RECOGNIZING PATTERNS OF DIVERGENCE AND CONVERGENCE IN A MATRIX OF LEXICOSTATISTIC RELATIONS

Gary Simons
Summer Institute of Linguistics
and Cornell University

0. INTRODUCTION[1]

The phenomenon of language convergence, though always considered anomalous by the standard divergence model of language change, is very real and can be seen in the cognate percentages of almost any lexicostatistic study. By convergence, I mean the process by which two varieties of speech in contact become more like each other. Divergence is the process by which two varieties of speech become less alike. This paper does not attempt to explain a convergence model (see Simons 1977), rather, it discusses how patterns of language convergence and divergence can be recognized in a matrix of lexicostatistic relations[2]. First, I discuss the "anomaly" of convergence. Then the different patterns of divergence and convergence are explained and illustrated. Finally, the method of recognizing the patterns is illustrated by analyzing some published matrices of lexicostatistic relations.

1. THE ANOMALY OF CONVERGENCE

The phenomenon of language and dialect chaining has long proved a stumbling block for lexicostatistic analysis (McElhanon 1971:136-140). The lexicostatistic model as developed by Swadesh allows only for the divergence of speech groups. One of the basic assumptions of the method is that the lexical replacement that occurs in one speech group is statistically independent of the replacement that occurs in a neighbouring, related group (Lees 1953). Thus any form of change brought about by the influence of a neighbouring language group, is not permitted by the model. However, such change (borrowing) does occur, and quite frequently. If these borrowings have sufficient time depth or do not include sounds that have undergone diagnostic sound changes, they cannot be detected by the comparative method. The result is lexicostatistic figures which are inflated and a language chain appears.

Conrad and Dye (1975:6) illustrate the anomaly of a language chain with the following example:

> Assume that B is 50% cognate with A and C, but A is only 40% cognate with C. Then three mutually contradictory patterns of divergence are implied by the traditional model. If one's conclusions are based on cognates with A, then C must have split off first. If based on cognates with B, all three diverged at once. Based on C, A must have split off first (see figure below).

TYPICAL COGNATE PERCENTAGES IN A
CHAIN OF THREE LANGUAGES

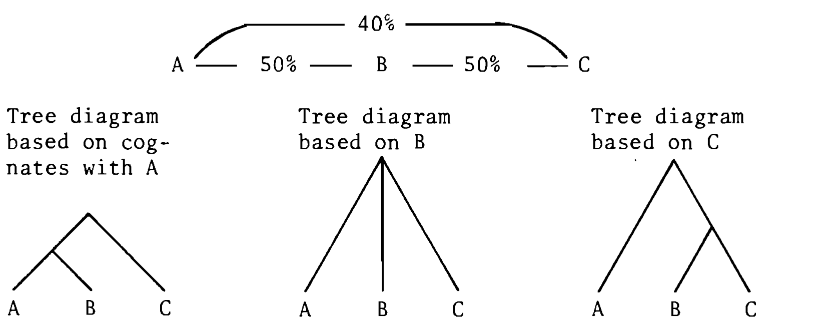


Figure 1

Conrad and Dye go on to explain that these relationships result not from the random divergence of speech groups, but from linguistic influence across language boundaries caused by social contact. This is what I mean by convergence relationships. Rather than attempting

The dominance pattern, too, causes problems for classification and subgrouping. The lexicostatistic relations of the central speech group have become so inflated through convergence, that it is generally impossible to determine its immediate genetic affiliations.

4.3 Sporadic convergence

The phenomenon I am calling sporadic divergence is discussed by Dyen (1963) in his paper "Lexicostatistically determined borrowing and taboo". He states that the lexicostatistic relations between a single language and all the languages in another subgroup of languages should be the same. This is a basic hypothesis of the divergence model. Dyen concludes that "every significant difference between lexicostatistical percentages implies a historical event determinable from the configuration of the percentages" (1963:66). The historical events he's talking about are those which result in borrowing and taboo. The discussion of taboo is beyond the scope of this paper. Suffice it to say that taboo results in an inflated rate of divergence. As Dyen shows (1963:64), taboo can be detected only by comparing the relationships of suspect and non-suspect languages to distantly related languages.

Borrowing, however, is exactly the kind of thing meant by convergence. To rephrase Dyen, any lexicostatistic percentage which is significantly higher than the diagnostic block for a divergence says it should be, indicates a convergence relationship between the speech groups involved. For instance, in Figure 20 below, there is a diagnostic block which defines a split between ABCD and EF. One percentage, however, the 76% between C and E, is inflated. This indicates a convergence relationship between C and E. It is an isolated instance of convergence in a whole matrix of divergence relationships.

SPORADIC CONVERGENCE IN A DIVERGENCE PATTERN

A					
90	B				
81	81	C			
73	73	73	D		
67	67	76	67	E	
67	67	67	67	81	F

Figure 20

The instance of sporadic convergence in a divergence pattern is unique among the convergence patterns in that the subgrouping of speech to develop the convergence model further in this paper, I will simply outline a method by which the effects of social contact can be discovered in a matrix of lexicostatistic relations. The means by which both patterns of divergence and convergence can be recognized will be discussed. Once the linguist has extracted the patterns of divergence and convergence from the matrix, he must look beyond the numbers to the actual data in an attempt to explain the social dynamics of the situation.

