جامعة نيويورك أبوظبي

NYU | ABU DHABI

PSYCH-UH 2218: Language Science

Class 19: Four case studies in language acquisition

Prof. Jon Sprouse
Psychology

Quick recap of the big challenge of language acquisition

The logical problem of language acquisition

Fact 1: Learning human language is learning the rules that generate a very large, probably infinite, set.

Fact 2: The evidence that children receive about those rules is finite.

Fact 3: In the absence of disease or abuse, all children succeed in learning language, and all succeed to the same degree.

Fact 4: Learning the rules that generate an infinite set from finite evidence alone requires both positive and negative evidence.

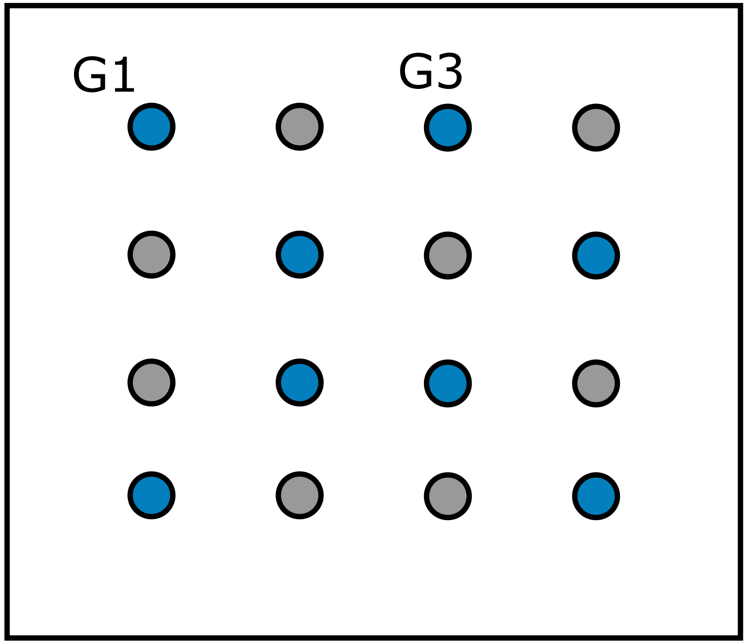
Fact 5: Children do not receive (or make use of) negative evidence.

Conclusion: Children must have some other mechanism that ensures that all children successfully learn language (the same way all learn to walk, see, etc).

Part of that endowment must be mechanisms that ensure successful acquisition

Again, this is obviously true. What is the point of a genetic endowment for an ability if it does not guarantee successful development of that ability?

But the logical problem of language acquisition makes it clear to us what those mechanisms must accomplish:



Hypothesis space

All learning requires a **hypothesis space** - the set of all possible hypotheses that the learner could entertain.

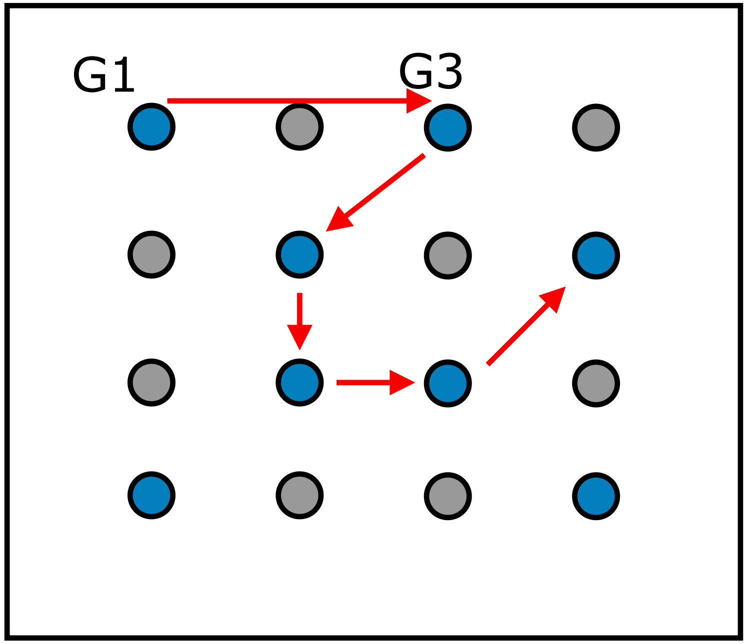
We can view this as a space of possible grammars that the child could hypothesize.

Part of the genetic endowment of language will be the fact that some grammars are possible and some are not. For example, a syntactic rule that doesn't follow X-bar theory is probably not possible. A phonology built on amplitude is probably not possible. This simplifies the task by reducing the number of hypotheses

Part of that endowment must be mechanisms that ensure successful acquisition

Again, this is obviously true. What is the point of a genetic endowment for an ability if it does not guarantee successful development of that ability?

But the logical problem of language acquisition makes it clear to us what those mechanisms must be:



Hypothesis space

All learning requires an algorithm for evaluating one hypothesis and adopting a new hypothesis based on evidence.

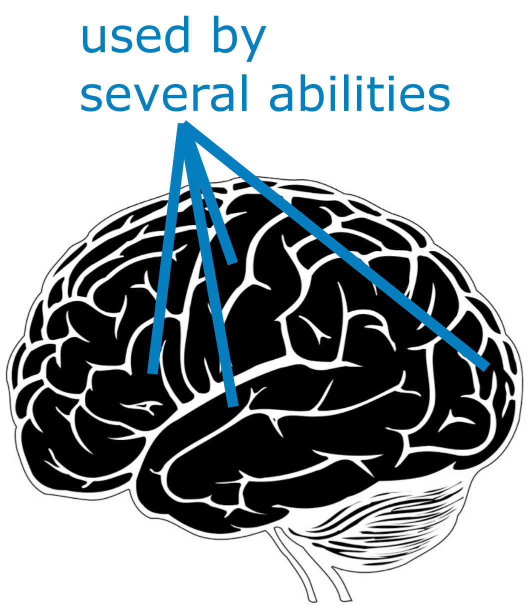
We can think of this as moving through the hypothesis space based on the evidence that children receive (positive evidence only).

Part of the genetic endowment for language will be a mechanism that prevents children from "getting stuck" when they use positive evidence, and ensures that all children succeed in language acquisition.

An extra dimension

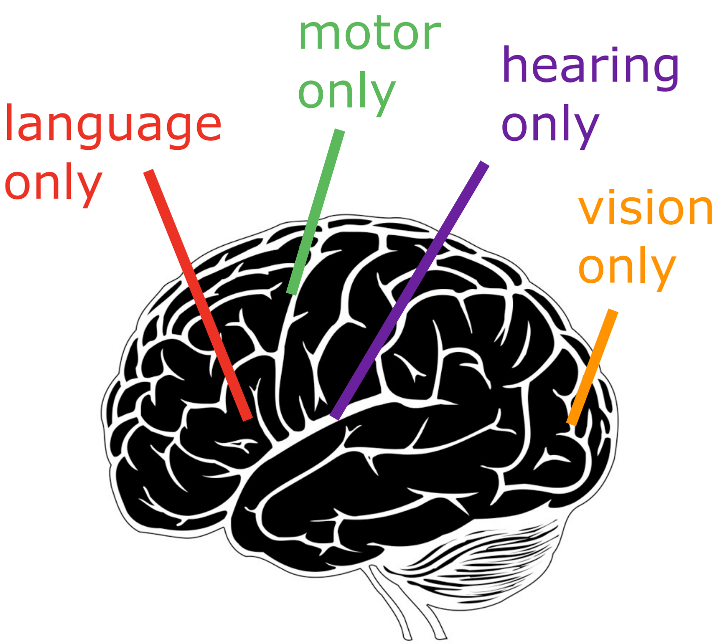
There are two possible types of mechanisms that could be part of our genetic endowment:

Domain-general mechanisms are used by multiple cognitive abilities.



Tracking probabilities may be domain-general.

Domain-specific mechanisms are used by one cognitive ability.



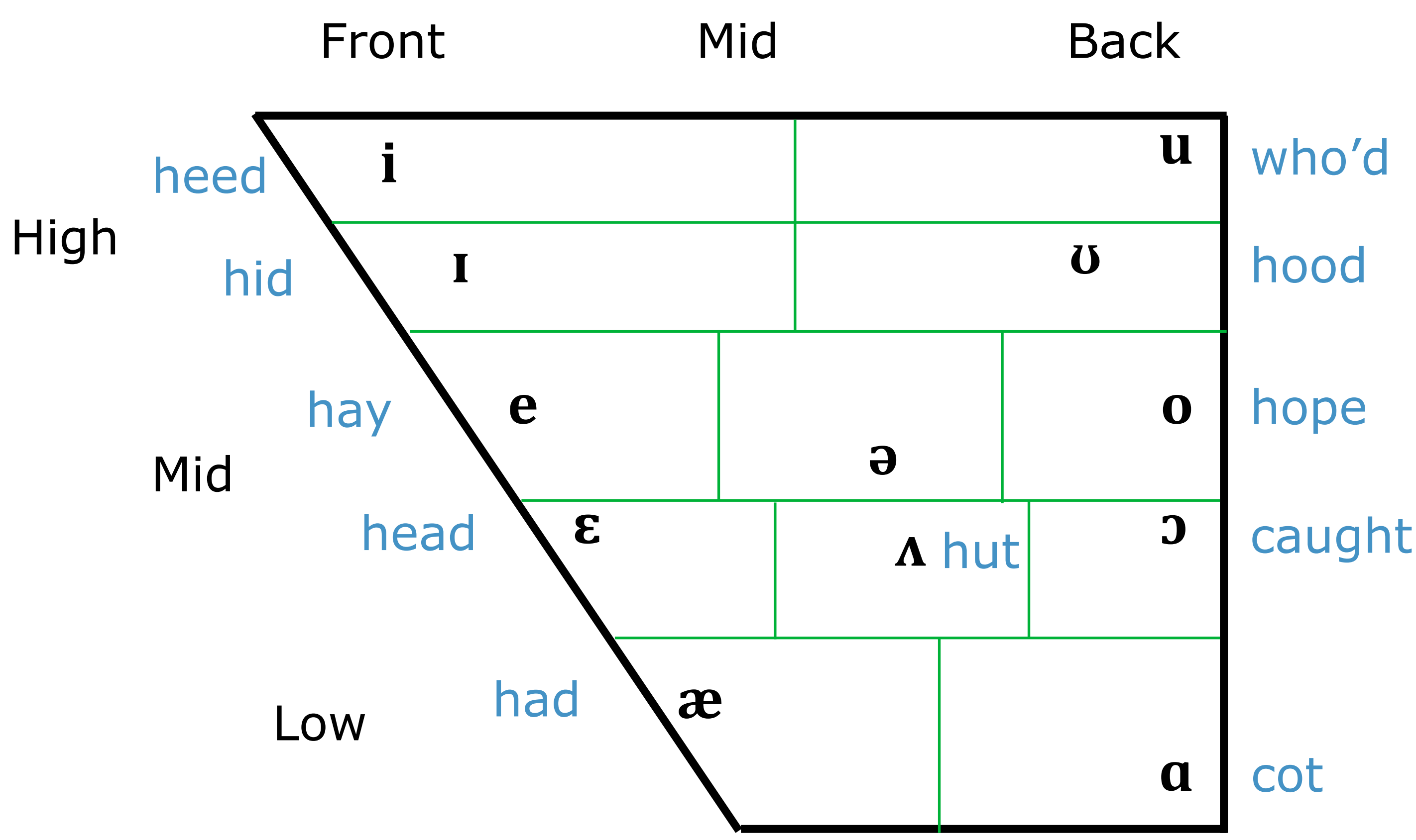
Something like X-bar theory might be domain-specific.

