Introduction to the language classification, universals, and typologies

- There are lots of languages in the world—between 4000 and 8000 (depends on what a language is vs. a dialect)

- These languages can be classified into subgroups a couple of ways:
  - Similar structural characteristics
  - Genetic relatedness
  - Areal proximity

- However, these methods of classification divide languages differently:

  **LATVIAN** (Indo-European; genetically related to ENGLISH)
  jāizgudro
  jā -iz -gudro
  must out figure
  ‘(One) must figure (it) out.’

  **SWAHILI** (not genetically related to ENGLISH)
  Maria anampenda Anna.
  Maria likes Anna
  ‘Maria likes Anna’

  **Point:** LATVIAN and ENGLISH though related differ structurally
  (A whole sentence can be expressed in a single word)
  SWAHILI and ENGLISH through unrelated share a structural similarity
  (Standard order of a sentence is subject verb object)

In this course we will be concerned with **structural classification**, or **linguistic typology**.
- Groups together languages that have similar structural characteristics:
  - Word order patterns
  - Word structure
  - Phonological systems
  ...

---

Certain structural characteristics turn out to hold of all languages: **absolute universals**

...or of most languages: **universal tendencies**

But most language universals are **implicational universals** which state that if a language has a certain structural property X, it will also have a second structural property Y.

Some quick examples:

- All languages have stop consonants (e.g., /p, t, k/ — no language lacks all three)
- Just about every language has the vowel /a/.
- If a language has nasal vowels (like FRENCH), it will have non-nasal vowels too.
- If a language allows syllables which start with sequences of consonants, it will also allow syllables with one or no initial consonant (*stripe, tape, ant*) [no language will have just *stripe*]
- If a language has only suffixes (no prefixes), it will have only postpositions (not prepositions)

We can classify languages along many different dimensions; here are a couple:

**Phonology**
- Vowel inventory
- Consonant inventory
- Tone and prosodic (stress) system
- Syllable types

**Morphology**
- Word structure

**Syntax**
- Word order (SVO [ENGLISH]; SOV [JAPANESE], …)
- Grammatical hierarchies (verbal agreement)

In each case, certain universals (absolute, tendencies, or implications) can be stated.

**Morphology.** Categories of word structure

- *Isolating* (or *analytic*) languages express each morpheme as its own word:

  **Mandarin**

  Ta chi fan le
  he eat meal past
  ‘He ate the meal.’

  Ta chi le fan.
  he eat past meal
  ‘He ate the meal.’

  (Compare ENGLISH, where past tense is part of the verb: *walk* vs. *walked*)
• Polysynthetic languages express entire sentences in one word

INUKTUT
Qasuirarsvigssarsingitlunarnarpuq.
Qasu -iir -sar -vig -ssar -si -ngit -luinar -nar -puq
tired not cause-to-be place-for suitable find not completely someone 3/sg
‘Someone did not find a completely suitable resting place.’

• Somewhere inbetween are agglutinating languages and fusional languages
  Words have several morphemes, generally a root and affixes

—Agglutinating languages have one function for each affix:

TURKISH
köy ‘village (singular)’
köy-ler ‘village (plural)’
köy-ler-in ‘of the villages’ (genitive plural)

—Fusional languages mark several functions with a single affix:

RUSSIAN
mi vid’im ruk-u
we see hand-fem/sg/Acc
‘We see a/the hand.’

Of course, most languages don’t fit completely in any of these types;
ENGLISH can be isolating (I will leave), agglutinating (re-en-act-ment),
fusional (him = masculine, third person, singular, direct object).

This probably means that this isn’t a good way to split languages
(although it is an established terminology)

Syntax
The most discussed classification system is one based on word order.
The order of the subject (S), verb (V), and direct object (O)
in a “normal sentence” (the basic order).

• The most common orders are SOV, SVO, and VSO
  Over 95% of the world’s languages have one of these as the basic order

TURKISH
Hasan öküz-ü al-di. SOV
Hasan ox-acc bought
‘Hasan bought the ox’
**English**
The athlete broke the record.  

**Welsh**  
Lladdodd y ddraig y dyn.  
‘The dragon killed the man.’

• Concerning “basic order”:  
  **English** is an SVO language: *John ate lunch.*  
  But not all sentences have this order, for example:  
  *Coffee I’ll drink, but tea I won’t.*  
  (that’s OSV)  
  Idea: This is a “marked” word order, associated with a “non-neutral”  
  meaning. You can only say *Lunch John ate* in certain special  
  situations (where you mean to contrast lunch with something else,  
  which John presumably didn’t eat).  
  *However* there are languages that allow very free re-ordering, in which  
  several different word orders are basically synonymous. **WALPIRI**  
  is one celebrated which allows almost any word order. What, then,  
  is “basic”?  

• There are a small number of VOS languages  
  **Malagasy**  
  Nahita ny mpianatra ny vehivavay.  
  ‘The woman saw the student.’