2. THE DIVERGENCE PATTERN

We begin by examining the minimal case of divergence, that of three speech groups.[3] In terms of a tree diagram, the situation is this:

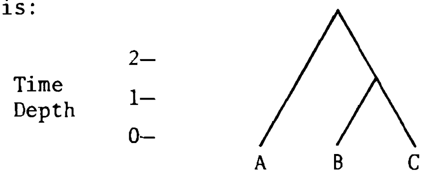


Figure 2

The same relations are shown in the following matrix where the number in each cell represents the number of time periods since divergence of the two speech communities.

A		
2	B	
2	1	C

Figure 3

By Swadesh's model, given the time depth of separation, *t*, and the rate, *r*, of retentions of basic vocabulary, the percentage of cognates, *c*, between two speech groups is predicted by the formula (Lees 1953):

$$c = r^{2t}$$

Given a retention rate of .95 per time period, the above matrix predicts the following cognate percentages:

A		
$.95^4$	B	
$.95^4$	$.95^2$	C

or

A		
81	B	
81	90	C

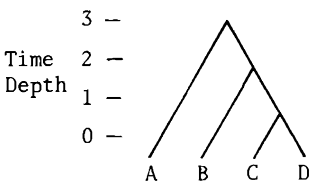
Figure 4

(As done in the above example, all lexicostatistic relations will be converted to percentages and rounded to the nearest whole number. In all the examples to follow, the retention rate of .95 per time period is used.

As seen in the example, the basic divergence pattern for three speech groups contains one high percentage and two percentages which are lower and equal to one another.[4] As more speech groups are added to the divergence pattern, the diagnostic feature of the pattern becomes apparent.

SUCCESSIVE DIVERGENCE INVOLVING FOUR SPEECH GROUPS

Tree



Lexicostatistic relations

in time depths

A			
3	B		
3	2	C	
3	2	1	D

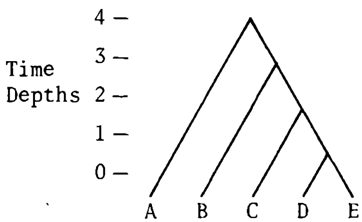
as cognate percentage

A			
73	B		
73	81	C	
73	81	90	D

Figure 5

SUCCESSIVE DIVERGENCE INVOLVING FIVE SPEECH GROUPS

Tree



A				
4	B			
4	3	C		
4	3	2	D	
4	3	2	1	E

A				
67	B			
67	73	C		
67	73	81	D	
67	73	81	90	E

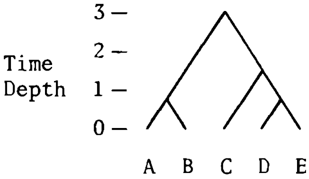
Figure 6

The diagnostic feature of the above divergence patterns, in which one group at a time splits from the others, is a whole column (or row) of adjacent cells filled by equal and low cognate percentages. The deeper the time depth, the lower is that percentage, and the longer is the column of equal percentages.

The general pattern emerges when we examine the case of more than one group splitting off together:

COMPOUND DIVERGENCE INVOLVING FIVE SPEECH GROUPS

Tree in time depths



A					
1	B				
3	3	C			
3	3	2	D		
3	3	2	1	E	

A					
90	B				
73	73	C			
73	73	81	D		
73	73	81	90	E	

Figure 7

In this example, the diagnostic pattern for the split is the block of equal and low percentages which is three cells high by two cells wide. To generalize, the diagnostic feature of a divergence is a rectangular block of equal and lower cognate percentages. The groups which have diverged are those which define the columns in the block, as against those which define the rows. That is, in Figure 7 above, there is a three by two block of equal and low percentages (i.e. 73). A and B define the columns of the block; C, D, and E define the rows of the block. Thus, this pattern indicates a divergence of groups A and B from groups C, D, and E.

DIAGNOSTIC BLOCK EXTRACTED FROM FIGURE 7

A	B	
73	73	C
73	73	D
73	73	E

indicates

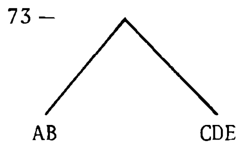


Figure 8

In figure 5 above, we see a three by one block of 73's. Group A defines the column; B, C, and D define the rows. Thus, this block is diagnostic of a split between A and BCD, at a time depth corresponding to 73% cognates. In the sub-matrix which remains for B, C, and D, we see a two by one diagnostic block of 81's. It is defined by B versus C and D, and thus indicates a split between B and CD at a later time corresponding to 81% cognates. All that remains is a one by one block defined by C versus D. It indicates a split between C and D at a time depth corresponding to 90% cognates.

DIAGNOSTIC BLOCKS EXTRACTED FROM FIGURE 5

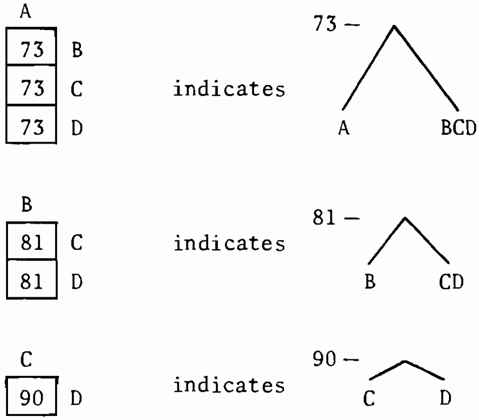


Figure 9

The example of figure 9 illustrates the procedure for extracting successive patterns of divergence from a matrix of lexicostatistic relations. First the diagnostic block with the lowest cognate percentage (which will also be the largest diagnostic block) is extracted. The diagnostic blocks in the remaining sub-matrices are then extracted in ascending order of cognate percentages; that is, the remaining diagnostic block with the lowest cognate percentage is always extracted next.

