25 November 2014
Linguistics 203

Aaron J. Dinkin
aaron.dinkin@utoronto.ca

Lecture 10
The prehistory of English: Indo-European

We've already seen some examples of **language families**—
groups of languages all descended from the **same ancestral language**,
via **different** language changes acting on **different regional dialects**.
Thus the **Romance** languages all descend from different dialects of **Latin**,
and the **Germanic** languages from different dialects of **Proto-Germanic**.

Latin and Proto-Germanic both date back approximately 2000 years.
But we can trace back language families **even further than those**!
Germanic, Latin, Greek, and more come from an **even more ancient** language,
which is known to us as **Proto–Indo-European** (PIE).
Thus Germanic, Romance, etc. are all **subfamilies** of the **Indo-European** family.

Proto–Indo-European was spoken **at least 5000 years ago** and **was not written**;
we have **no direct evidence of it**.
Everything we know about it is by **comparing the similarities and differences**
between the languages descended from it, and **reconstructing** what the
ancestral language must have been like.

Subfamilies of Indo-European include:

- **Germanic**
- **Italic** (including Latin and its descendants the **Romance** languages)
- **Greek** (a subfamily with only one language in it)
- **Celtic** (Irish, Scottish Gaelic, Welsh, and a few others)
- **Indo-Iranian** (languages of south Asia and Iran: Hindi, Persian, etc.)
- **Balto-Slavic** (languages of eastern Europe: Russian, Polish, Latvian, etc.)
- **Albanian**
- **Armenian**

The English linguist William Jones was the first to propose (in 1786!)
that the ancient Indian language **Sanskrit** was related to European languages.

> "The Sanskrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, **than could possibly have been produced by accident**; so strong indeed, that no philologer could examine them all three, without believing them to have **sprung from some common source**, which, **perhaps, no longer exists**: there is a similar reason, though not quite so forcible, for supposing that both the Gothic and the Celtic… had the same origin with the Sanskrit; and the old Persian might be added to the same family…"

Jones's key observations:

- **Systematic similarities** between languages are evidence of a common origin.
- The ancestral language **might itself be unknown**—
  e.g., Latin isn't derived **from** Sanksrit, or vice versa.

(This way of thinking about how **languages** change and are related to each other was very influential on Charles Darwin's theories of how **animal species** evolve!)

Here are some of the similarities Jones was referring to:

	Latin	Greek	Sanskrit
'father'	*pater*	*patēr*	*pitā*
'is'	*est*	*esti*	*asti*
'two'	*duō*	*duō*	*dvā*
'three'	*trēs*	*treis*	*trayas*
'ten'	*decem*	*deka*	*daśa*

The similarities exist because these groups of words are **cognates**—i.e., words in **related** languages derived from the **same** word in the **ancestral** language.

The Proto–Indo-European word for 'father' **underwent various changes** as different dialects of PIE evolved into Latin, Greek, and Sanskrit—
but **not so many changes** as to eliminate the similarity between them.

We can use similarities such as these to **reconstruct** facts about Proto–IE:
e.g., it's very likely that the PIE words for 'ten' and 'two' began with **d*.

We use an **asterisk** to mark words and sounds **reconstructed** for an ancestral language, to remind us that we don't actually have **direct** evidence for them.

still "guessing"

Cognates in related languages **aren't necessarily as similar** as those examples;
e.g., although the Latin, Greek, and Sanskrit 'father' cognates all start with *p*, the English cognate (i.e., *father*) starts with *f*.

所以形异而声同/近

But related languages have **regular patterns of sound correspondences**.
English words with *f* **regularly** have cognates in other IE languages with *p*:

English	Latin	Greek	Sanskrit
father	*pater*	*patēr*	*pitā*
foot	*pedem*	*poda*	*padam* (accusative case)
fish	*piscis*		
five		*pente*	*panca*
fire		*pȳr*	

This is because of the **regularity of sound change**:
When a language undergoes **changes in pronunciation**
the **same sound** in the **same context** changes the **same way** in **all words**.

This is **different** from semantic change, where "every word has its own history":
semantic changes affecting one word may not affect similar words similarly.

Sound change **only affects the sounds**, independent of what words they're in.

So patterns like the above let us say PIE **p* **regularly becomes *f* in English**,
and so English *f* **regularly corresponds** to *p* in cognates in Latin, Greek, etc.