As we uncover potential mechanisms that are part of the genetic endowment, we can also ask whether they are domain-general or domain specific.

Learning phonemes

Phonemes are about breaks in a continuous space

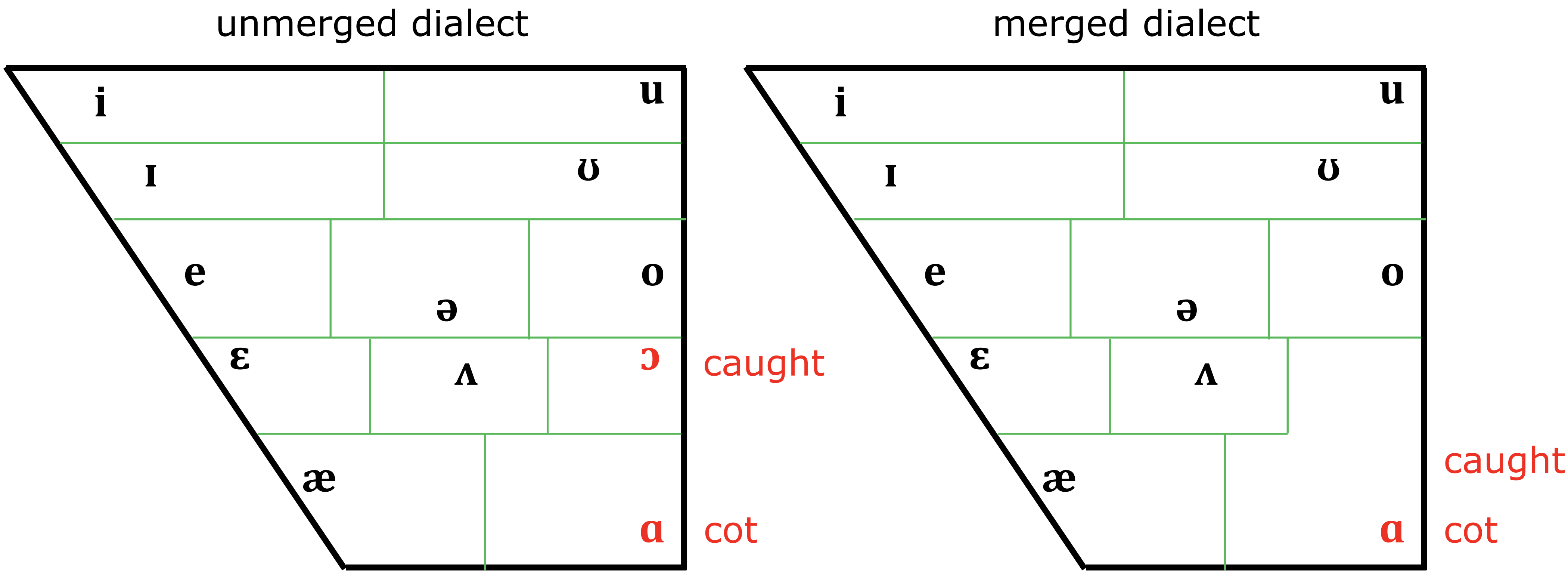
This is probably easiest to see with vowels. Vowels are just positions of the tongue inside of your mouth. This space is obviously continuous, like all spaces. When we say a language has certain vowels, we are really saying that a language has certain breaks in the vowel the space:



Differences between two vowel inventories

I will use two versions of US English to illustrate this because it is a minimal distinction. In one version, the words "caught" and "cot" are pronounced with distinct vowels. In the other, they are pronounced with the same vowel. (Or "dawn" and "don", or "hawk" and "hock".)

What this means in terms of space is that one version has a break that the other version does not have:



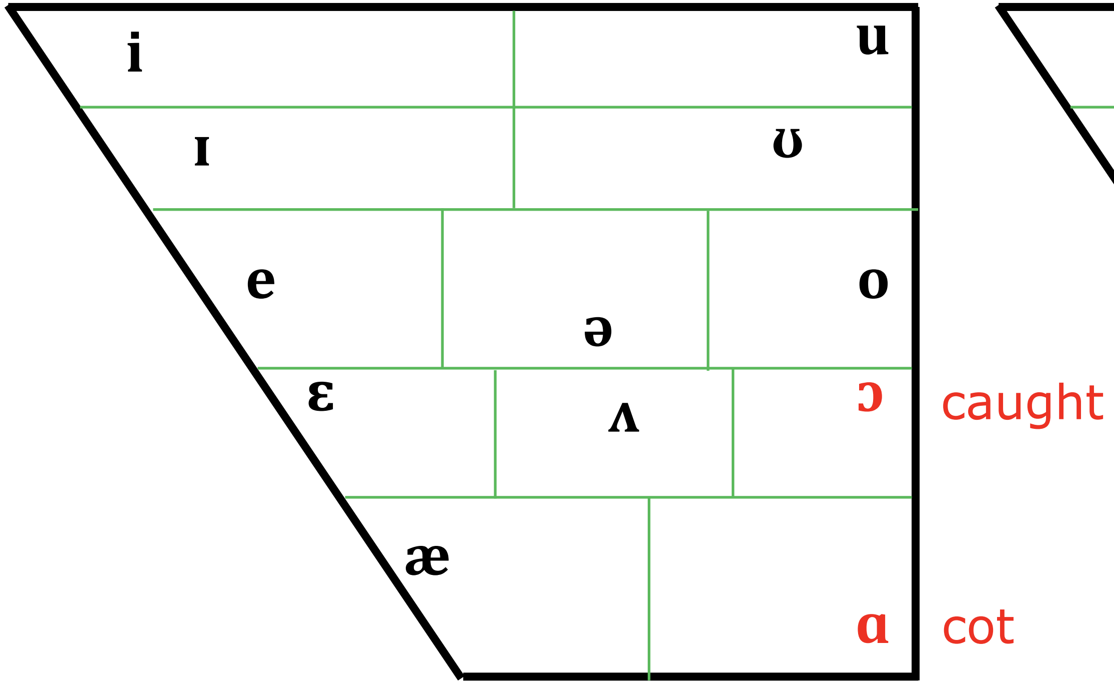
Children need to learn the breaks

Now, imagine a child trying to learn the vowel inventory of their language.

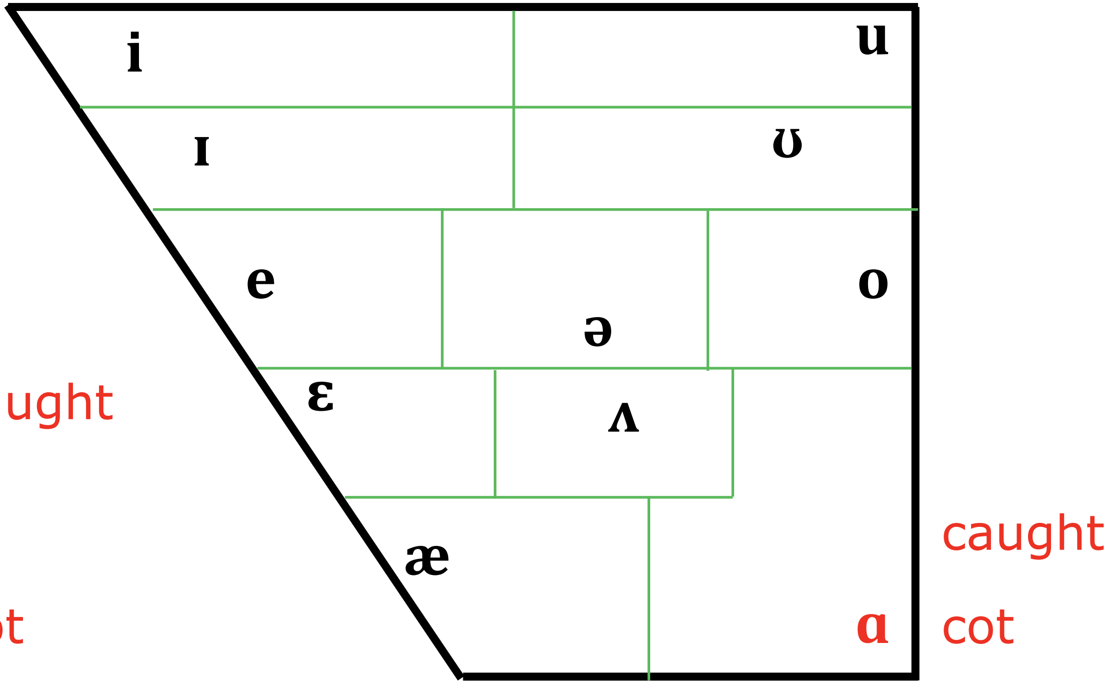
Viewed this way, what they need to learn is which breaks their language has.

