Ling 1 Notes

Yash Lala
<yashlala@gmail.com>

Course Logistics

We can attend any class, any discussion on any week. The discussion that we're enrolled in *will* be the one that we're being graded on, so talk to them.

Looks like the discussions just recap the course material.

Week 1: Language and Dialect

Day 1

Language:

1. A tool for communication.
2. A kind of knowledge (because we can't turn it off).

Components of language:

Phonetics
: An inventory of sounds that you have in your language.

Phonology
: Understanding of the sound patterns available in your language (eg: no words starting with "rt" in English.). How sounds interact with each other.

Morphology
: Knowledge of word structure, used to extrapolate plurals, etc (eg: wug → wugs).

Lexicon
: The list of all words you know.

Syntax
: Rules for arranging terminal and nonterminal symbols (CFGs, etc). Can be enormously complex, depends on words.

Semantics
: Assigning meanings to words and sentences, given a syntax tree.

Convention: We notate ungrammatical sentences with an asterisk.

What is necessary for language?

1. Independent of historical knowledge.

2. Independent from ability to write.
3. Independent of education level.

Notation:

Linguistic Competence
: An individual's knowledge of a language.

Linguistic Performance
: What actually comes out of someone's mouth. Affected by nerves, stutters, the body, etc.

Day 2

Universal Grammar
: The set of common properties of *all* human languages. All have lexicons, etc.

What properties do languages have in common?

1. The word form <⇒ word meaning mapping is arbitrary. They're not always independent (onomotopoea, etc).
2. Can assemble words into novel ideas; not just replicating old sentences.
3. Comprised of a set of discrete sounds (we use words instead of whale style oOoOoos / analog encodings).
4. All children acquire the languages of their environment. They do so spontaneously, in different phases.

Classifying approaches to grammar:

Prescriptive
: A formal set of rules, like middle school grammarians. What language "should" be.

Descriptive
: Linguistic model of mental language. What speakers actually *do*.

We focus on descriptive grammars in this course.

Week 2: Language and Dialect

Dialect
: Any variety of language that is shared by a group of speakers.

Language
: A collection of dialects.

Standard Variety
: An arbitrary dialect. Usually the dialect of the ruling class is adopted as the standard. Often has a written form, a formal grammar, and standardized spelling/writing. Taught in public life. *Prestige Dialect*.

Different classifications of dialects depending on originating cause:

Geographical Dialects

Based on geographical origin.

Social Dialects
: Based on social background.

Dialect terms:

Negative Concord
: Need two instances of a negative word to negate a sentence (need "double negatives").

Be-deletion
: AAVE deletes the *be* word. But the grammar otherwise matches.

Discourse Particles
: No direct semantic meaning, but have a pragmatic function (such as attitude). In English, say "well", "you know". In Hindi, "are".

Register
: The level of formality. *Low Register* means casual.

Case study of low register speech; the use of "like".

1. Quotative like: "She was like, *no way*...".
2. Approximative like: "She was carrying like 30 books".
3. Discourse marker like: Tells about connection between sentences. Semicolon style.
4. Discourse / Empathetic Particle like: Tells us we're emphasizing the next word or phrase.

   Linguistic Profiling
   : Using characteristics of a person's speak to identify them as members of a particular group.

Signed Languages

Signed languages
: Use different medium than spoken languages. Visual; use hands, face, and eyes instead of vocal tract and ears.

American Sign Language (ASL)
: Primary signed language in the USA and Anglophone Canada.

About 0.2-0.3% of children are born deaf.

Clearing up myths about signed language:

1. Sign language is not universal. There are around 300. Sign languages have dialects too, eg *Black ASL*.
2. Sign language is not purely iconic (not just acting out the objects). Tends to have more iconicity, but most signs are not guessable. Like onomotopoea.
3. Sign languages do not encode spoken languages. They are entirely orthogonal. Fingerspelling does exist (spells out the letters in English), but it's used mainly as import mechanism.
4. Signed languages aren't just strings of hand motions. They do have grammar.

ASL has inflections, unlike English. Can be used to express frequency of an action, etc.

5 basic parameters of sign language:

1. Shape of hand.
2. Place of articulation (ie location).
3. Movement.
4. Palm orientation (region of hand contacting body, orientation of hand to body, orientation of hands to each other).
5. Facial expressions.

Sign languages have physical analogies to phonetics. Only a restricted set of hand shapes are allowed.

Week 3: Brain and Language

Where in the Brain?

Cortex
: The surface of the brain. Decision making, language resides here.

Cerebral Hemisphere
: One of the two halves of the brain.

Contralateral Brain Function
: Each hemisphere controls the opposite side of your body.

Corpus Callosum
: Joins the two cerebral hemispheres. Allows communication between sides.

Localization
: Theory by Franz Joseph Gall. Proposed that different areas of the brain perform different actions; ie. that the brain is not homogenous.

Lateralization
: The localization of a certain function to one brain hemisphere. Language may be lateralized to the left hemisphere (even the interpreting portions of signed languages).

We believe language is localized to the left hemisphere.

Evidence:

1. **Dichotic Listening**: If you play 2 different sounds to 2 different ears, we hear one sound; the one coming from our right ear (ie. left hemisphere).
2. **Split-Brain Patients**: Communication between both sides of the brain is disabled. They can still draw what they see on their left side, but they can use language to describe only what's on their right side.
3. **Wada Test**: Anesthetize half of the brain. Can't give a linguistic response when we do the left half.

   Acquired Aphasia

Any language disorder that results from brain damage.

