

Historical Change summary

1. Language acquisition is surprisingly perfect...

"Plato's Problem" (a term from Chomsky):

- What's in the stimulus?
 - a) acoustic continuum between phonemes
 - b) no chanting of the phonemic inventory or list of phonological rules
 - c) no morpheme boundaries
 - d) no word boundaries
 - e) no syntactic trees
 - f) no indices indicating coreference
- How do we know so much, when we had so little evidence for this knowledge?
- Answer: much of our linguistic knowledge is innate. (Universal Grammar)
 - knowledge of the phonetic boundaries
 - a drive to discover word boundaries, a drive to build a lexicon
 - a drive to place phrase boundaries, innate knowledge of possible syntaxes
 - innate knowledge of Principle C, compositional semantics etc.

2. ...but also *unsurprisingly* imperfect!

- Still -- it's not surprising that learning might not be perfect even if input is perfect. Imperfect learning should still deposit the child's knowledge of language within the boundaries of UG. Result: **variation**.
- Also: language contact produces situations where a child's input might be diverse -- native speakers, non-native speakers. Result: **variation**.
- In a random walk of variation, the variant adopted by the majority of speakers may be different at different times. Result: **linguistic change**.
- The course of linguistic change may be different in different speech communities that were united but have since been separated. Result: **dialect/language distinctions**

Two speech communities that have seen a modest amount amount of distinct paths of linguistic change are commonly said to speak **dialects of the same language**.

But: *Dialect, language, even speech community* are *imprecise* concepts. The important point is that *each of us speaks a dialect of a language*.

It's *not* the case that some of us speak a language, and others a dialect. That usage is just propaganda:

"A language is a dialect with an army and a navy." -Uriel Weinreich (famous linguist)

- An expectation: linguistic change is always **grammatical change** or **lexical change**. It should reflect the structure of grammar and the nature of the lexicon.
- Grammatical change should look orderly: new phonological rules entering the language, old rules reordered, systematic changes in choices from the "syntactic menu". Yes!
- Uses of this fact:
 - Understanding how a language whose history is recorded evolved over time.
 - Understanding current linguistic change.
- Figuring out whether two languages are **related**. But what does this mean?

Something like:

*Once upon a time there was a speech community sharing a fairly uniform grammar that we can call the **proto-language** that was the ancestor of modern languages X and Y.*

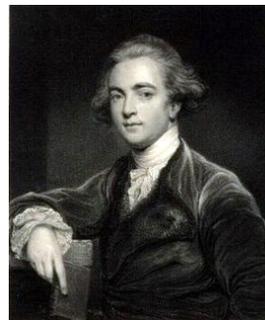
Historical change produced dialects of the protolanguage. These dialects themselves produced dialects, the process iterating until we reach a period where X and Y are two of the dialects that can trace their ancestry to the protolanguage.

3. Indo-European

[See textbook pp. 285-288 and 319-327.]

Case in point: a large group of languages spoken in Europe and Central Asia are related, stemming from a proto-language called **Indo-European**.

The discovery of Indo-European



Sir William Jones
(1746-1794)

This image is in the public domain.
Source: [Wikimedia Commons](#).

"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, though blended with a very different idiom, had the same origin with the Sanskrit; and the old Persian might be added to the same family, if this were the place for discussing any question concerning the antiquities of Persia."

SIR WILLIAM JONES
THE THIRD ANNIVERSARY DISCOURSE, ON THE HINDUS

Delivered 2 February, 1786, to the Royal Asiatic Society

- William Jones' observation was informal, and we can make a similar observation for ourselves, just by eyeballing relevant data. For example, the words for the first ten numbers strongly suggest that Sanskrit, Greek, Latin, Gothic, Old Irish, Lithuanian, and Old Church Slavonic share a common ancestor — but Basque, Hungarian and Turkish do not share that common ancestor. (My apologies for the small font size.)

(1) Numerals 1-10. Which languages are related?

	Sanskrit	Greek	Latin	Gothic	Old Irish	Lithuanian	Old Church Slavonic	Basque	Hungarian	Turkish
1	ékas	hei:s	u:nus	ains	oín	vienas	jedinŭ	bat	egy	bir
2	dvaú	dúo:	duo	twai	da	dŭ	dŭva	bi	kettő	iki
3	tráyas	trei:s	tre:s	þreis	tri	try:s	trĭje	hiru	három	üç
4	catvá:ras	téttares	quattuor	fidwor	cethir	keturi	četyre	lau	négy	dört
5	páñca	pénte	quinque	fimf	cóic	penki	peřĩ	bost	öt	beş
6	şát	héx	sex	saihs	sé	šeši	šestĩ	sei	hat	altı
7	saptá	heptá	septem	sibun	secht n-	septyni	sedmĩ	zazpi	hét	yedi
8	aşţáú	októ:	octo:	ahtau	ocht n-	aštuoni	osmĩ	zortzi	nyolc	sekiz
9	náva	ennéa	novem	niun	noĩ n-	devyni	devetĩ	bederatzi	kilenc	dokuz
10	daśa	déka	decem	taihun	deich n-	dešeimt	desetĩ	hamar	tíz	on

Cognate: a word in a language X that is hypothesized to share a common origin with a different word in language Y.

