جامعة نيويورك أبوظبي

NYU | ABU DHABI

PSYCH-UH 2218: Language Science

Class 8: Morphemes and Structure Building Rules

Prof. Jon Sprouse
Psychology

Morphology: shape and meaning

Words are (at least) a pairing of **sound** and **meaning:**

sound	**meaning**
[k æ t]	

Much like phonology is driven by puzzles (e.g., regularities in the distribution of sounds), **morphology** is driven by puzzles about the relationship between the meaning and the shape of words. By shape, we simply mean the sequence of phonemes that make up the word.

Morphology is the study of the shape of words.

The pairings are (mostly) arbitrary

For simple words (we will get better at defining this later), the pairing between sound and meaning is arbitrary. There is no reason why the meaning cat is paired with the sound cat in English. We can see this by looking at all of the different sounds that are paired with this meaning in different languages:

language	word
arabic	besseh
ethiopian	domadh
farsi	gorbeh
gaelic	piscin
hawaiian	popoki
icelandic	köttur
korean	koyangi
mayan	miz
swahili	paka
tamil	poonai
thai	maa-oh
tsalagi	we’sa

But some pairings are systematic

Let’s take a look at a classic puzzle in morphology. We can use this puzzle to motivate quite a bit of the theory of morphology:

What does this word mean?

unlockable

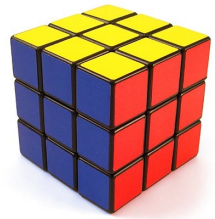
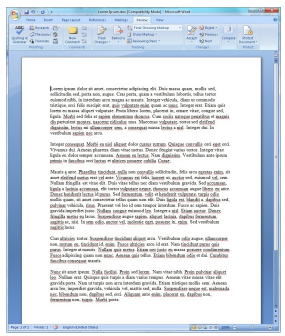
Meaning 1: Can’t be locked

Meaning 2: Can be unlocked

Why are there two meanings for this word? Why isn’t there one, or more than two? And why are they the meanings that they are?

But some pairings are systematic

Every word that has the form un-X-able, where X is a reversible verb, seems to have two meanings (no more, no less). And those meanings always seem to be the same two meanings:

	undoable	unfoldable	unlearnable
Option 1: Can't be X'ed			$\int_a^b f(x)dx$
Option 2: Can be un-X'ed			

This seems like something more than a coincidence. Morphology wants to find a way to explain this.

Insight 1: These words seem to be made of smaller parts

A theory of the parts of words.

Insight 1: These words seem to be made of smaller parts

If you look at all of these words together, a pattern emerges. They all seem to be built from the same two parts (un, able) and a different verb in the middle.

un-lock-able

un-do-able

This should spark an idea. If word meanings are built from their parts, then the similarities between these words can be explained: they are similar because they use the same parts (2/3!).

un-fold-able

un-learn-able

Compositional: In linguistics, we say that the meaning of a string is compositional if the meaning can be derived from the independent meanings of the parts (if the meanings are composed of the meanings of the smaller parts).

Compositionality is another great example of structure in the mind. We interpret meanings from smaller pieces, but don't ever realize it!

Defining the parts: Morphemes

Some words are clearly one piece:

fierce
desk
boot
at

These words are complete units. There are no sub-parts that have meaning outside of the word. So these are each clearly a separate lexeme. For example, the **erc** in **fierce** does not have an independent meaning.

However, if we start looking around we see that lots of words are made up of smaller pieces; and that those pieces seem to have regular meanings:

desk**s**
pencil**s**
boot**s**

The **s** in these words seems to mean something like **multiple**: we can call it **plural**.

jump**ed**
fail**ed**
label**ed**

Similarly, the **ed** in these words seems to mean something like **in the past**: we can call it **past tense**

preset
prepay
preboard

The **pre** in these words seems to mean something like **before**

Defining the parts: Morphemes

Morpheme: The smallest unit of language that carries a distinct meaning.

Some words are only a single morpheme:

fierce
desk
boot
at

Each of these words is a single morpheme - there is one unit in the word that carries meaning: the entire word itself

Some words contain two morphemes:

desks
pencils
boots

The **s** in these words is a morpheme.

The rest of the word is a morpheme too!

preset
prepay
preboard

The **pre** in these words is a morpheme.

The rest of the word is a morpheme too!

Bound vs Free morphemes

It is possible to investigate all of the types of morphemes in a language, and develop a theory of the types of morphemes. That theory is called a theory of **morphology** (the shape of words).

fierce
desk
boot
at

Morphemes that can be a stand-alone word are called **free morphemes**

---

desks
pencils
boots

Morphemes that only occur attached to a free morpheme are called **bound morphemes**

jumped
failed
labeled

preset
prepay
preboard

Practice identifying morphemes

I know it is not the most exciting task, but let's try to identify some morphemes. This is the basic step for studying morphology.

