

CAS LX 422/  
GRS LX 722  
Intermediate Syntax

# 2

Morphosyntactic features  
(2.1-2.4.1)

## In search of the atoms of the system

Syntax is—at least in large part—the study of the principles of sentence formation.

There are principles that govern which combinations of words are sentences of English. What is the “vocabulary” of these principles? What are they stated in terms of?

“Words” might be a good starting point.

## The atoms of the system

However, it seems that it isn't exactly the words—it is the *properties* each word has that seems to be basic. Verb or not a verb, plural or not plural...

- 1) Three dogs are here. One dog is here.
- 2) Three geese are here. One goose is here.
- 3) Three deer are here. One deer is here.

## Properties... features...

Words have *properties*. Like being a verb, or being plural.

“Plural” is an *abstract concept*—there is no direct map to morphology (*deer, geese, mice, feet, dogs, children, data*), but they all make the same demands of the verb.

## Properties... features...

Same “agreement” requirement, regardless of the actual morphological shape.

The abstract property of “plural” (or “singular”) seems to be what the grammar is sensitive to. That's smaller than a word.

(Morphosyntactic) features

## Agreement

In English, the subject and the verb of a sentence need to *agree* in number and (for *be*) person.

- 1) The dog wants food. The dogs want food.
- 2) The dog is hungry. The dogs are hungry.
- 3) I am hungry. We are hungry.

## Agreement & interpretability

If the subject is plural (has a plural feature) then the verb must take on a “plural” form.

It is crosslinguistically common to have this kind of agreement relation between subject and verb.

Intuitively, the plural feature is *interpretable* on the subject, contributes to the meaning, “belongs there” in some sense. On the verb, the (agreeing) plural feature is just a “reflection”, *uninterpretable*—much more on that later.

## Data from other languages

1) Il a dit qu'elle était malade  
he<sub>[3.sg]</sub> have<sub>[3.sg]</sub> said that she was ill  
'He said that she was ill.'

2) Ils ont dit qu'elle était malade  
they<sub>[3.pl]</sub> have<sub>[3.pl]</sub> said that she was ill  
'They said that she was ill.'

Why does it matter what other languages do?

## What are the features?

Some features—that is, some properties—seem to matter for the purposes of syntax, some don't. So, the identity of the features need to be part of our theory—features are just “properties”—but, the features that syntax relies on are the *relevant* properties.

We're looking for the minimal (least complicated) set of features that suffices to explain the grammar.

## What are the features?

No language says that subject and verb must agree in the feature [invented in early September], although there are things that have this property.

For the purpose of describing the grammar and explaining the syntactic principles, we don't care about [invented in early September].

We have evidence, though, that [plural] matters to syntax (at least in some way...)

## [plural]

We know number matters. In English, things can be singular or plural. So, a first guess is that nouns have either a [singular] feature or a [plural] feature.

Hypothesis:

[sg] and [pl] are features a word can have.

Prediction:

Four classes of words: [sg], [pl], [sg,pl], []

## Science

That thing we just did? It was science.

We had some observations, the existence of singular and plural forms—and they matter for the grammar.

We formulated a hypothesis.

We identified other facts that we expect to hold—the *predictions*—if the hypothesis is correct.

Now, we'll go back to the data to see if the predictions are borne out.

## Overgeneration

However—it turns out that the prediction is *not* met in the data.

The prediction is that there are four number classes of nouns, but English has only two.

This hypothesis *overgenerates*—it predicts the existence of the feature combinations, but it also predicts other feature combinations that don't exist. (Or, at least, that the syntax of English does not seem to make use of.)

## [plural]

So, we have a new set of observations, now including the fact that the two classes we identified before are the only classes there are.

And there's a simpler story we can tell, one that *predicts* exactly two classes.

[plural] for plurals, [] for singulars.

## Undergeneration

An analysis that says “All words are singular” *undergenerates*.

All predicted combinations are attested.

Some attested combinations are not predicted.

## Why we're using “features”

What we're trying to do is characterize the syntax of a language by identifying what properties (of words, generally) it is sensitive to.

Any given word has a bunch of properties. Some matter, and some don't.

“Word” is not really well-defined—what we think of as a word can have several sub-parts, and a “word” like *bank* is really a bunch of different words.

## Features as named properties

But we want to be precise, and so we will name these properties.

Words will have features, but also the grammar will refer to them (“the subject and the verb agree in number”).

Also, remember that we hope to be doing a *cross-linguistic* project, looking for rules and principles that hold in many/all languages.

## Formalizing properties

We hypothesized that English nouns are either plural or they're not. So, “singular” is defined in opposition to plural, it isn't a property of its own—only “plural” is.

Features are usually written inside brackets (a “feature matrix”) [..., plural, ...]

