Journal of Classification 13:27-56 (1996)

Quantitative Properties of the Evolution and Classification of Languages

Eric W. Holman

University of California, Los Angeles

Abstract: Statistical analyses of a published phylogenetic classification of languages show some properties attributable to taxonomic methods and others that reflect the nature of linguistic evolution. The inferred phylogenetic tree is less well resolved and more asymmetric at the highest taxonomic ranks, where the tree is constructed mainly by phenetic methods. At lower ranks, where cladistic methods are more prevalent, the asymmetry of well resolved parts of the tree is consistent with a stochastic birth and death process in which languages originate and become extinct at constant rates, although poorly resolved parts of the tree are still more asymmetric than predicted. Other tests applied to a sample of historically recorded languages reveal substantial fluctuations in the rates of origination and extinction, with both rates temporarily reduced when languages enter the historical record. For languages in general, the average origination rate is estimated to be only slightly higher than the average extinction rate, which in turn corresponds to an average lifetime of about 500 years or less.

Keywords: Phylogenetic tree; Cladistics; Language evolution; Birth and death process; Evolutionary rates.

This research was supported by a grant from the UCLA Academic Senate and by computer time from the UCLA Office of the Academic Computing. I thank Merritt Ruhlen, Joseph B. Slowinski, and Thomas D. Wickens for helpful information and suggestions.

Author's Address: Eric W. Holman, Department of Psychology, University of California, Los Angeles, California 90024, U.S.A. (iap8ewh@mvs.oac.ucla.edu)

1. Introduction

Comparisons of linguistic and organic evolution date back at least to Darwin (1871, p. 58), who stated: 'The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously the same.' This analogy attracted the attention of both biologists and linguists in the nineteenth century but suffered an eclipse during most of the twentieth. A recent renewal of interest is apparent in the collections edited by Hoenigswald and Weiner (1987) and Hawkins and Gell-Mann (1992).

One cause of the current revival is the increasing use of phylogenetic or cladistic classifications in biology since the work of Hennig (1950, 1966). Such classifications aim to present a complete phylogenetic tree rather than just a subset of nodes corresponding to relatively distinctive groups of organisms. Platnick and Cameron (1977) noted that phylogenetic classifications have long been the norm in historical linguistics, and that such classifications are constructed by similar methods in both biology and linguistics.

Another cause of renewed interest is the correspondence between the contents of linguistic and biological classifications reported by Cavalli-Sforza, Piazza, Menozzi, and Mountain (1988) and Cavalli-Sforza, Minch, and Mountain (1992). A phylogenetic tree of languages was found to be significantly correlated with a tree of ethnic groups based on genetic data. This result has generated considerable controversy because the trees overlap only at a level of analysis that is broader than most linguistic classifications and narrower than most biological classifications; both the validity of the classifications and the degree of their relationship have been debated (Bateman, Goddard, O'Grady, Funk, Mooi, Kress, and Cannell, 1990).

There are differences as well as similarities between biological and linguistic classifications. In biology, cladistic classifications include at most a few hundred species or groups of higher rank, and such classifications are scattered throughout a much larger system the rest of which has been little influenced by strictly phylogenetic methods. In linguistics, however, cladistic methods have always been dominant and have by now produced a comprehensive classification that includes practically all the five thousand or so known languages. Linguistics thus offers a unique opportunity to investigate what kinds of information are conveyed by a very large phylogenetic classification.

The most recent and complete classification of languages was published by Ruhlen (1987, 1991); the later edition (used here) differs from the earlier only in a few places at the highest taxonomic ranks. This classification was also used in the genetic studies of Cavalli-Sforza et al. (1988, 1992). Like other large linguistic and biological classifications, Ruhlen's is a compilation of smaller classifications previously constructed by specialists. The cladistic principles underlying most of these classifications were implemented mainly by intuitive methods; reliance upon intuition is also typical of large classifications in general (Holman, 1992).

Three aspects of Ruhlen's classification are analyzed. The first is the resolution of the classification as indicated by the number of branches per node. Maximum resolution would correspond to a binary tree with two branches at each node. The actual resolution is compared to results from a collection of conventional classifications, biological and nonbiological, described by Holman (1992). In these classifications, resolution was independent of the content of the classification but depended upon taxonomic rank, with nodes at the highest and lowest ranks having more branches than nodes in between. The generality of this pattern across classifications suggests an origin in cognitive processes of memory and judgment. The question now is the extent to which the same pattern also applies to a classification that differs not only in content but also in taxonomic method.

The next property is the asymmetry of the phylogenetic tree embodied in the classification. Research in biology has investigated whether phylogenetic trees are asymmetric enough to imply adaptive differences between their branches. The appropriate null hypothesis is the stochastic birth and death process: species split to form new species at a constant rate, species become extinct at another constant rate, and these events are independent of each other. Savage (1983) found that this model provides a reasonable first approximation to the distribution of branching patterns in the inferred phylogenetic trees of living species and groups at higher taxonomic ranks. Further investigation by Guyer and Slowinski (1991) and Heard (1992) revealed a tendency for phylogenetic trees to be more asymmetric than predicted; but according to Guyer and Slowinski, the asymmetry does not appear in the minority of trees that are based on relatively large amounts of data. These results suggest that the inferred phylogenetic trees are determined both by the actual evolutionary branching process, which follows the birth and death model, and by a taxonomic bias toward asymmetric trees, which becomes prominent when the data are weak. A similar birth and death model can also be defined with languages substituted for species. The alternative hypothesis does not involve adaptation, however, for as Greenberg (1959) pointed out, "...one language succeeds another, not because it is more advanced as a language, but for extra-linguistic reasons of military, economic or cultural superiority of its speakers." The question is therefore whether such differences are consistent enough to produce significant asymmetry in the classification of languages that cannot be explained in terms of taxonomic bias.

The last property to be studied is the relation between the classification and the historical record of languages. Extinct languages, like fossil organisms, are placed in the same taxonomic system as living ones. With its added time dimension, however, the taxonomic fossil record has produced information about evolutionary rates that is not available in classifications of living organisms. For instance, the average extinction rate of species turns out to be nearly the same as the average origination rate (Raup, 1978). Also, in contrast to the birth and death model, there is clear evidence for variation in extinction rates across time (for instance, Raup and Sepkoski 1982, Hubbard and Gilinsky 1992) and also some evidence for differences in evolutionary rates between groups of organisms (reviewed by Holman 1989). Whether similar information can be obtained from the historical record of languages is the last question to be addressed here.

2. Taxonomic Rank and Resolution

Ruhlen's classification includes 5220 languages, 4741 living and 479 extinct (not counting unclassified languages, pidgins and creoles, or invented languages); they are classified in a rooted tree with 2038 internal nodes. Each node corresponds to a group of languages that is assumed to include all the descendants of a single ancestral language; the classification does not (intentionally) allow groups from which some descendants of the common ancestor have been removed (such as the reptiles, which are defined as amniotes that are not mammals or birds). The taxonomic rank of a node is defined as the total number of nodes on the path from the root of the tree to the given node (including the root and the given node). Unlike ranks in nonlinguistic classifications, all ranks except the highest (Rank 1, for the root) are optional in the sense that they are not defined for all branches of the tree; in fact, the number of ranks ranges from 2 for branches leading to so-called language isolates up to 19 for some branches within the group of Bantu languages. Nodes at the same rank in the tree are intended to be comparable if and only if they are coordinate, that is, adjacent to the same node at the next higher rank; the ranks are not supposed to reflect the same degree of distinctiveness or importance across the whole classification (as are biological ranks such as genus).

The pattern of groups in the classification embodies Ruhlen's inferences about the sequence of branching events that produced the entire set of languages. As these inferences are in part a matter of judgment, their unavoidable fallibility may reduce the accuracy of the classification. The standard comparative method in linguistics, like cladistic procedures in biology, infers phylogenetic groups of languages from the possession of shared innovations inherited from their common ancestors. Because of limitations in data, however, groups are also sometimes based on overall similarity, as in phenetic rather than cladistic classifications. The latter situation is particularly likely at the highest taxonomic ranks, where extreme time depths obliterate most (some would say all) traces of common ancestry. Ruhlen (1991, pp. 380-389) indeed reports that some of his high-ranking groups are not generally accepted because they are beyond the reach of the comparative method, and Matisoff (1990) presents the arguments of the skeptics. At the other end of the taxonomic hierarchy, the distinction between languages and dialects is often based on an arbitrary threshold of similarity, as is the parallel distinction in biology between species and subspecies or varieties. Ruhlen describes (pp. 6-9) how some groups at the lowest ranks above individual languages are also based on arbitrary divisions of an underlying continuum.

Some evidence on the effect of these judgmental factors can be obtained by studying the resolution of the classification as a function of taxonomic rank. At the maximum resolution, a node with exactly two branches indicates that the ancestral language split into two descendants. A node with three or more branches, however, is ambiguous: either the ancestral language split into three or more descendants at about the same time, or else the available data are not sufficient to determine the actual sequence of branches (Ruhlen 1991, p. 256). This ambiguity cannot be resolved for most individual nodes; but in the aggregate, the classification is likely to be less securely based at ranks with many branches per node.

In fact, the number of branches per node does depend to some extent upon taxonomic rank. At Rank 1, the root of the tree has 16 branches -- 12 phyla and 4 extinct language isolates; fewer than 1% of the nodes in the entire classification have as many as 16 branches. At Rank 2, the phyla have a mean of 7.67 branches each, whereas the mean number of branches per node across all lower ranks is 3.53. With phylum as the unit of analysis, a t-test for matched scores was used to compare (at the 5% significance level, two tails) the number of branches per node at the second and lower ranks; scores were first transformed logarithmically to reduce skewness in the distributions. Nodes proved to have significantly more branches at Rank 2 than at lower ranks, $t(11) = 2.73$. Differences among the lower ranks were tested for the 11 phyla whose classifications extend below Rank 3. An analysis of variance on the log mean number of branches per node showed no significant differences among Ranks 3 through 5, $F(2, 20) = 1.11$; but another analysis of variance showed significant differences among Ranks 2 through 5, $F(3, 30) = 8.64$, confirming the previous t-test. Since fewer than 11 phyla are represented below Rank 5, another method was used to test for differences at the lowest ranks. Lowest-rank nodes were defined as nodes all of whose branches are individual languages, while intermediate-rank nodes were defined as nodes below Rank 2 some of whose branches lead to nodes at lower ranks; there are 1166 lowest-rank nodes and 859 intermediate-rank nodes in the entire classification. A matched t-test showed no significant differences in log mean number of branches between lowest-rank and intermediate-rank nodes, $t(10) = 0.56$.

