NUMERICAL TAXONOMY AND THE ANALYSIS OF MANUSCRIPT RELATIONSHIPS

KATHERINE O'BRIEN O'KEEFFE
Texas A&M University

ALAN R. P. JOURNET
Southeast Missouri State University

This study explores the application of numerical taxonomy to the analysis of relationships among manuscripts of a single text, in this case, Aldhelm's *Aenigmata.* The *Aenigmata,* a collection of one hundred Anglo-Latin riddles of the late seventh century, is preserved either whole or in part in twenty-eight surviving manuscripts which range in date from the eighth through the fifteenth centuries and come from scriptoria in England and throughout the continent. Its complex manuscript tradition, which makes traditional stemmatic analysis unsatisfactory, provides a useful illustration of classification by numerical taxonomy, a mathematical technique which has been developed to classify operational taxonomic units, that is, any group of items which exhibit quantitative or qualitative differences.

Rudolf Ehwald, the foremost editor of the text, did not attempt a stemma in his 1919 edition of Aldhelm's works.[1] Indeed, the stemma as produced by Lachmannian analysis has been seriously criticized as a method of analyzing manuscript relationships. It is liable, as E. Talbot Donaldson has charged, to the conscious, or unconscious, subjective judgments of the editor.[2] Just as liable to subjectivity is Dom Henri Quentin's technique of grouping manuscripts by threes, since the editor is free to choose how many agreements against disagreements are significant.[3] For traditional methods, the more con-

---

[1] Rudolf Ehwald, ed., *Aldhelmi Opera,* Monumenta Germaniae Historica, Auc. Antiq. 15 (Berlin: Wiedmann, 1919). The most recent editor, Fr. Glorie (*Aenigmata Aldhelmi,* CCSL, 133, pt. 1 [Turnhout: Brepols, 1968]), does not comment on the textual tradition. All references to the text of the *Aenigmata* are to Ehwald's edition unless otherwise specified.

[2] E. T. Donaldson, "The Psychology of Editors of Middle English Texts," in his *Speaking of Chaucer* (New York: Norton, 1970), p. 107. For a description of the Recensionist method as an answer to its critics, see Leonard E. Boyle, O.P., "Optimist and Recensionist: 'Common Errors' or 'Common Variations'?" in John J. O'Meara and Bernd Naumann, eds., *Latin Script and Letters A.D. 400-900* (Leiden: E. J. Brill, 1976), pp. 264-74.

[3] See Vinton A. Dearing, *Principles and Practice of Textual Analysis* (Berkeley: Univ. of California Press, 1974), pp. 11-12. See also G. P. Zarri, "Algorithms, *Stemmata Codicum* and the Theories of Dom H. Quentin," in A. J. Aitken, et al. eds., *The Computer and Literary Studies* (Edinburgh: Edinburgh Univ. Press, 1973), pp. 227-28.

Manuscripta, XXVII (1983), 131-145.

taminated a textual transmission, the more obviously necessary are editorial decisions to group manuscripts and the more subject these decisions are to error.

Ehwald's grouping of manuscripts of the *Aenigmata* demonstrates the problem of sorting out the relationships among manuscripts which transmit a contaminated tradition. In Figure 1, an illustration of the major groupings in the Preface to his edition of the *Aenigmata,* it is obvious that P^1, E, F^1, and F^2 shift groups depending on the basis for classification.[4] These difficulties, overlapping groups and variable bases for classification, present a problematic picture of relationships which might be illuminated by numerical taxonomy.

Numerical taxonomy offers several distinct advantages as a method of classifying manuscripts. The technique graphs the relationships among extant copies of a text in a phenogram. (Such a graph is non-directional and does not attempt to predict an "ancestral" form of the text.) Since the technique is better applied to a text with more than 100 variations, it is ideally suited to texts with contaminated or complex traditions. The greater the number of variations, the more consistent the results. Because the technique is computer-assisted, it can produce, rapidly and cheaply, a display of relationships not only for the full text but for any portion of it. Such flexibility allows discrete phenograms to be produced which would confirm or deny, for example, a suspected change in exemplar. The technique further allows for the placement of fragments within the phenogram. Since the mathematics of numerical taxonomy is rigorously objective and complete in its analysis, it meets the primary objections to stemmas constructed on "shared error." A further advantage is that numerical taxonomy may raise questions about the relationships among manuscripts analyzed by conventional means and suggest further areas to explore.

Before discussing the application of the technique in detail, we will present an outline of the steps involved in which we will simplify the problem by reducing the number of manuscripts under consideration to five (V,W,X,Y,Z), the number of variations to ten (1, 2, . . . 9, 10), and the possible variants for each variation to four (A,B,C,D). All variations are given equal importance.

