Glottochronology, Lexicostatistics, and Other Numerical Methods

Bibliography

C. G. Williams

Glottochronology, Lexicostatistics, and Other Numerical Methods

Glottochronology and lexicostatistics are closely related concepts. Indeed, for some authors, these terms are interchangeable, as they both have to do with the application of numerical methods to the lexicon. However, as the name itself implies, the emphasis of glottochronology is historical; it is on the estimation of the time depth which separates a pair of languages. Lexicostatistics is an outgrowth from glottochronology, and is a more neutral term. It is more concerned with using the quantitative similarity in the lexicon for descriptive purposes, such as the estimation of the degree of mutual intelligibility.

The idea traces back to Morris Swadesh (see Swadesh, Morris), who was inspired by the method of carbon-14 dating that was invented in chemistry. The method is based on a naturally radioactive carbon isotope that decays with time, with a half-life of 5,700 years. Its use in determining the date of when life ceased for any carbon-containing animal or plant has revolutionized the understanding of prehistory. The challenge is to find something in language that can be shown to decay in time in some lawful way. Swadesh chose to work with words. The fundamental idea in glottochronology is that the time depth separating two genetically related languages can be estimated by the degree to which they share commonly inherited words.

1. Words as Units of Measure
It is well known that old words frequently fall into disuse as culture changes. For example, once a society makes the transition from horse-drawn carriages to automobiles, the associated vocabularies will change as well. Such cultural vocabularies cannot be used for the general purposes of glottochronology, since the words will be found in some societies but not in others. Furthermore, the rates of change of these words will be dependent on cultural events and may fluctuate unpredictably.

On the other hand, Swadesh reasoned, there must be a 'basic' set of words which deals with universal phenomena that must be present in all languages. This basic lexicon includes words for parts of the body (head, hand), natural objects (moon, mountain), common activities (come, sleep), as well as several other categories (see the Appendix). It turns out to be quite difficult to determine such a basic lexicon. The list of over 200 words which Swadesh originally proposed was gradually whittled down, to the 100 shown in the Appendix.

See also: Ecstatic Religion.
However, even this much-reduced lexicon contains a number of problems. For instance, a language may not have words for such colors as green, red, or yellow, and may not refer to these colors explicitly. Or, a language may call ‘moon’ not by a single word, but by a phrase, such as ‘eye of the night.’ Words change both in sound and in meaning. Thus, German Hund and English ‘hound’ are noticeably similar in pronunciation. Yet, in terms of meaning, the German Hund actually corresponds to English ‘dog’, whereas ‘hound’ has become much more specialized in its semantic range, that is, in the particular category of dog to which the word refers.

In fact, typically, corresponding words do not have exactly the same semantic range across a variety of languages. The problem is how closely their semantic ranges must match before two words across two languages may be considered the same. This problem is much more acute with semantic matching than with phonetic matching, where the principles of phonetic change are much better understood. Furthermore, phonetic matching can be supported with whole sets of words which exhibit the same sound correspondence, whereas semantic change usually applies to single words.

2. Computing Time Depth
Putting aside for now these problems in designing the basic lexicon, Swadesh’s idea is to assume that these words in the basic lexicon get replaced at a relatively constant rate over long time spans, somewhat like carbon-14. The commonly used figure is a retention rate of 80 percent over 1,000 years, that is, \( r = 0.8 \).

It could be proposed that La and Lb are two languages which derive from an ancestral language, Lz. Suppose further that they diverged from each other at time \( T_d \), and that the two speech communities had no contact since then. If the time of divergence, \( T_d \), is 1,000 years ago, then La and Lb will each have preserved 0.8 of the original basic lexicon of Lz. If \( T_d \) is 2,000 years ago, then La and Lb will each have preserved 0.8 times 0.8, or 0.64 of the original lexicon. In other words, using \( P(\text{La}) \) to represent the percentage of the original lexicon preserved in La, one can write \( P(\text{La}) = r^t \), where \( t \) is the number of millennia of separation. Similarly, \( P(\text{Lb}) = r^t \).