Broca's Area
: A brain region near the front of the left hemisphere. Language may be localized here; when injured, it causes Broca's Aphasia.

Broca's Aphasia
: Sometimes called **Agrammatic Aphasia**. Labored speech (called **telegraphic** speech sometimes, because words are dropped). Can't find words — but most interestingly, cannot apply rules of syntax. Require simple syntactic structures (don't reorder subject-verb-object). Semantics are fine. Syntax is broken. Comprehension is mostly OK. Writing like they talk.

Wernicke's Area
: Another region of the left hemisphere. When damaged, it causes Wernicke's Aphasia.

Wernicke's Aphasia
: Sometimes called **Semantic Aphasia**. Very fluent speech that adheres to syntactical rules. Semantics, however, are broken. Word substitutions everywhere, can't fluid nonsense. Eg. "The only thing I can say again is madder or modder fish sudden fishing sewed into the accident to miss in the purdles.". Semantics are broken. Syntax is fine. Comprehension is severely impaired. Writing heavily impaired, write like they talk.

Conclusion: mental language may be comprised of different modules — the syntax vs semantics distinction is supported by different types of aphasia.

We can use fMRI machines to measure brain activity without quizzing aphasics. We ran these tests on speakers of signed languages; it looks like signed languages and spoken languages are processed in the same areas of the brain. Both active Broca's and Wernicke's area.

What happens if a multilingual person has a stroke?

Parallel Recovery
: Both languages recover.

Differential Recovery
: Some languages recover faster.

Selective Recovery
: Only some languages recover.

Successive Recovery
: One language is unlocked only after the other is restored.

Medieval doctors have 2 contradictory rules.

Pitre's Rule
: Not a hard rule, just a heuristic. The most recently used (most frequently + actively) language is recovered first in multilingual patients.

Ribot's Rule
: The mother tongue is restored first. Contradicts Pitre's rule.

Language vs Intelligence?

Is language correlated to intelligence? We look at several disorders.

1. **Specific Language Impairement (SLI)**: Affects some children. Difficulty with speech and language, but similar IQ.
2. **Williams Syndrome**: Affects other adults, genetic. Very fluent with speech and language, but lower IQ. Severely hampers spatial orientation abilities. Ability to feel music emotionally, quick ability to find homonyms (same word, different meaning).
3. **Linguistic Savants**: Low IQ, inability to understand numbers, spatial relationships, or time. Extensive vocabulary, complex sentences, extreme ability to learn different languages.

We conclude that intelligence and language skills are entirely independent.

Additional Disorders (Textbook)

Anomic Aphasia
: "Tip of the tongue" phenomenon affects these aphasics constantly.

Aphasia also tells us about word connections. Brain has ties between semantically related words, and between words of similar sounds.

Acquired Dyslexia
: Any reading disorder that results from brain damage of the left hemisphere.

Deep Dyslexia
: Produce semantically related terms when reading; can get to the correct lexical neighborhood, but can't acquire the correct word (eg. "North" → "North", "South", "East", "West"). Can't understand content words ("which", etc). Never swaps content words and words with meaning (eg. "which" != "witch" ever).

Surface Dyslexia
: Readers have to spell (sound) out words every time.

Additional Discorders (Case Study Reading)

Pure Word Deafness
: Also known as *auditory verbal agnosia*. The brain seems to filter environmental sounds from human speech and process them in separate areas; PWD is a disorder caused by brain damage to the auditory language processing system. Subjects can read and write language perfectly, can speak (self-generated speech) perfectly, but are unable to understand speech. They are capable of recognizing other environmental sounds. This is often caused by damage to both hemispheres of the brain (one can take over in the event of failure).

Auditory Agnosia
: The inability to comprehend and identify environmental sounds, combined with an ordinary ability to recognize speech. The "twin" of PWD.

Week 4: Phonetics and Phonology

When you know the *phonetic inventory* of a language, what do you know?

1. We can chop a stream of sounds into individual words using our *phonetic inventory*.

How is sound produced? We go over the vocal tract (should be able to recognize these on diagram):

Alveolar Ridge
: Top of mouth, behind teeth. Used when we say `t`.

Hard Palate
: Center of roof of mouth.

Soft Palate
: Also called *velum*. Farther back than the hard palate.

Uvula
: Flap of skin that hangs down behind the velum.

Pharynx
: Farther back in throat.

Glottis
: Farther back. Controls air flow; can hear the "stop" when we say "uh oh"

Why not just use spelling to notate the phonetic inventory of a language?

Orthography
: Conventional spelling system of a language.

Phonetics != spelling. The same symbol can represent different sounds when writing. The same sound can be represented by different symbols. Because of these inconsistencies, we use the *International Phonetic Alphabet (IPA)* to write down phonetic inventories. Some facts:

1. IPA is universal. Can be used to represent any human language.
2. IPA symbols are enclosed in square brackets.

We divide sounds into *consonants* and *vowels*. Won't go into precise definition here; generally consonants have some kind of constriction in air flow.

You'll have to remember all of the following sounds: cheatsheet can be found

Consonants

Consonants have three dimensions.

1. Place of articulation.
2. Voicing.
3. Manner of Articulation.

Consonants: Place of Articulation

Articulation
: When part of the mouth (usually the tip of the tongue) interacts with another area of the vocal tract.

