The Basis of Glottochronology
Author(s): Robert B. Lees
Source: *Language*, Vol. 29, No. 2 (Apr. - Jun., 1953), pp. 113-127
Published by: Linguistic Society of America
Stable URL: https://www.jstor.org/stable/410164
Accessed: 20-12-2018 22:48 UTC

THE BASIS OF GLOTTOCHRONOLOGY

ROBERT B. LEES
University of Chicago

[It is shown that a linguistic dating system can be set up on the basis of several explicit assumptions about morpheme decay. Thirteen sets of data, presented in partial justification of these assumptions, serve as a basis for calculating a universal constant to express the average rate of retention $\bar{k}$ of the basic-root-morphemes: $\bar{k} = 0.8048 \pm 0.0176$ per millennium, with a confidence limit of 90%. Finally an expression is derived for the sampling-error to be expected in the calculated time-depths of related dialects.]

0. INTRODUCTORY

0.1. Glottochronology. Many linguists have shown an increasing interest recently in the application of mathematical and, in particular, statistical methods to linguistic problems.[1] One such application is the use of statistical techniques to measure the degree of relatedness among cognate dialects.[2] It was in this connection that Swadesh first suggested in his article on Salish internal relationships the particular statistical method which has since been elaborated under the name GLOTTOCHRONOLOGY. The validity of the glottochronologic technique rests at present on the data and the mathematical derivation to be set forth in this paper.

0.2. Method. If (1) the morpheme inventory of a language, or a definable portion of it, is observed over a span of time, and if (2) the individual members of the inventory at a given time are identified as cognates of members at some previous time, and if (3) some statable regularity can be found in the time-rate at which members disappear from the inventory to be replaced by new items, then the number of items in a certain subset which are present at any one time can be used as a measure of time elapsed since some previous time-point for which a similar count is available. The members of the chosen subset may be likened

[1] See B. Trnka, *A tentative bibliography* (CIPL: Publication of the Committee on Linguistic Statistics, No. 1; 1950); H. E. Driver, Statistics in anthropology, *Am. anthr.* 55.4 ff. (1953), with bibliography.

[2] A. L. Kroeber and C. D. Chrétien, Quantitative classification of Indo-European languages, *Lg.* 13.83–103 (1937); id., The statistical technique and Hittite, *Lg.* 15.69–71 (1939); C. D. Chrétien, The quantitative method for determining linguistic relationships, *Univ. Cal. Pub. Ling.* 1.11 ff. (1948); M. Swadesh, Salish Internal relationships, *IJAL* 16.157 ff. (1950); id., Lexico-statistic dating of prehistoric ethnic contacts, *Proc. Am. Phil. Soc.* 96.457 ff. (1952); id., Mosan I: A problem in remote common origin, *IJAL* 19.26 ff. (1953); D. W. Reed and J. L. Spicer, Correlation methods of comparing idiolects in a transition area, *Lg.* 28.348 ff. (1952); R. B. Lees, Relationships among the Germanic dialects by a statistical method (unpublished, 1950); id., The genetic relation in linguistics (unpublished, 1952); C. F. Hockett, Linguistic time perspective and its anthropological uses (paper read before the summer meeting of the Linguistic Society, Bloomington, 1952); D. Taylor, Sameness and difference in two Island Carib dialects, *IJAL* 18.223 ff. (1952); E. Cross, Chronometric and telemetric determination of the relationship of Latin, Gothic, Old Church Slavic, and their present affiliates (paper read before the meeting of the Linguistic Society, Cambridge, 1952). This list is not exhaustive.

to the (indistinguishable) atoms in a given mass of a radioactive element; since the rate of disintegration is predictable at any time during observation of the sample, the mass (or number of remaining atoms) of this element remaining among the decay products at any time in the sample is a measure of how long the sample has been decaying. The analysis of decay products in mineral samples permits the calculation of the age of the earth's crust. Similarly, analyses of morpheme decay products should provide an absolute chronology for lexical history.

0.21. Morpheme decay. The total morpheme inventory of a language may contain over 100,000 identifiable items; of these, some smaller number, say 20,000 morphemes, comprises (according to most accepted estimates) the active colloquial vocabulary of an average speaker. Of the active colloquial morphemes, a few hundred may be affixes or other patterned units in a tight grammatical structure; these we shall disregard. The rest are active colloquial root-morphemes, each correlated with a small set of extralinguistic items in the culture, i.e. each with its own set of meanings.

These meanings may be of the most varied sort, differing from culture to culture; but certain meanings are likely to be found in all cultures, such as body parts, numerals, geographical terms, and simple activities.[3] We call the morphemes associated with these cultural universals the BASIC-ROOT-MORPHEME INVENTORY, and denote this set of several thousand items by *I*. Because of the universality of the meanings associated with the members of *I*, we may assume that the size of *I* remains nearly constant in time and from language to language.

