Linguistics 107—Intro to Indo-European

Lectures by Jay Jasanoff
Notes by Max Wang

Harvard University, Spring 2013

Introduction

Linguistics 107 is one of two introductory historical linguistics courses offered at Harvard University. The course covers the basics of Proto-Indo-European grammar, and provides a comparative survey of all the modern Indo-European language families.

These notes were live-TEXed, then edited for correctness and clarity. I am responsible for all errata in this document, linguistic or otherwise; any merits of the material here should be credited to the lecturer, not to me.

Feel free to email me at mxawng@gmail.com with any comments.

Acknowledgments

In addition to Jay for his ever-amusing lectures, I thank TF Laura Grestenberger for various clarifications to these notes (which cover lectures only; discussion and review sections are not included). Thanks goes also to several classmates whose notes I cross-referenced in the revision process.

Acknowledgment goes also to Zev Chonoles, whose online math lecture notes (`http://math.uchicago.edu/~chonoles/expository-notes/`) inspired me to post my own. I have also borrowed his format for this introduction page.

The page layout for these notes is based on the layout I used back when I took math notes by hand. The LaTeX styles can be found here: `https://github.com/mxw/latex-custom`.

Copyright

Lecture 1 — 1/28/13

A sample of IE language families include:

- Latin: French, Spanish, Italian
- Proto-Germanic: English, Dutch, German
- Proto-Indo-Aryan: Marathi, Hindi/Urdu, Gujarati, Panjabi
- Proto-Slavic: Russian, Polish, Czech
- Proto-Celtic: Irish, Scots Gaelic, Welsh

Lecture 4 — 2/4/13

	Latin	Greek	Sanskrit	PIE
"is"	est	esti	asti	*esti[1]
"I"	ego	egō	aham	
"eight"	octo	octō	aṣṭā	*oktō
"bone"	os	osteon	asthi	
"field"	ager	agros	ajra-	*agro-
"drive"	agō	agō	ajām	

PIE was originally reconstructed with an initial 'a' for all of the words above, since Sanskrit had mostly a's and was thought to be closer to PIE due to its age. However, this could not be explained with regular sound change—and sporadic sound change does not actually occur in reality.

Besides sound change, analogical change may occur when an older word deviates too much from a newer language system. For instance:

- Children sometimes say "foots" or "feets", basing their choice on ingrained rules for pluralizing English words.
- In OE, "book" pluralized as "beech", but in modern English, it pluralizes as "books".
- Houses is the only word in English to pluralize as /haʊzɪz/, although recently, it has been repaired analogically to /haʊsɪs/.

Indo-European Sound System

IE was stop-rich and fricative-poor.

Stops

	labial	dental	"palatal"	"velar"	labiovelar
vcls.	*p	*t	*ḱ	*k	*kʷ
vcd.	*b	*d	*ǵ	*g	*gʷ
vcd. asp.	*bh	*dh	*ǵh	*gh	*gʷh
			velar	uvular	

[1] An asterisk (*) denotes that a word is reconstructed.

Lecture 5 — 2/6/13

Consider the word for "brother" in these IE languages:

Skt.	Lat.	Grk.	Gmc.	Celt.	Slav.	PIE
bhrātar-	frāter	phrātēr	brother	bráthair	brat	*bh-

b^h is linguistically uncommon, and it's difficult to pronounce it accidentally—$b^h > b$ is much more likely. In Latin, the likely sound change was $b^h > p^h > f$.

PIE had only one type of fricative: /s/. When /s/ is next to another voiced consonant, it was also voiced as /z/ (more generally, this occurred whenever it was easier to pronounce a /z/). /s/ and /z/ here are allophones—different realizations of the same phoneme. We write this as s ∼ z.

PIE had two nasals, m ∼ m̥ and n ∼ n̥. Note that these are allophone pairs; for example, in PIE, we have

root	sg. acc.
*deiwo	*daiwom
*pod	*podm̥

where m ∼ m̥ is the accusative singular ending. PIE also has two liquids, r ∼ r̥ and l ∼ l̥, and two glides, y (or i̯) ∼ i and w (or u̯) ∼ u.

PIE had five vowels: two high vowels (i and u), two mid vowels (e and o), and one low vowel (a). These came in short (e.g., a) and long (e.g., ā) forms, which differed only in length.

Skt.	Lat.	Gmc.	PIE
pitar-	pater-	patēr	*-a-
shita-	status	statós	*-a-

PIE had diphthongs (i.e., two or more vowels forming one sound). Three notations are in use, e.g., for the English pronunciation of light, we might see in a reconstruction any of [lait], [layt], or [la i̯t].

Finally, PIE had an accent, which could stand on any syllable, e.g.,

English	PIE
brother	*bhrā́-
father	*patḗr

Morphology

In PIE, morphological changes relied heavily on suffixes (prefixes are rare), as well as ablaut, or vowel change. For instance, for the word "to ask", we have

grade	PIE root	Latin ex.
e	preḱ	precēs
o	proḱ	procus
0	pr̥ḱ	—

Conventionally, we refer to a root by its e-grade form. Consider another example, meaning "to proclaim solemnity":

PIE root	English ex.
seng^{w}h-	sing
song^{w}h-	sang

Lecture 8 — 2/13/13

Indo-Iranian

The Indo-Aryan (or Indic) and Iranian languages are more closely related than most branches of the Indo-European family—we thus group them together as one subfamily, Indo-Iranian. The word Aryan comes from *Ărya-, meaning peoples. The word Iran has a similar source—it comes from aryānām daḥyauš, meaning land of the Aryas.

Sanskrit, an Indo-Aryan language, is general accepted to be the most archaic of the attested Indo-European languages.[2] The oldest known form of Sanskrit was Vedic Sanskrit (sometimes just "Vedic"). The word "Veda" meant "body of knowledge" and referred to texts preserved by memory and an oral tradition. The basis for much linguistic context about Sanskrit comes from the r̥g-veda, a set of Vedic hymns.

The word "Sanskrit" comes from the Sanskrit saṃskr̥ta, meaning "perfect". Sanskrit developed further into a number of "middle Indic" languages, called Prakrits (meaning "assembled") before evolving into the modern Indic lanugages.

Indo-Iranian Sound Changes

In PIE, *e and *o are common, whereas *a is not. However, in Sanskrit, we see:

Lat.	Skt.
est	asti
octo	astā
ager	ajraa

[2]The most archaic modern IE language is Lithuanian.

In Indo-Iranian, and in no other Indo-European branch, *e, o, a > a.

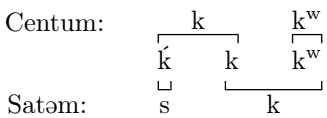
Consider also the word for "hundred" in the following IE languages:

PIE	Lat.	Sardinian	Skt.	Av.	Lith.
ḱm̥tóm	centum	ĥm̥tóm	śatam	satəm	šimtas

This represents a huge linguistic divide in the Indo-European language families. In Latin and other centum languages, the ḱ and k merged together as /k/, and the k^{w} remained a separate phoneme. In the so-called satəm languages (named after the Avestan word), the "palatal" ḱ pushed forward to become /s/ or /tʃ/ and the k^{w} became a /k/. Illustrating this latter change, we see

PIE	Lat.	OE	Eng.	Skt.
k^{w}od	quod	hwæt	what	kad

Pictorally, we have



This centum-satəm split divides the IE family roughly as follows:

Centum	Satəm
Italic	Indo-Iranian
Germanic	Balto-Slavic
Celtic	Aremenian
Greek	Albanian
	Tocharian

In Sanskrit, the grapheme 'c' is pronounced as /tʃ/. But this gives us an apparent inconsistency:

	Skt.	Lat.	PIE
100	śatam	centum	ḱ-
five	pañca	quinque	
	-ca	-que	k^{w}e-
what	kad	quod	k^{w}o-
	kravin	cruor	k-

However, based on the vowel difference betwen -que and quod, we were able to conclude that in the presence of *front vowels*, velar consonants were palatalized.

$$*k^{w}e > *ke > *če > ca$$

Note that the vowel change occurred last.

Lecture 10 — 2/20/13

So far we have seen three important sound changes in Indo-Iranian:

1. "Satəm-shifting": ḱ > ś and k, kʷ > k.
2. Law of palatals: k, g > č ǰ / _e, i
3. Vowel merging: a, e, o > a

As an example, consider the sound change for *pekʷeti:

$$*pek^{w}eti > peketi > pečeti > pacati$$

The same principle fails for *pekʷonti, however:

$$*pek^{w}onti > pekonti > pekonti > \sout{pakanti}$$

Instead, we have the following paradigm, which results from proportional analogy transforming k > c:

pacámi	pacámah
pacasi	pacatha
pacati	pacanti

Sanskrit also presents reduplication in present and perfect verb conjugations:

PIE	Skt.	pres.	perf.
deh$_3$	dā-	dadāti	dadau
	tr̥d-	tatarda	
	takṣ-	tatakṣa	

For the Sanskrit root dhā-, we would thus expect dhadhāti in the present—but instead, get dadhāti:

PIE	Skt.	pres.	perf.
dheh$_1$-	dhā-	dadhāti	dadhau
	dhr̥ṣ-[3]	dadharṣa	

This sound change is known as Grassmann's Law. When two aspirated consonants appear side-by-side, the first one loses its aspiration:

$$C^{h}\ldots C^{h} > C\ldots C^{h}$$

Grassmann's Law is, more generally, an example of dissimilation.

A few more examples of Indo-Iranian sound change:

	Sanskrit root		
Change	kr̥-	gam-	han-
(PIE)	*kʷekʷór-e	*gʷegʷóm-e	*gʷhegʷhón-e
1	kekore	gegome	gheghone
2	cekore	jegome	jʰeghone
G.L.	—	—	jeghone
3	cakāra	jagama	jaghana

[3] English: durst

Later in its history, Sanskrit saw jh > h in certain situations. For instance, *gʷn-énti, the 3 pl. of the PIE root *gʷen-, becamse ghnanti. However, the 3 sg. form under went the following changes:

$$*g^{w}hén\text{-}ti > ghenti > jhenti > jhanti > hanti$$

Finally, we consider the ruki rule[4], in which

$$s > ṣ \;/\; r\_,\ u\_,\ k\_,\ i\_$$
$$z > ẓ \;/\; r\_,\ u\_,\ k\_,\ i\_$$

that is, s and z become retroflex (via s > š > ṣ) after r, u, k, or i. Furthermore, ṣ and ẓ made neighboring dental stops into retroflex dental stops, e.g.,

$$dhr̥ṣ\text{-}ta > dhr̥ṣ\text{-}ṭa$$

Sanskrit also eventually eliminated all voiced sibilants, e.g.,

$$*ni\text{-}zdo\text{-} > niẓda\text{-} > niẓḍa\text{-} > nīḍa\text{-}$$

Sanskrit thus acquired a full set of retroflex dental stops which contrasted with the regular dental stops, e.g., in the minimal pair

$$sapta > satta > šat$$
$$ṣaṣṭi > ṣaṭṭhi > šaṭ$$

Sanskrit loan words from Romance languages use the regular dental stop, whereas English loan words used the retroflex stops—the English /t/ is alveolar rather than dental.

Finally, we briefly consider a phenomenon of syntax in Sanskrit regarding preverbs, or place adverbs. In Latin, these were originally used as separate modifiers of verbs, but eventually attached to the words:

dūcō	ductus
conducō	con-
adducō	ad-
abducō	ab-

English examples include such words as forego, withhold, and understand. In older Sanksrit, however, preverbs could still completely separate from their verbs, a phenomenon known as tmesis.

[4] "Ruki" was a Russian mnemonic, meaning "hands"

Lecture 12 — 2/25/13

In addition to the three sets of Proto-Indo-European stops—voiceless, voiced, and voiced aspirate—Proto-Indo-Iranian and the Indo-Iranian languages additionally had a set of voiceless aspirates:

p	t	k
b	d	g
bh	dh	gh
ph	th	kh

It is now generally accepted that PIE only had the first three sets of stops. Occurrences of voiceless aspirates are rare in PII and the etymology is thought to be based in consonant clusters ending in laryngeals rather than a PIE voiceless aspirate. For example, the 2nd sg. perf. -tha in Sanskrit is believed to come from

$$\text{vettha} < \underset{\smile}{\text{u}}\text{oid-th}_2\text{a}$$

We now turn to the study of the other half of the Indo-Iranian branch of IE—the Iranian languages. These include:

E. Iranian	W. Iranian
Avestan	Old Persian
Pashto	Middle Persian
Ossotic	Modern Persian
	Balochi
	Kurdish

Old Persian was written in a simple cuneiform and was the first Iranian language to be deciphered; however, the attestations we have are very formulaic (e.g., basic inscriptions by kings in royal courts). Avestan was much more linguistically energetic, and in general, Iranian languages were much less conservative than Sanskrit.

Iranian Sound Changes

Iranian, unlike Sanskrit, developed a very rich set of fricatives at the expense of the elaborate stop system. One basic change involved the loss of voiced aspirates in favor of voiced stops.

$$\text{bh, dh, gh} > \text{b, d, g}$$

For example,

Skt.	Av.
bharati	baraiṭi
dā	dā-
dha	dā-

Before consonants, voiceless stops would become fricatives

$$\text{p, t, k} > \text{f, θ, χ / \_C}$$

for instance,

Skt.	Av.
pra-	fra-
kratu-	χratu-
ă̆p-	āfš

In Avestan only, we also see the development of voiced fricatives:

$$\text{b, d, g} > \beta, \delta, \gamma \text{ / V\_}$$

such as in

Skt.	Av.
-tha	θa
khara-	χara-

Finally, another high-profile Iranian sound change is

$$\text{s} > \text{h / \_V}$$

as in

Skt.	Av.
santi	hənṭi

The word India comes from the Greek Indos. This comes from sindhu- in Sanskrit, meaning "river" or "flowing", which in Iranian became hindu- in Iranian by the s > h sound change. Since the Eastern Greeks spoke the Ionian dialect of Greek, with no *h*, "Indos" resulted.

Iranian Morphology

Avestan was attested in two forms: Old (or Gathic) Avestan and Young(er) Avestan. Old Avestan had all the innovative sound changes that Younger Avestan had, but it additionally retained morphological artifacts from PIE that even Sanskrit had lost.