There are 7 places of articulation. Going from the front of the mouth to the back:

1. **Bilabials** are produced by bringing both lips together. Eg: "pat", "bat", "mat".
2. **Labiodentals** are produced by touching the bottom lip to the upper teeth. Eg. "five", "view".
3. **Interdentals**: Tip of the tongue is between the two sets of teeth. Eg: "thigh", "thus". IPA symbols are sigma.
4. **Alveolars**: articulated at the alveolar ridge. Eg: "tie", "die", "night", "sip", "zip", "light", "rock".
5. **Palatals**: Raise the front part of the tongue against the hard palate. Eg: "mi_ss_ion", "_ch_eap", "_j_udge", "_y_ou". Weird IPA symbols here.
6. **Velars**: Back of the tongue raises to hit the soft palate (aka velum). Eg: "ba_ck_", "ba_g_", "ba_ng_". Think right at the end of the words.
7. **Glottals**: Produced by flowing air through the open glottis. Eg. "_H_igh". **Glottal Stops** are produced when the vocal cords stop the air at the glottis. Eg: middle stops of the word "uh-oh".
8. **Labio-Velar**: Bilabial and Velar. Produced by pursing lips and raising the back of the tongue to the velum. Eg: "_wh_y".

Consonants: Voicing

Voicing
: Are the vocal folds vibrating (*voiced*), or not (*voiceless open-glottis*)?

Voiceless Sounds
: Vocal folds are apart, air flows freely through the glottis. Eg. "_s_uper".

Voiced Sounds
: Vocal folds are relatively close together, causing vibrations. Eg. "bu_zz_".

Pinch your Adam's apple. You'll see if there's vibration or not.

Consonants: Manner of Articulation

How is the consonant actually produced?

Stops
: There is a complete obstruction of airflow somewhere in the vocal tract. Can be *oral* or *nasal*, depending on where the air eventually vents from. Eg: "p" is oral, "m" is nasal.

Oral Stop
: The velum raises itself to block the nasal cavity. This means that air is eventually vented from the mouth. Can really only hear these sounds when they're followed by something else. Eg. "ca_b_", "rai_d_".

Nasal Stop
: The velum raises itself to block the oral cavity. This means that air is eventually vented from the nose. Don't be confused — the vocal tract is blocked, even though air keeps going from the nose, so it's still a stop. Eg. "ca_m_", "rai_n_".

Aspiration

A brief puff of air that escapes after a stop is released, and before the vocal folds begin vibrating (for the vowel that comes after). Notated by a superscripted `h` in IPA. If you put your hand in front of your mouth, you will feel a puff of air. Eg: "pit", "tale", etc.

Voiceless oral stops are aspirated in **word-initial** (beginning of word) position. Eg: "_p_ool" is aspirated, "s_p_ool" isn't. Like the Hindi difference between "k" and "kh". Aspiration. Indians don't voice the start of words.

Fricatives
: Major, but not total, obstruction of airflow in the vocal tract. This causes a hissing sound. "_f_ish", "_v_ile", "_z_ipper, "_h_ouse", "_sh_oot", etc.

Affricates
: Very briefly stopping airflow completely, then slightly releasing closure. Like a stop, immediately followed by a fricative. Eg. "_ch_eap". "_J_oe" (they pronounce like a "djoe").

Liquids
: Minor obstruction of air, but it still flows through cleanly. Two subtypes.

Lateral Liquid
: Tongue is straight, and air passes around the sides of your tongue. Eg: "_l_eaf".

Retroflex Liquid
: Tip of the tongue is curled backwards, behind the alveolar ridge. "_r_yan".

Glides
: Very small obstruction of airflow. Articulators are slightly close. Two glides in English: "_y_ou", a palatal glide, and "_wh_at", a labiovelar glide. Very vowel-like. After the sound, your tongue quickly "glides" into place to produce the next vowel.

Trill
: One articulator touches another in a rapid-fire manner.

Retroflex Trilled
: India has this. "_r_oti". So does Spanish.

Dialects of the same language have different phonetic inventories.

Vowels

Again, several dimensions. We define 4:

1. **Tongue Position**.
2. **Tongue height**.
3. **Tongue Tenseness**.
4. **Lip Rounding**.

Vowels: Tongue Position

Front
: Eg. "beet", "bit", "bait", "bet", "bat".

Central
: Eg. "about", "but", "cot".

Back
: Back of tongue articulates with oral cavity. Eg: "boot", "book", "boat", "caught"

Say "beet", "caught", "boot". It'll give you a sense for the front → middle → back tongue position transtions.

Vowels: Tongue Height

High
: Eg: "beet", "bit", "boot". Say them: you'll see the tongue position changes in terms of front central back, but the tongue height stays high.

Mid
: Eg: "bait", "bet", "_a_bout", "but", "boat", "bought".

Vowels: Tenseness

Tense
: Pronounced with tense muscles and a narrow mouth. These are slightly higher than lax vowels. Notated in IPA as the standard vowels we learned in schools. Lax vowels are more esoteric.

Lax
: Muscles a bit more relaxed. These are slightly lower than lax vowels.

Lip Rounding

Can be rounded or unrounded. You get the idea.

English is weird here. It doesn't have any 2 sets of sounds that are only distinguished by their rounding. All back vowels are rounded in English.

Some dialects of American English have merged two vowels. This is called the *Caught-Cot Merger*. People west of the Mississippi have lost the distinction between the two — they round both.

Dipthongs
: Combine two vowels. Can also be a vowel + glide combo. There are at least 3 in English; `[aj]`, as in word "eye", word "boy", and "h_ou_se".

Natural Class
: A group of sounds that share particular features. Eg: we can group glottal and noiseless sounds into a natural class. Look for the features they have in common. Useful in understanding the differences between dialects. Can redefine these arbitrarily.

