CAS LX 522
Syntax I

Week 1b.
Morphosyntactic features
(2.1-2.4.1)

Where is English?

- When we speak of "English," what are we referring to?
- Every native speaker has a complete knowledge system of their language.
- As far as the grammar is concerned, it's all part of a native speaker's cognitive makeup. (Vocabulary is a different thing...)

I-language, E-language

- The notion of "English" is really an *external* notion. It's kind of an "average" of the properties of the (nearly identical) knowledge systems that the individual speakers in the community have.
- What we're interested in here, in a sense, are the properties of a single speaker's knowledge of language. We might call it "English" if that speaker is part of the "English" speech community. But it's really an individual's knowledge. It's just that the community by and large has the same knowledge.

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$_{[3.sg]}$ have$_{[3.sg]}$ said that she was ill
   'He said that she was ill.'
2) Ils ont dit qu'elle était malade
   they$_{[3.pl]}$ have$_{[3.pl]}$ 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 actual distinctions, but it also predicts other distinctions that don't exist.

[plural]

- So, we have a new set of observations, now including the fact that there are just two classes.
- 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.

Privative features

- There are (at least) two ways we can characterize features.
- Above, we did it one way—the feature [pl] is there on plurals, and not there on singulars. This kind of feature—which is either *there* or *not there*—is a *privative* feature.

Binary features

- We could also view a feature as having *values*.
- A *binary valued* feature could have either of two values. Usually "+" and "–".
- On this view, plurals have [+plural], and singulars have [–plural].

Which is the right way to think of features?

- We don't know from the outset which view is the best for describing syntax, we want to choose the one that best captures the generalizations we see.
- The two views *are not* indistinguishable. They *do* make different predictions. Specifically, about what syntax can "see."

The Hopi dual

1) Pam taaqa wari
   that man ran[sg]
   'That man ran.'
2) Puma taʔtaqt yuʔtu
   those man[pl] ran[pl]
   'Those men ran.'
3) Puma taaqat wari
   those man[pl] ran[sg]
   'Those two men 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 should predict the existence of a *fourth* number: singular, plural, dual, and neither singular nor plural.

The fourth number?

- There doesn't, however, seem to be a fourth number—across languages. There's really just the three kinds: singular, plural, and dual.
- Adger tells a story at this point: There is an additional constraint that every noun needs to have *some* number feature. I want to come back to this in a little while.

General structure of the account

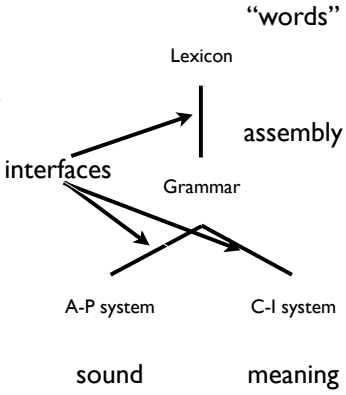
- Knowing a language is:
  i) knowing the "words"
  ii) knowing how to put them together
  iii) knowing how to pronounce them
  iv) knowing what they mean in combination

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*.



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.

Points of tension

- For English, it seems that independent [sg] and [pl] features is more complicated than we need—it seems to overgenerate.
- In the broader picture, Language needs to allow for independent [sg] and [pl] features in order to accommodate duals in, e.g., Hopi.

Tension

- We need a hypothesis about what is different in languages with no dual (e.g., English).
- Adger's suggestion: All languages have singulars, but in languages without duals, singular is the *default*, the "number for nouns not specified for number." The feature [sg] is *not recorded* in the English lexicon: *book* [], *books* [pl].
- So languages can differ in whether they record [sg] in the lexicon.

What are the features?

- Hard to say. A universal set, some used in some languages, but not others? Learned?
- Some features seem not to exist, why?
- Okham's razor: keep theories as simple as possible. Here, we want to define the simplest set of features we can get away with and still explain the data.

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? []?

Privative? Or binary?

- There's something kind of uncomfortable about saying the prepositions simply *lack* category features.
- We can soothe ourselves somewhat by adopting binary category features instead of privative features.
- Same predictions, but more in line with our intuition about what "category" should be.

[±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:

1) untie, unfold, unwrap, unpack

2) unhappy, unfriendly, undead

*3) *uncity, *uncola, *unconvention*

*4) *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 emtpy box
   'The beautiful girl put the black cat in the empty box.'

Lexical and 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).†

†*Telegram* (n.): An ancient form of texting.

Lexical and functional

- Sentences are held together by little "function words" as well. These are the *functional* categories. We will discuss these more later.

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

Functional categories/ syntactic "glue"

- 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].
  - 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").

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.

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].

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 axious 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.

  1) It is important that Pat votes.

  2) 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 [±N, ±V], which can characterize certain natural classes across categories.

Regrouping

- But there are many more than four categories.
- 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.)

Feature grouping

- Features themselves seem to be grouped.
- And this is the way we'll think of them for much of the course.
- Consider: category.
- [±N] is a feature, [±V] is a feature. There must be others to handle D, T, C, etc. But *together* they constitute the syntactic category.

Feature grouping

- We can write this like [Cat: +N–V], and in fact we'll generally just write [N] as a shorthand for that.
- We'll see other groupings. E.g., Number could be [Number:sg,pl] (or dual). Although number will in fact be part of a larger grouping including gender and person.