جامعة نيويورك أبوظبي

NYU | ABU DHABI

PSYCH-UH 2218: Language Science

Class 22: Language Diversity, Death, and Genesis

Prof. Jon Sprouse
Psychology

Navajo Code Talkers

Navajo (called Diné by the actual speakers) is a native american language spoken by about 150,000 members of the Navajo nation in the southwest US.

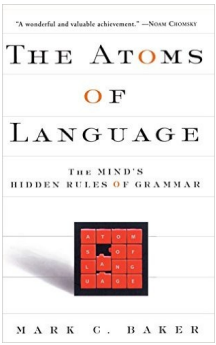
https://www.youtube.com/watch?v=VR13lIRLfic

During WWII, the Japanese easily broke US military codes. So in 1942, the Marines tried something a little different: they recruited Navajo to use their language to encode messages.

The Japanese never managed to decipher the language. The Navajo code talkers turned the tide of the war in the pacific.

The Code Talker Paradox

In his book *The Atoms of Language* (highly recommended!), linguist Mark Baker points out the apparent paradox of Navajo Code Talkers.



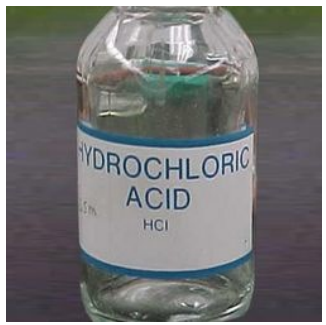
On the one hand, Navajo Code Talkers were able to accurately, and rapidly, translate from English to Navajo and back to English without losing any information. This suggests that the two languages are similar enough to directly translate from one to the other.

But on the other hand, Japanese cryptographers were not able to decipher the Navajo language. Remember, they had broken the very best codes the military had in a matter of months, but over the course of 3 years, they were unable to break the Navajo code. This suggests that Navajo and English were too dissimilar for the codebreakers to decode.

A chemical paradox

Baker makes a great analogy, and I am just going to steal it here. Chemistry is full of similar paradoxes, where substances with very different properties can nonetheless be converted into each other. Here is a classic example

HCl



NaOH

Hydrochloric acid and sodium hydroxide are two very caustic chemicals. They will each burn you if you touch them.

Water is necessary for us to survive, and table salt tastes awesome on food.

On the one hand, HCL and NaOH seem incredibly different from H_20 and NaCl.

H_20

NaCl

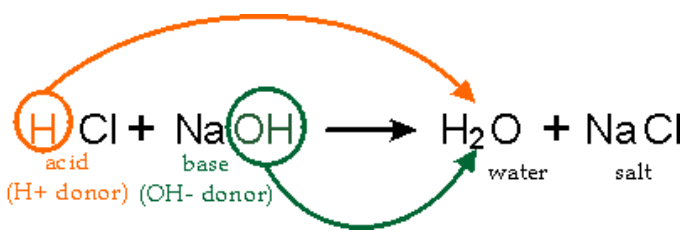
But on the other hand, if you combine HCL and NaOH in the right proportion, you will end up with H_20 and NaCl. You can convert from poison to food.

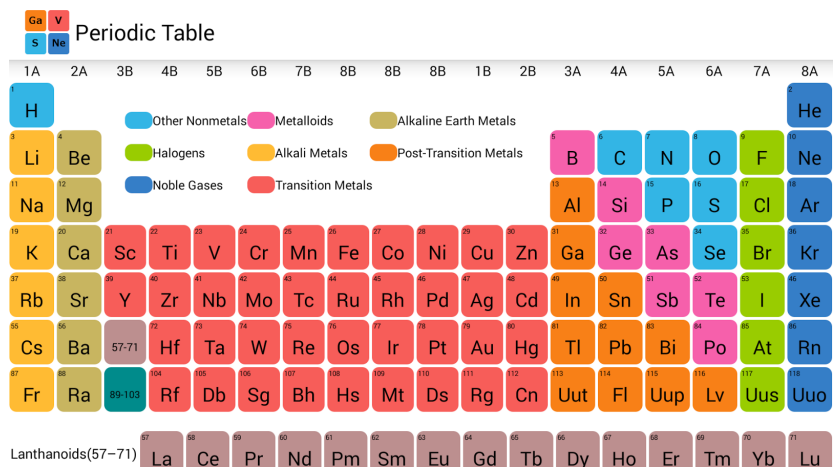
Solving the paradoxes

We all know how the chemical paradox is solved:

General Principles of Chemistry: Conservation of mass and energy, valences, etc.

A theory of elements: A list of the elements that can make up a substance, and the properties of those elements.





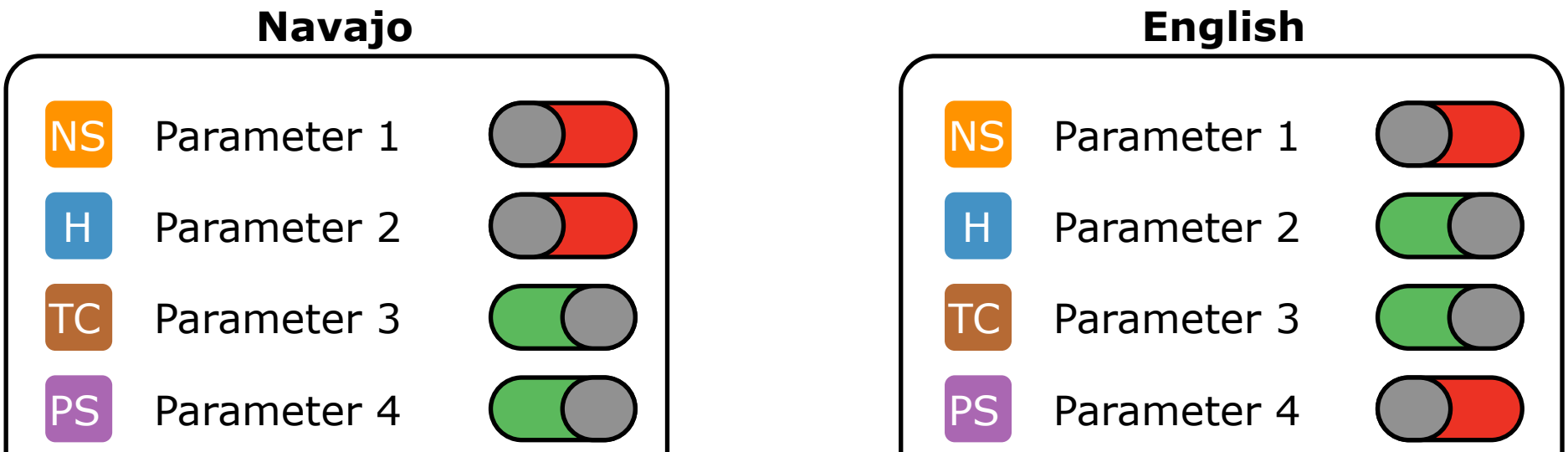
Solving the paradoxes

Linguists have proposed a similar solution to the paradoxes of linguistic variation. This theory is called **Principles and Parameters Theory**:

Principles: General principles that govern the way languages work. These properties are shared by all human languages.