Canadian Raising
: Dipthongs pronouncing "ai" and "ou" in Standard American English are raised. Doesn't occur in all words, only stopped before voiceless consonants.

Transcription
: rendering a word into IPA.

Syllable Nuclei
: Rules as to which sounds can be at the center of a syllable. In English, these are almost always vowels. We also allow liquids and nasals.

In American English, we delete the `[j]` (pronounced "yuh" as in "yuh boi") after a *word-initial alveolar consonant*. Americans say "dew" as "doo".

In New York English, we drop the `[r]` when it's in a word-final position, or when it precedes another consonant. The pronunciation of `[r]` is prestigious, and dropping the `[r]` is perceived as lowerclass, or nonstandard speech. Dropping this `[r]` has empirical links to socioeconomic status and relative formality. *Labov* did some studies in 1963. *Received Pronunciation* is the prestige dialect in British English. If you drop the `[r]`, it's higher class.

In conclusion:

- Differences in dialects are not random. They show phonetical trends.
- Linguistic differences between dialects may be related to non-linguistic factors such as social class. The phonetic markers of social class are arbitrary.
- The more marginalized the speakers, the more marginalized their speech forms become.

What do Phonologists ask each other?

- What's the organization of sounds in a given language?
- What the phonetic context (environment) that tells us when to drop things?

How does phonetics work for signed languages?

- They also have only some permissible hand shapes.
- Individual hand shapes are meaningless. They need to be composed into words.

We can define analogues to phonetic terms for a given set of signs. We call them *Parameters*, or *primes*. We'll talk about 4 parameters.

1. Handshape.
2. Place of Articulations.
3. Movement.
4. Palm Orientation.

Week 5: Morphology

Intro to Morphology

Morphology is the study of words. How are they represented in the minds of speakers?

What do you know when you "know" a word?

- Arbitrary sounds associated with a word.
- Meaning of the word.
- Word's *syntactic category* (noun, verb, etc).
- How to use the word in a sentence.

Knowledge of a word is independent of how the word is written, or its etymology.

Lexicon

Mental dictionary of all possible words.

Lexical Entry
: A single word in the lexicon.

Morpheme
: The smallest linguistic word that has meaning. All words contain at least one. Every morpheme is a discrete unit of language.

Open Class
: A category of morpheme that can be easily added to; eg. Lexical morphemes. We can add new nouns ("email") easily when we learn them).

Closed Class
: A category of morpheme that cannot be easily added to. Eg: nobody makes up new pronouns or prepositions (some people try, but it's strongly restricted).

Multiple types and subtypes of morphemesL

- **Free Morphemes**: Can stand on their own as a word. Eg: "fast" can be used in "faster".
- **Content/Lexical**: We can imagine the contents, they have a lexical meaning. Eg: "manatee". Most nouns and verbs fall into this category. This is an open class. Wernicke's Aphasia kills this, because it serves some syntactic purpose.
- **Functional**: Perform a grammatical purpose in a sentence. Eg: pronouns, articles ("as", "the", "they"), conjunctions ("and", "or"). This is a closed class. Broca's Aphasia kills this, because they serve some syntactic purpose.
- **Bound Morpheme**: Must be attached to a free morpheme. Eg: "-ed". Cannot stand alone; must be bound to a *root*. Also called an *affix*.
- **Inflectional**: Affect the grammatical (syntax tree) properties of the word they attach to. They do not typically change a word's category (noun, verb, etc). They refine meaning; add some information, but don't change core meaning. You can usually attach them to any word in a category (eg: "-ing" can be added to pretty much any verb). We have a list of 8 to memorize; look below.
- **Derivational**: Help derive new word categories. Usually changes both meaning *and* part of speech. Eg: "-ly", "re-", "-ish". Usually less versatile than inflectional, eg: "rerun" works, "retalk" doesn't.

Need to remember the following inflectional morphemes:

- "-s": Plural. He brought "dogs".
- "-s": Possessive. The book is "John's".
- "-ed": Past tense. "Walked".
- "-s": Third person singular present tense. He "runs".
- "-ing": Progressive tense. Johnny is "baking".
- "-en/ed": Past Participle. I have "eaten". Not past tense ("ate"); occurs after the phrase "I have".
- "-er": Comparative. "Bigger".
- "-est": Superlative. "Biggest".

Words have structure. They are comprised of morphemes. We can decompose them into morphemes via a word tree. Every intermediate state of a word tree should have its own meaning. Eg: "unmotivated" becomes "un"
"motivated", not "unmotivate" + "ed" (full decomp: "un" "motivate" "ed"). When decomposing a word into a tree, there are some steps:

1. Break the word down into morphemes.
2. Verify that each intermediate node is its own word.
3. Verify that each bound morpheme modifies the type of the word correctly (eg: "-ing" should apply to verbs. Can't do "dogging").

Some word tree decompositions are ambiguous. This usually corresponds to ambiguity in the word meaning itself. Eg:

- "unlock" + "able" := "You can unlock it".
- "un" + "lockable" := "You cannot lock it".

Affixation
: A process we use to build new words. The process of attaching bound morphemes to a root.

Prefix
: An affix that appears before a root.

Suffix
: An affix that appears after a root.

Infix
: An affix that attaches inside the root.

Circumfix
: An affix that surrounds the root on both sides.

Morphological Analysis

We can look at related sets of words (in an unknown language) to understand how phonemes work. Do note that the absence of a phoneme can also indicate something; indicate such a missing phoneme with $\emptyset$.

Consider the plural morpheme in English. Sometimes it's pronounced like a `s`, `es`, or a `z`. These are *allomorphs* (variant forms) of the morpheme. There are complicated rules regarding when to use these allomorphs.