For instance, we have the so-called amphikinetic PIE forms:

Strong	Weak
*dhégh-om-	*dhgh-m-ˊ
*pént-oh$_2$	*pn̥t-h$_2$-ˊ

We consider the expected sound change of the latter of these:

	nom.	gen.
PIE	*pent-oh$_2$-s	*pnt-h$_2$-és
IIr.	*pant-ā-s	*path-ás
Skt.	pantāḥ	patháḥ
Av.	pantå	paθō

The difference here between the nominitive and genitive is severe, and we might expect an analogical sound change to repair one of the forms. In Sanskrit, indeed, we actually get panthā̄h and patháh. But in Avestan, these morphologies and sounds are completely preserved from the PIE forms!

These nouns in PIE and IIr mean "path", but the etymology of the English word is unclear, since in Germanic, p changed to f. It is conceivable that our "path" is actually a loan word from Iranian.

In general, Indo-European languages have lots of compound words of the form

A.B "having B's that are A"

where A is a noun or adjective and B is a noun. These include English forms like redhead, or the Avestan words for three-jawed, etc. We see these compounds in the epithets of *The Odyssey*—e.g., grey-eyed Athena or rosy-fingered dawn.

The Sanskrit grammarians, who catalogued all types of compounds in Sanskrit, gave this compound a name—bahu-vrīhi, meaning "much rice". One such bahu-vrīhi is Zarathustra, coming from the Iranian zarat.uštra, meaning, "having camels that are old."

Lecture 13 — 2/27/13

Ancient Greek

The ancient Greeks had a strong sense of speaking the same language even though the language was broken up into countless dialects—each city-state has its own. *Attic* was the dialect of Athens, which later developed into Koinē (meaning "common") and eventually gave way to all the modern forms of Greek. Meanwhile, *Ionic* was used as the cover term for a lot of other, older dialects, including Homeric Greek. All the other dialects of antiquity have died out.

Attic was eventually adopted for written Greek—and thus is learned today for grammatical structure—and was later adopted for all speech as well.

Homer's epics were dated to around ∼800 BCE, but they reflect an older time period: the Bronze Age. This refers to the Mycenaean period of Greece (1400-1200 BCE). In the 1950's, Michael Ventris deciphered clay tablets associated with the Mycenaean language written in a script called Linear B. His decipherment demonstrated that Mycenaean was an archaic form of Greek—it is thus the oldest attested form of Greek. Nonetheless, Homeric Greek remains the main corpus of ancient Greek for modern study.

Linear B was a very inefficient representation of Greek, in which all characters represented consonant-vowel pairs, or just vowels. The invention of a vowel system, however, was a very Greek idea. The Greek alphabet was borrowed from the earlier Semitic alphabet. In both, each grapheme was taken to represent the first phoneme in its name. However, since Greek did not have a glottal stop, it took the initial consonant [ʔ] of the Semitic "alp" to be the vowel sound represented by α.

Greek vowels sometimes are accompanied by a breathing mark. The rough breathing ῾ mark is transcribed as h in Sanskrit, and indeed it is derived from upper left portion of the character 'H'. The smooth breathing mark ᾽ is transcribed as nothing.

Greek Sound Changes

Greek is the most conservative Indo-European language w.r.t. vowels, but was very innovative with consonants. In Greek, the voiced aspirates of PIE lost their voicing.

bh, dh, gh > ph, th, kh

Note that we transcribe "ph", but it represents a unitary sound p^h. Compare:

Skt.	Gk.
bharati	pherei
dadhāmi	tithēmi

Note that while Sanskrit keeps the PIE consonant, Greek retains the PIE vowel.

Recall also that Greek is a centum language. However, in Greek, the k^w sound was lost:

	PIE	Gk.
"and"	*-k^we	-te
five	*penk^{w}e	pente
gen. inter.	*k^wo-	pōs, pou[5]
which	*k^wotero-	poteros

The lip rounding needed to produce [p] and [k^w] are similar; they send the same acoustic signals and are easy to mistake; compare also

Gk.	Rom.
lingua	limbă

The loss of k^w follows these rules:

$$k^w > t \;/\; \_V_{[+front]}$$
$$k^w > p \;/\; \_V_{[-front]},\ \_C$$

In general, if a p precedes a sound that is not a front vowel, it's difficult to tell whether it was a "real" p or a p that came from k^w.

In Greek, the p is generally favored over the t, e.g., when it comes to analogical change. For instance, from the PIE *sek^w-, we have

[5] "how" and "where", resp.

	3rd sg.	3rd. pl.
PIE	sek^{w}eti	sek^{w}onti
Gk.	hepetai	hepontai

where we get hepetai instead of hetetai by analogy.

Also, although Greek lost k^{w}, it could still be found in Mycenaean, which had glyphs for pe, te, and qe, the lattest of which was pronounced as k^{w}e-.

Greek also presented an unconservative treatment of non-stop consonants. For instance, we see in the above an example of

$$\text{s} > \text{h} \ / \ \#\_$$

More strikingly, many non-stop consonants were completely lost between vowels:

$$\text{s, y, w} > \emptyset \ / \ \text{V\_V}$$

This led to particularly conspicuous constructs in Greek; for instance, consider the following derivatives of the PIE for "fame"[6]

PIE (gen.)	Skt.	Gk.
*ḱléwos	śravaḥ	kleos
*ḱléwesos	śravasaḥ	kleeos

This side-by-side occurrence of two vowels is known as hiatus.

Greek also had many stems with a y, which was problematic. Consider the verb form of philos, phile-ye/o- (the -ye/o- being a thematic suffix):

phileō	phileomen
phileeis	phileete
phileei	phileonti

Such verbs are known as contract verbs; the /eo/ sound contracts to ü.

Lecture 15 — 3/4/13

Recall (or note) that in Sanskrit, the syllabic consonants underwent the following sound changes:

$$\underset{\circ}{\text{r}} > \underset{\circ}{\text{r}} \qquad \underset{\circ}{\text{l}} > \underset{\circ}{\text{r}} \qquad \underset{\circ}{\text{m}} > \text{a} \qquad \underset{\circ}{\text{n}} > \text{a}$$

Contrast this with the changes in Greek:

	Skt.		Gk.	
PIE	Change	Example	Change	Example
$\underset{\circ}{\text{r}}$	$\underset{\circ}{\text{r}}$	dhṛ$\underset{\circ}{\text{s}}$u-	ra/ar	thrasús
$\underset{\circ}{\text{l}}$	$\underset{\circ}{\text{r}}$	pṛthu-	la/al	platús
$\underset{\circ}{\text{m}}$	a	dasa	a	deka
$\underset{\circ}{\text{n}}$	a		a	a- (negation)

[6] An *s*-stem; the root remains across declensions and an -es suffix is added.

[7] Reduplication occurs mostly in the perfect, and only sometimes in the present.

In Greek, the only tolerated terminal consonants were *n*, *s*, and *r*. For instance, the PIE accusative ending -*m* became -*n* in Greek, though it remained -*m* in Sanskrit.

	Gk.	Skt.
nom.	hippos	aśvaḥ
acc.	hippon	aśvam
nom.	theā	
acc.	theān	—m
nom.	polis	
acc.	polin	—m

However, we also see the following:

PIE	Gk.	Skt.
*ph$_2$ter-$\underset{\circ}{\text{m}}$	patera	pitaram

We expect to find *pitara* in Sanskrit since $\underset{\circ}{\text{m}} > \text{a}$, but analogic change yields *pitaram* instead.

Reduplication is also preserved in Greek:

	Root	Pres.	Perf.
Skt.	dā-	dadāti	dadau[7]
Gk.	dō-	didōmi	dedōke
Skt.	dhā-	dadhāti	dadhau
Gk.	thē-	tithēmi	tethēke

Based on our final example, we can deduce that

1. Greek had a version of Grassman's Law.
2. In Greek, devoicing of voiced aspirates occurred before Grassman's Law.

This tells us that Sanskrit and Greek had completely separate instances of Grassman's Law that happen to be typologically identical.

Consider as another example the word buddha, which comes from the Sanskrit budh-ta

	Skt.	Gk.
3 pl. aor.	budhanta	(e)puthonto

Knowing what we know about Grassman's Law and Greek's sound change for aspirates, we can reconstruct the PIE word as **bhudhonto*.

In Greek, only five distinct cases were preserved: the nominative, accusative, genitive, dative, and vocative. The instrumental and locative cases fell into the dative, and the ablative fell into the genitive—this phenomenon is known as syncretism ("falling together"). In PIE, the genitive and ablative often shared declensions, so this syncretism is not too surprising.

Greek also featured an odd grammatical rule, e.g. in the clause

αστρα προ-βεβηκε

from L252 of The Iliad, Book X, II. The word αστρα is the nom. pl. nt., meaning "stars"; but the word προβεβηκε is the 3rd. sg. perf. of the verb meaning roughly "to move along". Typically, plural subjects take plural verbs and singular subjects take singular verbs, but neuter plural subjects also take singular verbs in Greek.

This phenomenon also occurs in Hittite and Old Avestan—it is so strange that this is very unlikely to be coincidental, and we reconstruct it in PIE. In general, highly irregular features that are preserved across the branches of a language family are also constructed in the root language, since its loss is much more likely than its spontaneous formation in multiple completely separate contexts.

These neuter plurals in Greek had *collective nouns* as their origin, and collective nouns ended in *-a*, much like feminine nouns—and hence, they shared the same verb agreement, which is to say, singular.

Greek also had a definite article, "the":

Gk.	Masc.	Fem.	Nt.
Nom.	ho	hē	to
Acc.	ton	tēn	to
Gen.	tou	tēs	tou
Skt.	Masc.	Fem.	Nt.
Nom.	sa	sā	tat
Acc.	tam	tām	tat
Gen.	tasya	tasyās	tasya
OE	Masc.	Fem.	Nt.
Nom.	se	sēo	þæt
Acc.	þone	þā	þæt
Gen.	þes	þære	þes

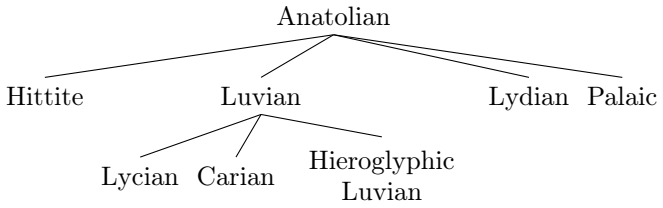
The consistent irregularities in the nom. masc. and nom. fem. certainly go back to a PIE demonstrative, meaning "that". (Note that Greek had a definite article, whereas in Sanskrit, the above paradigm is produced for the demonstrative.)

Lecture 16 — 3/6/13

From Egyptian, Assyrian, and Babylonian sources, we know of a kindgom, Ḫatti, uncovered in the 19th century in Boğazköy (meaning "throat-ville"), a town in modern Turkey. This king, we know, communicated with the Egyption pharoah, referring to the pharoah as his brother.

In 1906, the royal archive of the Hittites was discovered in Boğazköy, written in a cuneiform writing system which was neither Babylonian nor Egyptian. This was deciphered by Hrozny in 1915-1917, which began the study of the Hittite language.

Hittite's language family, the Anatolian family, was organized as follows:



Anatolia refers to the geographic region of Asia Minor. In classical times, the phrase "rich as Croesus" emerged, referring to the extreme wealth of King Croesus of Lydia, a kingdom of Asia Minor. The Lydian language was eventually displaced by Greek.

All Anatolian languages are now dead languages. Hittite, attested in stone and clay tablets from 1500 BCE, remains the oldest attested Indo-European language.

Hittite employed a cuneiform writing system. In addition to phonemic symbols, however, Hittite also made use of ideograms. Ideograms convey their meaning upfront, but do not represent any sound; they played an important part in cuneiform writing throughout history.[8] This practice, however, means that there are Hittite words for which we have no sense of pronunciation.

Consider now the following Hittite passage, which began Hrozny's decipherment of the language:

nu BREAD-an ez-za-at-te-ni
wa-a-tar-ma a-ku-ut-te-ni

Sentences in Hittite seemed often to begin with *nu*, which probably meant "and" or "now". Hrozny also recognized that *-ma* was a particle that could attach to a number of words.

The key to Hrozny's decipherment was in realizing that there were two parallel constructs in this sentence, of the form

_____ __-teni

Moreover, *ez-za* resembles the Latin *edo*, and *a-ku* resembles the Latin *aqua*; *wa-a-tar*, of course, resembles "water". *-an* resembles the PIE accusative ending -(V)*m*, which we know became *-an* in Greek. Likewise, *-te* is like the PIE second person verbal suffix. The *-an* suffix of the BREAD ideogram was a phonetic complement indicating declension; BREAD-aš would be the singular nominative. Hrozny thus was able to deduce the meaning of this sentence: "You eat bread, you drink water"; this paved the way for his ultimate decipherment of the language.

The Hittite empire was destroyed in 1200 BCE.

[8] In English, the ampersand "&" is an ideogram, derived from the Latin *et*; it is a graphical contraction of the letters 'e' and 't'.

Hittite and PIE

Hittite helped to confirm the PIE laryngeal theory. The Hittite ḫ was shown to be the reflex of h_2, e.g.,

Hitt.	Eng.
tarḫ-zi	s/he conquers
ḫanza	

Surprisingly, Hittite is much less morphologically complicated than we would expect based on its proximity to PIE:

- Hittite has no dual.
- Hittite retains only the present stem and has only two verb tenses, present and past, constructed with primary and secondary endings, respectively.
- Hittite verbs have only two moods, indicative and imperative—the subjunctive and optative are missing.
- Hittite kept only five PIE cases (but innovated its own instrumental and ablative).
- Hittite declensions were less complicated than those in PIE.

Hittite did, however, retain a distinction between active and middle voices.

Hittite's writing system did not distinguish between *k* and *g*, and due to its other limitations, dummy vowels were also often required:

PIE	Hitt.
g^wén-ti	ku-en-zi
g^whn-énti	ku-na-an-zi

Since *-zi* was pronounced as *-tsi*, we also see here that

$$t > ts \ / \ \_i$$

Hittite is also the only PIE language which retains *r*/*n*-stems. Consider the words in PIE and Hittite for the English "water", which in Old Norse was *vatn*.