If the morphemes correlated with a certain subset of cultural universals in some language at a given time is compared with the corresponding morphemes correlated with the same meanings in the derivative cognate language at some later time, many corresponding morphemes will be found to be cognate; but a certain number may not be cognate.[4] In the latter case, certain morphemes of the original set have disappeared and have been replaced by new, non-cognate morphemes. This temporal decrease in the size of the original subset is called MORPHEME DECAY.

The reasons for morpheme decay, i.e. for change in vocabulary, have been classified by many authors;[5] they include such processes as word tabu, phonemic confusion of etymologically distinct items close in meaning, change in material culture with loss of obsolete terms, rise of witty terms or slang, adoption of prestige forms from a superstratum language, and various gradual semantic shifts, such as specialization, generalization, and pejoration.

[3] Probably no meaning is completely universal: it would not surprise us to find languages containing no morpheme correlated with *man*, *head*, *three*, or *sky*, or languages in which *water* is represented by an affix (already excluded from consideration). We hope that such cases are so rare and so randomly distributed that our sampling process will show no significant effects from the non-universality of any one meaning.

[4] By cognate we mean, of course, derivable one from the other by the use of a systematic set of phoneme correspondences, furnished by the traditional comparative method as applied to the language family in question.

[5] See L. Bloomfield, *Language* 392–495 (New York, 1933); E. H. Sturtevant, *Introduction to linguistic science* 123–53 (New Haven, 1947).

0.22. Word-lists. In order to keep the material down to manageable size, and to render the material in one language comparable to that in another, the basic-root-morpheme inventory is sampled in the following way. A small set of basic morphemes (say 200) is selected from the inventory of some control language (say English), and each item in it is translated into the common colloquial expression of the test languages. These translations will then comprise, for the most part, root-morphemes which can be compared by the usual etymological techniques. Corresponding terms in two test languages will be either cognate or non-cognate, the latter label including terms borrowed by one language from the other. It is assumed that all the various causes of morpheme decay add up in both languages to some total amount of change which is dependent only upon the length of time during which these causes have been active.

Sometimes it will be difficult to determine whether two corresponding forms should be considered cognate root-morphemes. For example, the later form may be a fusion of several etymologically different roots; or it may represent either the retension of an old root or, just as plausibly, the borrowing of a similar root from some neighboring cognate dialect. The glottochronologic method depends upon the determination of retained original roots, and all ambiguous cases such as these are therefore ignored. It has been found that the number of items that must be omitted for this reason from a list of 215 basic-root-morphemes seldom runs higher than ten or fifteen.

1. Rate Equations

1.1. Morpheme decay law. We have described the basic-root-morpheme inventory I of a language as some large set of morphemes expected to remain approximately constant in size for all times and all languages. The identity of its members is continually changing as morphemes are lost and replaced by new, non-cognate morphemes. Let V be the number of items in I, and let R be the time-rate at which these items are exchanged.

As a BASIC WORKING HYPOTHESIS we shall assume that THIS RATE R IS ALSO CONSTANT IN TIME. This assumption will have to be verified by the data to be collected for calculating the value of the rate-constant (§3.2).

Consider some arbitrarily chosen sample subset S of the V items in I for a given language, containing N_0 items at the time t_0, of which only N items are left at some later time t. Now while in one unit of time, R items have been replaced in I, only $(N/V)R$ items are lost from S; for we may expect the random sample to lose the same FRACTION of items as the total population does in one unit of time. This is the rate at which S is decreasing in size, or:

$$-\frac{dN}{dt} = \frac{N}{V} R = \lambda N, \tag{1}$$

where λ is the rate-constant, and $\lambda = R/V$. Solving for N as a function of time:

$$N = N_0 \, e^{-\lambda t}, \tag{2}$$

where e is the well-known mathematical constant $= 2.718$. This is the BASIC RATE EQUATION for morpheme decay, and is identical in form to that for simple radioactive decay. If we can evaluate the rate-constant λ, Eq. 2 tells us how many morphemes in a sample S are left, out of some original number N_0, after time t has elapsed.[6]

Conversely, solving Eq. 2 for t gives us:

$$t = -\frac{1}{\lambda} \ln \frac{N}{N_0}. \tag{3}$$

This permits us to calculate how long it has taken a language with a morpheme decay rate-constant λ to decrease the size of a sample S of basic-root-morphemes from N_0 to N. We may call this value of t the TIME-DEPTH of the dialect stage which still had all the N_0 items, if we count the number N which are left from the present time (say 1953).

1.2. Time-depth equation. Suppose now that there are two genetically related dialects at time t, both assumed to have descended from some single common ancestor. Glottochronologically interpreted, this means that if we examine the identity of the items in a sample S of basic-root-morphemes in the two dialects back through time, there is a time t_0 when the two become identically the original N_0 items of the proto-language.