The manuscripts are examined one variation at a time. Each variant per variation is assigned a letter (A,B,C, or D). These data are recorded in a matrix of manuscripts and variations (Figure 2). The manuscripts are taken a pair at a time and a similarity co-

[4] Ehwald, *Aldhelmi Opera,* pp. 41-56.

efficient (S) is calculated for each pair: $S = \frac{m}{m + r - i}$ when m is the number of matching variants, r is the number of remaining variants, i is the number of impossible comparisons, and $(m + r)$ is the total number of possible comparisons. Those variations for which one or both manuscripts either lack a variant or have an indecipherable one are recorded under i.

The range of pair-wise manuscript similarity coefficients so computed is then displayed as a matrix (Figure 3). The diagonal (V-V, W-W, etc.) contains only 1 since these comparisons are identical. The comparison V and W over 10 variations shows 5 matching variants, leaving 5 remaining and zero impossible comparisons ($\frac{5}{5+5} = .5$) V and Z, meanwhile, show 2 matching and 8 remaining with 1 impossible comparison ($\frac{2}{2+8-1} = .22$). The procedure adopted here is the sample matching coefficient of Sneath and Sokal, being the simplest similarity coefficient for data sets where variations may have several (i.e. more than two) variants.[5]

This matrix of similarities is generated in a manner similar to that outlined by Colwell.[6] Now, however, the procedure is taken considerably further than Colwell's matrix. Here we perform a cluster analysis which produces a visual graphic display of the family relationships among manuscripts.

The first step is to identify the highest similarity coefficient not in the diagonal. In this example, the highest coefficient is V-W at 5. These manuscripts are now combined as a new unit V/W. Similarity coefficients between V/W and the remaining manuscripts are now recomputed as the average of V and W each against each of the remaining manuscripts (Figure 4). Thus, V against X was .2 and W against X was .1, so V/W—X is not .2 + .1 ÷ 2 = .15. The remaining coefficients (V/W—Y, V/W—Z) are computed in identical fashion.

Again the matrix (Figure 4) is examined for the highest similarity coefficient. In this example it is Y—Z (.33) which now forms the

---

[5] Peter H. A. Sneath and Robert R. Sokal, *Numerical Taxonomy* (San Francisco: W. H. Freeman, 1973). A discussion of the influence of different similarity coefficients on the pattern of manuscript relationships appears in Alan R. P. Journet and Katherine O'Brien O'Keeffe, "Exploring the Application of NT-SYS for Grouping Manuscript Records of a Medieval Text," in Sarah K. Burton and Douglas D. Short, eds., *Sixth International Conference on Computers and the Humanities* (Rockville, Md.: Computer Science Press, 1983), pp. 311-317.

[6] Ernest C. Colwell, "Method in Establishing Quantitative Relationships Between Text-Types of New Testament Manuscripts," in his *Studies in Methodology* in *Textual Criticism of the New Testament* (Grand Rapids: Eerdmans, 1969), pp. 56-60.

unit Y/Z, and a new matrix is computed with this unit compared to each other manuscript and unit as outlined above (Figure 5). Again, the highest similarity coefficient is identified, X—Z/Y at .165. This is now combined as the unit X/Y/Z, with the final matrix being calculated (Figure 6).

There are several techniques available for clustering.[7] That discussed in this example is the unweighted pair group method using arithmetic averages (UPGMA). One alternate technique is the weighted pair group method using arthmetic averages (WPGMA). To illustrate the difference we can examine the final stage of clustering indicated above. Using the UPGMA, we compute the similarity coefficients of the matrix (Figure 6) as follows: X/Y/Z against V/W is the average of X—V, X—W, Y—V, Y—W, Z—V, and Z—W = approximately .155. Using WPGMA, On the other hand, this similarity coefficient would be the average of Y/Z—V/W and X—V/W = .1625.

The point to make here, without dwelling on the full range of clustering techniques available, is that different techniques produce slightly different similarity coefficients which may produce slightly different family groupings. There are no clear rules for determining which technique to use. The most frequently employed technique, however, is the UPGMA used in this study.

The final step in this process is to display the manuscripts and their connecting points (unifying similarity coefficients) as a graphic phenogram (Figure 7). This is an objective display of the relationships among the manuscripts under examination. In the phenogram, the units to be clustered, the manuscripts, appear in an array along the top axis with a line descending from each. The vertical axis represents the degree or level of similarity among manuscripts over the set of variants employed in the analysis. The horizontal lines, connecting, near the top, individual manuscripts and, lower down, groups of already clustered manuscripts, locate the level of similarity exhibited by these manuscripts or clusters of manuscripts. The range of similarity is from 1, implying two manuscripts are identical, to zero, implying complete dissimilarity. The levels of similarity of manuscripts or groups of manuscripts displayed here is that cal-

[7] Cf. Sneath and Sokal, *Numerical Taxonomy*, passim. An examination of different clustering techniques and their effects on manuscript relationships appears in Journet and O'Keeffe, ICCH p. 312. For an analysis of manuscripts using ordination see John G. Griffith, "Non-Stemmatic Classifications of Manuscripts by Computer Methods," in *La Pratique des Ordinateurs dans la Critique des Textes*, Colloques Internationaux du C.N.R.S., 579 (1973), pp. 73-86.

culated by the clustering program, NT-SYS.[8]

The technique of numerical taxonomy does not render the editor or paleographer redundant. It is not a substitute for decision making since judgment must be exercised not only beforehand in the construction of the algorithm but continually in choosing variations, variants, and their weights. Numerical taxonomy's claim to objectivity lies in its ability to reduce, as far as possible, individual subjective judgments by setting out rules for decision-making in advance. Once decisions on variations, variants, and weights are completed, the mathematics of the technique is purely objective.

