Phrase trees

PSY 200

Greg Francis

Lecture 26

Dr. Francis says something new!

Purdue University

Language

- Conveys information
- Allows us to know about things we have never experienced
  - moon flights
  - mating habits of tigers,…
- How do we do it?
- Two key aspects

Purdue University

Library of Babel

- Library of Babel https://libraryofbabel.info
- Every combination of 3,200 characters (about a page of text)
  - It includes:
    - » Gibberish (mostly)
    - » All songs
    - » All essays (includes those you wrote)
    - » An accurate description of everything you will ever do
    - » A proof that P=NP (if it exists)
    - » Lies about you and your mother
- Grammatically corrects phrases are small subset of the possibilities

Purdue University

Library of Babel

- Example texts

Symbols and grammar

- Symbols
  - words are arbitrary
  - the sound “dog” has nothing to do with dogs
  - compare driving on parkway to parking on driveway, blueberries and cranberries, hamburger…
- Grammar
  - the order of words matters
  - Dog bites man. vs. Man bites dog.

Purdue University

Grammar

- Discrete combinatorial system
  - combinations of words
- How many combinations?
- Grammatically corrects phrases are small subset of the possibilities
  - Even so, it allows you to communicate almost *everything*
- If interrupted in the middle of a sentence, you have (approximately) 10 choices for the next word
- If sentences average around 20 words, that means there are around 10^20 unique sentences

Purdue University

Grammar

- But in fact, there are infinitely many different sentences
  - there is no limit to how long a sentence can be
- For any sentence I give you, you can always make it longer by adding something like
  - Professor Francis said that, "…."

Purdue University

Vastness

- It is amazing how powerful language is
- You have probably never heard the following sentence
  - moreover, it is probably its first utterance in human history, but you understand it anyhow

Purdue University

Grammar

- You not only understand language, you *sense* when a sentence is ungrammatical
  - Is raining.
  - The child seems sleeping.
  - Sally poured the glass with water.
  - It's a flying finches, they are.
  - Rarely is the question asked: Is our children learning? (a joking George W. Bush)

Sometimes you still understand what was meant!

Purdue University

Grammar

- You can also have sentences without meaning that are perceived as grammatical
  - Colorless green ideas sleep furiously.
  - If we don't succeed, we run the risk of failure. (a not joking Dan Quayle)
  - 'Twas brillig, and the slithy toves
    Did gyre and gimble in the wabe:
    All mimsy were the borogoves,
    And the mome raths outgrabe.

Purdue University

Grammar

- These properties of language suggest that your knowledge about language grammar is a basic component of language systems
- It is *distinct* from both meaning and understanding
- Much of linguistics explores the rules of language
  - we are interested in how people perceive grammar
  - this is different from the grammar rules you may have learned in school!
    - » Which often focus on forming sentences that are *easy* to understand

Purdue University

Modern linguistics

- Noam Chomsky used the properties of grammar to demonstrate that language is quite different from other types of learning that might occur
  - it's not like learning to play a piano
  - or learning about statistical regularities in the environment (stimulus-response)

Purdue University

Nonsense sentences

- Think about the sentence
  - Colorless green ideas sleep furiously.
- What is the probability that in normal life you would hear the word "green" follow the word "colorless"?
  - it must be close to zero
- But we recognize it as a grammatically correct sentence!

Purdue University

Statistics

- If you just learned the statistical combinations of words, you might think something like this was a grammatical sentence

> House to ask for is to earn our living by working towards a goal for his team in old New York was a wonderful place wasn't it even pleasant to talk about and laugh hard when he tells lies he should not tell me the reason why you are is evident

Purdue University

Statistics

- The previous paragraph creates coherent groups of 4 words at a time (generator made sure 4 words were with high probability)
- Maybe by including a larger number of words grouped together you can insure that every sentence is appropriate
- Actually you cannot
  - Because sentences have no maximum length

Purdue University

Long-term dependencies

- Language has rules that determine what types of words can be used and when
- A word choice early in a sentence can have an effect at the end of a sentence

How Ann Salisbury can claim that Pam Dawber's anger **at not receiving** her fair share of **acclaim** for Mork and Mindy's success derives from a fragile ego *escapes* me.

1) "at not receiving" --> noun "acclaim"
2) "anger" --> "derives" (singular)
3) "How" --> "escapes" (number)

Purdue University

Long-term dependencies

- Chomsky demonstrated that long term dependencies can be *very* long
  - Consider "If…then…" and "Either…or…" sentences

If the girl eats ice cream, then the boy eats hot dogs.

Either the girl eats ice cream, or the boy eats hot dogs.

Purdue University

Recursion

- In fact, *any* sentence can go inside the "if…then" part of a sentence
  - embed a sentence in a sentence
- Thus the following is a (ugly) valid sentence

Either if the girl eats ice cream, then the boy eats ice cream, or if the girl eats ice cream then the boy eats candy.

  - recursion cannot be learned by statistics, it has to be based on rules

Purdue University

Phrases

- Every sentence is built out of phrases

The happy boy eats candy.

The first three words form a unit called a noun phrase (NP)

The happy boy

What identifies a noun phrase?