Parameters: A finite set of options or settings that determine how languages can vary.

Though parameters are analogous to atoms in the work that they do, it is probably easier to think of them like a series of settings on an iphone. You can turn each one on or off, creating a distinct way for your iphone to work.



Some examples of Principles

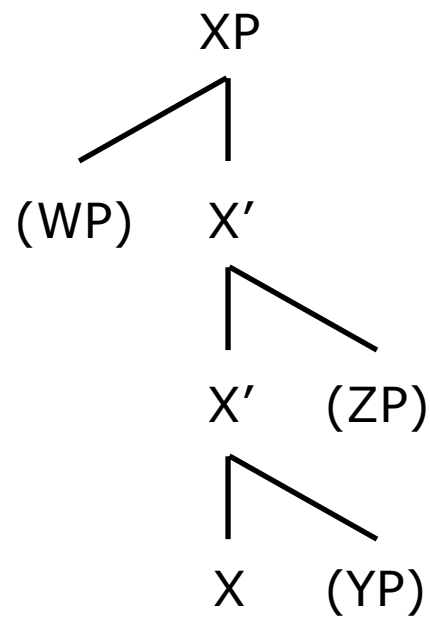
An example of a syntactic principle: X'-Theory

As far as we can tell, all of the languages we've studied in detail (around 2000) use hierarchical structure that fits X'-Theory:

XP → (WP) X'

X' → (ZP) X' (ZP)

X' → X (YP)



A phonetic principle: Articulatory Features

As far as we can tell, all of the languages we've studied in detail (around 2000) use the same set of articulatory features to create phonemes:

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Different languages use different subsets of consonants (and vowels), but all use the same articulatory features to create those consonants (and vowels)

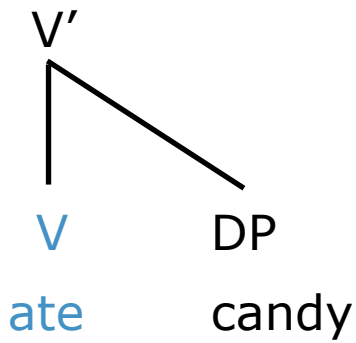
Some examples of Parameters

The Null Subject Parameter

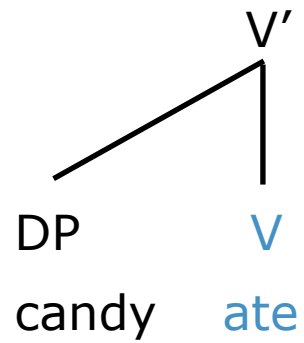
Head Parameter

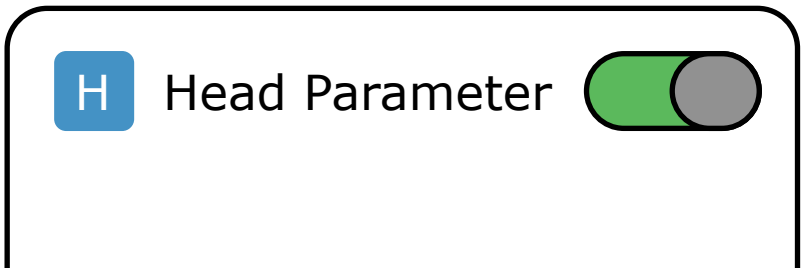
The Head Parameter determines whether heads come first in their phrases (head-initial), or whether they come last (head-final):

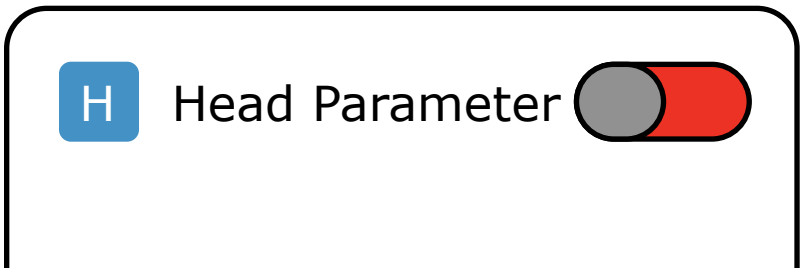
English: John ate candy



Japanese: John candy ate







The Null Subject Parameter

If you speak Spanish or Italian, you may have noticed that subjects of sentences in these languages behave a bit differently than subjects in English (and French).

All four languages allow a subject to be present:

English	John will arrive.	Spanish	Juan llegará.
French	Jean arrivera.	Italian	Gianni verrà.

But, in Spanish and Italian, if the subject has already been discussed, you can omit the subject completely. In English and French, you can’t do this. You have to use a pronoun:

English	He will arrive.	Spanish	Llegará.
French	Il arrivera.	Italian	Verrà.

The Null Subject Parameter - weather

Certain verbs about the weather don't really have an entity that is the subject, but in English and French, you have to use an empty subject "it". We call this an expletive because it has no meaning (this is a different sense of expletive than the senes that means "curse word").

English

It is raining.

It is snowing.

French

Il pleut.

Il neige.

Spanish and Italian don't ever use a subject for these verbs:

Spanish

Llueve.

Nieva.

Italian

Piove.

Nevica.

The Null Subject Parameter - post-verbal

And there is more. In Spanish and Italian, you can put the subject after the verb. But it looks like you can't do this in English and French:

English	*Will arrive John.	Spanish	Llegará Juan
French	*Arrivera Jean.	Italian	Verrà Gianni.

In fact, it is a bit more complicated than this. You can put the subject after the verb in English and French under certain circumstances, but you have to put an expletive subject in the specifier of TP!

English: There is a man in the garden.

French: Il est arrivé trois hommes.

It is arrived three men

These subjects don't refer to anything. The semantic subjects are a man and three men.

The Null Subject Parameter

So it looks like these four languages split into two types:

English/French

Always have a subject.

Spanish/Italian

Omit known subjects.

Omit weather subjects.

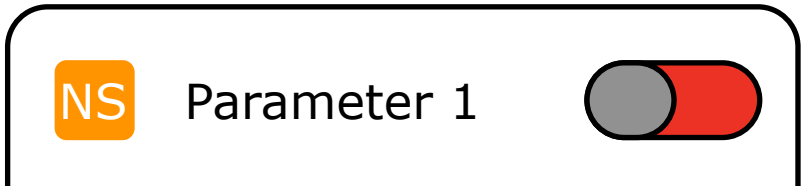
Omit subjects when the subject is after the verb.