So for plurals, we wrote “plural” (or “pl”), and for singulars we conserved our energy and wrote nothing.

## Binary features

Another way we could have done it is to have a dedicated “slot” in the feature matrix for [plural], and have two variants:

[+plural] (plural)

[-plural] (not plural)

Features formalized this way are *binary features*. The ones from before (there or not there) are called *privative features*.

## Which is better? Privative or binary?

The difference between privative and binary features is very subtle, and mainly comes down to what allows us to state generalizations we see most simply.

The basic difference is: what if there is a grammar rule/principle that affects only singular nouns? How do you describe the conditions under which the rule is used?

## The Hopi dual

- 1) Pam wari  
that ran[sg]  
'He/she ran.'
- 2) Puma yùutu  
those ran[pl]  
'They (plural) ran.'
- 3) Puma wari  
those ran[sg]  
'Those (two) ran.'

## Hopi morphology

In Hopi, the dual is expressed by *combining* singular and plural.

Unlike what we observed about English— for Hopi, we have kind of an explanation of this if we analyze dual as [+pl, +sg] (or as [pl, sg]).

So, we seem to need to specify [±sg] for Hopi, but not for English.

## Overgeneration?

The Hopi dual can be nicely described as being [+plural, +singular].

So for Hopi we need both [±plural] and [±singular] (or the privative analog).

Which by itself predicts the existence of a *fourth* number: singular, plural, dual, ...and neither singular nor plural. Yet Hopi has no fourth class of this sort.

## The fourth number?

In fact, across languages, there doesn't seem to be a fourth number. There's really just the three kinds: singular, plural, and dual.

If we use privative [sg] and [pl] features, we could suppose that there's a principle of language that prevents a “numberless noun” and treats any such noun it encounters as singular. (Also: mass nouns? They act singular in some respects in English: *soup is tasty*.)

## Category

Syntax is concerned with *distribution*.

Words seem to come in distributional classes.

One class of words can appear after the possessive pronoun *my* (*my book*, \**my at*, \**my quickly*, \**my explode*, \**my purple*). The *nouns*.

One class of words is compatible with past tense. The *verbs*.

One class of words is compatible with comparative (*happier*). The *adjectives*.

## Category

Words can be separated into classes: noun, verb, adjective, preposition, etc.

Classes also vary with respect to the kind of morphological endings they can have, and so forth. (*Arrival*, *replacement*, *destruction*; *widen*, *computerize*)

## Distribution examples

- They have no *noun*.
- They can *verb*.
- They are *adjective*.
- Very *adverb*, very *adjective*.
- So long as it makes sense (e.g., with gradable adjectives; #*they are very absent*).
- Right *preposition*. (*right over the house*)

## Nouns and verbs

Nouns have a category feature [N].

*Books* [N, pl]

Verbs have a category feature [V].

*Complained* [V]

Two independent features.

Four predicted categories.

## [N], [V], [N,V], [ ]

So, nouns are [N], verbs are [V].

What might [N,V] be? Maybe adjectives are a bit “nouny” and “verby” at the same time.

And the fourth possibility? [ ]?

The other basic category would presumably be prepositions.

But, really? [ ]? Well, we’ll switch to binary notation for now, to soothe the nerves.

## [±N, ±V]

The [±N, ±V] category system may seem a bit “out of the blue.” But it does yield some descriptive benefit. To wit:

Consider what *un-* can attach to:

*untie*, *unfold*, *unwrap*, *unpack*

*unhappy*, *unfriendly*, *undead*

\**uncity*, \**uncola*, \**unconvention*

\**unupon*, \**unalongside*, \**unat*

## [±N, ±V]

Basically, it applies to (reversible) verbs and adjectives, but not to nouns or prepositions.

Well, what are those?

## Russian case

Case is a morphological form nouns take on depending on where they are in the sentence (subject vs. object). English pronouns show this distinction: *I like her, she likes me*. Some languages (like Russian) show differing case forms on all nouns.

When Russian nouns are modified by an adjective, the adjective is *also* marked for case.

## Russian case

What gets marked for Case in Russian?

1) Krasivaya dyevushka vsunula  
beautiful girl put

chornuyu koshku v pustuyu korobku  
black cat in empty box  
'The beautiful girl put the black cat in the empty box.'

## Categories: Lexical vs. functional

Nouns, verbs, adjectives, adverbs: These are *lexical* categories. They carry significant and arbitrary meaning, and they are *open-class* (new ones can be invented).

But not all words are of this kind (except maybe those on telegrams).<sup>†</sup>

<sup>†</sup>Telegram (n.): An ancient form of texting.

## Functional categories/ syntactic "glue"

Sentences are held together by little "function words" as well. These are *functional* categories.

