

*Introduction to Language - Lecture Notes 3B*

**In Search of the First Language**

**(a Nova Video)**

Notes written by Prof. Russ Schuh, UCLA

**Note 1:** A brief introduction to the methods of linguistic reconstruction was given in class before the film. This discussion is not reproduced here. For an exciting introduction to the history of human populations, you may read *The Great Human Diasporas*, by L. L. Cavalli-Sforza and F. Cavalli-Sforza (Perseus Books) [obviously this is not in any way a required reading for this course].

**Note 2:** a complete transcript is available online at <http://www.pbs.org/wgbh/nova/transcripts/2120glang.html>

<b>Aharon Dolgopolskij:</b>	Nostraticist, Haifa Israel
<b>Luigi Cavali-Sforza:</b>	geneticist, Stanford university
<b>Joseph Greenberg:</b>	linguist and anthropologist, Stanford University
<b>William Labov:</b>	sociolinguist, University of Pennsylvania
<b>James Matisoff:</b>	Sino-Tibetan scholar and linguist, UC Berkeley
<b>Colin Renfrew:</b>	Indo-Europeanist, Cambridge University
<b>Donald Ringe, Jr.:</b>	historical linguist, University of Pennsylvania
<b>Merritt Ruhlen:</b>	freelance linguist, Stanford, California
<b>Vitaly Shevroskij:</b>	Nostraticist, University of Michigan
<b>Chris Stringer:</b>	paleoanthropologist, Museum of Natural History, UK

Some basic concepts

- (1) “Genetic” relationship between languages—the family tree model.
- (2) Well-established language families; controversial language groupings, especially in the Americas.
- (3) Comparative reconstruction: using *regular sound correspondences* to demonstrate language relatedness and to reconstruct a proto-language.
- (4) The method of mass comparison as a way to establish genetic relationships and critiques of this method.
- (5) Genetic evidence used in conjunction with linguistic evidence to support groupings of peoples.
- (6) The gradual but documentable nature of linguistic change over a single generation.
- (7) Limits on the time depth for establishing linguistic relationships.
- (8) Some controversial proposals for deep genetic groupings: Nostratic, proto-Human.

“There are more than 5000 languages spoken across the face of the earth. Could all these languages ever be traced back to a common starting point? Was there a time when the people of the world spoke one tongue?”

According to Biblical legend (Genesis 11), all people originally spoke the same language, but they displeased God by trying to build a tower to heaven. In order to put a stop to this, God made the people unable to communicate by making them speak many unintelligible languages.

Babel: • film: from Hebrew *balbel* ‘to confuse’

- *American Heritage Dictionary*: from Hebrew *ba\bbhel* ‘gate of God’, the name of the place (thought to be Babylon, i.e. Hebrew *babel*)

## Language relatedness and the comparative method

- Language family tree (Ruhlen): family is the “tree”, languages are the “leaves”
- Obvious and not so obvious similarities between languages:
  - Arabic & Hebrew (both Semitic)
  - English & Punjabi (both Indo-European)
- To show that similarities between languages can be traced to a common origin rather than to chance, one must show SYSTEMATIC PATTERNS to the similarities (Renfrew, Ringe):
 

• Indo-European numbers ;	English	<i>tooth</i>	Hindi	<i>dant</i>
	"	<i>ten</i>	"	<i>das</i>
	"	<i>two</i>	"	<i>do</i>
- *Comparative method*: sorting out systematic differences allows us to classify languages into families and sub-families and to reconstruct ancestral proto-languages (film goes through the major groups of Indo-European)
  - fascination of comparative linguistics is establishing a proto-unity from apparent great diversity (Matisoff)—film illustrates the origin of this diversity through change over time by using the Lord’s Prayer at different stages of English
  - *Indo-European*: about 100 languages spoken by over half the world’s people; best studied family, in part because it has the greatest number of written records providing direct evidence for changes that have taken place
  - *Sino-Tibetan*: about 250-300 languages spoken by 1/4 the world’s people
    - Matisoff project at UC Berkeley: “thesaurus” of Sino-Tibetan vocabulary, with words organized by semantic field; words culled from dozens of dictionaries and put into database; patterns of sound and meaning emerge, illustrated with EYE: *mik*, *smik*, *myak*, etc. showing up in word for ‘eye’ but also ‘eyelash’, ‘blind’, ‘eye matter’, ‘evil eye’, etc.
    - seeking relationships among languages, e.g. Thai resemblances to Sino-Tibetan seem to be borrowing from Chinese, not genetic inheritance; this can be illustrated with Thai *taa* ‘eye’ which resembles the Austronesian root *\*mata* rather than the Sino-Tibetan root *\*mik*; thus must sort out resemblance due to inheritance vs. those due to borrowing or chance; the deeper embedded in history, the harder to sort out

