SPATIAL DISTANCE AND LEXICAL REPLACEMENT

L. L. Cavalli-Sforza, Stanford University
William S-Y. Wang, University of California, Berkeley

The 'stepping-stone' model has been used to study genetic similarity of peoples as a function of the geographical distances separating them. A gene is represented by one of several alleles, much as a meaning is represented by one of many words. Based on this parallel, the stepping-stone model is here applied to a body of linguistic data from a chain of Micronesian islands. The logarithm of the lexical similarity, when plotted against geographical distance, shows a pronounced upper concavity not found in the genetic investigations. This deviation from linearity is largely caused by non-homogeneity of replacement rates in the words studied. Another contributing factor is that the effect of distance on lexical similarity is much greater for the eastern islands in the group.

Rates of replacement in space, computed here, show a significant positive correlation with those in time, computed for a related language or for different groups of languages.*

1. INTRODUCTION. When two speech communities are separated, their languages develop in different ways. The rate at which differences accumulate depends largely on the extent to which the populations continue to interact. The less the interaction, the more rapidly the two languages diverge. Nowadays, contact is greatly enhanced by mass media and fast transportation. But until very recent times, it was the movement of populations and of individuals that generated these interactions.

Population movement can affect linguistic change in many ways. For instance, mass movements are important, especially for the macro-geographic distribution of languages. The relocation of a group (or a splinter of a larger group) to a new area is part of the history of settlement of that area, and of the origin of the local language. Usually, however, the history of settlement is complicated and poorly known; pre-existence or later arrival of groups with different languages adds complexity to the picture. But local linguistic differentiation at a micro-geographic level must be affected especially by migration of single individuals or families. Individual migration is a constant phenomenon in all societies, including traditional ones—though it is usually more intense in economically more advanced societies. Relocation of individuals can result from marriage, or can be related to work or other causes; and travel without relocation can be a common event. In earlier times, individual movement must have provided an important social exchange for language.

Following the classic work of Hägerstrand 1967, geographers have shown that social exchanges of all sorts tend to decrease regularly with increasing distance. Quantitative study usually supports the validity of simple mathematical functions for the dependence of social contact on spatial distance. Because individual migration is usually maximal at short distances, one can expect it to exercise most of its influence on linguistic differentiation at a micro-geographic level.

According to the classical 'wave theory' of diffusion of linguistic change, innovations spread from their place of origin in concentric, circular waves, generating the 'isoslosses' which have been so frequently described in linguistic geography. If each new, independent innovation spreads independently from earlier ones, the result should be a highly varied set of intersecting isoslosses. Geographic, political, or social barriers which generate discontinuities in social (and therefore also linguistic) exchange put constraints on this free variation; and specific patterns of isoslosses may emerge in response to orographic, hydorgraphic, or political geography. But wherever these constraints do not dominate the picture, there should be greater similarities, on average, between two languages spoken at a shorter geographic distance. This expectation was recognized early by the proponent of wave theory, J. Schmidt, in 1872. A succinct statement of the theory was given by Bloomfield (1933:317):

'Schmidt showed that special resemblances can be found for any two branches of Indo-European, and that these special resemblances are most numerous in the case of branches which lie geographically nearest each other. . . . Different linguistic changes may spread, like waves, over a speech-area, and each change may be carried out over a part of the area that does not coincide with the part covered by an earlier change. The result of successive waves will be a network of isoslosses. Adjacent districts will resemble each other most; in whatever direction one travels, differences will increase with distance, as one crosses more and more isoslosselines.'

Over the past century, the study of linguistic relationships has been dominated by genetic models aimed at constructing tree diagrams, even though the weaknesses of such models have often been noted (cf. Southworth 1964); more recently, attempts have been made to overcome some of these weaknesses (cf. Hsien 1973, Krishnamurti et al. 1983). By contrast, very little effort has been made at quantifying the effect of geographic distance on linguistic variation. This imbalance is being corrected, however, as scholars once again concern themselves with 'locating language in time and space' (cf. Labov 1980).