1) I expect that the CEO will want to retire.

Determiners: *the, a(n), some, every, that, ...*

Pronouns: *you, him, they, my, your, ...*

Infinitival to: *to*

Auxiliaries/modals: *have, be, do, can, should, ...*

Complementizers: *that, for, if, ...*

## Determiners

Determiners generally come before a noun, and come in a few different types. There are differences between the types, though for now we'll lump them together. Category: [D]. (Or maybe [+N, -V, +Functional]?)

Articles: *the, an*

Quantificational determiners: *some, most*

Interrogative determiner: *which*

Demonstratives: *that, this*

Possessive pronouns: *my, your, their*

## "Pre-noun things" vs. determiners, adjectives

Can we lump determiners together with adjectives?

They both come before nouns.

They both seem to "modify" the noun.

If we didn't need both categories (if they don't matter for syntax/distribution), we'd have a simpler theory putting them together.

*Tall building, that building, a building, my building.*

## Determiners vs. adjectives

1) The big fluffy pink rabbit

2) \*The my rabbit

3) \*The that rabbit

4) \*Every my rabbit

Determiners cannot co-occur with other determiners, must precede any adjectives.

Adjectives can occur with other adjectives.

To properly describe the distribution of these elements, we really need to separate them into two classes. Lumping them together will not give us a simpler descriptive systems.

## Pronouns

Pronouns differ from nouns in a couple of ways (example: case marking), and should be considered a *functional* category.

The pronouns of English express *person*, *number*, and *gender*.

1st person: *I, me, we, us*

2nd person: *you*

3rd person: *he, she, him, her, they, them, it.*

## Pronouns are Ds.

We'll come back to this again later on, but we will treat pronouns as having category [D], like, say, *the* or *which*.

1) We linguists must stick together.

## Auxiliaries and modals

Different from verbs: *have, be, do, will, can, might, must, should, could, would, ...*

In questions, auxiliaries "invert" with the subject, verbs don't.

*Will you leave? Can you leave?*

*Do you leave often?*

*\*Leave you often?*

## Auxiliaries and modals

Auxiliaries occur before *not*, verbs don't

*You will not leave. You did not leave.*

*\*You left not.*

Notice the extra *do*—"do-support"

Auxiliaries are responsible for things like tense, mood, modality, aspect, voice.

We abbreviate their category as [T] ("tense"). (again, maybe [-N,+V,+Functional?])

## Infinitival to

1) I like to go to the movies.

Kind of looks like a preposition, but it's not. Prepositions take nouns, *to* as a P has a kind of contentful meaning (endpoint of a path). Infinitival *to* takes (bare) verbs only, means nothing (apart from "untensed").

It might be more like a modal: *To* and modals (*can*, *might*, *should*) seem to appear in the same place (between the subject and a bare verb form).

## Infinitival to

1) I like that John can pick up his own dry-cleaning.

2) I'd like for John to pick up his own dry-cleaning.

*Spoiler:* FOR NOW, we will consider all of these (modals, auxiliaries, and *to*), to be category T. Before long, though, we will have reason to break this up. In fact, as it happens *none* of them will be exactly category T, but to see why we need to do some more.

## Complementizers

1) Pat will leave.

2) I heard that Pat will leave.

3) I wonder if Pat will leave.

4) I am anxious for Pat to leave.

It is perfectly possible to *embed* a sentence inside another one. When we do this, it is indicated with a *complementizer* (introducing a *complement clause*). Category: [C]. (maybe [-N,-V,+Functional]? Though the fourth is elusive.)

## The P for v. the C for

*For* is of course a preposition (*I looked for you for three hours*), but not when it is introducing clauses.

- He headed right for the back row.
- \*He'd like right for the class to be over.
- \*He expressed interest in the class to be over.
- Who would you vote for in the election?
- \*Who are you anxious for to win the election?

## The D that v. the C that

Same kind of thing holds for *that*.

1) I liked that movie.

2) I heard that movie involved guinea pigs.

Sometimes you can replace *for* clauses with *that* clauses.

3) It is important that Pat votes.

4) It is important for Pat to vote.

## Regrouping

Lexical categories:

N: noun, V: verb,

A: adjective, P: preposition

We started a feature decomposition of these by proposing that they are labels for feature bundles like  $[\pm N, \pm V]$ , which can characterize certain natural classes across categories.

## Regrouping

But there are many more than four categories, though some of these might be subcategories.

Aux: auxiliary, C: complementizer, Adv: adverb, D: determiner, PRN: pronoun, T: modals?, ...

So, we would need more features to make all of the distinctions. We won't pursue that, however—we'll just use the labels like N, V, A, P, D, T, C, etc.)