Sibilants
: Have a buzzing, hissing, quality.

Inflectional morphemes usually come after derivational morphemes in a word. Eg: "paint" + "er" + "s".

Reduplication
: Form new words by duplicating part of an existing word. Eg: in Tagalog, repeat the first syllable to make future tense.

Contrastive Focus Reduplication

A form of reduplication in English. Eg: "I'm not *awake* awake". Puts focus on the most prototypical, or most stereotypical, sense of the word.

In conclusion: When we "know" a language, we know:

1. Words.
2. Morphemes/allomorphs.
3. Properties of morphemes.
4. How to combine morphemes.

We've also discovered:

1. There are multiple types of morphemes.
2. Morpheme inventory is dialect-dependent.

Week 6: Syntax

Syntax
: Study of how phrases and sentences are constructed.

Constituents
: units into which words are grouped. Sentences may have more than one possible grouping, which corresponds to more than one interpretation. Eg: "The little man jumped" has "the little man" as a constituent.

Auxiliary
: The modifiers in a constituent.

When we know a language, we know permissible word orders. Varies between languages. English is SVO (subject, verb, object). Tatar is SOV, etc etc. This is syntax, not semantics. Colorless ideas, blah blah. You know.

Permissible word order is a function of syntactic knowledge. All languages distinguish between the same few syntactic categories. All have nouns, verbs. However, adjectives etc may not exist.

Verb
: Action, state of being.

Noun
: Person, place, thing.

Some people try to statically type these. They fail. Eg: "The assassination of Malcolm X". Noun or verb? Instead, duck type. Syntactic category is determined by how it's acting in the sentence. Expand the parse tree, do type inference.

How do we check if a phrase is a constituent? Three tests — if it passes one, it will pass.

Base cases: Every sentence is a constituent. Every word is a constituent.

1. Replacement tests:
    - *Pronoun replacement*: try replacing with a pronoun. If it's still grammatical and has the same basic meaning, then the substituted phrase was a constituent (noun phrase).

    - *Do so / Do too replacement*: Targets verbs. Replace a string of words with "do so" or "do too". If it works, then the string is a constituent (a verb phrase). Eg: "The old man *found a dollar*" ⇒ "The old man did so". Eg: "Julia *called the governor of Kansas*" ⇒ "Bill did too".

2. Stand Alone test: If a string can stand on its own as the answer to a question, it's a constituent. Eg: "The girl ran in the rain" ⇒ *Who ran in the rain?* "The girl".
3. Movement tests:
    - *Fronting*: If you can move the string to the front of a phrase. Eg: "The girl ran in the rain" ⇒ "In the rain, the girl ran". Great for prepositional phrases.
    - *Clefting*: Can you feed contiguous parts of the sentence into "It is/was X that Y"? Works well for NP and PP, not as much VP. It's the William Riker phrase.

We can break sentences down using these tests. The structure is usually similar for sentences in the same language — we can express this via *phrase structure rules*. Eg: `Sentence -> NounPhrase VerbPhrase`. Now for some rules (parentheses ⇒ optional):

- `Sentence -> NounPhrase VerbPhrase`
- `NounPhrase -> (determinant) (Adjective) Noun (PrepositionPhrase)`. Eg: `(the) (cute) dog (on the couch)`.
- `VerbPhrase -> VerbPhrase PrepositionPhrase` Eg: `(shot the wookie) (with the blaster)`
- `VerbPhrase -> Verb (NounPhrase) (PrepositionPhrase) (Adverb)`. Eg: "carrying him (to the store) slowly".
- `PrepositionPhrase -> Preposition (NounPhrase)`. Eg: "to the store".

Apparently, we just have to modify the *order* of elements on the RHS to draw a tree for another language. TODO: Do they claim the nodes themselves never change?

Recursively break down sentences to create a *Phrase Structure Tree*. We believe that speakers unconsciously decompose speech into these trees.

Some tree terms:

Phrase Structure Tree
: Basically a syntax tree from CS.

Node
: A vertex in the tree.

Domination
: Every node dominates the nodes below it in the hierarchy. If you can go down from X to Y, then X dominates Y.

Sisterhood
: Two nodes that are children of the same parent.

Head
: The lexical item that determines the syntactic category of a phrase. Eg: "An extinct bird species of South America" → "species". Prepositional phrases have the preposition as the heads, and verb phrases have the verb. The head of a noun phrase is a noun.

Complement

Something that gives more information about the head. Complements are usually sisters to the head. Eg: "ate a bagel" → "a bagel".

Structural Ambiguity
: Ambiguity in a sentence that arises due to an multiple possible PST decompositions.

Enumeration of some syntactic differences between dialects:

Habitual Be
: Used in AAVE. Used to express a habit. SAE's "John happy" means he's happy right now; AAVE's "John be happy" means John is generally a happy person. You can also use a different alternative: You can say "John working now", but not "John be working now", that's ungrammatical. It would just be "John be working".

Multiple Modal Constructions
: Used in Southern English. Use different combinations of modals in a sentence. Eg: "We might can go up there on Saturday" (but can't say "can might").

Subject Auxiliary Inversion
: Used in AAVE. SAE says "No dog bit you", AAVE uses "Didn't no dog bite you".

Negative Inversion
: Used in many dialects. If you start something off with "never", you invert. "Can't nobody beat them" works. Try swapping it; "Nobody can't beat them" doesn't make sense. When you swap the two, you invert the second one.

