We’re studying the properties of peoples’ unconscious knowledge of language

- You have unconscious knowledge of language.
- We can gain access to it via judgments.
  - Is this sentence a sentence of your language (say, English)?
  - In what situations would such a sentence be appropriate?
  - What can the sentence mean?

(1) a. John read the book.
   b. * Book the read John.
   c. * John the book read.

(2) a. John is too stubborn to talk to Mary.
   b. John is too stubborn to talk to.

(3) A: Do you want some coffee?
   B: Coffee keeps me awake.

(4) a. Someone told Mary to read every book by Chomsky.
    √ ‘There is a person x such that x told M to read every book by C’
    √ ‘For every book by C, someone told M to read that book’
    (cf. We have something for everyone.)
    —I know John told Mary to read Knowledge of Language, and
    Susan told Mary to read Manufacturing Consent...
    ...but did anyone tell Mary to read Syntactic Structures?
    —Yes of course; at some point or other,
    someone told Mary to read every book by Chomsky.
   b. Someone told me Mary read every book by Chomsky.
    √ ‘There is a person x such that x told me M read every book by C’
    * ‘For every book by C, someone told me that M read that book.’

(5) a. I think (that) John fixed the car with a hammer.
   b. What do you think John fixed with a hammer?
   c. How do you think John fixed the car?
   d. I wonder whether John fixed the car with a hammer.
   e. * What do you wonder whether John fixed with a hammer?
   f. * How do you wonder whether John fixed the car?

But how do we know that some of these are bad sentences of English?

(6) a. John read the book.
   b. Mary thinks that John read the book.
   c. Mary thinks John read the book.
   d. What did John read?
   e. Who read the book?
   f. What does Mary think John read?
   g. What does Mary think that John read?
   h. Who does Mary think read the book?
   i. * Who does Mary think that read the book?

The point: Pretty much all English speakers agree about this. It wasn’t taught to any of us in our English classes.

If we just went by what seem to be the rules, we should be able to say (6i).

- Sentences can be embedded with or without that.
- Questions are formed by putting the wh-word at the beginning.

So how do we know it? Basically: We know it because we’re human.

- Part of what it is to be human is to have language.
  - no other animals have language (even if exposed to it).
  - all (normal) humans have language

- Our faculty for language is innate, specified genetically.
  - Kids acquire (complicated) language quickly.
  - Kids all end up knowing stuff like ‘you can’t have the word that if you ask a question about the subject of an embedded sentence’ even though the evidence doesn’t lead them to that conclusion.

- Kids do learn their language—they don’t come with English built in.
  - The faculty for language only allows limited types of languages to be learned. Languages differ in vocabulary and languages differ in certain aspects of their grammars, but those differences are only of a small number of allowable kinds: parameters of variation.
Big picture: When we study syntax, we are partly working toward a characterization of this knowledge of language that speakers possess (unconsciously) about their language—and ultimately toward a characterization of the human faculty for language itself.

The data for syntax is generally (native speaker) judgments about the acceptability of sentences.

• We are characterizing the knowledge of language of that person.
  (knowledge of language in an *internal* sense)
• Communities whose members share very similar knowledge of language are said to speak the same language.
• For us, a “property of English” is a property of language knowledge of an individual speaker who would be classified as a member of the English speech community.

If a native speaker accepts I want to quickly go in. as a good sentence, that is a data point. Prescriptively, it may have a split infinitive and a dangling preposition, but we are in the business of describing (characterizing) native speaker knowledge (not helping the native speaker get a job as an editor).

We study the speaker’s knowledge abstracting away from what s/he may accept when drunk (or tired, or asleep, or in a coma, …); we are concerned with native speaker competence (not performance).

Phrase structure

Knowledge of language includes a lexicon (vocabulary) and rules of combination.

(First pass) The lexicon contains:
  pronunciation
  meaning
  kind of word (verb, noun, etc.)