To conclude this section on the divergence pattern, we will illustrate the above procedure on a matrix of data collected in a recent lexicostatistic study of the dialects of the Biliau language on the Rai Coast of Papua New Guinea.[5] The five dialects are Biliau (Bl), Yamai (Ym), Yori (Yr), Galek (Gl), and Suit (St). On the Swadesh 100-word list, the following lexicostatistic relations were computed:

THE FIVE DIALECTS OF THE BILIAU LANGUAGE

Bl				
96	Ym			
89	89	Yr		
77	75	74	Gl	
74	76	75	92	St

Figure 10

The three possible interpretations of the convergence pattern are summarized in Figure 15 below. The double-headed arrows indicate the convergence relationships.

POSSIBLE HISTORICAL INTERPRETATIONS OF THE BASIC CONVERGENCE PATTERN

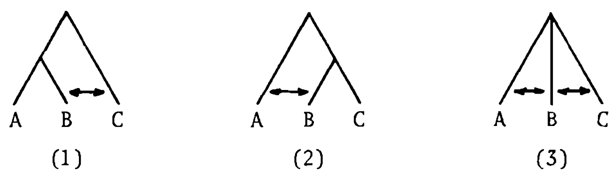


Figure 15

The lexicostatistic data alone cannot resolve the problem of the true historical relations in the case of the convergence pattern. One must look to oral tradition, known external history, and detailed linguistic comparison. In the preliminary lexicostatistic study we must then be content to symbolize the convergence pattern as the ill-formed tree of Figure 14.

4. PATTERNS OF CONVERGENCE IN RELATIONS INVOLVING MORE THAN THREE SPEECH GROUPS

There are three patterns of convergence commonly found in relations involving more than three speech groups—chaining, dominance, and sporadic convergence. The basic convergence pattern involving only three speech groups could in fact be an instance of any one of these patterns. Only when the convergence relations are put into a larger context can the different patterns be distinguished. The same patterns that exist between individual speech groups can be seen to exist between subgroups of the groups in the matrix. These relations are discovered by reducing the matrix to its subgroups.

4.1 Chaining

In the chaining pattern, the speech groups are situated geographically in a roughly contiguous arrangement. A speech group has contact relationships with the two groups on either side of it in the chain. This results in the linguistic convergence of adjacent groups. McElhanon (1970:218) has documented a dialect chain involving more than 70 villages from five languages of the Huon Peninsula of Papua New Guinea. In this case, contact was along the lines of a major trade route. The geography of the chaining pattern is diagrammed in figure 16. The double-headed arrows represent convergence relations.

GEOGRAPHY OF THE CHAINING PATTERN

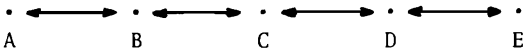


Figure 16

The chaining pattern results in the following lexicostatistic relationships. A speech group will show its highest cognate percentages with its neighbours which are adjacent in the chain. As groups get further and further apart along the chain, the lexicostatistic relation between them gets lower and lower. When the lexicostatistic relations of a chain are represented in a matrix in which adjacent speech groups in the chain are placed in adjacent rows and columns of the matrix, the pattern which emerges is as follows. The highest lexicostatistic relations will occur on the diagonal and the lowest relations will occur in the corner. As one proceeds from the corner to the diagonal, the matrix values get successively larger. This is the diagnostic shape for the chaining pattern. It is shown below in Figure 17.

THE CHAINING PATTERN

In terms of time depths

A				
1	B			
2	1	C		
3	2	1	D	
4	3	2	1	E

In terms of cognate percentages

A				
90	B			
81	90	C		
73	81	90	D	
67	73	81	90	E

In terms of a tree diagram

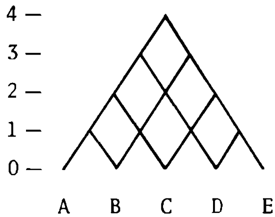


Figure 17

In the chaining pattern it is impossible to posit a classification or subgrouping of the sppech groups on the basis of lexicostatistic data alone (McElhanon 1971:136-140). The problem of historical interpretation for the basic convergence pattern (see Figure 15) is compounded many times over in the chaining pattern. Lamb (1959:41ff) has suggested that no classification or subgrouping be attempted in these cases, that "no taxon is to be recognized at the intermediate rank". Such a method is rarely applied, however. In positing boundaries, many investigators accept the boundaries recognized by the local inhabitants (McElhanon 1970:218, Cook 1966). Others attempt to resolve the indeterminacies by looking at phonological or structural data (Wurm 1961:21, McElhanon 1970:228-229). Nevertheless, in the preliminary lexicostatistic study, no subgroupings can be posited. We must be content to diagram the chaining pattern as the fan-shaped tree of Figure 18.