This is *not* the same analysis you did in grammar school!

Purdue University

Phrases

- *All* noun phrases obey certain rules
  - *rewrite rules*

    NP-->(det)A*N
  - *NP -- noun phrase*
  - *det -- determinant: "the", "a", "an"*
  - *A -- adjective*
  - *N -- noun*
  - *( ) -- optional*
  - ** -- as many as you want*

the happy boy

the boy

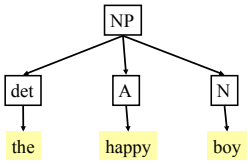
John

the tall slender woman

Purdue University

Phrase tree

- It helps to describe rules as phrase trees
- Specifies both *what* can be used in the phrase and *where* it must be used



Purdue University

Phrases

- Similarly, there are rules for all sorts of phrases in a language
- There may be many ways to rewrite a phrase!

S-->NP VP

VP-->V NP

S -- sentence
NP -- noun phrase
VP -- verb phrase

Purdue University

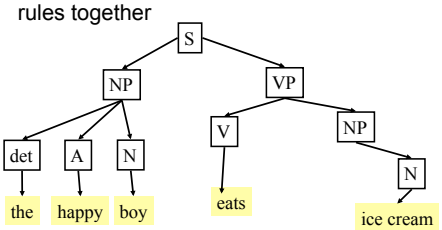
Lexicon

- We also need a mental dictionary (lexicon) that specifies parts of speech
  - N --> boy, girl, candy, hot dogs, ice cream,…
  - V --> eats, likes, bites,…
  - det --> a, the, one,…
  - A --> lucky, tall,...

Purdue University

Phrase tree

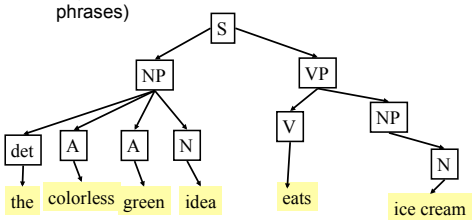
- With rewrite rules and a mental dictionary, you can create a sentence by linking the rules together



Purdue University

Sockets

- In a phrase tree, a phrase is like a component that snaps into the right place
  - any appropriate phrase works! (even nonsense phrases)



Purdue University

Usefulness

- It is important to appreciate how the phrase tree approach simplifies the description of language
- Consider how we learn a new word and know how to use it
- If you learn that a word is a noun, you can *immediately* use that noun in many different ways

Purdue University

Learning phrases

- You do not have to relearn the role of the word "boy" for each use

The *boy* eats candy.

I like the happy *boy.*

I gave the new *boy* a cookie.

The happy *boy*' s cat eats candy.

Purdue University

Long term dependencies

- Phrase trees have no problems with long-term dependencies and recursion
- The rewrite rules provide the *structure* needed to insure the right if-then combination

S--> either S or S

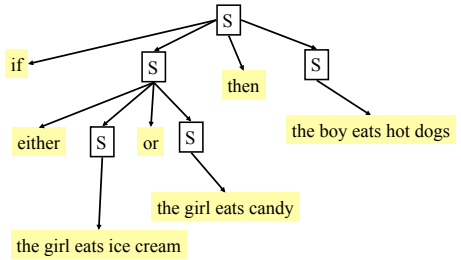
S--> if S then S

S -- sentence
either -- the word "either"
or -- the word "or"
if -- the word "if"
then -- the word "then"

Purdue University

Phrase tree

- A phrase tree can handle this type of sentence



Purdue University

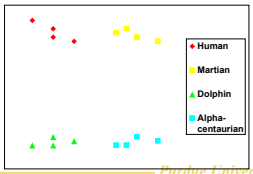
Significance

- Rules and phrase trees allow us to identify fundamental characteristics about how humans communicate
- Consider all the ways you might communicate
  - Morse code, 0-1' s, English, Spanish, tapping toes, beeps,..
  - an infinite number of ways to create a language

Purdue University

Language similarity

- *All* human languages are **very** similar, compared to the possibilities
- In some sort of *language space* all our 6000 languages are clustered together



Purdue University

Language universals

- There are several types of universals
- For example, in English the normal pattern of sentences is
  - Subject-Verb-Object
  - (There are exceptions: "*A bear he shot.*")
- This pattern is true for most of the world's languages
  - 98% of languages have the Subject before the Object (the Verb location varies across languages)
  - 80% of languages have the Subject before the Verb (the Object location varies across languages)

Purdue University

Language universals

- Most language universals involve a co-appearance of linguistic features
- For example, if a language's preferred word order is Subject-Object-Verb
  - the language is likely to form questions by adding some words at the *end* of the question
- If a language's preferred word order is Subject-Verb-Object (like English)
  - the language is likely to form questions by adding some words at the *beginning* of the question
  - "Where did he…?", "When did they…?"

Purdue University

Conclusions

- Language consists of
  - symbols (words)
  - grammar (rules)
- Language is best described as phrase trees
  - explains long term dependencies
- Language universals

Purdue University

Next time

- Words
- Mental lexicon
- Morphology
- Structure
- CogLab on Word superiority due!

- *What is the plural of "walkman"?*

Purdue University