Parts of speech. Noun, verb, preposition, adverb/adj, … These fit in different places in a sentence.

(7) The {man / puppy / *yellow / *swim / *under the table} ate lunch.

Person, place, or thing?

(8) The destruction of the rainforests worries many people.

| some tests for part of speech (in English) from Carnie (2001). |
|-----------------------------|-----------------|
| Nouns                       | Traditionally: Person, place, or thing. |
| Distributionally:           | • the subject or object of a sentence |
|                             | • modified by adjectives |
|                             | • follow determiners (the, a, this) |
|                             | • marked with case, number (singular, plural), or gender endings |
|                             | • can have derivational endings like -ment, -ness, -ing, -er. |
| Frame: X is a pain in the neck. |
| Verbs                       | Traditionally: Action (sometimes state) |
| Distributionally:           | • the predicate of the clause |
|                             | • modified by adverbs and take auxiliaries (have, be) |
|                             | • follows subject, precedes object |
|                             | • has tense (-ed), aspect (-en) endings |
|                             | • can be negated |
| Frame: They can X or They X-ed the banana. |
| Adjectives                  | Traditionally: State (modifying), qualities, attributes |
| Distributionally:           | • follows very |
|                             | • modifies noun (and follows determiner) |
|                             | • can have derivational endings like -ish or -some. |
| Frame: She is very X or I want the X book. |
| Adverbs                     | Traditionally: Modifier of anything other than a noun |
| Distributionally:           | • can have -ly ending |
|                             | • can appear at the beginning or very end of a sentence |
| Frame: Bill treats Fred X or X the students do the homework. |
Sentences have structure. Consider…

(9) This Sentence
is made up of the words
this, student, can, solve, the,
and problem. Each word has
a category (Det, N, Aux, or
V).

Parts of the sentence (strings of words) function together as a group in certain respects. These are constituents.

Some ways in which they function together: —displacement
—deletion
—coordination
—replacement
—stand alone together

This student is a constituent.

(10) a. She can solve the problem. (replacement)
    b. [This student] and Bill can solve the problem. (coordination)
    c. It is [this student] who can solve the problem. (displacement)
    d. —Who can solve the problem? —This student. (stand alone)

Same for the problem.

(11) a. This student can solve it. (replacement)
    b. This student can solve [the problem] and the mystery. (coordination)
    c. It is [the problem] that this student can solve. (displacement)
    d. [The problem], I believe this student can solve. (displacement)
    e. —What can this student solve? —The problem. (stand alone)

(12) [This student] can solve [the problem].

Solve the problem is also a constituent:

(13) a. This boy is determined to solve the problem and
    [solve the problem] he will. (displacement)
    b. John can’t [solve the problem] but this student can […] (deletion)
    c. This student will [solve the problem] and win a prize. (coordination)
    d. This student can [solve the problem] and so can I. (replacement)
    e. —What can this student do? —Solve the problem. (stand alone)

Sentences have hierarchical structure—constituents within constituents.

Not everything is a constituent. Student can is not a constituent.

(14) a. [solve [the problem]]
    b. solve the problem

[the problem] is a constituent inside another constituent [solve the problem].

Nor is solve the.

(15) a. * —This who/what solve the problem? —Student can (*stand alone)
    b. * The farmer can’t solve the problem but this […] solve it. (*deletion)
    c. * This one solve the problem. (*replacement)
    d. * This student can and farmer can’t solve the problem (*coordination)
    e. * Student can is what this solvethe problem. (*displacement)

(16) a. * —What can this student do (to) problem? —Solve the. (*stand alone)
    b. * This student can solve the and ridicule a problem. (*coordination)
    c. * Solve the is what this student can problem. (*displacement)

(17) [this student] can [solve [the problem]].

