti with meatballs.
 - salad.
 - abandon.
 - a fork.
 - a friend.

CS 5233 Artificial Intelligence

Natural Language – 23

Indexicality and Anaphora

Indexicality refers to the situation during the communication (place, time, speaker/hearer, etc.).

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

Anaphora is using pronouns to refer back to entities previously introduced.

- 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*.

CS 5233 Artificial Intelligence

Natural Language – 24

Metonymy and Metaphor

Metonymy is 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 is the “non-literal” usage of words and phrases:

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

CS 5233 Artificial Intelligence

Natural Language – 25

Noncompositionality

Noncompositionality refers to combinations of words whose meanings are difficult to derive from the individual words.

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

CS 5233 Artificial Intelligence

Natural Language – 26