Models of the isolating effect of distance that have been developed and applied in biology lend themselves to use in linguistics; and our paper is devoted to applying them to a selected set of linguistic data, which fit the requirements of the simplest theoretical model fairly closely. Among the advantages to be derived by the use of this model are the possibility of obtaining simple measures of the effect of distance, and that of comparing synchronic and diachronic differentiation. In §2, we present the model in its original biological context. It will then be applied to linguistic data in later sections.

2. THE MODEL OF ISOLATION BY DISTANCE. The biological model used here was first developed by Malécot 1959, and later independently by Kimura & Weiss 1964. The name 'isolation by distance' goes back to related earlier work of Wright 1943, 1946. In the simplest form of the model, the distribution of the population is hypothesized as concentrated in communities, which are located geographically in a linear succession. Kimura & Weiss proposed the term 'stepping-stone model', suggesting the image of a series of steps that traverse a pool.

* This manuscript was completed while the authors were Fellows at the Center for Advanced Study in the Behavioral Sciences. We are grateful for financial support provided by NIH grant GM 20467, NSF grant BNS 76-22943, an Exxon Education Foundation grant, and Alfred P, Sloan Foundation grant 82-2-10.
Exchange between them is determined by the movement of individuals from one village to a neighboring one. For the necessities of mathematical modeling, hypotheses are extremely simple: e.g., individual communities are usually composed of the same number of individuals; and the proportions of migrants is always the same for each village. The aim of the mathematical theory is to predict how the average similarity to be expected between villages (communities) is affected by the distance between villages (measured in number of steps). It is intuitively clear that the following forces will operate:

(a) The rate at which innovations (in biology, mutations) arise: if this is low, villages will tend to remain more similar. If it is high, every village may accumulate some innovation of its own in a fairly short time, and local differentiation increases.

(b) The rate of migration, i.e., of exchange between neighbors: if this is low, villages will tend to differentiate more easily. If it is high, innovations will be more easily shared with neighbors. In the first generation, only the nearest neighbors are reached; but in later ones, innovations can spread to others, and eventually permeate the whole chain of villages. Thus, while mutation has a differentiating effect, migration has an opposite, homogenizing effect between villages.

Let us consider a given gene which can exist in a number of different states called ‘alleles’. For simplicity, let us consider at first one gene per individual. Different individuals may have different alleles, and a community will be characterized by the frequency with which a given allele is represented in it: i.e., there may be a percentage $p^1$ of individuals with allele no. 1, $p^2$ of individuals with allele no. 2, etc. A mutation generates a new allele, which will be transmitted to the progeny and which represents a biological innovation. At the beginning, it will be present only in one individual; but after generations, it may be present in many individuals, and may spread to neighboring communities through migration.

It is clear that, the closer communities are, the greater the similarity between them will be; but we are interested in predicting more accurately how their similarity will vary with distance. The first question is how to measure the similarity. One possible method is Pearson’s correlation coefficient $r$. Other measures—such as the proportion of shared cognates, familiar to linguists—are even simpler.

The general shape of the relationship between similarity $r_{ij}$ of two villages $i,j$ and their geographic distance $d_{ij}$ was found to be represented, when the forces of mutation and migration reach equilibrium, by a negative exponential:

\[ r_{ij} = e^{-kd_{ij}} \]

Here $r_{ij}$ can be expressed by a correlation coefficient, or more simply by the proportion of shared cognates, and $k$ is a constant that depends on the two forces just mentioned. To test the graphic fit of the hypothesis to the data, it is simplest to plot the logarithm of the similarity measure vs. the distance:

\[ \log r_{ij} = -kd_{ij} \]

We expect a straight line going through the origin, with slope equal to $k$, the replacement rate (the reciprocal of a retention rate). The mathematical theory shows that the value of $k$ should be

\[ k = \frac{2\sqrt{u/m}}{m} \]

Here $m$ and $u$ are the migration and mutation rate respectively. Thus the replacement rate $k$ will be greater with greater mutation, and smaller with greater migration. We measure $m$ as the proportion of adult inhabitants entering a village at every generation, and $u$ as the probability of an innovation (mutation) appearing in each generation per individual. This model has been extended theoretically to cover geographically less restrictive situations (cf. the summary in Cavalli-Sforza & Bodmer 1971); the examination of many human populations shows that formula 2, and therefore formula 1, fit a great variety of real genetic situations (Morton 1982).