4.2 Dominance

The dominance pattern occurs when a number of speech groups show convergence with one speech group. In terms of geography, the one speech group is generally central to the others and would have some sort of political, economic, or social dominance in the area. The geography of the dominance pattern is diagrammed in Figure 18.

GEOGRAPHY OF THE DOMINANCE PATTERN

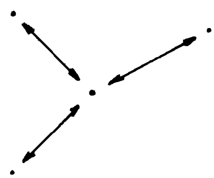


Figure 18

The dominance pattern results in the following lexicostatistic relationships. Most of the speech groups in the area will show their highest lexicostatistic relation with the central speech group. Those that don't have their highest relation with the central group, will at least have a relatively high relation with it. This appears in the matrix of cognate percentages as a row or column with consistently high values. This row or column of high values is the diagnostic characteristic of the dominance pattern. Figure 19 below illustrates a hypothetical case in which B, C, and D have a divergence pattern, but show a dominance pattern with A.

THE DOMINANCE PATTERN

In terms of time depths

A			
1	B		
1	3	C	
1	3	2	D

In terms of cognate percentages

A			
90	B		
90	73	C	
90	73	81	D

In terms of a tree diagram

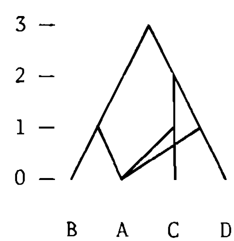


Figure 19

The dominance pattern, too, causes problems for classification and subgrouping. The lexicostatistic relations of the central speech group have become so inflated through convergence, that it is generally impossible to determine its immediate genetic affiliations.

4.3 Sporadic convergence

The phenomenon I am calling sporadic divergence is discussed by Dyen (1963) in his paper "Lexicostatistically determined borrowing and taboo". He states that the lexicostatistic relations between a single language and all the languages in another subgroup of languages should be the same. This is a basic hypothesis of the divergence model. Dyen concludes that "every significant difference between lexicostatistical percentages implies a historical event determinable from the configuration of the percentages" (1963:66). The historical events he's talking about are those which result in borrowing and taboo. The discussion of taboo is beyond the scope of this paper. Suffice it to say that taboo results in an inflated rate of divergence. As Dyen shows (1963:64), taboo can be detected only by comparing the relationships of suspect and non-suspect languages to distantly related languages.

Borrowing, however, is exactly the kind of thing meant by convergence. To rephrase Dyen, any lexicostatistic percentage which is significantly higher than the diagnostic block for a divergence says it should be, indicates a convergence relationship between the speech groups involved. For instance, in Figure 20 below, there is a diagnostic block which defines a split between ABCD and EF. One percentage, however, the 76% between C and E, is inflated. This indicates a convergence relationship between C and E. It is an isolated instance of convergence in a whole matrix of divergence relationships.

SPORADIC CONVERGENCE IN A DIVERGENCE PATTERN

	A	B	C	D	E
B	90				
C	81	81			
D	73	73	73		
E	67	67	76	67	
F	67	67	67	67	81

Figure 20

The instance of sporadic convergence in a divergence pattern is unique among the convergence patterns in that the subgrouping of speech groups can still be inferred, whereas with the other convergence patterns no complete subgrouping is possible. For instance, in Figures 13, 14, and 15 which illustrate the basic convergence pattern, it is impossible to say whether A split from BC, or AB split from C, or A, B, and C split simultaneously. Thus no subgrouping may be posited. In Figure 20, however, we can infer that ABCD split from EF and that the convergence relation between C and E is a secondary development. We may therefore posit two subgroups - ABCD and EF.

A sporadic convergence can also appear in a convergence pattern like chaining, however. In this case we can look behind the sporadic convergence to see the underlying pattern, but still no subgrouping of the chain is possible. Figure 21 represents a chaining pattern. The 81% between A and E is inflated. This represents a disproportionately strong convergence relation between A and E.

SPORADIC CONVERGENCE IN A CHAINING PATTERN

A					
90	B				
81	90	C			
73	81	90	D		
81	73	81	90	E	
60	67	73	81	90	F

Figure 21

4.4 Convergence patterns between subgroups

Thus far we have discussed the different convergence patterns as they exist between individual speech groups. They can be seen to exist between subgroups of speech groups as well. Consider the matrix in Figure 22 below.

THE BASIC CONVERGENCE PATTERN AMONG SUBGROUPS

In terms of time depths

A					
1	B				
2	2	C			
2	2	1	D		
3	3	2	2	E	
3	3	2	2	1	F

In terms of cognate percentages

A					
90	B				
81	81	C			
81	81	90	D		
73	73	81	81	E	
73	73	81	81	90	F

Figure 22

We first spot the two by two diagnostic block of 73's. As discussed in section two, this indicates a split between the speech groups defining the rows of the block and those defining the columns. In this case, a split of A and B from E and F is indicated. However, C and D are still unaccounted for. We see that the relationships of C to A, B, E, and F, are identical to those of D to the same four groups. Thus we can assume, that at one time C and D were the same speech group. The data then indicate three early speech groups: AB, CD, and EF.

We will now reduce the matrix to what the relations between the subgroups AB, CD, and EF are. To determine the relations of the subgroup AB, the columns A and B are averaged and combined into one column. Then the rows A and B are averaged and combined into one row. The cell where A and B intersect disappears in the process. This process is repeated for all the subgroups in the matrix. Figure 23 illustrates the reduction of the matrix in Figure 16. First the columns are reduced, and then the rows.