In summary, nodes at the top two ranks tend to have more branches than nodes at lower ranks, which in turn are fairly homogeneous across ranks. A large number of branches at the highest rank was also found in most other large classifications by Holman (1992), and may therefore be a general characteristic of taxonomic judgment, independent of content or method. Large numbers of branches per node at the second-highest rank, however, were not generally observed in other classifications, and thus may specifically indicate the difficulty of inferring relatively ancient groups of languages. The absence of a difference in resolution between lowest-rank and intermediate-rank nodes suggests that groups at the lowest rank are not appreciably affected by the arbitrary distinction between languages and dialects. The homogeneity of resolution at low ranks also stands in contrast to the larger number of branches per lowest-rank node found in most other classifications; this contrast may be related to the fact that in language classifications, unlike classifications with mandatory ranks, the lowest rank is different on different branches. Finally, the mean of 3.53 branches per node at ranks below the second is lower than most of the mean numbers of branches per nodes at these ranks in other classifications according to Holman's (1992) Table 2; this difference is consistent with the inference that the resolution of conventional classifications reflects properties of memory and categorization rather than attempts to maximize resolution as in cladistic classifications.

3. Asymmetry and the Birth and Death Model

Next, the asymmetry of the inferred phylogenetic tree of languages is compared to predictions from the birth and death model. Asymmetry depends upon the distribution of languages across branches. For nodes with exactly two branches, the expected distribution has been derived from the birth and death model by Farris (1976) and also by Slowinski and Guyer (1989); related results on the expected distribution of trees are presented by Harding (1971) and Page (1991). If there is a total of n living languages on both branches from a node, then there are $n - 1$ possible distributions of numbers of languages across the branches: a given branch may have 1, 2, 3, and so on up to $n - 1$ languages, with the remainder on the other branch. Farris showed that each of these distributions is equally likely under the null hypothesis of the birth and death model, conditional upon the existence of exactly n living languages in the whole group.

The statistical implications of this prediction can be illustrated by applying it to Niger-Kordofanian, the only language phylum that has two branches with living languages on both branches. The smaller branch is Kordofanian, with 32 of the 1064 living languages in the phylum. The probability of obtaining a distribution at least this asymmetric from a birth and death process is therefore 64/1063, or about 0.06. This borderline result exemplifies the contention of Farris (1976) and Slowinski and Guyer (1989) that even a highly asymmetric distribution may be consistent with the null hypothesis.

To test the birth and death model at lower ranks in the classification, the following procedure is used to combine the results from individual nodes. The test is applied to nodes with at least four living languages, since there is only one possible distribution of two languages across two branches, and the two possible distributions of three languages across two branches are equally asymmetric. The asymmetry of the distribution for an individual node is measured by the coefficient of variation (C.V.), defined as the standard deviation divided by the mean; for n languages of which k are on the smaller branch, the C.V. is equal to $1 - 2k/n$. The obvious summary statistic for an entire set of nodes is the mean C.V. To test the model, the observed mean C.V. is compared to the expected distribution of the mean C.V. for a set of nodes the same as the observed nodes except assumed to satisfy the null hypothesis. The expected distribution is approximated by the Monte Carlo method of directly generating a large number of samples from the distribution. In each Monte Carlo run, each real node is matched to a Monte Carlo node with the same number of languages (n), which is then split into two branches according to the null hypothesis by sampling at random an integer from 1 to $n - 1$ inclusive; the numbers of languages on the branches are equal to the sampled integer and the difference between n and the sampled integer, respectively. The mean C.V. for all the Monte Carlo nodes is calculated in each run. A distribution of mean C.V. is formed using 1000 such Monte Carlo runs, and the 2.5 and 97.5 percentiles of the distribution are found. If the observed mean C.V. falls outside these limits, the null hypothesis is rejected at the 5% level. A mean C.V. above the upper limit indicates significantly more asymmetry than predicted by the model; a mean C.V. below the lower limit indicates significantly more symmetry than predicted.

Figure 1 presents the results of this test at each rank in the classification that contains nodes with two branches and at least four languages. The dotted lines give the values of the mean C.V. beyond which the null hypothesis is rejected. They are closest together at intermediate ranks, which contain the largest numbers of nodes. The solid line gives the observed mean C.V. It is slightly below the upper rejection limit at Rank 2, confirming the previous test on Niger-Kordofanian, and slightly above the upper limit at Rank 3, indicating significant asymmetry at this rank. Below Rank 3, the mean C.V. is between the limits, as predicted by the birth and death model.

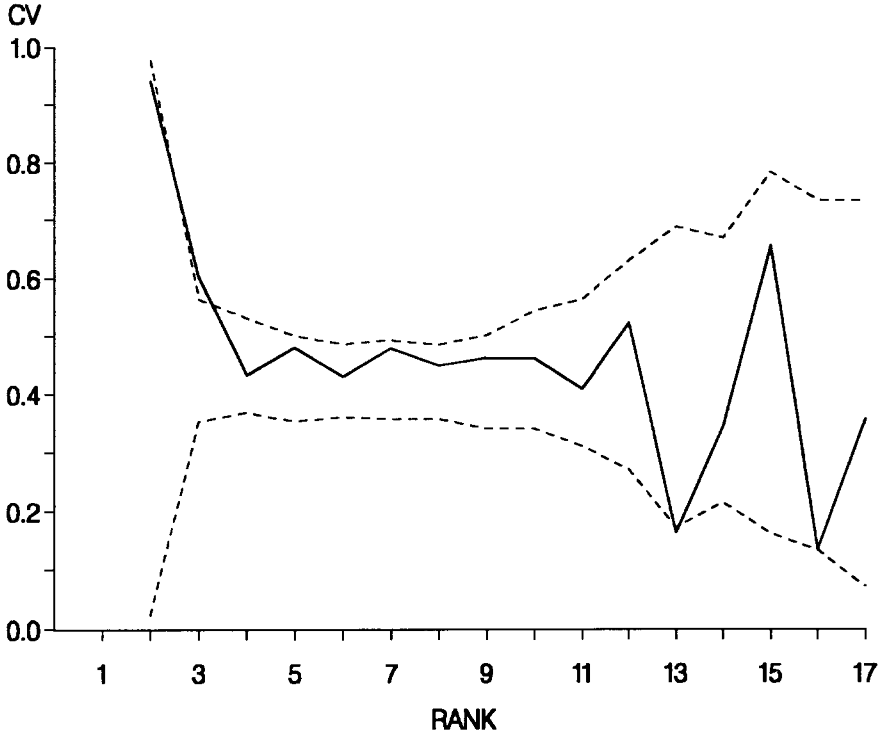

Figure 1. Solid lines give the mean C.V. for nodes with two branches, as a function of taxonomic rank; dotted lines give the limits for rejecting the null hypothesis.

An obvious limitation of these results is that they apply only to nodes with exactly two branches. In the entire classification, 46% of the nodes have two branches, 22% have three, and 32% have four or more. The test can be extended to nodes with more than two branches by a direct generalization of Farris's (1976) argument. If a single language at time 0 has any living descendants (which may include the language itself) at a given later time t, then as Farris pointed out, the number of those descendants is geometrically distributed. Specifically, let $P(k)$ be the probability of exactly k descendants, given at least one; then,

$$P(k) = (1 - B)B^{k-1} , \tag{1}$$

where B is a function of the time t and the rates of origination and extinction. Now suppose that each of m languages at time 0 has at least one living descendant at time t, and let $P(k_1,k_2, \ldots ,k_m)$ be the probability that the first language has k_1 descendants, the second has k_2 descendants ,..., and the last has k_m descendants. Let n be the total number of descendants, that is, the sum of k_1 through k_m. By the birth and death model, the descendants of each ancestral language are distributed according to (1), and the distributions are independent; it follows that

$$\begin{aligned} P(k_1,k_2,\ldots,k_m) &= (1-B)B^{k_1-1}\,(1-B)B^{k_2-1}\cdots(1-B)B^{k_m-1} \\ &= (1-B)^m\,B^{n-m}\,, \end{aligned}$$

which depends upon the number of ancestors and the total number of descendants, but not upon how the descendants are distributed among the ancestors. Moreover, the conditional probability that the ancestral languages have respectively $k_1,k_2,\ldots,k_m$ descendants, given that they have a total of n descendants, is equal to $P(k_1,k_2,\ldots,k_m)$ divided by the probability of having a total of n descendants, which also does not depend upon how the descendants are distributed among the ancestors. Thus, if a group of n living languages is divided into m branches descended from m contemporaneous common ancestors, then each possible distribution of numbers of languages among the branches is equally likely. This generalizes Farris's result from two to any number of branches. It is not necessary to assume that the m ancestors diverged at exactly the same time, but only that they all diverged from each other before any of them started splitting into descendants.

The generalization of the theoretical result implies a corresponding generalization of the statistical test. The test is applied to nodes with at least two more living languages than branches, since the distributions of fewer languages are equally asymmetric. The C.V. of the number of languages per branch can be defined for nodes with any number of branches, and so can the mean C.V. across nodes. The distribution of the mean C.V. under the null hypothesis can be approximated using the same Monte Carlo method as before, except that a Monte Carlo node with n languages is divided into m branches by sampling $m-1$ integers at random, without replacement, from 1 through $n-1$ inclusive; the sampled integers are ordered from smallest to largest, and the numbers of languages per branch are equal to the smallest integer, the differences between adjacent integers, and the difference between n and the largest integer.