Principles and Parameters Theory can account for all three of these properties by postulating a single parameter: **The Null Subject Parameter**

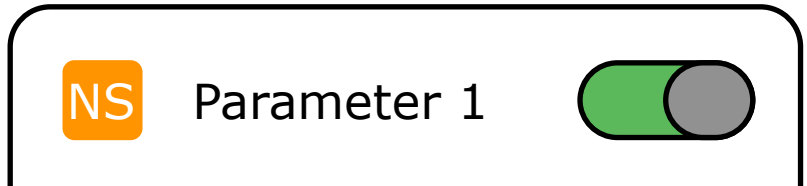
Null Subject Parameter

If set to no, a subject is always required in (finite) sentences. If set to yes, subjects are not required if they can be recovered from the context.

English/French



Spanish/Italian



The Null Subject Parameter

So it looks like these four languages split into two types:

English/French

Spanish/Italian

This is the power of a parameter. If you ever studied these languages in school, you would probably just have to memorize these facts about the languages.

But as linguists, we can see that they all derive from a deeper fact — the Null Subject Parameter!

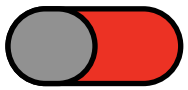
Principles and Parameters Theory can account for all three of these properties by postulating a single parameter: **The Null Subject Parameter**

Null Subject Parameter

If set to no, a subject is always required in (tensed) sentences. If set to yes, subjects are not required if they can be recovered from the context.

English/French

NS Parameter 1



Spanish/Italian

NS Parameter 1

Principles and Parameters Theory is a very powerful idea

The first power of P&P is the ability to capture abstract universal properties of language as principles. The fact that all human languages look roughly the same in terms of things like Phrase Structure is a great example of **structure in the mind**.

The second power of P&P is the ability to **capture large amounts of variation with very few parameters**. If the number of parameters is P, and each parameter has 2 values, then the number of languages that can be created is defined by this equation: number of languages = 2^P

2 parameters = 2^2 = 4 languages

3 parameters = 2^3 = 8 languages

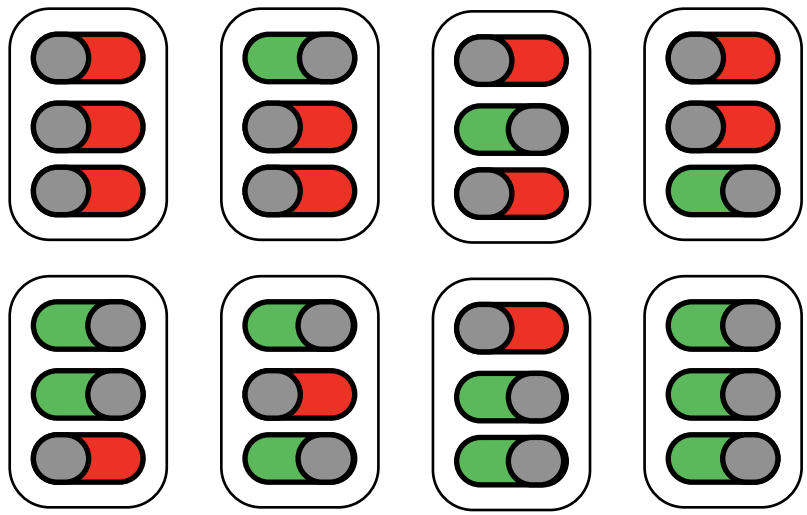
4 parameters = 2^4 = 16 languages

8 parameters = 2^8 = 256 languages

16 parameters = 2^{16} = 65536 languages

32 parameters = 2^{32} = 4 billion languages

Here are 3 parameters:



Do parameters capture all of the variation?

It is important to note that parameters are only intended to capture systematic variation. By systematic, we mean variation that has relatively few options (perhaps only two) and has far reaching consequences.

There is variation between languages that is not systematic (i.e., has lots of potential values, and does not have far reaching consequences). This variation must be captured some other way… perhaps simply through memorization.

For example, the variation between languages is fairly large when it comes to which phonemes they use.

They all choose from the same set of consonants, but the specific choices vary!

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

How many languages are there?

What is a Language?

Recall from our last lecture that we've settled on a **scientific definition** of what it means to be a human language (this might not be an exhaustive definition, but it is a good working definition for this course).

A human language will have the following properties:

1. It will have its own phonology.

2. It will have its own morphology.

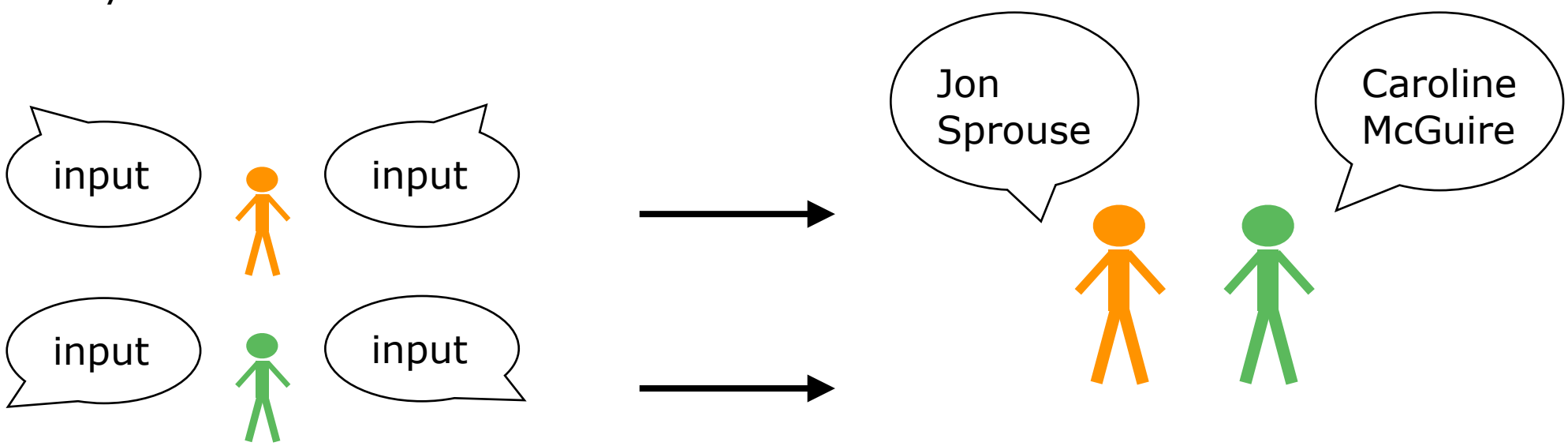
3. It will have its own syntax.

4. It will be acquired using known mechanisms of language learning.

How many languages are in the world according to the scientific definition?

It is very tricky to count languages using the **scientific definition**.

Because every human being hears slightly different input, it is logically possible that every human being will acquire a slightly different phonology, morphology, and syntax.



This means that, according to the scientific definition, it is possible that every human being might speak a different language! (I speak Jon Sprouse, and you speak your own language.) So there could be 8,000,000,000 languages spoken in the world today!