In the identification of variations we distinguish two classes of variants.[9] Class I, those variants which conceivably are the product of a scribe's conscious choice, comprises grammatical changes; substitution of words, and shifts in word order. Variants produced by grammatical change would include, for example, the alternation of "spoliabor" and "spoliabar" in *Aenig.*, 100, 52, "Pallida dum iugiter dapibus spoliabor opimis." A clear example of a substitution of words occurs in *Aenig.*, Pref., 18, "Iamdudum cecinisse prisci vexilla tropei," where some manuscripts read "prisci," others read "celebris," and others "almi." The variation "munera"/"numera" in *Aenig.*, Pref., 15, "Inspirans stolidae pia gratis munera menti" illustrates the importance of adhering to the categories of the algorithm. Although only one manuscript reads "numera," and although "numera" quite possibly is the result of a reversal of "n" and "m," nevertheless, because "numera" is grammatically appropriate and a legitimate word, it must be counted a substitution in words, not an accident. Possible causes of variants are irrelevant here. The last variant in this class is a shift in word order, here illustrated by *Aenig.*, 100, 7, "Nam Deus ut propria mundum dicione gubernat," where a group of six manuscripts reads "mundum propria."

Class II, those variants which most probably were produced

---

[8] NT-SYS is a program package written to perform a wide range of multivariate statistical manipulations involved in numerical taxonomy. Developed by F. James Rohlf, State University of New York at Stony Brook, John Kishpaugh and David Kink of Quantra Development Corporation, the package employs a simple control language and performs these computations rapidly and, therefore, cheaply. All the computations presented and discussed in this paper were performed for a total cost of less than five dollars on the Texas A&M University computer system consisting of an Amdahl/470-V8 and an Amdahl/470-V6-2. Computer file storage of the data costs a fraction of a cent per month.

[9] A variation in a text is composed of two or more variants. See Dearing, *Principles*, p. 6.

unconsciously, includes differences in spelling and scribal errors (i.e. transpositions of letters, haplography, dittography, separation/combination, and so on). For example, in *Aenig.*, Pref., 4, "Horrida nam multans torsisti membra Vehemoth," the variation "vehemoth"/"beemoth"/"behemoth"/"behemoh" is a result of spelling practice. The reading "pandererum" in P^1 and P^4 for "pandere rerum" (*Aenig.*, Pref. 7) is a nonsense word produced by combination and eyeskip. These two classes are tied to the system of weighting in the procedure. To avoid skewing the relationships among the manuscripts, we have adjusted for local spelling habits, routinely ignoring alternations of "i" and "y," "e" and "æ," and the presence or absence of initial "h."

Two further principles in the definition of variations and variants must be mentioned. First, the number of variants within a variation must be kept as small as possible, even if to do so requires introducing extra variations. As the number of variants per variation approaches the number of manuscripts, clustering becomes impossible. Second, although Ernest Colwell has argued against the inclusion of any variation where there is almost complete agreement among manuscripts, we have chosen to include such variations on the principle that such oddities, while useless to the textual editor, are important clues for the paleographer.[10]

For this investigation in adapting numerical taxonomy, we have limited the data to textual differences in order to compare the phenograms with Ehwald's groupings. However, the technique is equally adaptable to additional kinds of data and can accommodate variations which include paleographic information such as columns per page, marks of punctuation, gaps in the text, foliation, lineation, and so on.

All variations may not be equally significant. The analyst of text and manuscripts may wish to focus on the relationships among manuscripts *qua* records of texts or may wish to focus on the relationships among significant textual differences. To reflect the differing importance of variations, a numerical taxonomy may assign weights to classes of variants. Although there is no theoretical limit to the number of different weights, parsimony urges few classes. Further constraints require that categories for weighting be objective, spelled out beforehand, and not subject to debate. This project uses three weighting systems we have devised to accommodate the classes of variants. In method A, all variants are equally important, from grammatical changes to nonsense words. Each has a value of 1. In

[10] Colwell, *Studies*, p. 57.

method *B*, only those variations in class I are considered. The third method of weighting (C) compromises between these extremes by considering all variants but by assigning variants of class I a double weight.

Aenigmata 100, 11. 5-7 demonstrates the principles of selection we have followed:

Pervigil excubiis: numquam dormire iuvabit,
Sed tamen extemplo clauduntur lumina somno;
Nam Deus ut propria mundum dicione gubernat. . . .