- So Sanskrit *dvaú* 'two' and Greek *dúo*: 'two' are cognates.
-
- But simple identification of words that look similar (what I was calling "eyeballing") is not enough to establish that two languages are related according to this scenario:

1822: Jacob Grimm formulates "Grimm's Law", which describes a regular correspondance between the stop consonants of Latin, Greek and Sanskrit and consonants in Germanic languages. Pay particular attention to p>f, t>θ and k>h (the most famous cases — final exam fodder):

(2) Grimm's law

	Latin	Greek	English		Latin	Greek	English	
p~f	<i>ped</i> -e	<i>pod</i> -e	<i>foot</i>	b~p	<i>labium</i>		<i>lip</i>	
	<i>pisc</i> -is		<i>fish</i>		d~t	<i>duo</i>		<i>two</i>
	<i>pater</i>	<i>pyr</i>	<i>fire</i>			<i>deca</i>	<i>deka</i>	<i>ten</i>
t~θ	<i>tres</i>	<i>treis</i>	<i>three</i>	g~k	<i>genu</i>		<i>knee</i>	
	<i>dent</i> -e	<i>dont</i> -e	<i>tooth</i>		<i>gno-sco</i>		<i>know</i>	
	<i>cord</i> -e	<i>kard</i> -ion	<i>heart</i>	p ^h ~b	<i>fer</i> -o	<i>pher</i> -o	<i>bear</i>	
<i>can</i> -e	<i>kuon</i>	<i>hound</i>	<i>frater</i>		<i>phrater</i>	<i>brother</i>		
<i>cannabis</i>	<i>kannabis</i>	<i>hemp</i>	t ^h ~d		<i>erythr</i> -os	<i>red</i>		
<i>cornu</i>		<i>horn</i>		<i>vidua</i>	(w) <i>eitheos</i>	<i>widow</i>		
<i>centum</i>	<i>hekaton</i>	<i>hundred</i>	g ^h ~g	<i>hostis</i>		<i>guest</i>		

- Reality check: Do regular sound correspondences result from historical change? Look for cases where we know the languages are related. We do indeed see the regular sound correspondences that we expect to find:

(3) We know these languages are related!

French	Italian	Spanish	Portuguese
cher	caro	caro	caro
champ	campo	campo	campo
chandelle	candela	candela	candeia

Note: French written "ch" is [ʃ].

Here's a partial list of Indo-European languages:

(4) **Some Indo-European languages (incomplete list)**

Albanian: Geg, Tosk dialects

Armenian

Balto-Slavic:

Baltic: Latvian, Lithuanian

Slavic: *East* (Russian, Ukrainian, Byelorussian), *West* (Czech, Slovak, Polish, Sorbian/Wendish, Old Prussian), *South* (Bulgarian, Macedonian, Serbo-Croatian, Slovenian)

Celtic:

Brythonic: Welsh, Breton, Cornish

Goidelic: Irish, Scottish Gaelic, Manx

Continental: Gaulish

Germanic:

East Germanic: Gothic

Scandinavian: Icelandic, Faroese, Norwegian, Danish, Swedish

West Germanic: German, Dutch, Frisian, English, Afrikaans, Yiddish

Hellenic: Greek

Anatolian: Hittite, Luwian, Lydian, Carian, Palaic

Indo-Iranian:

Iranian: Farsi [Persian], Pashto [Afghanistan], Kurdish

Indic: Hindi-Urdu, Bengali, Punjabi, Marathi, Gujarati, Romany

Italic: *Ancient:* Latin, Oscan, Umbrian.

Modern (= Romance) Spanish, Portuguese, Catalan, French, Occitan, Provençal, Romansch, Sardinian, Italian, Romanian, Haitian

Tocharian: Tocharian A, Tocharian B

- Most experts think: the parent language split well before 2000 BC.
- Possible homesite: "Kurgan culture" east of the Dniepr river, modern-day Ukraine. But there is hot debate about this. See, for example, a long series of blog postings in reply to an article in *Nature* here:

<http://geocurrents.info/cultural-geography/linguistic-geography/mismodeling-indo-european-origin-and-expansion-bouckaert-atkinson-wade-and-the-assault-on-historical-linguistics>

- Spread west around 4000-3500 BC; in Northern Iran 3500-3000BC; entered Greece, Italy, Western Europe 3500 BC. [speculation: do not memorize]

5. Good and Bad methods in historical linguistics

All languages change, and change is found in all branches of language:

- 1. lexicon: new words, borrowing
- 2. phonology: new rules, loss of old rules
- 3. morphology: new morphemes
- 3. syntax: changes in parameter settings, kinds of movement, etc.s

Relatedness is best established on the basis of the lexicon, not syntax:

- Because there are an unbounded number of possible words, finding the same sets of words in multiple languages is low-probability.
- There is a substantially smaller number of possible syntactic systems, so finding the same syntax in multiple languages is higher-probability: recall the V2 example.