How many languages are in the world according to the colloquial definition?

Obviously, it doesn't make sense to count languages if every person speaks a different language. So for today's purposes, we are going to set aside the scientific definition, and instead use a **colloquial definition**.

I am not going to define this explicitly, because colloquial definitions are not intended to be precise. Instead, let's just rely on our intuitions about how normal people talk about language, and ask: How many languages are spoken in the world today?

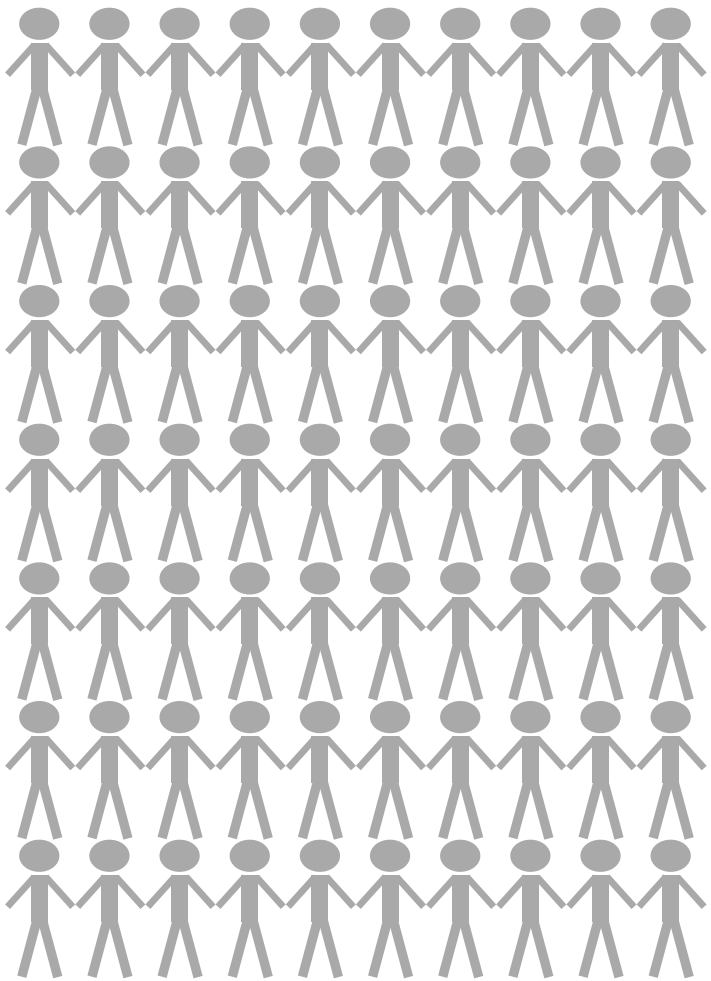
Approximately 7,000

That feels like a lot, doesn't it? If someone were to ask you if we have a lot of linguistic diversity, you would probably say yes, wouldn't you?

When children stop acquiring a language: Endangered Languages and Language Death

The distribution of languages in the world

Here is a representation of the 7,000,000,000 people in the world. Each stick-figure represents 100,000,000 people. (This information is from ~2015.)

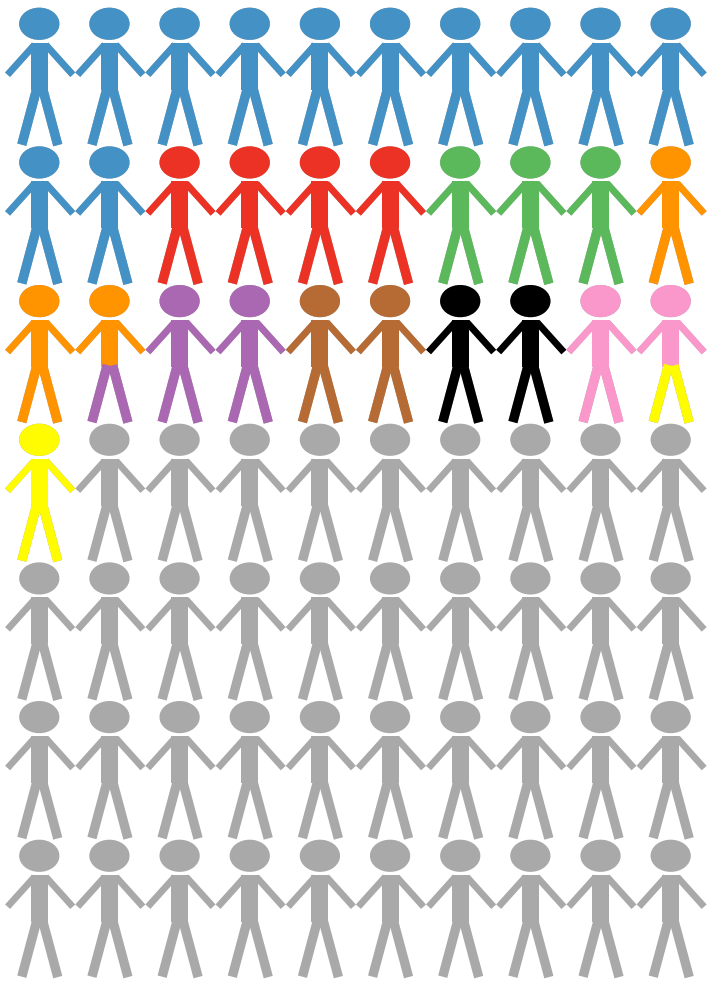


The top 9 most spoken languages in the world are:

1. Chinese
2. Spanish
3. English
4. Hindi
5. Arabic
6. Portuguese
7. Bangla
8. Russian
9. Japanese

The distribution of languages in the world

Here is a representation of the 7,000,000,000 people in the world. Each stick-figure represents 100,000,000 people. (This information is from ~2015.)