	PIE	Hitt.
nom.	wód-r̥	wātar
gen.	wéd-n̥-s	wetenaš
loc.	wed-en	

We see the same preservation in the Hittite words for embassy:

uppeššar
uppeššnaš

Hittite is a centum language, and we have, for the 3rd sg. of the PIE word "to lie",

PIE	*kéi-
Skt.	śaye
Gk.	keitai
Hitt.	kittari

We can further contrast Hittite with Luvian, one of its sister languages:

PIE	Hitt.	Luv.
k^w	kuiš[9]	kuiš
ḱ	kittari	ziyar
k	kišai	kiš

Luvian, unlike Hittite, retains all of PIE's tectals.

One idea which enjoyed a certain vogue was that Anatolian was not a branch of PIE at all, but rather, that both were sister branches of an earlier language, Proto-Indo-Hittite. This let all the existing theory of PIE stand, and in this sense was kind of a lazy solution (though lazy does not necessarily mean wrong). However, this theory has not stood the test of time; Anatolian likely branched off from PIE early, but nonetheless is considered derivative.

Lecture 18 — 3/11/13

Unlike most IE languages, Hittite has a single glob gender (masculine/feminine) along with a neuter, and doesn't have the same agreement requirements. Perhaps the three-gender system is an IE innovation, in support of Proto-Indo-Hittite.

A clitic is a little word that attaches to other words and cannot exist on its own. The term generalizes two other words. The first is enclitic, which refers to a little word that attaches to a word on its left, the "host"; the attachment is made for purposes of accent or stress. Meanwhile, Romance languages often have indirect pronouns which are proclitic; they come to the left and must appear with a verb.

Arguably, the English [tə] which makes verbs infinitive is not a word in context, but a clitic. The English morpheme *-n't* is also a clitic.

Early IE languages, and Hittite in particular, had many clitics, such as in *nu-kán* and *našta*—both are the particle *nu* along with an enclitic, *-kán* or *šta*. Consider also the following lines:

nu-kan ^{MUŠ}Illuyankan kuenta
n-an-kan kuenta

Object pronouns in Hittite are made much like in French; they are clitics that have to attach to something. In particular, we have *-an* (him, her), *-at* (it), and *-uš* (them). Thus, our second line reads "he killed him".

In Hittite, we find that clitics are often in the second position in the sentence. We see a similar phenomenon in Greek:

astra de dē probebēke

The first *de* must go where it is in the sentence; were there more clitics, they would chain together in that position. The observation that clitics tend migrate to the second position in the sentence is known as Wackernagel's law, and it applies to non-IE languages as well.

The clitic *de* is phonologically part of the subsequent verbs; recall that this splitting of preverb from verb is known as tmesis.

Primary and Secondary Endings

Proto-Indo-European verb conjugations used one of two kinds of endings. Primary endings typically denoted the "real present"—things happening right now—whereas secondary endings were used for other purposes. Thus far, our primary endings have always been the secondary endings plus a particle *-i*, the *hic et nunc*, or "here and now", particle. In Sanskrit and Greek, for instance, we have:

	Primary	Secondary
Skt.	bharati	(a)bharat
	bharanti	(a)bharant
Gk.	pherousi	epheron
	pheronti	epheront

Not all languages keep the *-i* particle; Latin, for example, seems to have lost it for the most part.

Note, however, that this picture of primary and secondary endings is only the case for the active. The endings in the middle are more complicated. We have:

	Primary	Secondary	Meaning
Gk.	hepetoi	(e)hepeto	"to follow"
Skt.	sacate	asacata	"to follow"

The secondary ending for the 3rd pl. in Greek was *-nto*; this bears similarity to the active paradigms. Note also that the standard form of *hepetoi* was *hepetai*; likewise, the Sanskrit *sacate* arose form the sound change ai > e. Sanskrit and Greek thus offer no surprises with respect to primary and secondary endings in the middle.

Turning to Hittite, however, we see a different story:

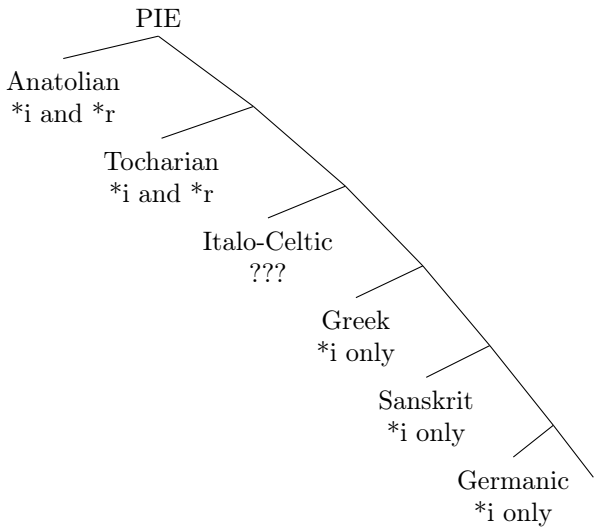
	Primary	Secondary	Meaning
Hitt.	zaḫḫiyattari	zaḫḫiyatta	"to fight"
	kittari	kitta	"to lie"

In Hittite, we see a middle suffix *-ta* alongside a primary particle *-ri*; this results from *-ta-ri* < *-*tor*.

With this data, we can revise our rule for primary and secondary endings. While the *hic et nunc* particle appears everywhere in the active, in the middle, some languages have a primary form which is the secondary form plus an *-r*. We see this in Hittite, and it is also the case in Tocharian, where we have

	Primary	Secondary
3rd. sg.	-tär	-te
3rd. pl.	-ntär	-nte

Our picture of Indo-European, with respect to these primary and secondary mediopassive endings, thus looks like this:



Hittite also had two different ways of inflecting active verbs—it had two distinct sets of primary and secondary endings.

	"to seize"	"to take"
1 sg.	epmi	dāḫḫi
2 sg.	[*epši]	dātti
3 sg.	epzi < *-ti	dāi

These were known as the mi-conjugation and the ḫi-conjugation, respectively. There was no difference in meaning between them; they were simply a grammatical fact. No other IE language has anything obviously like this distinction; however, it does seem to be related to the perfect in other languages:

	PIE	Gk.
1 sg.	*-h_2e	-a
2 sg.	*-th$_2$e	-tha
3 sg.	*-e	-e

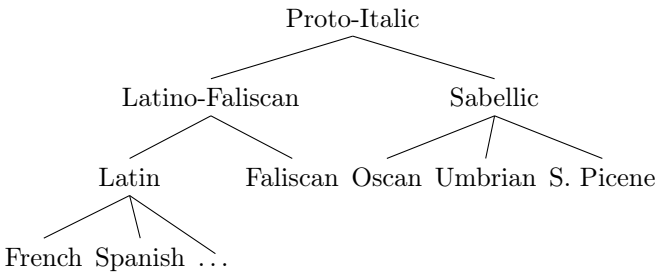
These endings and Hittite's ḫi-conjugation are somehow connected, but that's as far as the consensus goes. This is one of the many unresolved problems in Hittite.

Lecture 19 — 3/13/13

Italic

Unlike Greek, Latin was not the sole, nor even the main language, of its family. Latin was cousin to another Italic language called Faliscan, and Latino-Faliscan was one of two main branches under Italic. The other was Sabellic, which was much more far-reaching than Latin. Under the

Sabellic branch, South Italy had Oscan (more archaic), and North Italy had Umbrian (the best attested) and S. Picene.



Other cousins of Latino-Faliscan and Sabellic include Sicel (gave Sicily) and Venetic (Venice). Sicily is thought to be Italic, but there is not enough evidence to be sure; Venetic was very likely Italic, but not part of the two main branches.

Early Latin was very different from modern Latin—basically unrecognizable. Moreover, Italic languages are much less archaic and conservative than Greek, Sanskrit, and Hittite.

Italic was a centum family, and preserved the basic, unaspirated PIE stops:

$$\text{p, t, k, k}^{\text{w}} > \text{p, t, k, k}^{\text{w}}$$
$$\text{b, d, g, g}^{\text{w}} > \text{b, d, g, g}^{\text{w}}$$

Meanwhile, the voiced aspirates lost their voicing and eventually gave way to fricatives:

$$\text{bh, dh, gh, g}^{\text{w}}\text{h} > \text{p}^{\text{h}}, \text{t}^{\text{h}}, \text{k}^{\text{h}}, \text{k}^{\text{wh}} > \text{f}, \theta, \chi, \chi^{\text{w}}$$

Contrasting against Sanskrit, we have

Skt.	Gk.
bhrātar	frāter
bharati	ferō

However, we also have

Skt.	Gk.
dhūma	fūmus

where we would expect þūmus. However, [θ] and [f] are similar sounds, and Latin underwent the sound change

$$\theta > \text{f}$$

As another example, we have

Skt.	Gk.	Lat.
dhā-	thē-, (e)thēke	*theci > feci

Another Latin sound change was

$$\chi > \text{h}$$

This change persists across the Romance languages; we have

Latin	French	Italian	Spanish
habēre	avoir	avere (ho, ha)	haber

Contrasting against PIE and other IE languages, we see examples

PIE	Lat.
*ghos-ti	hostis
"stranger"	"enemy"

and

	PIE	Gk.	Skt.	Lat.
"goose"	*ǵhă̆ns	χην	haṃsa	*hanser > anser

Continuing with our fricative reflexes, we have

$$\text{g}^{\text{w}}\text{h} > \text{k}^{\text{wh}} > \chi^{\text{w}} > \text{f}$$

with, for example,

$$*\text{g}^{\text{w}}\text{hén-ti} > *\text{g}^{\text{w}}\text{en-d-} > \text{fendō}$$

In the middle of words, fricatives often became voiced; however, sound change continued past this event to yield unexpected reflexes. For instance, we have

PIE	Gk.	Skt.	Lat.
$\text{h}_1$rudh-ro-	eruthros	rudhira-	rubro-

In Latin, the following transformations occurred:

$\text{h}_1$rudh-ro-
rudhro-
ruþro-
rufro-
ruβro-
rubro-

with nom. ruber and acc. rubram.

Take as another example the PIE root

$$*\text{sneig}^{\text{w}}\text{h-}$$

One form of *sneig$^{\text{w}}$h- was *snoig$^{\text{w}}$h-o-s, a noun meaning "mass of sticky material" or "stickiness". In daughter languages, this manifests as

PIE	Skt.	Gmc.	Eng.	Lith.	OCS
*snoig$^{\text{w}}$h-o-s	sneha-	snaiwa-	snow	sniegas	sněgŭ

meaning "snow" (except the Sanskrit *sneha-*, which means "love").

The other form of this root was the root noun, *sneig$^{\text{w}}$h-s, with genitive *snig$^{\text{w}}$h-es. These forms were inherited by Latin and Greek. In Greek, νιφα was the word for snow in the accusative.

In Latin, on the other hand, we had, in the nominative,

$$\text{nix} < \text{-g}^{\text{w}}\text{hs}$$

and in the accusative, from the PIE genitive,

$$\text{nivis [nɪwɪs]} < \text{-Vg}^{\text{w}}\text{hV-}$$

In turn, we had

$$\text{nivis} > \text{nivea} > \text{nivya, neige}$$

Without the subsequent vowel or -*s*, we also see in Latin

$$\text{ninguit} < \text{*sni-n(e)-g}^{\text{w}}\text{h-}$$

Note that sound changes in Latin do not always generalize to Italic. In Oscan and Umbrian, for instance, the action is in the labiovelars:

$$\text{k}^{\text{w}},\ \text{g}^{\text{w}} > \text{p, b}$$

and we have

Lat.	Osc.
quis	pis
quod	púd

Unlike Greek, Italic languages are not dialects; there was no unity among Italic speakers.

We examine one final phenomenon of Italic sound change. This is rhotacism, which means "turning into an *r*". In Latin, we have

$$\text{s} > \text{r} \ /\ \text{V\_V}$$

probably via s > z > r. For instance, the Latin for flower is given by

nom.	gen.	adj.
flōs	flōris	flōrālis

Rhotacism did not occur without the presence of vowels; we can contrast, for instance,

*ges- > *gesō > gerō
*ges-to-s > gestus

Compare these words for "race, kind" which are familiar in modern English:

	nom.	gen.
PIE	*ǵénh$_1$-os	ǵenh$_1$-es-os
Skt.	ganaḥ	janasaḥ
Gk.	genos	geneos
Lat.	genus	generis

Lecture 21 — 3/25/13

PIE had a word accent which could stand on any syllable of the word:

g$^{\text{w}}$hén-ti	g$^{\text{w}}$hn-énti
memón-e	memnḗr

Greek and Sanskrit preserved this free accent, but Latin did not. Latin had a different way of accenting words, which we know about in a lot of ways, e.g., the way in which the Romance languages carried on these stress rules.

In Latin:

- If a word has two syllables, accent the first.
- If a word has three or more syllables, accent the second-to-last syllable if the syllable is long, else the third-to-last syllable.

For instance, compare

nom.	acc.
uérităs	ūrităátem

In Romance languages, the word for truth is derived from the Latin accusative; the last syllable is dropped, and we end up with a terminal stress:

French	Spanish	Italian
vérité	verdad	verità

There was also an intermediate step between the PIE accent system and the classical accent system, in which initial syllables were always accented, which gives us a three step evolution of the Latin accent system:

1. PIE free accent
2. initial accent
3. classical accent

the latter two steps occurring in the history of Latin. The classical accent is typically preserved for Latin loan words in English.

In English, we cannot readily imagine any vowel in the second syllable of an initial-accent word besides a schwa. For instance, the word

monitor

must, in English's stress system, admit a pronunciation in which the *i* is a schwa. Indeed, post-stress vowels are (almost) always reduced. A similar phenomenon in Latin leads to the following picture, when preverbs are attached to verbs:

	faciō	factus, factum
per	perficiō	perfectus
re	reficiō	refectus
con	conficiō	confectus

At some point in its history, Latin had words of the form pérfaciō, preserving the root verb. However, post-stress vowels weakened just like in English—only afterwards did the classical accent rule arise, which produces the forms in the table above.