Some tree terminology:

(18) A
    B
    C
    D
    E

A, B, C, D, E are labels on the nodes of the tree. C defines a constituent containing D and E. A node dominates the nodes in its constituent. A immediately dominates B and C in (18), because there is a single line connecting them. Nodes which dominate nothing are terminal nodes; nodes which dominate other nodes are nonterminal nodes. Two nodes immediately dominated by the same node are sisters.
We can label the kinds of the terminal nodes in (17); what about the nonterminal nodes? [this student] and [the problem] are each made of a determiner and a noun. We’ll call these noun phrases (NPs).

**Phrase structure rule:** “The combination of a determiner and a noun is an NP”
or “an NP consists of a determiner and a noun.”

(19) \[ \text{NP} \rightarrow \text{Det} \quad \text{N} \]

(student and problem are nouns (N), and the and this are determiners (Det)).

The next-biggest constituent is solve [the problem], a verb and a noun phrase. This is a verb phrase (VP).

(20) \[ \text{VP} \rightarrow \text{V} \quad \text{NP} \]

So far:

(21) \[ \text{S} \rightarrow \text{NP} \quad \text{Aux} \quad \text{VP} \]

The last one (which we can even read off the tree):

(22) \[ \text{S} \rightarrow \text{NP} \quad \text{Aux} \quad \text{VP} \]

This system of rules (where you start with S) is said to generate the tree in (21):

(23) a. \[ \text{S} \rightarrow \text{NP} \quad \text{Aux} \quad \text{VP} \]
b. \[ \text{VP} \rightarrow \text{V} \quad \text{NP} \]
c. \[ \text{NP} \rightarrow \text{Det} \quad \text{N} \]
d. \[ \text{Lexicon:} \quad \text{N: student, problem} \]
   \[ \text{V: solve} \]
   \[ \text{Aux: can} \]
   \[ \text{Det: the, this} \]

And it also generates trees for other sentences:

(24) a. The student can solve the problem.
b. This student can solve this problem.
c. The student can solve this problem.

And clearly if we added more words to the lexicon (23d) we could get many more sentences, like The cat will chase three dogs.

But the system also generates sentences that aren’t so good.

(25) a. The problem can solve this student.
b. This problem can solve this student.
c. The problem can solve the student.
d. This problem can solve the student.

If this system is supposed to characterize our knowledge of English (all and only the sentences it generates are English sentences), it overgenerates in that it can form sentences which are in reality not good English sentences, and it undergenerates in that it can’t form sentences which are good English sentences (like Where did you go yesterday?).

So, we will try to improve the system to make it more correct in its predictions.

First: (Some [small] steps toward) dealing with the undergeneration problem.

**Structure of the verb phrase**

(26) a. The boy cried.
b. The dog ran.

transitive verbs

(27) \[ \text{VP} \rightarrow \text{V} \quad \text{NP} \]

Allows either of:

\[ \text{Intransitive verbs} \]

(transitive (e.g., hit) (e.g., fall)
Verbs with a PP complement

(28) a. The girl knocked on the door.
   b. The teacher hinted at the solution.

(29) \[ PP \rightarrow \text{P} \quad \text{NP} \]

(30) *Lexicon: P: on, at, …*

(31) \[ VP \rightarrow \text{V} \quad \left\{ \begin{array}{l} \text{NP} \\ \text{PP} \end{array} \right. \]

(32) 

\[ \text{VP} \quad \text{knock} \quad \text{PP} \]

\[ \text{P} \quad \text{NP} \]

\[ \text{Det} \quad \text{N} \]

\[ \text{the} \quad \text{door} \]

(33) a. The boy sent the letter to his mother.
   b. The student brought a sandwich to the dean.

(34) \[ VP \rightarrow \text{V} \quad \text{NP} \quad \text{PP} \]

(35) a. The student said that the boy sent the letter to his mother.
   b. The boy said that the dog ran.

