CAS LX 522
Syntax I

3

Features (II), constituents, Merge
(2.4.2-3.4)

Reaching agreement

English speakers agree about this:

1) Three dogs are here. One dog is here.

2) *Three dogs is here. *One dog are here.

What's wrong with the second set of sentences? How do we distinguish good ones from bad ones (which we *can* do)?

Classes of subjects and verbs

The standard answer is that the subject and verb must *agree*. What does that mean?

There are two different kinds of subjects and two different kinds of verbs. Subjects of one kind can only appear with verbs of one kind, subjects of the other kind can only appear with verbs of the other kind.

Plural subjects

What seems to differentiate them is that subjects of one class are all *plural*, and subjects of the other class are not.

So, nouns that are plural… that have the property of being plural… that have the [plural] feature… require a verb chosen from one class. Other nouns require a verb chosen from the other class.

"Plural" verbs

There's nothing in particular we can grab onto that differentiates the classes of verbs, except that one class occurs with plural subjects, and the other occurs with non-plural subjects.

So, we'll call the verbs that go with plural subjects "plural"—why not? We can say that these verbs also have the property… the feature… [plural].

Agreement

By naming the features that way, we can state the requirement simply, by defining a notion of "agreement" (which basically means: "matching"). If we call [plural] (or lack of [plural]) a specification for *number*, we can say that…

Subjects and verbs must agree in number.

Simple.

The reason for going over all that is to kind of drive home the idea that positing a [plural] number feature and stating a rule about agreement is about as simple as it can be. There's pretty much no other way to describe this effect that isn't just equivalent. So, if agreement is part of syntax (and let's say that it is), we're already off and running with rules/constraints sensitive to features.

Categories

We also approached the idea that if we're going to describe the syntax of a language, we need to divide words into classes, *syntactic categories*, that determine where they can appear.

Lexical: N: noun, V: verb, A: adjective, P: preposition

Functional: D: determiner, T: tense/inflection, C: complementizer

Categories

There are ways in which some behave alike, we saw some things affect verbs and adjectives but not nouns or prepositions. So, verbs and adjectives should share some property. Same for nouns and adjectives. We can call the verb-adjective property [+V] and the noun-adjective property [+N].

Functional categories seem like they might be "functional versions" of lexical categories, so D might be a [+functional] N.

English pronouns

The English pronouns make several distinctions over and above a singular/plural distinction.

One distinction is *person*, which is sensitive to who is talking and to whom.

English (and most languages) distinguish three persons.

	singular	plural
first person	I	we
second person	you	you
third person	he/she/it	they

English pronouns

We could model person with [1], [2], and [3]—except that that predicts eight distinctions, and we have only three.

	singular	plural
first person [1]	I	we
second person [2]	you	you
third person [3]	he/she/it	they

English pronouns

Rather, we want to use two features, which only predict four. Slightly better.

By eliminating [3], we predict a system like that below—as well as a [1,2] combination that is not morphologically distinguished in English.

	singular	plural
first person [1]	I	we
second person [2]	you	you
third person [3]	he/she/it	they

English pronouns

What about [1,2]? There's no special pronoun form, but what would it *mean*?

Well, [1] is the speaker, [2] is the person being spoken to. So [1,2,pl] would be *we* (including you). Not the same as [1,pl], *we* (excluding you).

	singular	plural
first person [1]	I	we
second person [2]	you	you
third person [3]	he/she/it	they

English pronouns

Some languages distinguish inclusive and exclusive *we* morphologically, e.g., Dakota.

No languages seem to distinguish 8 persons.

	singular	plural
first person [1]	I	we
second person [2]	you	you
third person [3]	he/she/it	they

Gender

Many languages distinguish nouns on the basis of "gender" as well.

English: *she/he/it* (3rd person pronouns)

Gender often comes in 2-3 flavors (masculine, feminine, neuter) which often corresponds roughly to biological gender where applicable.