Throughout the time interval $t - t_0$ the two dialects have been diverging. By divergence we mean here that the particular items selected for replacement in S by new, non-cognate forms have not been the same for the two dialects. Some items will still be shared by the two, some will have been lost by one dialect, some by the other, and some by both.

1.21. Independence. By introducing two more fundamental hypotheses we shall be able to use these data (on the number of shared items left) to measure the time during which the two dialects have been diverging. First, as a SIMPLIFYING ASSUMPTION (which unfortunately it will prove difficult to verify), we shall suppose that THE SELECTION OF ITEMS FOR REPLACEMENT IN ONE DIALECT IS STATISTICALLY INDEPENDENT OF THE SELECTION IN THE OTHER DIALECT; that is, that the two selections are random with respect to one another. This may be stated mathematically in the following way: dialect A preserves N_a items out of the original N_0 after time t; it therefore preserves the same fraction N_a/N_0 out of any random subset of S—in particular, the subset of N_b items retained by dialect B; i.e. the number of items retained by A among those also retained by B is $N_s = (N_a/N_0)N_b$. Finally, the FRACTION of the N_0 original items still SHARED by dialects A and B after time t is:

$$F_0 = \frac{N_a}{N_0} \cdot N_b \cdot \frac{1}{N_0} = \frac{N_a N_b}{N_0^2}. \tag{4}$$

1.22. Interlingually constant rate. The second hypothesis which we shall introduce here is one that we shall later attempt to verify empirically: namely, that the rate-constant for morpheme decay has the same value for both dialects A and B: i.e. $\lambda_a = \lambda_b$.

1.23. Dating equation. Now since A and B have been diverging for the same

[6] We shall even try to show later that the rate-constant may be considered the same for all languages.

length of time, and since we have assumed that $\lambda_a = \lambda_b$, we see from Eq. 2 that $N_a = N_b = N$, and we may rewrite Eq. 4 to read:

$$F_s = \left(\frac{N}{N_0}\right)^2. \tag{5}$$

Substituting in Eq. 2, we obtain:

$$F_s = e^{-2\lambda t}. \tag{6}$$

Eq. 6 enables us to calculate the fraction of items in S shared by the two dialects after time t, provided we know the value of the rate-constant λ.

As before, we may solve Eq. 6 for t to obtain the time-depth equation for two cognate dialects:

$$t = -\frac{1}{2\lambda} \ln F_s. \tag{7}$$

It will prove convenient to express the rate-constant in Eq. 7 somewhat differently. We shall determine empirically the average fraction of retained items per 1000 years. Measuring time in millenia, we may write this from Eq. 2:

$$\frac{N_1}{N_0} = e^{-\lambda} = k. \tag{8}$$

Then substituting our new constant k for λ in Eq. 7:

$$t = \frac{\ln F_s}{2 \ln k} = \frac{\log F_s}{2 \log k}.^{7} \tag{9}$$

2. Empirical Data

In order to evaluate the rate-constant λ, or its equivalent, k, counts were made on a sample S of basic-root-morphemes for a number of languages to determine the fraction of the original N_0 items retained after the language had evolved through a time t. These data were then expressed as fraction-retained-per-millenium, or k, and an arithmetic mean was calculated. The data were also subjected to several statistical treatments to attempt a verification of our hypotheses.

2.1. Word lists. The basic-root-morpheme sample for each language tested was obtained by translating each of 215 English words into the most common colloquial term of that language. The first studies of this kind were made with the list used by Swadesh[8] to measure the rate-constant of English. Since there

[7] We use ln for logarithms to the base e, log for base 10. Eq. 9 is identical with Swadesh's equation

$$i = \frac{\log C}{2 \log r}$$

as given in *IJAL* 16.161 (1950).

[8] Ibid. 161 (§2.2, Par. 1 and 3). The English words of Swadesh's list were originally chosen to be representative of universal semantic areas, to be relatively stable and resistant to culture changes, and to be easily found in the lexicons of many languages. Swadesh has since suggested a slightly modified word list, *Proc. Am. Phil. Soc.* 96.456–7 (1952).

is no a-priori reason to prefer one set of basic English morphemes to another, and since altering the word list might possibly introduce an unknown variable, all the data published here were obtained with the same list. The effects of changing the word list will have to be studied in the future.[9]

For each language tested, one list was prepared for an older stage and another for a more recent stage; in each case there was an independent way of dating the vocabulary of both stages. Corresponding morphemes were then compared by specialists in the language family involved, and word pairs were marked as cognate, non-cognate, or indeterminate. Occasionally it was impossible to translate an item, especially for the older stage; such items were simply omitted. Uncertainties and omissions never reduced the total of any list to fewer than 200 items.

2.2. Data. At present, 13 word counts have been prepared for this study; the following list shows the languages compared, the number of words and cognates (with percent of cognates in parentheses), and the persons responsible for the test.