However, since there has been no contact between these two languages, it can be assumed that neither language influenced the other with respect to which words to preserve. The percentage of the shared basic lexicon must then be the product of the two independent retentions. If \( P(\text{La}, \text{Lb}) \) is used to represent the percentage of the original basic lexicon preserved in both languages, the formula for glottochronological reasoning may be written as follows:

\[
P(\text{La}, \text{Lb}) = r^{2t},
\]

\[
\text{and } t = \frac{(\log(P(\text{La}, \text{Lb})))}{2*(\log r)}.
\]

From (1), it is easy to compute that if \( t = 2 \), then \( P(\text{La}, \text{Lb}) \) is approximately 41 percent. If \( t = 3, 4, \) and 5, then \( P(\text{La}, \text{Lb}) = 26 \) percent, 17 percent, and 11 percent respectively. Conversely, given \( P(\text{La}, \text{Lb}) \), for any pair of related languages, it is possible to compute \( t \), from which \( T_d \) can be obtained, the time at which these two languages diverged. Therefore if 11 percent of commonly preserved basic lexicon in two languages is found, it would follow from glottochronological reasoning that these two languages diverged 5,000 years ago.

The challenge of estimating the time depth of language divergence is indeed a fascinating one, with profound implications for the study of human prehistory. Swadesh’s pioneering work has inspired a great deal of research on language prehistory. Unfortunately, the results obtained have been uneven, indicating that the methods of glottochronology developed so far are not totally reliable. A balanced survey of this body of research is Embleton (1986), where a detailed bibliography is presented. This article will briefly mention four problems with glottochronology, not to detract from the considerable merits of Swadesh’s pioneering work, but in the hope that better and better methods will be invented to meet this exciting challenge.

3. Four Problems with Glottochronology
The challenge of designing a universal basic lexicon, mentioned above, is one which some investigators do not believe is possible. Teeter (1963) has written, for example, ‘The closer we get to universal validity, the fewer items we have. My own opinion is that there will be no items at all on the perfect list.’ Perhaps the binary distinction basic:nonbasic is too crude and culture-bound to be useful. It may very well turn out, however, that different categories of words within a language change at different rates, and that there are many such categories. This is a question that Swadesh himself considered in a later article (1955). Such differential rates of change in the lexicon are predicted by the theory of lexical diffusion, which was advanced in 1969 (see Wang 1991:3). They have been abundantly documented in the history of English by Ogura (1987:64).

In any case, from the several dozen glottochronological studies which have been published up to 1992, it seems that the retention rate proposed by Swadesh of around 80 percent is too low. It remains to be determined whether the higher retention rate is to be solely ascribed to continued contact for certain situations, or whether the estimate should be increased universally. Only many empirical studies of well-documented cases can eventually answer this question.

A second, related problem has to do with a uniform rate of change assumed for all languages. Linguists have long observed, though not quantified, that languages differ appreciably in their rates of change. In the Germanic family, Icelandic is often remarked to be relatively conservative, especially when contrasted with an innovative language such as English. The received opinion, therefore, is on the side of Ellegård when he wrote: ‘It seems to me that Swadesh is somewhat rash in assuming a uniform rate of development.’ Indeed, a better research strategy would be to devise methods for empirically computing various \( r \)’s for different languages, as well as for different categories of words, rather than assuming a uniform value.

A third problem is the assumption of the sudden and total break-off of contact of the diverging languages. It is much more usual that contact between languages is maintained to varying degrees even after the speech communities have largely separated from each other. Such contacts have the effect that the retention of words from the original lexicon is not completely independent in the two languages.
Consequently, the value of $t$ computed from the glottochronology formula is frequently too small. A fundamental difficulty in all historical linguistics is the separation of traits which are transmitted vertically through time from those transmitted horizontally from contemporaries. Glottochronology is also vulnerable to this difficulty.

A fourth problem is a logical one. Glottochronology works on languages one pair at a time, and cannot make use of crucial information contained in other related languages. Often, the relation that a third language, $L_c$, has with $L_a$ and $L_b$ can be highly informative on the relation between the other two languages. New information comes with each new addition to the group of languages to be compared. In expanding their concern with larger groups of languages, linguists are at once returning to the traditional way of representing language families with tree diagrams, and linking their interests with developments in biology.

4. Parallels with Biological Evolution

It should come as no surprise that there are significant parallels between the evolution of language and that of organisms. Darwin made special note of these in his *Descent of Man* (1871). His contemporary, Schleicher (see Schleicher, August), to whom the first systematic use of family tree diagrams in linguistics is credited, was deeply impressed by these parallels as well. This suggests that some of the methods developed for evolutionary studies in one area may be applied fruitfully in the other area.