The strict validity of formula 3 in linguistic applications is limited by the fact that it is derived on the condition of transmission which is valid for genes, i.e., strictly from parent to child. This is not valid for linguistic innovations, which may spread between speakers of a great diversity of relationships: parent to child, teacher to student, and among neighbors or friends. Further theoretical work is necessary for a complete validation of formula 3 in linguistic applications; but formulas 1–2, as well as the opposition of effects of the rates of migration and mutation, indicated by their appearing in the numerator and denominator of $k$ in 3, respectively, should remain valid.

3. The Micronesian data. For a first application of this theory, we looked for a situation satisfying that of the simplest model: linearity of geographic distribution, discreteness, and approximately equal dimensions. The closest approximation we could find to this ideal situation is a chain of islands in Micronesia. The Pacific islands are the latest region of the world populated by man. The difficulty of reaching them was increased when, some 11,000 years ago, the level of the sea rose about 100 meters, because of the melting of the great ice caps that covered much of the northern hemisphere. Authorities generally agree that the Austronesians left their homeland in southern China six or seven thousand years ago. Some of them settled in Micronesia and small parts of Melanesia; others journeyed on and fanned out in a remarkable arc, reaching from Easter Island, off Chile, all the way to Madagascar, off Africa.

To these parts of the world they brought varieties of the Austronesian language, which may ultimately be related to the Tai languages of southern China and continental Southeast Asia. Probably as a result of the great distances separating the islands, as well as the sparse island populations, the linguistic diversity of Austronesian is remarkably high. It is estimated that Island Southeast Asia and Oceania contain some 1400 indigenous languages, out of some 7000 estimated for the entire world. This ratio can be seen as high when we consider that the total number of speakers for this area is only some 150 million, in a world population of four billion. Roughly calculated, 4% of the world’s population speak 20% of its languages.
The area of special interest here is the horizontal chain of the Caroline Islands, one of the four large island chains of Micronesia. Our data come from 17 different islands in this chain (see Figure 1), mostly atolls with limited pop-

![Figure 1. Geographic location of the 17 Micronesian islands in Table 1.](image)

ulations. They are listed in Table 1, together with their site numbers and population sizes. These islands lie along a narrow band some seven degrees above the equator, and span approximately 1500 miles. It is likely that they were uninhabited before the arrival of the Austronesians. Although these islands may have been known to Europeans since around 1500, the dominant influence during this century was that of Japan—until 1945, when they came under American trusteeship. Politically, the Truk group of islands on the east and the Yap group on the west are important. Linguistically, the languages have been grouped under the label 'Trukic', which is a branch of Nuclear Micronesian. Bender 1971 provides a useful account of the languages of this region.

The linguistic data on the 17 islands were first gathered by E. Quackenbush in 1966. In his 1968 dissertation, these data were reported together with a set of phonological correspondences for the islands, from which one can reconstruct the proto-forms. A team led by Byron M. Bender at the University of Hawaii have analyzed these data for cognate relationships, with the aid of a computer. We have modified the data somewhat in accordance with our own requirements.

Essentially, the data consist of a list of 571 glosses, and their corresponding words as elicited from a consultant for each of the 17 sites. To reduce the data to a form suitable for quantitative analysis, it is necessary to discuss some of the cases how the glosses are reflected at the 17 sites. To aid in this discussion, we will use the abbreviations G for gloss, S for site, and M for morpheme: thus the abbreviation 'G(99)' refers to the gloss 'full (stomach)'; and 'G(100)' refers to the gloss 'full (box)'.