1. Old English of 900–1000 A.D. : Modern English; 209 words, 160 cognates (76.6); R. B. Lees and J. H. Sledd. — 2. Plautine Latin of 200 B.C. : early Modern Spanish of 1600 A.D.; 200 words, 131 cognates (65.5); D. Griffin. — 3. Plautine Latin : Molière's French of 1650 A.D.; 200 words, 125 cognates (62.5); D. Griffin. — 4. Old High German of 800–900 A.D. : Modern German; 214 words, 180 cognates (84.2); R. B. Lees and G. J. Metcalf. — 5. Middle Egyptian of 2100–1700 B.C. : Coptic of 300 A.D.; 200 words, 106 cognates (53.0); K. Baer. — 6. Koine Greek of 250 B.C. : Modern Athenian Greek; 213 words, 147 cognates (69.0); E. P. Hamp and B. Einarson. — 7. Koine Greek : Modern Cypriote; 211 words, 143 cognates (67.8); E. P. Hamp and B. Einarson. — 8. Ancient Classical Chinese of 950 A.D. : Modern Mandarin; 210 words, 167 cognates (79.6); C. Y. Fang, M. Swadesh. — 9. Old Norse of 800–1050 A.D. : Modern Swedish; 207 words, 176 cognates (85.0); R. B. Lees and G. Franzen. — 10. Classical Latin of 200 B.C. : Modern Tuscan; 210 words, 144 cognates (68.6); J. Corominas and H. Noce. — 11. Classical Latin : Modern Portuguese; 210 words, 132 cognates (62.9); J. Corominas. — 12. Classical Latin : Modern Rumanian; 209 words, 117 cognates (56.0); J. Corominas. — 13. Classical Latin : Modern Catalan; 208 words, 126 cognates (60.6); J. Corominas.

2.3. Rate-constants. These data can now be tabulated, together with the calculated rate-constants and time-depths (counted from the mid-points of the ranges in dates). The rate-constant k is fraction-retained-per-millenium, and t is given in millenia.

LANGUAGE	F_s	t	k
1. English	.766	1.0	.766
2. Spanish	.655	1.8	.790
3. French	.625	1.85	.776
4. German	.842	1.1	.854

[9] To save space we will not list the 215 words used, but a copy of the work-sheet may be obtained from the author upon request. All the original data are available for inspection.

Language	F_s	t	k
5. Coptic	.530	2.20	.760
6. Athenian	.690	2.07	.836
7. Cypriote	.678	2.07	.829
8. Chinese	.796	1.0	.795
9. Swedish	.850	1.02	.854
10. Italian	.686	2.15	.839
11. Portuguese	.629	2.15	·.806
12. Rumanian	.560	2.15	.764
13. Catalan	.606	2.15	.793
		Mean $\bar{k}$ =	.8048 ± .0176/mill.

The limits of error on the mean rate-constant $\bar{k}$ were calculated as the 9/10-error in the mean, using small-sample methods.[10]

We take this to mean that on the average about 81% of the basic-root-morphemes of a language will survive as cognates after 1000 years, for all languages, at all times.

3. Verification of Hypotheses

We must now see if these empirical data can be used not only to provide a value for an average rate-constant, but also to verify the hypotheses proposed for justifying the calculation of a mean rate-constant for all languages.

3.1. Interlingually constant *k*. The theory mentioned in §1.22 assumes that the rate-constant λ for morpheme decay is the same in two dialects. The obtained values of k ($= e^{-\lambda}$, Eq. 8) range from 0.760 for Coptic : Egyptian up to 0.854 for both German : Old High German and Swedish : Old Norse, with a standard-deviation of 0.0342.[11] To what extent are we justified in assuming that all of this variation is random sampling error, and that, if it were not for these supposedly random perturbations, each of the thirteen languages would show the same rate-constant?

3.11. The χ^2-function. To answer this question it is not sufficient to look at the differences between the expected and the observed result ($E - O$), for there is

[10] That is,

$$0.0176 = \frac{z\sigma}{\sqrt{n-1}},$$

where σ is the standard-deviation of the k's by the small-sample formula

$$\sigma = \frac{\Sigma\delta^2}{\sqrt{n-1}},$$

$n = 13$ is the number of k's averaged, and z is the unit-deviate for 12 degrees of freedom on the 't'-distribution for the 90% level of confidence. We have used the 9/10-error as a measure of uncertainty throughout this study.

[11] The standard-deviation is the root-mean-square deviation about the mean:

$$\sigma = \sqrt{\frac{\Sigma_i(k_i - \bar{k})^2}{n-1}}.$$

no way to evaluate the significance of a larger or smaller difference. To be sure, there will always be a discrepancy; we seek merely to know what maximum discrepancy may be tolerated in any calculations that use our data. To evaluate the discrepancy between expected and observed results we may employ a certain function of the differences $(E - O)$, a function whose mathematical and statistical properties have been studied and tabulated.[12]

This function is known as CHI-SQUARE, and it is given by:

$$\chi^2 = \sum_i \frac{(E_i - O_i)^2}{E_i} \tag{10}$$

where the O_i are a set of observations on the outcome of some experiment, and the E_i are the expected values as predicted by some theory. For every number of independent observations, and every value of χ^2 for them, there is a certain known probability that a value of χ^2 as great or greater could have been obtained by chance alone. If our calculated value of χ^2 is small enough that this probability P is high (say 50%), then we have reason to believe that the discrepancy between E_i and O_i was due only to random factors, to be expected in any physical data. But if our value of χ^2 is so large that P is very low (say 1% or less), then the discrepancy probably did not result from chance alone in our sample experiments, but represents a 'real' difference between observation and theory.