How do you think they go about doing that?

unmerged dialect

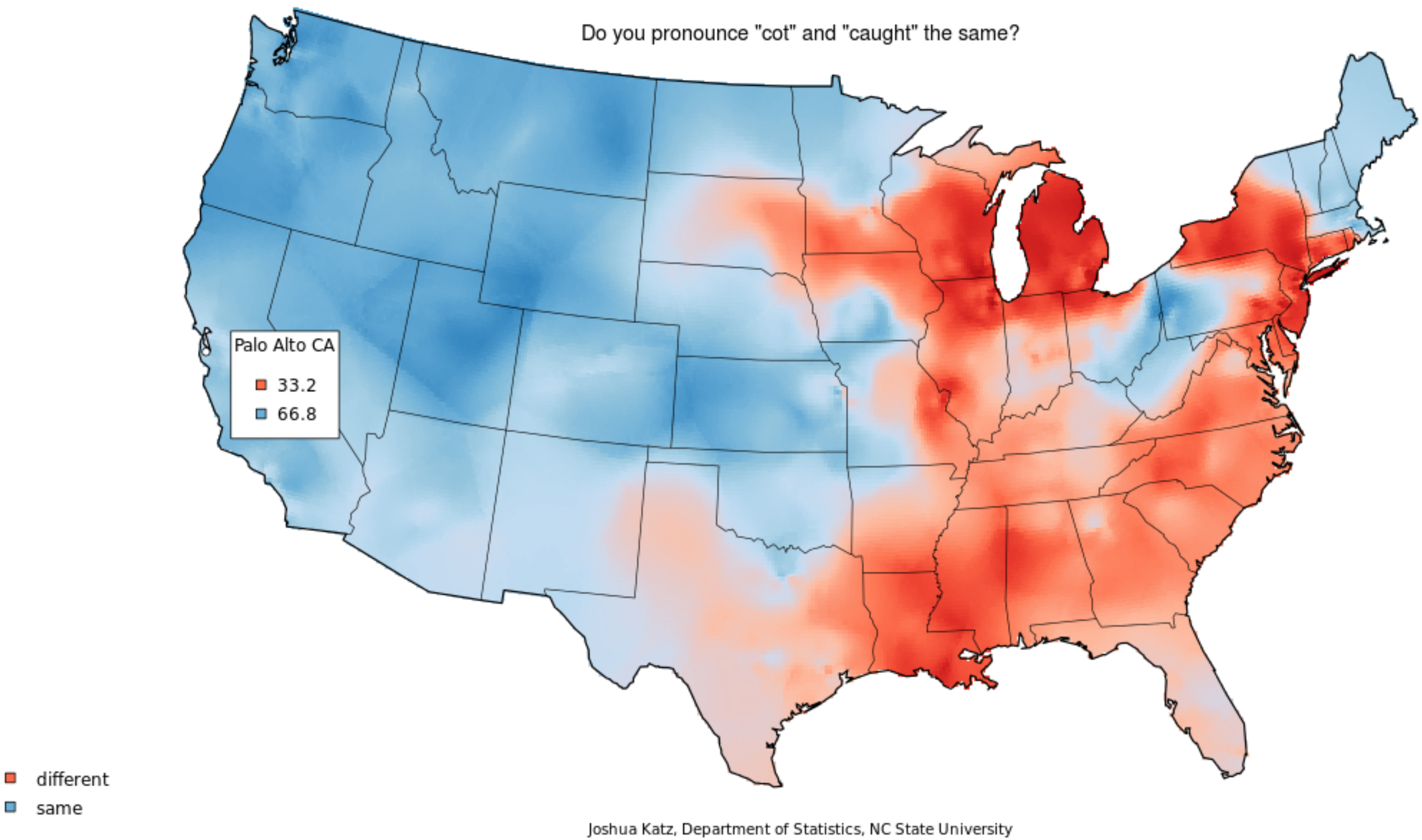


merged dialect



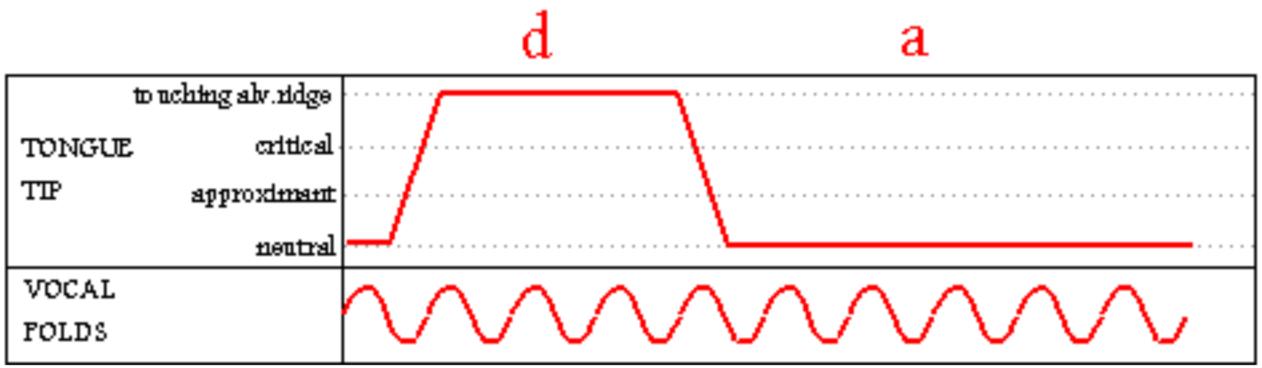
The distribution of these two versions of English

This is just fun to look at… I have no major point with this.

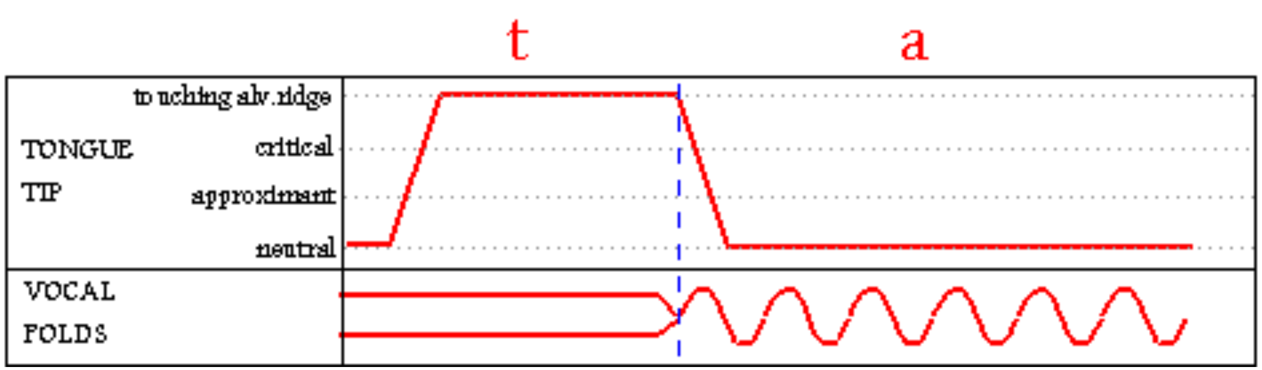


All of the articulatory features are physically continuous!

Voiced stop: the vocal folds start vibrating <u>at the same time the air constriction begins</u>



Voiceless stop: the vocal folds start vibrating <u>as the air constriction ends</u>