Person, number, gender

So, to describe the English pronoun system (that is, to predict what the subject pronoun will sound like), we need to pay attention to person (I vs. you), number (I vs. we), and gender (he vs. she).

Yet—there aren't so many pronouns. It's just "you" whether the addressee is male, female, or neuter.

How does this work?

What features does a subject pronoun need to have to sound like "he" (third person, masculine, singular)? To sound like "they" (third person, plural)? To sound like "we" (first person, plural)?

Suppose we want to write the rules down. If a subject pronoun has features…it sounds like…otherwise, if…

Is there something special about third person, masculine, singular? About "you"?

Terminology: Phi-features (ϕ-features)

Collectively, person, number, and gender features are referred to as ϕ-features.

These are the features that are generally involved in subject-verb agreement, across languages.

We group them together because they seem to have their effects together (that is, not separately).

Case features

English pronouns change form also depending on where they are in the sentence.

He left. I saw him. He saw me.

The information about syntactic position is encoded by case features.

In English, case is only visible on pronouns.

In many other languages, case is visible on all nouns (and sometimes on words modifying nouns, like adjectives or determiners).

Case names

In English, we distinguish nominative (on subjects), genitive (on possessors), and accusative (elsewhere)

Singular			Plural		
Nom	Acc	Gen	Nom	Acc	Gen
I	me	my	we	us	our
you	you	your	you	you	your
he	him	his	they	them	their
she	her	her	they	them	their
it	it	its	they	them	their

Features & pronunciation

Lexical items are bundles of features. Like [Acc, 1, sg, PRN].

The syntactic system arranges these lexical items into sentences, which have some pronunciation, and some meaning.

The pronunciation of [Acc, 1, sg, PRN] is "me"

Singular			Plural		
Nom	Acc	Gen	Nom	Acc	Gen
I	me	my	we	us	our
you	you	your	you	you	your
he	him	his	they	them	their
she	her	her	they	them	their
it	it	its	they	them	their

Features & pronunciation

Not every distinction:

Only 3rd person singular distinguishes gender.

2nd person does not distinguish number or between Nom and Acc.

3rd person singular feminine doesn't distinguish between Acc and Gen.

The structure of the paradigm can give us clues as to how the interface rules work.

Singular			Plural		
Nom	Acc	Gen	Nom	Acc	Gen
I	me	my	we	us	our
you	you	your	you	you	your
he	him	his	they	them	their
she	her	her	they	them	their
it	it	its	they	them	their

Verbal features

Some features are specific to verbs.

[past], for example, differentiating *write* from *wrote*, *kick* from *kicked*. This is a *tense* feature. Semantically, it relates the action/state described by the verb to another point in time (such as, the time of utterance).

Some languages have a special form of the verb for the future as well; [future].

Verbal features

We can characterize present tense as being non-past, non-future.

In English, future is expressed in other ways, with a modal (*will*) or with the verb *go*. English does not seem to make use of the [future] feature; in English there is just past and non-past.

Cf. duals and the use of [sg] on nouns.

Realizing verbal features

Let's think a bit about regular verbs. Past tense of *kick* is *kicked*. No matter what the subject is. Present tense of *kick* is *kick*—except when the subject is third person singular, then it's *kicks*.

Gender never seems to have an effect in English. But what makes third person singular special? (We've seen this question already...)

Be

One verb is special: *be* seems to inflect more than other verbs. I am, it is, they are, I was, it was, they were. Gender still has no effect, but how might we understand the forms of *be* as compared to the forms of regular verbs?

Still a number distinction in the past. And a person distinction in the present. "You" again is a bit special.

Participles

English verbs can also take on a *participle* form: *writing, written*.

Not tense, but *aspect*.

The *-ing* form ("present participle") comes after *be*, indicating a continuing event.

The *-en* form ("past participle") comes after *have*, indicating a completed event.

Tense can still be expressed—on the auxiliary: *I have written, I had written, I am writing, I was writing*.