Additional evidence of this intermediate accent shift include

	verb	past part.
	caedō	caesus
in	incīdō	incīsus

and

*kekaid-
*cecīdī
mātri-cīda
homi-cīda

Note the vowel change of

$$*ai > *ei > \bar{\imath}$$

Our first example also reveals a strange alternation between *-d-* and *-s-*. Other examples of this alternation abound:

verb	nom.
cadō	cāsus
sedeō	sessus
mittō	missus
uideō	uīsus

The past participle of many words has a similar form, but is more regular:

verb	past part.
dicō	dictus
canto	cantātus

These are derived from the PIE past participle suffix, and also exhibit vowel weakening; that is, we have

$$\text{-tus} < *\text{-tos} < *\text{-to-}$$

We look to the word for "to see" in PIE for an explanation:

$$*\text{weid-}, *\text{woid-}, *\text{wid-}$$

Its past participle should, in theory, have been

$$*\text{wid-tó-s} > *\text{wittó}$$

However, we can't have the voiced *-d-* followed by the unvoiced *-t-*. Instead, we might expect [tt], but this is actually not very common. Turning to the various IE daughter languages, we find

PIE	*wid-tó-s > *wittó
Skt.	vittá-
Av.	vista-
Gk.	aistos < awistos < n̥-wistos
Hitt.	-zt- [tst]
Lat.	uīsus < uīssus

Note that in Hittite, the *-zt-* is pronounced as [tst], e.g., *ezzatteni* is pronounced as *etsteni*, or, earlier, *et teni*. We therefore reconstruct the PIE pronounciation of *wittó with Hittite's [tst], as this is the most likely to produce all the different treatments we see, by simplification of the consonant cluster in different ways.

Let us examine another oddity of Latin, shown by:

	verb	nom.
Lat.	iubeō	iussus
Skt.	yudh-	

The verb results from the sound change

$$\text{dh} > \text{th} > \theta > \text{f} > \text{ƀ} > \text{b} \;/\; \text{V\_V}$$

while the nominative comes from

$$\text{iussus} < \text{yudh-to} < \text{-[tst]}$$

Verbal Morphology

In Latin, the PIE aorist and perfect tenses merged into a single "perfect" tense. Meanwhile, like PIE, Latin had a present stem was was used to produce the present and imperfect tenses.

In PIE, recall that the present and imperfect tenses were constructed from the present stem using the primary and secondary endings. As we know, Greek and Sanskrit preserved these paradigms.

	Present	Imperfect
PIE	*bhéreti, -onti	
Skt.	bharati, -anti	abharat, -an
Gk.	phérei	ephere

Latin, and most other IE languages, however, did not. Latin innovated its own imperfect tense, e.g., in the following paradigm for "to say":

Present		Imperfect	
dīcō		dīcēbam	
dīcis		dīcēbās	
dīcit	dīcunt	dīcēbat	dīcēbant

Historically, the *-ba-* suffixes meant "to be", so dīcēbam, for example, very literally meant meant "I was saying." Though they were no longer free morphemes in Latin, they came from the PIE "to be", which was

$$*\text{bhuH}$$

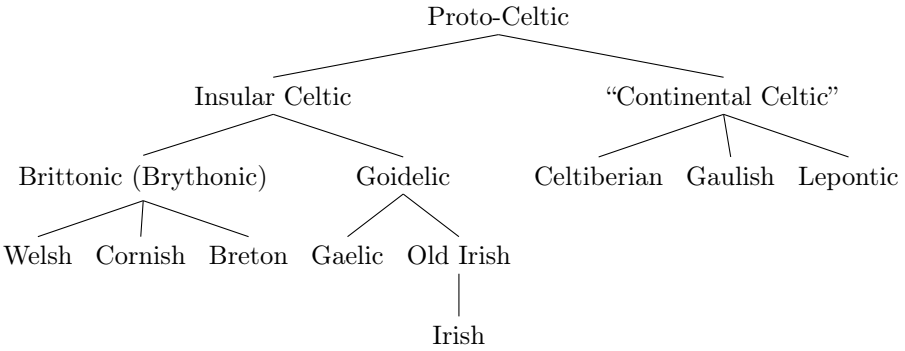
where H is probably h_1 or h_3.

Latin retains no trace of the PIE imperfect. It also innovated a future tense:

Future
cantābō
cantābis
cantābit

Romance languages kept the *-b-* in the imperfect, but not in the future, for which they instead used the infinitive.

Lecture 22 — 3/27/13



The names of a number of areas throughout Europe saw some degree of Celtic influence. For instance, the modern city of Milan < Medio|lanum, meaning "middle plain" (Celtic loses the *p* which is retained in the English "plain"). The Anatolian region of Galatia, meanwhile, is a cognate of Gaul.

Celtic languages in general retained many cases from PIE. These were inflectionally similar to cases retained in other IE languages, e.g., the relative pronoun

	Celtiberian	Skt.
nom.	ios	yas
acc.	iom	yam
dat.	iomui	yasmai

Old Irish

The Irish developed a writing system of notches inscribed on the edges of upright gravestones. Known as Ogham, these were the earliest attestations of Old Irish. These inscriptions often took the form "Of *X*", in the genitive, commemorating the individual interred there.

Ogham is significant in our linguistic understanding of Old Irish because, like Latin and Greek, or other Celtic languages like Celtiberian, Ogham largely preserves the PIE inflectional endings it inherited from Proto-Celtic, which were later lost in Old Irish. For example, one such inscription was

TRIA|MAQA|MEOLAGNI

which consists of *tria*, the gen. pl. of "three"; *maqa*, the gen. pl. of "sons"; and *meolagni*, the gen. sg. of the name Meolagnos. In Old Irish, however, where these endings were lost, this would be rendered as

tre mac Mailáin

Old Irish Sound Changes

Old Irish is very difficult to learn because of a number of rules governing the combination of words, e.g.,

do beir	to bring
ni tabari	to not bring

In Old Irish, stops became fricatives intervocalically:

$$\text{t, d, g, } \ldots > \theta, \eth, \gamma \ / \ \text{V\_}$$

This process is known as lenition; for instance:

	Pre-OIr.	OIr.
father	*atīr	athair
tribe	*tōta	túath

Nasals following a vowel and followed by a voiceless stop become a double voiced stop, which usually simplified to a single voiced stop with compensatory lengthening of the vowel:

$$\text{VC}_{[+\text{nasal}]}\text{C}_{[-\text{voice}]} > \bar{\text{V}}\text{C}_{[+\text{voice}]}$$

This is called nasalization, and we have, e.g.,

	Pre-OIr.	OIr.
100	*canton	cét [kēd]

Finally, *-s-* was lost word-medially, with

$$\text{s} > \text{h} > \varnothing \ / \ \text{V\_V}$$

The confusing bit, however, is that Old Irish liberally applied these sound changes across word boundaries.

	Pre-OIr.	OIr.	Irish
their tribe	*eisan tōta	> a túath [dúath]	a dtúath
his tribe	*esa tōta	> a thúath	a thúath
her tribe	*esas tōta	> a túath	a túath

	Pre-OIr.	OIr.
their fish	*eisan ēskas	> a n-íasc
his fish	*esa ēskas	> a íasc
her fish	*esas ēskas	> a [h]íasc

All of the "little words" mutate according to these rules:

	"son white"	"hand white"
Pre-OIr.	*mak$^{\text{w}}$os windos	*lāma winda
OIr.	mac(c) find	lám ḟind[10]
MW	mab gwyn	llaw wen

[10] $\dot{X}$ means silent X.

Lecture 23 — 3/29/13

(These notes are from section, not lecture.)

Functions of the PIE Middle

The middle voice in Proto-Indo-European has a variety of functions:

- passive (questionable?), e.g., in Latin,

inflection	Lat.	
3 sg. pass.	initur	"it is gone"
3 sg. pass.	traditur	"it was traded, related"
3 sg. act.	tradit	"he relates it"

- reflexive: e.g., in Greek,

Gk.		voice
louō	"I wash"	act.
louomoi	"I wash myself"	mid.

- self-benefactive:, e.g., in Sanskrit,

Skt.		voice
yajati	"sacrifices"	act.
yajate < *-tai (?)	"sacrifices for herself"	mid.

- anticausative, or inchoative. Consider an example in English of anticausative usage of "to break":

The window broke.	anticausative
John broke the window.	causative

- deponent, i.e., middle/passive form with active function: e.g., in Latin,

*sequō > sequor "I follow"

Recall that we reconstruct the PIE middle with a *-*r* ending. In general, when we have a situation like the following among attested sibling languages:

1		2		3	
sg.	pl.	sg.	pl.	sg.	pl.
A	A	A	B	B	B

we want to reconstruct the proto-languages as

*0	
sg.	pl.
A	B

In an unrelated footnote—in PIE, the singular is the strong form of a verb, with the plural being the weak form. The motion of an accent between the suffix in the strong form and the ending in the weak form is called hysterokinesis.

[11] The lenided form is different.

Lecture 24 — 4/1/13

The earliest Celtic languages that have been discovered are sort of garden-variety IE languages—not drastically innovative. However, before the most well-attested languages like Irish and Welsh arose, some very dramatic changes occurred. Final syllables were lost, and the institution of mutations—the most prominent such change—emerged. Irish and Welsh also feature very innovative consonant systems—in both languages, for example, *t* can alternate with θ and *d*.

Let us turn now, however, to some of the more basic Indo-European characteristics of the Celtic family. Celtic was a centum language:

OIr.	Celtib.
cét	cantom

Some Celtic languages kept the k^w which is typically preserved in centum languages, but others have a *p* in its place:

	Irish	Welsh
who	cía	pwy
4	cethir	pedair[11]

Likewise, for the word "and", Celtiberian had *-cue*; Gaulish had a *p*.

There used to be a significant linguistic distinction between these groups, which were known as P-Celtic and Q-Celtic. P-Celtic included Gaulish and Brittonic (which included Welsh), and the remaining languages were Q-Celtic. However, the $k^w > p$ change among these subbranches were different events; P-Celtic and Q-Celtic don't describe a genetic relationship, merely a language feature. This is seen in other IE families as well:

Lat.	Osc.
quis	pis

In Celtiberian, the k^w was retained, but in Old Irish, it eventually lost its labialization and became [k], written *c*. This was a later-occurring change; in Ogham, MAQAS was still pronounced [makwas].

In Latin, dh > f; however, in Old Irish,

d, dh > d

This makes Old Irish one of the few IE languages where the voiced stops and voiced aspirates merged. It is impossible to discern the reconstructed PIE consonant given an Old Irish *d*:

	OIr. PIE
10	deich < *d
fire	daig < *dh

Relationship to Italic

Italo-Celtic is the second hyphenated IE branch we have encountered thus far, the other being Indo-Iranian (and the third is Balto-Slavic). The jury is out (which is to say, the jury has gone home for good without returning a verdict) on precisely how Italic and Celtic relate. It is much less obvious than other hyphenations, but these two branches do share some remarkable features which are probably better explained by a short period of common development between the languages before branching off.

In Celtic, the Indo-European *p* disappears. However, we have an odd exception:

	PIE	OIr.	Welsh	Lat.
5	*penk^{w}e	cóig	pump	quīnque

We don't expect *p* to give k^w in Latin in general, and from this picture, we can deduce that a sound change occurred during the Italo-Celtic stage, before Celtic lost the *p*:

$$p\_k^w > k^w\_k^w$$

	PIE	I-C	Lat.	Welsh
"to cook"	pek^{w}-	k^{w}ek^{w}-	coquó	pobi[12]

Recall also our primary and secondary endings in PIE:

	primary	secondary
active	-ti, -nti	-t, -nt
middle	-tor, -ntor	-to, -nto

Hittite, as we know, preserves *-tor* and *-ntor*, and so does Italo-Celtic. Other families, however (including all later branches) have *-toi* and *-ntoi* in the middle instead:

Lat.	OIr.	Gk.	Skt.
sequitur	sechethar	hepetai	sacate

Italic and Celtic also share the way they mark superlative adjectives. In PIE, the superlative is made by

*-is-to- > -est (Eng.)

The comparative, meanwhile, is marked by

*-is- > -er (Eng.)

Greek, Sanskrit, and Germanic tend to preserve this suffix for the superlative.

Latin, however, does not make superlatives in this way, and instead we have

Lat.	super.
fortis	fortissimus < *is-amo-
maximus	magissamus

In Celtic, the same suffix is used! In Old Irish, for the word "old" for instance, we have

OIr.	super.
sen	sinem < senisamos

This appears to be an innovation of Italic and Celtic together.

[12] The *b* is the lenition of a *p*.

Celtic and Old Irish Syntax

Old Irish is a typical IE language in that it makes heavy use of prefix verb forms:

to bring	to give	to bring out, to say
berid	do.beir	as.beir

and

to take	to find
gaibid	fo.gaib

Insular Celtic languages share an innovation that marks them as related: they both become VSO languages. Many IE languages are SOV, and English is SVO. The verb gets fronted in insular Celtic, originally for emphasis, but eventually becoming the standard syntactic ordering. This is not normally a feature of IE languages.

This becomes interesting when compound verbs (preverb + verb) occur at the beginning of a sentence:

berid in fer claideb
"carries the man a sword"

do.beir in fer claideb
"gives the man a sword"

Suppose that we then wish to construct the sentence, "The man gives me a sword." Irish prefers to use a clitic or unstressed version of the pronoun right next to the verb, resulting in an infixed pronoun as follows:

do-m.beir in fer claideb
"gives me the man a sword"

This sort of thing can't be done in English or in the Romance languages, but in Old Irish, preverbs were still separate from their verbs (recall that this is called tmesis):

do . . . ber

The placement of the pronoun falls naturally in the second position, as described by Wackernagel's law!

Historically, what happened here is that the pronoun was concatenated after the prefix, and only afterwards was the verb glommed on. Thus, the pronoun will induce initial mutation in the base verb. Since the *-m-* is a reduced form of the pronoun *me*, the pronunciation is actually [dom-veir], although the spelling remains unchanged.