The data show that, for G(99), Site 1 has the morpheme /matY/; i.e., M(S1) = /matY/. For the same gloss, M(S2) = /mat/; M(S3) = M(S4) = /matY/, M(S5) = M(S6) = M(S7) = /mat/; M(S8) through M(S13) = /mat/ as in S2, and finally M(S14) through M(S17) = /met/. Since /matY/, /mat/, /mat/, /mat/ and /mat/ are all cognates, no lexical change has taken place for G(99). For this gloss, the input file for the quantitative analysis will take the following form, where the 17 repetitions of 'one' show that all the sites have cognates:

\[
\text{G(99)} \quad 11111 \quad 11111 \quad 11111 \quad 11111
\]

For G(100), the situation is more complex: M(S1) = /wulo/; M(S2) = /wug/, M(S3) through M(S7) = /ssog/, M(S8) = /ssogo/, and M(S9) through M(S17) = /wur/. Here two sets of cognates exist: /wulo/ = /wug/ = /wur/ on the one hand, and /ssog/ = /ssogo/ on the other. They are entered in the input file as follows:

\[
\text{G(100)} \quad M1 \quad 11000 \quad 00001 \quad 11111 \quad 11111
\]

\[
\text{G(100)} \quad M2 \quad 00111 \quad 11100 \quad 00000 \quad 00
\]

We can state the analogy of genes and alleles with linguistic elements in a number of different ways. For present purposes, we will use the following convention: a gloss is taken as the equivalent of a specific gene, and sets of cognatic morphemes of that gloss are taken as the alleles. There are differences between the evolutionary and analytic processes in linguistics and in biology.

1 See Bender 1971 and Quackenbush 1968 for more background on these Micronesian languages. We wish to thank both these scholars for sharing with us their linguistic expertise on this part of the world, as well as the data on which the present study is based. We are particularly indebted to Bender for his detailed comments on an earlier version of this paper.
which might seem to make the analogy inapplicable, but they are actually not of great importance. For instance, linguists often examine one `typical' consonant per population. This is done for practical reasons, but also because it is implicitly assumed that the information supplied by one individual is almost identical to that supplied by another. In fact, individual variation exists, but can be neglected at a first level of approximation. The existence of a period in which both the old and the new form of a word co-exist in a population is stressed by Sankoff 1971. Biologists always study a sample of many individuals from a community, perhaps for only one gene (though usually for several, the results to be used jointly). Linguists always study many words (glosses); and the multiplicity of glosses can, to some extent, replace the multiplicity of individuals in providing measures of similarity between communities.

This possibility of establishing a formal similarity between the process of lexical change and the substitution of amino acids in proteins and of nucleotides in DNA, or in general of one allele by another, is of considerable interest: a substantial fraction of the theoretical machinery developed for allelic gene substitution can be adapted to it. Naturally, caution should be exercised: the two processes, biological and linguistic, show important differences. We have mentioned the differences in transmission and in sampling schemes. Other differences could be mentioned, but the basic applicability of the method is not affected.

4. TEST OF THE SIMPLE RELATIONSHIP. Among 17 islands, 136 pairs of islands exist. Their geographic distances and proportions of shared cognates are shown in Table 2. In Figure 2, these data are presented graphically with a logarithmic

**Figure 2.** Relationship between proportion of cognates (shown in logarithmic scale on the ordinate) shared by two islands, and their distance in miles (Micronesian data of Table 1). Circles indicate island pairs showing the given percentage for cognates and distance; when one of the islands in the pair considered is no. 12 (Salpan), the circles are black. Formula 1 should give a straight line, which is obviously not observed. The fitted theoretical curves are explained in the text and in Appendix II. Note that the theoretical solid curve was calculated including Salpan.
scale in the ordinate. If formula 1 held, the observations should cluster around a straight line; but they obviously don’t. The theoretical curves drawn in Fig. 2 were obtained with another formula, to be explained in §5.