One such experiment at applying a model developed by Cavalli-Sforza and Wang (1986). The model was called isolation-by-distance, envisages the distribution of a human population as concentrated in communities, which are located in a linear succession. The aim of the model is to predict the average genetic similarity between these communities on the basis of the geographical distances separating them. This model has been successfully applied to many parts of the world. The genetic similarity, $G$, it turns out, is related to the distance, $D$, by a negative exponential. This may be written as the following equation, where $k$ is a constant that is determined by other factors:

$$G = e^{-(kD)}.$$

(2)

In the linguistic experiment, the quantity $G$ is taken to mean the degree to which two communities share vocabulary. Instead of genetic similarity, therefore, the quantity investigated was lexical similarity. When the model was applied to lexical data from Chinese dialects, no simple relation could be observed between $G$ and $D$, presumably because of the complex geographical as well as historical relations among these communities. The model was also applied to a body of lexical data from a chain of island communities in Micronesia in the South Pacific. Here, the relation of the negative exponential is found, similar to that of the genetic studies. The linguistic curve is somewhat different from the genetic one because of the differential replacement rates of individual words, a problem which has already been alluded to above.

It is clear that there can be complications in linguistic data which the simple isolation-by-distance model cannot capture, as with the Chinese dialects. On the other hand, the Micronesian data provide an excellent analog to the genetic studies. Lexical replacement there is a function of geographical distance, much as in the case of the replacement of alleles in genetic interactions. So, biological methods can be usefully applied in linguistics, if the proper analogs are chosen. One fundamental difference between linguistic evolution and many types of biological evolution is that the former contains extensive amounts of horizontal transmission. One receives genetic materials only from one’s parents; while linguistic materials typically come from many sources in one’s environment: parents, teachers, peers, and mass media.

5. Unrooted and Rooted Trees

There are two classes of methods which will be considered here for estimating time depths: distance methods and parsimony methods. They have in common the goal of sub-grouping the taxa in the most informative configuration, whether these taxa are languages or biological objects. All these methods require intensive computation when the number of taxa becomes large. However, most of them have been programmed for microcomputers, so that the most tedious aspects of repeated comparisons and arithmetic can be relegated to machines.

To obtain some idea of the various processes of computation, it is necessary first to define some basic terms associated with tree diagrams. A tree is a collection of connected branches, nodes, and tips. Tips are distinguished by the fact that they are each connected to only a single branch. Tips are typically labeled by the taxa being analyzed; in this case, by languages. A tree is unrooted if one of the nodes is a root. The present discussion will consider only those unrooted trees where each node is connected to exactly three branches.

In the context of these constraints, it can be seen that, given three tips, there is exactly one unrooted tree. This is to say that there is only one way of representing three languages with an unrooted tree, with a unique node, $n_1$, in the middle of the tree. The languages may be called $A$, $B$, and $C$, and the branches connecting them $a$, $b$, and $c$. Now, one may want to add a fourth language, $D$, to the tree, together with its branch, $d$. One can, for instance, attach the branch $d$ to the branch $a$, creating a new unrooted tree, with a new node, $n_2$. The new tree may be represented in the following linear notation:

$$(n_2(A, D), n_1(B, C)).$$

(3)

Or, one can attach $d$ to $b$, or to $c$, creating a new unrooted tree in each case, which may be represented respectively as:

$$(n_2(B, D), n_1(A, C)).$$

and

$$(n_2(C, D), n_1(A, B)).$$

(4)

Because the original three-language tree has three branches, there are exactly three ways to form a four-language tree. Each of the three four-language trees formed in this way has five branches, since $d$ has been added and since $n_2$ has bisected an original branch. If another language were added to the group, say $E$ with its branch $e$, this now gives 15 ways of doing so. One can attach $e$ to any of the five branches, creating a new node $n_3$, and one can do so on any of the three four-language trees. Therefore there are exactly 15 unrooted trees for five languages. Furthermore, it is known that each of these new trees has seven branches.
If \( E \) is added as a neighbor of \( A \), the linear representation of that unrooted tree is:

\[
(n2)(3(A, E), D), n1(B, C)).
\]

(5)

The procedure for forming unrooted trees should now be clear. Because of the constraint that each node connects exactly three branches, an unrooted tree representing \( L \) languages will always have exactly \( 2L - 3 \) branches, shown as \( B \) below. To compute \( U(L+1) \), the number of unrooted trees for \( L+1 \) languages, multiply \( B(L) \) and \( U(L) \). An equivalent way of computing \( U(L) \) is to take the cumulative product of all the odd integers up to and including the \((L-2)\)th one. For example, if \( L=9 \), then \( U(9) \) is

\[1^*3^*5^*7^*9^*11^*13,\]

which is 135,135. This is to say, there are 135,135 ways of representing 9 languages with unrooted trees.