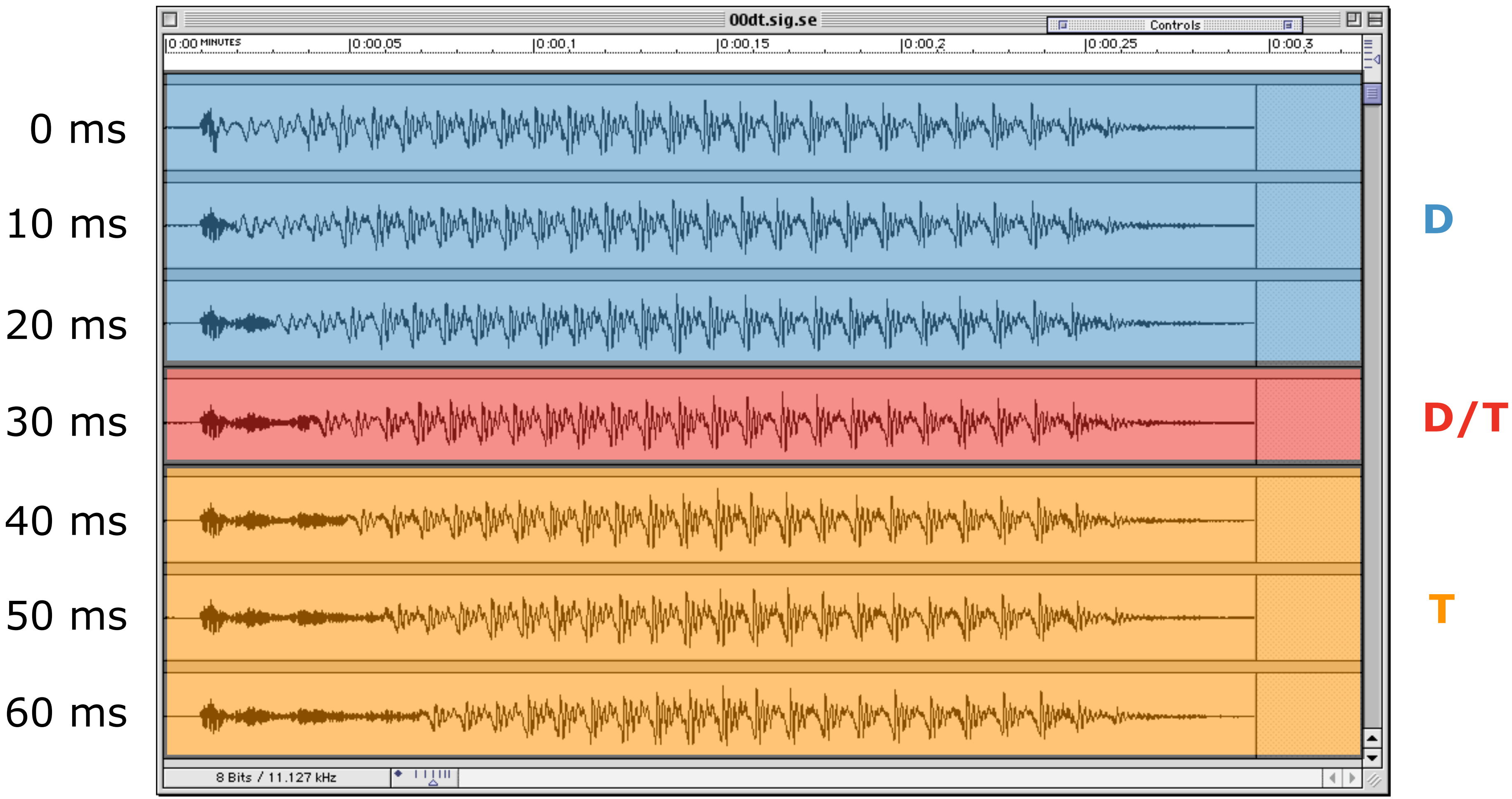


Voice Onset Time is continuous because time is continuous!

Voice Onset Time: A measure of the time between the start of an utterance and the start of the vocal fold vibration

The continuum of voice onset time

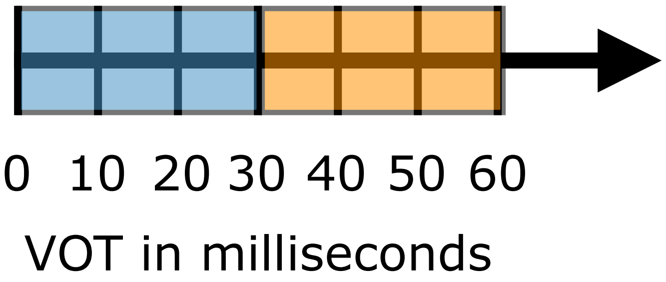
We can create examples of VOT along the continuum. Here are 7 examples, ranging from 0ms VOT to 60ms VOT:



Even though VOT is a continuum, we perceive it as two <u>categories</u>: D and T.

Making categories out of the continuum of voice onset time

When children learn language, they have to learn to break the continuum of VOT into two categories, D and T. This is the same idea of category that we've seen before - sounds within a category are treated as identical even though they differ in details.



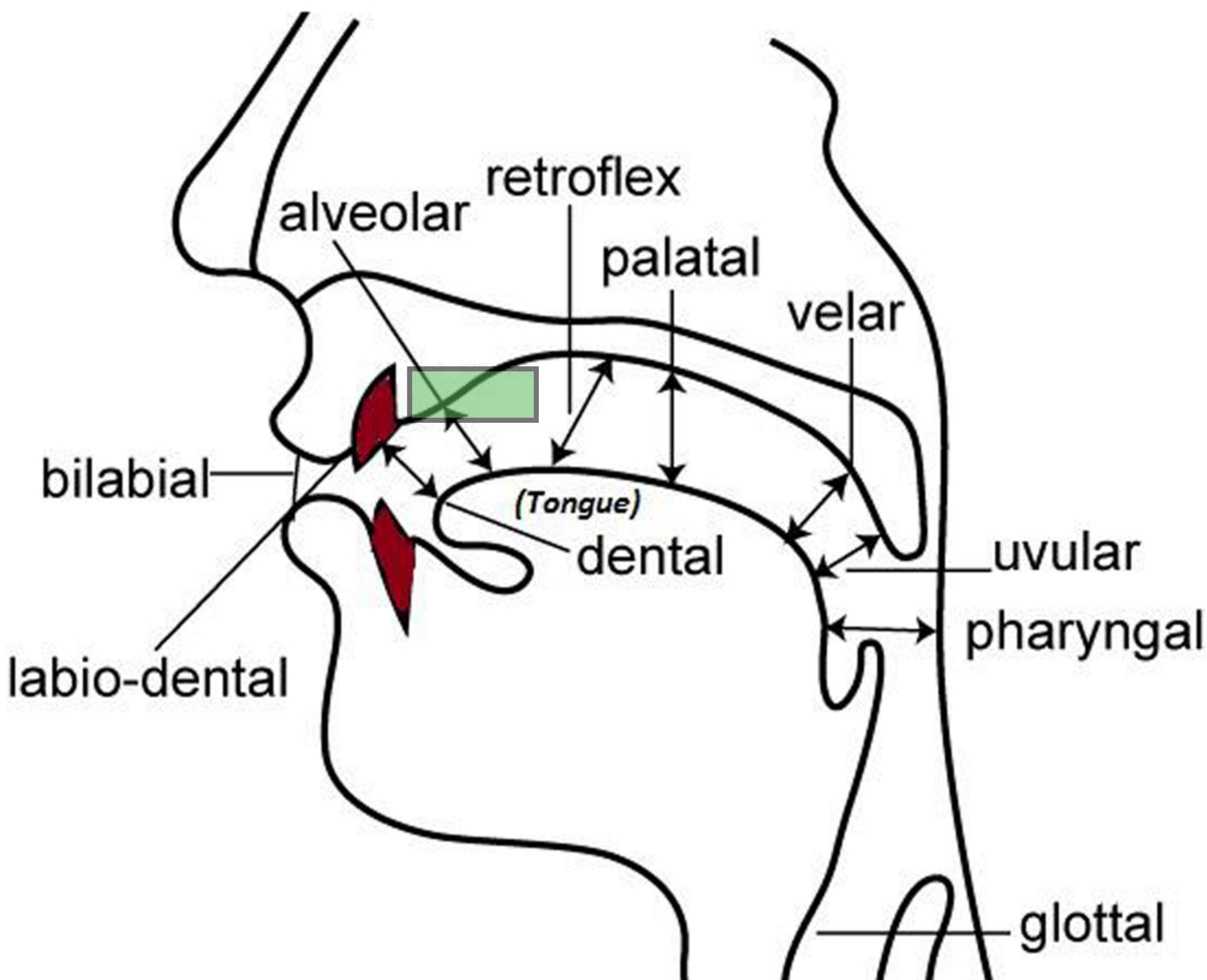
And just to be absolutely clear, the fact that you perceive the low VOT sounds as similar to each other, and the high VOT sounds as similar to each other, has nothing to do with them having similar VOTs. We can show this be creating pairs of sounds that differ by the same VOT, for example, 20ms.

1. 0 ms versus 20 ms	20 ms difference	SAME
2. 20 ms versus 40 ms	20 ms difference	DIFFERENT
3. 40 ms versus 60 ms	20 ms difference	SAME

The continuum of place of articulation

Just like voicing is continuous because time is continuous, place of articulation is continuous because physical space is continuous.

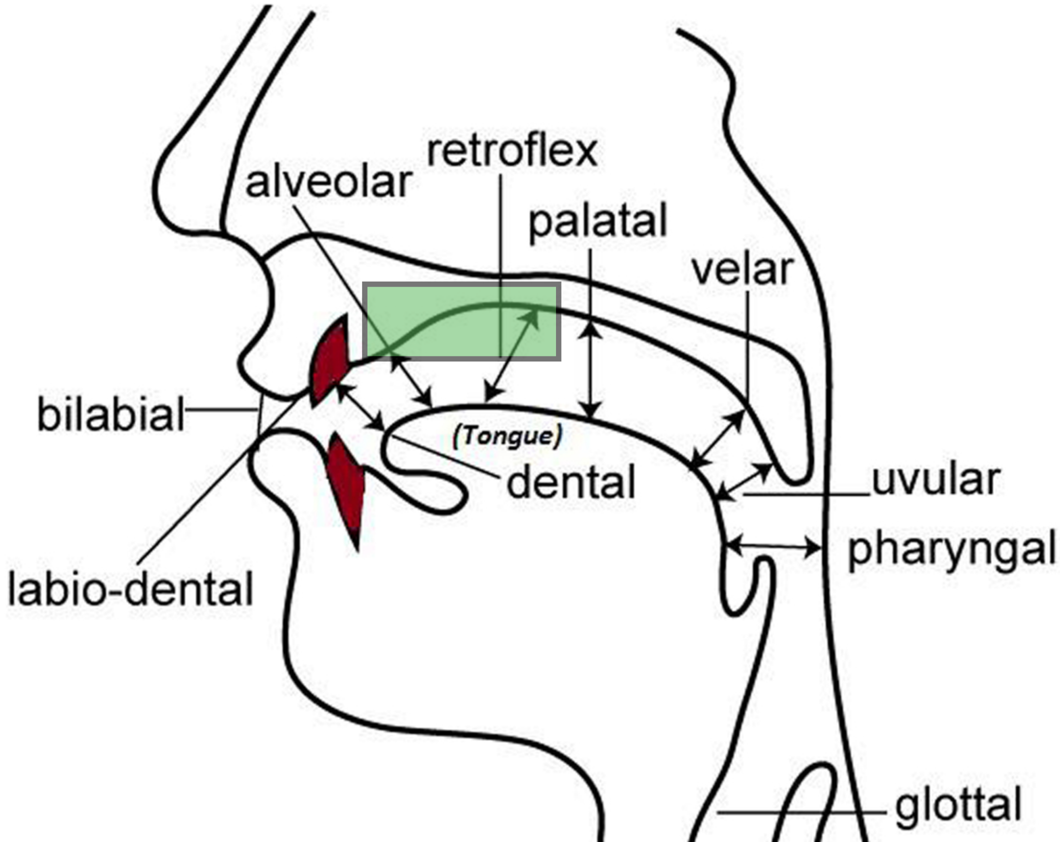
We talk about locations like “alveolar”, but really it is a region of space.