The possibility that some islands might be especially responsible for the curvature was tested by taking only the points which include one specific island, doing this in turn for all islands, and evaluating the residuals around the curve fitted in Fig. 2. Most islands gave results close to the common curve, others diverged from it; the worst offender was the island of Saipan. The pairs which include Saipan are those shown with black circles instead of white circles in Fig. 2; these make it graphically clear that observations which include Saipan are almost all located above the curve, and hence indicate a bias in the data.

There are three good reasons for this to have happened: (a) A smaller number of glosses is given for Saipan than for any other island, and the choice is clearly not random: in Quakenbush’s list of the words from 319 to 519, 183 are missing. (b) In addition, we have learned (Frederick H. Jackson, p.c.) that the Saipan consultant was not as good a choice as in the other islands, because of being young and of having originated, in part, from another linguistic group. Quakenbush himself takes a very cautious attitude toward the data he obtained for Saipan: ‘Saipanese Carolinian is not well integrated into the continuum, but the partial data collected on Saipan is included as a matter of interest’ (22–3). (c) Finally, it should be noted that ‘Saipan, previously unhabituated by Trukic speakers, is known to have been settled by speakers from a number of the Trukic islands included in this study within the last 100 years’ (B. W. Bender, p.c., based on Spoehr 1954.)

This last reason is perhaps the most compelling, since the distances calculated for Saipan data are actually larger by an important factor than they would be if there had not been a very recent evacuation of some of the islands. As a consequence, all the points involving Saipan represented on the graph were moved toward the right—i.e., to distance values higher, on average, than the unknown ones at which they should have been placed. This can justify qualitatively the abnormal position above the curve of almost all the points involving Saipan. If the Saipan inhabitants came originally from one specified Trukic island, one might attempt a correction; but since they came from various islands, the correction can only be approximate. It has therefore been decided to eliminate Saipan from the rest of the investigation.

Partial use of these observations was made in previous research by Bender, who gave a matrix of cognate percentages based on 176 words chosen (on the basis of lexicostatistical lists) from the 571 words given by Quakenbush for six of the seventeen islands. There is a good correlation between Bender’s results and ours (Pearson’s coefficient of correlation $r = 0.94$), although his percentages are higher than ours (by 8.4% on average). The existence of a substantial variation in rate of change of different words, to be discussed below, is probably responsible for this difference, at least in part.

Imaiizumi & Morton 1970 examined Bender’s relationship of cognate percentages and distances for the matrix of six islands—as well as, separately, another matrix of cognates between eight more widely scattered Micronesian islands, reported by Bender on the basis of information from Dyen 1956. Imaiizumi & Morton stated that a straight line fit these data well, but they reported no test of goodness of fit.

The discrepancy between Imaiizumi & Morton’s findings and ours may result in part from the way of computing the percentage of cognates. The original list of 571 words increased to 647 with our procedure, since we broke up compound words into their component morphemes; treated independently, these showed a greater variation in substitution rate than the rest. Even by testing them separately, it became clear that compound words can account for only a small part of the deviation. We conclude that the data show a distinct deviation from linearity—i.e. from formulas 1 or 2—which is clearly manifest only if one considers the whole or most of the Quakenbush sample of words and islands; but it is shown in a more indirect way with the restricted sample first given by Bender, and used by Imaiizumi & Morton.

The main reason for the discrepancy between Fig. 2 and the result of Imaiizumi & Morton is in the particular set of islands chosen in their investigation. The six islands they used would give a straight line with our percentages of cognates; but the line would not pass through the origin, as it is expected to do on the basis of formula 2. Thus, even with the six-island sample, there is a deviation from the expected relationship; strictly speaking, this is one of proportionality, not of linearity.