Figures 2 and 3 present the results of the generalized test for nodes with three branches and nodes with four or more branches, respectively. The curves for nodes with three branches are similar in shape to the previous curves for nodes with two branches, although the rejection limits are farther apart because of the smaller numbers of nodes involved. The mean C.V. indicates significant asymmetry at Rank 2, even though there is only one phylum with three branches, Khoisan. Otherwise, the mean C.V. is within the limits except at Ranks 6 (slightly above) and 14 (below); there are no nodes with

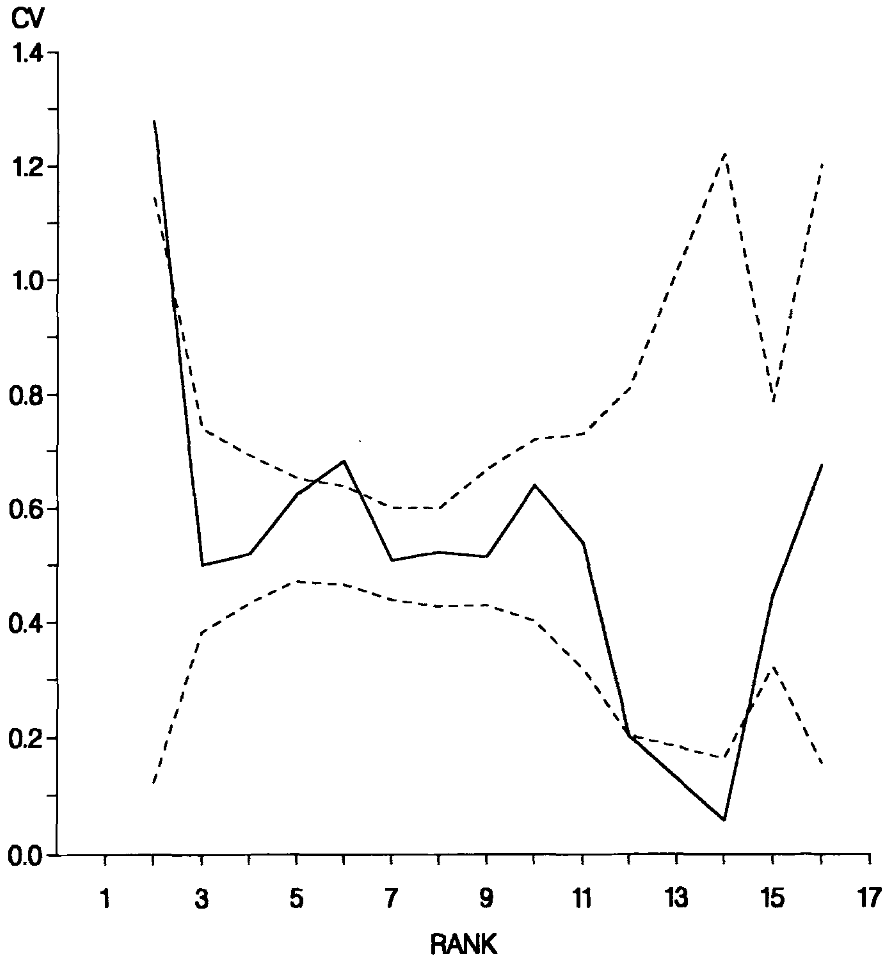

Figure 2. Solid line gives the mean C.V. for nodes with three branches, as a function of taxonomic rank; dotted lines give the limits for rejecting the null hypothesis.

three branches at Rank 13. For nodes with four or more branches, the shape of the curves is much the same as before, but the data are displaced upward relative to the predictions at most ranks because of asymmetric distributions. The mean C.V. is consistent with the null hypothesis for the root of the tree at Rank 1; it is far above the upper limit at Rank 2 and remains above the limit through Rank 9, after which it fluctuates across both limits.

Putting all the tests together, the clearest pattern is a dependence of the results upon the number of branches per node: the null hypothesis is much more successful at nodes with two or three branches than at nodes with four

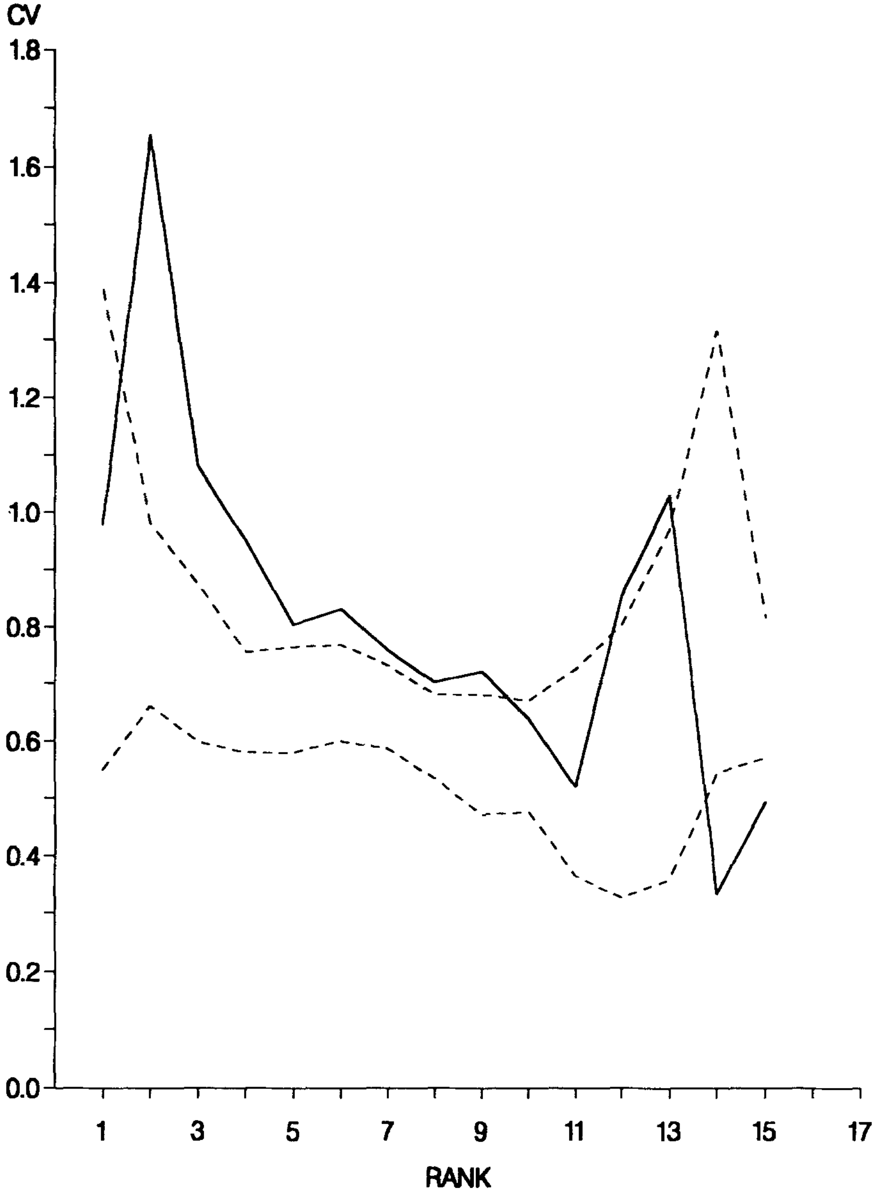

Figure 3. Solid line gives the mean C.V. for nodes with four or more branches, as a function of taxonomic rank; dotted lines give the limits for rejecting the null hypothesis.

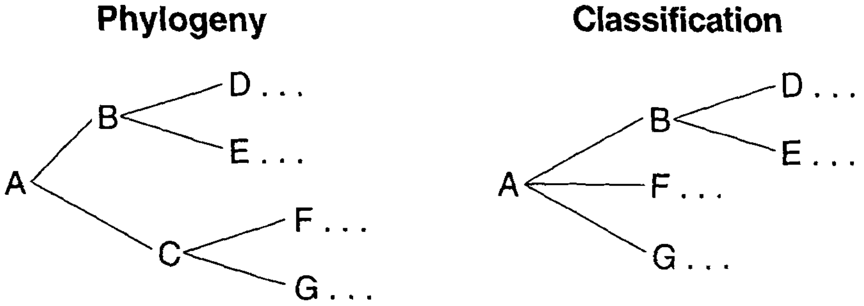

Figure 4. True phylogeny (left panel) and a possible classification (right panel) of the same set of languages. Letters indicate languages; dots indicate possible descendants.

or more branches. This pattern of results may mean that the test is more powerful at nodes with more branches; but to the extent that the power of the test is related to the width of the interval between the rejection limits, differences in power are not prominent. The observed results may also mean that the inferred phylogenetic tree really is more asymmetric at nodes with more branches. A node may have relatively many branches either because the ancestral language split into several descendants during a relatively short time interval, or because the exact order of the branching events is still unknown. The first alternative does not explain the data. There is no good reason why four or more languages that diverge at about the same time would produce significantly different numbers of descendants, while two or three such languages would produce more nearly equal numbers of descendants. The second alternative is more plausible, as illustrated in Figure 4. The true phylogenetic tree here is binary: language A splits into B and C, then B splits, followed by C. Since the order of the latter two events is not necessarily inferable by cladistic methods, however, the classification may merge node C with A to produce a node with three branches. The test would then incorrectly assume that B, F, and G were all separate languages at the same time, and the observed tree would probably be too asymmetric because actually B had already split before F and G existed as separate languages. Thus, the asymmetry of the tree tends to be increased by the merger of nodes that are adjacent in the phylogeny but not adjacent in time. There is only one way (shown in Figure 4) that a merger of this sort can produce a node with three branches, and no way that a merger alone of any sort can produce a node with two branches; but the number of possible mergers increases rapidly as the number of branches per node increases to four and above. This increase can explain the observed asymmetry of the tree at nodes with four or more branches.