“Understanding why words are similar is essential to determining relationships among [language] families.”

## Language families

- (According to film) Most linguists believe about 200 *families* can be identified
- Some of the main families are Indo-European, Sino-Tibetan, Austronesian, Afroasiatic, Altaic, Dravidian, and Australian Aboriginal
- American Indian languages are especially complex and problematic

Excursus on Salish, illustrating the problem in studying American Indian languages:  
 Estimated that at the time of arrival of Lewis and Clark in 1805, there were about 23 Salishan languages with 100,000 speakers over 22 million acres. Jesuits forbade use of native languages when they brought Western education. Sarah Thomason is working with people to make a record of the language and to work toward reviving its active use.

- only about 1/3 of 1600 original American Indian languages are still spoken; classifications have proposed up to 200 families in the Americas

## The method of mass comparison for establishing genetic relations: African languages and American Indian languages

- Greenberg: the classification of African languages
  - Compared “the elements that are most stable over time”, viz. basic vocabulary items across many languages (pronouns, body parts, common natural features); languages down left side of notebook page, words in columns across pages
  - Greenberg came up with four families: Afroasiatic, Niger-Congo, Khoisan, and Nilo-Saharan [film claims that the first three were already recognized; this is not true—RGS]
  - (Ruhlen) Implications for African prehistory: Bantu languages of Eastern and Southern Africa are related to “Semi-Bantu” languages in Nigeria; historical inference from the language classification is that ancestors of Bantu-speaking people lived in what is now Nigeria and spread south and east
- Greenberg: the classification of American Indian languages
  - Using same methods as for Africa, came up with just three families: *Eskimo-Aleut* [across northern North America and Greenland], *Na-Dene* [Alaska, Western Canada, US Southwest—includes Navajo and Apache, for example]; *Amerind* [EVERYONE else in both North and South America and the Caribbean]
  - Critics of Greenberg and his methodology: (Matisoff) “unscientific” or “pre-scientific”; no methodological constraints; given the time depths involved, words resembling each other may be unrelated whereas related words could be changed so much that the relationship is unrecognizable.

Excursus on language isolates which resist classification into larger groups: Basque—why has it survived? [The point of this section is that we have examples of languages which defy classification despite the fact that they and neighboring languages are well-studied. DNA evidence supports the distinctiveness of the Basques vis à vis their neighbors. The film suggests that it may have been the ancestors of the Basques who were responsible for ancient cave paintings in Europe.]

- Genetic evidence for the Greenberg classification of American Indian languages: Geneticist Cavalli-Sforza claims that DNA evidence from the Americas supports the Greenberg linguistic classification.
  - Matisoff: Even if there were 3 migrations, how do we know the migrants themselves were a unified linguistic group? How do we know that there were not many languages which underwent pidginization? Matisoff notes that were “fossils” of modern North Americans studied by scientists in the future, great diversity would be found, but there is great linguistic uniformity. *Language can change within a generation.*

**Language change within a generation (Labov in Philadelphia)**

- longitudinal study of Philadelphia residents to trace change in their speech over time
- it seems to be the people at the “center” of society (as opposed to the upper or lower “fringes”) where change takes place; you need to know who influences whom in order to understand where changes originate
- changes in vowels in one generation: has been going on in English for centuries (the examples are given for your convenience only; you will not be tested on them, nor on the phonetic alphabet used here)
- Examples:
 

/e/	‘snake’	[snejk]	>	[snijk]
/æ/	‘bad’	[bæd]	>	[bʔ <sup>2</sup> d]
/u/	‘two’	[tu]	>	[tʔu]
/o/	‘go’	[gou]	>	[gæu]
/aw/	‘out’	[aut]	>	[ʔut]
/ay/	‘sight’	[sait]	>	[sʔit]
- enough accumulated small changes over time lead to great changes which make intelligibility across communities difficult and eventually to different languages (cites intelligibility study between speakers in Chicago, Birmingham, and Philadelphia)

“If English shows significant change within a single decade, the implication for linguists who are trying to study a language believed to have been spoken 15,000 years ago are enormous.”

