

# **Introduction to Syntax**

## **Lecture Notes**

Guilherme Zeus Dantas e Moura  
[gdantasemo@haverford.edu](mailto:gdantasemo@haverford.edu)

Haverford College — Fall 2021  
Last updated: September 8, 2021

This is Haverford College's undergraduate LING H113, instructed by Amanda Payne. All errors are my responsibility.

Use these notes only as a guide. There is a non-trivial chance that some things here are wrong or incomplete.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	What do we study in syntax?	4
1.2	‘The grammar’	5
1.2.1	The Wug Test	5
1.3	Areas of Linguistics	6
<b>2</b>	<b>Syntax</b>	<b>8</b>
2.1	Syntactic primitives	9

# 1 Introduction

## 1.1 What do we study in syntax?

Syntax is all about sentence *structure*. Syntax is about *form*, and semantics is about meaning. This has a few implications.

A *sentence* is just a string of words. It doesn't necessarily have a meaning assigned to it. (A *proposition* is a statement that has meaning and can be true or false.) We often represent sentences with written text for ease of analysis, but you can study the syntax of spoken or signed language.

(1) It is raining today.

(1) is an example of a sentence in English. Most English speakers agree that (1) is *grammatical*. That means it follows the rules of English syntax. We haven't talked about what exactly those rules are yet, but if you are an English speaker, you have some intuitions about what is grammatical and what is not. This data — sentences and whether they are grammatical or not — is our subject of study as syntacticians. Sometimes we can generate the data ourselves! But sometimes we can't, or shouldn't.

(2) Íbsaan dháabata autóbusaa jira.

(2) is a sentence from the language Oromo. It is also grammatical.

Some sentences are *grammatical* — they have a form that follows the rules — but considered unacceptable. This is a major source of confusion for our data! Here are some such sentences:

(3) That person is a real \*\*\*\*\*. (Slurs are socially unacceptable, but typically grammatical.)

(4) #My toothbrush sneezed. (Semantically unacceptable!)

Finally, grammar in linguistics is *descriptive*, not *prescriptive*. That means if a sentence would be naturally uttered by a speaker of a language, the sentence is considered grammatical for our data! This includes things like:

(5) What should I talk about? (Ending a sentence with a preposition happens all the time in speech.)

(6) Who is this for? (Using 'who' where you might have been taught to use 'whom'.)

Ungrammatical sentences are marked with \*. In-between or ‘marginally grammatical’ sentences are marked with a ?. Sentences that are semantically odd are marked with #.

- (7) \*Ungrammatical is sentence like this.

Grammaticality is not binary. Speakers of the same language can have different grammars, and therefore different judgments. Sometimes you might not even be sure if a given sentence is grammatical for you or not!

2021-09-01

## 1.2 ‘The grammar’

*Grammar* in linguistics is what we call *descriptive*, not *prescriptive*. That means if a sentence would be naturally said by a speaker of a language, the sentence is considered *grammatical* for our data. This includes:

- (8) What should I talk about? (Ending a sentence with a proposition happens all the time in speech.)
- (9) Who is this for? (Using ‘who’ where you might have been taught to use ‘whom’.)

Judgments about grammaticality are usually innate. That means you can judge if a word or sentence “sounds good” in your language without any formal education or training. You don’t need to know how to read or write to have linguistic judgments. In fact, very young children also have intuitions about language.

### 1.2.1 The Wug Test

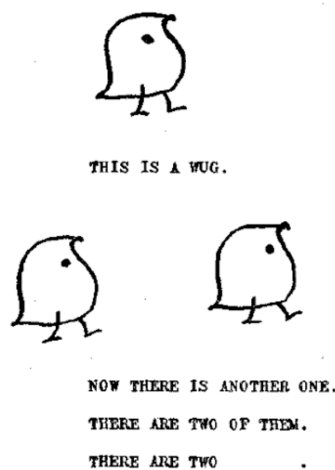


Figure 1.1: The Wug Test (Berko, 1958)

The Wug test is checking the ability to generate a regular plural — which is the suffix (-s) in English — so this is in the realm of morphology and not necessarily syntax — although there is a gray area between morphology and syntax, since how we define a single ‘word’ is not super clear.

However, language users have intuitions in all areas of the grammar — which will serve as a handy overview of generative linguistics.