Now our theory predicts the outcome of thirteen probability experiments, wherein N_0 morphemes are allowed in each case to fall by chance into two categories, RETAINED and LOST. It tells us that for a given time t, the number retained is

$$N = N_0\, e^{-\lambda t}$$

and the number lost is

$$N_0 - N = N_0(1 - e^{-\lambda t}).$$

The actually observed outcomes for the 'retained' category are found in the table in §2.2, and the number in the 'lost' category is obtained in each case by subtraction. There are then 26 comparisons of observation and theory from which a calculation of χ^2 may be made. The value of χ^2 was computed for our thirteen sets of data:

$$\chi^2 = 29.5.$$

Finally, from statistical tables of the χ^2 function we see that for 12 degrees of freedom[13] there is a probability of 0.01 for obtaining a χ^2 at least as great as

[12] The author gratefully acknowledges the invaluable help that he received, in preparing this formulation, from W. Kruskal and his associates in the Department of Statistics, University of Chicago.

[13] The number of INDEPENDENT observations used to calculate χ^2, however, is only 12, because the 13 in the 'lost' category may be determined from N_0 and the value in the 'retained' category in each case, and one last value may be determined from the remaining 13 and their mean (or its equivalent, the mean k), leaving altogether only 12 DEGREES OF FREEDOM, not 26.

26.1. Our value is even larger. We must conclude then, that if the χ^2-test is valid for these data, there is very probably a greater-than-chance discrepancy between our theory and our data.

Before accepting this conclusion, however, we must note several important qualifications. First of all, the value of χ^2 depends upon the size of E_i and O_i; in our case these numbers are rather large, in the neighborhood of 100 (40 to 170). This fact may account to some extent for the high value of χ^2.

Second, some of the 'real' discrepancy between E_i and O_i may be contained in the values of the parameter t; surely we cannot claim to have dated all the word lists very accurately. Slight errors in dating our test lists will show up as errors in the predicted values E_i, which will not be due to a failure of our theory.

Third, we must remember that the χ^2-test is by no means the only judge of a good fit; even with a high χ^2, a theory may, within quite respectable limits, be a valuable tool for certain calculations.

Finally, there is some possibility that the value of $\bar{k}$, the mean rate-constant used to calculate the E_i, was faulty. We obtained it by averaging the values of k; perhaps a better value (say, a maximum-likelihood value) would have been some other function of the individual k's, say $e^{-\lambda}$. (The equations for a minimum-χ^2 and for a maximum-likelihood calculation were both too difficult to solve analytically.)

3.12. The 9/10-error. Another indication of the reliability of our assumption of interlingually constant k is the sampling-error in $\bar{k}$ itself. We have calculated the standard-error of the k's by small-sample methods:

$$\sigma_{\bar{k}} = \frac{1}{n-1}\sqrt{\Sigma\delta^2} = 0.00987$$

and the 9/10-error in the mean:

$$9/10\text{-error } (\bar{k}) = dk = z_{13}\sigma_{\bar{k}} = \pm\ 1.782\ \sigma_{\bar{k}} = \pm\ 0.0176$$

where z_{13} is the unit deviate at $P = 0.1$ for 13 observations on the 't'-distribution. This number, the 9/10-error, means the following: if the fluctuations among the values of k are truly random, and if we draw a large number of such samples of 13 from among the whole population of possible languages, and compute similar sample means $\bar{k}$ for each, then on the average 90% of these sample means will lie within 0.0176 on either side of the mean we obtained, viz. $\bar{k} = 0.8048$.

Now if 90% of all sample means can be expected to lie within 2.2% of our mean $\bar{k}$, we seem justified in assuming some considerable central tendency among the k's. It is quite certain that they are NOT all identical, and indeed, statistical theory would predict just this fact; but all we desire is that the k's be sufficiently close together so that our assumption that they are all equal will not introduce an intolerable error into further calculations.

3.2. Temporally constant *k*. In §1.1 we stated our first working hypothesis, that the rate-constant R (or λ, or k) is constant in time for any language. In order to check this assumption it would be necessary to measure the rate-constant for a given language at various periods in its history. Since we are reluctant to accept any data for time-depths of less than 500 years, it is very difficult to find a language for which word lists could be prepared at 500-year intervals over several thousand years.