In evolutionary studies, rooted trees are more familiar. In such trees, there is a root, which is distinguished by having two outgoing branches. All other nodes are still connected by three branches, one incoming from the root, and two outgoing toward the tips. Earlier, the constraint was placed on unrooted trees of each node being connected to exactly three branches. Now, it can be seen that for rooted trees this constraint results in all trees being binarily branching. This may be a simplification of historical situations where three communities diverge from each other at exactly the same time. However, such situations are probably not common; in any case, the simplification is not critical.

Given an unrooted tree, it may be rooted by inserting the root on any of its branches. Thus, using the table above, \( R(L) = B(L) * U(L) \). Or equivalently, to compute \( R(L) \), take the cumulative product of all the odd integers up to and including the \((L-1)\)th one. For \( L=9 \), \( R(9) \) is then the product of all the first 8 integers, which turns out to be greater than 2,000,000.

This is to say that, given 9 languages, there are over 2,000,000 ways of subgrouping them. Just enumerating these trees is a formidable computing task, let alone examining them for historical informativeness. Access to computer programs is therefore indispensable for such research in historical linguistics. Some of the programs which have been developed by biologists that linguists will find useful are those by Felsenstein (PHYLP), Maddison and Maddison (MacClade), and Swoford (PAUP), as listed in the Bibliography.

6. Distance Methods

A pioneering effort towards providing methods for analyzing genetic distances among groups of human populations is that of Cavalli-Sforza and Edwards (1967). Many of the concepts subsequently taken for granted in phylogenetic analysis were first presented in that paper, such as the quest for the minimal tree (parsimony) and the requirement that the branch lengths be additive (additivity). Since that time, several related methods have been suggested, and programs for some of them have become available on microcomputers. The earlier methods often yield trees in which all tips are equidistant from the root, such as Average Linkage and UPOMA, which stands for Unweighted Pair-Group arithMetic Average. Since these earlier methods assume constant rates of change, they will not be considered here. Other methods do not make this assumption and allow the root-to-tip distance to vary. One such method is Neighbor Joining (Saitou and Nei 1987), which will be applied to some linguistic data below.

The raw data for distance methods is the pairwise similarity among the languages. For example, take the relations among seven languages. Some quantifiable dimension of these languages would be chosen, and their degrees of similarity along this dimension indicated in a matrix such as the following.

<table>
<thead>
<tr>
<th></th>
<th>XM</th>
<th>MX</th>
<th>GZ</th>
<th>NC</th>
<th>CS</th>
<th>SZ</th>
<th>BJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xiamen</td>
<td>100</td>
<td>68</td>
<td>63</td>
<td>64</td>
<td>61</td>
<td>59</td>
<td>56</td>
</tr>
<tr>
<td>Mexian</td>
<td>100</td>
<td>79</td>
<td>77</td>
<td>72</td>
<td>73</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Guangzhou</td>
<td>100</td>
<td>78</td>
<td>76</td>
<td>77</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanchang</td>
<td>100</td>
<td>88</td>
<td>84</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Changsha</td>
<td>100</td>
<td>86</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suzhou</td>
<td>100</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These data represent the major urban dialects of Chinese. The numbers are taken from Xu (1991: 422). They represent the percentages of shared basic lexicon between all pairs. Thus it can be seen that the numbers along the main diagonal of the matrix are all 100 percent, since each dialect is being compared with itself. The similarity between Xian and Beijing, shown in the upper right corner of the matrix, is the lowest, that is, 56 percent. Using the formula for glottochronology derived earlier, the time depth for the divergence of Xiamen and Beijing can be computed at around 1,500 years. This suggests that the two speech communities separated some time around 500 ad, which is too recent on the basis of other language-internal as well as external evidence. As mentioned earlier, such errors of underestimating \( t \) are often due to the fact that the communities have maintained contact after the initial separation.

The similarity numbers of the above table, which reflect affinity, can be easily transformed to dissimilarity numbers, which reflect distance. One way to do this is simply to subtract each percentage from 100, so that the distance between Beijing and Xiamen becomes 100 - 56, or 44. The distance matrix composed of such distance numbers can then be analyzed into an unrooted tree by some computer program. Such a distance matrix for the seven Chinese dialects has been constructed, and analyzed with the Neighbor Joining method.