Proto–Indo-European is reconstructed as having **3 sets of stop consonants**:
voiceless, **voiced**, and **voiced aspirate**".
(**Aspiration** is an *h*-like puff of breath after a stop.)

	labial	alveolar	velar	labio-velar
voiceless	**p*	**t*	**k*	**kw*
voiced	**b*	**d*	**g*	**gw*
voiced aspirate	**bh*	**dh*	**gh*	**gwh*

These underwent a **distinctive series of changes** in the **Germanic** subfamily,
distinguishing it from the other Indo-European families.
These patterns were formulated as a **rule** by **Jakob Grimm** in 1819,
and are collectively known as **Grimm's Law**:

- **voiceless** stops become voiceless **fricatives**
- **voiced stops** become **voiceless** stops
- **voiced aspirated** stops become normal unaspirated **voiced stops**

Proto–Indo-European				Germanic		
**p*	**t*	**k*	becomes	*f*	*θ*	*h*
**b*	**d*	**g*	becomes	*p*	*t*	*k*
**bh*	**dh*	**gh*	becomes	*b*	*d*	*g*

The results of Grimm's Law are most visible **at the beginnings of words**;
in other contexts, other sound changes often interfere with or obscure its results.

Influences of Grimm's Law

The discovery of Grimm's Law was **very influential** in the history of linguistics:

- It demonstrated Germanic was part of the Indo-European family
- It showed that family relationships could be established not just by **direct similarity** but by **consistent correspondences** between **different sounds**
- It showed that **sound change is regular**, affecting **all words** in which a given sound appears.

In Greek and Latin, the **voiceless** and plain **voiced** stops remained **unchanged**,
so the correspondences with English are easy to see:

changed

English	Latin	Greek
three	*trēs*	*treis*
thin	*tenu-is*	
hound	*can-is*	*kyōn*
heart	*cor*	*kard-iā*
two	*duō*	*dyo*
ten	*decem*	*deka*
kind	*genus*	*genos*
carve		*graph-ein* ('write')

remained

(Examples of English *f* corresponding to Latin/Greek *p* are on the previous page.
There aren't a lot of good examples of English *p* corresponding to Latin/Greek *b*;
The sound **b* seems to have been very rare in Proto–Indo-European for some reason.)

what changes as well is voiced aspirates (in Germanic, Latin & Greek)

The PIE **voiced aspirates** changed to **other sounds** in Latin and Greek as well:
in Greek, they became **voiceless aspirates;** in Latin they became **fricatives**.
(Well, that's what happened **at the beginning of a word**, anyway.)

PIE	*bh	*dh	*gh
Germanic	*b*	*d*	*g*
Latin	*f*	*f*	*h*
Greek	*ph*	*th*	*ch*

Cognates demonstrating these correspondences:

English	Latin	Greek
bear 'carry'	*fer-ō*	*pher-ō*
brother	*frāter*	*phrāter*
do	*faci-ō*	*ti-thē-mi* 'put'
dusk	*fusc-us* 'dark'	
gold	*helv-us* 'yellow'	*chlor-os* 'greenish-yellow'
garden	*hort-us*	

.. .. have is not cognates

Some **words that look like cognates** aren't necessarily actually cognates,
as demonstrated by the **regular** correspondences.
English *have* and Latin *hab-* **both mean 'have'** and look very similar,
but via Grimm's Law, Germanic *h* and Latin *h* don't correspond to each other,
so the two morphemes can't actually be cognates.
(*Have* is actually cognate with *cap-* 'take', and Latin *hab-* is cognate with *give*,
obeying the Grimm's Law correspondences.)

exceptions

and more exceptions on the other hand

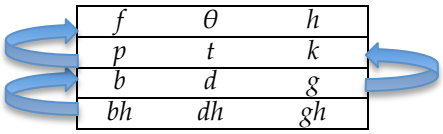
On the other hand, the correspondences can **also** demonstrate how words that
might **look** very different are actually **quite closely related**:
Greek *kannabis* (borrowed into English as *cannabis*, meaning 'marijuana') and
English *hemp*, referring to the same plant, have **no sounds in common**,
but they are **exact cognates** of each other.
Recognizing that English *h* and *p* correspond to Greek *k* and *b* via Grimm's Law
makes this cognate relationship easier to understand.
Obviously the English word *hemp* has experienced **more** sound changes that just Grimm's Law—
vowel shifts and deletions occurred, and also **place assimilation** of **n* → *m* / __*p*.
This is also one of the **few good examples** of Indo-European ***b* becoming *p* in Germanic.**
Kannabis* itself was **probably a loanword in Proto–Indo-European—
demonstrating that **borrowing** has been part of language **as far back as we can reconstruct**.

By knowing the sound changes and correspondences, it is often possible to
deduce which of two morphemes is native and which is borrowed.
E.g., consider the **doublets** *foot* and *ped-* (as in *pedal* and *pedicure*).
We see a correspondence between *f* and *p*, and between *t* and *d*;
these are the correspondences we'd expect to see under Grimm's Law,
if *foot* is a **Germanic** words and *ped-* a cognate from some other IE branch.
So we can infer that *foot* is the native Germanic word for 'foot',
and *ped-* is **borrowed** into English.