5. VARIATION OF REPLACEMENT RATES FOR INDIVIDUAL WORDS. It is known that variation of replacement rates exists for individual words over time (e.g. Swoodesh 1955, Joos 1964, Kruskal et al. 1973, Sankoff 1973). We conjectured that a similar variation might be responsible for the deviation from proportionality of the log cognate-distance curve of Fig. 2. The procedure used to test the importance of this possible source of deviation went through the following steps: first, evaluate the rate of replacement for each word; second, study the variation of individual replacement rates; finally, test the hypothesis that it can fully explain the curve observed in Fig. 2.

We evaluated separately the rate of replacement of each word on the assumption of negative exponential decay, as in formula 1. The method of interpolation was maximum likelihood (see Appendix 1, below).

Figure 3 shows the distribution of $k$ values obtained for all words (glosses) which had more than one reflex. The distribution is somewhat skewed. For 190 words, about 30% of the total, we find only one reflex, and thus a replacement rate of zero. They have not been included in Fig. 3, and the distribution is thus truncated at zero. Analysis showed that the distribution of the words is neither a truncated normal nor a truncated asymmetric one (such as the log normal); its shape is more similar to, but not identical with, gamma distribution (cf. Sankoff 1973).

The number of reflexes is very highly correlated with replacement rates. Figure 4 shows how the latter vary with the number of reflexes. The replacement rate is closely proportional, on average, to the number of reflexes per word, minus one. The only substantial deviation is for the class of words with
the greatest number of reflexes (9+). The number of reflexes might be used as a simple, indirect estimate proportional to the rate of replacement of individual words; but it is not as accurate as the approach used here.

6. Does the variation of replacement rate of individual words fully account for non-linearity? When the distribution of individual replacement rates is known, and the replacement of each individual word as a function of distance is a negative exponential (as in formula 1), then one can predict the shape of the curve of the global proportion of cognates to be expected for all words considered together—i.e. the empirical data of Fig. 2. If we consider five distributions of individual replacement rates with different individual variation, the theoretical curves of the average proportion of cognates change from the single exponential (as in formula 1, when there is zero variation of individual replacement rates) to curves with upper concavity, as in Figure 5. The greater the individual variation of replacement rates, the greater the concavity.

Figure 4. Abscissa: number of roots (reflexes) per word. Ordinate: average replacement rate per mile. Bars are two times standard errors.

Figure 5. Effect of the variation of replacement rate \( k \) of different glosses on the global proportion of shared cognates (represented on the ordinate on a logarithmic scale). The variation is given by the quantity \( n \) in the gamma distribution assumed for the variation of \( k \), in formulas where \( n \) is equal to \( 1/V/CV \); and CV is the coefficient of variation of \( k \), values (standard deviation divided by the mean). For ease of comparison, the distances of the various curves are scaled so that all curves cross at \( y = 50\% \) (from Cavalli-Sforza & Feldman 1981).

The theoretical curves shown in Fig. 5 are calculated on the assumption that the individual replacement rates follow the gamma distribution, described in.
Appendix II. The principle, however, is general. The logarithm of the average proportion of cognates, plotted against distance, is a straight line only if the variation of individual replacement rates is zero (or in practice negligible); it has upper concavity which tends to increase with increasing individual variation, whatever the distribution of individual replacement rates. Qualitatively at least, therefore, the upper concavity noted in the average proportion curve of Fig. 2 could result from variation of individual replacement rates. But can this entirely account for the observed concavity?

One route we followed was to calculate, from the distribution of individual replacement rates (Fig. 3), the curve of the average proportion of cognates; we then compared it with the observed ones. In Fig. 2, the solid curve is calculated by assuming a gamma distribution of replacement rates (see also Sankoff 1973), and the dotted curve by using the observed distribution of replacement rates for individual words. The latter, in particular, seems to give a satisfactory fit; but the scatter of the observed points around the straight line is not negligible, even if one excludes from consideration (for the reasons mentioned earlier) the results obtained from Saipan.

