TRADITIONAL DRAMA STUDIES

VOLUME 1 1985

The Problem of Analysis of Traditional Play Texts: A Taxonomic Approach

PAUL S. SMITH

The study of traditional play texts, like many other forms of textual traditions, has suffered over the years from what I refer to as "the like syndrome". By this I imply that researchers have often been prepared to make statements concerning the relationships amongst pairs or groups of texts in terms of some personal value judgement.

Whilst such value judgements may partially work with a small number of plays and a well defined set of criteria for comparing texts, as the corpus of material increases, the application of the rules becomes difficult. This applies particularly if all pairs of texts are to be considered. As a consequence, resulting conclusions as to the distribution of plays, causes of textual similarity/dissimilarity must necessarily be suspect.

In the past, this problem has been partially overcome by using a limited number of criteria for matching texts - usually the presence of particular characters or plots, and attributing the texts exhibiting such features to one of a number of predetermined categories.[1] Whilst such an approach is useful as an initial data reduction device, unfortunately by imposing external categories and not allowing the data to generate the classificatory scheme,[2] inflexible generalisations tend to occur. Similarly, it is difficult to determine the extent of the overlapping amongst the predetermined categories, and so assess the criteria which make them part of a common single genre or separate them as distinct entities.

It is apparent, therefore, that to further the interpretation of the many texts collected to date, a more suitable method of analysis needs to be employed, allowing the data to determine the categories. This paper attempts to fulfil this requirement by firstly examining the possibility of utilising a taxonomic classificatory technique and, secondly demonstrating the possible use to which results derived from this method could be put.

For the purpose of this study, it was decided to analyse a selection of texts of the type of plays referred to as the Derby Ram, Derby Tup or T'Owd Tup. This decision was taken for several reasons. Firstly, although it would have been ideal to examine the texts of the plays listed in English Ritual Drama,

to undertake to analyse even a ten percent sample of the 474 full texts listed would be well outside the scope of an exploratory study; secondly, the Tup plays have comparatively short texts, thus speeding analysis; thirdly, whilst they exhibit textual variation amongst themselves, they still retain a high degree of similarity; fourthly, as this particular type of text, so far as is known, has only been transmitted by oral means, no chapbook versions having been recorded, in subsequent interpretation of the results, the problems of taking into account alternative and multiple media of transmission are alleviated.

The Derby Ram play has been performed for at least the last hundred years in South Yorkshire and the adjacent counties. The distribution of the recorded performances of this tradition is set out in Cawte's Ritual Animal Disguise.[3]

As the prime focus of this study is on the text of the plays, prior to the analysis, consideration must be given to the selection of texts in terms of their being complete and accurate documentations. It is, of course, very difficult to judge whether in fact any one text is complete. For example, several accounts of performances imply more text was used than was noted or printed. One such example is to be found in Thomas Ratcliffe's description of the performance of T'Owd Tup in Worksop prior to 1898:

> "The owner fails to make a bargain and he says 'Then Ah'll hay sum mutton fer my supper'. The fourth actor is a butcher and he comes in to 'Stick t'tup' and is allowed to do so after he has detailed his ability for the butchering business."[4]

These plays are by nature additive texts, in that verses/lines can be added or omitted at will, allowing the play to be lengthened or shortened to suit the social context of the individual performance, without altering the sense of the play.[5] To confuse matters further, it may be that the same selection of verses was not necessarily used in each performance, but rather the performers drew on a corpus of material.[6] Similarly, there is also the problem that the recorders may have consciously or unconsciously omitted material or, made errors in the notation of texts. There is unfortunately little that can be done to overcome these inherent problems, other than acknowledging that what we are examining is no more than the documentation of a series of single performances.

In selecting plays for analysis, all items with texts of a fragmentary or ambiguous nature were discarded. This process left 16 texts,[7] referring to the 14 locations identified on Map 1. There were, in fact, several further texts which could have been used. However, they referred to performances at locations already listed and it was considered that it would be of more value to include a representative sample of plays throughout the area, rather than concentrating attention on the temporal change which had taken place at any one location. However, at two locations, Staveley and Carlton-in-Lindrick, additional texts were incorporated to aid the analysis of theory concerning the static-temporal transmission of such plays.

Compared to the current sampling methods employed by behavioural scientists, this selection of data by the elimination of unsuitable texts, is to say the least, crude. However, as the data was not gathered in any systematic manner in the first place and is consequently subject to several biases which directly affect the spatial and temporal distribution of the plays,[8] it can therefore be considered that these texts comprise a representative enough sample for our purposes.