Grimm's Law is a type of sound change known as a **chain shift**:
a series of sound changes where one change leaves a **gap** in the system, and then another sound moves to **fill** that gap, and so on.

So in Grimm's law…

- **voiceless stops** become **fricatives**, leaving a **gap** for the voiceless stops;
- the **voiced stops** become **voiceless** and **fill that gap**, leaving a new gap;
- the **voiced aspirates** become plain **voiced stops**, and fill **that** gap.

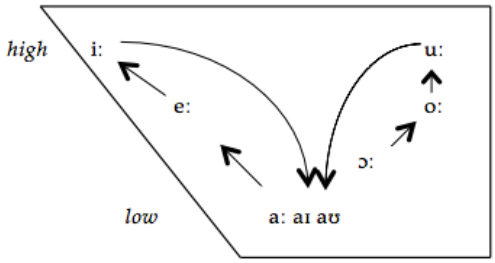


f	*θ*	*h*
p	*t*	*k*
b	*d*	*g*
bh	*dh*	*gh*

The **Great Vowel Shift** between Middle and Early Modern English was **also a chain shift**:

- **long high vowels** become **diphthongs**;
- **long mid vowels** become **high**;
- and **long low and lowish vowels** become **mid**.

It's also "filling the gap"



Many more changes occurred between Proto–Indo-European and Germanic; Grimm's Law is just the **most sweeping** and **most distinctive**.

Latin and Greek also obviously underwent many sound changes, differentiating them from Proto–Indo-European and each other.

The **different sets of sound changes** they experienced led Latin and Greek to have somewhat **different overall inventories of sounds**.

This often allows us to tell **which language a morpheme was borrowed from**.

Greek **developed** sounds represented by *ph*, *th*, *ch*, *y*, *z* through sound change; it **lost** the sounds represented by *j*, *qu*, *v* (actually [w]), which existed in PIE; and **unlike** Latin, ancient Greek **never developed** *f*.

So borrowed morphemes containing ***j*, *qu*, *v*,** and ***f*** are generally **from Latin** and those containing ***ph*, *th*, *ch*, *y*, *z*** are very likely to be **from Greek.**
(Well, not counting ones borrowed from sources **other** than Latin or Greek, of course.)

This is helpful to know because **most of the time**, English prefers to combine Greek morphemes with other Greek morphemes, and Latin with Latin.

Thus we have words like *metaphor* and *transfer*, not "*metafer*" and "*transphor*": *meta-* and *-phor* are **Greek** for 'beyond' and 'carry'; *trans-* and *-fer* are Latin.
(Exceptions exist, though; *television* and *monolingual* are recognizable ones.)

Very little is known about the **people** who spoke Proto–Indo-European;
the best guess is that they **lived** somewhere near the present-day Ukraine,
and their **language** spread out from there over several millennia.

But it's possible to get some clues about where and how they lived based on the **words of their language** we can reconstruct.
E.g., we can reconstruct morphemes for 'snow', 'salmon', and 'birch tree',
so it's likely they lived in a part of the world where those could be found.
(Those are **sneigwh-*, **laks-*, and **bherəg-*, if you're interested.)
The words **kwekwlos* 'wheel', **arətrom* 'plow', and **dyeus* 'god', among others,
tell us something about their technology, agricultre, and religion.

The Proto–Into-European **language** appears to have been heavily **inflected**—
with several **verb tenses** inflected for subject agreement,
two or three **genders**, and eight or nine **noun cases**—
and to have had relatively **free word order**.

Earlier descendants of PIE, such as Latin, Greek, and Sanskrit, tend to **share** many of these features; later descendants may have lost many of them.
E.g., modern **English** is very **weakly** inflected, with no grammatical gender, no case, and only one subject-agreement affix.

Some **morphological** and **allomorphic** patterns that were **productive** in PIE leave only a few **relics** hanging on in English.
These include some we've seen before—
nasal infixation seems to have been a **usual** way to mark the present tense,
and *e~o~*zero–grade **ablaut** was quite regular as well.
In English, these exist in only a **few** morphemes—
some borrowed from Latin and/or Greek (*tag-~tang-*; *gen-~gon-~gn-*),
a few native (*stood~stand*; *sing~sang~sung*).

Note that PIE (and other reconstructed ancient proto-languages) are **neither more simple nor more sophisticated** than modern languages.
When we look deep into the past, we see **neither** a more primitive stage of evolution **nor** some ideal linguistic situation that we have degenerated from.
All languages, **present and past**, are equally sophisticated in their **structure** and **capacity to communicate**; they're all manifestations of the **same basic human cognitive capability** to use language.