But we should be careful. Processes like borrowing complicate the picture:

(5) **A sound change in the history of Italian**

Latin	Italian				
1. flamma	fjamma	'flame'	9.	kla:ma:re	kjamare 'shout' (L), 'name' (I)
2. flokkus	fjokko	'flock of wool'	10.	kla:rus	kjaro 'clear' (L), 'bright' (I)
3. flumen	fjume	'river'	11.	ekkle:sia	kjeza 'assembly'(L), 'church' (I)
4. ple:nus	pjeno	'full'	12.	klawstrum	kjostro 'enclosure'
5. plakere	pjatfere	'please'	13.	klawdere	kjudere 'close'
6. plumbus	pjombo	'lead' (= Pb)	14.	glakies	gjatfio 'ice'
7. plu:ma	pjuma	'feather'	15.	glu:to(ne)	gjottone 'glutton'
8. klavis	kjave	'key'	16.	plakare	plakare 'calm'
			17.	floridus	florido 'flowery'

4. Fun Facts about the Indo-Europeans that I crazily forgot to tell you

1. They had domesticated animals. *Evidence:* cognate words for dog **kwon*, horse **ekwo*, cow **gwo* and pig **suH*. [asterisk = hypothesized form in Proto-Indo-European]
2. They had grain. *Evidence* - cognate words **yewo*
3. They had the wheel: **kwekwlo* (cf. Greek *kuklos*, from which *cycle* is borrowed)
4. They had bee-keeping, since they have cognate words for honey: **melit*
5. Numbers up to 100, but not 1000.

- Example (5) shows a rule "l→j / C_V" entering, then leaving the language.
- Some words (*biancho*, *schivo*) were borrowed from other languages (Germanic and Slavic [via Greek], respectively) while this rule was active. Others (*florido*, *placare*) were borrowed once the rule had left the language.
- The rule left the language once its effects were no longer transparent to a language-learning child.

And there are many ways to go wrong when doing historical linguistics.

(6) Which two languages are related?

	A	B	C
'two'	er	erku	duo

- **Answer:** B (Armenian) and C (Greek).
- "Looking similar" is neither necessary nor sufficient to establish a genetic relationship among languages. In fact (and this is indeed weird), Armenian systematically shows [erk] where Proto-Indo-European is hypothesized to have shown [dw]:

(7) Armenian-Greek sound correspondences

	Armenian	Greek
'two'	erku	duo
'fear'	erki-	dwi-
'long'	erkar	dwa:ron

- In class, we saw a number of examples of "false cognates", just to warn you against accepting theories of historical relatedness on the basis of "eyeballing". For example:

Mbabaram *dog* 'dog' vs. English *dog* 'dog' — where we *know* the languages are unrelated, and Babaram *dog* is a cognate with words in nearby languages that have an additional first syllable lost in Babaram: Yidiñ *gudaga* and Dyirbal *guda*. (You can find more information about these languages on the web.)

And then you were further warned against "mass comparison" of lexical items that look alike and have vaguely similar meaning. That was the big slide (which I won't reproduce here) with words for 'milk', 'suck', 'chew', 'breast', 'neck', 'throat' etc. — all of which have an initial [m] (maybe onomatopoeic, as some of you remarked) followed by an [l] somewhere and a velar or uvular consonant.

6. History of English

Old English: 450-1100 AD

Languages spoken in present-day England were very much Germanic in appearance. SOV, verb-second, case-marked noun phrases [see overheads]

Middle English: 1100-1500

[1066 - Norman invasion of England, England under French domination]
Loss of verb-second, case-marking, V-to-I

French influence on lexicon: [blue = skipped this in class, alas]

- *pork, beef, veal, mutton, venison* (French) vs. *swine/sow, cow, calf, sheep, deer* (Germanic) [see textbook, p. 270, table 7.32]
- *-able* (from French)
- Loss of some words:
wer 'man' (cf. Latin *vir* as in English *virile*) - still found in *werewolf*
rice 'domain' (cf. *bishopric*, German *Reich* -- related to *reach*)
- Meaning change:
hund 'dog' > *hound*
docga 'mastiff' > *dog*
- Sound change:

1. Loss of [x] (*right, night*)

2. Very late in ME period: **Great Vowel Shift**
[book, pp. 258-259]

Image removed due to copyright restrictions. To see a graphic of the Great Vowel Shift, please go to: <http://users.ipfw.edu/thompsoc/LingL103/overheads/vowelshift.gif>.

3. Order of loss of [x] w.r.t. GVS: [x] was lost when GVS was still active, vowel lengthened, then GVS applied. Thus: /rixʰt/ > /ri:t/ > /rajt/.

- Syntactic change: V2 > SVO with V-movement to I (see overheads)
-

Modern English: 1500-present

- Loss of V-to-I.
- Definitive loss of case.

MIT OpenCourseWare
<http://ocw.mit.edu>

24.900 Introduction to Linguistics
Fall 2012

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.