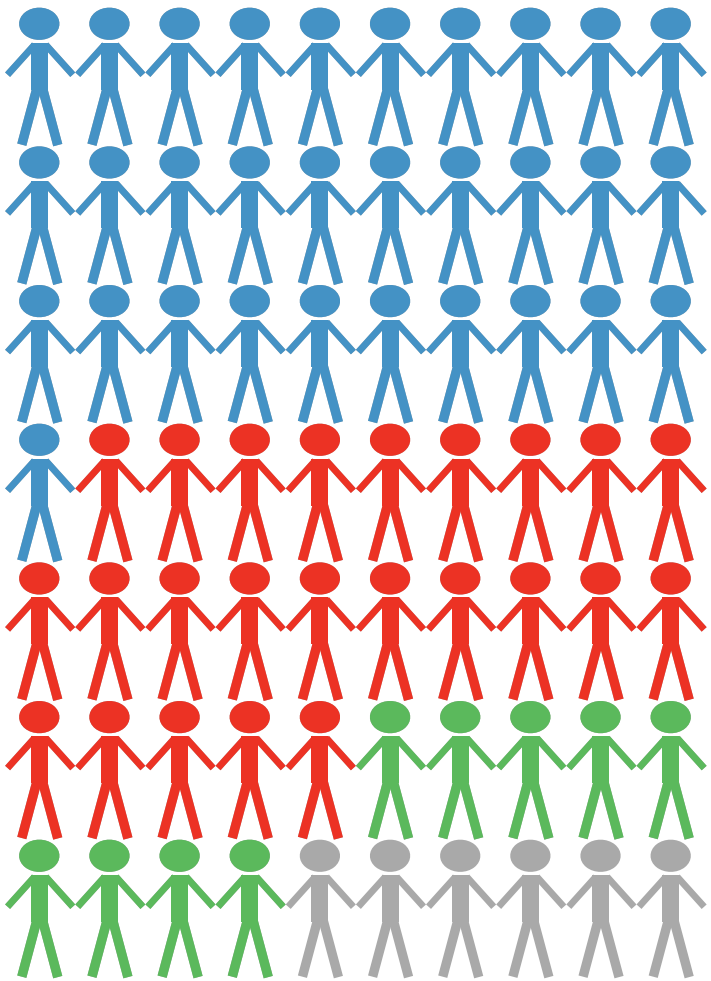
The top 9 most spoken languages in the world are:

1. Chinese	1,200,000,000
2. Spanish	400,000,000
3. English	300,000,000
4. Hindi	250,000,000
5. Arabic	250,000,000
6. Portuguese	200,000,000
7. Bangla	200,000,000
8. Russian	150,000,000
9. Japanese	150,000,000

The big 9 account for 44% of the world's population!

It is even worse than that

Here is a representation of the 7,000,000,000 people in the world. Each stick-figure represents 100,000,000 people. (This information is from ~2015.)



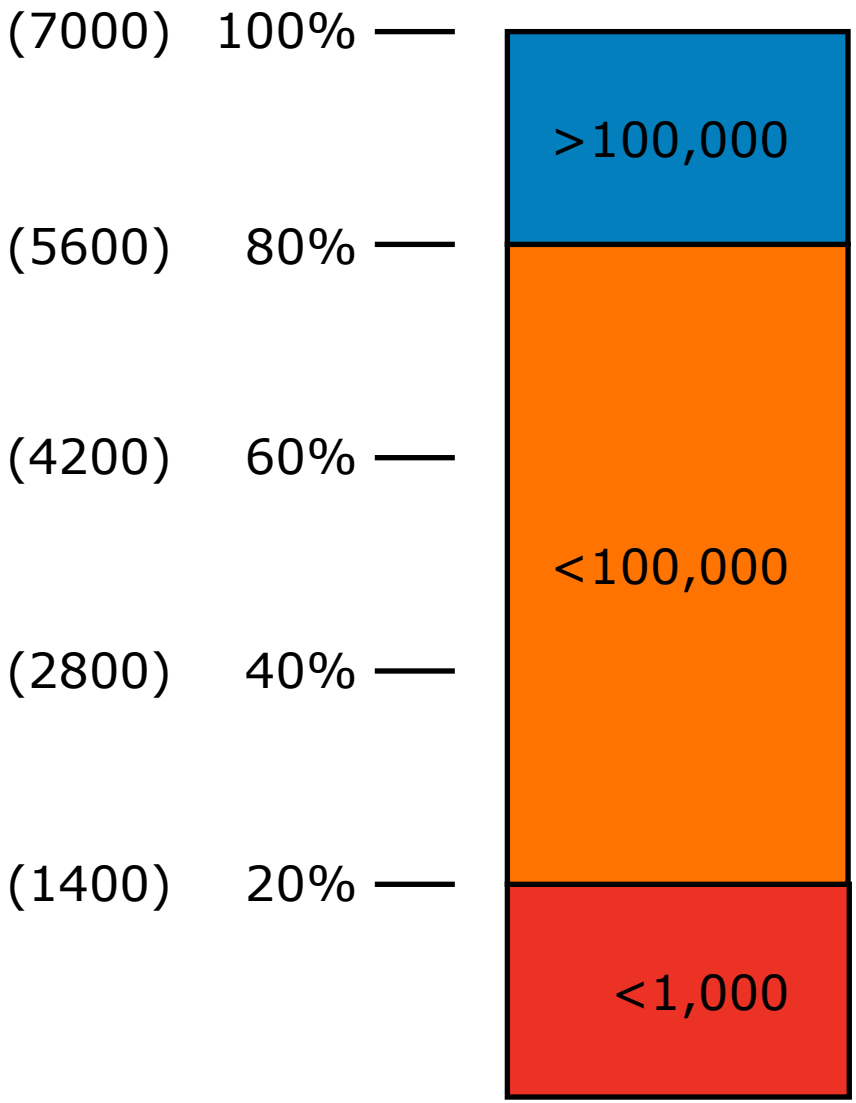
And if we look at a few hundred more languages:

The big 9	3,100,000,000
The next 80	2,400,000,000
The next 300	900,000,000

With just 400 languages, we cover over 90% of the world's population. 6% of the world's languages take up 90% of the population!

http://www.ethnologue.com/statistics/size

Many languages have **very** few speakers



20% of the world's languages have robust populations (100,000 speakers or more).

80% of the world's languages have less than 100,000 speakers (we are including the 20% with less than 1000 here). To put this in perspective, NYU NY has 50,000 students.

20% of the world's languages have less than 1,000 speakers. That is half the size of the NYUAD student body.

Language Death

Why do we care? Languages are dying!

A language is alive if there are living human beings who speak that language natively.

A language dies when the last native speaker of that language dies.

Language death occurs when parents stop teaching their language to their children. This is an issue of power, not about science or linguistics. There is nothing about the grammatical structure of a language that causes its death.

Power comes from politics and economics. Power disparities can be helped along by widespread acts of devastation, such as war, disease, and natural disasters, that either reduce the population, or force the population to move/ merge with another group.

An example from Connecticut

Connecticut used to be home to at least 5 Native American languages.

At least 4 of the languages are extinct: Mahican, Quiripi, Mohegan-Pequot, and Nipmuc.