(36) \[ S' \rightarrow \text{that} \quad \text{S} \]

(37) \[ VP \rightarrow \text{V} \quad S' \]

Rule (37) is recursive; once we have included it, our system can generate an unbounded number of sentences, since every S contains a VP which may itself contain an S which contains a VP which may itself contain an S…

(38) John said that Mary claimed that she heard that Bill insinuated that Sue suggested that the dog ran.

(39) a. The student said the boy sent the letter to his mother.
   b. The boy said the dog ran.

(40) \[ S' \rightarrow \quad \text{(that)} \quad \text{S} \]

(41) \[ VP \rightarrow \text{V} \quad \left\{ \begin{array}{l} (\text{NP}) \quad (\text{PP}) \\ S' \end{array} \right. \]

and so on…

Structure of the noun phrase

(42) a. the destruction of the city
   b. the claim that Mary left
   c. kittens

(43) a. \[ \text{NP} \rightarrow \text{Det} \quad \text{N} \quad \text{PP} \]
   b. \[ \text{NP} \rightarrow \text{Det} \quad \text{N} \quad S' \]
   c. \[ \text{NP} \rightarrow \quad \text{(Det)} \quad \text{N} \]

(44) \[ \text{NP} \rightarrow \quad \text{(Det)} \quad \text{N} \quad \left\{ \begin{array}{l} \text{PP} \\ S' \end{array} \right. \]

(45) a. John’s cancellation of the party.
   b. the student’s claim that Bill left.

(46) a. * John’s the cancellation of the party.
   b. * This the student’s claim that Bill left.

(47) \[ \text{NP} \rightarrow \quad \left\{ \begin{array}{l} \text{Det} \\ \text{NP’s} \end{array} \right. \quad \text{N} \quad \left\{ \begin{array}{l} \text{PP} \\ S' \end{array} \right. \]

and so on…

Co-ordinate structures

(48) The boy and the girl solved the problem.

(49) 

\[ \text{NP} \quad \text{and} \quad \text{NP} \]

\[ \text{Det} \quad \text{N} \quad \text{boy} \quad \text{Det} \quad \text{N} \quad \text{girl} \]

Coordination: Combining two categories of the same type, yielding a category of the same type.
Mary knocked [ [on the door] and [on the window] ]

and so on...

**Grammatical functions**

Subject and Object are both NPs.

The **subject** of the sentence S is the NP which is a daughter of S: [NP, S]
The **object** of the sentence S is the NP which is a daughter of its VP: [NP, VP]
The **predicate** of the sentence S is the VP which is a daughter of S: [VP, S]

**The Lexicon**

We’ve done something about the *undergeneration* problem—the system now generates many more grammatical sentences—but we’ve made the *overgeneration* problem even worse—the system now generates scads of ungrammatical sentences as well.

Part of our knowledge of English is the knowledge that different verbs have different requirements about their objects (or *complements*).

**Subcategorization restrictions**

The verb *give* needs an object NP and a location PP. Like *put*.
The verb *sleep* needs no objects. Like *cry*.

Individual information about lexical items belongs in the lexicon.

a. *give* [V; — NP PP]
b. *sleep* [V; —]
c. *kick* [V; — NP]
d. *think* [V; — S’]
e. *rely* [V; — PP]

**Lexical Insertion Rule:**

Insert lexical item X under the terminal node Y, where Y corresponds to the categorial properties of X, and YP corresponds to the subcategorization properties of X.

**Selectional restrictions**

Sincerity and Stonehenge are nouns, but they aren’t the right kind of nouns in (55b–c).

**Derived categories**

a. *translation* < translate + -ion  
(b- ion changes V to N)
b. *sheepish* < sheep + -ish  
(-ish changes N to Adj)
c. *colonize* < colony + -ize  
(-ize changes N to V)

If we list *translation* and *translate* separately, we miss a generalization (there’s an *-ion* nominalization form for a lot of different verbs).
Morphological subcategorization: This happens before syntax starts; translation is just an N as far as syntax is concerned. The morphological combination system is “between” the lexicon and syntax.