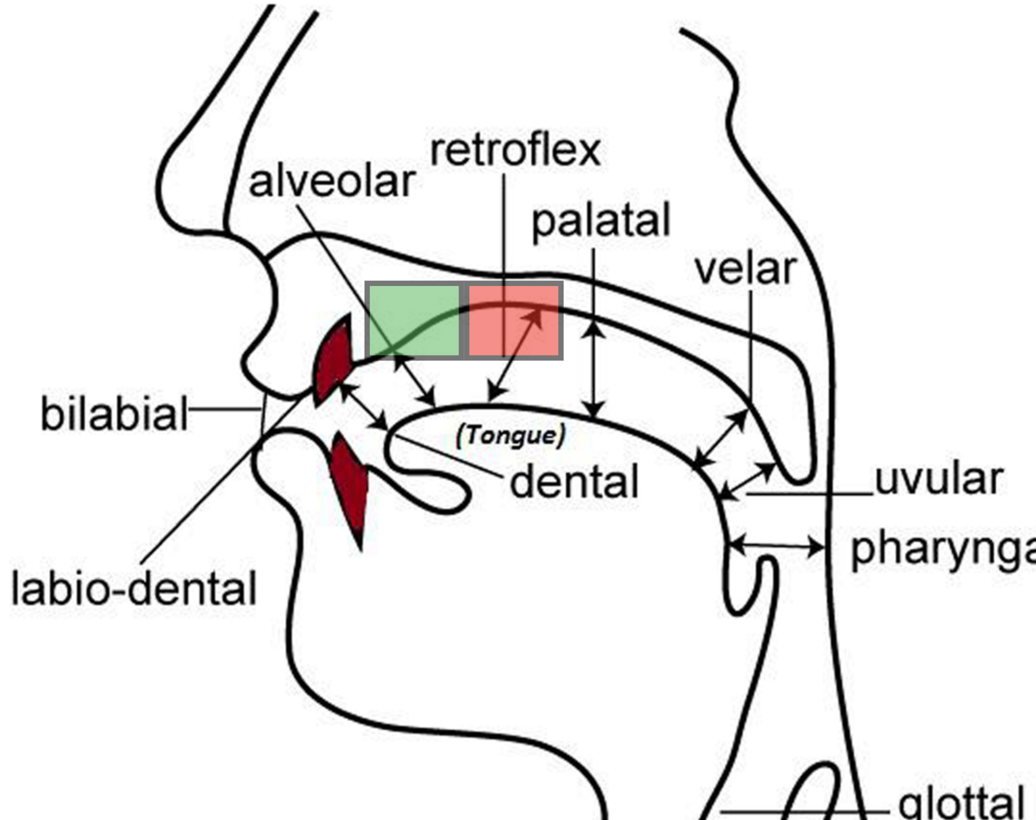
The continuum of place of articulation

And just like voicing, children learning language must learn to break the continuum into <u>categories</u>.

In English, children learn to make no distinction between alveolar and retroflex places of articulation. For example, every voiced stop in this region is perceived as d.



In Hindi, children learn to make a distinction between alveolar and retroflex places of articulation. Hindi speakers have both alveolar d and retroflex d.



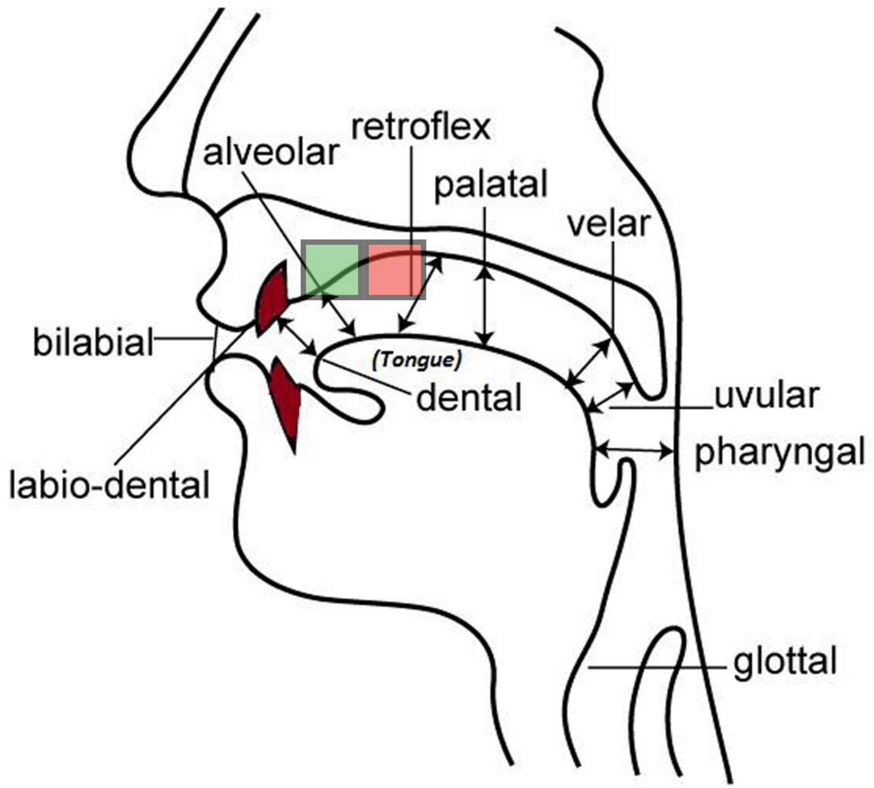
The continuum of place of articulation

Here are eight examples that span the continuum from pure alveolar d to pure retroflex d.

/d/

1.
2.
3.
4.
5.
6.
7.
8.

/ɖ/



For comparison, here is d and t again.

/d/

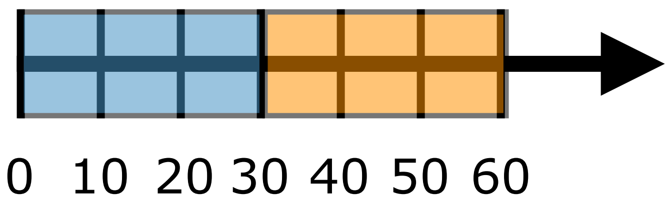
1.
2.
3.
4.
5.
6.
7.

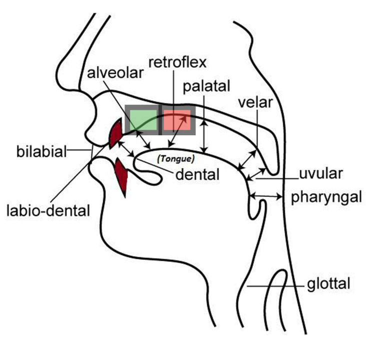
/t/

And here is the shocking part…

Children appear to be born with the ability to discriminate every sound difference found in human languages. They are universal listeners. This means they have all possible breaks already in their minds.

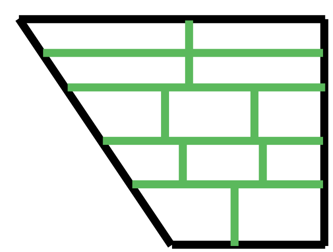
For VOT, this means that they already have the d/t boundary in place.





For place of articulation, this means that they already have the alveolar d / retroflex d boundary in place.

For vowels, this means that they already have the cot / caught boundary in place.



Over time, children lose the ability to discriminate speech sounds that aren’t in the language being spoken around them. Only the category boundaries in their language remains. This whole process takes about 10-12 months.

How did we figure all of this out?

If you have ever played with an infant from 0-12 months, you know that they aren't really doing much that looks like language. You certainly can't ask them whether they can discriminate different sounds. So how did we figure all of this out?

The answer is something called the **Conditioned Head-Turn Procedure**.

Here is Janet Werker explaining the task. She is the one who first determined that children lose the ability to discriminate sounds by 10-12 months of age!

https://www.youtube.com/watch?v=CSMjKDZvNWA&ab_channel=EricaHolt

As you can see, the Conditioned Head-Turn procedure takes advantage of children's desire to see novel fun things, and their ability to turn their head. The idea is that we can train them to expect a novel fun thing after a change in the (boring!) sound being played in the background. If they can hear the difference in the sound being played, they turn their head to look for the fun thing. If they can't hear the difference, they don't turn their head!

The other side of the coin: babbling

Experimental procedures like the conditioned head-turn procedure let us see what children can do during language comprehension. But what about language production?

Age

6 months: Babbling begins. Babbling at this age tends to be repetitive (ba ba ba ba ba), and does not necessarily correspond to the language being spoken by adults! (https://www.youtube.com/watch?v=Zmf1kpXRlJg&ab_channel=LauraMcGarrity)

6-10 months: Over time, babbling starts to show variability (ba bi da di do), and slowly starts to take on more and more characteristics of the language being spoken by adults.)