MATRIX REDUCTION

A					
90	B				
81	81	C			
81	81	90	D		
73	73	81	81	E	
73	73	81	81	90	F

column reduction $\Longrightarrow$

A	
B	
81	C
81	D
73	81	E
73	81	F

row reduction $\Longrightarrow$

AB		
81	CD	
73	81	EF

Figure 23

The resulting reduced matrix is the familiar basic convergence pattern discussed in section three. This time the convergence is between groups of speech groups rather than individual speech groups. We may now expand the generalization previously made about extracting diagnostic blocks -- the groups which have diverged are those which define the columns in the block, as against those which define the rows. We now add: any speech groups that remain unaccounted for represent the central group in a convergence pattern. Figure 24 illustrates.

EXTRACTING THE DIAGNOSTIC BLOCK IN A BASIC CONVERGENCE PATTERN

	A	B	C	D	E	F
B	90					
C	81	81				
D	81	81	90			
E	73	73	81	81		
F	73	73	81	81	90	

indicates

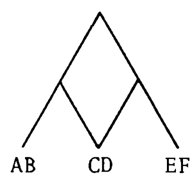


Figure 24

Matrix reduction may reveal any of the three other convergence patterns, as well as a divergence pattern. For instance, it is possible that a subgroup of languages could have a sporadic convergence relation with another subgroup during the dispersal of a whole language family or stock. A whole subgroup of languages could also show a dominance relation in a similar dispersal situation. McElhanon (1970:224) gives the reduction of a matrix of the relations between 39 dialects of the Huon peninsula to the relations between the eight major subgroups they divide into. This matrix is given below in Figure 25. It shows a clear chaining pattern near the level of language families. Note that the groups LM and NOP show spordic convergence. F and NOP may also show sporadic convergence.

CHAINING BETWEEN HUON PENINSULA SUBGROUPS

ABCDE							
33	LM						
44	41	F					
36	37	60	GI				
29	31	41	58	HJK			
26	38	37	34	32	NOP		
22	27	30	32	33	38	QRS	
17	20	24	27	26	27	34	TU

Figure 25

5. FINDING THE PATTERNS IN A MATRIX

We have discussed five patterns of divergence and convergence and how they may be recognized in a matrix of cognate percentages.

(1) Divergence - recognized by a block of adjacent lower and equal values (a diagnostic block).

(2) Basic convergence - recognized by two higher values and one lower one.

(3) Chaining - recognized by the highest values on the diagonal and the lowest values in the corner.

(4) Dominance - recognized by a row or column of values which are consistently higher than any of the other rows or columns.

(5) Sporadic convergence - recognized by a value which is abnormally high in the midst of lower values.

We also noted that these same five patterns may be found to exist between subgroups if the matrix is reduced to the average relations between subgroups.

When a matrix of lexicostatistic relations is first compiled, these patterns may not be apparent. To find the patterns, one must permute, that is, change the order of, the rows and columns of the matrix until the relevant patterns in the matrix appear. Figure 26 illustrates a matrix permutation. The order of B and D has been switched in the second matrix. Note that a chaining pattern appears after the matrix is permuted. The relations haven't been changed, only the order of the display.

A MATRIX PERMUTATION

Original Matrix

A			
73	B		
81	90	C	
90	81	90	D

With B and D permuted

A			
90	D		
81	90	C	
73	81	90	B

Figure 26

A very helpful aid in determining how to permute the matrix is to first compute the average shared cognates of each speech group. The average shared cognates of a speech group is the average of the cognate percentages which that speech group has with all other groups in the matrix. Thus if A has 90% cognates with B, 80% cognates with C, and 70% cognates with D, the average shared cognates for A is (90% + 80% + 70%)/3 or 80%.

The average shared cognates relates to the divergence and convergence patterns in the following ways. In a divergence pattern, the first speech group to diverge from the rest will show the lowest average shared cognates. This is illustrated in Figure 27. The average shared cognates for each speech group is represented by the number to the right of the corresponding row and column label.

AVERAGE SHARED COGNATES AND THE DIVERGENCE PATTERN

The Matrix

A 81				
90	B 81			
81	81	C 78		
73	73	73	D	73

The Tree

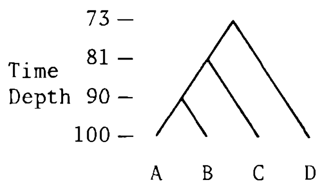


Figure 27

In a chaining pattern, the speech groups in the center of the chain show the highest average shared cognates and the speech groups at the ends of the chain show the lowest average shared cognates. This is illustrated in Figure 28.

AVERAGE SHARED COGNATES AND THE CHAINING PATTERN

The Matrix

A	78				
90	B	84			
81	90	C	86		
73	81	90	D	84	
67	73	81	90	E	78

The Tree

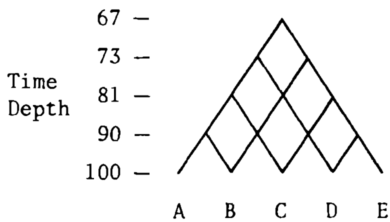


Figure 28

We note here that the speech group in the center of the matrix has the highest shared cognates and the speech groups on the edges of the matrix have the lowest average shared cognates. The overall pattern for the average shared cognates is that starting at one edge they get successively larger until the center is reached and then they get successively smaller.