Participles

Adger's proposal:

Present participle: [V, part] (*writing*)

Past participle: [V, part, past] (*written*)

I distance myself from that because it is not at all clear that the [past] feature in Adger's past participle has anything in common with the [past] feature in an actual past tense. A better name would be, e.g., [perf].

Bare verb/infinitive

I want to win the lottery.

The bare form of the verb (often appearing after *to*) is the *infinitive*.

We will assign infinitive forms the feature [Inf].

The fact that the infinitive is a bare verb (no suffixes or other inflection) in English may be something of a coincidence. Other languages mark the infinitive with a special verb form, on a par with participles or tensed verbs.

Verb agreement

Verbs very often (across languages) agree with the subject in ϕ-features as well.

I eat bagels.

He eats bagels.

They eat bagels.

However, again: *eat* isn't really "plural" in any sense. Plurality is a property of the subject, but it is *reflected* in the morphology of the verb.

Verb agreement

In English, only *finite* verbs show agreement (those that are not infinitives or participles).

In fact, only *present tense* verbs do, with the single exception of the *copula* (*be*).

In other languages, agreement sometimes appears on other forms. Participles, for example, sometimes agree with their object. Infinitives very rarely agree with anything.

Summary:

Categories: N, V, A, P, D, T, Aux, C

Nominal features: case ([nom],[acc], [gen]), ϕ-features: person ([1],[2]), number ([pl]), gender ([fem])

Verbal features: tense ([inf],[past]), aspect ([part], [perf])

General structure of our account

Knowing a language is:

 knowing the "words"

 knowing how to put them together

 knowing how to pronounce them

 knowing what they mean in combination

The lexicon

To construct a sentence, we start with the "words" and put them together.

We can describe the knowledge of the words of a language as being a list, a mental *lexicon*.

Interfaces

We can view a "word" as a bundle of features, as defined by its properties. The grammar *assembles* words into sentences. The sentences are *interpreted* and *pronounced*.

"words"

Lexicon

assembly

interfaces

Grammar

A-P system   C-I system

sound   meaning
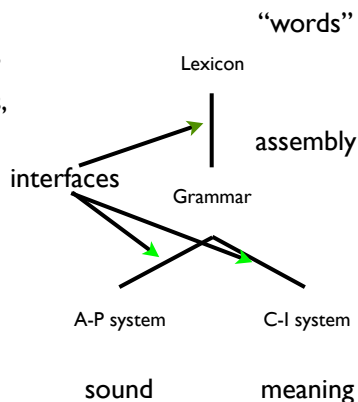
Interfaces

The assembly process is the grammar proper.

The system that interprets sentences is another cognitive module ("conceptual-intensional system") concerned with meaning, reasoning, etc. It interprets the constructed sentence at the *interface*.

The system that determines the pronunciation of sentences is yet another cognitive module ("articulatory-perceptual system"), interpreting the constructed sentence at its interface.

Lexical items

Part of our language knowledge is the knowledge of the lexicon.

The lexicon is a list of the "words"

More accurately, it is a list of the things sentences are made of.

It is traditionally considered to be where "unpredictable" information is stored. The sound, the meaning, the grammatical category, and other features.

Features of lexical items

A lexical item is a bundle of properties. It is a meaning, linked with instructions for pronunciation, linked with syntactic properties like category.

We represent these properties as *features*.

Features of lexical items

Any given lexical item has:

Semantic features
Phonological features
Syntactic features

When it comes to syntax, syntactic features certainly matter. But no language seems to arrange its sentences such that words that start with *t* are first.

Hypothesis: Syntax can only "see" syntactic features.

The structure of sentences

1) You will give it to her

2) You will give the book to your roommate

3) You will give the book about syntax to your roommate's sister

Someone doing the giving

Something changing hands

Someone receiving the thing

Sentential players

It's like there's a "spot" for each of these players:

1) __ will give __ to __

And it doesn't matter whether the "player" is described with one word, two words, or several words.

Constituents

Each "unit" of this sort is what we'll call a *constituent*. We enclose them in brackets to indicate that the words form (and behave as) a unit.