WH movement
: SAE. "Mary will read that book". But if we ask a question, we get "What will Mary read?". First, "mary will", second "will mary". A special case of subject auxiliary inversion in SAE that occurs with words that start with a "wh".

Conditional Inversion
: Again, SAE. You can invert the conditional part of an conditional clause (eg: assertive "You had arrived on time" becomes "Had you arrived on time, you would have X"). Prefixing with an explicit "if" statement, no inversion.

Fuck Inversion
: British English. "John is a nice guy". Response: "Is he fuck a nice guy!". Means he's emphatically not a nice guy. Notice — the "he is" becomes "is he". It's another subject auxiliary inversion.

In short: Subject Auxiliary inversion happens in tons of dialects. It happens differently all the time, and differs heavily between dialects always.

Week 7: Language Acquisition

How do children acquire the first language?

1. *Who is their first teacher?*

   Nobody, really. Attempts to formally educate children just don't play a role in primary language acquisition. It's social, primarily. They just seem to pick it up.

2. *What is the input data?*

Children only hear grammatical sentences from adults. They use this *positive evidences* as training data. They have no negative evidence (only exposed to grammatical from adults!).

3. *How do they interpret novel data?*

   Children are exposed only to a small number of sentences. They learn how to generate many many sentences.

4. *What errors do children make?*

   They make "errors", but they're usually very systematic errors. They understand regular words, but have difficulty with irregular rules — they understand the patterns, but have to be taught the manual (dumb) exceptions.

Conclusion, 4 facts:

1. No instruction.
2. No negative training data is available.
3. Small input, large output.
4. Non adult-like errors.

Called the *Logical problem of Language Acquisition*. Central theme is *Poverty of the Stimulus*; there doesn't seem to be enough data to understand a language.

Three theories:

1. *Imitation Model*: Kids imitate what adults say. Element of truth; even kids that are 10 minutes old mimic the faces the adult makes. It's not conscious, because they don't know they have a face. But we don't think this applies to language, because they generate sentences that they haven't heard ("ungrammatical" ones, in the same patterns). They start from some sort of simplified sentence, instead of mimicking prepositions etc. Fail.
2. *Correctional Model*: Children learn language when people correct the wrong sentences, and reinforce the correct ones. This also fails. Adults correct them inconsistently, and don't reinforce anything in a systematic matter. Statistically adults approve of sentences independently of grammar; they focus on truth of meaning etc. Fail.
3. *Innateness Model*: Noam Chomsky. He believes language has 2 components: an innate component called *Universal Grammar* (biologically hardwired, linguistic categories, etc), and environmental components that have to be learned (how nouns are represented, particular syntax trees, etc — all the stuff that varies between languages).

Some more thoughts:

In villages, there's a cultural idea that adults just don't talk to kids. At all. Unless they have to give a warning, etc. The kids still learn. Inuit moms don't even talk (even as play) to kids, but they still learn. They seem to learn just by listening to adults.

Input must occur before a certain age (*critical period*) — after that, it's hard to acquire language to native levels. This critical period is amazingly powerful. They raised kittens in environments with only horizontal stripes, and they lost the ability to see vertical objects at all. This inability didn't disappear with time. The initial parameters were too tuned. Case study of Victor (raised by wolves) — even after 5 years, he could only say "milk" and "oh my god". Could understand to a limited extent, but that's it. Grunts and howls mostly. Another more recent case ("Genie", child abuse, locked in a closet). By age 17, had vocabulary of a 5 year old. Couldn't use grammatical morphemes, no syntax. Her language processing happened in the

right hemisphere. But again — we don't know if it generalizes. These kids are generally very screwed up. Eg: "Chelsea" was born deaf, but raised as mentally retarded. She was taught language in her 30s. 2000+ word vocabulary, but unable to interpret any syntax.

Similar trends for language acquisition; age of first exposure is a better indicator of accent than number of years speaking the new language.

Similar trends for brain damage; if aphasia is early enough, the right hemisphere can take over entirely.

Week 8: Languages in Contact

Language Contact
: 2 or more languages / dialects come into extended contact with each other for a long time. Forces one of several outcomes.

1. Widespread Bilingualism and Code Switching.
2. Selection of a *Lingua Franca* (used to communicate between people of different languages). Could be one of the languages in contact, could just be some other language.
3. Language Creation: Creation of Pidgins and Creoles.
4. Language Shift / Endangerement / Death: One language is wiped out.

Language Creation

Borrowing
: Copying from another language. Can borrow words, phonetics ("dj" sound is from French), morphology (eg: "-able" from French, "flu schmu" from Yiddish), or syntax ("Why" vs "How Come", Yiddish).

Adstrates
: Languages come in contact, have equal prestige.

Superstrate Language
: Languages of dominant group.

Substrate Language
: Language of less dominant group.

Lexifier Language
: The input language that provides most of the basic vocabulary (lexicon). Usually the superstrate language.

Substrate Language
: The input language that provides most of the basic phonological and grammatical features. Usually the substrate language. TODO: YO YO YO WE SAID SUBSTRATE TWICE

First, we talk about Pidgins and Creoles.

Pidgins

Simplified languages used in specific interactions such as business, service, trade. Not used at home, specifically just to talk to people who normally speak another language. No "native speakers". Grammars are always simpler than those of the source languages. Highly restricted vocabulary, usually no words for abstract concepts (because no need — just doing trade). Not mutually intelligible with their source languages.