10-12 months: The sounds created during babbling only come from the adult language. This is the last babbling stage before true words are spoken (around 12 months). (https://www.youtube.com/watch?v=sMaxy8uaJjY&ab_channel=LauraMcGarrity

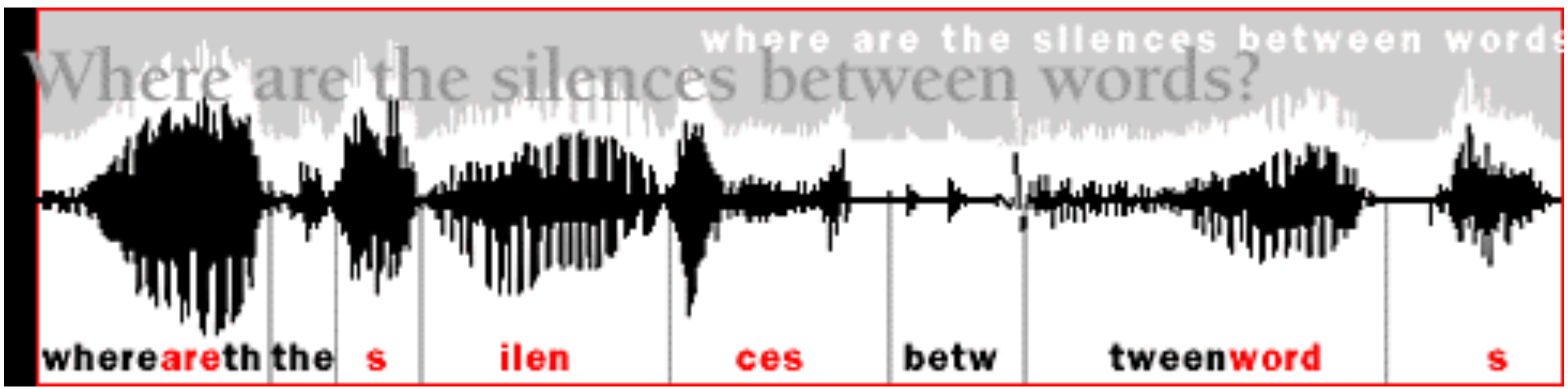
And here is a bonus: two twin boys babbling with each other: https://www.youtube.com/watch?v=_JmA2ClUvUY&feature=related

Learning the phonetic representation of morphemes

The word segmentation problem

You may recall from the first section of class that we learned that there is no obvious way to identify individual speech sounds in a stream of speech.

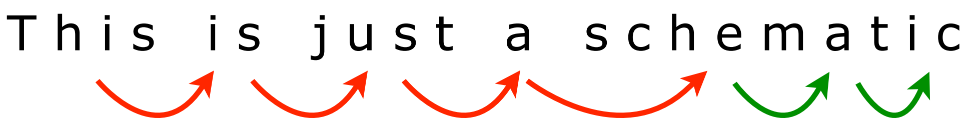
Well, this problem scales up to words too. The stream of speech is a continuous modulation of amplitude and frequency. There are no obvious breaks in the physical signal that correspond to breaks between words.



The **word segmentation problem** is the fact that children must somehow decide where the breaks are between words in the speech stream, despite the fact that there are no physical breaks in the stream (i.e., they must segment the speech stream into words)

Transitional Probabilities of syllables

One popular proposal is that children track the probability that certain syllables appear in a sequence. The idea is that syllables that are part of the same word will frequently follow each other — because they are in a word together! But syllables that are in different words will have a low probability of following each other - because it is just an accident of that one sentence that they are next to each other!



We call this the **transitional probability** - it is the probability of transitioning from one specific syllable to the syllable that comes after it.

It is really easy to calculate (on a computer). You simply find every instance of a syllable in a corpus, and then look at the syllable that comes after it each time. You then pick one syllable like is, and divide the number of times you see that syllable after this by the number of times this appears:

$$\text{transitional probability (this is)} = \frac{\text{\# of “this is” sequences}}{\text{\# of this}}$$

It doesn't work!

Gambell and Yang 2006 took real child-directed speech (from the CHILDES project that we mentioned last time!), converted it to IPA, and built a computational model to mimic a child learning word boundaries from transitional probabilities of syllables.

What percentage of the actual word breaks did it identify correctly?
This is called recall:

$$\frac{\text{true positives}}{\text{true positives + false negatives}} = \frac{\text{breaks-it-got-right}}{\text{total-number-of-real-breaks}} = \boxed{23.3\%}$$

What percentage of the breaks it guessed were real breaks?
This is called precision:

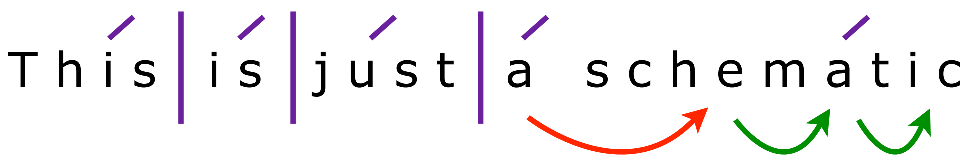
$$\frac{\text{true positives}}{\text{true positives + false positives}} = \frac{\text{breaks-it-got-right}}{\text{total-number-of-guesses}} = \boxed{41.6\%}$$

This is absolutely terrible. Transitional probability alone is not sufficient to learn word boundaries.

Adding primary word stress

Every word carries one and only one primary stress. If we postulate that a child can detect primary word stress (an innate ability?), and if we postulate that children know that words carry only one primary word stress (innate knowledge!), then we let them use this information to detect word boundaries:

The learning algorithm could work like this:
1. Postulate a word boundary when two adjacent syllables both have stress.
2. Use transitional probability for all other sequences.



What we can see right away is that this will capture a good portion of single syllable words. So we expect the success rate to go up!

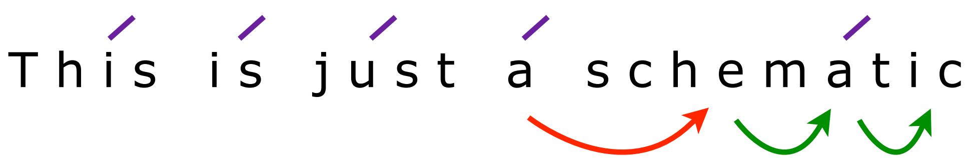
What percentage of the actual words did it identify correctly? 71.2%
recall: TP/(TP+FN)

What percentage of its guesses were real words? 73.5%
precision: TP/TP+FP

Does that solve it?

This is beyond this class, but identifying stress is not easy. There is no single acoustic correlate of it — sometimes it is amplitude, sometimes it is a change in frequency, sometimes it is a change in vowel quality. So this really just kicks the can down the road.

And we have no idea if children actually use stress…



So there is a lot of work left to do. But this shows the way that building in some basic innate knowledge (the genetic hypothesis/nativism) can help.

Learning the semantic representation of morphemes

Nouns seem like they should be easy to learn, but they are not!

Let's say you see this:

And I say "glorp".

What do you think "glorp"
is referring to?

It could be any number of things!

It could be the full bear.

It could be a piece of the bear, like its arm or eyes.

It could be a property of the bear, like soft or brown.

It turns out that children appear to have a bias to associate words with **whole objects**, not subparts of them. So words for whole objects tend to be learned before words for smaller parts of objects.

How do children generalize from one instance to another?

Let’s say you show this array to a child, and ask them if there are any glorps. Which they choose can tell you what they think the word means. This study was done in Japanese, where there are no grammatical differences between proper names and other types of nouns (Imai and Haryu 2001):

just this bear = proper name

(stuffed) brown bear

(stuffed) bear - no color specified

(stuffed) animal in general

2yr	4yr
11%	18%
24%	24%
54%	57%
8%	0%

How do children generalize from one instance to another?

Let's say you show this array to a child, and ask them if there are any glorps. Which they choose can tell you what they think the word means. This study was done in Japanese, where there are no grammatical differences between proper names and other types of nouns (Imai and Haryu 2001):

54%

(stuffed) bear - no color specified

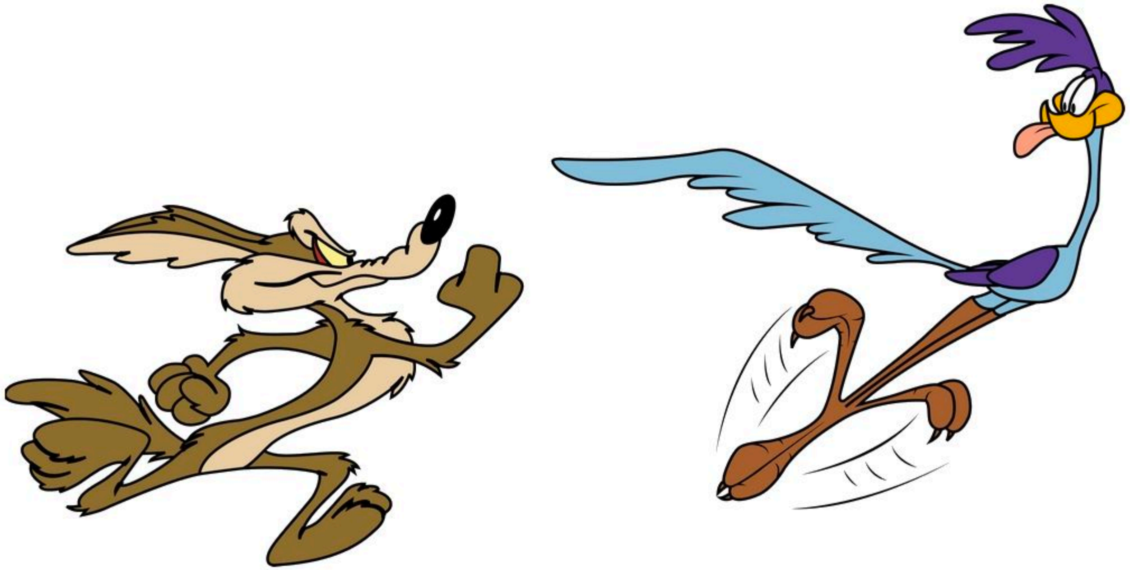
This is an open area of investigation, but it seems that children are biased to **basic level categories** like "bear" over more general categories like "animal" or specific categories like "brown bear". But what counts as basic vs general/specific is hard to define a priori. More work is needed!

Verbs are even harder than nouns

The challenge with verbs is that they have very subtle meaning differences that **cannot be easily deduced from real-world context.**

Let’s say you see this scene, and I say “blicking”.

It could mean chasing, or it could mean fleeing.