As another example, we have

ni.gaib [g]
he does not take

ni-m.gaib [ɣ]
he does not take me

The Old Irish third person pronoun was just a vowel that tended to contract with the vowel before it:

ní.gaib [ɣ]
he does not take it

ní.ngaib [g]
he does not take him

The word "to find" in Old Irish was given by

foigaib < *wo-gabiti

and we have

the men who take God
**indi wiri gabinti-yo Déwan*
"the men they take who God"

the men who find God
**indi wiri wo-yo-gabinti Déwan*

This illustrates the migration of the clitic to occupy the second position in the clause ("who take God"), depending on the presence of a preverb. The transition to Old Irish then yielded

take *ind fír gaibet Dé* [g]
find *ind fír fo gaibet Dé* [ɣ]

To determine what happens in each of these cases, it is sufficient to learn the verb, then the special relative form for the third singular, and the one for the third plural. The clitic goes in the second position—if there's a preverb prefix, it goes after that, but otherwise it has to go after the verb.

Irish preserves the separateness of prefixes in compound verbs that other languages don't. It also shows the Wackernagel's law of clitics more than any other language does.

Lecture 25 — 4/3/13

Study of the Indo-Europeans

The study of the Indo-European people is inherently political—hopelessly intertwined with political, social, and ideological forces that have moved people throughout Western history since the 19th century. People tend to see the Indo-Europeans as a vehicle for their own vision for how history was.

The early 19th century is when the study of Indo-European linguistics first became a serious pursuit, and since then it has always borne some of the historical context of its origins. The 18th century had just ended; industrialization was emerging; superstitions were high; and social, economic, and political mobility were still rather low. This led to a sense of a new beginning.

The French Revolution showed that a national group could form a modern state, and could remove the old institutions and the old forms of property and power. Napoleon Bonaparte brought great swaths of Europe under French influence. The Napoleonic Wars aroused a new sense of European nationalism; this is the origin of modern nationalism, that we are destined to come together and join as nations. The thought, "If the French can do it, why can't we do it?", pervaded countries like Germany, the Netherlands, and Italy.

Indo-Europeans and Nationalism

As nationalism grew up, and different people started asserting their rights *qua* their nation, they became more interested in studying their origin and reclaiming their "true essence". French cultural domination extended to court manners, cultivated literature, etc. In the rising nationalism, it became popular to discover what sort of aspects of, say, German or Italian culture weren't reflected in French culture. People became interested in studying their real background, in spite of the French and in spite of, say, the Roman Catholic tradition or of local princes.

The Brothers Grimm collected folk tales as part of the project of rekindling an interest in the authentically German part of the German tradition; the same thing happens in all European countries to some extent. These were interests that had not yet emerged during the 18th century.

When the idea of the Indo-Europeans came along, any European national group not interested in tracing their history through the Latins or Greeks suddenly had an alternative origin that they could study. If the Indo-European languages are related, then there must have been people who spoke Indo-European. Those people must have been, in some sense, the founders of the Greeks and the Romans and the other Southern European peoples, but also those of Northern Europe. Indian culture suddenly became important in studying European heritage. People across Europe began to use the Indo-Europeans to back up for their cultural origins, or, to put it more crudely, national claims.

Indo-European Supremacism

The tone of nationalism in the new century begins with noble aims of discovering culture and heritage; in the second half of the century, however, it turned to reasserting rights against other nations whose rights were deemed less valuable.

In the 19th century, the European world was also beginning to penetrate Asia. Portuguese and Dutch settlements sprung up in the "New World". The Industrial Revolution began, and European technological and scientific progress outstripped the progress of other parts of the world. Europeans were increasingly in the driver's seat—and as usually happens when one finds oneself in this position, one begins to question one's place in the world and whether or not it is just, and usually concludes that it is.

Indo-Europeans were used to justify that the European peoples were of higher stature than other peoples—Indo-European peoples had contributed more to modern science and had conquered the world, and it was easy and flattering to conclude that their superiority was due to the Indo-European element, which just made them better than the rest of the world.

It was argued that modern Greeks weren't "real" Greeks because they weren't fair-haired and grey-eyed, and these Nordic characteristics were taken to be the "essential" Aryan. This, as it turns out, was not the right choice for the Indo-Europeans' traits, but it was chosen for its purity—its difference from non-European characteristics. Those who no longer had Nordic traits were believed to have intermixed too much with native races and to have lost their Nordic purity.

Everyone held these sorts of views during this time period; that is to say, there weren't bad people who thought these self-aggrandizing things and good people who didn't—everybody did. This supremacism found its way to America as well, although there the views were held much less aggressively.

The usually identified site of Indo-European civilization was Northern Europe.

Linguistic Justification

Starting in this time, everybody also began to talk routinely about the Aryans. Aryans were said to be successful because they had a superior physique and a superior language.

The word Aryan comes from the Skt. *ārya*, meaning "us":

- Old Persian had a cognate *āriya*, from *ariyānām dahyanš*, "Persian kingdom".
- The word was also similar to *Ériu* and *Erenn*, the Old Irish meaning Ireland, but we know now by regular sound change that this is impossible.

  Ériu had a Welsh cognate *Iwerddon*, from which we can determine that the actual starting point for the word must have been a nominative *Iweriia̯ū̯*, *Īweriia̯onos*. In Welsh, also, *-ri̯- > -dd* [ð]. In turn, we have *Īweriia̯ū̯* < *pīweriyo-*, meaning "fertile" or "fat", from **pi̯er*.

Ultimately, we find that Ireland means "fertile place", and has no relation to Aryan; Aryan had no more claim to what the Indo-Europeans called themselves than any other name.

Another interesting linguistic association was that of the swastika. The swastika enters into European iconography before it has any Nazi associations, as a purely "Aryan" symbol.

swastika < svāstika (Skt.)
< svasti-, "welfare" or "health"
< su-, "good", "well" (Gk. *eu-*) + asti, "be-ing"

It was decided by someone that the swastika represented an Aryan solar system, sacred to the Aryans.

Everything to do with the study of the material culture, the history, and the whereabouts of the Indo-Europeans is tied up with this whole century of wishful-thinking that the Indo-Europeans begat an inherently superior people.

Lecture 27 — 4/8/13

Linguistic Paleontology

A favorite candidate region for where the Proto-Indo-European people lived was East Germany—in the area which is now Poland, in Europe. This was well in line with Germanic nationalism, but how can we verify the factuality of this claim?

In determining where the Proto-Indo-European people lived, we use "linguistic paleontology"—we look for PIE words which describe their environment. For instance, we have:

- "beech": Lat. *fāgus*, Gk. *phēgós*
- "salmon": Ger. *Lachs*, Lith. *lašis*, Toch. B *lākso*

The limited geographic distribution of the beech and salmon places the Proto-Indo-European homeland in the Northern European plain. This suggests an identification of the PIE people with the Corded Ware culture—so named for their pottery, which was patterned by pressing string into the raw clay.

Linguistic paleontology suffers from a serious shortcoming, however— misapplication of these various names arises as the PIE language and its people migrate. For instance:

- "robin": The European robin is not related to the American robin. The word came across the Atlantic with the Puritans, who had a conception of what the robin looked like—but they were actually two different species.

- "buffalo": Refers to the American bison, but the European water buffalo.
- "possum": Am. marsupial animals, Eur. squirrel-like animals.

How can we know that the beech or the salmon actually referred to those particular species? In fact, we can't, and to draw such a conclusion would be incorrect:

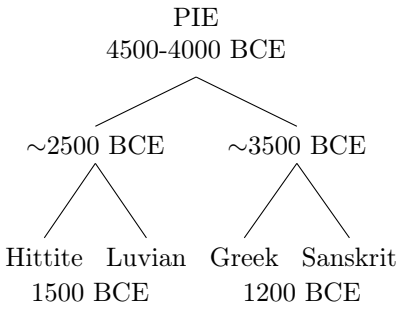
- Gk. gen. *phēgós* refers to the Valonia oak, not to a beech tree.
- Ger. *Lachs* refers to the Atlantic salmon, which has a limited geographic range, but the brown trout is very similar and lives all over the place—the Toch. gen. *lākso* simply means "fish", or perhaps "salmonid fish".

Locating and Dating the Proto-Indo-Europeans

We really only have a very few hard facts when trying to date and place PIE culture.

Although the IE languages are part of the same family, they are really very different, and we can use their various linguistic artifacts to date the Proto-Indo-Europeans. Romance languages, for instance, have a time-depth of at most 2000 years, at which point they converged on their common ancestor, Latin. Sanskrit is attested from approximately 1200 BCE; Greek starts at least with Homer, also around 1200 BCE.

Hittite is dated around 1500 BCE; Luvian is in the same family but doesn't even begin to look like Hittite. Proto-Anatolian diverged probably no later than 2500 BCE, and we expect that Greek and Sanskrit diverged at least 2000 years before they were attested. This gives us the following rough picture:



The Corded Ware culture dates to around 2500 BCE, which is too late for them to have been the Proto-Indo-Europeans.

Turning to PIE culture, we know that the PIE people loved horses. The word for horse, $*h_1 ek^w o$- (which comes out as *eoh* in OE) is actually derived from the PIE for "fast". Horses were the charismatic megafauna of the Proto-Indo-Europeans, like the lions or bald eagles of today. They appeared frequently in names as bahu-vrihi compounds, like

Chrysippus
Xanthippe[13]

in Greek, where -*hippos* meant "horse". In Sanskrit, we had names

śvetāśva-
śyāvāśva-

The Proto-Indo-European people also had wheeled vehicles—they had a great many words relating to wheels and wheeled vehicles. The best known invention of the wheel, however, is concurrent with the development of wheeled vehicles in Mesopotamia—located in southeastern Europe, north of the Caucasus—around 3500 BCE.

This seems somewhat problematic, since the PIE people are dated to at least 500 years prior. However, the speakers could still be geographically close together at 3500 BCE.

The PIE word for wheel is suspected to be a late coinage:

$*k^we$-k^wl-o-

which is a reduplicated form of the word meaning "to turn", i.e., implying "round and round." In Sanskrit, this became

$*k^wek^wlo$- > *keklo-
> *čeklo-
> *čakla-
> čakra- (Skt.)
> cakka-
> cāk (Hindi)

and is the origin of the borrowed "chakra". In Greek, we have

κύκλος

which, in its traditional transcription, is *cyclos*-. In Germanic, we have

$*k^wek^wlo$- > *hwehwla-
> *hwehla- (PGmc.)
> hwēol (OE)
> hwēl/wheel (ME)
> wheel

The invention of the wheel has to be linguistically later than when the Proto-Indo-European language split up. Hittite doesn't have most of these works, but it retains a few of them (e.g., yoke, etc.). Thus, the invention of the wheel must have come in when Indo-European was already a dialect continuum, rather than a single cohesive language.

[13] The wife of Socrates.

PIE Archaeology and Migration

There are many archaelogical remains in eastern Europe of what are called the Kurgan cultures (referring to large, earthen burial mounds). Archaeologist Marija Gimbutas argues that the Proto-Indo-Europeans stemmed from Kurgan culture. The geography is correct, and there are archaeological connections. Descriptions of graves from early Indo-Aryans suggest that they probably practiced Kurgan burials—they certainly piled up earth around graves and had chambers within.

The Yamna culture (3500-2200 BCE), also known as the pit-grave culture, might have been Indo-European. It replaced the Sredny Stog culture, which was located around the Dneiper River, and is the best single contestant for who the Proto-Indo-Europeans were. While it is hard to get from the Yamna to Europe, it is easy to get to India and Iran.

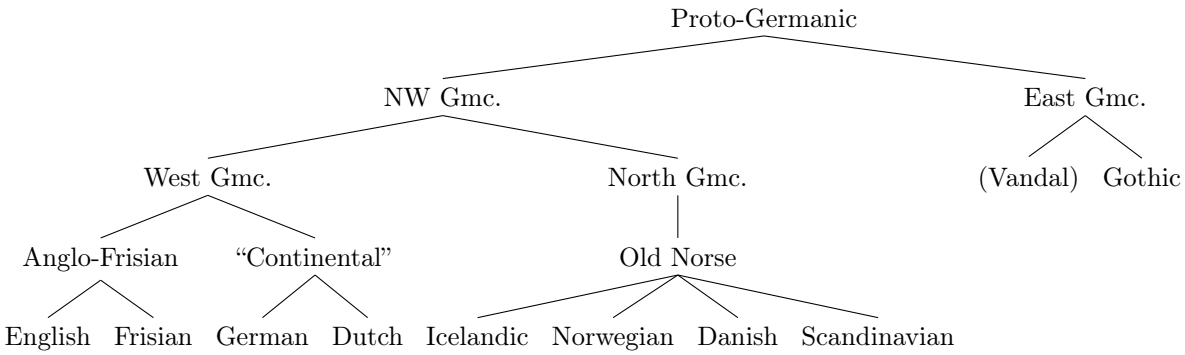
Gimbutas thought that the Yamna people were warlike, had bronze weapons, and conquered Europe. She identified with the "Old Europeans", an older matriarchical society in mainland Europe, and projected male qualities onto Indo-European ancestors. This picture has the Indo-Europeans coming from the steppes and involves the standard migration theory of conquest and population replacement.

There exists a rival theory which also fits (maybe a little suspiciously too well) an archaelogical trend not to see conquest as the means of replacement. This is Colin Renfrew's Anatolian hypothesis. Under this hypothesis, we date the Proto-Indo-Europeans to about the time that agriculture was invented; their spread dates from 7500 BCE and simply accompanied the spread of agriculture. This theory is not linguistically accepted because the date is so early; the diffusion of Indo-European groups does not corroborate with this hypothesized slow diffusion into southern Europe.

Very little else of Indo-European culture is known besides a general idea of a sky god and a dawn goddess, basic notions of social organization, and the beginnings of a reconstructible poetic tradition, with similar scanning devices as those seen in later IE languages.

Lecture 28 — 4/10/13

The original Germanic speakers were the traditional barbarians. They were less intellectually cultivated than the Romans and Greeks were (or believed themselves to be). In the first centures AD, the Germanic people expanded into what is now West Germany, and also expanded slowly southeast into modern Romania and Ukraine.



The quintessential East Germanic people are the Goths; the most important such languages is Gothic. Gothic is the flagship language of the Germanic family—it is the most attested and archaic. We have Gothic primarily from a 6th century attestation of the New Testament, originally translated in the 4th century AD.

The oldest attestation of all Germanic languages is from the 2nd century AD, written in the runic alphabet in Pre-Old Norse (Old Norse actually being 13th century Icelandic). The runic alphabet was written with slanted characters (e.g., ᚠᚾ) so as not to split the grain of the wood they were written on.