The process for identifying morphemes requires thinking about other words that might contain the morpheme, and asking yourself if there is a shared meaning for that potential morpheme across the words that share it.

word	**morphemes**	**types**	**evidence of regularity**		
happiness	happy-ness	1 free 1 bound	happy	sadness stillness	—
innumerable	in-numer-able	3 bound	insecure indelicate	enumerate numerical	debatable laughable
reception	re-cept-ion	3 bound	rewind return	deception inception	intuition omission

The morpheme "cept" is tricky. It is not clear that it has a stable meaning. This is an example that motivates our more abstract definition of morpheme, and the idea that semantics are optional for morphemes. I'll mention more later!

Roots and Affixes

root: A morpheme that other affixes attach to. It typically (though not always) contributes the core lexical meaning of a complex word. A root can be free or bound. Because other morphemes attach to it, its position in a word can vary (first, middle, last).

affix: A morpheme that attaches to other morphemes. Affixes are always bound. Affixes can attach before (prefix), after (suffix), within (infix), or around (circumfix) other morphemes.

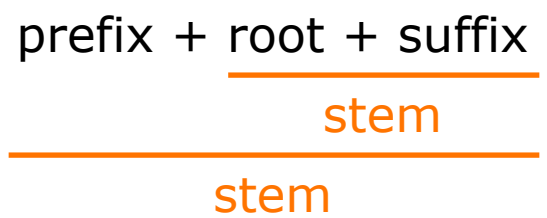
	prefix	**root**	**suffix**	free/bound?
happiness	-	happy	ness	free root
innumerable	in	numer	able	bound root

Another term: stem

root: A morpheme that other affixes attach to. It typically (though not always) contributes the core lexical meaning of a complex word. A root can be free or bound. Because other morphemes attach to it, its position in a word can vary (first, middle, last).

affix: A morpheme that attaches to other morphemes. Affixes are always bound. Affixes can attach before (prefix), after (suffix), within (infix), or around (circumfix) other morphemes.

stem: Sometimes you will see the word "stem". A stem is a root plus one or more affixes. (It is a metaphor with a plant.)



Roots and stems are sometimes in opposition because some affixes can only attach to roots, while others can attach to both roots and stems.

Types of affixation

In general, there are four types of affixes:

prefix: a bound morpheme that appears before the root or stem

preset
repay

suffix: a bound morpheme that appears after the root or stem

jumped
failing

infix: a bound morpheme that appears inside of a root

Bontoc (Philippines) uses an infix to distinguish nouns and verbs:

fikas ‘strong’
kilad ‘red

fumikas ‘to be strong’
kumilad ‘to be red

circumfix: a bound morpheme that appears around a root (or stem)

Chickasaw (Native American) uses a circumfix for negation:

chokma ‘he is good’
lakna ‘it is yellow’

ikchokmo ‘he is not good’
iklakno ‘it is not yellow’

A pseudo-infix in English

English has something similar to an infix. In US English it is the "F-word". In UK English it also includes the "b-word" based on the word "blood". I'll use the more polite version here in class. The process seems to still work:

Let's take a look at some examples:

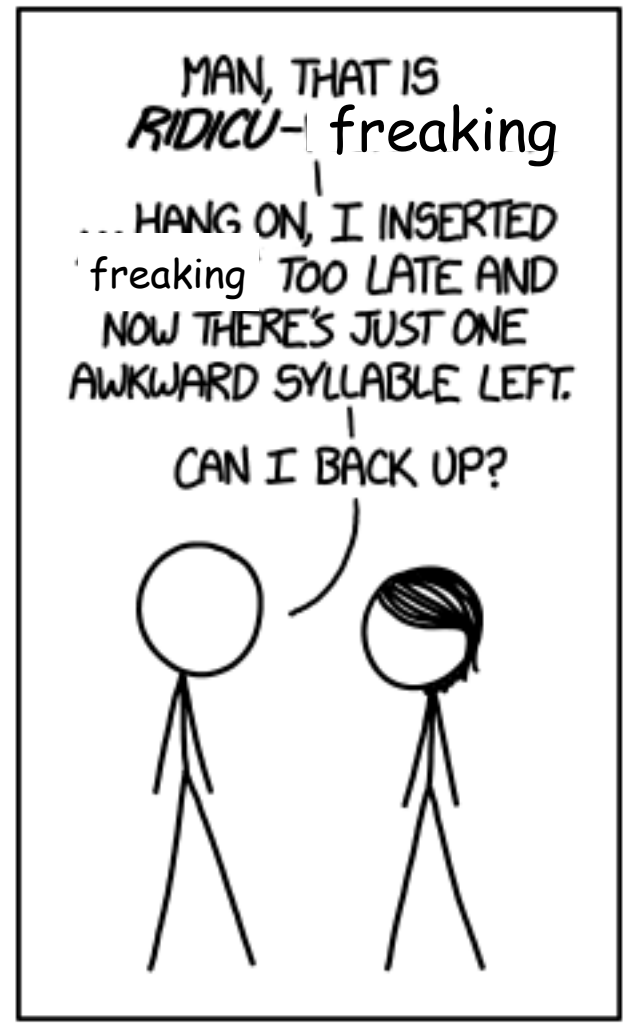
fantastic → fanfreakingtastic

absolutely → absofreakinglutely

What is interesting about this is that there is a rule for where in the word you can place "freaking".

*fantasfreakingtic *absolutefreakingly

What do you think the rule is?

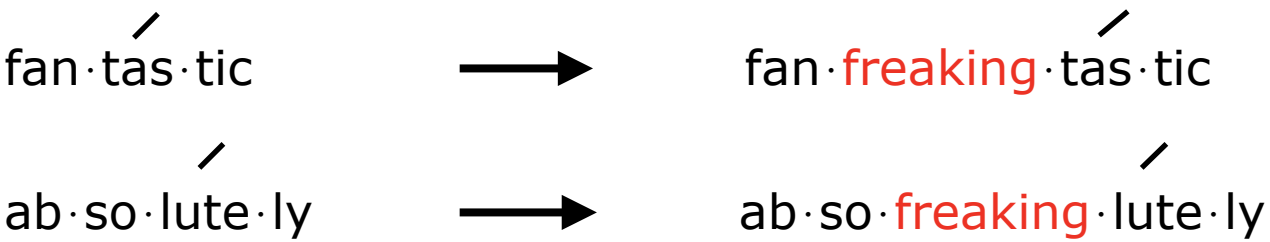


http://xkcd.com/1290/

The rule for inserting “freaking"

The rule is based on **word stress**. Word stress is the extra acoustic prominence that we give to certain syllables inside of words.

The “freaking” insertion rule: The word “freaking” can only be inserted in the position immediately before the primary stressed syllable.



What is so interesting about this is that there is no chance that anyone ever explicitly taught English speakers this rule. That is true for all phonological and morphological rules, but this one is much clearer because it is about an obscenity. Parents and teachers typically don’t teach about obscenities. So this is something that English speakers learn as part of language acquisition!

Here’s a question - Why did I call **freaking** a “pseudo” infix?

Compounding: combining two roots

Compound words are words that are composed of two (or more) roots

Compounds can involve almost any type of root, but as a quick example, we can look at noun-noun compounds:

Novel compounds are compounds that you make up on the fly. They tend to have a fully compositional meaning. They also tend to have two (or more) possible meanings:

cookie chair	A chair made of cookies / shaped like a cookie. A chair for (eating?) cookies.

Lexicalized compounds are compounds that have become stored in memory as complete units. They tend to be written as a single unit without a space. They tend to have one meaning, which may not be completely compositional any longer:

teacup cupcake flagship

Compounding: combining two roots

Compound words are words that are composed of two (or more) roots

Languages vary on which roots can participate in compounds. English is particularly permissive when it comes to compounds — it allows almost any type root to combine with any other. Your language may be more restrictive about the kinds of roots that can participate in compounding.

N N	N V	N A
dog bed	brainwash	stone cold
V N	V V	V A
think tank	blow dry	feel-good
A N	A V	A A
greenhouse	slow dance	blue-green
P N	P V	P A
backseat	downvote	overblown

Notice that orthography (spelling) does not tell you whether a word is a compound or not. Sometimes they are spelled as one word, sometimes with a space, and sometimes with a dash. You have to use your theory to identify them.