• And a *very* small number of OVS or OSV languages  
  **Hixkaryana**  
  Kana yanimno biryekomo.  
  ‘The boy caught a fish’  

  **Apunrinâ**  
  Anana nota apa.  
  ‘I fetch a pineapple’

**Some word order universals**  
• If a language has VO word order (SVO, VOS),  
  it will have prepositions (not postpositions)  
• If a language has OV word order (SOV, OVS),  
  it will *probably* have postpositions (not prepositions)
• PPs (prepositional phrases, or postpositional phrases) almost always precede
  the verb in OV languages, and usually follow the verb in VO languages.
• Manner adverbs (like fast) overwhelmingly precede the verb in OV languages,
  and generally follow the verb in VO languages.
• In possessives, OV languages overwhelmingly have possessor-noun order;
  VO languages usually have noun-possessor order.
  (possessors are usually called genitives).

English (SVO) has both genitive-noun and noun-genitive orders:
  Mary’s friend genitive-noun
  A friend of Mary(’s) noun-genitive
  (But the universals are tendencies; allowing for exceptions)

**Grammatical hierarchies**
Some implicational universals are best stated in terms of hierarchies.

A common hierarchy is the “accessibility hierarchy” of nouns:

\[
\begin{array}{ll}
\text{most accessible} & \text{subject} \\
\text{least accessible} & \text{direct object} \\
\end{array}
\]

For such hierarchies, if a grammatical process applies to one member, it will
also apply to things higher on the hierarchy. For example, **agreement**:

• In many languages, the verb agrees (is marked with features of) the subject:

  **Spanish**
  - Juan partició
  - Juan leave-3/sg/past
  - ‘Juan left.’

• In some languages, the verb agrees with the object too:

  **Swahili**
  - Juma a-li-wa-piga watoto
  - Juma 3/sg-past-3/pl-hit children
  - ‘Juma hit the children’

• But no language agrees with the object and **not also** the subject.

_Later in the course_, we will talk about relative clauses with respect to a hierarchy
of this kind.
In a hierarchy like this, the direct object is more marked than the subject. Hierarchies are used to express degrees of markedness, where the existence of a (relatively) marked option in a language implies the existence of a (relatively) unmarked option. Markedness is also often measured in terms of frequency. More languages will take unmarked options than take marked options; or, within a language, a language will more often use an unmarked option than a marked one.

**Case marking and subjects**
Nouns in many languages are “case-marked”, meaning that it is inflected differently depending on whether it is, for example, a subject or an object. ENGLISH pronouns show case-marking:

She believed me. & She (3rd, singular, feminine) has subject case (nominative), me (1st, singular) has object case (accusative).

I believed her. & I (1st, singular) has nominative case, her (3rd, singular, feminine) has accusative case

And in many languages, this applies to every noun (not just pronouns).

**Nominative/accusative languages.**
ENGLISH is a nominative/accusative language, in which the subject of an intransitive verb is treated the same as the subject of a transitive verb, and different from the object of a transitive verb: (specifically, nominative case in the first two, accusative case in the third)

Mark believed her.

She ran.

She believed Mark.

**Ergative/absolutive languages.**
Another class of languages, **ergative languages**, treat the subject of an intransitive verb and the object of a transitive verb the same, different from the subject of a transitive verb. BASQUE is one such language, many Australian languages are also ergative. Paraphrasing, this would be like:

Mark believed her.

Her ran. (or maybe even ran her).

She believed Mark.
Explaining Universals

Typological studies have isolated certain uniformities across languages. But why do languages share universal properties (or conform to universal implications)?

Functional pressures, phonology

- Many explanations in phonology rely on functional pressures;
  - vowel systems maximize perceptual distance, /a i u/ are maximally “far apart” phonetically
  - languages with complex consonants like [tf] always have the simpler consonants [t] and [f]; if a language has the “harder” (more marked) consonants, it will also have the “easier” ones.

Genetic origin

- One tempting possibility is to suppose that properties of an older language got “passed on” to historically related languages. But it is hard to make this argument work, since genetically related languages differ structurally/typologically and typologically similar languages are not genetically related.
- Later in the semester, we’ll discuss creoles, languages for which there is (claimed to be) no “parent language”, yet the universals hold of these languages too.

Facts about the world and/or physiology

- Later in the semester, we’ll see a hierarchy of color terms that seems to hold across languages. But is this a fact about language or is it a fact about perception (i.e. how our optical system perceives color)? Deep questions… are these different things? Can similar explanations be advanced about the correlation between the order of possessors and nouns and the order of verb and object?
- Is the tendency for the subject to come before the object due to the fact that the subject is usually the causal agent, conceptually prior? If so, how can languages differ?

“Language instinct”

- People are genetically “wired” to acquire and use language; it’s part of the structure of the human brain, and that structure imposes certain restrictions on what form languages can take. The universals we observe give us clues to the structure of this innate language system.

The basic program for language universals:

- Typological studies provide generalizations and universals (provided the typological categories are right)
- Theories and analyses of language(s) can be evaluated with respect to how well they predict the universals to hold.