Germanic Sound Changes

Germanic was a centum language.

The iconic sound change of the Germanic family was described by Grimm's law, originally called the "First Sound Shift". Grimm's law is one of the reasons people first got interested in sound change. Coupled with the knowledge that Germanic is a centum family, Grimm's law gives us:

$$\begin{aligned} &\text{p, t, k, k}^{\text{w}} && > \text{f, þ, h, w} \\ &\text{b, d, g, g}^{\text{w}} && > \text{p, t, k, kw} \\ &\text{bh, dh, gh, g}^{\text{w}}\text{h} && > \text{b, d, g, gw} \end{aligned}$$

The *gw* further differentiates into just *g* or *w*.

As examples, we have

$$\acute{k}m̥tom > \text{hundred}$$
$$\acute{k}uwon > \text{hound}$$

Comparing with Latin, we have

Lat.	Eng.
pater	father
ped-	foot
cornu	horn

Wulfila, the inventor of the Gothic alphabet, had a letter for the treatment of k^w: ƕ, called the *hwair*.[14]

Proto-Indo-European words rarely have *b*'s, especially as initial letters. Germanic words beginning with *p* tend to actually be borrowed from Latin words (perhaps via French) with *b*, which became *p* by Grimm's law.

One notable exception is

	PIE	Gmc.
5	*penk^{w}e	> f-
4	*k^{w}etwores	> f-

In general, numerals are often influenced by neighboring numerals because they so often appear together (e.g., in counting). Thus, the word for "four" begins with *f*- in all Germanic languages, by analogy with "five".

We turn to some more examples. In other IE languages, the word for "knee" is given by

$$\text{genu, gonu, jānu} < \acute{g}$$

Similarly, we have the Gk. *gnosko* for "to know". This demonstrates *gn-* > *kn-*; the *k* is lost in Modern English, save for in words like "acknowledge", where it is protected by the prefix.

We also have

	PIE	Eng.
"woman"	*g^{w}en-h$_2$, g^{w}n-eh$_2$	> queen
	*dheh$_1$-	> do
	*dhers-	> dare

In **ǵhans-* > goose, the nasal is lost because of its participation in the *-ns-* cluster. In

$$\text{*g}^w\text{hor-mo- > warm (Gk. thermos)}$$

we have an example of $g^wh > w$. Finally, with

Gmc.	Skt.
bindan	bandhati

Grimm's law appears to require a *p* in Germanic—but we recall Grassman's law, which takes $C^h \dots C^h > C \dots C^h$. Indeed, it is the case that *bandhati* < **bhandhati*, which resolves the issue.

We also examine a few cases which don't quite follow Grimm's law. In

Lat.	Gk.	Skt.	Gmc.
stāre	stā-	st-	stand

there is no change $t > \theta$—this change did not occur after fricatives.

One particularly serious exception to Grimm's law is as follows:

Goth.	fadar	broþar
OE	fæder	brōþor
Lat.	pater	frāter

The *-d-* here was pronounced as [ð]—but this leads to a problematic differentiation of

$$t > \text{þ, ð}$$

We see a similar problem in

	OE
pres.	weorþan < *wert-
pret. pl.	wearþ
pret. pl.	wurdon
past. pt.	worden

Similar altercation arises for *f*, ƀ and for *h*, ɣ. This critical exception to Grimm's law is resolved by a later deduced sound change known as Verner's law.

Lecture 30 — 4/15/13

In Vedic Sanskrit, the word accent could still stand on any syllable, and was inherited from the PIE accent (it was frozen later in Sanskrit's history):

PIE	Skt.	Gmc.
ph$_2$tér	pitár-	fadar
bhréh$_2$tēr	bhrā́tar	broþar

Until Verner's law was discovered, nobody thought that the accent could be relevant to sound change. Verner's law explains that the fricatives resulting from Grimm's law vocalized intervocalically *unless* the word accent fell on the preceding syllable. As a result, there was no longer a good and bad treatment for Grimm's law, but instead a good treatment with the accent before the consonant, and a good treatment without the accent.

[14] Wulfila's name means "little wolf", and the diminuitive ending is the same as that of Attila—i.e., Attila the Hun—whose name meant "little father", and had a connotation much like "big daddy" might today.

Umlaut

One of the most characteristic features of Germanic is what is called umlaut, the general term for phenomenon where a sound is altered by another sound, typically another sound in the next syllable.

Umlaut is an assimilation process. If a vowel in one syllable is followed by a high front vowel in the next syllable, the vowel becomes a little higher and fronter, e.g.,

$$[ui] > [üi]$$

Consider, for example, the PIE root noun *$m\bar{u}s$, meaning "mouse". This has nom. sg. *$m\bar{u}s$ and nom. pl. *$m\bar{u}s$-es. In Germanic, these changed as follows:

	sg.	pl.
PIE	*mūs	*mūs-es
PGmc.	*mūs	*mūsiz
WGmc.	mūs	mūsi
OE	mūs	mȳs
ME	mūs	mīs
MnE	[maws]	[majs]

Note that in Proto-Germanic, final -s > -z, and the -z was then lost in West Germanic. Between West Germanic and Old English, then, we see the effects of umlaut. Specifically, it is a case of i-umlaut—$\bar{y} = \ddot{u}$ and changes as a result of i-coloring. As another example, consider

ME	fōt	fōtiz
	foot	fēt (alt. *feet*)

When o was followed by i, it became $\ddot{o}$ [eu].

There were processes called i-umlaut in German, Old Norse, and Old English. The only Germanic language without i-umlaut is Gothic, another way in which Gothic is the most conservative Germanic language. All modern Germanic languages have their own forms of umlaut, and even different forms of i-umlaut.

One such alternate umlaut is u-umlaut, which causes vowels to front and to lift. For instance:

	sg.	pl.
PGmc.	*land	*landu
ON	land	lond

Some modern Germanic languages have the most complex vowel systems in the world because of varying umlaut rules that have acted over time.

Umlaut and Ablaut

We can describe *umlaut* as sound *alteration*; complementarily, we can call *ablaut* sound *alternation*.

PIE ablaut may go back many years to some umlaut process—we don't know how exactly it arose. We don't get back to an origin language when we reconstruct PIE; we just get an earlier ancestral language to all our modern ones.

Consider the PIE infinitive *sed-, meaning "to sit". We will consider its present and causative forms (the latter is formed by taking the o-grade and adding a thematic suffix).

	pres.	caus.
PIE	*sédyeti	*sod-éye-ti
PGmc.	*sitjan	*sat-ja-n
OE	sittan	settan[15]

This example shows both umlaut and ablaut—the -ja- has a short e sound, and induces i-umlaut. Likewise, the root *men-, *$moneye/o$ > Lat. $moneō$ (to cause to think).

Germanic Morphology

When we look at Germanic, it is clearly more evolved than most of the other languages we have studied. In Indo-European, there were three verb forms which could serve as the past tense: the aorist, the perfect, and the imperfect. Recall these:

- The *aorist* is the simple past tense, and uses its own stem.
- The *imperfect* is the continuous past tense, and is just the present stem with secondary endings.
- The *perfect* refers to a state resulting from having just completed an action, and involves reduplication.

Sanskrit and Greek retain all three (although in Greek, the perfect is never really used in a past tense sense). In Latin and Celtic, however, the aorist and the perfect merge:

pres.	simple past	PIE
dūcō	dūxī	*deuk-s- (s-aorist)
cadō	cecidī	(reduplicated perfect)

The Germanic languages are the first branch we've seen in which all three past tenses merge, resulting in only the simple past tense. The great winner of the PIE morphological forms is the perfect—every simple past form in Germanic is traced back to the PIE reduplicated perfect. For example,

"bite"	pres.	sg. perf. (past)	pl. perf.
PGmc.	*bheid-e/o-	*bhebhóid-e	bhebhid-n̥t
OE	bītan	bāt	biton

[15] The word "set" has the sense of "to cause to be in a sitting position.

Note that the sg. perf. is not preserved in the modern English paradigm—when we say the simple past tense "bit", we are using the stem form that was originally the plural of the perfect. This form eventually generalized to the entire past paradigm.

The transparency of the early Germanic past tense goes back to the PIE perfect; for instance, we have

	e-grade	o-grade	0-grade
OE	sing	sang	sungon

In this case, the singular (sang) won out as the past tense—sung is left over in the past participle.

All primary verbs in Germanic—called strong verbs—work like this. But how do things work if you're a secondary verb in Germanic—that is, a verb derived from another verb or a noun? What if you're not **sitjan*, but **satjan* < **sod-eye/o-*? Since **satjan* didn't have a perfect in PIE, any Indo-European language having to make a perfect or aorist with it would have to innovate one.

"to set"	OE	PGmc.
sg.	settan	< satjan
pl.	sette	< satidē

This innovated *-d-* marks the standard English simple past tense, or the weak preterite. The strong/weak terminology comes from Grimm, who thought that changes due to vowel change were somehow more vigorous than changes by suffix. There are lots of verbs which were once weak that are now strong due to analogy.

As we move on to Balto-Slavic, we note one pattern of endings across the IE languages for nouns in the dative or instrumental case:

	dat., instr.
Skt.	-bhiḥ, -bhyaḥ
Lat.	-bus
Arm.	-b-

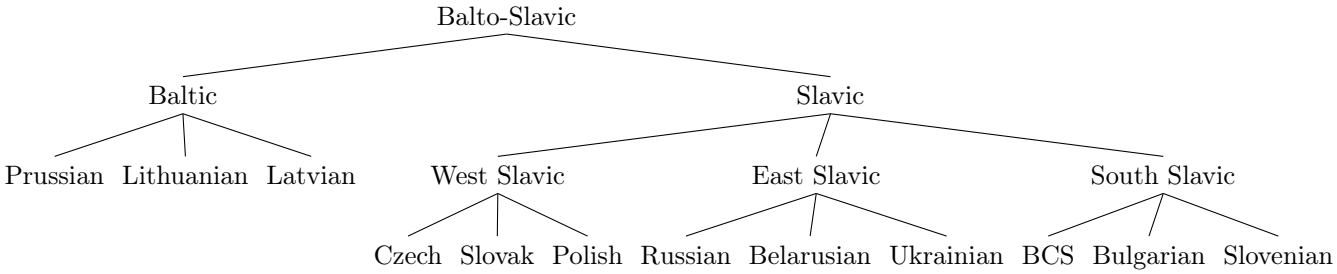
In Germanic and Balto-Slavic, however, these stand-out nominal endings all have an *-m-*. We don't know what exactly causes this.

Lecture 31 — 4/17/13

The Balto-Slavic hyphenation is much less controversial than Italo-Celtic, but somewhat more controversial than Indo-Iranian. It's not that it's not obvious that they go together; linguistically, it is, but there are many political concerns.

For instance, BCS stands for Bosnian, Croatian, Serbian. For linguistic purposes, all three are the same language; however, they are not the same language for their speakers. (Perhaps the most notable linguistic difference, however, is that some are written in Cyrillic while others are not).

Baltic national feelings also are not assuaged by being part of the Balto-Slavic family. Consequently, Lithuanian and Latvian linguists tends to argue against the hyphenation.



Prussian is nowadays connotated with the landowning military class of historical Prussia which staffed the German army. The real Prussians, however, were lost in the Late Middle Ages due to German expansion into their territory; they were conquered in a series of crusades. Various works of Christian content (e.g., catechisms) in the 16th century, all by non-native Prussian-speaking Germans, are attested. This constitutes the full corpus of Prussian, however, since it was extinct by the 17th century as the Prussian people were gradually organized into serfdom.

Baltic preserves a lot of Proto-Balto-Slavic; Slavic, on the other hand, is a lot like a Baltic dialect that then did five hundred other things to make it less archaic.

Many words for "king" are derived from the tribes that were ruled over, e.g.,

"king" < OE cyning < PGmc. *kuningaz < *kunja

**kunja* also gives the MnE "kin". Other examples include

OE þēoden < *þeod, "people"
Welsh brenin < *brigantīnos

Some Proto-Germanic words got borrowed into Baltic (and into Finnish):

Lith.	"priest"	kunịgas
Russ.	"prince"	knjaz′ (kn′az′)

Stripping off various Slavic sound changes gets us closer to the Baltic language:

knjaz′ < OCS kŭnędzĭ

Balto-Slavic Sound Changes

Balto-Slavic is a satəm language:

	PIE	Lith.	OCS
"100"	*ḱm̥tom	šimtas	sŭto

Note that *-om* ⊁ *-as* is not a sound change—it makes no sense. However, neuter stems don't end in *-s*; neuter o-stems end in *-o-m*. Masculine nouns, however, have the paradigm

nom.	-o-s
acc.	-o-m
ins.	-o-h_1

Lithuanian must have merged the neuter into the masculine, much like the Romance languages.

Lithuanian is also the only IE language that gives us the -m̥- in the PIE *ḱm̥tom.

	Lat.	Skt.	Ir.
"100"	centum	śatam	cant

Balto-Slavic and Germanic

Balto-Slavic languages possess a number of Germanic-like features. In both Germanic and Balto-Slavic, for example, the short vowels *o* and *a* merge.

	Goth.	Lith.	OCS	Lat.	Gk.
"night"	nahts	naktīs	noštĭ	nox	
"plow"	arjan	ariu	orjǫ	arātrum	aróō

Neighboring regions often tend to affect one another's languages. As another example of this, most IE languages point to a set of grammatical endings that contain a *-bh-*:

	PIE
ins. pl.	*-bhis
dat., abl. pl.	*-bh(y)os

However, in Germanic and Balto-Slavic, we see

	Gmc.	Lith.	Russ.
ins. pl.	*-m	*-mis	*-mi
dat. pl.		*-ms	*-m
ins. sg.		*-mi	

[16] In OCS, *-mn̥tís does not occur as a free word; rather, we have this derived word, meaning memory.

Nobody's ever really explained this change. It's not a sound change; rather, it's what's called an isogloss, a shared feature that extends over a set of related languages.