Another weaker pattern in the results is their dependence upon taxonomic rank: the null hypothesis is less successful at Ranks 2 and 3 than at other ranks. At Rank 1, the success of the null hypothesis may be coincidental because there is no reason to assume that the ancestors of all the language phyla were individual languages simultaneously. At Rank 2, the mean C.V. is higher than at other ranks, and is significantly different from the null hypothesis with the borderline exception of the one phylum with two branches. At Rank 3, the null hypothesis is rejected at nodes with two branches as well as at nodes with four or more branches, but not at nodes with three branches. The power of the test seems, from the width of the interval between the rejection limits, to be if anything lower at these ranks than at intermediate ranks that contain more nodes. Since higher ranks correspond to earlier times, it is possible that there were more extreme historical events at earlier times to produce the more asymmetric branches observed at the higher ranks; but this hypothesis is hard to test because the relevant times predate recorded history. Two taxonomic explanations are also possible. First is the greater difficulty of correctly inferring the remote relationships represented at higher ranks. This difficulty is mentioned by Ruhlen (1991, p. 380) and corroborated by the large number of branches per node at Ranks 1 and 2. More mistakes in the order of events, as in Figure 4, could explain the greater asymmetry at high ranks for nodes with three or more branches; but the asymmetry at nodes with two branches would require an additional type of error, such as the presence in the classification of nodes that do not occur in the phylogeny. The other taxonomic explanation is variability in the threshold for distinguishing languages from dialects. Differences in threshold among groups at Rank 4 and above would inflate the number of languages in some groups and deflate the number in others, thus increasing the asymmetry at nodes above Rank 4. This pattern of variation could occur if it is harder to maintain a consistent threshold across groups of distantly related languages than across groups of closely related languages, or if specialists working on different groups of languages use thresholds that are consistent for each specialist but different between specialists.

The asymmetry of the inferred phylogenetic tree is consistent with the birth and death model at nodes below Rank 3 with fewer than four branches. Since these are the nodes most likely to be valid, the present results are similar to the findings of Guyer and Slowinski (1991) that well- corroborated portions of phylogenetic trees of animals and plants are consistent with the model. The observed effect of taxonomic rank is also similar to Savage's (1983) report of less asymmetry in cladistic trees than in phenetic trees, given that cladistic methods tend to be replaced by phenetic methods at the highest ranks of linguistic classification.

4. Evolutionary Rates and the Historical Record

The obvious next step is to estimate the rates of origination and extinction of languages. Evolutionary rates cannot be inferred from purely taxonomic data, however, because the classification describes the current distribution of languages but not how long it took to reach its present state. Even the ratio of the extinction rate to the origination rate, although independent of the time scale, is not reflected in the classification; and the ratio cannot be estimated from the proportion of classified languages that are extinct, because the sample of extinct languages is much less complete than the sample of living ones. To answer questions such as these, further data must be sought in the historical record.

The birth and death model makes many predictions about the branching process of language evolution as a function of historical time and the rates or origination and extinction. These predictions can be compared to the historical record to estimate the evolutionary rates; deviations from the predictions would indicate fluctuations in evolutionary rates. The main limitation on such comparisons is the incompleteness of the data. Most languages have no recorded history at all, and most historically attested languages have no records of their origination or extinction. Moreover, the historically recorded languages are a highly selected sample, biased toward languages whose speakers were numerous or powerful; such languages are expected to have atypically low extinction rates. Under these conditions, the most representative estimates of evolutionary rates will be obtained from predictions that require the least amount of historical data about individual languages and thus permit the largest and least biased sample of languages.

To state one set of predictions, let λ and μ be the rates of origination and extinction, respectively. For a single language alive at time 0, let $Q(k,t)$ be the probability that the language has k living descendants at time t. Bailey (1964, p. 94) shows that according to the birth and death model,

$$Q(k,t) = (1-A)(1-B)B^{k-1}, \quad \text{for } k \geq 1 , \tag{2}$$

$$Q(0,t) = A, \quad \text{for } k = 0 ;$$

$$\text{where } A = \frac{\mu(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu} , \quad B = \frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu} .$$

The quantity B is exactly the same in (2) as in (1). In the first line of (2), the function $Q(k,t)$ differs from $P(k)$ in (1) in that $Q(k,t)$ is an unconditional probability whereas $P(k)$ is conditional upon $k \geq 1$; also, the dependence upon time is made explicit in $Q(k,t)$ because time is important in historical data. In the second line of (2), $Q(0,t)$ gives the probability that the language has become extinct without living descendants at time t; the equation shows that this probability increases monotonically with increasing t to approach an asymptote of 1 if $\lambda \leq \mu$ or μ/λ if $\lambda > \mu$. It is worth emphasizing that the time 0 in these equations does not have to be the time at which the language originated. Instead, the clock can be started at any point in the lifetime of the language and the predicted results will be the same because the rates of origination and extinction remain constant over time according to the model.

To compare the predictions in (2) with historical data, the time points 0 and t were chosen in such a way as to maximize the number of languages in the sample: 0 was the date of the language's first appearance in the record (not the language's origination), and t was the present. The date of first appearance was defined as either the date on which the language was first reduced to writing, or the date of the earliest surviving material written in the language, whichever is known more exactly. Glosses and word lists, as well as connected texts in the language, were accepted as written material, but proper names were not. The two exceptions to this definition were Sanskrit and Avestan, whose written records are later than their extinction as spoken languages; their dates of first appearance were defined instead as the dates of composition of the Vedas and the Gatha, respectively. For most historically attested languages, the date of first appearance can be determined to the nearest century or better, a degree of accuracy that cannot be approached by other time points such as the date of origination. Indeed, the main ambiguity for some languages is whether they were distinct languages rather than dialects of the same language at their times of first appearance. At the other end of the trajectory, the time point t was chosen to be the present because the proportion of known languages is much higher for languages living at the present than at any earlier time. The number of living descendants of each language in the sample was determined from Ruhlen's classification. The main ambiguity here is whether the living languages are direct descendants as opposed to more distant relatives of a given earlier language.

The Appendix presents the date of first appearance, the number of living descendants, and the name of these descendants as grouped in Ruhlen's classification, for each language that satisfies the following four requirements. First, the language is listed in Ruhlen's classification. Second, the historical records of the language include more than proper names. Third, the language is not a descendant of another language that entered the historical record earlier and is therefore already listed in the Appendix. This requirement is imposed because the unit of analysis is the language plus all its descendants, not just the individual language. Fourth, the date of first appearance is before 1491 (500 years before the publication of Ruhlen's classification, considered to be the present). This requirement is imposed because earlier dates are more informative about evolutionary rates than are later dates, and also because first appearances before 1491 are more likely to involve materials written by native speakers of the language rather than by foreigners such as colonists or missionaries.

The languages in the Appendix are subject to the same questions about the distinction between languages and dialects and the validity of taxonomic groups that were previously described for languages in general. In some respects, the problems are less severe because the written record provides information that is not available for most other languages. For instance, there is little controversy about the phylogenetic reality of most of the groups in the Appendix. In other respects, however, the problems are exacerbated by the introduction of the time dimension. The distinction between languages and dialects must be addressed at the time of a language's first appearance in the record as well as at the present time; and the problem of grouping is complicated by the question of whether the relation between old and young languages within a group is less direct than ancestor to descendant. These ambiguities can have sizeable effects on some entries in the Appendix. For instance, if the Slavic languages are considered to have been mutually comprehensible dialects in the ninth century, then Old Church Slavonic has 13 descendants rather than four, and the separate entries for Russian, Czech, and Polish must be deleted. Fortunately, because of the relatively small number of languages in the Appendix, it is practical to estimate evolutionary rates and test the birth and death model under alternative assumptions about the more ambiguous ancestor-descendant relationships.

An even more obvious problem with the Appendix is its incompleteness and selectivity. As a quantitative estimate of incompleteness, the living descendants in the Appendix include only 234 of the 4741 living languages in Ruhlen's classification; in other words, fewer than 5% of known languages can be traced back as far as 500 years in the written record. The problem is compounded by selectivity; historically recorded languages are not a random sample, but rather are selected according to the sorts of cultural and demographic factors that are also expected to influence the birth, survival, and death of languages. This problem is less amenable to statistical solutions than the others described previously, and must therefore be addressed in the interpretation of the statistical results.

The asymptote of $Q(0,t)$ can be inferred from the fate of the oldest languages in the Appendix. In fact, the eight oldest languages have no living descendants. It follows that $Q(0,t)$ must approach an asymptote near 1. If the asymptote is as low as 0.69, then the probability of no descendants from the eight oldest languages is less than $(0.69)^8$, which is about 0.05. Thus, a 95% confidence interval for the asymptote ranges from 0.69 to 1. Moreover, the known biases in the sample of recorded languages would tend if anything to increase the probability of living descendants. An asymptote of $Q(0,t)$ near 1 must therefore apply to languages in general, not just recorded languages. It follows from (2) that the origination rate of languages is unlikely to be much higher than the extinction rate.

The solid line in Figure 5 is an estimate of $Q(0,t)$ based on the data in the Appendix. Time since first appearance was divided into 500-year intervals up to 2000 years, then a 1000-year interval up to 3000 years, and then a final interval of all longer times. The horizontal axis gives the mean time since first appearance of the languages in each interval, and the vertical axis gives the percentage of the languages that are extinct without living descendants. The dotted line in the figure is the prediction of $Q(0,t)$ by the birth and death model. To make the prediction, maximum likelihood estimates of the rates λ and μ were obtained from the data in the Appendix using a grid search that directly maximized the likelihood function implied by (2). The estimated rates turned out to be 1.18 and 0.60 per millennium for λ and μ respectively. The ratio of these rates implies an asymptote of 0.51 for $Q(0,t)$, which is below the confidence interval obtained earlier. Moreover, the shape of the predicted curve is inconsistent with the data: the predicted $Q(0,t)$ is too high at first and too low later. To test the statistical significance of the difference in shape, the languages in the Appendix were divided at the median date of first appearance (430 AD) into older and younger languages; the predicted numbers of older and younger languages with and without living descendants were then calculated from (2) with λ and μ set equal to 1.18 and 0.60 per millennium. Table 1 presents the observed and predicted numbers; the Pearson $X^2(2)$ is 11.16, which indicates a significant discrepancy.