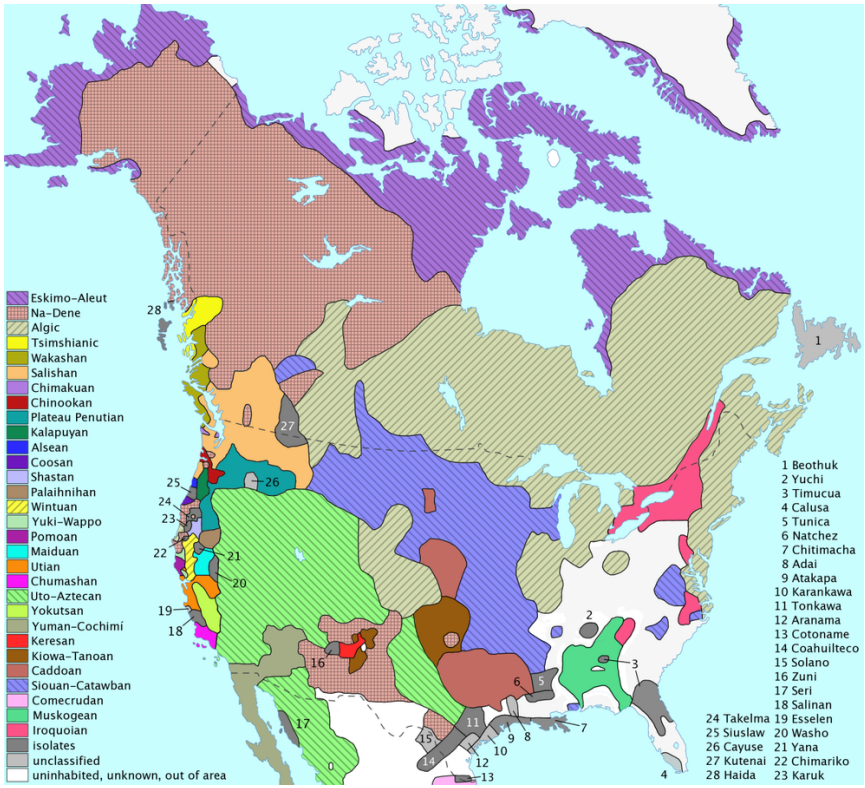
Munsee is only spoken by a few elders that live on a reservation in Ontario, Canada.

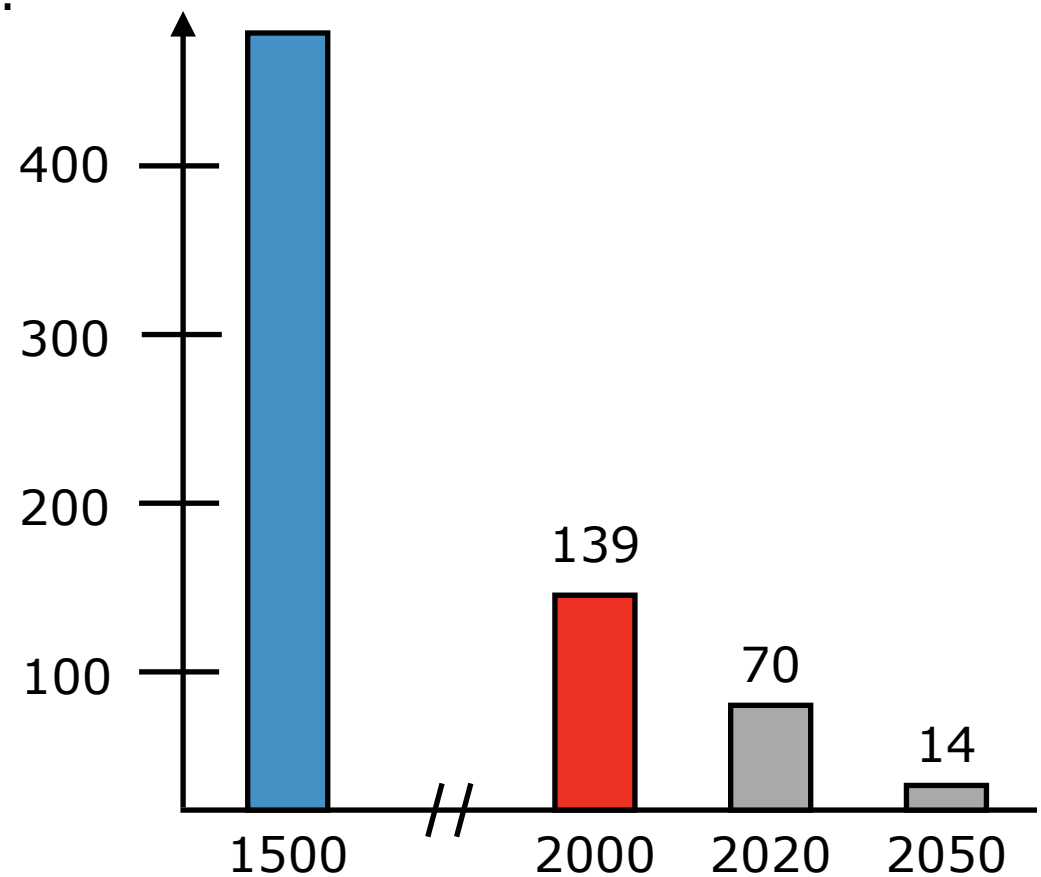
We even know the precise moment that the Mohegan-Pequot language died. It died in 1908 when the last known native speaker died. Her English name was Fidelia Fielding, and her Mohgegan-Pequot name was Dji'ts Bud dnaca.

This is not limited to CT. Most Native American Languages have died or are dying

It is estimated that there were around 500 Native American languages spoken in North America in the year 1500.

This picture shows language families, not individual languages. Each family consists of several languages (the same way that French, Spanish, and Italian are in the Romance language family):





Language Death is caused by humans,
not by (a property of the) languages

What happened?

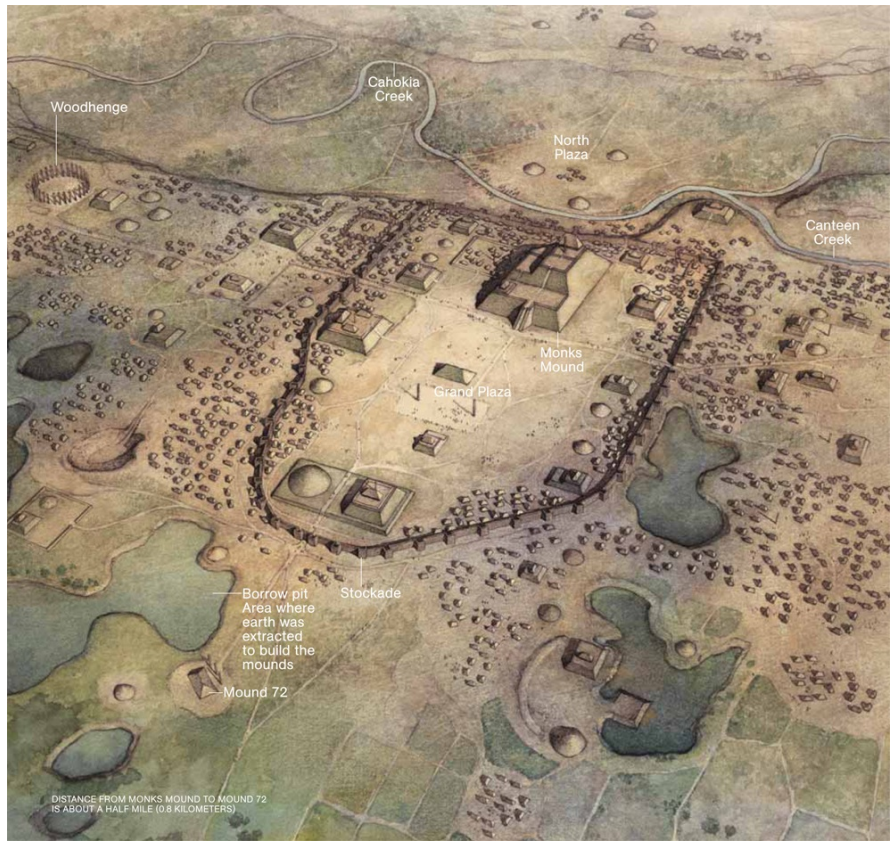
Reports from early European explorers indicate that Native American societies were impressive.