Germanic and Balto-Slavic languages also all use nasal endings to make derived words (e.g., to awaken < wake). For example, Lithuanian makes use of a nasal infix:

Lith. bud-, bundu

Shared Characteristics of Baltic and Slavic

The thing that gives Balto-Slavic their own distinctive character as an IE branch is their presentation of the most complicated prosodic system in the IE family—i.e., those phonological characteristics other than segmented, sequential sounds (e.g., length, accent, and intonation).

Any long vowel or diphthong in Proto-Balto-Slavic, as well as clusters like -*er*C-, could either have rising (acute) or falling (circumflex) intonation.

ế	éi	érC
ề	èi	èrC

These are phonologically relevant—for instance, in Serbo-Croatian, long vowels with rising intonation get shortened, but not those with falling intonation.

One final major sound change relates to syllabic consonants:

R̥ > iR

For instance, we have

	PIE	OE	Lith.	OCS	Russ.
"mind"	*-mn̥tís	mynd	mintìs	pa-mętĭ[16]	pámjat′

Note again that Lithuanian has not only conservative phonology but also noun morphology; it keeps the PIE endings in a recoverable form.

Lecture 33 — 4/22/13

Baltic Morphological Innovations

The Baltic language family has comparatively conservative phonology—it keeps final syllables, final *-s*, and most of the PIE consonants. It also has a conservative nominative system, retaining all PIE cases except for the ablative. The neighboring language Estonian—similar to Finnish, which has many, many cases—influenced Baltic in keeping its cases, by placing it in an environment where many cases were promoted. This is a common linguistic phenomenon; imperfect bilingual speakers of collocated languages can effect change.

Old Lithuanian had a number of other cases, including an allative (motion towards; produced by fusing a postposition meaning "to" to the main verb) and two others. It also keeps the dual number. However, when it comes to verbs, Baltic is a lot less conservative.[17]

Baltic languages have done something that even English doesn't do—they've identified all 3rd person forms: 3 sg. = 3 du. = 3 pl. Consider, for example, the Lithuanian paradigm for "to be":

esmi	esme
esi	este
esti	esti

This occurs in regular and irregular verbs—all of them—in Baltic, but it's something Slavic didn't do.

In Baltic, you also don't see PIE perfects or aorists. Consider, for instance,

$$\text{degti} < *\text{dheg}^{\text{w}}\text{h-}$$

the Lithuanian verb for "to burn" (infinitives are marked with *-ti* in Lithuanian). This inflects as

pres.	past.
dega	dego
	degé

None of these forms goes back to PIE verbal forms.

Slavic Sound Change

Slavic exhibits some of the most dramatic phonological developments of any IE language family we've seen. Slavic, like Sanskrit, observed a form of the ruki rule:

$$\text{s} > \text{š} > \text{x } [\chi] \ / \ \text{r\_, u\_, k\_, i\_}$$

Balto-Slavic and Indo-Iranian were probably neighbors at some point, hence their sharing of this feature.

In Russian, all prepositional locative plurals end in *-x*: either *-ax* or *-ix*. These go back to the PIE locative ending **-su*. The ruki rule gives us *-ix* < **-isu*, but we have nothing that would give *-ax* ≮ *-asu*—this form arose by analogy to its *-i-* counterpart.

Unlike in Baltic, we do observe a reflex of the PIE *s*-aorist in Slavic. OCS in particular retained a lot of PIE's verbal morphology, and we have:

	pres.	aor.
"to lead"	vedǫ	věsŭ
	*wedh-	*wēdh-s-
"to convey in a vehicle"	vezǫ	věxŭ
	*weǵh-	*wēǵh-s-

The general mark of the aorist in OCS ends up being *x*, however, because it is analogized even to cases where one would expect the *s*-aorist.

Note the observance of the ruki rule in the second example; it comes into play after *gh* > *k*. Also, note the change *gh* > *z* characteristic of satəm languages.

Slavic is highly palatalized:

$$\text{k} > \text{č} \ / \ \text{\_e, i}$$

that is, preceding front vowels. For instance, we have

	PIE	Slav.
"eye"	*oko	> Russ. oči
"woman"	*$\text{g}^{\text{w}}\text{enh}_2$	> žena, OPr genna

Consider also the Russian for "to be able to", i.e., the auxiliary "can":

mogǫ	možemŭ
možeši	možete
možetŭ	mogǫtŭ

Note that we have *ž* only before *e*, not befofe *ǫ*. This palatalization happened only in Slavic, however, and not Baltic:

PIE	Lith.	Slav.
k^{w}etwor-	keturi	četyre

Unlike in Sanskrit, the ruki rule in Slavic is not complicated by significant vowel change—but it is only the First Palatalization in Slavic. At a later point in time, Slavic underwent

$$*\text{oi}, *\text{ai} > \text{ě } [\bar{\text{e}}]$$

This produced new front vowels that led to the Second Palatalization:

$$\text{k, g} > \text{c } [\text{ts}], \text{ ʒ } [\text{dz}]$$

Our first example is the PGmc. and PBS for "entire, healthy":

$$*\text{koilo} > *\text{kělo-} > \text{čělŭ- (Slav.)}$$
$$> \text{hāl (Gmc.)} > \text{whole (Eng.)}$$

Note that the *wh-* is a misspelling that English took on for some reason. Consider also the PIE for "dry":

$$*\text{sauso-} > \text{sausas (Lith.)}$$
$$> *\text{sauxo} > *\text{suxŭ} > \text{suxói (Russ.)}$$

This second example is cognate with the English word "sear". Note also, from this example, that Slavic loses the final *-s*, which Lithuanian retains.

All syllables in Proto-Slavic were open—however, once this was established, a number of changes happened going into later Slavic languages that made them very rich in closed syllables. Let's look at what happens when we attach a preverb in a Baltic language:

[17] As an aside, English is the only IE language left where the IE *w* is still a *w*.

Lith.	inf.	1 sg. pres.
"to take"	im-ti	imu
"to occupy"	už-im-ti	už-imu

In OCS, this preverb took the form

$$\text{uz} > \text{wuz} > \text{vŭz}$$

We would expect > *vŭz-ĭmti*, but instead, we nasalize the vowel and delete the nasal to avoid the closed syllable:

$$\text{inf.} \mid \text{už-im-ti} > \text{vŭz-ĭmti} > \text{vŭz-ęti}$$

Meanwhile, in the present, we already have all open syllables, so our changes are minimal:

$$\text{1 sg. pres.} \mid \text{vŭz-imǫ}$$

In Russian, however, the picture changed dramatically:

inf.	vŭz-ęti > vŭz′-ęt′i
	> vŭz′-at′i
	> vz′-at′
pres.	vŭzĭmǫ > vŭz′ĭmu
	> vŭz′mu
	> voz′mu

Going into Russian, we start from the end of the word and delete all odd-numbered short vowels; then, the surviving short vowels underwent

$$\text{ŭ, ĭ} > \text{o, e}$$

Lecture 34 — 4/24/13

Armenia got Christianity pretty early; the Bible was translated into Armenian around 400 CE, and it is our earliest Armenian text. It was written in an alphabet originally designed by St. Mesrop (who was also the designer of the Georgian alphabet). This alphabet doesn't look familiar except for a few letters that resemble their Greek analogues, but it's basically an altered Greek alphabet. Armenian has always been easy to read.

Since the 5th century, it has been the classical Armenian period. Today, Armenian splits as two dialects: Eastern Armenian and Western Armenian. Eastern Armenian is the language of the current Republic of Armenia; Western Armenian is the Armenian of the vast Armenian diaspora.

Armenian has its own distinctive Christian and artistic and cultural traditions. It's a language that has gone off in a different direction than the other IE languages, and has taken turns especially in its phonology—Armenian is known for its weird sound changes.

[18] May be borrowed from Iranian.

Armenian Sound Changes

Armenian is a satəm language (it borrowed heavily from Iranian), but the Armenian word for "hundred" is unknown. In fact, we only have about 400 Armenian words—so consider instead the word for "heart":

	PIE	Arm.
"heart"	*ḱerd-, ḱr̥d	sirt

Note the sound change from the PIE *ḱ-* and *-d*. Armenian underwent a massive consonant change, similar to Germanic under Grimm's law (although much less famous).

Let's begin with the PIE voiceless stops. In general, these went to voiceless aspirates:

Change	PIE	Arm.	
p > h	*ph$_2$ter >	hayr	"father"
	*pod >	otn	"foot"
t > t‘ [t^h]		t‘e	"that"
ḱ > s		sirt	"heart"
k$^{(w)}$ > k‘, č‘	*k^{w}etwor- >	č‘ork‘	"four"

This was an intermediate step on the way to having voiceless fricatives (cf. Germanic). Note that the shift from *p* proceeded as $p > p^h > h$, and it was sometimes lost before certain vowels—much like in Celtic. Also note that the ‘ marks aspiration.

Like in Germanic, voiced stops became voiceless stops in Armenian:

Change	PIE	Arm.	
d > t	*dek̑m̥ >	tasn	"ten"
ǵ > c [ts]		cin	
g$^{(w)}$ > k	*g^{w}enh$_2$ >	kin	"woman"

There are no good examples for *b* since the PIE *b* itself barely existed.

Finally, voiced aspirates became voiced stops:

Change	PIE	Arm.	
bh > b	*bhereti >	berem	"to bear"
dh > d	*dheh$_2$ >	dir	"put"
ǵh > j [dz]	*ǵhes- >	jeṙn	"hand"
		jiwn	"winter"
g$^{(w)}$h > g	*g^{w}hen- >	gan[18]	
g$^{(w)}$h > ǰ	*g^{w}hen-ye-ti >	ǰnǰem	"to beat"

For the last change, we have

$$\text{*g}^w\text{en-ye-} > \text{ǰinje-} > \text{ǰnje-} > \text{ǰnǰem}$$

Under the skin, these changes are quite like what Germanic did. Half a century ago, it was in vogue to say that PIE looked more like this, and that all the other IE languages changed.

The voiced stops are pronounced differently in different Armenian dialects, including the modern Eastern and Western. Western Armenian has the normal pronunciations, but in Eastern Armenian, [b, d, g] are now actually the voiceless aspirates, [p^h, t^h, k^h]. There is enough variance in the phonology of the voiced stops, though, that we can't really tell whether or not the voiced aspirates > voiced stops sound change actually occurred—Armenian may have retained the voiced aspirates.

Armenian Morphology

Does Armenian have the -bh- or -m- nominal endings? The former—they have a firm -bh- reflex:

Arm.	root	inst. sg.	inst. pl.
"father"	hayr	harb	harbk‘

Armenian is also the third branch of the IE family, in addition to Greek and Indo-Iranian, that uses the PIE augment. Recall the Sanskrit and Greek:

"to carry"	Gk.	Skt.
"carries"	φέρει	bhárati
"carried"	ἔφερε	ábharat

The Armenian aorist adds endings to the stem:

beri	berak‘
berer	berēk‘
eber	berin

The rule in Armenian is that you keep the augment if the word is one syllable; originally, we expect all these forms had the *e*-augment. We suspect that Armenian was geographically close to Indo-Iranian and Greek for this reason.

Consonant Clusters

Armenian is famous for its weird phonology. When Armenian finds consonant clusters, unexpected things happen. Consider the word for "two" in Armenian:

$$*\text{dw}\bar{\text{o}} > \text{erku}$$

This is actually an instance of a completely regular sound change,

$$\text{dw} > \text{erk}$$

which proceeds along the following lines:

- *d* became *r* via an intermediate flap, but Armenian did not tolerate initial *r*'s and prepended an *e*-.
- *w* > *g* > *k*; the second change was due to the consonant shift.
- $\bar{o} > u$ in general.

This is an exact cognate to other IE words for "two" (including the English), even though it has no resemblance!

In Armenian, the word for "mother" has very clear etymology:

$$*\text{m}\bar{\text{a}}\text{t}\bar{\text{e}}\text{r} > \text{mayr}$$

Let us look at the sound changes that the PIE **bhrātēr*, "brother", underwent:

*bhrātēr > *brayir	consonant and vowel shift; t > t^h > θ > y / V_V
> *erbayir	addition of *e*; metathesis of Cr > rC
> *elbayir	dissimilation of multiple *r*'s
> *ełbayir	l > ɬ/ _C ("dark" l) ɬ may have been [ɣ]
> *ełbayr	[yexp^{h}áyr]

Lecture 36 — 4/29/13

North of the Himalayas bordering northern India, in the southern part of China's Xinjiang province, is the Tarim Basin, the desert region surrounding the Tarim River. This was an important part of the Silk Road coming from China. The routes bordering the north and south of the basin are nowadays a "wild country" only navigable by archaeologists, and home to the Turkic Uygur (or Uighur) people.

The Tarim Basin, however, was also a pre-Uygur center of distinctive Buddhist culture: oasis towns dotted the basin's rim. In the 20th century, these towns were found to house lots of old literary documents (c. 6th-9th century) from many different languages, preserved and protected by the desert sand. Two of these languages, discovered along the upper route, were obviously Indo-European. These were Tocharian A (the Eastern Tocharian) and Tocharian B (the Western Tocharian, which was slightly more archaic).[19] Both Tocharian languages eventually died out and were replaced by the unrelated Uygur, a Turkic language.

Tocharian belongs in the same time period as the earliest Old English and as Old High German. It was written in an Indian-based alphabet, which was known and studied; its texts were mostly translated from Buddhist texts that were known in other languages. Consequently, the Tocharian languages were pretty quickly readable and were soon observed not to be identifiable with any of the other IE families.

[19] Tocharian is actually a misnomer, based on an incorrect supposition as to who the Tocharian people were.

The Tocharians obviously separated from the main mass of IE pretty early and went their own way. Somewhere along the way, however, they acquired a great deal of lexical vocabulary that isn't recognizable as *anything*—though some Tocharian words are Turkic in origin and other words are Chinese, many have no apparent etymology whatsoever.

Tocharian Sound Changes

Tocharian did many things that are atypical of other IE languages. One very prominent such change is that Tocharian merged the three sorts of PIE stops—voiced, voiceless, voiced aspirate—into a single stop system.

	TB	PIE	
"father"	pācer	< *p-	
"tall"	tapre	< *d-	< *dhub-ro-, "deep"
"brother"	procer	< *bh-	
"hundred"	kante	< *ḱ-	
	ákau	< *ǵ-	Lat. agō
"arm"	pokai	< *ǵh-	Gk. pēkhus

Tocharian dentals are the only exception—we have

$$\text{d} > \text{ts}$$

although the rest just come out as > *t*. We know Tocharian didn't have these lost sounds in part because the Indic scripts did have letters for them, e.g., for *b*. Also, note that the form of "hundred" classifies Tocharian as a centum language.