1. 5, excubiis; excupiis V.
iuvabit/ A,N,P,O,L,S^2,V,P^4,F,F^1,F^2; iuuauit K,B; iuuabat L^1; iubavit P^1; iuabit P^3; valebo B^1,C,[E]; erasure E.

1. 6, extemplo/ A,K,N,B,L^1,L,S^2,V,P^3,F; extimplo B^1,C,O,E,P^1,P^4, F^1,F^2; extymplo P. clauduntur/ codd.; cluduntur K; claduntur P^4.

1. 7. propria mundum/ A,K,N,B,B^1,C,L^1,S^2,V,P^3,F,F^1,F^2; mundum propria P,O,E,P^1,L,P^4.
gubernat/ codd.; gubernabat B (L^1 undecipherable).

This apparatus reflects all the variants in the three lines across all of the manuscripts used.

The first variation, "excubiis"/"excupiis," reflects a difference in spelling (or, one might argue, a letter reversal). As such, it is a member of class II and is weighted 1 in methods *A* and *C* and is disregarded in method *B*. The alternation of "iuuabit"/"iuuavit"/"iuuabat"/"iubavit"/"iuabit"/"valebo" presents an interesting problem insofar as it combines different classes of variants. On the principle of parsimony, we set up three variations. In the first, the variants are the choice of either "iuuo" or "valeo." Hence "iuuo": A,N,P,O,L,S^2V,P^4,F,F^1,F^2/"valeo" B^1,C,[E].[11] Manuscripts K,B,L^1, and P^3 have no entries because their variants are nonsense words and to interpret them would be to classify by assertion. E receives no entry because it has been erased at this point. This variation is composed of a substitution of words (class I) and is weighted 1 in method *A* along with all other variants, is weighted in method *B* exclusive of all variants in class II, and is given double weighting in the compromise method *C*. But the alternations on this word also comprise a variation based on grammatical form. We construct a second variation "iuua*bit*"/"iuva*bat*"/"vale*bo*" on the basis of per-

[11] The designation [E] refers to manuscript E, Oxford, Bodleian ms. Rawlinson C. 697, originally from northeast France (ix^3), as corrected in England sometime in the second half of the tenth century. In effect [E] becomes a separate manuscript (taxonomic unit).

missible grammatical form. Again, for K,P^3,B,L^1, and E we enter no comparison because of impossible grammatical forms and erasure respectively. This variation, also of class I, is weighted 1 in *A*, considered exclusively in *B*, and weighted 2 in *C*. The final variation comprises nonsense words. Here the variation is "iuuavit" K,B; "iubavit" P^1; "iuabit" P^3; for all other manuscripts we enter no comparison. This variation from class II is weighted 1 in *A* and *C* and is ignored in *B*.

The variations, "extimplo"/"extemplo," "excubiis"/"excupiis," and "clauduntur"/"cluduntur"/"claduntur," are composed of variant spellings and are weighted as members of class II. The variation "propria mundum"/"mundum propria" is a shift in word order and is weighted as a member of class I. Finally, the variation "gubernat"/"gubernabat" is weighted in class I as a grammatical change. Since the reading in L^1 is indecipherable, it is assigned the designation "no comparison."

There remains now only to discuss the phenograms and their relation to Ehwald's grouping of the manuscripts. Ehwald collated twenty-five manuscripts for his edition:

Leningrad, State Public Library, Q.v.I.15	A	(Lenin 0)
Leningrad, State Public Library, F.v.XIV.1	R	(Lenin 1)
Karlsruhe, Badische Landesbibliothek, Aug. LXXXV	K	(Karls)
Vatican, Pal. lat. 1753	N	(Naz)
London, B. L. Royal 15 A.xvi	B	(Brit 0)
London, B. L. Royal 12 C.xxiii	B^1	(Brit 1)
St. Gallen, Stiftsbibliothek, ms. 242	S^1	(Swiss 1)
St. Gallen, Stiftsbibliothek, ms. 1394	S	(Swiss 0)
Paris, B. N. lat. 16700	P	(Paris 0)
Leipzig, Karl Marx University, Rep. I/74	O	(Leipzig)
Oxford, Bodl. Rawlinson C.697	E	(Oxfd 0)
Paris, B. N. lat. 2339	P^1	(Paris 1)
Leiden, Bibliotheek der Rijksuniversiteit, Voss. lat. Oct. 15, XII	L	(Leid 0)
Bremen, Universitaet Bremen, ms. 52	S^2	(Swiss 2)
Vatican, Reg. lat. 2078	V	(Vatican)
Paris, B. N. lat. 7540	P^2	(Paris 2)
Paris, B. N. lat. 8440	P^4	(Paris 4)
Einsiedeln Ms. 302	e	
Cambridge, C. U. L. Gg. 5. 35	C	(Camb)
Paris, B. N. lat. 2773	P^3	(Paris 3)
Brussels, Bibliothèque Royale Albert I^{er} ms. 4433-4438	F	(Bruss 0)