The same analyses were also conducted on two alternative versions of the data, which are based on different criteria for ancestor-descendant relationships. A lenient criterion produced the following changes in the Appendix. Runic is the ancestor of West Germanic as well as North Germanic (Penzl 1972) and therefore has 12 descendants; German, English, and Frisian are eliminated as separate entries; Old Church Slavonic is the ancestor of Slavic and has 13 descendants; Russian, Czech, and Polish are eliminated; Welsh is the ancestor of Cornish, which is eliminated. In this case, λ and μ are 1.29 and 0.67 per millennium and $X^2(2)$ is 11.84, still a significant discrepancy. A strict criterion produced the following changes. Old Libyan is not the ancestor of Berber and therefore has no descendants; Epigraphic South Arabian is not the ancestor of Ethiopic and has only one descendant (South Arabian); Geez first appears c. 300 AD and has two descendants (North Ethiopic); Amharic first appears c. 1350 AD and has one descendant (Amharic); Sogdian is not the ancestor of Yaghnobi; Ju-chen is not the ancestor of Manchu. In this case, λ and μ are 1.10 and 0.64 per millennium, and $X^2(2)$ is 10.68, again significant. Although other versions of the data are obviously possible, it seems unlikely that any plausible alternative would be consistent with the birth and death model.

Table 1

Observed (Predicted) Numbers of Older and Younger Languages with and without Living Descendants.

	Date of First Appearance	
	Before 430 AD	After 430 AD
With Living Descendants	16 (22.88)	35 (27.59)
Without Living Descendants	25 (18.12)	5 (12.41)

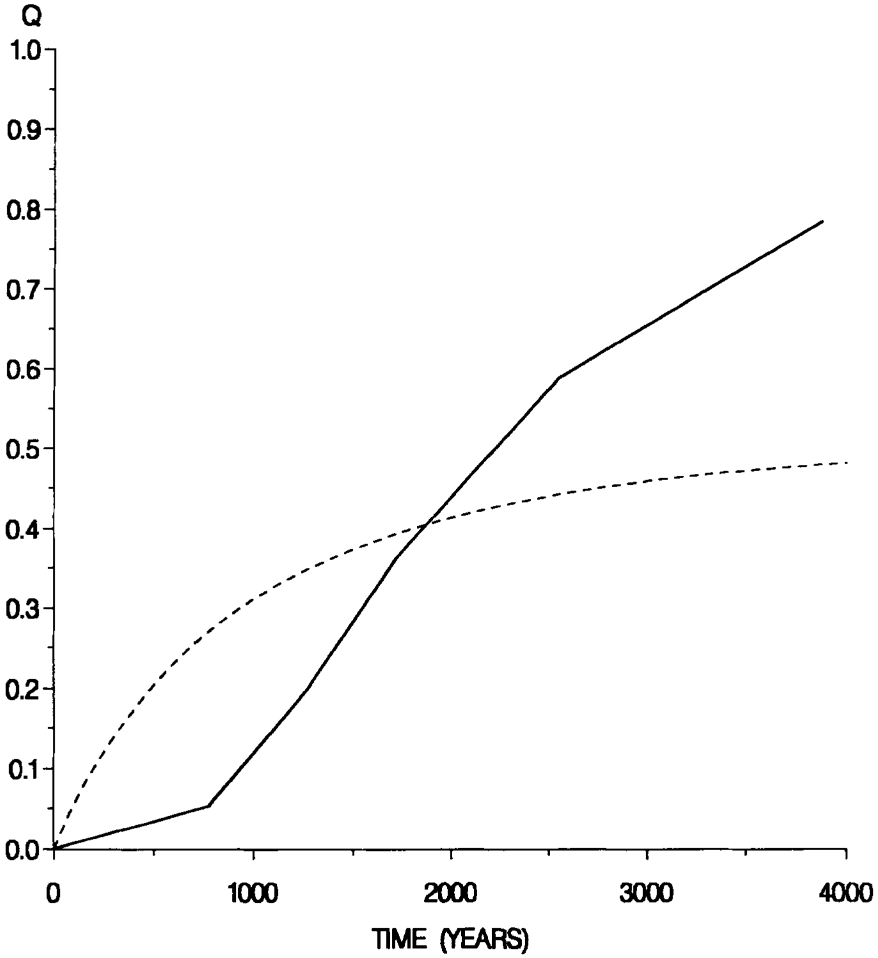

Figure 5. Predicted (dotted line) and observed (solid line) percentage of languages extinct without living descendants, as a function of time since first appearance in the record.

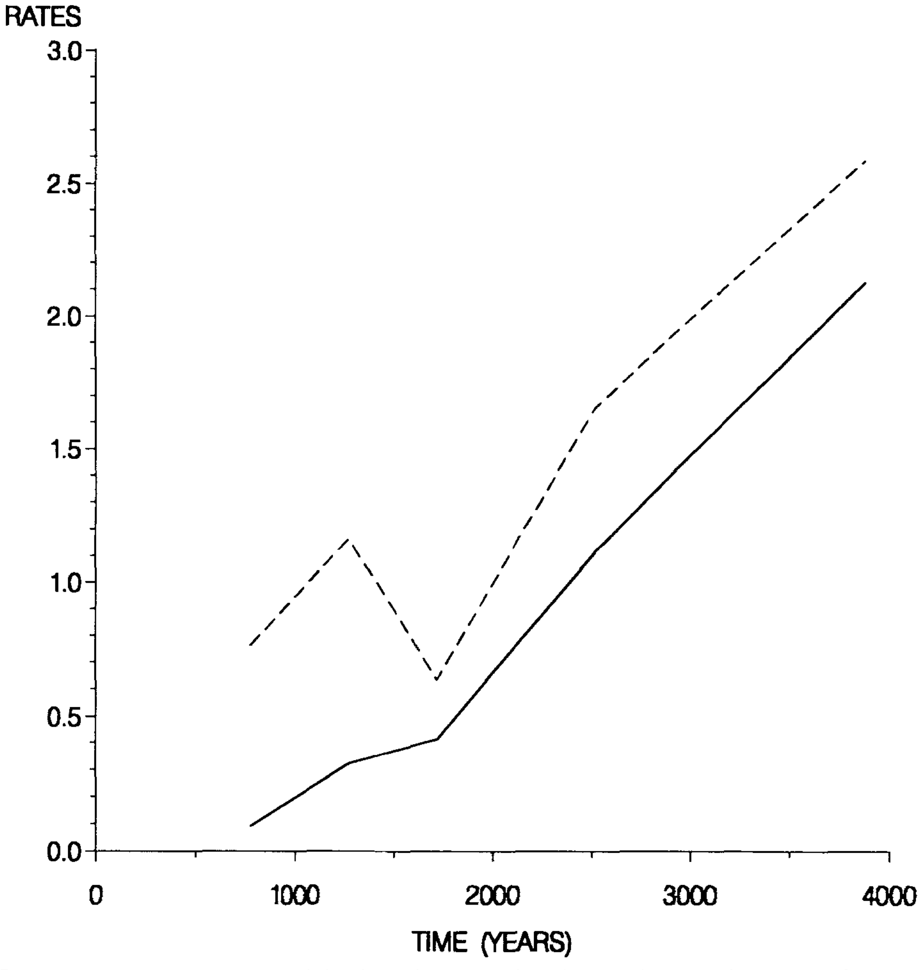

Figure 6. Estimated rates of origination (dashed line) and extinction (solid line) as a function of time since first appearance in the record.

The deviations in the shape and asymptote of the predicted curve imply that contrary to the assumptions of the birth and death model, the evolutionary rates of recorded languages must vary as a function of the time since the languages' initial appearance in the record. To study this variation, time since first appearance was divided into the same intervals as in Figure 5, and separate maximum-likelihood estimates of λ and μ were obtained for the languages in each interval. Figure 6 presents the results. The horizontal axis gives the mean time since first appearance, and the vertical axis gives the estimated evolutionary rates. Both λ (dashed line) and μ (solid line) increase steeply with time since first appearance, contradicting the constancy assumed by the birth and death model. Moreover, the observed increase in rates explains the discrepancies in Figure 5. For the first 1500 years or so after a language enters the record, its rate of extinction is relatively low in Figure 6 and so is its probability of extinction in Figure 5, but then an increasing extinction rate produces a rapidly increasing extinction probability. As both λ and μ increase, so does the ratio μ/λ, which determines the asymptotic probability of extinction; the ratio for the oldest languages is 0.82, well within the confidence interval for the asymptote. The evolutionary rates shown in Figure 6 are probably not representative of languages in general, even during historical times. The discrepancy is most obvious for the youngest group of languages in the figure, whose records cover the last 500 to 1000 years. The estimated extinction rate of these languages is close to 0, but the extinction rate of languages in general during the same time period was probably above average if anything, as many Amerind and Australian languages were extinguished by European colonization toward the end of the period. The results in Figure 6 are better explained by the hypothesis that the sample of historically recorded languages is biased toward low rates of origination and extinction. Suppose that the languages in an ancestor-descendant lineage fluctuate over time in their rates of origination and extinction. If (for any reason) a language is more likely to enter the record when its evolutionary rates are lower than average, then the purely statistical phenomenon of regression to the mean implies that the evolutionary rates of the language and its descendants will tend to revert toward average values thereafter. For instance, if languages tend to enter the record when they have many speakers which in turn causes low evolutionary rates, then eventually the number of speakers will return to average and so will the evolutionary rates. Regression to the mean can in fact be more directly observed for other characteristics of recorded languages besides evolutionary rates: for example, many descendants of old written languages such as Sanskrit are no longer widespread literary languages. Since older languages (and their descendants) have more time to return to average rates than do younger languages, it follows that evolutionary rates estimated from older languages are closer to average than rates estimated from younger languages. Thus, the increasing functions in Figure 6 represent this return of the estimated rates toward the average rates for languages in general.