Intuitively, the method may be visualized as follows, starting with a hypothetical tree shaped like a star. This tree has a single node in the middle, \( n_0 \), with branches radiating from it leading to the seven tips. The question is which two languages should be grouped together such that the total of all the branch lengths can be minimized. A node \( n_1 \) is inserted above these two tips, and a pseudo-tip \( t_1 \) is posited, such that the distance between \( n_1 \) and \( t_1 \) is the average of the distance between \( n_1 \) and the two original tips. The result of these steps is a new star, now with six tips and two nodes, and the procedure can start again.

This process is iterated until an unrooted tree is produced, in which each node is connected to three branches. The distance between any pair of languages is the sum of
the distances of the branches separating them. This tree aims to satisfy two conditions: the pairwise distances should approximate the values in the original distance matrix; and the total branch lengths should be minimized. From this tree, one could construct another distance matrix, call it the output matrix, and compare it with the original input matrix. Typically, these two matrices will not be identical, and a difference matrix may be computed. To a certain extent, it may be conjectured that the differences between the input and output matrices are due to effects other than strict inheritance. In this regard, the difference matrix may provide a first step toward separating vertical transmission from other effects.

Various suggestions have been made for finding the root for an unrooted tree, such as the one obtained for the Chinese dialects by the Neighbor Joining method. A common procedure is to place the root at the midpoint of the longest path separating two tips. This procedure produces

![Figure 1. Rooted tree of seven Chinese dialects, obtained by the Neighbor Joining method. (Data from Xu 1991.)](image)

the rooted tree as shown in Fig. 1. The subgrouping indicated in Fig. 1 has some interesting features. The fact that Xiamen is most distantly related to the other six dialects will not surprise scholars in Chinese linguistics; Xiamen preserves many archaic traits which have all been replaced in dialects.

More surprising is the fact that Suzhou and Changsha are grouped closely together. Although Chinese linguists recognize that the region containing these two cities was more linguistically homogeneous in ancient times, this region has been repeatedly split apart by paths of migration since around 1000 AD. The fact that they are closely associated may be due to the use of the basic lexicon, which is more conservative than other words. A similar analysis of these seven dialects, using a larger lexicon, does not show a similar degree of affinity between these two cities.

An important contribution of such distance methods over that of glottochronology is that all the languages are calibrated against each other on the tree, rather than only one pair at a time. So, once the absolute date of one split can be determined, say on the basis of historical or textual evidence, the other splits on the tree can all be inferred.

7. Character State Methods

An important limitation of distance methods is that they apply to only a single dimension. Languages typically change along numerous different dimensions, in the lexicon, phonology, morphology, syntax, etc. Some of these changes may be related to each other, others may not. In some cases, the direction of change may be known, say from x to y; in other cases, it may not. Furthermore, some dimensions may be binary, such as voicing for obstruents, while other dimensions may involve more states, such as vowel raising from a to e to i.

Even though character state methods were developed by biologists in the context of analyzing molecular sequence data, which comprise unidimensional data, these methods allow for the treatment of multidimensional data of the kind found in language change. Each of these dimensions of change may be called a character. Thus, for example, vowel raising may be used as a character for studying language history. Furthermore, if it is known that the direction of change has been a => e => i, then it is a three-state directed character. As several characters are combined in one phylogenetic analysis, there is the further complication that some of these have greater diagnostic value than others, in which case they may be weighted differently. In the early 1990s, efforts are being made to discriminate among characters on the basis of their internal compatibilities; see Meacham (1992). It may be assumed that the subset of characters which exhibits the greatest degree of internal consistency will also provide the most plausible phylogeny.

The fundamental idea in character state methods is to find the most parsimonious tree, that is, a tree which implies the minimum number of changes in all of the characters. The branches of such trees are not differentiated for length, unlike those from distance methods. But a branch may or may not be marked for a change, and it is these marks which are to be minimized.

This idea was independently tested in linguistics for subgrouping South Central Dravidian languages, by Krishnamurti, et al. (1983), although they used different methods of computation from those used in character state analysis, such as that computerized by Swafford (1991). Their data comprise 63 cognate words in six Dravidian languages (see Fig. 2), with respect to a particular set of sound changes. Since sound changes typically proceed by lexical diffusion,
it turns out that the languages vary according to which subset of these 63 words has changed. Thus, the situation may be analyzed as one of 63 two-state, directed characters, where the states are u for unchanged and c for changed.