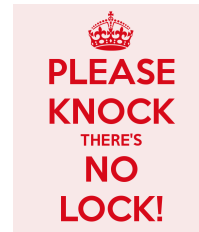
Insight 2: These words are ambiguous

A theory of the rules that build complex words

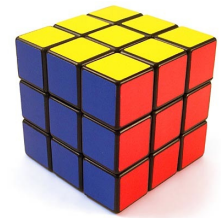
Insight 2: These words are ambiguous

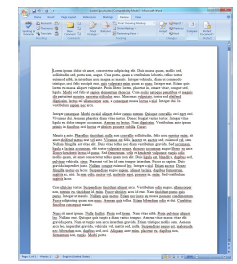
Ambiguity: In linguistics, we say that the meaning of a string is ambiguous if there is more than one possible meaning.

un-lock-able

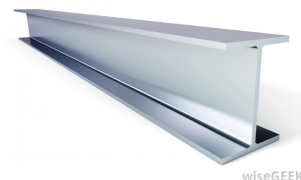


un-do-able





un-fold-able

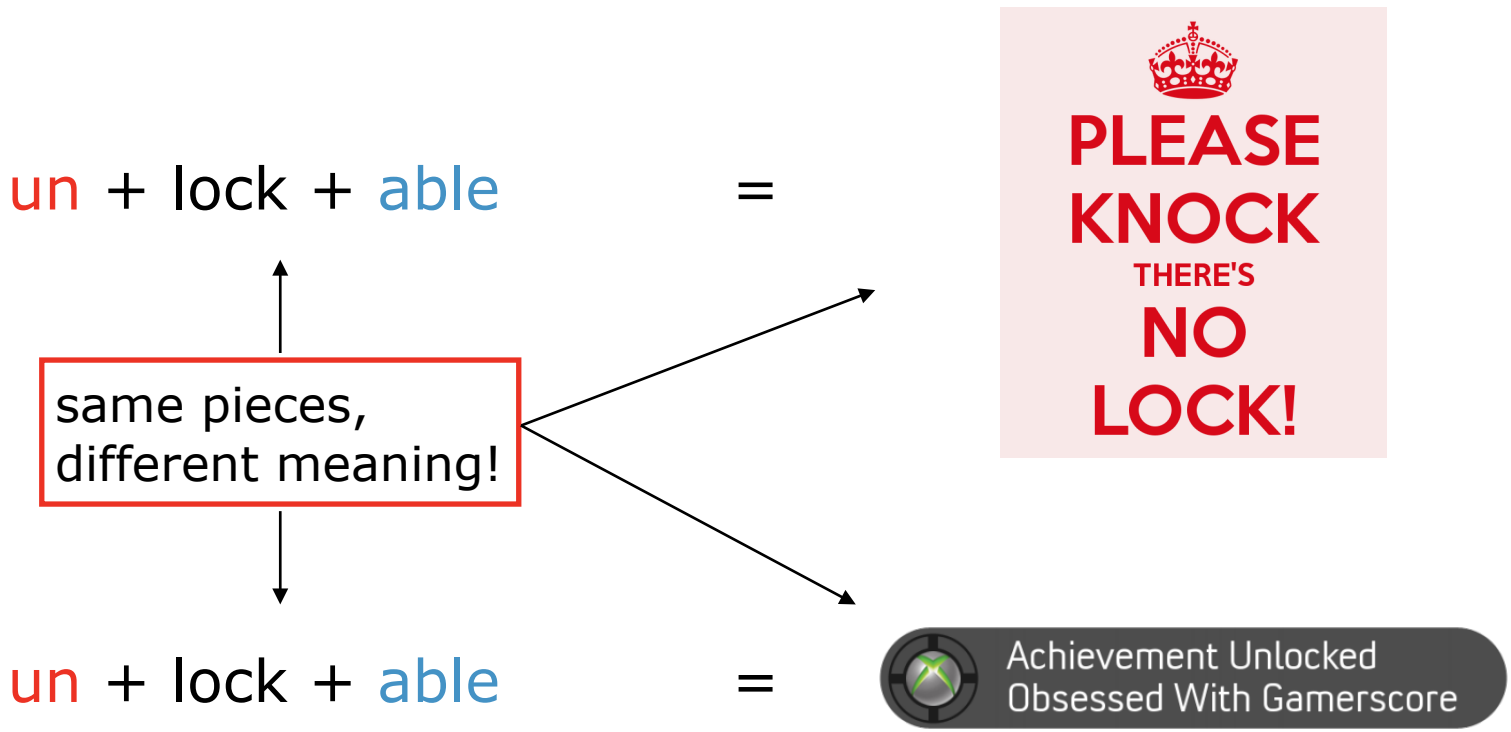


un-learn-able

$$\int_a^b f(x)dx$$

The puzzle of ambiguity

Ambiguity raises a real puzzle for compositionality: How is it that two meanings can come from the same pieces?

