LINGUISTICS 407
Historical Linguistics LECTURE #7

GLOTTOCHRONOLOGY

Lexicostatistics: The study of vocabulary statistically for historical inference.

M. Swadesh (1949) proposed a method for determining the time when two related languages became independent: **GLOTTOCHRONOLOGY**.

Glottochronology (one of several possible lexicostatistic methods):
Method of lexicostatistics for determining degrees of relationship between languages, based on counting the number of cognates in a particular set of vocabulary items.

Assumption: the common words in languages are maintained at a definite rate, i.e., some parts of the vocabulary are much less subject to change than other parts.

The basic (core) vocabulary consists of words for concepts assumed to be a necessary part of all human cultures. The semantic field represented by the lexical items includes pronouns, numerals, adjectives (e.g., *big, long, small*), kinship terms (*mother, father*), living beings (*dog, louse)*, body parts (*head, ear, eye*), events and objects in nature (*rain, stone, star*), and common activities (*see, hear, come, give*).

There exist several versions of Swadesh's list: the two of them that are most commonly used are the 100-word list and the 200-word list.

Swadesh made several further assumptions:

i. the rate of retention of vocabulary items in the basic core is constant through time;

ii. the rate of loss of basic vocabulary is approximately the same in all languages; the rate of loss was arrived at by testing lexical loss in languages with a long series of texts (e.g., Latin and the Romance languages). On the basis of theses tests, there is a 80-85% loss over a thousand ys.).

iii. Modern English preserved 80% of the basic vocabulary in AD 991, and 64% at the beginning of our era.
If we date late PIE at 3000 BC, modern English has maintained 30% of the basic core vocabulary.

English and German:
German: *Tier, vier, Kopf, ich, Sonne*
English: *Animal, four, head, I, sun*

60% agreement!

Expected rate of loss:
if both maintained 80%bof the basic vocabulary (though not necessarily the *same* 80%), after one thousand years the common basic core vocabulary would be 64% -- if 80% is the rate of maintenance.

iv. if the percentage of cognates within the core vocabulary is known for any pair of languages, the length of time that has elapsed, since the two languages began to diverge from a single language, can be computed by a formula (devised by Robert Lees):

$$t = \frac{\log c}{2 \log r}$$

t: the time depth is equal to the log of the % of cognates (**c**), divided by twice the log of the assumed % of cognates retained after a millennium of separation (**r**).

Applying the formula to the examples from English and German:

$$t = \frac{\log 60\%}{2 \log 85\%} = \frac{-0.511}{2x\ -\ 0.163} = 1{,}561$$

Thus, according to the formula, English and German separated appr. 1.561 x 1,000 years ago – about AD 430.

Important: the date of invasion of England by the Angles and Saxon is AD 440, thus the result above corresponds to the historical information.

At least three problems have to be considered before glottochronology can be accepted as a valid technique:

i. Is the list representative?

ii. Are the words contained in the list really the common core shared by all languages?

iii. Is it a fact that the items in the list are not culture-bound?

Criticisms of glottochronology

i. the existence of a basic vocabulary is culture-bound, e.g., in the cultures of South Asia some items referring to natural objects (such as natural objects, e.g. *sun*) are not part of the core vocabulary: they belong to borrowed religious vocabulary.

ii. duplication of elements in the core vocabulary in some languages, e.g., in Navaho, the words *this, that* designates five items, similarly, the words *tree, seed, grease, eat, kill* correspond to several items, none of them can be matched with the English vocabulary list.

iii. the rate of retention is not constant from language to language, e.g. the rate of loss in English is slower than the rate of loss in English, Lithuanian and Greek is much lower than in the Romany dialect of Armenian. In Icelandic, the rate of retention is very high (due to the long period of isolation and maintaining old culturalpractices).

iv. recognition of true cognates is often difficult:

as far as the phonetic shape of words is concerned, there are two possibilities:

a. either true cognates may not be recognizable because of sound changes (e.g., English *tooth* = German *Zahn*)

b. chance similarities may be interpreted as evidence for the presence of the true cognate pair (e.g. Latin *habere* and German *haben* (Gmc **khaben*) both mean 'to have' but are not cognates)

Result: overestimation of cognates results in underestimation of time depth and vice versa.

Lexicostatistics applied to determine the oldest relationships among language families

H. Pedersen proposed a proto-language from which Indo-European, Hamito-Semitic, Uralic, Altaic and possibly other language families have descended: **NOSTRATIC** (see Lecture #3)

iv. A. Dolgopolsky proposed a "mathematically rigorous procedure" to examine the Nostratic hypothesis. He made a list of 15 items assumed to be essential in every language: elements with "stable meaning" are compared (see Lecture #3).

first-person marker, two, second-person marker, who/what, tongue, name, eye, heart, tooth, verbal negative, finger/toe, nail, louse, tear (noun), water, dead.

Difficulties with the list:

- taboo words -- the word *tongu*e has several forms in IE languages;

- 1st and 2nd person markers may not be stable, as there are languages where they don't even occur (e.g., Japanese, Chinese);

- the verbal negative may be replaced, e.g., French:

> French *ne* (< Latin *non*) was originally the only negative, and was placed before the verb. Later it became reinforced by the word *pas* 'step' after the verb, which was added for emphasis: "not a step" (compare English *not at all*). Over the centuries, the emphatic negative became the norm: *Je ne sais pas*.
> Eventually, the reinforcement *pas* came to be thought of as the main negative, and this has led to the omission of *ne* in casual speech: *Je sais pas*.

Conclusion: when considering the rate of retention after six millennia, *the number of maintained cognates is comparable to that of items similar by chance.*

The reliability of glottochronology as a method for establishing time depths remains to be confirmed.