Brussels, Bibliothèque Royale Albert I^{er} ms. 10615	F^1	(Bruss 1)
Brussels, Bibliothèque Royale Albert I^{er} ms. 9799	F^2	(Bruss 2)
Wolfenbuttel, Herzog August Bibliothek, Gudianus lat. 331	G	(Guelf)
Vatican, Pal. lat. 591	p	(Pal)

To these we have added Leiden, Voss. Q. 106 (L^1, Leid 1) and the Miskolc fragment (M, Misc.)[12] In collating we have limited ourselves to reading the text as originally copied or as corrected by the hand of the scribe. Later corrections are ignored. Thus the phenogram of the connection between L (Leid 0) and P^1 (Paris 1), while showing a fair degree of correlation, does not reflect that L, as Ehwald has rightly observed, was probably corrected from P^1. The specific exception to this rule is Oxford, Bodl. Rawl. C. 697 (E, Oxfd 0), whose corrected version ([E], Oxfd 1) was included to determine whether the revisions made in England would cause [E] to group separately. There was not a sufficient number of variants to cause a shift.

To demonstrate the adaptability of numerical taxonomy to the analysis of manuscript relationships, we have collated the beginning and end of the *Aenigmata:* the Preface through Riddle 7, and the last riddle, "Creatura." Not all manuscripts appear in both sets of phenograms since several manuscripts are either fragments or only transmit selections from the *Aenigmata.*

In general, the six phenograms (Figures 8–13) confirm Ehwald's judgments on the relationships among the manuscripts he used. The two English manuscripts, B^1 (Brit 1) and C (Camb), group.[13]

---

[12] Z. Mady, "An VIIIth Century Aldhelm Fragment in Hungary," *Acta Antiqua Academiae Scientiarum Hungaricae,* 13 (1965), 441-53. Cf. A. E. Lowe, ed., *Codices Latini Antiquiores,* II, suppl. (Oxford: Clarendon Press, 1972), item 1792. There are two omissions from our collation. The xerox of R (Lenin 1) we received arrived lacking the Preface, and we have been unable to obtain a copy of Einsideln 302. Fr. Glorie (CCSL, 133, pt. 1) lists two other manuscripts without folio numbers: Vatican, Ottob. 35 and Paris, B. N. lat. 8069, both probably from the list in Max Manitius, *Geschichte der Lateinischen Literatur des Mittelalters,* bd. 1 (1911; rpt. Munich: C. H. Beck, 1965), p. 138. Vat. Ottob. 35 is certainly a ghost. We have not seen Paris, B. N. lat. 8069. St. Gallen, Stiftsbibliothek, ms. 273 preserves only Riddle 100, l. 83, and we have omitted so brief a fragment.

[13] That is, except for Figure 12, weighting method *B.* Here C (Camb) groups at a low level, probably because C has been modified idiosyncratically by the scribe after f. 398^v.

F^1 (Bruss 1) groups with its apograph F^2 (Bruss 2); P^1 (Paris 1) and L (Leid 0) group as well. K (Karls), N (Naz), and F (Bruss 0) group in Figures 11—13 only. P (Paris 0) groups with P^4 (Paris 4). V (Vat), p (Pal), and 0 (Leipzig), all idiosyncratic copies, group at very low coefficients of similarity. Ehwald's judgment that the two Leningrad manuscripts, A and R (Lenin 0, Lenin 1), were not related, contrary to earlier analysis, is confirmed by the placement of R far away from A (Figures 8—10). Further, the grouping of M (Misc) and K (Karls) at a high degree of similarity in Figures 11—13 confirms Mady's argument that the Miskolc fragment was textually most similar to K.[14]

There are, however, two important respects in which the phenograms challenge Ehwald's conclusions. In his Preface to the edition, Ehwald noted that there were two recensions of the riddles and that A,F^1, and F^2 (Lenin 0, Bruss 1, Bruss 2) represented the final recension. Ehwald's assertion of similarity among these manuscripts is questionable, however, at least given the opening and closing sections of the text. The phenograms demonstrate that there is no close connection between A and F^1F^2 and illustrate as well that A is a highly individual manuscript which groups with other manuscripts at a very low level of similarity. There is no evidence of two clear recensions in these phenograms.

A second challenge to Ehwald's grouping is posed by the phenograms' presentation of the relationships among the line shift manuscripts. Ehwald grouped manuscripts E, L, O, P, P^1, and P^4 (Oxfd 0, Leid 0, Leipzig, Paris 0, Paris 1, Paris 4) on the basis of their placing lines 61—67 in Riddle 100 between 11. 43 and 44. The phenograms in Figures 11—13 show two groups: ELOP^1; L^{1}PP^4.[15] In Figures 8—10, E affiliates not with the other line shift manuscripts but with A (Lenin 0). This shift of families suggests a switch in exemplars, a suspicion supported by a clear change in scribal hand in E after l. 20 of f. 14^v.