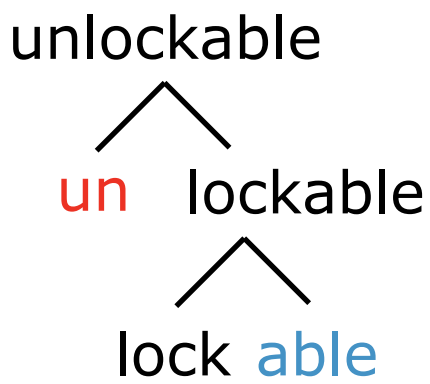


If meaning comes from the parts (compositionality), then when we have the same parts, we should get the same meaning, right???

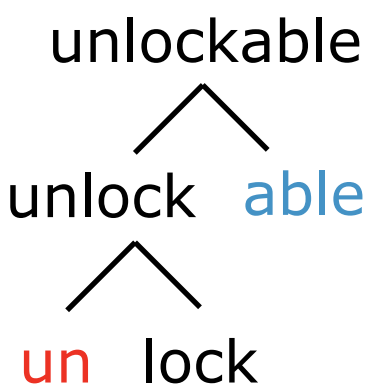
Ambiguity through hierarchical structure

Instead of throwing out compositionally, we save it by saying that complex words are compositional, and that the word has **hierarchical structure.** It is a difference in the structure that leads to a difference in the meaning!

Meaning 1: lock+able, then un + lockable



Meaning 2: un+lock, then unlock + able



Hierarchical structure: smaller units are combined to form larger units.

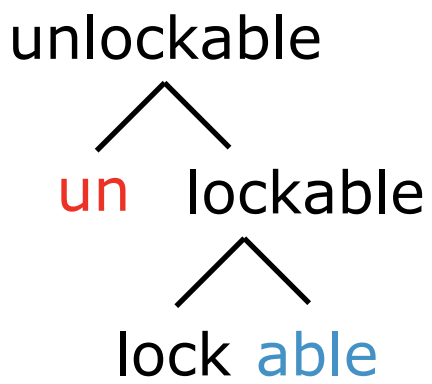
We can use **trees** to demonstrate the hierarchical structure.

Two items that combine are linked with two lines that converge into a node. We label that node in order to show that they formed a new unit with certain properties.

Ambiguity through hierarchical structure

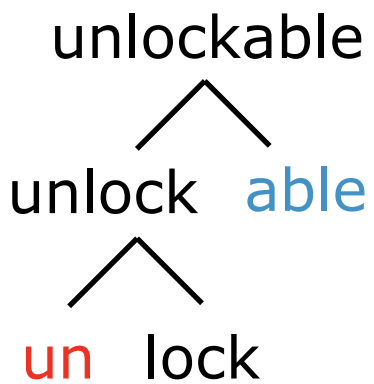
Instead of throwing out compositionally, we save it by saying that complex words are compositional, and that the word has **hierarchical structure.** It is a difference in the structure that leads to a difference in the meaning!

Meaning 1: lock+able, then un + lockable



=

Meaning 2: un+lock, then unlock + able

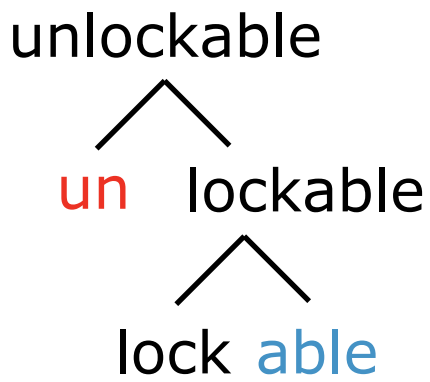


=

Structure-building rules

We can capture the hierarchical structures of words with structure-building rules. These rules combine two objects together to yield a third (larger) object:

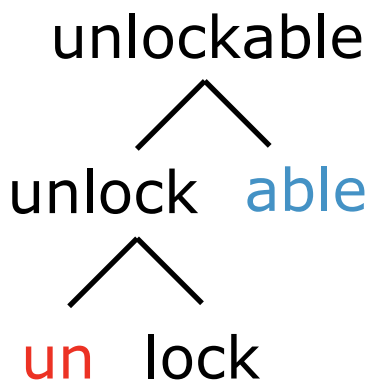
Meaning 1: lock+able, then un + lockable



un + lockable → unlockable

lock + able → lockable

Meaning 2: un+lock, then unlock + able



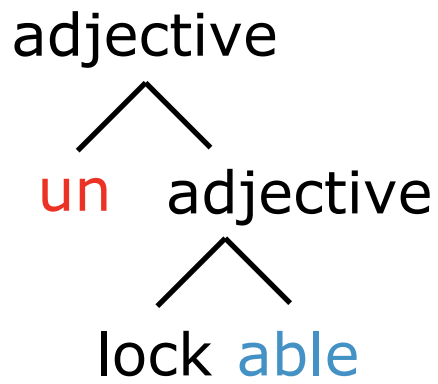
unlock + able → unlockable

un + lock → unlock

Making the rules more general

These rules are specific to the word unlockable. But we saw earlier that this ambiguity is true of other words (undoable, unlearnable, etc). We can make the rules more general by replacing the non-affixes with parts of speech:

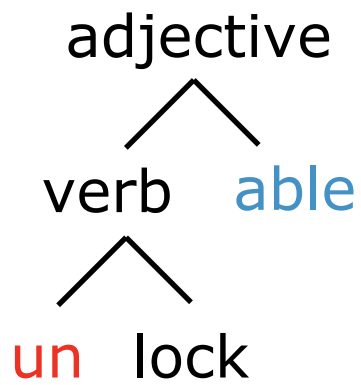
Meaning 1: lock+able, then un + lockable



un + ADJ → ADJ

VERB + able → ADJ

Meaning 2: un+lock, then unlock + able



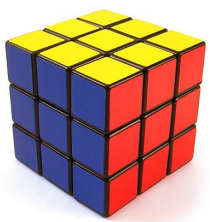
VERB + able → ADJ

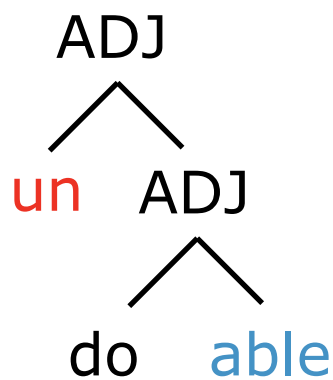
un + VERB → VERB

Applying the rules to other words

To see that this general form works, let's try other words. First, let's try the word **undoable**.

Meaning 1: do+able, then un + doable

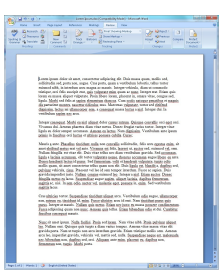


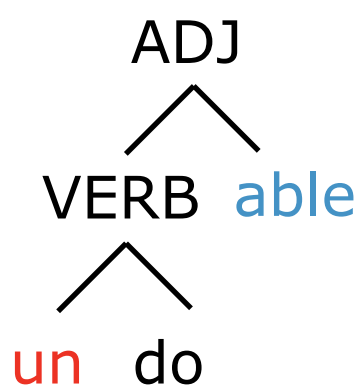


un + ADJ → ADJ

VERB + able → ADJ

Meaning 2: un+do, then undo + able





VERB + able → ADJ

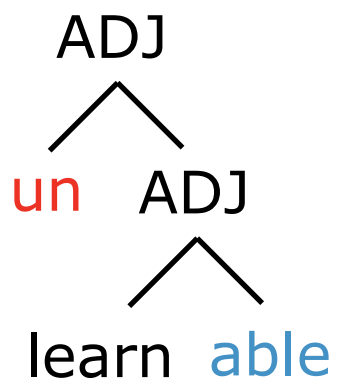
un + VERB → VERB

Applying the rules to other words

To see that this general form works, let's try other words. Next, let's try the word **unlearnable**:

Meaning 1: learn+able, then un + learnable

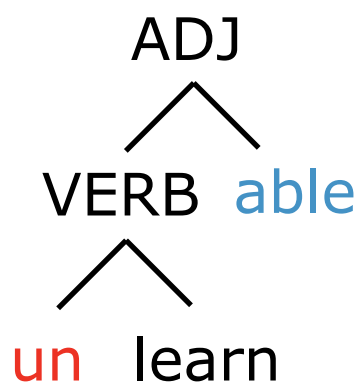
$$\int_a^b f(x)dx$$



un + ADJ → ADJ

VERB + able → ADJ

Meaning 2: un+do, then undo + able



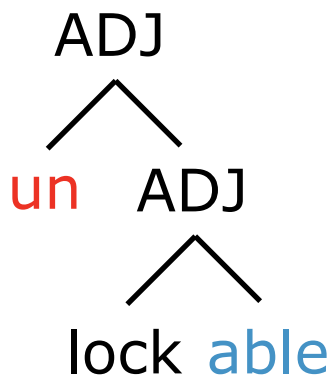
VERB + able → ADJ

un + VERB → VERB

The standard form of the rule

The format of the rules on the previous slides is very easy to understand. However, there is a more standard format for the rules that reverses the order of the left/right sides, and removes the plus sign.