1) [You] will give [the book] to [your roommate].

A significant property of language is that these units can be arbitrarily complicated:

2) [You] will give [it] to [Ed's roommate's sister's friend]

Arbitrarily complicated

- [Ed's roommate's sister's friend]

This has sub-units within it:

- [[Marge]'s friend]
- [[Ed's roommate's sister]'s friend]

And within that:

- [[[Marge]'s sister]'s friend]
- [[[Ed's roommate]'s sister]'s friend]
- [[[[Ed]'s roommate]'s sister]'s friend]

In general, it looks like wherever a <u>name</u> can go, so can [<u>name</u>'s <u>noun</u>].

[<u>name</u>'s <u>noun</u>]

Wherever a <u>name</u> can go, so can [<u>name</u>'s <u>noun</u>].

1) I gave the book to Homer.
2) I gave the book to Bart's father.
3) I gave the book to Lisa's brother's father.

This replacement rule is *recursive*. The thing we are replacing is also contained in the thing we replaced it with.

Groups of groups of groups

Sentences are made of *grouped* words. These groups can be contained in other groups, arbitrarily deep. A group of this kind: a *constituent*.

Constituents can contain constituents that can contain constituents, etc.—The structure of a sentence is *hierarchical*.

Constituents behave as a unit…

Constituents

Functioning as a unit…

- The students did <u>their syntax assignment</u>.
- The students did <u>the crossword puzzle</u>.
- <u>John</u> did the crossword puzzle.
- <u>The crossword puzzle</u> is what John did.
- *<u>Crossword puzzle</u> is what John did <u>the</u>.
- John <u>likes</u> the crossword puzzle.
- John likes the <u>jigsaw</u> puzzle.
- John likes the <u>theater</u>.

Finding constituents

How do we find constituents in a sentence? For many of them, we can guess, but a guess isn't evidence.

The *structure* of a sentence has consequences.

To find the constituents (to determine the structure) we test for the consequences.

Constituency tests

Replacement test

Fragment test

Ellipsis

Clefting

Movement test

Replacement test

A constituent is a group of words which function as a unit. If you can *replace* part of the sentence with another constituent (the smallest constituent being a single word), this tells us that the replaced section of the sentence is a constituent.

- This isn't foolproof, but it usually works if you try to keep the meaning as close as possible.

Replacement test

1) The students left.
2) **They** left.
   The students is a constituent.
3) The students will eat the sandwiches.
4) They will eat the sandwiches.
5) The students will eat **them**.
6) The students will **dine**.

- [The students] will [eat [the sandwiches]].

Sentence fragment test

Generally, only constituents can be used in the fragmentary response to a question.

- Who will eat the sandwiches?
   The students.   *Students will eat the.*
- What will the students do?
   Eat the sandwiches.   *Eat the.*
- What will the students eat?
   The sandwiches.
- [The students] will [eat [the sandwiches]].

Ellipsis test

If you can *elide* a string, it qualifies as a constituent.

- Ellipsis is really deletion of a string from a sentence. Sometimes this is "repaired" by using the verb *do*, something which we will seek to explain at a later point.

The professors will eat the sandwiches, and then..

The students will.

The students will eat the cookies, and then…

*The professors will sandwiches.

> **WARNING: Passing a constituency test constitutes evidence for a constituent. Failing a constituency test tells you little—there may be other reasons for the ungrammaticality.**

Movement (topicalization) test

Sometimes you can "move" a string of words to the front of a sentence (then generally interpreted as the topic of the sentence). When you can, you've found a constituent.

- The sandwiches, the students will eat _.
- Eat the sandwiches, the students will _.
- The students, they will eat the sandwiches.
- *Students will, the eat the sandwiches.
- *Students, the will eat the sandwiches

   - Failing a constituency test isn't evidence *against* constituency!

Clefting test

Like the movement test, if you can fit your string into the frame *it be X that S* (where you move the string X from inside S), X is a constituent.