To turn our attention to the development of a taxonomic classification within any one genre of fixed narrative texts,[9] such as traditional plays, in both spatial and temporal terms, there should exist a high degree of affinity between them.[10] It would seem apparent, therefore, that if the extent of the affinity between the texts comprising our data set can be measured, it could be used as a basis for implementing taxonomic classification techniques such as cluster analysis. In addition, such techniques could provide data in a suitable form for the subsequent testing of hypotheses.

Scholars in the field of language studies and particularly those involved with examination of textual relationships and the authorship of anonymous texts, have for some time successfully been utilising cluster analysis techniques.[11] However, except for the studies by Morgan and Shaw on English dialect,[12] little consideration has been given to the possibility of applying such techniques to the analysis of cultural traditions.

Everett suggests that clustering techniques can be used in seven basic ways:

1. Finding a true typology
2. Model fitting
3. Prediction based on groups
4. Hypothesis testing
5. Data exploration
6. Hypothesis generation
7. Data reduction[13]

and it is primarily with uses, 1, 4 and 7 that we are concerned.

The basic principles of cluster analysis set out by Sokal and Sneath[14] require that the objects to be clustered are represented numerically in the form of similarity or dissimilarity coefficients calculated for each pair of "individuals" - in our case, pairs of plays. These coefficients then provide the base from which to construct clusters of "individuals" which are related at differing levels of similarity.

The method used to calculate the similarity/dissimilarity coefficients is to draw up a table listing "characteristics" and indicating the presence or absence in each "individual" of these "characteristics".

For example, if we wish to study the relationship between five phenomena, such as shopping centres, A, B, C, D and E, using as "characteristics" the types of shops present and/or absent in each centre, the following table (Fig. 1) showing how the different "characteristics" were distributed amongst the centres would be drawn up. A stroke is used to indicate the presence of a "characteristic", a cross its absence and a zero to indicate that no comparison was possible. This latter situation arises in the case of Centre A, which does not have "characteristic" 4. It cannot consequently be recorded for the tied "characteristic" 5. Such a table would then form the initial binary data matrix.

Fig. 1 Hypothetical Binary Data Matrix

"Characteristics"	Shopping Centres				
	A	B	C	D	E
1. Post Office	/	/	/	/	X
2. Furniture Store	X	/	X	/	/
3. Grocers	/	X	/	/	/
4. Bank	X	/	/	/	/
5. Bank (Overseas Dept.)	0	/	X	/	X
6. Newsagents	/	X	/	/	/
↓	↓	↓	↓	↓	↓
60. Book Shop	X	/	X	/	/

The presence and/or absence of the total number of "characteristics" for any pair of centres can then be scored in a two-by-two table:

		A: +	A: -
B	+	a	b
	-	c	d

Using one of many similarity coefficients devised,[15] the scores can then be transformed to similarity coefficients and recompiled as a matrix showing the extent of similarity for all pairs of centres in the data set. This similarity coefficient, then, is the measure of affinity.

The question we can now pose is, if this method is to be adopted, what aspects of the texts under consideration can be used as "characteristics" which will suit the data input requirements of cluster analysis?

Literary and linguistic applications of cluster analysis have used many different and varied units to measure the similarity between any pair of forms. In the majority of studies, however, some measure of either word, word combination or motif comparisons have been used, and it is this approach to measuring similarity which has been adopted in the present study.

Theoretically, play texts can be seen as comprising short speeches/verses which can be broken down for analysis in three main ways:

1. Word units
2. Line units
3. Speech/verse units

These units can be used as the equivalent of Sokal's "characteristics". In addition, in considering the extent of the affinity existing amongst these units, they can be "matched" with each other at two different levels;

1. "Identical unit match". Here the word, line or speech units in any two texts are matched character for character.
2. "Semantic unit match". In this case the word, lines or speech units, although not having the same text, are taken to have an identical meaning.

In illustration of the latter case, when comparing texts at a "semantic line match" level, a line such as:

"They were of football size"

is taken as being identical to:

"For they were of football size"

A combination of the three different units and two levels of matching gives the possibility of six different measures of similarity between texts, and these are summarised in Fig. 2.

Fig. 2 Possible measure of textual similarities

Levels of similarity \ Units	Word	Line	Speech / Verse
Identical	Identical word match	Identical line match	Identical speech match
Semantic	Semantic word match	Semantic line match	Semantic speech match

It is apparent, however, that with specific data sets, advantages accrue from using certain of these measures as opposed to others. For instance, the

identical word matching approach, although of little use when comparing plays from distinct traditions, is very useful if serial recordings of one team of performers are being analysed.