Perhaps more troubling is the consideration that this procedure may be circular: we assumed that cognates of individual words have a negative exponential decay with distance, but it was difficult to test this hypothesis on the very small sample of data offered by each individual word. We therefore tested this crucial hypothesis more directly by grouping words into classes, depending on their individual replacement rate, and then plotting the logarithm of the average proportion of cognates for each class vs. distance. Each class should give a straight line, because the individual variation within each class is now considerably reduced. The results are shown in Figure 6.

In spite of some scatter, one can recognize a reasonable (though not perfect) fit to a straight line for the words which have high replacement rates, giving the highest slopes. One can show that the fit is still especially poor for the class of words with smallest replacement rate (the highest line with the flattest slope)—and also, though to a lesser extent, for the next class. The class with smallest replacement rate is represented in more detail in Figure 7 (p. 52). It is mostly composed of words with only two reflexes, and it is the class of the most highly conserved words (after those that have only one reflex).

A careful analysis of Fig. 7 showed the clear cut effect of a geographic factor which had not so far been considered. The map of the islands in Fig. 1 shows that islands west of the dotted line are farther apart. From Table 1, we see that they also have smaller populations. Thus the effect of distance here could be confounded with that of other factors. In Fig. 7, comparisons between islands belonging to the western group (sites 1–8 in Tables 1–2) are shown as solid circles; those between islands belonging to the eastern group (9–17 in Tables 1–2) are shown as open circles; and comparisons between an eastern and a western island are shown as triangles. It is clear that most pairs of islands from the western group fall near the average line. The eastern group tends to show a higher slope; and the east-west values tend to follow a straight line with a flatter slope. Thus the effect of geographic distance on linguistic differentiation is modified by other factors—geographic, social, or both.

A similar situation, but more blurred, is observed for words of the next higher class of replacement rates (not shown here); and with a still higher replacement rate (also not shown), the distinction of west and east tends to disappear completely. A simple interpretation is that inhabitants of the western group, more sparse and somewhat less densely inhabited, are prepared to travel longer distances for establishing contact with individuals from other islands. Thus an east-west difference in customs may affect mobility and social exchange. It is difficult to say whether this is an adaptation to the need of maintaining social interaction in spite of extreme isolation in the western islands; or a carry-over from previous customs, if there is ethnic heterogeneity among the inhabitants of the two groups of islands; or a consequence of long-established trade routes.
average in different language groups. Our results provide quantitative verification of the dialect geographer's 'chaque mot a son histoire'—a dictum usually attributed to Jules Giliérón, though traceable to Hugo Schuchardt (see Malkiel's historical account, 1967). Indeed, much of the recent work on lexical diffusion, partly anthologized by Wang 1977, rests on the diachronic truth of this dictum. The differential rate of change of each word appears to correlate with its frequency of usage—a correlation conjectured a century ago by Schuchardt, but recently verified quantitatively by Phillips 1984.

On this basis of our results, replacement rates in space correlate with those in time almost as much as can be expected for different language groups in time. This is analogous to a comparison of the synchronic variation (in space) with the diachronic variation (in time). There seems to be a good correspondence between diachronic and synchronic replacement rates. An advantage of measuring the rate of evolution in space is that it is easier to get the data needed for that measurement. One also can further explore the relative usefulness of wave theory vs. other interpretations of language evolution. It is perhaps worth adding that the causes of variation of replacement rates of individual words may be of considerable interest in offering possibilities for understanding some factors of linguistic evolution.

Appendix I

Maximum likelihood interpolation of individual replacement rates

If the chance of a particular word being cognate is $e^{-px}$ in a pair of islands at distance $x$, the first derivative of the likelihood $L$ is

$$A_1 = - \Sigma \Sigma (1 - e^{-px}) e^{-px}$$

Here $p$ is 1 for pairs of islands having the same cognate, zero otherwise; and $x$ is extended to all pairs. The variance of the estimate of $p$ is

$$A_2 = \Sigma \Sigma (1 - e^{-px})^2$$

Rigorously, this is valid only for independent pairs and a binomial distribution of cognates. It is necessary to iterate for obtaining $p$; and for convergence, it is convenient to choose a very small value of $k$.

