LINGUISTICS 407
Historical Linguistics **LECTURE #3**

GENEALOGICAL CLASSIFICATION

THE FAMILY TREE THEORY

↓

It was formulated by August Schleicher (1871).

The theory assumes that languages change in regular, recognizable ways (the *Regularity Hypothesis*) and because of this, similarities among languages are due to a "genetic" relationship among those languages (the *Relatedness Hypothesis*).

In order to fill the particulars of such a relationship among those languages, it is necessary to *reconstruct* the hypothetical *parent* language from which the related languages are derived.

In keeping with the analogy of language relationships to human families, the theory makes use of the terms *parent* (or *mother*), *daughter*, and *sister* languages.

For example, in the family tree of IE, French and Spanish are *sisters*, both are *daughters* of Latin; Germanic is the *mother* of English.

The family tree model shows the *direction of change* and the *relations among languages*; the older stages of the languages being located higher in the tree and direct descendants being linked to their ancestors through straight lines (or branches).

Study the Indo-European family tree on p. 10.

Shortcomings of the family tree model resulting from considering a language as a biological organism:

i. faulty views with regards to language change

- each language forms a uniform speech community without internal variation and without contact with its neighbour languages (e.g., it was assumed that all speakers of Latin spoke exactly the same way at the time when French and Spanish split off);

- the split of a parent language into its daughter languages is an abrupt occurrence, happening without intermediate stages.

The above two assumptions are false: no language is uniform or isolated from others, and languages do not split apart abruptly but rather drift apart indiscernibly, starting as dialects and only ending up as separate languages after years of gradual change.

(In fact, the dividing point between two "dialects" and "two languages" is often impossible to locate and is often obscured by non-linguistic (social, political) factors.)

Languages, unlike animals or trees, do not have an independent existence: they are sets of conventions. Changes are introduced by their speakers and not by the language itself.

- distinct branches or sub-branches of the family tree should not undergo change caused by another branch (or sub-branch) that separated from the stem earlier (e.g., English should not be influenced by Latin -- yet, there are many words borrowed from Latin into English).

ii. misconceptions that result from names of successive stages of a language (e.g., OE, ME, NE -- these terms suggest that NE is a direct descendant of OE; however, modern standard English developed from the London dialect, while OE has come down to us in a West Saxon form).

Because of its simplicity the family tree has remained a useful model -- all genealogical relationships have been based on it and terminology associated with this model is widely used.

THE INDO-EUROPEAN LANGUAGE FAMILY

PIE was spoken between 5,000 and 6,000 years ago, probably in Southern Russia, possibly by a people of the so-called Kurgan culture (there are several other theories).

Study 2.65-2.72, pp. 45-48.

1. The Indo-Iranian subgroup

INDIC:

- *Rigveda* (the earliest Indic text; before 1000 BC);
- the language of the Rigveda and the Vedas has become obsolete -- commentaries and grammars were prepared by priests. The result of these

linguistic interpretations was the development of a standardized language that did not undergo change; it was completely regulated and described in detail: *samskrta* or *Sanskrit*.

- beside the *samskrta* (regulated, cultivated, correct) language there existed spoken languages called *prakrta* (popular, natural) or *Pankrits* languages.
- Three stages of Indic:

a.	Vedic Sanskrit or Vedic (1,200-800 BC)	Old Indic
b.	Classical Sanskrit (400 BC)	
c.	Pankrits	Middle Indic

- From the end of the Middle Indic period (before 1,000 AD) there are documents written in languages referred to as Apabhamsas (off-branchings):

MODERN INDIC LANGUAGES/DIALECTS (developed from the Apabhamsas)	Hindi
	Urdu (a version of Hindi spoken in Pakistan)
	Bengali
	Gujerati
	Marathi
	Punjabi
	Sinhalese (spoken in Ceylon)
	Romany (spoken by the Gypsies)

IRANIAN -- Study 11.1, p. 226.