In a dominance pattern, the central speech group will show the highest average shared cognates. This is illustrated in Figure 29.

AVERAGE SHARED COGNATES AND THE DOMINANCE PATTERN

The Matrix

A	90			
90	B	79		
90	73	C	81	
90	73	81	D	81

The Tree

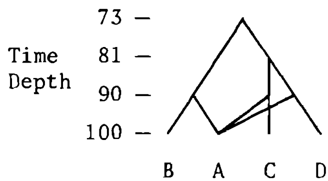


Figure 29

A sporadic convergence generally cannot be spotted by the average shared cognates. It is found by comparing the suspect percentage to the adjacent percentages in the matrix. When a sporadic convergence is found, however, it is generally good to recompute the average shared cognates for the two groups as though the inflated convergence percentage were normal. A sporadic convergence tends to inflate the average shared cognates and thus makes it less clear what the underlying pattern is.

6. EXAMPLES

I will now illustrate the above methods for recognizing patterns of divergence and convergence by analyzing some matrices that have been published in the literature on Papua New Guinea languages. In offering these analyses and tree diagrams, I in no way mean to suggest that this is the way it really was historically. I mean only to suggest that if we take the lexicostatistic data at face value, these are the conclusions indicated. To make strong historical inference requires corroborative evidence from the phonologies, grammars, oral traditions, and external history.

6.1 Kiwai family

To illustrate a divergence pattern with sporadic convergence, the following matrix is taken from S.A. Wurm's *Notes on the linguistic situation in the Trans-Fly Area* (1971:139). It gives the lexicostatistic relations between members of the Kiwai family. The average shared cognates is given to the right of the language names.

KIWAI FAMILY

Island Kiwai 62					
66	Wabuda 59				
66	60	Sisiame 65			
58	59	84	Pirupiru 63		
58	52	50	52	Kerewo 53	

Figure 30

We have seen that the optimal ordering of the matrix generally puts the lowest numbers in the corner and the highest numbers on the diagonal. Also, the groups with the lowest average shared cognates generally go on the edges. For these reasons we will swap Island Kiwai and Wabuda. The resulting matrix is shown in Figure 31. Figure 32 shows the diagnostic blocks that can be extracted from the matrix.

KIWAI FAMILY

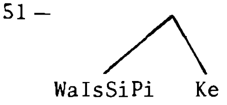
Wa 59				
66	Is 62			
60	66	Si 65		
59	58	84	Pi 63	
52	58	50	52	Ke 53

Figure 31

DIAGNOSTIC BLOCKS EXTRACTED FROM FIGURE 31

Wa	Is	Si	Pi	
52	58	50	52	Ke

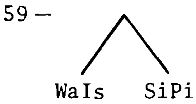
indicates



with sporadic convergence between Is and Ke

Wa	Is	
60	66	Si
59	58	Pi

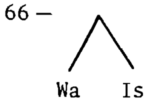
indicates



with sporadic convergence between Is and Si

Wa	
66	Is

indicates



Si	
84	Pi

indicates

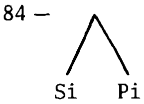


Figure 32

We see that all the basic patterns are divergence, with two sporadic convergences. Figure 33 gives the resulting tree. The sporadic convergence relations are marked with the double-headed arrows. Again I must emphasize that I do not mean to suggest that these are the five historical relations. I am suggesting only that this is the analysis indicated by these lexicostatistic data. If one chooses not to regard the 6% difference between 52% and 58%, or between 60% and 66% as significant, then a different analysis will result.

KIWAI FAMILY: TREE DIAGRAM

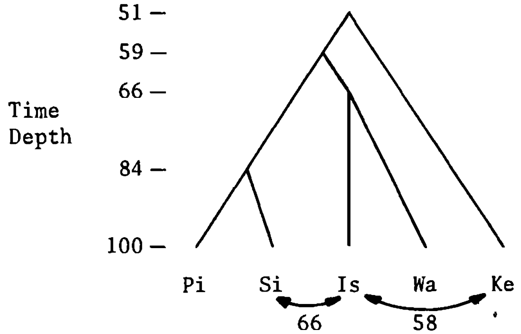


Figure 33

6.2 Dialects of Southern Kiwai

The following matrix of the relations between five dialects of the Southern Kiwai language is given to illustrate a chaining pattern. The chaining pattern does not show up clearly in the matrix as published (Wurm 1973:221). This matrix is reproduced in Figure 34 with the average shared cognates.

DIALECTS OF SOUTHERN KIWAI

SC	86				
91	EC	87			
85	83	DA	83		
84	89	82	IK	86	
82	84	82	88	DO	84

Figure 34

In a chaining pattern the average shared cognates is the highest in the middle of the chain and the lowest on the two edges. In this case, the lowest average shared cognates is in the middle of the matrix. Thus DA needs to be moved to the edge. The other clue is that the lowest percentages should be in the corner. Thus the two 82's in the DA column need to be moved to the edge. SC and EC will be moved one column to the right, and DA will be moved to the left most column. The resulting matrix is shown in Figure 35. A perfect chaining pattern becomes apparent in the data.