**Some controversial language groupings: Nostratic**

- Nostratic hypothesis (Shevroschkin): established in Russia by Vladislav Illic'-Svitic' in 1960's
- Using 6000 year old reconstructed languages, reconstruct “precise sound correspondences” back 15,000 years, e.g. compare reconstructions for ‘water’ in several families: proto-Nostratic \*wete > \*wod/\*vesi/\*ved/etc. > water/voda/etc.
- Dolgopolskij: over 1000 reconstructed Nostratic roots back 15,000 years

Dolgopolskij’s hunting scenario in Nostratic: “From the telescope of the vocabulary we can discern a hunter who is following (\*desa) the tracks (\*goki, \*guja, \*miryo) of a beast ... is casting a spell (\*korwo, \*s'ubya), trying to hit (\*taba), and is afraid of missing (\*mera). Among the animals he hunts (\*hakra, \*harka) there are different kinds of antelopes (\*oryo, \*gula, \*gurha). He knows a lot about the anatomy of animals: ‘meat’ (\*siwya), ‘marrow’ (\*ayΩo), ‘spleen’ (\*lepa, \*payge). Some words are connected with spiritual culture, such as the meaning ‘to make magic, to use magical forces’ (\*arba).”

- Most linguists are skeptical that this can be done (Ringe): Accumulated changes beyond 10-12,000 years mean that virtually nothing will be left of original vocabulary—none of the well-studied language groups go

beyond this. Apparent resemblances could be a result of many factors other than genetic inheritance. With few if any original elements left, there is no way to tell the difference between inheritance and chance.

### **Some controversial language groupings: “The Mother Tongue” (proto-Human)**

- Though accepted linguistic reconstructions go back only about 10,000 years, human fossils go back 100,000 years and more. Could fossil records shed any light on when language originally emerged? (Stringer) Modern language must have been in existence *at least* 40,000 years ago; language would have been necessary for the known features of culture (symbolic systems, art, etc.) found at that time.
- Some linguists believe elements of the “Mother Tongue” can be unearthed 40,000 to 100,000 years ago: (Ruhlen) *tik* ‘one, finger’, *pal* ‘two’, *milik* ‘milk, suckle, breast, throat’ are some of the elements that come up in reconstructed language family after family.
- However, few linguists take these results seriously: (Dolgopolskij) “Serious reconstruction” would be needed, not vague resemblances; (Ringe) languages probably all had a common source, but this is merely a *belief* in the absence of objective proof; (Matisoff) there is a big difference between a reconstruction to 10,000 years and one to 20 or 40 or 200,000 years—such reconstructions are just a dream.

### Some issues for study and discussion:

- (1) The *comparative method* (using inter-language comparisons to establish systematic correspondences) and the method of *mass comparison* are generally treated as two differing ways of working out genetic relationships between languages. What types of data are used for these methods? What are some of the methodological techniques of these two approaches? Do these methods have the same goals? Are they in conflict? What criticism(s) can be leveled at either or both?
- (2) Why do some linguists believe there is a temporal limit beyond which it is impossible to provide any convincing evidence of genetic relationship? About how far back is that temporal limit? How is this putative limitation reflected in present-day classification of the world’s languages?
- (3) What sorts of classifications have been proposed for American Indian languages? What controversies have arisen in classification of these languages?
- (4) What is Nostratic? Who are some of the main proponents of the Nostratic hypothesis? What types of reconstructions have been claimed for Nostratic?
- (5) What range of views can be found among linguists on the reconstructibility or even the existence of “proto-Human”?