Creoles

Once second generation children start to acquire pidgins as their native languages, they turn it into a Creole language, transforming it into a full language with complex grammar. Beware — many Creole languages have the word "Pidgin" in their name. Most present day creoles came about through colonization. List of some Creoles:

- Jamaican Creole English (Patois).
- Tok Pisin (literally Talk Person)

Properties of both Pidgins and Creoles:

- Simplified consonant clusters. Eg: "Strong" becomes "trong", "dust" becomes "dus".
- "Stopping" of fricatives. Eg: "think" becomes "tink", "they" becomes "dey". This tends to be very variable; doesn't happen all the time.
- Cross-linguistic *common* sounds (eg: no implosives, where you breathe in) are preserved.
- Pidgins are never tonal. Tonal creoles are known to exist, but are very very rare.
- Pidgins don't have inflectional morphemes (so past tense suffixes, etc don't exist). Sometimes use reduplication to avoid needing morphemes.
- Pidgins often have flexible word orders. Creoles have more fixed word orders.
- Pidgins and Creoles tend to have Subject-Verb-Object (SVO) orders. Usually true for pidgins, but pidgins are more flexible in terms of orders anyways.
- No complex sentences, no nestings in Pidgin.
- Pidgins don't have definite or indefinite articles ("the mall, a mall"; just "go to mall").

Language Endangerment and Death

About 7000 languages in the world, but distribution is very skewed. 5% of the languages serve 94% of the people in the world. 497 languages have less than 50 languages left. The median language size is 3000 speakers.

Only languages with >100 million speakers:

- Mandarin
- Spanish
- English
- Bengali
- Hindi
- Portugese
- Russian

- Japanese

Language diversity is strongly clustered. Hotspots in New Guinea, Indonesia, India, and Mexico

Colonization kills languages. North American and Indigenous Austrailian are about to die — as are about 90% of the world's languages.

Language Death
: When the last speaker of a language dies.

Language Shift
: Prior to language death, when the language of a community shifts from one to another.

Why do we care?

- We need diversity, analogies to ecosystem diversities.
- We lose access to history about speakers (can no longer do linguistic analysis to figure out who a group traded with, etc).

Why do languages die?

- Earthquake kills the speakers. Disease.
- Economic exploitation, genocide.
- Cultural assimilation.

Language shift and language loss are inevitable without explicit language maintainance by members of the community.

Factors in Language Shift:

- Accept ideas of linguistic diversity, reduce linguistic pressures on minority communities.
- Examine socioeconomic factors. Eg: what language is required for jobs, etc?
- Demographic Factors: Can the speakers isolate themselves from the majority? Eg: Amish speak Pennsylvania Dutch.
- Attitudes towards the language by minority speakers.
- Speakers in minority languages must have say in the educational systems.
- Give endangered language speakers access to tech. Internet allows isolated speakers to communicate with each other.

Week 9: Language Change Over Time

Romance Languages
: "From Rome", evolved from Latin as the Roman army spread across Europe. Italian, French, Catalan, Spanish, and Romanian.

Why do languages become more diverse over time?

Some old folks thought that they start as dialects, then people make them more different to keep a unique identity. This is a false hypothesis. Nobody plans this.

The true process is analogous to evolution. Variation is inherited across generations, and is "caused" primarily by separation of population subgroups. They're not exposed to a centralizing core, so they are free to vary more.

Can languages evolve arbitrarily? No. Language instinct means that some components are innate, and don't change. Other components do change.

- Every language talks about nouns vs words. Every language can express qualities.
- The hierarchical nature of language (syntax trees) doesn't change.
- Speech: The speech stream is a limited inventory of distinct sounds broken into distinct units (syllables). It's *discrete* (no whalesong).
- *Duality of Patterning*: A set of meaningless sounds that we can group together to make meaningful symbols. True for all human languages.
- No language intersperses whistles or claps. We do have clicks though.
- No language would use a single sound for a single word (just pigeonhole here, FFS).
- The specific words do change.

We'll talk about 2 specific changes that happened to English.

Early Middle English Vowel Shortening (1100s): Long vowels become short if they're followed by 2 consonants or 2 unstressed syllables. Eg: "cepan" keeps the long e, "cepte" loses the long e, following the consonant rule.

The Great Vowel Shift (finished by 1400). The forward and back vowels all become higher. The highest vowels start to dipthongize; flat I becomes ay, `u` becomes `aw` (house used to be "hoose", but is now "house"). End result: the vowel glyphs English uses are pronounced entirely differently than other European languages. The correspondence between glyphs and sounds is messed up, English becomes much harder to learn. "house")

Genetically Related Languages
: Languages are genetically "related" if their history can be traced back to a single ancestor.

Comparative Method
: Scientific method to show two languages are related. Requires systematic *sound* similarities.

Mass Comparison Method
: Statistical method looking for resemblances among words (doesn't bother about sound similarities). You have to be careful to avoid culturally related words, stick with basic vocabulary (eg: Sugar was introduced to diets very recently. When the product was borrowed into other cultures, so was the name).

False Correspondence
: False positive in the comparison method.

Missed Correspondence
: False negative in the comparison method.

Trivia: nobody's been able to relate Greek to any other language in the Indo-European language family.

Week 10: The Origin of Language

How did language originate — how did it evolve from our earlier hominid ancestors? This is the theme of the week.

Three main classes of hypothesis.

Cries of Nature Theories

Cries of Nature Theories: Very popular in the 19th century.

Imitation Theory
: Language started by imitating environment sounds.

Spontaneous Vocalization
: Evolved from noises that we made while exerting ourselves. Sometimes called the Pooh Pooh or Yo He Ho theory.

Association with Gesture
: Language developed via association with physical movements (eg: "ta ta" to say goodbye).