The atypical evolutionary rates of historically recorded languages do not account for the significant asymmetry at high ranks in Ruhlen's classification. At Rank 2, the one phylum with three branches (Khoisan) is significantly asymmetric in its distribution, and the one phylum with two

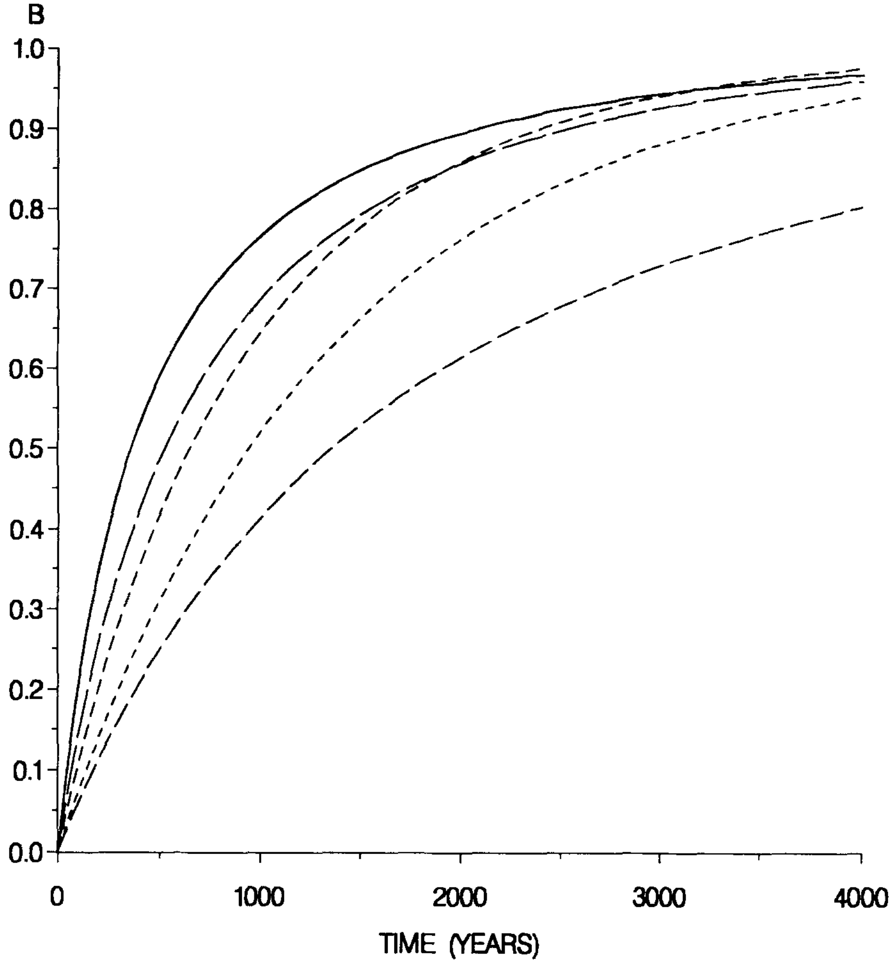

Figure 7. *B* as a function of *t* in (2), for rates of origination and extinction estimated at various times since first appearance in the record; the curve with the shortest dashes refers to the shortest time since first appearance, and the curves with successively longer dashes refer to longer times.

branches (Niger-Kordofanian) is nearly as asymmetric; but neither phylum contains any of the recorded languages in the Appendix. At Rank 3, the 24 nodes with two branches are significantly asymmetric; but the mean C.V. is not significantly greater in the six groups that contain languages in the Appendix than in the 18 other groups, $t(22) = 1.40$.

This null result can be explained by noting that the distribution of living languages is determined by the quantity B in (1) and (2); B in turn depends upon the magnitude of λ and μ and also strongly upon their difference $\lambda - \mu$, which appears in the exponential terms of (2). In Figure 6, λ and μ both increase with time since first appearance, but their difference decreases if anything; these trends may therefore tend to cancel each other with little net effect on B. To test this explanation, Figure 7 shows B as a function of t for each of the pairs of rates in Figure 6; the curve with the shortest dashes refers to the youngest set of languages, and curves with successively longer dashes refer to successively older sets. The curves are fairly similar to each other; and the most aberrant curve refers to the middle set of languages, suggesting that the differences among the curves are largely random variation. Thus, the changes in evolutionary rates shown in Figure 6 have little effect on the distribution of living languages according to Figure 7, despite their strong effect on the probability of extinction in Figure 5.

5. Discussion

The present results offer some methodological information about linguistic taxonomy. Except at the top two ranks, Ruhlen's classification has fewer branches per node than the noncladistic classifications studied by Holman (1992), consistent with the aim of cladistic methods to maximize resolution. The excessive asymmetry of the tree at nodes with four or more branches, however, suggests that a substantial proportion of such nodes reflect the merger of nodes that are actually distinct in the phylogeny, rather than the rapid divergence of several languages from the same ancestor. Thus, the resolution of the classification is still limited not only by the structure of the underlying phylogeny, but also by the lack of sufficient data to infer the phylogeny completely.

Two lines of evidence suggests that limitations on the data are particularly severe at the top few ranks of the classification. One is the resolution of the classification: although a large number of branches at Rank 1 is typical of classifications in general, the relatively large number of branches per node at Rank 2 is more specific to languages. The other line of evidence is the relatively high asymmetry of the tree at Ranks 2 and 3, even at nodes with high resolution. These results can be compared to the consensus of linguists about the limits of their data. Nichols (1990, p. 477) defines a stock as 'the oldest grouping reachable by application of the standard comparative method'; she gives a (nonexhaustive) list of 100 stocks, all but eight of which are also included in Ruhlen's classification. The mean rank of the stocks in the classification is 6.00 for the 65 stocks in the Amerind phylum, and 3.41 for the 27 stocks not in Amerind. Except in Amerind, therefore, the limit of the comparative method is just below the ranks at which the classification shows less resolution and more asymmetry. The limit in Amerind, however, is significantly different, $z = 10.83$, suggesting unusual evolution or taxonomy in this group.

The distinction between languages and dialects at the bottom of the classification seems to involve fewer problems than the construction of groups at the top. Two aspects of the data indicate that linguists can maintain a consistent distinction between languages and dialects at least within groups at or below the fourth rank in the classification. First, nodes at the lowest ranks, where the distinction between languages and dialects is most important, have no more branches than nodes at higher ranks. Second, nodes with two or three branches at Rank 4 and below show no more asymmetry than is predicted by the birth and death model, which assumes a consistent distinction between languages and dialects. At higher ranks, however, the increase in asymmetry may indicate variation among groups in the threshold between languages and dialects.

The data from the historical sample cast some light on the biases in the historical record. The bias toward a low extinction rate is consistent with the relative prominence of most written languages. The bias toward a low origination rate is consistent with the evidence of Bergsland and Vogt (1962) and Diebold (1964) that the replacement rate of words in written languages is lower than in unwritten languages, since dialects must diverge in vocabulary as they become separate languages. The type of fluctuation in rates that can be inferred from the historical data turns out to have little effect on the asymmetry of the phylogenetic tree. It is therefore impossible to determine without further data whether similar fluctuations also characterize the evolution of the vast majority of languages that have left no historical record.

The historical biases imply that most languages evolve at higher average rates than might be inferred from historical calibrations. How much higher is hard to estimate from the available data, since the functions in Figure 6 show no sign of approaching an asymptote before the historical record runs out. The last points in the functions, for the oldest recorded languages, are best interpreted as estimated lower bounds on the average rates for unrecorded languages. Since the reciprocal of a rate is equal to the average time between events, the estimated bounds on rates correspond to average times of about 400 years between originations and 500 years from origination to extinction, and therefore the average times for unrecorded languages are if anything shorter. These times are far below the lifespans of most historically recorded languages, but they nevertheless exceed the few generations in which creole languages have been known to evolve. With respect to rates of evolution, the seemingly atypical creole languages may actually be more representative of languages in general than are the more familiar languages in the historical record.

The inferred fluctuations in the evolutionary rates of recorded languages may be related to historical phases of growth and decline. If languages tend to enter the record when the political or ethnic groups of their speakers are relatively prominent, then the subsequent regression to average rates corresponds to a phase of decline. Although regression to the mean is statistically inevitable, its time scale is not. The observed time scale of a few thousand years (at least) exceeds the known longevity of nearly all political groups, while the longevity of ethnic groups is harder to estimate. Further investigation of extremely long-term historical processes may be warranted.

Although the average rates of origination and extinction are hard to estimate, their relative magnitude can be inferred with more confidence. On the one hand, the extinction of the eight oldest recorded languages implies a probability near 1 that a language will eventually become extinct without living descendants. Consequently, the average origination rate cannot be much higher than the average extinction rate. On the other hand, the number of currently living languages is undoubtedly larger than the number of primordial languages, and thus the number of living languages must have increased over time on the average. Consequently, the average origination rate cannot be any lower than the average extinction rate. These two conclusions together imply that the average rates are similar in magnitude, with the origination rate slightly higher.

The present results allow some new comparisons between linguistic and organic evolution. The clearest similarity appears in the patterns of asymmetry in phylogenetic classifications of living languages and species. In Ruhlen's linguistic classification, as in the biological classifications studied by Guyer and Slowinski (1991), the portions of phylogenetic trees with the strongest empirical support are consistent with the birth and death model, but less securely established portions of trees tend to be more asymmetric than predicted. Moreover, as in the biological classifications studied by Savage (1983), cladistic methods are associated with less asymmetry than are phenetic methods. These patterns suggest a similarity not only in the underlying evolutionary process but also in the superimposed taxonomic biases.

Comparison of estimated evolutionary rates yields expectedly huge differences in absolute values, with average lifetimes measured in centuries for languages according to the present analysis, and in millions of years for species according to Raup (1978). The relation between the rates of origination and extinction turns out to be very similar, however, with the origination rate slightly higher than the extinction rate for both languages and species. Equilibrium models such as Sepkoski's (1984) attribute the near equality of rates for species to an increase in competition as species become more numerous. A type of competition between languages can be inferred from the fact that most people speak only one language, but quantitative models for such competition have not been worked out.