Perhaps the only language for which this could be done is Assyro-Babylonian, which covers about 5000 years of written records. But it is difficult to obtain word lists by the chosen method from existing Assyriological materials, and the author has not yet been able to assemble the necessary data.[14]

Some evidence tending to justify the assumption of a temporally constant k is seen among the thirteen determinations of §2.3, where we can find no discernable correlation between t and k.

3.3. Independence of selection. Our other basic assumption (in §1.21) is that the selection of items for replacement in the basic-root-morpheme inventory of one dialect is statistically independent of the selection in another dialect. Now we know of course that for two diverging dialects there can be no complete statistical independence, for the internal forces (whatever they are) which originally caused morpheme decay in the undifferentiated proto-language will certainly continue to affect the two in the same way, at least initially. Furthermore, although the two speech communities are geographically separated, so long as they remain in communication we shall expect certain external factors as well to affect them equally.

The only convenient way to test this hypothesis is to calculate the time-depth of related dialects which began diverging at a known date. If we find any discrepancy in time-depth, we shall ascribe it to a lack of independence.

3.31. German/English time-depth. When we compare the word lists for Modern German and Modern English, we find 124 cognates (58.5%) in a total of 212 words. By means of Eq. 9 we can calculate the time-depth:

$$t = \frac{\log .585}{2 \log .805} = 1.236 \text{ millennia.}$$

Counting 1,236 years back from 1952, we would predict that German and English began to diverge in basic-root-morpheme inventory about 716 A.D. But since the Germanic invasions of Britain began about 449 (though there was probably considerable traffic and intercommunication up to the year 600), our estimate would seem to be too late: the Middle German dialects which were the main source of Modern German must have separated from the northern dialects which were transplanted to Britain several centuries at least before our date.[15]

Before we ascribe this deviation to lack of independence between the two dialects, we must assess the limits of error in our answer to see if the allowable range does not perhaps include the historical date (see §4.3).

3.32. Turkish/Azerbaijani/Uzbek time-depths. A similar calculation was made

[14] Members of the Assyriology staff in the Oriental Institute, University of Chicago, may be able to supply the necessary word lists.

[15] One could object that part of this discrepancy may be due to failure of our first hypothesis of interlingually constant k. But if we recalculate, using the individual rate-constants for German and English instead of the mean constant $\bar{k}$, we reduce the date by only 27 years, to 689 A.D.

for modern Istanbul Turkish and modern Azerbaijani.[16] The word list contains 209 morphemes and 166 cognates (79.4%), giving a time-depth of 0.526 millenia. This would date the split of Osmanli and Azerbaijani about 1424 A.D. The Turks took Constantinople in 1453, but had been in Anatolia since about the year 1000. There must have been some considerable intercommunication between the Anatolian and the Caucasian Turks over a period of many years, and this may account for our late estimate.

A second word list was prepared for Turkish and for Ferghana Uzbek.[17] On the basis of 177 usable words, of which 117 (66.2%) were identifiable cognates, the calculated time-depth was 0.954 millenia. This would indicate that the Osman tribes may have separated from their Uzbek speaking relatives about the year 1000, which compares favorably with their date of entry into Anatolia.

3.4. Miscellaneous calculations. Swadesh has calculated an Eskimo/Aleut time-depth and compared it with a carbon-14 dating of the earliest Aleutian settlement. Both dates were in the neighborhood of 3000 years.[18]

Taylor and Swadesh[19] have counted morphemes retained in Modern Carib from the Dominican Carib of 1650 A.D.; they find 93.5% cognates left in the modern dialect. Using our mean rate-constant we would put the split at 1640.

The author has prepared a word-list for French and English to estimate the time-depth of Germanic/Romance. The list of 202 morphemes shows 56 cognates (27.7%), dating the separation of these two branches at 1000 B.C. Now Trager and Smith have proposed[20] that Southwest European may have split off from General European (to leave North European) between 1800 and 1500 B.C., in any case not much later. Our late date may very well reflect some contact between Romance (or Italic) and Germanic after that time, and between Norman French and Old or Middle English.

The author has also estimated the fraction of shared cognates for Modern Gujerati and Modern Rajasthani[21] as 104/191 or 54.5%, which would put the split of these two Indic dialects at about 550 A.D.

A. Barrera Vasquez (Education Clearing House, Unesco), prepared word lists for the Mayan of Yucatan, using missionary materials from 1540–1700 and the modern language. There were 203 cognates out of 212 words (95.8%), giving a time-depth of 200 years. This is probably too short a time to insure any great precision for the estimate.

To observe the effect of including in the word list a large proportion of non-basic items, the entire English vocabulary of C. D. Buck's *Dictionary of selected*

[16] The Azerbaijani word list was provided by F. W. Householder Jr., of Indiana University.

[17] The Uzbek words, taken from the author's field notes, were supplied by Rusi Nasar of Margalan, Özbekistan.