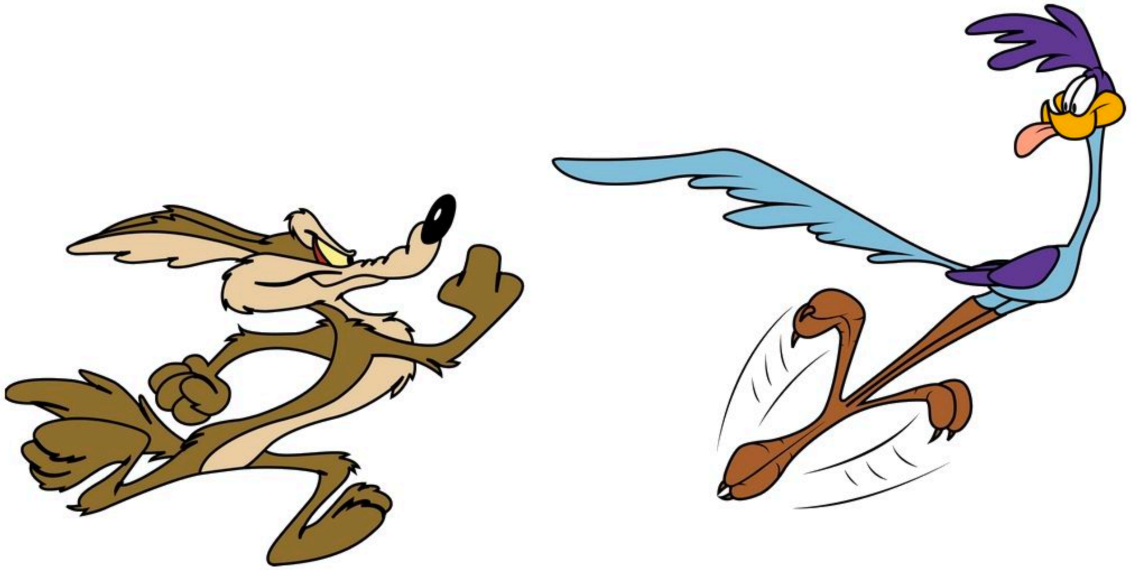


The issue is that any scene that is compatible with chasing will also be compatible with fleeing (if the objects in it are living, animate beings). So a child can’t possibly figure out the correct meaning of blicking from the scene.

This is not a small problem. There are many such pairs: buy/sell, give/receive, etc. And there are other pairs that show subtle semantic differences of other kinds (like look vs see). The bottom line is that verbs have subtle meanings!

Verbs are even harder than nouns

One interesting solution to this problem was proposed by Barbara Landau and Lila Gleitman (1985). They suggested that children could use **syntax** to help them learn verb meanings.



If I say the full sentence:

Roadrunner blicks coyote.

And if a child already knows that the subject of a sentence is the agent of the action…

Then the child can deduce that the meaning of blicks is flee, because the sentence is about the roadrunner being the agent of the action.

This is called **syntactic bootstrapping**. The word bootstrapping is metaphorical - syntax is the metaphorical bootstrap that they use to pull on the metaphorical boot of verb meaning. Crucially, it is another proposal for innate knowledge! (And there is some evidence that children can do this as early as 2 years: see Fisher et al. 2020 for some examples.

Learning syntactic rules

Head movement seems to occur in all yes-no questions

The head-movement transformation appears to be part of the process for forming all yes-no questions in English:

Declarative	**Question**
Lisa is running.	Is Lisa running?
Lisa can juggle.	Can Lisa juggle?
Lisa will read.	Will Lisa read?
Lisa has left.	Has Lisa left?

So it is going to be a pretty important grammatical rule for children to learn!

Looking at head-movement more deeply

So far we’ve defined head-movement as “move is”. But what happens when there is more than one “is” in the sentence?

Lisa is thinking that Mary is smart.

Which “is” are we allowed to move in English (to form a question)?

In this sentence you can move the first one:

Is Lisa <is> thinking that Mary is smart?

But you can’t move the second one:

*Is Lisa is thinking that Mary <is> smart?

This suggests that the definition of the head-movement transformation is more complicated than just “move is”. It has to be something like “move the first is”.

Testing a new definition (a new theory)

So let’s see if we can test the theory that the definition of head-movement is “move the first is”.

Theory to be tested: In English, you move the first is.

Here is a new sentence to test it on:

The student that is happy is smart.

You CAN’T move the first one: *Is the student that <is> happy is smart.

But you CAN move the second: Is the student that is happy <is> smart.

This sentence falsifies our theory. So we need a new one.

Quick review of the facts that our theory needs to capture

For some sentences, you move the first instance of is.

Lisa is thinking that Mary is smart.

Is Lisa is thinking that Mary is smart?

For other sentences, you move the second instance of is.

The student that is happy is smart.

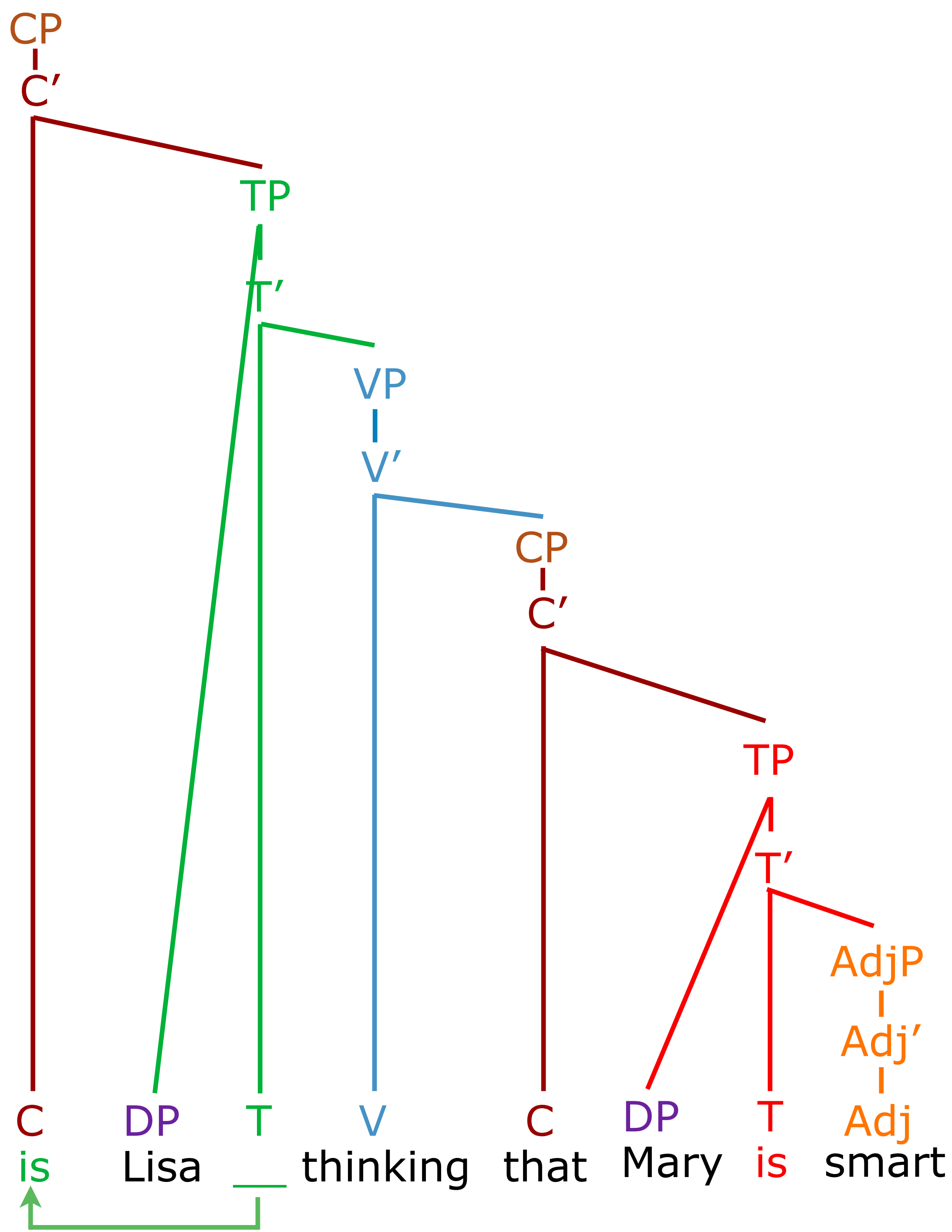
Is the student that is happy is smart.

So we need a theory that **does not rely on linear order**. Because the linear order doesn't seem to be the deciding factor.

The answer is structure, of course!