In terms of obtaining the most accurate measure of similarity, probably the "identical line match" approach is the most applicable. This proves more useful than working at the speech/verse unit level, for as the size of the "units" being matched is increased, so the possibility of finding identical matching units decreases.

As the aim of this exploratory study is to examine the possible utilisation of this type of technique, it was decided to calculate the similarity coefficients on semantic speeches/verses matching. This would utilise the principle of this type of approach, whilst at the same time reducing the possibility of error, as the number of units, and consequently decisions to be made, would be reduced.

Having decided on the nature of the "characteristics" to be used, a list of the individual "characteristics" in the corpus of texts was drawn up. Proper names and numbers were not taken into account when matching, as at the semantic match level the existence of a name or number in the text was seen as being the matching requirement. In the case of identifying distinct verses, no problems were encountered; in the case of the variable length speech sections, however, this proved a little more difficult. The most adequate method was found to be to treat them in terms of the overall actions expressed by that speech. Furthermore, because of the peculiar nature of the choruses in the song section, the majority having no semantic content, the only way found to account for them as "characteristics" was in terms of being "vocables", those choruses having nonsense forms:

e.g. "Failey-failey laddie fol larie lay"

and "non-vocables", those choruses having identifiable words:

e.g. "Poor old tup, poor old tup"

From the texts, 52 "characteristics" were identified, and the 16 texts were scored against these for presence or absence against. This then produced a 16 x 52 cell binary data matrix. The next stage was to decide on the type of similarity coefficient to be used. This coefficient demonstrated the measure of affinity between any pair of texts.

The CLASP Cluster Analysis Program,[16] had been selected to calculate the similarity matrix from the binary data matrix, as this program was capable of accommodating a large matrix, i.e. up to 6000 cells, and calculating several different types of similarity coefficients and derived clustering algorithms. A qualitative similarity coefficient was selected, as this took into account not only the extent of the positive, but also the negative matches in the binary matrix. It was considered important to take this factor into account for, in this instance, as we are dealing with additive texts where the text length can vary without affecting the overall sense of the play. Consequently, it was felt that the sections of text omitted should have equal weight to those included. This point is, however, debatable and for different data sets it may be more applicable to score just the positive matches.

From this similarity matrix, it was possible to derive a variety of types of hierarchical clustering routines: nearest neighbour; single linkage analysis and centroid sorting. The actual selection of any one hierarchical clustering alogrithm is somewhat arbitrary. However it was considered that the median sorting and nearest neighbour routines would not be suitable for this purpose because of their tendency for "chaining".[17]

Of the remaining three algorithms, it was further decided to reject the average linkage analysis routine, primarily because the main objective of this analysis was to obtain maximum definition between clusters to aid subsequent analysis and interpretation. Consequently, it was decided to use the centroid analysis algorithm, as this produced the best clustering definition.

The CLASP program was then run using the qualitative version of the equation:

$$Sij = \sum_{k=1}^{p} Sijk \sum_{k=1}^{p} Wijk$$

where Sijk = 1 if the two individuals i and j are the same for the k^{th} character, and Sijk = 0 if they differ. From the 16 x 52 cell binary data matrix a 16 x 16 cell similarity matrix was calculated (See Fig. 3).

The centroid clustering algorithm was next computed from the similarity matrix and a list of the text clusters and the hierarchy of their groupings produced, (see Fig. 4).

In addition, the dendrogram illustrating the taxonomic relationships amongst the 16 texts was derived, (see Fig. 5). This presents a graphic representation of the levels of textual affinity between the plays. To aid identification of discrete clusters and identify "maverick" plays, a cut-off convention was then calculated as the between and within group mean similarity coefficient[18] which was, in this case, 75 percent. This is indicated on Figs. 4 and 5 by the continuous line labelled M.S.C.