[18] *Proc. Am. Phil. Soc.* 96.452–3 (1952). The difficulty with this example is that (1) we cannot be sure that the owners of the organic matter from which the C-14 was taken spoke Eskimo-Aleut, and (2) we do not know how long after or before the deposition of the sample the two branches of Eskaleut split apart.

[19] Taylor, *IJAL* 18.229 ff. (1952).

[20] G. L. Trager and H. L. Smith Jr., A chronology of Indo-Hittite, *SIL* 8.61 ff. (1950).

[21] Using as informants Nataraj Vashi of Bombay and Kumar Sumer Singh of Jaipur.

synonyms in the principal Indo-European languages was examined for cognates retained from Old English. The root-morphemes include many from such semantic areas as religion, government, and military life. Out of 1010 words used, 573 (56.8%) are cognate. Assuming a time-depth of 1000 years, this figure gives a retention rate of 0.568/millennium, nearly 30% lower than our mean rate-constant 0.805.

In Buck's chapter on body parts and functions alone, there are 68 cognates out of 97 words. The indicated rate of 0.701 for these items is still 13% below our mean. This may perhaps reflect a large incidence of tabu among the items of this semantic area.

3.5. Homogeneity. The randomness or representativeness of our word list was checked once by splitting the German/English list (§3.31) into two parts in a statistically arbitrary way, viz. alphabetically. The fraction of shared items for the two halves was calculated as follows: items 1–112, 65 cognates, $F_s = 0.580$; items 113–213, 59 cognates, $F_s = 0.590$. This close check would imply that the list used is at least homogeneous.

4. Limits of Error

We come finally to the most important consideration of all. Assuming that the rate-constant for morpheme decay is real and has the value that we obtained for it (§2.3), and assuming further that it is interlingually and temporally constant, with what degree of precision does this method allow us to specify our data?

4.1. Errors in F_s and k. In §3.12 we showed that we can estimate the limits of sampling-error in our mean rate-constant by assuming that the differences among the individual k's are random (and normally distributed). We expressed this as the 9/10-error:

$$dk = \pm 0.0176$$

Now the fraction of shared items F_s obtained from correlated word lists for two related dialects is a proportion drawn on a sample of some 200 items, presumably representative of the entire basic-root-morpheme inventory of each language. We shall accordingly expect to find also a sampling error in F_s. This may be expressed as a standard-error in the proportion F_s:

$$dF = \sqrt{\frac{F_s(1 - F_s)}{m}} \tag{11}$$

4.2. 9/10-error in t. Finally, since both of these sampling-errors will contribute to errors in t, we can calculate from them a 9/10-error in t:[22]

$$dt = \pm 0.0946 \sqrt{t^2 + \frac{1569}{m} \cdot \frac{1 - F_s}{F_s}} \tag{12}$$

where m is the number of words in the list used.

[22] From Eq. 9,

$$t = \frac{\ln F_s}{2 \ln k},$$

4.3. Range in time-depths. We may now return to the calculated time-depths in §3.3 and §3.4, and compute the limits of error in t.

The time-depth for English/German was given (§3.31) as 1.236 millenia. From Eq. 12 we calculate the 9/10-error in t:

$$dt = \pm 0.246 \text{ mill.},$$

and then we may state:

$$t = 1.236 \pm 0.246 \text{ mill. at the } 90\% \text{ level.}$$

On a calendar scale, German and English may be said to have separated lexically somewhere between 470 and 962 A.D., with 90% assurance within the limitations of our theory. This means that if we recompute the time-depth for German/

---

and expanding in a Taylor's series about the point (k_0, F_0):

$$t - t_0 = \frac{\partial t}{\partial k}(k - k_0) + \frac{\partial t}{\partial F}(F - F_0) + \cdots,$$

where the partial derivatives are to be evaluated at the point (k_0, F_0). Neglecting terms of higher order, multiplying out, and collecting constant terms:

$$t = \frac{\partial t}{\partial k} k + \frac{\partial t}{\partial F} F + \left(\frac{\partial t}{\partial k} k_0 + \frac{\partial t}{\partial F} F_0 + t_0\right).$$

Now since var $(c) = 0$, var $(x + y) =$ var $(x) +$ var (y), and var $(cx) = c^2$ var (x), the variance in t is given by:

$$\text{var}(t) = \left(\frac{\partial t}{\partial k}\right)^2 \text{var}(k) + \left(\frac{\partial t}{\partial F}\right)^2 \text{var}(F),$$

or, since var $(x) = (\sigma_x)^2$,

$$\sigma_t = \sqrt{\left(\frac{\partial t}{\partial k}\right)^2 \sigma_k^2 + \left(\frac{\partial t}{\partial F}\right)^2 \sigma_F^2}.$$

Substituting for the partial derivatives

$$\frac{\partial t}{\partial k} = -\frac{t}{k \ln k} \quad \text{and} \quad \frac{\partial t}{\partial F} = \frac{1}{2F \ln k}$$

and also the expression for

$$\sigma_F^2 = \frac{F(1 - F)}{m},$$

we obtain:

$$\sigma_t = \frac{\sigma_k}{k \ln k} \sqrt{t^2 + \frac{k}{4\sigma_k m} \cdot \frac{1 - F}{F}}.$$

Substituting for the values $k = 0.8048$ and $\sigma_k = 0.00987$, and converting to the 9/10-error by (9/10-error) $= 1.645\ \sigma$:

$$dt = \pm\ 0.0946 \sqrt{t^2 + \frac{1569}{m} \cdot \frac{1 - F}{F}},$$

where m is the number of items in the word list.