Tocharian underwent a far, far reaching palatalization to dissimilate these merged stops. For instance, we have

$$\text{pacer} < \text{patēr}$$

where the *c* was palatalized from the *t* preceding a front vowel. Other palatalized sounds in the Tocharians include:

$$\text{c, ś, ṣ, ly, ñ}$$

Tocharian's vowel changes are complicated, but they're mostly figured out now—though not entirely. Many vowels became ə, which is unfortunately transcribed as *ä*. Like schwas in general, they tend to get lost, especially between two consonants as part of an open syllable.

$$\text{nekäta} > \text{nekta}$$

This is known as syncope.

Both Tocharians are full of syncope procedures. In Tocharian A, these are so ubiquitous that they produce very odd looking words by typical IE consonant-vowel structure:

TA	Etym.	
tpär	TB tapre	"tall"
ype		"country"
wtaṣ	< *dwito-	"again"
rse		"hate"
ñkät		"god"
tkaṃ	< *dhǵh-	"earth"
ylār		"fragile"
wmār		"jewel"
rtär	< *$h_1$rudhro-	"red"
nkiñc		"silver"

Note that most of these words don't have clear etymologies—i.e., all the ones for which we haven't listed one. Some have guesses, but they are far from definite.

Tocharian Morphology

Another bizarre feature of Tocharian is the creation of secondary cases. Most IE languages lose cases, but this tendency is reversed in Tocharian. Tocharian started life like a normal IE language: it lost the ablative, dative, and instrumental, really only keeping the nominative, genitive, accusative, and vocative. Finding themselves in a case-rich neighborhood, the Tocharians reversed gears and proceeded to create new ones, by taking the old accusative and proceeding to fix postpositions to them.

In Tocharian B, where we have

	"elephant"	"horse"
root	oṅkolmo	yakwe
obl. pl.	oṅkolmaṃ	yakweṃ

Let us consider, as an example, the comitative case, which denotes accompaniment. This is produced simply by affixing the postposition *-mpa*, meaning "with" to the oblique (acc.) form of the noun, e.g.,

oṅkolmaṃ-mpa
elephants-with

This construction varies between Tocharian A and B. Interestingly, there is a choice of whether to use the secondary form more than once in a phrase of related things. For example, we can have

oṅkolmaṃmpa yakweṃmpa
elephants-with horses-with

but we could also simply have

oṅkolmaṃ yakweṃmpa
elephants horses-with

or

oṅkolmaṃmpa yakweṃ
elephants-with horses

One or the other *-mpa* can be dropped, as in the English, "with elephants and horses" rather than "with elephants and with horses." This is known as group inflection, and the same is true of adjectives.

Tocharian A is the only IE language in which there are not only separate forms for he and she, but also separate gendered first-person pronouns:

"I" (masc.)	"I" (fem.)
näṣ	ñuk

Historically, *näṣ* probably meant "me", but it was repurposed.

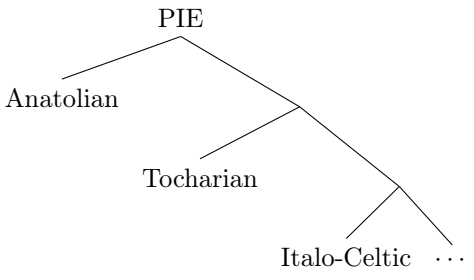
Tocharian and Indo-European

Tocharian was originally thought to be Italic or Celtic that strangely migrated to Asia. The basis of thinking this was twofold: first, Tocharian was a centum language located among satəm languages; and second, it had *-r-* in the verbal middle.

	TB "lead"
3 sg. act.	āśäṃ < *age-
3 sg. mid.	āśtï < *agetor

Originally, only Italic and Celtic had a primary *-tor* in the middle, so placing new languages with the *-r-*'s was hard. Tocharian and Hittite helped make it clear that these *-tor* endings were the correct PIE setup. When mutually distant branches share features, we set up the ancestor with that feature and say that the other branches situated closer together were the innovators.

This gives us a picture of the order of departure of child languages from PIE:



The easiest feature to illustrate this with is the PIE *s*-aorist. In Sanskrit and Greek, the way to make aorists (which is productive at least in Gk.) is to do:

Gk.		Skt.	
phileō, "to love"		ayā, "to ride/go"	
ephilē-sa	-s-	ayā-sam	-s-
-sas		-sīḥ	
-se		-sīt	

[20]The *ë* is pronounced as ə.

This is just called the first aorist in Greek, and there are analogues of this in Celtic, Slavic, Welsh, Latin (as the perfect), etc.

In Tocharian, however, the *s* doesn't run through the entire paradigm. If we want to make the *s*-preterite of a verb in TB, we have

nek-, "destroy"	sg.	pl.
1st person	nekwa	nekam
3rd person	neksa	nekar

Here, we only see the *s* in the 3 sg. Similarly, in Hittite, we have

nai-	sg.	pl.
1st	neḫ̬ḫun	neyawen
3rd	naiš	nair

The reconstructed *s*-aorist probably started life with *-s-* only in the 3 sg.—after Hittite and Tocharian left, it got extended to the entire paradigm.

Lecture 37 — 5/1/13

Albanian and Other Languages

There is only a small percentage of the PIE lexicon that survies in Albanian. For one example, we have

	Alb.
"king"	mbret
"kingdom"	mbrëterësh[20]

which are cognates to the Eng. *emperor* and Lat. *imperator*. Albanian has quite a bit of morphology, but it doesn't trace back easily to the PIE stuff. It is similar to English in this way—modern, with distance from PIE—but unlike English, it doesn't have languages like Gothic or Old English to bridge the gap.

There are lots of other cases where we have no record of languages spoken by ancient groups. We don't really know what the Macedonians spoke—the ruling groups all spoke Greek, but that doesn't mean these people (like Alexander the Great) all spoke Greek, or that their grandfathers spoke Greek. Probably, they spoke a Macedonian dialect of Greek, but also a native language that was not Greek.

These ancients are extremely uncurious about languages—they tell us stories about what elephants they saw, but no Greco-Roman author ever undertook to tell us how such and such a barbarian language worked. Other people whose languages we don't know include those of Asia Minor, and of Northern and Central Europe.

We know of several peoples of Sicily whose languages don't seem to be Indo-European. The Etruscan

people neighboring Rome (cf. Tuscany) spoke an unknown language—we know enough to think it's *not* Indo-European.

In Britain, near the beginning of Early Middle Ages, were the Picts (the most famous Pict was Macbeth, though he may not have spoken Pictish). There are claims that the Picts were some kind of Celts, but also other claims that the Picts were Indo-European but not Celtish, or not European at all.

Let's talk about the common tendencies of all Indo-European languages. For the most part, we haven't followed these languages down to the modern period, so let's talk about some later changes.

Later Morphological Changes

Apart from Balto-Slavic languages, we see almost no nominal or verbal inflections today. The work of case inflection can be easily done by adpositions and word order instead—consequently, early IE languages have much freer word order than modern languages. Meanwhile, auxiliary morphosyntactic verbs are the biggest replacement for verbal inflection.

English actually has a more complex tense-aspect system than most older IE languages, e.g., Sanskrit. This arose as in the following:

"I have the exams graded."
"I have graded the exams."

The latter came to be a paraphrase for the former—though this is a fairly shallow change, and happened recently.

	Itl.
"I have seen him."	l'ho veduto
"I have seen her."	l'ho veduta

We can categorize languages based on their morphological traits (though these categorizations may not necessarily reflect historical relationships).

- analytic—e.g., English. Using separate little auxiliaries or adpositions.
- synthetic—e.g., Sanskrit. Morphology is glommed on with morphology.

In analytic languages, verbs gain inflection from auxiliaries and adverbs, while nouns gain inflection from adpositions.

Consider these inflections in various Romance languages of the Lat. infinitive *cantāre*, "to sing":

	"I will sing"
Lat.	cantābō
Fr.	chanterai
Sp.	cantaré
It.	canterò

In Portuguese, you can stick pronoun objects between the stem and the r-ending. Indeed, as this might suggest, the Romance future is actually derived from

$$\text{cantāre} + \text{habeō}$$

Other verbs can also be used periphrastically to make new tenses.

Later Sound Changes

One modern sound change that is common among IE languages is the tendency for stops to become fricatives:

$$\text{VC(V)} > \text{VF(V)}$$

Consonants are obstructions of air, and there are two axes that distinguish consonants from vowels: obstruction and voicing. (Vowels are usually voiced, though there are exceptions in Japanese). Consonants will tend to move towards vowels intervocalically, gradually losing these traits.

$$\text{stop} > \begin{matrix}\text{fricative}\\ \text{voiced stop}\end{matrix} > \text{voiced fricative} > \text{glide} > \varnothing$$

Among the Romance languages, we see this in

	"father"	Consonant
Lat.	patre(m)	
It.	padre	vcd.
Sp.	padre	vcd. fric.
Fr.	pere < peðrə < padre < patre	∅

Similarly, among Sanskrit's descendents, we have

	"foot"	"four"
Skt.	pāda	catvārah
Hindi	pai	cār
Bengali	pa	cār

and in general, we see in these languages the same kinds of morphological and phonological reductions as we do in the Romance languages w.r.t. Latin.

All of these sound changes are known as weakening or lenition of consonants after vowels. Palatalization, meanwhile, is a kind of assimilation: sounds changing to match other sounds. In palatalization, the consonants get closer to the vowels they precede.

Lecture 38 — 5/3/13

PIE Review

Recall that in PIE morphology, both nominal and verbal, we have thematic stems which end in the thematic vowel *-e/o-*. Note that the thematic vowel is totally separate from ablaut, even though its variation looks the same—and, in fact, thematic stems don't undergo ablaut:

Case	"horse"
nom.	*$h_1$ék$^{\text{w}}$-o-s
acc.	-o-m
gen.	-o-s(yo)
inst.	-o-h_1

The thematic vowel varies with the paradigm:

"carry"	
*bhér-o-h_2	-o-mes
-e-si	-e-te
-e-ti	-o-nti

For athematic verbs, the nom./acc. use strong stem, and the other forms use the weak stem:

Case	*ph_2-tér / *ph_2-tr-′
nom. sg.	*ph_2-tér-s > *ph_2-tḗr
acc. sg.	*ph_2-tér-m̥
nom. pl.	*ph_2-tér-es
gen. sg.	*ph_2-tr-és
dat. sg.	*ph_2-tr-éi
inst. sg.	*ph_2-tr-éh_1
	"father"

Let's consider some words relating to the mind.

"mind" *mén-ti- / *mn̥-téi-

mén- meant "bring to mind"; meanwhile, *-ti-* marked abstract nouns. We also have

"to think" *mn̥-ye/o-

where *men-* is the root aorist.

As we can see here, in PIE, stems consists of a root plus a suffix or multiple suffixes—everything to the left of the inflectional suffix is the stem. For instance, **ph₂tér* has a root **ph₂-*, which doesn't mean anything by itself (compare to *cran-* in *cranberry*), with a suffix **-tér* relating to family members (mo*ther*, bro*ther*) or agent nouns.

Some other PIE morphologies:

- root present:

  "to be" $h_1$és-ti / h_1s-énti

- nasal infix roots, **-né-* ∼ **-n-*, e.g., the nasal infix present **yeug-*

"to yoke"	*yu-né-g-	*yu-n-g-
	*yu-né-g-ti	*yu-n-g-énti

  Most alternating stems don't like to have multiple full grades—some predecessor to PIE had both accent vowels, which surfaced as *e/o*, and unaccented vowels, which got deleted.

- neuter nouns: nom. and acc. are the same.

IE Families Review

We recap some of the major features which we use to draw distinctions and commonalities among the various IE language families:

- The centum / satəm isogloss.
- *-r* endings in the middle—this is a primary particle, which gives *-tor* in the middle; in the present, we have *-ti*. We see this in some IE languages:

Lat.	Toch.	Hitt.
-tur	-tär	-ta-ri

- Grassman's Law,

  $$C^{h}\ldots C^{H} > C\ldots C^{H}$$

  occurs independently in Skt. and Greek.

- The ruki rule in Indo-Iranian and Balto-Slavic:

  $$s > š \;/\; r\_,\ u\_,\ k\_,\ i\_$$

- *bh-* vs. *m-* endings. The inst. pl., and the dat. and abl. pl. (and inst. sg. in some languages), have odd, heavier, more massive endings in many languages—Sanskrit, Irish, Latin, Celtic, traces in Greek. In the Germanic and Balto-Slavic languages, we have *m-* instead:

Skt.	Lith.
-bhis	-mis

  These endings have English reflexes: *seldom*, *whilom* (obsolete, meaning "sometimes"; the dat., inst. pl. of *while*).

- Merging of a/o in the northern languages and Indo-Iranian.

We also review some general tendencies of IE sound changes.

- Sound change tends to affect the end of the word more than other places (unless the language has a fixed final stress). Most languages have serious constraints on what consonants can end a word—e.g., no voiced consonants in German or Russian.

	[t]	[d]
Russ.	gorod	goroda
Gm.	Rad	Räder

- Many IE languages lost the final *-s*, either as a sound change or just by haphazardly losing it—e.g., Lat. *lupus*, acc. *lupum*, has no trace of the final consonant.
- Types of vowel changes include lowering, raising, syncope, apocope (i.e., terminal syncope), umlaut, and compensatory lengthing. As an example of the latter, we have:

$$*\text{dheh}_1\text{-tor-} > \text{dh}\bar{\text{e}}\text{-tor-}$$

This is the strong stem agent noun of "put".

- *y* and *w* are often lost (*w* may become *v*). *l* and *r* are typically retained.
- The syllabic $\underset{\circ}{l}$ and $\underset{\circ}{r}$ generally became consonant-vowel pairs:

Skt.	Celt.	Lith.	Gm.	Hitt., Arm., Gk.
$\underset{\circ}{r}$, v$\underset{\circ}{r}$ka-	ri	ir	ur	ar (also Gk. ra)