Fig. 2 illustrates the situation with four rooted binary trees for a single word, where the tips are labeled by the six Dravidian languages. In addition to u and c, the letter 0 indicates that the word in question has been lost. Thus, in tree 1, the word has been lost in Konda; hence Konda will not figure in the analysis. The grouping in tree 1 turns out to be the most parsimonious for this word, since only one change needs to be posited, that for the ancestral language of Kui, Kuvi, Pengo, and Manda. Trees 2 and 3 each require two changes in character state, while tree 4 requires four changes. The best way to represent the phylogeny of these six languages, therefore, is the most parsimonious tree, examined for all 63 words.

8. Conclusion
Glottochronology was invented in the 1950s for dating pair-wise language separations. The method did not receive wide usage because of several basic weaknesses inherent in it. Since its invention, various powerful methods have been advanced, primarily from biological research, which construct phylogenetic trees. Distance methods facilitate work with single dimensions of change, and yield trees with branch lengths which reflect differential degrees of independent evolution. Character state methods allow the combination of many dimensions of change into a single phylogenetic framework, and yield trees that minimize the number of changes.

Since these methods were invented for research on phylogenetic problems very similar to those which linguists study, their utility to linguistics is obvious. It may be expected that the twenty-first century will see increasing application of these methods in work on language prehistory, as well as greater interaction among scholars who study the evolution of organisms and those who work on the evolution of language. From such interaction, much will be learned about the many parallels between the two modes of evolution as well as about their fundamental differences.

Appendix: Basic Lexicon Used in Glottochronology

1. all 21. ear 41. horn 61. nose 81. stone
2. ashes 22. earth 42. I 62. not 82. sun
3. bark 23. eat 43. kill 63. one 83. swim
4. belly 24. egg 44. knee 64. person 84. tail
5. big 25. eye 45. know 65. rain 85. that
6. bird 26. fat 46. leaf 66. red 86. this
7. bite 27. feather 47. lie 67. road 87. thou
8. black 28. fire 48. liver 68. root 88. tongue
9. blood 29. fish 49. long 69. round 89. tooth
10. bone 30. fly 50. loose 70. sand 90. tree
11. breast 31. foot 51. man 71. say 91. two
12. burn 32. full 52. many 72. see 92. walk
13. claw 33. give 53. meat 73. seed 93. warm
14. cloud 34. glad 54. moon 74. sit 94. water
15. cold 35. green 55. mountain 75. skin 95. we
16. come 36. hair 56. mouth 76. sleep 96. what
17. die 37. hand 57. name 77. small 97. white
18. dog 38. head 58. neck 78. smoke 98. who
19. drink 39. hear 59. new 79. stand 99. woman
20. dry 40. heart 60. night 80. star 100. yellow

Bibliography
Ellegård A 1959 Statistical measurement of linguistic relationship. Lg 35: 131–56
Felsenstein J 1990 PHYLIP Manual, version 3.3. University Herbarium, University of California, Berkeley, CA
Gudchinsky S C 1956 The ABCs of lexicostatistics (glottochronology). Word 12: 175–210
Krishnamurti B, Moses L, Danforth D 1983 Unchanged cognates as a criterion in linguistic subgrouping. Lg 59: 541–68
Meacham C A 1992 Evaluation of phylogenetic information content of characters of primitive angiosperms by probability of compatibility. Unpublished manuscript
Ogura M 1987 Historical English Phonology: A Lexical Perspective. Kenkyusha, Tokyo
Swadesh M 1951 Diffusional cumulation and archaic residue as historical explanations. Southwestern Journal of Anthropology 7: 1–21
Wang W S-Y 1991 Explorations in Language. Pyramid Press, Taiwan
Xu Tongqiang 1991 Lishi Yuyanxue (Historical Linguistics). Shangwu Yinxihuguan, Beijing

W. S.-Y. Wang

Goffman, Erving (1922–82)
Erving Goffman, a Jewish Canadian, spent three decades studying face-to-face interaction through naturalistic observation, completing 11 radically influential books and significant articles and chapters. Goffman’s essays, his preferred style of presentation, influenced sociologists, anthropologists, linguists, and psychiatrists. He was not uniformly respected and his status was even varied within sociology, although his contributions to microsociology, deviance, and medical sociology were unmistakable.

Goffman’s signature was controversy. His methods, theoretical persuasion, political stance, the import of his ideas, and, curiously, his personal lifestyle were religiously criticized. Attention to Goffman since his demise, however, has clarified the cornerstone of his career as microsociological constraint and marked much of the criticism as wrong-headed. Goffman is sometimes inappropriately described as a ‘symbolic interactionist.’ Symbolic interactionism emphasizes the SELF as a being constantly interpreting events (see Symbolic Interactionism). While for Goffman (1959: 1450...