Aside: What is generative linguistics? This is a particular approach to studying language which assumes humans have a *generative capacity*, i.e., the ability to generate linguistic forms, using some internal set of rules. This is often contrasted with *functionalist* approaches, which care more about the function of linguistic elements rather than the generative algorithm or set of rules used to produce them. These two approaches are actually compatible in most cases, but many colleges/linguists fall into one “camp” and might malign the other — if they think about them at all. We will be looking at things through a generative lens in this class, but that doesn’t mean it’s the only way!

### 1.3 Areas of Linguistics

**Phonology.** The study of sound patters — or patters in parameters of signs, for signed languages.

(10) blick

(11) ?bnick

(12) \*bfick

**Morphology.** The study of word-internal structure — like prefixes, suffixes, in-freakin’-fixes, etc. Not so different from syntax.

(13) run

(14) runs

(15) run-(n)er

(16) \*run-s-er

**Syntax.** The study of sentence structure — like nouns, verbs, etc. More on this later.

**Semantics.** The study of meaning as encoded in language.

(17) I just stopped smoking. Now my lungs are feeling fresh.

(18) I just stopped smoking. #But in fact, I have never smoked.

(19) #This jar is green and not a jar.

## 1 Introduction

**Pragmatics.** The study of meaning in context — social context, language context, cultural context.

(20) *Don't tell your friends about your indigestion, "how are you?" is a greeting, not a question.*

**Others.** Of course there are many other 'named' areas of study in linguistics: socioling, anthroling, acquisition, compling, neuroling, psycholing, internet ling, phonetics, conversation analysis, philosophy of language, etc. — but they are often overlapping/combined with the study of those various areas of the grammar listed above.

## 2 Syntax

Syntax is about how sentences are put together. In order to actually talk about the structure of a sentence, we need some categories, because there are too many words to list out individual rules. For example:

- [Noun Verb] makes a fine sentence.
  - (1) Cats talk.
  - (2) Cats walk.
  - (3) Dogs talk.
  - (4) Dogs walk.
- But [Verb Noun] doesn't really (unless these are commands).
  - (5) \*Talk cats.
  - (6) \*Walk cats.
  - (7) \*Talk dogs.
  - (8) \*Walk dogs.

So rather than listing  $\{\text{Cats, Dogs, ...}\} + \{\text{Walk, Talk, ...}\} = \text{English sentence}$ , we can use the shorthands “Noun” and “Verb”.

2021-09-08

However, this rule *overgenerates*, i.e., it produces sentences which are considered ungrammatical — it produces too many sentences. For instance, (9) and (10) are produced by this rule that we might want to exclude from our grammar.

- (9) \*Dog run.
- (10) \*Dogs runs.

Of course, if we said our entire grammar was produced this rule, that would be a massive *undergeneration*, since there are *many* grammatical sentences of English which aren't generated by it, like (11) and (12).

- (11) Happy dogs run.
- (12) Dogs always run.



## 2.1 Syntactic primitives

If we take the rule as our starting point, it's clear that we have two types of changes we need to make: we somehow need to introduce either more specific sub-types of nouns and verbs, to rule out things like (9) and (10), and we need to add more rules, to cover things like (11) and (12). Then, given more data, we might need to refine those new rules, too. This is a lot of what syntax is: coming up with rules, looking at data to see if the rules generate the data, and if not, changing the rules and starting again! In order to come up with these rules, we need some *syntactic primitives*: basic categories of words that we can all agree on. So let's try it.

Now, upon combining our lists, we've probably got a lot of familiar terms like Nouns, Verbs, Adjectives, Adverbs, Prepositions, Determiners — this is a solid foundation! We use these terms in syntax often, as they are extremely common across languages and can roughly be described in terms of their distribution in addition to their function. For instance, in English, *nouns* are the types of things that can be pluralized, or the things that can follow a determiner. They often represent 'people, places, and things' — but not always. So rather than a meaning-based category, we want a distribution-based category.

You might have come up with a bunch of other terms that we may or may not end up using for our rules in this class — things like conjunctions, particles, participles, or interjections. We will certainly talk about them as they become relevant to describing data as we see.

Finally, two noun-type things that we ought to consider specially: *pronouns* and *proper names*. Are these distinct syntactic categories from *noun*? We have special names for them, but do they have an identical distribution to nouns like *cat(s)*, *dogs(s)*? What can we do with these?