- It's the sandwiches that the students will eat _.
- It's the students that _ will eat the sandwiches.
- It's eat the sandwiches that the students will (do) _.
- *It's students eat that the _ will the sandwiches.
- *It's eat the that the students will _ sandwiches.

Finding constituents

Tests: Replacement, (ellipsis,) movement, clefting, fragment.

Some to try:

- Two African swallows can carry a coconut.
- A cat was walking down the street.
- A creature was stirring up trouble.
- Flying planes can be dangerous.
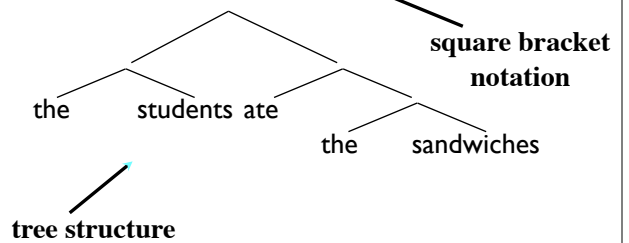
And all through the house...
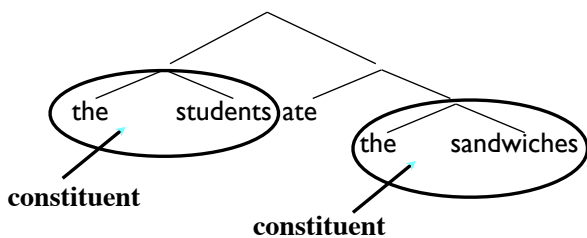

Bonus: the breakdown.

## Trees, hierarchy, and constituency

- [The students] [ate [the sandwiches]]

square bracket notation

the   students ate

the   sandwiches

tree structure

Trees, hierarchy, and constituency

- [The students] [ate [the sandwiches]]

the   students ate

the   sandwiches

constituent

constituent

Trees, hierarchy, and constituency

- [The students] [ate [the sandwiches]]
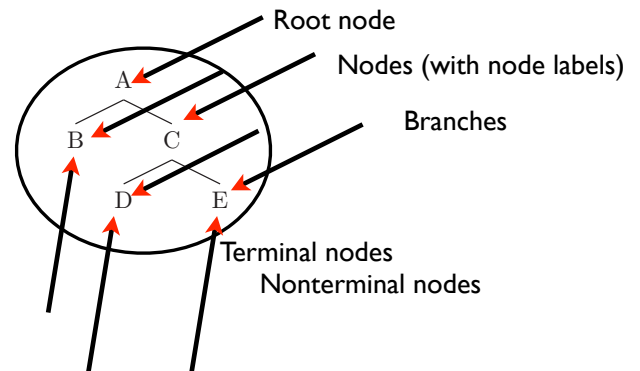
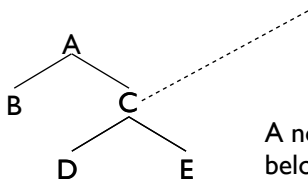the   students ate

the   sandwiches
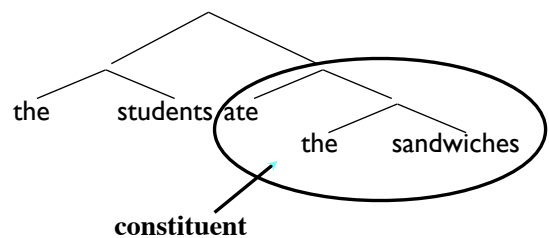
constituent

The triangle

Sometimes, when the internal constituency is unknown or unimportant to the current discussion, a triangle is used instead.

the students   ate   the sandwiches

Trees

Root node

Nodes (with node labels)
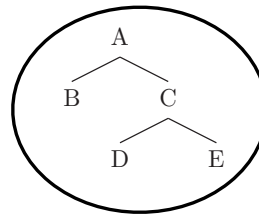
Branches
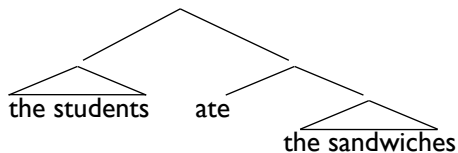
Terminal nodes
Nonterminal nodes

Tree relations

A node X **dominates** nodes below it on the tree; these are the nodes which would be pulled along if you grabbed the node X and pulled it off of the page.