Meaning 1: can't be locked

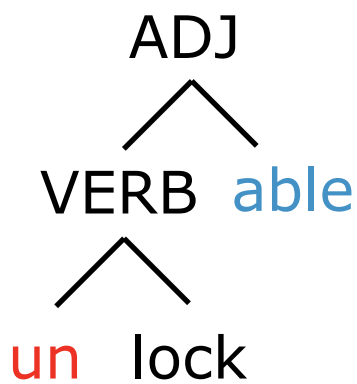


ADJ → un + ADJ

ADJ → VERB + able

In this format, we read the arrow as "rewrites to".

Meaning 2: can be unlocked



ADJ → VERB + able

VERB → un + VERB

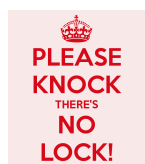
Hierarchical Structure and Ambiguity

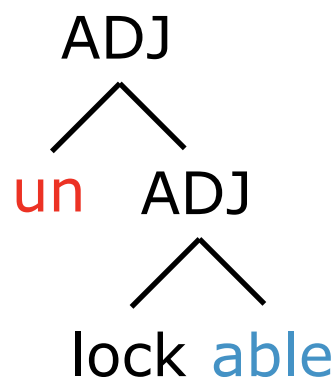
And here is the big payoff from structure-building rules. The two meanings come from two different hierarchical structures, which we get through the application of different rules. In this case, from three rules, we get two distinct meanings. Both meanings use the "able" rule. But they each use a different "un" rule, and use it in a different order.

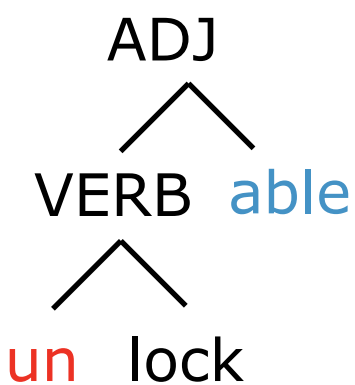
ADJ → un + ADJ

ADJ → VERB + able

VERB → un + VERB







Rethinking arbitrariness

The pairing of sound and meaning for **individual morphemes** is arbitrary. Our example "cat" is a single morpheme, so its pairing is arbitrary.

language	**word**
arabic	besseh
ethiopian	domadh
farsi	gorbeh
gaelic	piscin
hawaiian	popoki

But once that relationship is established, the presence of that morpheme in a **multi-morphemic** word will have **systematic effects due to compositionality**.

cat + s =

An application of a theory of morphemes

A real-world application of a theory of morphology

How many words do the “Eskimo” have for snow?

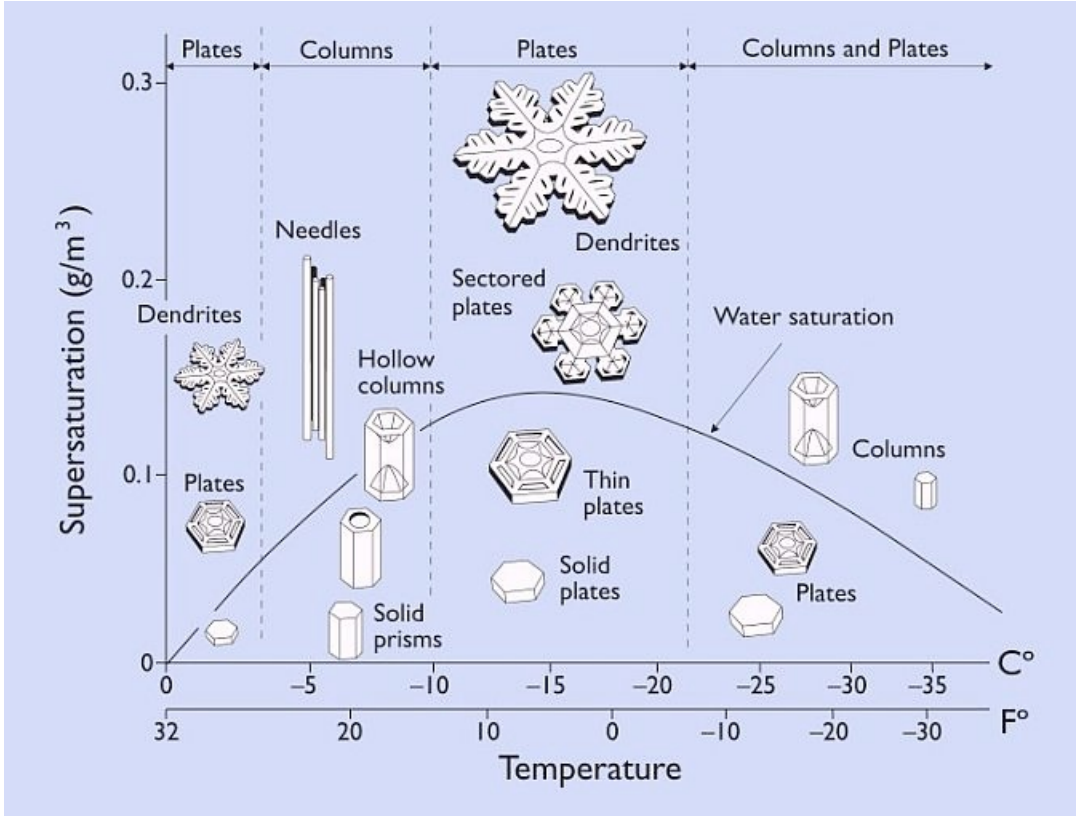
We should actually call them **speakers of Inuit-Yupik/Aluet languages.** Here I used the term “Eskimo” just to identify the saying/meme. Let’s say Inuit from now on!