Old Iranian > Avestan and Old Persian

- The book *Avesta* is a sacred book of the Zoroastrian religion; the poems of *Gathas* are from 1,000 BC
- Old Persian is documented from inscriptions (by Darius and Xerxes); among these the most important is the Behistan (Iran) inscription → trilingual text (Old Persian, Akkadian and Elamite)!

They illustrate the close relationship between Old Persian, Avestan and Sanskrit.

Middle Iranian: 300BC - 900 AD

↓

Pehlevi or Middle Persian (the language of the Persian Empire)
Sogdian
Saka or Scythian (texts discovered recently!)

Iranian languages at present:

Balochi (West Pakistan)
Pashto or Afghan (Afghanistan)
Farsi or Persian (Iran)
Kurdish (Western Iran, Iraq, Turkey and the former Soviet Union)
Ossetic (Northern Caucasus)
Yaghnobi

Study 10.1-10.4, pp. 202-203.

1. The Indo-Iranian subgroup **(see the complete list on p.10)**
2. The Armenian subgroup **(Study 16.1-16.8, pp. 382-383)**
3. The Albanian subgroup **(Study 19.1-19.5, pp. 446-448)**
4. The Balto-Slavic subgroup **(Study 18.1, p. 414)**
5. The Greek subgroup **(Study 12.1-12.10, pp. 248-252)**
6. The Italic subgroup **(Study 13.1-13.4, pp. 274-276)**
7. The Celtic subgroup **(Study 14.1-14.2, pp. 309-310)**
8. The Germanic subgroup **(Study 15.1-15.4)**
9. The Tocharian subgroup **(Study 17.1- 17.6, pp.400-402)**
10. The Anatolian subgroup **(Study 9.1 -9.4, pp. 170-172).**

The discovering old documents/texts in the 20th century contributed greatly to the reconstruction of PIE -- now we can reconstruct PIE as of a period before 3,000 BC.

Tablets in Crete (**Study 12.1, p. 248**)

- 1400 to 1200 BC
- a variety of Greek (Mycenaean Greek) – we thus now have early texts for Greek, similar in age to those in India.

Chinese-Turkestan Buddhist writings from the sixth to eighth centuries: information on Tocharian!

it preserved IE palatals as *k* before back vowels (the change of some IE palatals to sibilants took place *after* the Hittites left the IE community! Significance: we now know the earlier state of PIE!

Study 17.1, p. 400.

Boğazköy (Turkey) discovery: 10,000 cuneiform tablets – Hittite (1700-1200 BC), another IE language was identified!

Study 9.1-9.3, pp. 171-172.

The satem-centum subdivision

This classification is based on the sibilants (in three branches) vs. velars (in the remaining branches) in IE languages.

Gloss: *ten*

Skt. daśa Av. dasa Arm. tasn OCS deseti Lith. dẽʃimt

vs.

Gk. déka Lat. Decem OIr. deich Goth. taihun

Earlier assumption: PIE had split into two groups (eastern and western groups).

However, because there is no additional evidence for the split, it is not accepted today that there was a single predecessor for *satem* and *centum* languages -- there is only one such change, and even that does not show up everywhere it would be expected (e.g., in the Baltic languages not all palatalized velars have become fricatives).

Anatolian languages, Tocharian – the change from *k* > *sibilant* did not take place because they had left the IE speech community before the sound change took place.

Significance of the satem/centum subdivision: it assists in the establishing of relationships to *other* language families.

Example: the FU language family was still unified when this change took place in IE (around 4,000 BC):

Hu	=	Fi	>	PFU	=	PIran.		
száz		*sata*		*śata		* śata	> PIE **kmto-*	'hundred'
szarv		*sarvi*		* śarwe		* ś*arva*	> PIE **krwo-*	'horn'

We may adopt the view that by the fourth millenium BC certain independent languages had already emerged in IE -- using this as a starting point from which it is possible to postulate the dates of the FU period.