The singlemost difficulty with the phenograms is posed by manuscripts which change groups. In part the different groupings between Figures 8—10 and Figures 11—13 can be traced to fragmentary manuscripts: R, P^2, G, p, and S^1 (Lenin 1, Paris 2, Guelf, Pal, Swiss 1) which, by disappearing after the early part of the text reflected in Figures 8—10, allow for new groupings in Figures 11—13. Shifts in exemplar may also be responsible for alternate

---

[14] Mady, "Aldhelm Fragment," p. 446.

[15] Leiden, Voss. Q. 106 (L^1, Leid 1), a manuscript unknown to Ehwald, exhibits the line shift.

groupings. Such is probably the case with E (Oxfd) and F (Bruss 0).[16] Where manuscripts group differently within a set of phenograms, as with S^1 and S^2 (Swiss 1, Swiss 2) in Figures 8–10, there may be two possible explanations. The first is a problem of distortion in the clustering which becomes pronounced when several manuscripts have almost identical coefficients of similarity. Comparing the phenograms to the similarity matrices may clarify relationships when distortion seems to have occurred.[17] The second explanation is more appropriately a reminder. Shifting affiliations may not be products of distortion, but may reflect a change in relationships determined by weighting different classes of variants. Where, for example, method *B* produces a significantly different phenogram from those produced by *A* and *C*, the researcher must carefully review precisely what it is that he or she wishes the phenograms to illustrate.

Numerical taxonomy is a useful and adaptable tool in the analysis of the relationships among manuscripts of a text. As a first step toward a new edition of an already edited text, the phenograms may ask questions of the previous stemma or choice of a base text and offer suggestions on the affiliations of previously uncollated manuscripts. In the first editing of a text, it is an appropriate starting place for the grouping of manuscripts. The rapid retrieval of information and the manipulation of subsets of information which the computer and the NT-SYS package make possible allow the researcher to examine and test data in a variety of ways both efficiently and inexpensively.[18]

---

[16] There is a change of scribe in F (Bruss 0) after f. 65v.

[17] If there is disagreement among phenograms in the location of manuscripts, distortion may be an explanation. During clustering two relatively similar manuscripts may become grouped with different clusters, thus exaggerating in the phenogram the difference exhibited in the matrix of similarity coefficients. This distortion is suggested if the level of connection of two manuscripts, or, rather, the clusters in which they have been grouped, is much lower in the phenogram than in the matrix of similarity coefficients. When distortion is suspected, the actual level of similarity of manuscripts is that exhibited in the matrix of similarity coefficients.

[18] Support for this project has been provided by a Texas A&M University Mini-Grant, an Andrew W. Mellon Grant to work in the Vatican Film Library at Saint Louis University, and a grant from the Southeast Missouri State University Grants and Research Funding Committee. Some aspects of this project have been presented in papers delivered at the St. Louis Conference on Manuscript Studies, the Sewanee Medieval Colloquium, and the Southeast Missouri State University Science Symposium. We would like to acknowledge Phil Spector, Institute of Statistics at Texas A&M University, for his assistance with programming problems.

KP^1SNF (Tract and *Aenigmata*)
$P^3B^1CEF^1F^2$ ("Cuiusdam Scoti de Alphabeto")
$EPLOP^1P^4$ (11. 100, 61-67 placed between 100, 41-44)
AF^1F^2 (final recension)

FIGURE 1. Ehwald's Grouping of MSS

	V	W	X	Y	Z
1	A	B	A	C	D
2	A	A	B	D	B
3	A	B	A	C	D
4	A	B	C	D	D
5	A	A	B	B	—
6	A	A	B	A	A
7	A	A	B	B	A
8	A	B	B	C	C
9	A	B	C	C	D
10	A	A	B	C	D