They report impressive cities housing thousands (and even tens of thousands) of residents.

They report advanced forrest clearing techniques (usually with fire) that allowed native societies to create lands that allowed them to cultivate both crops, and large (roaming) bison herds.

They even built giant ceremonial structures, like the earthen pyramid in Cahokia, IL, which covers 14 acres and reaches 100 feet high (surrounded by a city of 40,000 people!)

It is estimated that there were between 20 million and 100 million Native Americans in the US prior to colonization.



Cahokia, IL

What happened?

At least when I was in school, we didn't talk much about the death of 20 million to 100 million Native Americans. I think we were left with the impression that some died in wars, some died of disease, and the rest intermarried with Europeans… it was all very vague.

But the first answer is **disease**. The earliest explorer/conquerors brought Old World diseases like small pox and the bubonic plague with them. This decimated the native population, because they had no natural immunity to these new diseases (this is a National Geographic film based on Guns, Germs, and Steel; we won't watch it here):

https://www.youtube.com/watch?v=iR8OaVDB3_E&index=2&list=PLhzqSO983AmHwWvGwccC46gs0SNObwnZX

By the time colonists began arriving in earnest, the damage had already been done. A number of early colonists remarked at how cities already existed, ready for the colonists to move in (Plymouth rock was a previous Native American town). They remarked at how the forests in the US were open, allowing their carts to pass right through (the Native Americans cleared it!). And how fields were arranged in square shapes, ready for crops to be planted.

The colonists could walk right in and start using the Native American towns and fields because most Native Americans had died from diseases years earlier.

And what happened after that?

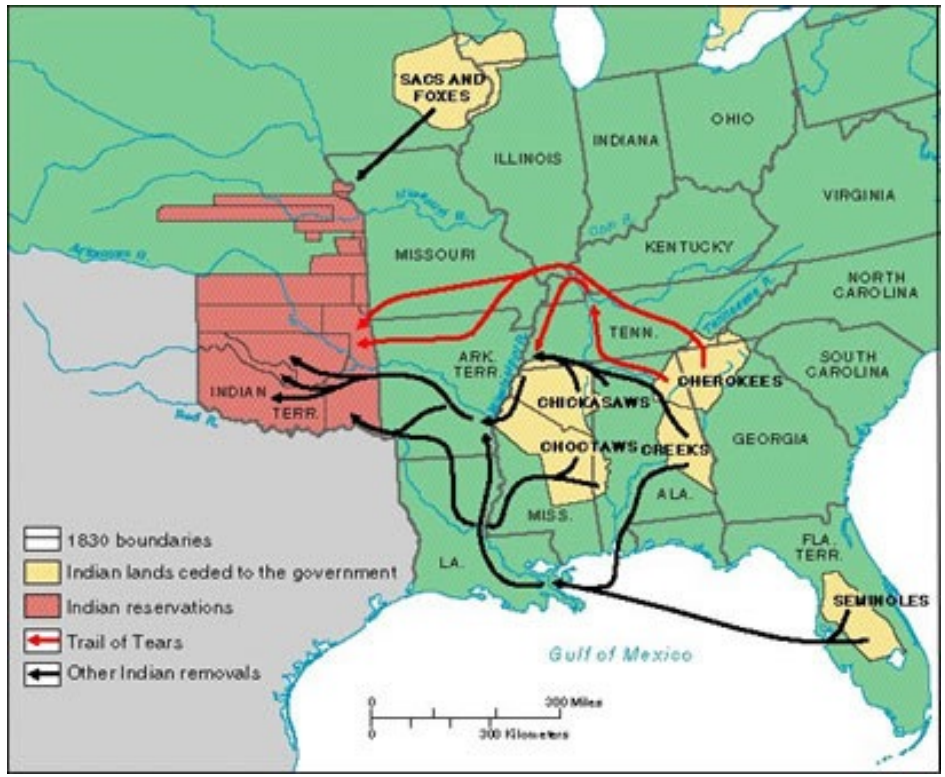
Some researchers estimate that 95% of Native Americans died from the diseases that early explorers/conquerors/colonists brought with them.

But that means that anywhere from 1M to 5M survived. What happened to them?

Forced relocation:

Native Americans east of the Mississippi were forced to "trade" their lands for (less valuable) lands west of the Mississipi (e.g., the Indian Removal Act of 1830).

I put trade in quotes because it was not a trade. Their lands were stolen from them by the government (which used the army to do it).



And what happened after that?

Boarding schools:

From the 1860s - 1960s, Native Americans were coerced into sending their children to special boarding schools. These schools attempted to "assimilate" the children into European culture by eliminating their language, appearance, dress, and customs. They were taught only English, Christianity, and European-centric history.

Carlisle PA c.1900

Carson City, NV
1949

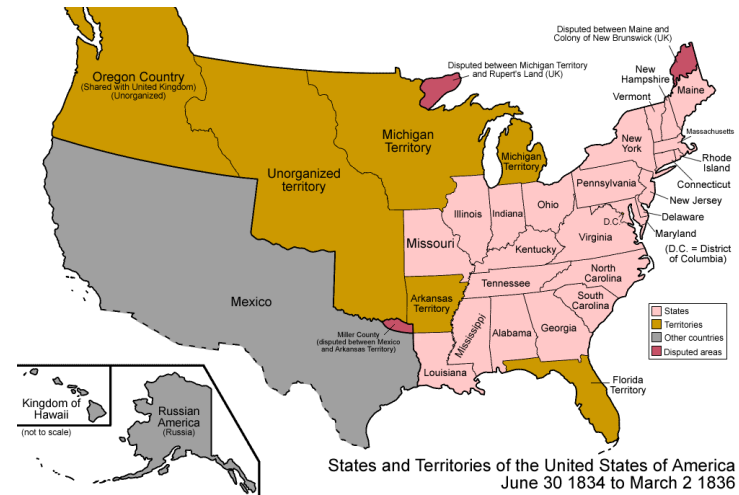
It is not just history. It is today.