English many times, with other word lists and mean rate-constants, 90% of these dates will fall between 470 and 962.

Now this range may be reduced indefinitely by accepting lower and lower degrees of assurance. For example, we may say these two languages began diverging lexically between 666 and 766 A.D., but now with only 24% assurance.

Similarly the other data in §3.32 and §3.4 may be expressed as follows, in each case with 90% confidence:

Turkish/Azerbaijani	1277–1571 A.D.
Turkish/Uzbek	775–1217 A.D.
French/English	1509–491 B.C.
Gujeratí/Rajasthani	258–836 A.D.
Maya of Yucatan	1678–1796 A.D.

4.4. Relative error. In §4.2 we showed how to estimate the 9/10-error in t. This number increases in absolute value from 0 years at $t = 0$, to ± 1000 years at $t = 5{,}600$ years, rising then ever more rapidly; at $t = 20{,}000$ years, the 9/10-error in t is equal in magnitude to t itself. We may also express the error as a fraction of t, i.e. as a RELATIVE-ERROR. When t is very small the relative-error $|dt|/t$ is large, dropping rapidly to a minimum of 17% at $t = 3{,}200$ years and then rising again to 100% at $t = 20{,}000$ years.

Taking our value of $\bar{k} = 0.8048$, and assuming $m = 215$ words, we may calculate the values of t, $|dt|$, and $|dt|/t$ corresponding to various values of F_s:

F_s	t	$\|dt\|$	$\|dt\|/t$	F_s	t	$\|dt\|$	$\|dt\|/t$
.99	23	26	115.0	.45	1,850	339	18.3
.95	119	61	50.9	.40	2,120	380	17.9
.90	244	90	36.8	.35	2,440	425	17.4
.85	377	115	30.5	.30	2,790	480	17.2
.80	517	139	26.9	.25	3,220	548	17.0
.75	666	163	24.5	.20	3,730	630	16.9
.70	826	193	23.4	.15	4,390	750	17.1
.65	1,000	215	21.4	.10	5,340	935	17.5
.60	1,182	241	20.4	.05	6,950	1,312	18.9
.55	1,385	270	19.5	.02	9,060	2,020	22.3
.50	1,608	302	18.8	.01	10,690	2,780	26.0

From this table of relative-errors we see that the optimum values appear between 1,000 and 10,000 years of time-depth, where $|dt|/t$ is less than 23%. The relative-error in this time range could be reduced to less than 9.5% if we would accept a 50% confidence-level (sometimes called the probable-error), instead of 90%.

4.5. Reduction of error. In this section we shall indicate briefly a few improvements which might be made to reduce the errors in the glottochronological method. By reduction of error we mean, of course, improvement of the precision (that is, the confidence) with which we may assert our predictions, not improvement of the accuracy, which depends upon the quality of the data used and upon the theory itself.

4.51. Increase in the quantity of data. Since the sampling-error in a statistic varies inversely with the number of observations used in determining the value of the statistic, we can reduce the size of the error by using more data. By using twice as many words in our list (i.e. 430 instead of 215), we can reduce the error in t by at least 20–25 %, for now

$$dt = \pm .0946 \sqrt{t^2 + \frac{785}{m} \cdot \frac{1 - F_s}{F_s}}.$$

Similarly, by obtaining ten more determinations of k to include in our average $\bar{k}$, even though this causes no decrease in the dispersion of the k's about their mean, we can still cut in half the sampling-error in $\bar{k}$.

The 9/10-error in t will then decrease by 30–35 %:

$$dt = \pm .0473 \sqrt{t^2 + \frac{3138}{m} \frac{1 - F_s}{F_s}}.$$

4.52. Improvement of the word list. It may be possible to define certain classes of morphemes which exhibit different and perhaps more stable rates of decay. For example, if our word lists consisted only of morphemes referring to body parts, the decay-rate-constants for the test languages might all group more tightly about their mean than do those which we have determined. They would then provide a better measure of time-depth, or at least a more precise measure, i.e. a measure with lower sampling-error.

If there are classes of basic-root-morphemes which decay at different rates, the assumption of a temporally constant k for our word list (which presumably contains members of several such classes) is clearly false. Through a given time period, as the more resistant classes survive the less resistant, their concentration in the sample will increase, and the value of k will rise.