A final point of similarity between languages and species is the fact that the historical and fossil records reveal variations in evolutionary rates not apparent in classifications of living languages and organisms. Beyond this point, however, linguistics diverges from paleontology in the nature of the observed variations in rates. Languages show temporary decreases in evolutionary rates correlated with their appearance in the historical record. In contrast, the fossil record shows temporary increases in extinction rates for most species during certain time intervals (Hubbard and Gilinsky 1992), and also some differences in evolutionary rates among taxonomic groups of organisms (Holman 1989). The differences between languages and fossils may be related to differences in the unit of analysis for which suitable data are available. The unit in the present analysis of linguistic data was the individual language and all its descendants, whereas the usual unit in paleontology is the genus or family. This difference is in turn a consequence of the fact that phylogenetic classifications are the norm in linguistics but much less common in biology, where traditional classifications are based on taxonomic ranks rather than phylogenetic trees. It is still unknown why the taxonomic differences arose in the first place.

6. Appendix

Language	Date of First Appearance	Living Descendants	Reference
Sumerian	c. 3100 BC	0	Coulmas (1989, p. 73)
Ancient Egyptian	c. 3000 BC	0	Coulmas (1989, p. 60)
Eblaic	c. 2500 BC	0	Pettinato (1981, p. 12)
Akkadian	c. 2350 BC	0	Coulmas (1989, p. 80)
Hurrian	c. 2300 BC	0	Houwink ten Cate (1986)
Elamite	c. 2250 BC	0	McAlpin (1981, pp. 59-60)
Hittite	c. 1700 BC	0	Houwink ten Cate (1986)
Palaic	c. 1700 BC	0	Houwink ten Cate (1986)
Classical Greek	c. 1400 BC	2: GREEK	Diringer (1968, pp. 116-119)
Luwian	c. 1400 BC	0	Houwink ten Cate (1986)
Ugaritic	c. 1375 BC	0	Gordon (1965, p. 1)
Archaic Chinese	c. 1200 BC	8: CHINESE	Keightley (1978, pp. 175,228)
Sanskrit	c. 1100 BC	48: INDIC	Emeneau (1966)
Phoenician	c. 1020 BC	0	Gibson (1982, p. 12)
Classical Hebrew	c. 950 BC	1: Hebrew	Gibson (1971, p. 1)
Old Aramaic	c. 860 BC	2: ARAMAIC	Gibson (1975, p. 1)
Moabite	c. 850 BC	0	Gibson (1971, p. 71)
Avestan	c. 750 BC	0	Vendryes and Benveniste (1952, p. 27)
Epigraphic South Arabian	c. 750 BC	13: SOUTH SEMITIC	Van Beek (1961)
Etruscan	c. 690 BC	0	Morandi (1982, p. 27)

Latin	c. 625 BC	16: ROMANCE	Morandi (1982, p. 47)
Faliscan	c. 610 BC	0	Morandi (1982, p. 54)
Lydian	c. 600 BC	0	Houwink ten Cate (1986)
Sabellian	c. 600 BC	0	Morandi (1982, p. 63)
Old Persian	c. 520 BC	2: PERSIAN	Diringer (1968, p. 138)
Oscan	c. 390 BC	0	Morandi (1982, p. 115)
Umbrian	c. 390 BC	0	Morandi (1982, p. 73)
Gaulish	c. 250 BC	0	Vendryes and Benveniste (1952, p. 53)
Tamil	c. 250 BC	3: TAMIL-IRULA	McAlpin (1981, p. 18)
Meroitic	c. 170 BC	0	Shinnie (1967, pp. 58,134)
Old Libyan	149 BC	30: BERBER	Rössler (1958)
Parthian	c. 20 AD	0	Oranskij (1977, p. 102)
Bactrian	c. 120 AD	0	MacDowell and Taddei (1978)
Khwarezmian	c. 150 AD	0	MacKenzie (1969)
Sogdian	196 AD	1: Yaghnobi	Harmatta (1979)
Runic	c. 200 AD	5: NORTH GERMANIC	Antonsen (1975, p. 29)
Classical Arabic	328 AD	3: ARABIC	Diringer (1968, p. 211)
Gothic	c. 350 AD	0	Jensen (1970, p. 484)
Cham	c. 360 AD	3: CHAM-CHRU	Coedès (1964, pp. 94-96)
Classical Armenian	406 AD	1: Armenian	Diringer (1968, p. 250)
Georgian	c. 410 AD	1: Georgian	Jensen (1970, p. 445)
Kannada	c. 450 AD	2: KANNADA	McAlpin (1981, p. 18)
Tocharian A	c. 500 AD	0	Lane (1986)
Tocharian B	c. 500 AD	0	Lane (1986)
Mon	c. 600 AD	2: MONIC	Coedès (1952)
Khmer	629 AD	1: Khmer	Diringer (1968, p. 317)
Telugu	633 AD	2: TELUGU	McAlpin (1981, p. 18)
Classical Tibetan	639 AD	2: TIBETAN	Diringer (1968, p. 278)
Saka	c. 650 AD	0	MacKenzie (1969)
Malay	682 AD	3: LOCAL MALAY	Coedès (1964, p. 156)
German	c. 700 AD	5: CONTINENTAL WEST GERMANIC	Salzer and Von Tunk (1972, p. 23)
Japanese	712 AD	2: JAPANESE-RYUKYUAN	Miller (1967, p. 32)
Old Turkic	716 AD	30: COMMON TURKIC	Poppe (1965, p. 60)
English	737 AD	1: English	Wrenn (1967, p. 96)
Irish	c. 750 AD	2: GOIDELIC	Thurneyson (1946, pp. 4-9)
Welsh	c 750 AD	1: Welsh	Jackson (1953, pp. 42-45)
Javanese	c. 778 AD	1: Javanese	Diringer (1968, p. 334)
Breton	c. 850 AD	1: Breton	Jackson (1953, p. 62)
Old Church Slavonic	863 AD	4: SOUTH SLAVIC	Jensen (1970, p. 490)
Bali	896 AD	1: Bali	Coedès (1964, p. 239)
Cornish	c. 900 AD	0	Jackson (1953, p. 59)
Frisian	c. 900 AD	1: Frisian	Markey (1976, pp. 53-54)
Czech	1057 AD	2: SOUTH WEST SLAVIC	Schamschula (1990, p. 32)
Burmese	c. 1084 AD	1: Burmese	Jensen (1970, p. 389)
Kurdi	c. 1100 AD	1: Kurdi	Oranskij (1977, pp. 140,181)
Korean	1104 AD	1: Korean	Poppe (1965, p. 76)
Russian	c. 1113 AD	3: EAST SLAVIC	Terras (1991, pp. 36-39)
Ju-chen	1119 AD	1: Manchu	Kane (1989, p. 9)
Basque	c. 1150 AD	1: Basque	Michelena (1964, pp. 49-51)
Hungarian	c. 1150 AD	1: Hungarian	Nemeskürty (1983)
Classical Mongolian	1204 AD	12: MONGOLIAN	Weiers (1967)

Estonian	c.1220 AD	1: Estonian	Sauvageot (1952, p. 282)
Moso	c.1225 AD	2: MOSO	Prunner (1967)
Lao	c.1250 AD	1: Lao	Diringer (1968, p. 325)
Polish	1270 AD	2: NORTH WEST SLAVIC	Krzyzanowski (1978, p. 13)
Thai	1283 AD	1: Thai	Coedès (1964, p. 360)
Old Prussian	c.1300 AD	0	Schmalsteig (1976, p. 68)
Sundanese	1333 AD	1: Sundanese	Coedès (1964, p. 430)
Vietnamese	1343 AD	1: Vietnamese	Coulmas (1989, p. 114)
Komi	c.1375 AD	1: Komi	Sauvageot (1952, p. 284)
Albanian	1462 AD	1: Albanian	Vendryes and Benveniste (1952, p. 45)

References

ANTONSEN, E. H. (1975), *A Concise Grammar of the Older Runic Inscriptions*, Tübingen: Max Niemeyer Verlag.

ANDERSON, S. (1974), "Patterns of Faunal Evolution," *Quarterly Review of Biology, 49*, 311-332.

BAILEY, N. T. J. (1964). *The Elements of Stochastic Processes, with Applications to the Natural Sciences*, New York: Wiley.

BATEMAN, R., GODDARD, I., O'GRADY, R., FUNK, V. A., MOOI, R., KRESS, W. J., and CANNELL, P. (1990), "Speaking of Forked Tongues: The Feasibility of Reconciling Human Phylogeny and the History of Language," *Current Anthropology, 31*, 1-24.

BERGSLAND, K., and VOGT, H. (1962), "On the Validity of Glottochronology," *Current Anthropology, 3*, 115-153.

CAVALLI-SFORZA, L. L., MINCH, E., and MOUNTAIN, J. L. (1992), "Coevolution of Genes and Language Revisted," *Proceedings of the National Academy of Science, 89*, 5620-5624.

CAVALLI-SFORZA, L. L., PIAZZA, A., MENOZZI, P., and MOUNTAIN, J. (1988), "Reconstruction of Human Evolution: Bringing Together Genetic, Archaeological, and Linguistic Data," *Proceedings of the National Academy of Science, 85*, 6002-6006.

COEDES, G. (1952), "A propos de Deux Fragments d'Inscription Récemment Découverts à P'ra Pathom (Thailande)," *Comptes Rendus des Séances de l'Académie des Inscriptions et Belles-Lettres*, 146-150.

COEDES, G. (1964), *Les Etats Hindouisés d'Indochine et d'Indonésie*, nouvelle édition, Paris: Editions E. de Boccard.

COULMAS, F. (1989), *The Writing Systems of the World*, London: Basil Blackwell.

DARWIN, C. (1871), *The Descent of Man, and Selection in Relation to Sex*, Vol. 1, London: John Murray. (Reprinted 1981, Princeton: Princeton University Press).

DIEBOLD, A. R., JR. (1964), "A Control Case for Glottochronology," *American Anthropologist, 66*, 987-1006.