There are nearly 5 million people in the US who identify as Native American or Alaskan Native.

There are 574 federally recognized tribes, and 326 federally recognized reservations. (https://www.bia.gov/frequently-asked-questions)

This is a map of recognized land from https://www.data.gov/climate/tribal-nations/tribal-nations-maps

You may notice that there are just a few very small dots in the east. That is the Indian Removal Act (1830) at work. Here is a map of the states (in pink) in the US around 1830.



Language death is still going on, perhaps even faster than ever

We like to think that these periods of our history are over, but they are happening right now, both here in the US, and in other places in the world:

https://vimeo.com/19712297

Languages are dying at an unprecedented rate:

Some linguists estimate that by the year 2100, over 3000 languages will die. That is nearly 50% of the world's languages.

Some linguists estimate that a language dies, on average, every two weeks.

And COVID-19 likely increased the danger to many languages because it disproportionately kills older people, and for many endangered languages, the only speakers left are the elders of the populations.

What do we lose?

Socio-cultural history:

Culture, customs, oral histories, etc.

These are examples of what humanity can do. They are part of what it means to be human. Once they are gone, they are lost forever.

Cognitive scientific knowledge:

Linguistic diversity is a window into the human mind (and brain). Each language is an example of what the human mind can do. Once we lose those examples, they are gone forever.

Language Genesis (Birth)

Language Genesis:
Nicaraguan Sign Language

Sign languages provide a unique opportunity for us to watch languages being born. We call this **language genesis**.

Linguists were able to watch language genesis happen in real time recently in Nicaragua.

pre-1970: Before the 1970s, there were no schools for the deaf in Nicaragua. Deaf children were simply raised by their hearing parents, with no real interaction with other deaf children (or adults).

When deaf children are not exposed to an existing sign language, they tend to create something called **home sign** in an attempt to communicate with their parents and family.

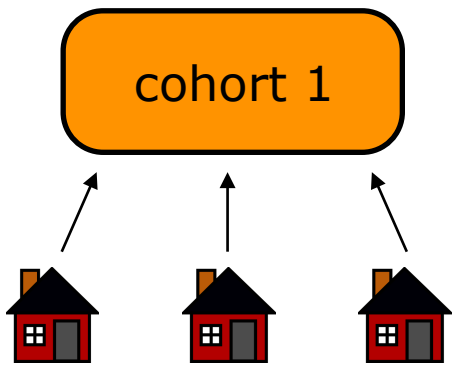
Home sign systems do not have complex phonology, morphology, or syntax. Typically the children develop a hundred or so signs (mostly nouns and concrete actions) to help them communicate their needs to their parents.

Language Genesis:
Nicaraguan Sign Language

In the 1970s, the Nicaraguan government created a special school for deaf children. This brought together all of the deaf children (and their home sign systems) for the first time!

This schools attempted to teach the students **spoken Spanish**. As you can imagine, this wasn't successful. But it is a typical approach of institutions back then (and even in some places to this day).

Outside of the classroom (on buses and in playgrounds), the children taught each other their home signs. As a group they were able to create a more complex system of communication. We call this group **cohort 1**.

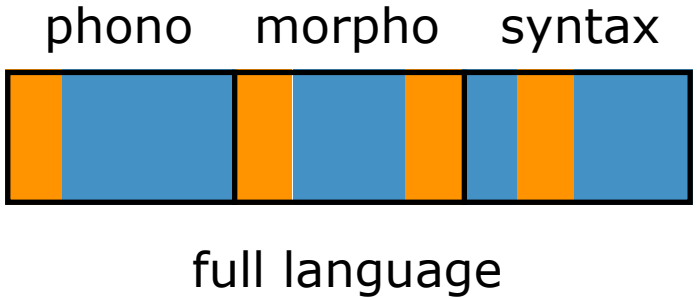


The communication system developed by cohort 1 was more complicated than home sign, but it was not yet a fully fledged language.

Language Genesis: Nicaraguan Sign Language

By the early 1980s, cohort 1 was joined by a new group of deaf children. We call this group **cohort 2**.

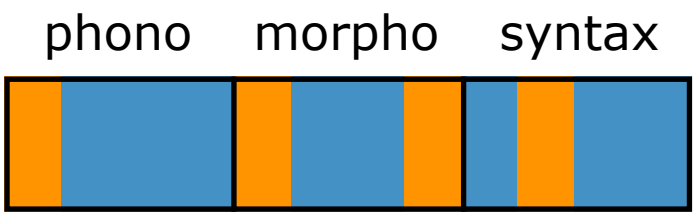
Cohort 2 was made up of very young children (e.g., 5 years old). These children were exposed to the complex sign system of cohort 1 during their critical period… and something amazing happened. They turned the cohort 1 system into a fully fledged sign language with complex phonology, morphology, and syntax!



One way to think about it is that cohort 2 took the cohort 1 system, and filled in the blanks to turn it into a full language!

How did they fill in the blanks?

The mystery here is how the children were able to fill in those blanks!



full language

Remember, the new children in cohort 2 did not have any exposure to full sign languages either. Somehow they were able to fill in the blanks with knowledge that they were never taught!

This is a very compelling argument for innate knowledge. Those children were able to fill in the blanks because they had some innate knowledge of how language works. Now, we don't know whether this knowledge is domain-specific or domain-general… but scientists are working on that.

This is also another example of the critical period. Cohort 2 was able to do this because they were exposed to the cohort 1 system during their critical period. But why couldn't cohort 1 continue to turn their system into a full language? Cohort 1 signers were not able to take their system any further because they aged out of the critical period!

Nicaraguan Sign Language

Here is a talk on youtube by Prof. Ann Senghas, one of the primary researchers working with signers in Nicaragua, about the evolution of Nicaraguan sign language across the cohorts:

https://www.youtube.com/watch?v=FTPGmKoDk0Y&ab_channel=UniversityofCaliforniaTelevision%28UCTV%29

(You don’t have to watch this, I just want to make it available to you!)

Al-Sayyid Bedouin Sign Language

Another example of the relatively recent genesis of a language is the Al-Sayyid Bedouin Sign Language. Al-Sayyid is a village in the Negev dessert in Israel. It has an abnormally high rate of congenital deafness - 150 out of 4000 or so residents were born deaf (due to recessive genes that were carried by the founders of the village).

The first deaf children were born in the 1930s. These children created home sign systems that have evolved into a full sign language (used by both the deaf and some of the hearing residents of the village). Much like Nicaraguan Sign Language, the relatively recent genesis has attracted quite a bit of attention from language researchers.

This is a documentary about one resident's controversial decision to get a cochlear implant for his son. I haven't watched it in detail, so I can't endorse its content. But it contains conversations with signers, so it shows real ABSL!

https://www.youtube.com/watch?v=kBjIIcVMdRQ&t=3383s&ab_channel=FirstHandFilms