Acts as a unit. Is a constituent.

Tree relations

A node X *immediately dominates* a node Y iff X dominates Y and is connected by only one branch. Or, X is mother of Y.

Nodes X and Y that share the same mother are sister nodes.

Verbs and substitution

One of the ways we know a verb is a verb (category) is by observing that it can substitute for other verbs.

Pat likes to sing. Pat likes to drive.

Pat bought a book. *Pat bought (a) sing.

Pat likes to eat sandwiches.

*Pat unpleasant to eat sandwiches.

So is *eat sandwiches* a verb?

Well, kind of, yes.

It's a constituent, a phrase, that has the properties a verb does. A *verb phrase*.

The making of a phrase

We're trying to characterize our knowledge of syntactic structure.

Our grammatical knowledge is a system (we can judge new sentences).

All things being equal, a theory in which the system is simpler (needed fewer assumptions) is to be preferred over a theory that entails more complex one.

The making of a phrase

In that spirit, we know that a phrase differs from a word in that it *contains* words (or other phrases).

- We've seen that when words are combined into a phrase, the phrase inherits the properties of one of the things we combined. (The phrase has a head).

- Suppose: a **phrase** can arise from **merging** two words together, with one taking priority. In a way, attaching one word to another.

The making of a phrase

What will Pat do?
- sing
- eat sandwiches

What does Pat like?
- to eat sandwiches
- to sing
- [to [eat sandwiches]]

So, a phrase can also arise from combining *to* and a verb phrase, to make a bigger phrase.

Merge

So, let's go for the simplest theory of structure we can (and only move away from it if the simplest theory won't work)

A phrase is a syntactic object formed by combining (*merging*) two syntactic objects, with the properties inherited from one of them (the *head* of the phrase).

A word is a syntactic object.

Merge, in the abstract

B

D   E

A good way to think about this is that we have a number of syntactic objects lying around on a workbench of sorts.

We use the operation Merge to assemble them together into one syntactic object.

Merge, in the abstract

B   C

          D   E

We combine D and E using Merge to form a combined syntactic object.

We need to call our new object something, so we call it C.

C is now a syntactic object (containing D & E).

D and E are now "off the table"—we can't Merge D with anything because it's inside C. ("Merge only combines objects at their root nodes").
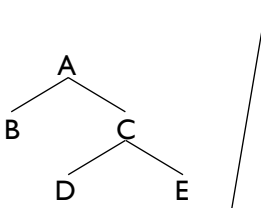
Merge, in the abstract

          A

B   C

          D   E

Since C is now a syntactic object, we can combine C with the other syntactic object, B, to form a new syntactic object we'll call A.

Now, all we're left with is the single syntactic object A.

Merge, in the abstract

A
B   C
D   E

When two objects are Merged, one of them is the **head**, the most important one.

The head determines the properties of the constituent— that is, the features of the head **project** to become the features of the whole combined object.
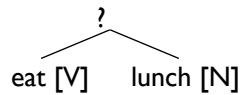
Trees and constituency

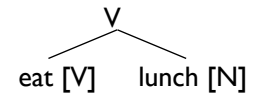- Pat will eat lunch.
- Pat will dine.

eat [V]   lunch [N]

Trees and constituency

- Pat will eat lunch.
- Pat will dine.

?
eat [V]   lunch [N]

Trees and constituency

- Pat will eat lunch.
- Pat will dine.

V
eat [V]   lunch [N]

So how do we know which is the head?

When we Merge two things, one is the head, and determines the properties of the resulting syntactic object.

The next thing we'll turn to is the question of how the syntactic system knows which is the head.