FIGURE 2. Sample Matrix of MSS and Variations

	V	W	X	Y	Z
V	1	.5	.2	.1	.22
W		1	.1	.1	.22
X			1	.2	0
Y				1	.33
Z					1

FIGURE 3. Matrix of Similarity Coefficients

	V/W	X	Y	Z
V/W	1	.15	.1	.22
X		1	.2	0
Y			1	.33
Z				1

FIGURE 4. Matrix Combining 2 MSS with Highest Similarity Coefficients

	V/W	X	Y/Z
V/W	1	.15	.16
X		1	.165
Y/Z			1

FIGURE 5. Matrix Combining MSS with 2 Next Highest Similitarity Coefficients

	V/W	X/Y/Z
V/W	1	.155
X/Y/Z		1

FIGURE 6. Final Matrix

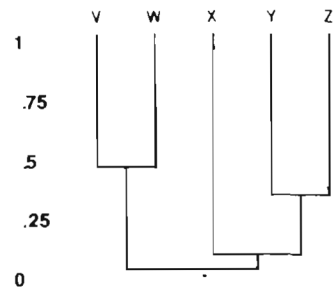

FIGURE 7. Connections among MSS Displayed in Phenogram

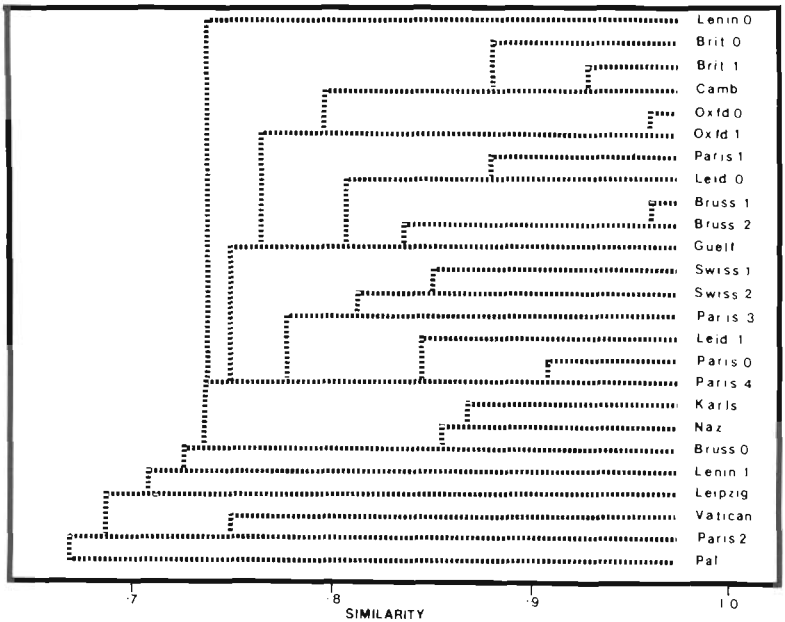

FIGURE 8. Phenogram of *Aenigmata:* Preface and Riddles 1-7, 72 lines, Method A