Appendix II

Effect of the variation of individual replacement rates on the curve

If the individual rate is of a negative exponential relating $p$ to $x$, as in formula $A_1$ (2), varies according to a known distribution function $g(x)$, then the proportion of cognates $p$ averaged over all words as a function of distance $x$ is given by

$$A_3 = \int e^{-px} g(x) dx$$

Here the integral is extended to all values of $k$. This is the well-known Laplace transformation. Thus one can infer the observed curve of $p$ vs. $x$ from the distribution of $k$.

For some theoretical distribution functions, the Laplace transformation leads to a simple relationship of $p$ with $x$. Suppose that $k$ varies from zero to $\infty$ according to the gamma distribution (also called chi-square or Pearson type 3):

$$A_4 = C e^{-bx} x^{-\gamma}$$

Here $C$ is a constant (depending on $b, c$); $x$ is the base of natural logarithms; $b$ is a constant depending on the scale of distance used; and $\gamma$ is a parameter equal to the reciprocal of the square of the variation coefficient (standard deviation/mean of $k$). The Laplace transformation of $A_4$
according to A3 leads to

\[(A5) \rho = (1 + bx)^{-n}\]

(see also Sankoff 1971).

This is the function used in Fig. 2 to fit the solid curve. Theoretical curves are shown in Fig. 3 for various values of \(n\). The mean and measures of variation of \(k\) are connected to \(b\) and \(n\) as estimated in A5 by these relationships:

\[
\text{(A6) mean} = \frac{ab}{n}
\]

\[
\text{standard deviation} = \frac{bn}{\sqrt{n}}
\]

\[
\text{variation coefficient} = \frac{1}{bn}
\]

It can be seen that, the smaller \(n\) is, the more concave is the curve of \(p\) vs. \(x\), and the more skewed the distribution of \(k\). With \(n\) large, the variation coefficient of \(k\) is small—and curve A5 leads to a negative exponential, as in formula 1.

A5 has proved fairly flexible, and shows a not unreasonable fit to the data of Fig. 2. Unfortunately, however, our observed distribution of \(k\) is not in a good fit with A3, in part because of truncation. In the present case, the fit of the \(p\) vs. \(x\) curve with A5 is useful only for descriptive purposes; it does not accurately predict the mean and variance of \(k\), as would have been desirable. Sankoff 1973 discusses the use of the gamma distribution in some glottochronological examples, for the variation of the replacement rate (in time). We have tried distributions other than the gamma (e.g., a truncated normal, whose Laplace transformation is also easily found), but we obtained only slightly better results, not sufficiently accurate for most purposes.

Since the observed distribution of \(k\) cannot be easily reduced to a simple theoretical one (to explore whether it can account for the deviation from linearity), one must use the observed distribution of \(k\) values in A3, replacing the integral with a sum over all observed values of \(k\). The result is given as the dotted curve in Fig. 2. The fit of the observed data seems reasonably satisfactory; it would seem that this method gives a positive answer to the question whether the variation of \(k\) can explain the deviation from linearity. There is some circularity in this procedure, however, since the \(k\) value of each word has been estimated assuming an exponential relationship—an assumption which is not easy to test directly for each word. We have therefore proceeded to group words with similar replacement rates; and to test the linearity of the decay of similarity for each group, as described above.

REFERENCES


DyEN, ISIDORE. 1956. Language distribution and migration theory. Lg. 32:611–27.


KRISHNAMURTI, Bh.; Lincoln MOORE; and DOUGLAS DANKERTH. 1983. Unchanged cognates as a criterion in linguistic subgrouping. Lg. 59:541–68.

KUSHKA, Konstantin; Isidore Dyen; and F. Black. 1973. Some results from the vo-