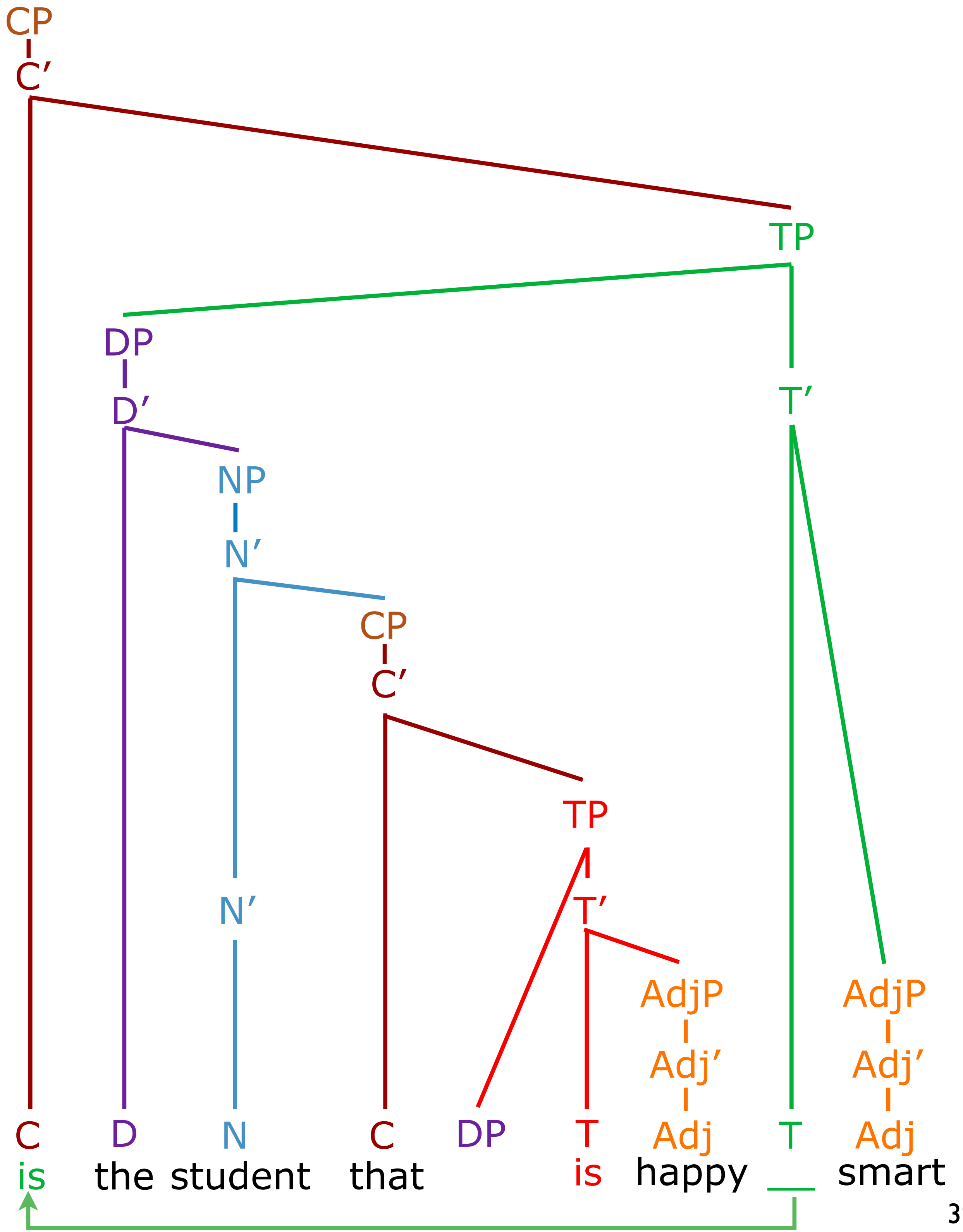
Here is sentence 1. If we draw a tree, we can easily see that the “is” that we move is the structurally higher one. The one in the matrix clause!

The “is” that we cannot move is the structurally lower one. The one in the embedded clause!



Here is sentence 2. If we draw a tree, we can easily see that the “is” that we move is the structurally higher one. The one in the matrix clause!

The “is” that we cannot move is the structurally lower one. The one in the relative clause, which is another form of embedded clause!



The correct theory

So now we see that the correct theory is something like "move the is that is in the matrix clause"

Correct theory: Move the is that is in the matrix TP

When we apply this to our two test sentences, we can see that it works:

Here you move the matrix clause is, but not the embedded clause is:

matrix TP "is" embedded TP "is"

Is Lisa <is> thinking that Mary is smart?

Here you move the matrix clause is, but not the relative clause is:

relative clause "is" matrix TP "is"

Is the student that is happy <is> smart.

Structure Dependence

So in the end it looks like the "move first" theory doesn't work, but the "move the matrix IP is" does work:

Move first theory: Move the first is.

Move matrix TP theory: Move the is that is in the matrix TP

One interesting difference between these two theories is that the first one only makes reference to the **linear order** of the words in the sentence. It doesn't make reference to the hierarchical structure of the sentence at all.

But the second one makes reference to the **hierarchical structure** of the sentence. It makes a distinction between the is that is in the matrix TP, and any is's that are in other clauses.

For this reason, linguists call the second (correct) definition of head-movement a **structure dependent** rule. The transformation is defined in terms of the hierarchical structure of the sentence (e.g., matrix vs embedded TP). So it is a structure dependent rule.

Learning Structure Dependence

OK, so what is the big deal? Well, children need to learn the **correct definition of head-movement** in order to be able to create English questions.

We have seen two theories that they could try. How do they decide which one to try?

Move the first is:

Pros: Simpler (not structure dependent)

Cons: Ultimately incorrect

Move matrix TP is:

Pros: Ultimately correct

Cons: More complicated (structure dependent)

A possible (but incorrect!) learning theory

Here is one possible theory of how head-movement could be learned:

Step 1: Children notice that questions in English are formed by moving is.

Is Lisa <is> running?

Step 2: Children postulate the hypothesis that "move first" is the correct theory. They choose this one first because it is simpler, and because it works for a lot of questions in English:

Is Lisa <is> thinking that Mary is smart?

Step 3: At some point, children notice a sentence that is incompatible with "move first". So they switch to the hypothesis that "move matrix" is the correct theory:

Is the student that is happy <is> smart.

Why do we think that this theory incorrect?

Step 1: Children notice that questions in English are formed by moving is.

Step 2: Children postulate the hypothesis that “move first” is the correct theory. They choose this one first because it is simpler, and because it fits with a lot of questions in English:

Step 3: At some point, children notice an example that is incompatible with “move first”. So they switch to the hypothesis that “move matrix” is the correct theory.

The problem with this theory is that successful learning requires hearing sentences like the following in order to notice that there is a problem with the move first theory:

Is the student that is happy <is> smart.

Legate and Yang (2002) looked at over 20,000 questions that were spoken to a child in the CHILDES database… and they found precisely zero questions of the critical type! So if children relied on hearing this sentence to learn questions, **they would never learn how to form this question correctly!**

Why do we think that this theory incorrect?

Step 1: Children notice that questions in English are formed by moving is.

Step 2: Children postulate the hypothesis that “move first” is the correct theory. They choose this one first because it is simpler, and because it fits with a lot of questions in English:

Step 3: At some point, children notice an example that is incompatible with “move first”. So they switch to the hypothesis that “move matrix” is the correct theory.

This theory has a chronological prediction — If children entertain the move first theory, then at some point during language acquisition, they should think that English involves moving the first “is”. We can look for evidence of this!

So how do children learn this?

Crain and Nakayama (1987) performed experiments to try to get children to produce yes-no questions in an attempt to see if they ever entertained the "move first" theory.

The prediction is that if children do entertain the hypothesis that "move first" is correct early in acquisition, then at some point early in acquisition they should produce sentences that follow the "move first" theory. These sentences will look like errors to us:

This is ungrammatical in adult English, but it is predicted to be spoken by children if they believe that "move first" is correct early in acquistion.

"Is the girl who <is> skating is tall?"

To test this prediction, Crain and Nakayama recruited a group of children ages 3;2 - 4;7, and played with puppets to try to get them to create yes-no questions. Then they looked to see if any of the yes-no questions showed the "move first" pattern. They tried really hard: they elicited 81 yes-no questions from the children.

Here is an example

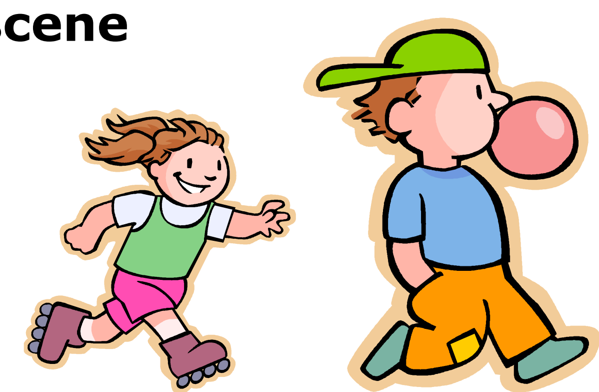
Experimenter: "Hey [child's name], look at that girl who is skating. Do you think she is tall?"

Child: "No! She isn't tall!"

Experimenter: "I wonder if Jabba thinks she is tall. Ask Jabba if he thinks the girl who is skating is tall."

Child: …… [creates the question] ……

scene



Jabba Puppet

This was 1987!

The results

It should go without saying that children this young do make mistakes. In fact, they make more mistakes than correct responses. But the critical question is what type of mistakes do they make?

Do they make mistakes that suggest the “move first” theory? Or do they make other types of mistakes?

	Child Response	Results
adult-like	“Is the girl who is skating tall?”	38%
move first error	“Is the girl who skating is tall?”	0%
mistakes	“Is the girl who is skating is tall?”	37%
mistakes	“Is the girl who is skating, is she tall?”	12%
mistakes	other types	13%

What does this mean for learning?

These facts (the corpus facts and the experimental facts) seem to suggest that our learning theory is wrong:

Step 1: Children notice that questions in English are formed by moving is.

Step 2: Children postulate the hypothesis that "move first" is the correct theory.

No. Children do not seem to ever entertain the "move first" hypothesis. (Crain and Nakayama 1987)

Step 3: Children switch to the hypothesis that "move main clause" is the correct theory when they notice an example that is incompatible with "move first"

No. Children do not seem to ever hear sentences that would show that "move first" is wrong. (Yang and Legate 2002)

So how could children possibly learn the correct definition ("move matrix TP") given all of this?

A nativist theory might work

One possibility that might work is to postulate that children know (innately) that all transformations must be structure dependent.

Step 1: Children notice that questions in English are formed by moving is.

Step 2: Because children know innately that all transformations must be structure dependent, even a simple sentence is evidence that head-movement targets the matrix clause is:

matrix clause is

Is Lisa <is> running?

Step 3: Therefore, as soon as children notice that a transformation is necessary, they will know the correct definition (the structure dependent definition).

Putting together all we saw in our whirlwind tour…

Some research programs for the genetic hypothesis

For learning phonemes:

Which boundaries do children know at birth? Are there more than there needs to be (distinctions that languages don't use)? Are there any that must be learned?

For learning the phonetic representation of morphemes:

Do children use stress for word segmentation? If so, what is the acoustic cue? Do they know the unique stress constraint innately?

For learning the semantic representation of morphemes:

What are the biases for learning nouns - whole word, basic-level (others are mutual exclusivity, shape)? Do children have some innate knowledge of syntax (agents, etc)?

For learning syntactic rules?

Do children know that syntactic rules are structure-dependent innately? Do they ever entertain hypotheses that are not structure dependent?