DIALECTS OF SOUTHERN KIWAI

DA	83				
85	SC	86			
83	91	EC	87		
82	84	89	IK	86	
82	82	84	88	DO	84

Figure 35

6.3 Tinputz and Petats families

These data illustrate the basic convergence pattern both between individual speech groups and between subgroups. A dominance relation may be seen also. The matrix, taken from Allen and Hurd (1965:21), is given in Figure 36 with average shared cognates.

TINPUTZ AND PETATS FAMILIES

Teop	38						
53	Hahon	35					
67	71	Timputz	36				
21	16	17	Halia	30			
17	15	16	34	Solos	30		
27	19	15	56	65	Petats	37	
40	36	28	34	30	40	Saposa	35

Figure 36

The first thing we see is the 3 by 3 diagnostic block of low numbers between the first three languages and the second three languages. This should be placed in the corner. This is done by inserting Saposa between Timputz and Halia, as in Figure 37.

Te	38						
53	Hh	35					
67	71	Ti	36				
40	36	28	Sa	35			
21	16	17	34	Hl	30		
17	15	16	30	34	So	30	
27	19	15	40	56	65	Pe	37

Figure 37

By the diagnostic block we see that there are two subgroups of three languages each. To put the subgroups in their optimal ordering we put the average shared cognates in ascending order from edge toward center. This involves moving Te and Pe towards the center as shown below in Figure 38. The diagnostic patterns are extracted in Figure 39, and the tree assembled in Figure 40. Nothing but the most general inference about linguistic history could possibly be made from this tree. As illustrated in Figure 15, there are three possible historical explanations of a basic convergence pattern. In this tree, containing three basic convergence patterns, there is a whole host of possible histories. Without corroborating evidence the tree must remain multiply ambigous.

Hh	35						
71	Ti	36					
53	67	Te	38				
36	28	40	Sa	35			
19	15	27	40	Pe	37		
16	17	21	34	56	Hl	30	
15	16	17	30	65	34	So	30

Figure 38

DIAGNOSTIC PATTERNS EXTRACTED FROM FIGURE 38

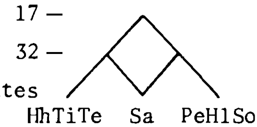
Hh	Ti	Te	
36	28	40	Sa

19	15	27	40	Pe
16	17	21	34	Hl
15	16	17	30	So

reduces to

HhTiTe		
35	Sa	
17	35	PeHlSo

which indicates



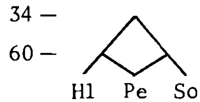
with sporadic convergence of Te and Pe, and perhaps Te with Sa and Pe with Sa

Pe		
56	Hl	
65	34	So

equivalent to

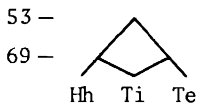
Hl		
56	Pe	
34	65	So

which indicates



Hh		
71	Ti	
53	67	Te

indicates



Te	Sa		
27	40	Pe	
		56	Hl
		65	So

indicates dominance of Pe

Ti		
67	Te	
	40	Si
	27	Pe
	21	Hl

indicates dominance of Te

Figure 39

TIMPUTZ AND PETATS FAMILIES TREE

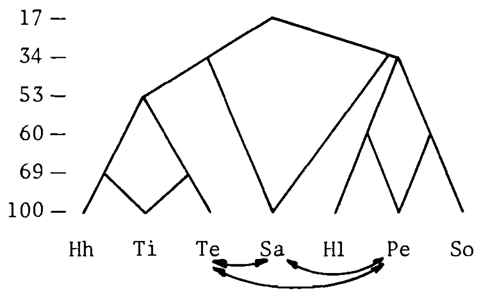


Figure 40

6.4 Ndu family

As a final example we shall look at the Ndu family as reported by Laycock (1965:181). The original matrix is reproduced in Figure 41 with average shared cognates.

NDU FAMILY

Ngala 33							
31	Nyaura 45						
33	49	Manambu 46					
33	48	52	Yelogu 48				
33	48	43	45	Yangoru 49			
33	45	43	46	62	Kwusaun 48		
35	45	48	55	56	55	Maprik 53	
33	50	52	56	55	54	78	Wosera 54

Figure 41

By the average shared cognates we see that Maprik and Wosera should be swapped to the center of the matrix as shown in Figure 42. The diagnostic patterns are extracted in Figure 44 and the tree assembled in Figure 45.

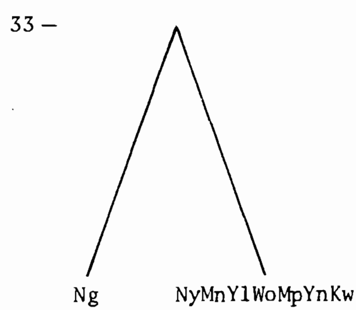
Ng 33							
31	Ny 45						
33	49	Ma 46					
33	48	52	Yl 48				
33	50	52	56	Wo 54			
35	45	48	55	78	Mp 53		
33	48	43	45	55	56	Yn 49	
33	45	43	46	54	55	62	Kw 48

Figure 42

DIAGNOSTIC PATTERNS EXTRACTED FROM FIGURE 42

Ng	
31	Ny
33	Mn
33	Yl
33	Wo
35	Mp
33	Yn
33	Kw

indicates



Ny		
49	Mn	
48	52	Yl
50	52	Wo
45	48	Mp
48	43	Yn
45	43	Kw

indicates

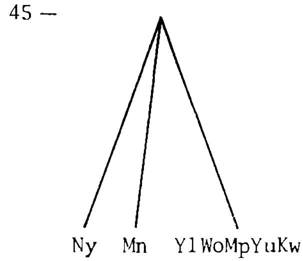


Figure 43

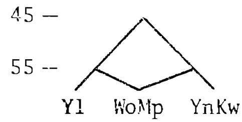
At first glance the above block appears diagnostic of NyMn versus YlWoMpYnKw, but since the relation between Ny and Mn is as low as all others in the block it must represent a three way split. The difference between values (i.e. 43, 45, 48, 49) are too insignificant to posit a more refined relationship.