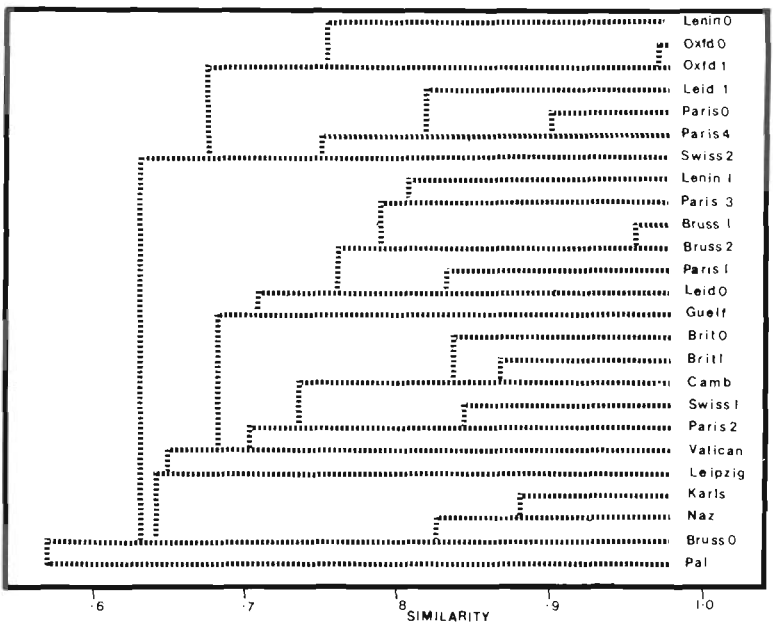

FIGURE 9. Phenogram of *Aenigmata:* Preface and Riddles 1-7, 72 lines, Method B

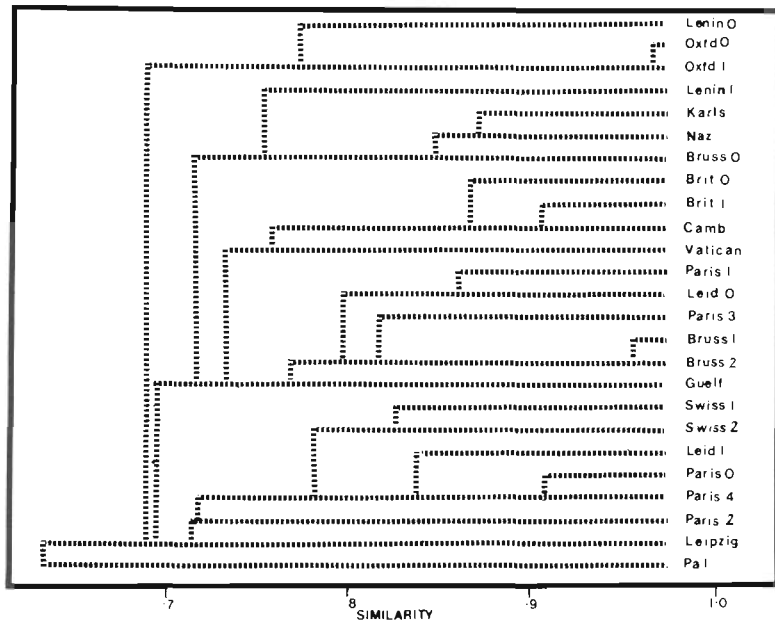

FIGURE 10. Phenogram of *Aenigmata:* Preface and Riddles 1-7, 72 lines, Method C

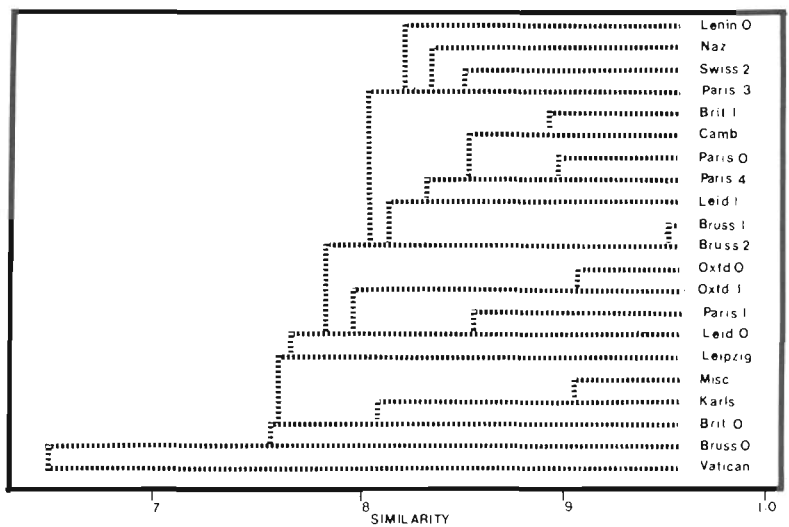

FIGURE 11. Phenogram of *Aenigmata:* Riddle 100, "Creatura," 85 lines, Method A

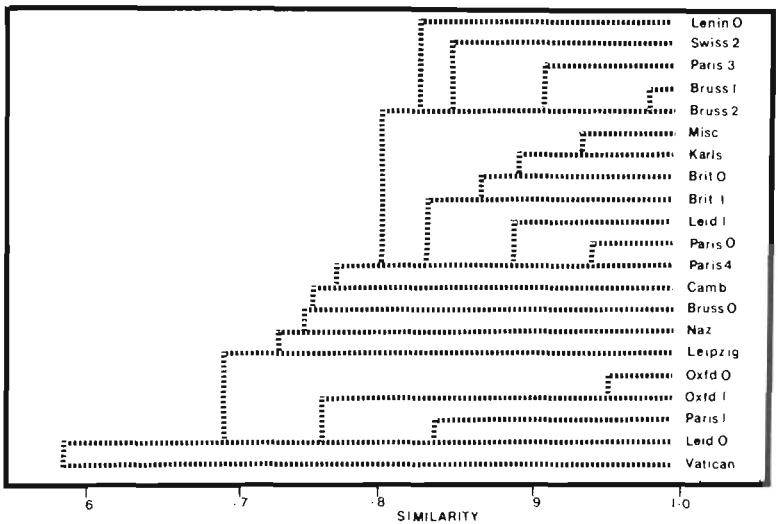

FIGURE 12. Phenogram of *Aenigmata;* Riddle 100, "Creatura," 85 lines, Method B

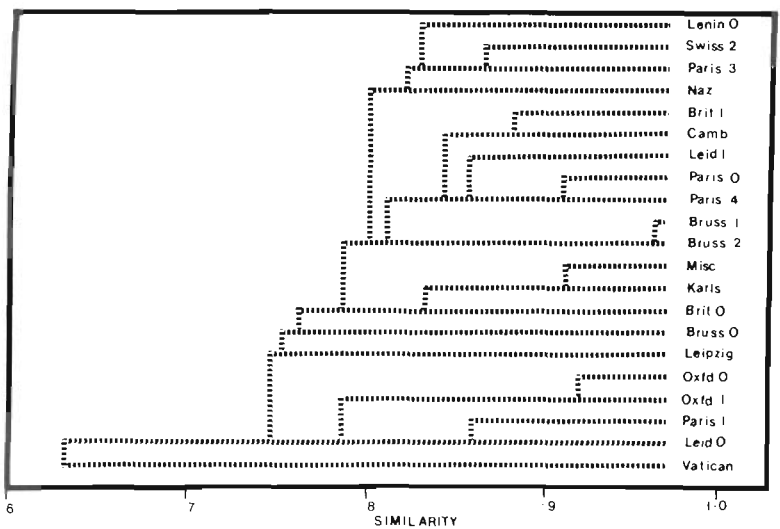

FIGURE 13. Phenogram of *Aenigmata:* Riddle 100, "Creatura," 85 lines, Method C