Other language families:

1. The Afro-Asiatic (=Hamito-Semitic) family :

   Five branches: (i) Egyptian, (ii) Berber, (iii) Cushitic, (iv) Chadic, (v) Semitic.

2. Sino-Tibetan family:

Sinitic subgroup: Mandarin, Wu, Xiang, Gan, Hakka, Min, Cantonese.

Tibeto-Burman subgroup – many languages are still undescribed.

The proto-language of these two families: Afro-Asiatic and Sino-Tibetan, has not been reconstructed. Each has languages not yet described.

3. Other African language families:

Only recently they were attempted to be classified genetically.

Chari-Nile (or Macro-Sudanic family), including Dinka and Shilluk;
Niger-Congo family, including Bantu and Swahili;
Khoisan family, including Bushman and Hottentot.

4. Uralic language family (see the Handout).

5. Altaic: Turkic, Mongolian and Manchu-Tungus; the genetic relationship of these language groups is uncertain, "Altaic" is a label of convenience.

6. Dravidian family: Telugu, Tamil, Kannada, Malayalam, Brahui.

7. Austric family: three subgroups.

(i) Austro-Asiatic (Munda and Mon-Khmer subgroups; Vietnamese)
(ii) Tai-Kadai
(iii) Austronesian -- about 500 languages! Languages with most speakers are Bahasa-Indonesia and Malay.

CONNECTIONS BETWEEN LANGUAGE FAMILIES

a. FU and IE (see also the example above!)

Hu	=	Fi	>	PFU	=	PIE	
méz		*mesi*		**mete*		**medhu*	honey
név		*nimi*		**nime*		**nomn-*	name
ház		*kota*		**kota*		**koto-*	house
-		*jyvä*		*jɛva		**jevo*	seed
vezet		*vetä-*		**wetä-*		**uedh-*	lead, pull
visz		*vie-*		**wike-*		**wegh-*	carry, take

b. FU and Altaic

Hu	=	Fi	>	PFU	=	Turkic	
al-		*ala-*		**ala*		*al*	bottom
em-		*ime-*		**ime*		*äm*	suck
nyal-		*nuole-*		**n´olo-*		*jala-*	lick
tø		*tyvi*		**tü °e*		*tü °-*	base

c. FU and Yukaghir (a Paleo-Siberian language, spoken by 300 people in North-East Siberia)

Hu	=	Fi	=	Yukaghir	
eme		*emä*		*emei*	mother
kéreg		*keri*		χar (skin)	bark of tree
lélek (soul)		*löyly* (steam)		*lul*	breath
mond		*manaa*		*mon-*	say
név		*nimi*		*neve, nim*	name

Similarities in grammatical elements:

FU = PIE: for example, the Accusative *-m* and the Ablative *-ta*

FU = Altaic: for example, the Genitive *-n* the Lative *-k*

FU and Yukaghir: several pronouns, suffixes

Problem: If it were only a question of vocabulary, we could regard certain words as loanwords (however, some basic vocabulary elements not likely to be borrowed!)

Grammatical elements are not easily taken over from one language to another -- it involves lengthy period of coexistence!

Can we consider these coincidences as evidence of early genealogical relationship between these language families?

> Linguistic relationship can be proved only by comparing entire language systems and several hundred etymologies.

Such comparative material does not exist in FU-Altaic, FU-IE or FU-Yukaghir connections.
(The number of FU-IE etymologies does not extend 37, of which 12 are doubtful).

> While we are still unable to prove the relationship between language families, one thing is certain: *these language families did not exist in isolation.*

THE NOSTRATIC HYPOTHESIS (MACRO-FAMILIES)

H. Pedersen (19th century): the founder of the Nostratic theory.

Pedersen proposed that Indo-European was related to several in a family he termed Nostratic; originally Nostratic included Indo-European, Semitic, Uralic, Yukaghir, Altaic and Eskimo-Aleut.