Goo Goo Theory
: Language developed by adults imitating baby sounds.

Have obvious shortcomings:

- Language isn't imitative.
- Naming things doesn't account for language structure.
- Language is used in *addition* to cries of nature, not *instead of*.
- Cries of nature and languages are controlled by different areas of the brain.

Social and Psychological Advance-Related Theories

Posit that language developed for social and cultural reasons, as opposed to biological reasons.

- Increasingly complex social organization: Language was needed to support increased densities of humans (how to trade, etc?).
- Cooperative Hunting and Gathering: Needed to coordinate hunts, etc.
- Tool Use:

Have obvious shortcomings:

- Humans somehow developed advanced biological structures needed for language, then suddenly realized that they possessed them? Kinda stupid.
- This would link the properties of languages and the complexity of social and cultural development. This isn't the case. "Changes in the structure of the American Family didn't affect linguistic capabilities..." — Phillip Leiberman.

Darwinian Theories

Based on three observations:

1. Complex systems evolve through incremental changes that have adaptive value.
2. Language is a complex biological system.

Problem: Why don't lemurs demonstrate less complicated language systems, just as flatworms have primitive eyes that detect light or dark?

Answer: Lemurs, etc aren't our ancestors, they're our relatives. Why are animal communication systems and human communication systems so radically different? Because we genocided all of the intermediary systems nearby us; with the extermination of other hominids, humans are the only real power left.

The Darwinian theory is widely accepted.

Evolution of the Larynx

Homologies
: Homologous structures, due to the same origin.

Analogies
: Independently evolved.

Exaptation
: An organ that gets one purpose is adapted for another purpose.

Larynx
: Somtimes called Adam's Apple. Act as the vibrating lips or reed to break up the lungs' uninterrupted stream of air, allows to generate sound.

Evolution of the vocal tract (Larynx) was critical to development of language. Larynx transformed significantly in hominids! Larynx used to be way up in the nose, it made a massive descent down to the throat (gives more space). Comes with great cost; food gets lodged in we choke on our food because of this (Can't breathe and swallow at the same time, the two tubes merge for us. All other animals have separate tubes. Babies have separate tubes, their laryxes descend when they're young).

Lungs evolved from the swim bladder of the fish. *Functional branch point* here — switch from "used to stay upright" vs "used for oxygen processing". We see this transition in the lungfish. Originally, larynx was just a valve to let air into the swim bladder. Now, it's actively involved in letting air into the lungs. See — lungfish, all other animals. Eventually, larynx was used to generate sound as well. See the male frog — it breathes in, then bubbles air through the larynx to make its croak. Gradually, communication was prioritized; horses have a glottis as large as their tracheal opening (maximize airflow, no bottleneck), but humans have way smaller tracheas. We took the tradeoff for speech vs aerobic efficiency.

Compare the larynxes of hominids:

- Chimpanzees: *Standard Plan*. Larynx behind tongue, up high. Tongue is very flat in the mouth, they can breathe and swallow at the same time.
- Neanderethals: *Near Standard Plan*. Communicated, but slower, with more errors. No round tongue, neck much shorter (we have long neck to get the larynx more space).

Controversial.

- Standard Human: We've seen this.

How do we know the biology of earlier hominids?

1. Reconstruct from fossils.
2. Homologous organs.
3. *Ontogeny Recapitulates Phylogeny*. Development of an individual organism from youth tends to mirror the development of the species (eg: we have gills when we're small).

Evolution of the Brain

We have to look at monkeys. There are no brain fossils.

Compare Squirrel monkeys with Vervet monkeys (New World vs Old World). Squirrel Monkeys are born with intrinsic vocal repertoire. Their communication refers mainly to their emotional states. Vervet Monkeys have to learn from their elders, and get a more complex vocabulary as they age. They can describe their surroundings.

Campbell's Monkeys (Old World, West African Tropical Forest) have 4 main calls. Three indicate predator sounds. They also have a fourth that can be used to negate (indicates that it's a non-predator alarm), a suffix ○○ that changes the meaning of the word (seems to reduce urgency), and the order of combined calls is always deterministic. They can also combine these calls to refer to more complex situations; chaining a particular three refers to "falling wood".

We can also use the ontology vs phylogeny mapping. Trace evolution of speech as they grow: first they babble, then they have short phrases of related words, then they start to use syntax, etc.

1. Babbling.
2. One-word phase.
3. Two-word phase.
4. Telegraphic: they talk about what's going on. "Can't push it".

We believe that the primates lack a *Theory of Mind* — the ability to attribute mental states to ourselves and others, the ability to understand that others have different desires. Seems necessary to use languages. Remember the television cubby experiment. Usually occurs around 4 years old in humans.

But later on, we notice hominids spending immense amounts of time creating beads (useless for survival). Suggests some cognitive elaboration that we believe was for communicative reasons.

Noam Chomsky and Steven J Gould doesn't like this Darwinian Theory. He doesn't think that everything has to arise step by step, with some sort of obvious incremental adaptational ability. Both think it's an emergent property of brain density, that end up being useful for language.

Finals Review

Acronym for Constituency Tests: FC PDS (Fuck Police Departments, #NWA, etc etc)

- Fronting
- Clefting
- Pronoun
- Did so
- Standalone

Interesting notes:

- Added to the syntax tree. Practice without looking at it?
- There's a new rule for VP: `VP -> VP PP`. So the NP VP PP sentence is actually a 2 level tree...they pack it under the verb for some reason. TODO: make acronym for tree.

Last updated 2022-09-09 09:45:26 PDT