First, let's confront our biases

Why do we think that Inuit-Yupik speakers would have lots of words for snow?

Do we think that they are excellent snow researchers, and therefore need a precise vocabulary for the different types of snow formations?

No, we don't. So this is not a comment on their scientific interest in snow. Could it be something negative?



First, let's confront our biases

Why do we think that Inuit-Yupik speakers would have lots of words for snow?

Or could it be a subtle form of language-oriented prejudice?

Language prejudice is something that we will discuss in more detail later in the semester. But for now, I want you to be aware that it exists, and show you a little bit about how we can apply our scientific theories of language to claims about languages that may have a root in prejudice.

Now let's try to answer the question for English...

How many words for snow are there in General American English?

snow

blizzard

flurry

You would all agree that these are different words for snow.

slush

sleet

avalanche

But what about these?

These are debatable. The problem with these is that their **meaning** is not very close to the "snow" **prototype**.

Now let's try to answer the question for English...

How many words for snow are there in General American English?

snow
blizzard
flurry

You would all agree that these are different words for snow.

slush
sleet
avalanche

But what about these?

These are debatable. The problem with these is that their **meaning** is not very close to the "snow" **prototype**.

snows
snowed
snowy
snowing

And what about these?

These certainly have the canonical "snow" meaning. But here we are seeing a complication imposed by language itself -- words can show up in **different forms**. We know this now as the affixation of morphemes to the root word snow. **Should we count these or not?**

So how many "words" for snow are there in Inuit ("Eskimo") languages?

There are around 12 distinct morphemes for snow-like phenomena, including both canonical snow (snow, blizzard, flurries), and related things (slush, sleet, etc). **This is not much more than English!**

The Inuit languages have about 280 different grammatical forms of each word (e.g., snows, snowed, snowing) that can be formed through affixation. **This is MUCH more than English, and is due to the grammatical properties of the languages.**

So there are two answers. If we only count distinct morphemes then the number is very close to English (~12). If we count the different forms that come from affixation, then then number is ridiculously large (>1000). But the large answer is simply a grammatical fact of the language, not an indicator of precision in discussing snow!

Some examples that complicate the idea of a morpheme.

Morphemes without meaning

Here is our definition of **morpheme:** The smallest unit of language that carries a distinct meaning.

We will keep this definition for this class, but I want you to know that in cutting edge language science, we need to look for ways to accommodate morphemes that don't appear to have a meaning.

Latinate roots:	**ject:**	reject, inject, eject, object, subject, …
	ceive:	receive, conceive, deceive, perceive, …
	tain:	obtain, detain, pertain, contain, maintain, …

False negatives:	unkempt	But is there kempt?
	disgruntled	But is there gruntled?
	inert	But is there ert?
	nonplussed	But is there plussed?
	ruthless	But is there ruth or ruthful?

Morphemes without meaning

Another famous one:

overwhelmed underwhelmed But is there whelmed?

And here is one that has changed meaning:

horror horrible horrific ← These all have the horror meaning.

terror terrible terrific ← But this one is positive!

All of these examples show that in some cases it is the full word that carries the meaning, not the individual morphemes. This is a puzzle for our theory of morphemes.

Morpheme boundaries can change

In some cases, the boundary is moved to a different spot:

	original	reanalysis	examples
hamburger:	hamburg-er	ham-burger	burger, cheesburger, etc.
alcoholic:	alcohol-ic	alco-holic	shopaholic, workaholic, etc.
helicopter:	helico-pter	heli-copter	helipad, quad-copter, etc.

And in other cases, the boundary is eliminated (these are called **blends**):

breakfast + lunch → brunch

smoke + fog → smog

motor + hotel → motel