Yl			
56	Wo		
55		Mp	
45	55	56	Yn
46	54	55	Kw

reduces to

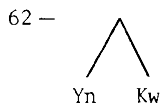
Yl		
55	WoMp	
45	55	YnKw

which indicates



Yn	
62	Kw

indicates



Wo	
78	Mp

indicates

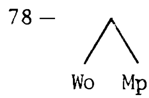


Figure 44

NDU FAMILY TREE

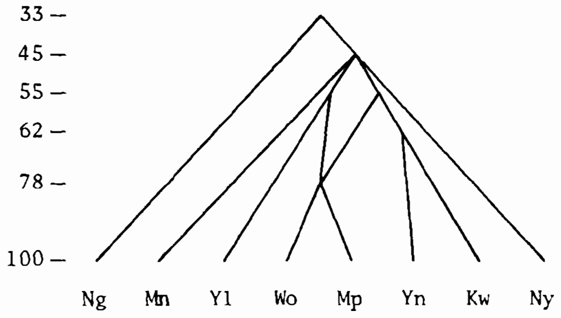


Figure 45

The symbolization involving Wo and Mp may prove confusing. It does not indicate that Wo split from Yl and Mp split from Yn, then Wo and Mp converged only to split again. It represents the basic divergence pattern between the three groups Yl, WoMp, and YnKw. Until the split at 78%, WoMp is a unity.

This tree differs in a few points from that given by Laycock (1965:187). I don't feel that the lexicostatistic figures are significantly different enough to posit three stages in the dispersal of Mn, YlWoMpYnKw, and Ny (of Laycock's *Beta, *Gamma, *Delta). Laycock gives a straight divergence pattern for the splitting of Yl, WoMp, and YnKw, with YnKw splitting from YlWoMp. However, the data show that WoMp is as equally related to YnKw (45) as it is to Yl (45). Thus we must posit a convergence pattern for the dispersal of Yl, WoMp, and YnKw.

NOTES

[1] I am indebted to T. Wayne Dye for first introducing me to the idea of a convergence model for lexicostatistics. I wish to thank Robert Conrad for his helpful comments on an earlier draft of the paper.

[2] The patterns of divergence and convergence were discovered and explored through experimenting with computer programs executed on the ETP 8L Electronic Text Processor. The divergence pattern was explored by experimenting with a program written to simulate the Swadesh model of lexicostatistic divergence. The convergence patterns were discovered and explored by using a matrix permutation program to analyze dozens of lexicostatistic matrices.

[3] Here, and throughout the rest of the paper, I use "speech group" to mean any type of linguistic grouping, whether it be communalect, dialect, language, or even language family.

[4] The special case of a three-way split should be mentioned. In this divergence pattern, for the minimal case of three speech groups, all three percentages in the matrix are equal. The diagnostic feature for the three-way split is then an L-shaped pattern of equal percentages.

[5] My wife, Linda, and I collected these data during a period of field work in the Biliau language, September - October 1976.

BIBLIOGRAPHY

Allen, Jerry and Conrad Hurd, 1965. *Languages of the Bougainville District*. Port Moresby: Department of Information and Extension Services.

Conrad, Robert and T. Wayne Dye, 1975. Some language relationships in the Upper Sepik region of Papua New Guinea. *Pacific linguistics A* 40.

Cook, Edwin A., 1966. Narak: language or dialect? *Journal of the Polynesian society* 75:437-44.

Dyen, Isidore, 1963. Lexicostatistically determined borrowing and taboo. *Language* 39:60-66.

Lamb, Sydney M. Some proposals for linguistic taxonomy. *Anthropological linguistics* 1(2):33-49.

Laycock, D.C., 1965. *The Ndu language family (Sepik District, New Guinea)*. *Pacific linguistics C* 1.

Lees, Robert B., 1953. The basis of glottochronology. *Language* 29:113-27.

McElhanon, K.A., 1970. Lexicostatistics and the classification of Huon Peninsula languages. *Oceania* 40:214-31.

———, 1971. Classifying New Guinea languages. *Anthropos* 66:120-44.

Simons, Gary, 1977. Toward a convergence model for lexicostatistics. Ms.

Wurm, S.A., 1961. The linguistic situation in the Highlands Districts of Papua and New Guinea. *Australian territories* 1(2):14-23.

———, 1971. Notes on the linguistic situation in the Trans-Fly area. *Pacific linguistics A* 28:115-72.

———, 1973. The Kiwaian language family. In *The linguistic situation in the Gulf District and adjacent areas, Papua New Guinea*, edited by Karl Franklin. *Pacific linguistics C* 26:219-60.