DIRINGER, D. (1968), *The Alphabet: A Key to the History of Mankind*, 3rd edition, London: Hutchinson.

EMENEAU, M. B. (1966), "The Dialects of Old Indo-Aryan," in *Ancient Indo-European Dialects*, Eds., H. Birnbaum and J. Puhvel, Berkeley and Los Angeles: University of California Press, 123-138.

FARRIS, J. S. (1976), "Expected Asymmetry of Phylogenetic Trees," *Systematic Zoology, 25*, 196-198.

GIBSON, J. C. L. (1971), *Textbook of Syrian Semitic Inscriptions, Vol. 1: Hebrew and Moabite Inscriptions*, Oxford: Clarendon Press.

GIBSON, J. C. L. (1975), *Textbook of Syrian Semitic Inscriptions, Vol. 2: Aramaic Inscriptions*, Oxford: Claredon Press.

GIBSON, J. C. L. (1982), *Textbook of Syrian Semitic Inscriptions, Vol. 3: Phoenician Inscriptions*, Oxford: Claredon Press.

GORDON, C. H. (1965), *Ugaritic Textbook (Analecta Orientalia, 38)*, Rome: Pontifical Biblical Institute.

GUYER, C., and SLOWINSKI, J. B. (1991), "Comparisons of Observed Phylogenetic Topologies with Null Expectations among Three Monophyletic Lineages," *Evolution, 45*, 340-350.

GREENBERG, J. H. (1959), "Language and Evolution," in *Evolution and Anthropology: A Centennial Appraisal*, Washington: Anthropological Society of Washington, 61-75.

HARDING, E. F. (1971), "The Probabilities of Rooted Tree-Shapes Generated by Random Bifurcation," *Advances in Applied Probability, 3*, 44-77.

HARMATTA, J. (1979), "Sogdian Sources for the History of Pre-Islamic Central Asia," in *Prolegomena to the Sources on the History of Pre-Islamic Central Asia*, Ed., J. Harmatta, Budapest: Akadémiai Kiadó, 153-165.

HAWKINS, J. A., and GELL-MANN, M., Eds. (1992), *The Evolution of Human Languages (SFI Studies in the Sciences of Complexity, Proceedings Vol. X)*, Reading, MA: Addison-Wesley.

HEARD, S. B. (1992), "Patterns in Tree Balance among Cladistic, Phenetic, and Randomly Generated Phylogenetic Trees," *Evolution, 46*, 1818-1826.

HENNIG, W. (1950), *Grundzüge einer Theorie der Phylogenetischen Systematik*, Berlin: Deutscher Zentralverlag.

HENNIG, W. (1966), *Phylogenetic Systematics*, Urbana, IL: University of Illinois Press.

HOENIGSWALD, H. M., and WIENER, L. F., Eds., (1987), *Biological Metaphor and Cladistic Classification: An Interdisciplinary Perspective*, Philadelphia: University of Pennsylvania Press.

HOLMAN, E. W. (1989), "Some Evolutionary Correlates of Higher Taxa," *Paleobiology, 15*, 357-363.

HOLMAN, E. W. (1992), "Statistical Properties of Large Published Classifications," *Journal of Classification, 9*, 187-210.

HOUWINCK TEN CATE, P. H. J. (1986), "Anatolian languages," in *Encyclopedia Britannica*, 15th edition, Vol. 22, 607-612.

HUBBARD, A. E., and GILINSKY, N. L. (1992), "Mass Extinctions as Statistical Phenomena: An Examination of the Evidence Using X^2 Tests and Bootstrapping," *Paleobiology, 18*, 148-160.

JACKSON, K. (1953), *Language and History in Early Britain*, Cambridge, MA: Harvard University Press.

JENSEN, H. (1970), *Sign, Symbol and Script: An Account of Man's Efforts to Write*, 3rd edition, tr. G. Unwin, London: Allen and Unwin.

KANE, D. (1989), *The Sino-Juchen Vocabulary of the Bureau of Interpreters (Uralic and Altaic Series, 153)*, Bloomington, IN: Research Institute for Inner Asian Studies.

KEIGHTLEY, D. N. (1978), *Sources of Shang History*, Berkeley and Los Angeles: University of California Press.

KRZYZANOWSKI, J. (1978), *A History of Polish Literature*, tr. D. Ronowicz, Warsaw: PWN - Polish Scientific Publishers.

LANE, G. S. (1986), "Tocharian Language," *Encyclopedia Britannica*, 15th edition, Vol. 22, 684-685.

MACDOWELL, D. W., and TADDEI, M. (1978), "The Pre-Muslim Period," in *The Archaeology of Afghanistan*, Eds., F.R. Allchin and N. Hammond, London: Academic Press, 233-299.

MACKENZIE, D. N. (1969), "Iranian Languages," in *Current Trends in Linguistics, Vol. 5*, Ed. T.A. Sebeok, The Hague: Mouton, 450-477.

MARKEY, T. L. (1981), *Frisian (Trends in Linguistics State-of-the-Art Reports, 13)*, The Hague: Mouton.

MATISOFF, J. A. (1990), "On Megalocomparison," *Language, 67*, 106-120.

McALPIN, D. W. (1981), "Proto-Elamo-Dravidian: The Evidence and its Implications," *Transactions of the American Philosophical Society, 71(3)*.

MICHELENA, L. (1964), *Textos Arcaicos Vascos (Biblioteca Vasca, 8)*, Madrid: Ediciones Minotauro.

MILLER, R. A. (1967), *The Japanese Language*, Chicago: University of Chicago Press.

MORANDI, A. (1982), *Epigrafia Italica*, Rome: "L'Erma" di Bretschneider.

NEMESKÜRTY, I. (1983), "The Middle Ages," in *A History of Hungarian Literature*, Ed., T. Klaniczay, tr. I. Farkas, Budapest: Corvina Kiadó, 17-32.

NICHOLS, J. (1990), "Linguistic Diversity and the First Settlement of the New World," *Language, 67*, 475-521.

ORANSKIJ, I. M. (1977), *Les Langues Iraniennes*, tr. J. Blau, Paris: Librairie C. Klineksierk.

PAGE, R. D. M. (1991), "Random Dendrograms and Null Hypotheses in Cladistic Biogeography," *Systematic Zoology, 40*, 45-62.

PENZL, H. (1972), "Old Germanic Languages," in *Current Trends in Linguistics, Vol. 9*, Ed., T.A. Sebeok, The Hague: Mouton.

PETTINATO, G. (1981), *The Archives of Ebla*, Garden City, NY: Doubleday.

PLATNICK, N. I., and CAMERON, H. D. (1977), "Cladistic Methods in Textual, Linguistic, and Phylogenetic Analysis," *Systematic Zoology, 26*, 380-385.

POPPE, N. (1965), *Introduction to Altaic Linguistics (Ural-Altaische Bibliothek, 14)*, Wiesbaden: Otto Harrassowitz.

PRUNNER, G. (1967), "Die Schriften der Nicht-Chinesischen Völker Chinas," *Studium Generale, 20*, 480-520.

RAUP, D. M. (1978), "Cohort Analysis of Generic Survivorship," *Paleobiology, 4*, 1-15.

RAUP, D. M., and SEPKOSKI, J. J., JR. (1982), "Mass Extinctions in the Marine Fossil Record," *Science, 215*, 1501-1503.

RÖSSLER, O. (1958), "Die Sprache Numidiens," in *Sybaris: Festschrift Hans Krahe*, Wiesbaden: Otto Harrassowitz, 94-120.

RUHLEN, M. (1987), *A Guide to the World's Languages, Vol. 1: Classification*, Stanford, CA: Stanford University Press.

RUHLEN, M. (1991), *A Guide to the World's Languages, Vol. 1: Classification*, Stanford, CA: Stanford University Press.

SALZER, A., and VON TUNK, E. (1972), *Geschichte der Deutschen Literatur*, 3rd edition, Vol. 1, Zürich: Stauffacher-Verlag.

SAUVAGEOT, A. (1952), "Langues Ouraliennes," in *Les Langues du Monde*, Eds., A. Meillet and M. Cohen, Paris: Centre National de la Recherche Scientifique, 279-318.

SAVAGE, H. M. (1983), "The Shape of Evolution: Systematic Tree Topology," *Biological Journal of the Linnean Society, 20*, 225-244.

SCHAMSCHULA, W. (1990), *Geschichte der Tschechischen Literatur, Vol. 1.*, Köln: Böhlau Verlag.

SCHMALSTEIG, W. R. (1976), *Studies in Old Prussian*, University Park, PA: Pennsylvania State University Press.

SEPKOSKI, J. J., JR. (1984), ''A Kinetic Model of Phanerozoic Taxonomic Diversity. III. Post-Paleozoic Families and Mass Extinctions,'' *Paleobiology, 10,* 246-267.

SHINNIE, P. L. (1967), *Meroe: A Civilization of the Sudan,* New York: Praeger.

SLOWINSKI, J. B., and GUYER, C. (1989), ''Testing the Stochasticity of Patterns of Organismal Diversity: An Improved Null Model,'' *The American Naturalist, 134,* 907-921.

TERRAS, V. (1991), *A History of Russian Literature,* New Haven, CT: Yale University Press.

THURNEYSON, R. (1946), *A grammar of Old Irish,* revised and enlarged edition, tr. D. A. Binchy and O. Bergin, Dublin: Dublin Institute for Advanced Studies.

VAN BEEK, G. W. (1961), ''South Arabian History and Archeology,'' in *The Bible and the Ancient Near East: Essays in Honor of William Foxwell Albright,* Ed., G. E. Wright, Garden City: Doubleday, 229-248.

VENDRYES, J., and BENVENISTE, E. (1952), ''Langues Indo-Européennes,'' in *Les Langues du Monde,* Eds., A. Meillet and M. Cohen, Paris: Centre National de la Recherche Scientifique, 1-80.

WEIERS, M. (1967), ''Die Entwicklung der Mongolischen Schriften,'' *Studium Generale, 20,* 470-479.

WRENN, C. L. (1967), *A Study of Old English Literature,* London: George C. Harrap & Co.