During the 1960's the theory further developed: Indo-European was related to Afro-Asiatic, Kartvelian, Uralic, Altaic and Dravidian -- over 700 etymologies, involving both grammar and lexicon exist now to support this hypothesis.

Two approaches by nostraticists:

- The proto-languages of the proposed families are compared and earlier proto-forms are hypothesized

- elements with "stable meaning" are compared, e.g.: 15-words list by Dolgopolsky: first-person marker, two, second-person marker, who/what, tongue, name, eye, heart, tooth, verbal negative, finger/toe, nail, louse, tear (noun), water, dead.

Nostraticists place great emphasis on only comparing families that have been reconstructed -- thus some families were not taken into account.

J. Greenberg, on the other hand, has approached the question of IE relatives from a totally different perspective: multilateral comparison!

↓

compares morphemes with similar sounds and meanings, but makes no attempt to generate sound correspondences.

Nostratic and multilateral approaches are very often confused with one another.

Nostratic: employs the comparative method;
importance of regular sound correspondences!

Multilateral comparison: does not try to prove that X is related to Y, but attempts to classify the words languages, without excluding those that do not have a reconstructed protolanguage yet.

Recently a family that is similar to Nostratic -- but with important differences -- has been proposed by Greenberg: Eurasiatic (see the Handout). Evidence for this family is provided by 64 grammatical and over 500 lexical etymologies.

Greenberg: proposed the classification of the indigenous languages of the Americas. He divides these languages into three families only:

Eskimo-Aleut
Na-Dene
Amerind -- this is a controversial concept; it is based on over 450 etymologies, and it includes most indigenous languages from North and South America. (See the Handout)

Sapir 1915: Soon after he proposed the Na-Dene family he became aware that this family shared many striking parallels with Sino-Tibetan languages. This idea has been further researched in the 1980s: Sapir's idea has been supported by additional evidence, but this question is not yet resolved.

Problems with long-range multilateral comparisons:

- unsystematic similarities between language families do not provide sufficient proof of genetic affinity, because unsystematic similarities can be found for any pair of languages.

- However, the comparative method does not allow us to research earlier periods (i.e. periods before PIE times); different methods are being employed.

Why the time-depths accessible to the comparative method are limited?

Answer:

Beyond a certain time-depth lexical replacement erases cognate sets to such a degree that there is insufficient material to work out sound correspondences or morphological paradigms.

Study 1.18-1.19, p. 13.

THE WAVE THEORY

Family trees represent languages as uniform entities, and show splits of languages as clear-cut and immediate.

Recognition that linguistic innovations take time to operate -- and may be diffused from one area, speaker or word to another -- resulted in an alternative theory → the *wave theory*.
The wave theory was first suggested by Johannes Schmidt (1972).

Linguistic innovations spread from one language or dialect through contact on the part of speakers of neighbouring languages and dialects; languages often share innovations that cannot be attributed to a common ancestor.

Two languages *A* and *B* may show identical sound changes. *A*, however, may show another change that is specific to language *A* and that must have preceded the development that is identical to the one in *B*.

A change common to more than one language (whether they are related or not) that has not been inherited from a common language, is termed *parallel development*.

For example, umlaut in the Gmc languages is not a PGmc innovation, but seems to have spread through the Gmc dialects after the period of differentiation. How do we know this? Gothic, the earliest attested Gmc language does not exhibit umlaut.

The process of umlaut is found in both North Germanic and West Germanic languages; the wave-like spread of this innovation is suggested by the fact that dialects spoken in the south of Germany do not exhibit as many umlauted forms as dialects in the north of Germany.

The family tree theory and the wave theory are not mutually exclusive hypotheses. Many changes that can be reconstructed in PGmc are reflected in *all* Gmc languages (e.g., the merger of PIE *o* and PIE *a* as PGmc *a*). Others, such as umlaut, have spread throughout the family.

The two hypotheses are complementary. Care must be taken to differentiate between changes that reflect genetic relationship and those that reflect diffusion when languages are being compared for purposes of historical inference.

Study the Handout!