Fig. 3 Similarity matrix of affinities amongst texts

TEXT	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A	100,0															
B	98,1	100,0														
C	71,2	69,2	100,0													
D	88,5	86,5	67,3	100,0												
E	86,5	84,6	76,9	86,5	100,0											
F	96,2	98,1	67,3	84,6	82,7	100,0										
G	67,3	65,4	61,5	67,3	69,2	63,5	100,0									
H	65,4	63,5	63,5	73,1	71,2	61,5	78,8	100,0								
I	75,0	73,1	76,9	71,2	80,8	71,2	61,5	59,6	100,0							
J	80,8	78,8	71,2	84,6	82,7	76,9	82,7	84,6	71,2	100,0						
K	73,1	71,2	67,3	76,9	71,2	69,2	67,3	69,2	67,3	80,8	100,0					
L	71,2	69,2	69,2	71,2	76,9	67,3	61,5	67,3	65,4	75,0	67,3	100,0				
M	71,2	69,2	69,2	67,3	73,1	67,3	53,8	55,8	80,8	71,2	63,5	69,2	100,0			
N	80,8	78,8	82,7	76,9	82,7	80,8	67,3	73,1	75,0	80,8	76,9	78,8	71,2	100,0		
O	78,8	76,9	73,1	82,7	84,6	75,8	84,6	82,7	73,1	98,1	78,8	73,1	69,2	78,8	100,0	
P	90,4	88,5	69,2	90,4	88,5	86,5	73,1	71,2	73,1	86,5	75,0	73,1	69,2	78,8	84,6	100,0

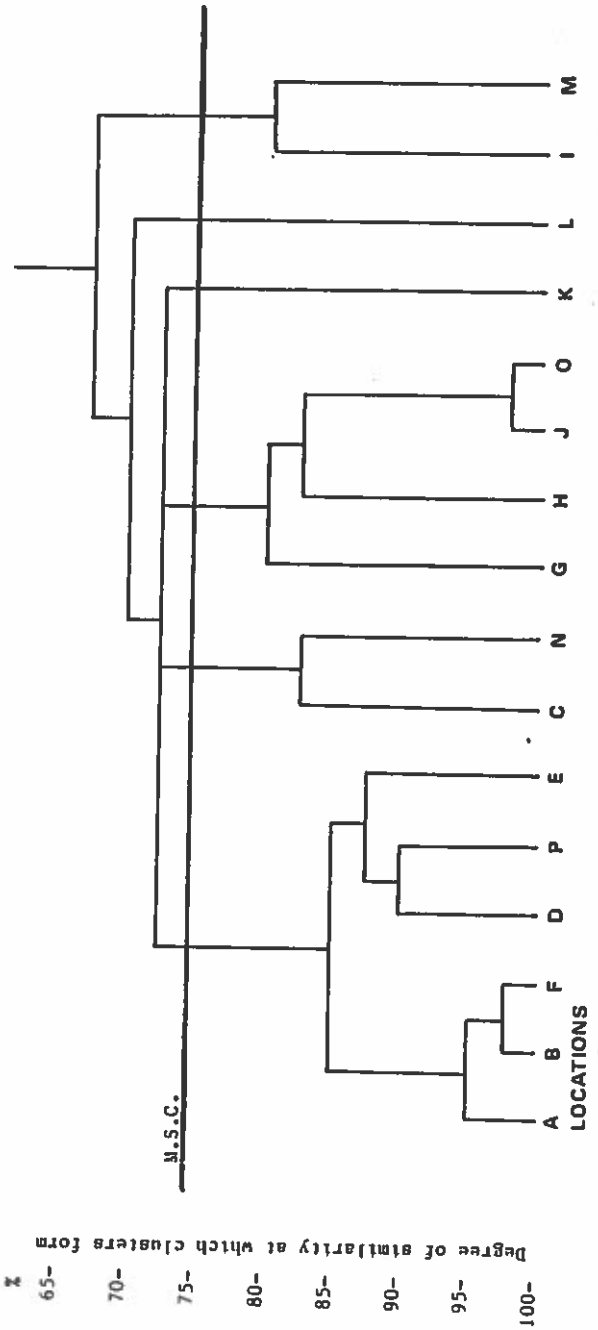
Fig. 4 Centroid Sorted Hierarchical Clusters at 2.5% Intervals

```
HIERARCHICAL CLUSTERS

LEVEL 97.5
**    2   6
**   10  15
LEVEL 95.0
**    1   2   6
**   10  15
LEVEL 92.5
**    1   2   6
**   10  15
LEVEL 90.0
**    1   2   6
**    4  16
**   10  15
LEVEL 87.5
**    1   2   6
**    4  16   5
**   10  15
LEVEL 85.0
**    1   2   6   4  16   5
**   10  15
LEVEL 82.5
**    1   2   6   4  16   5
**    3  14
**    8  10  15
LEVEL 80.0
**    1   2   6   4  16   5
**    3  14
**    7   8  10  15
**    9  13
LEVEL 77.5
**    1   2   6   4  16   5
**    3  14
**    7   8  10  15
**    9  13
LEVEL 75.0
**    1   2   6   4  16   5
**    3  14
**    7   8  10  15
**    9  13
LEVEL 72.5
**    1   2   6   4  16   5   3  14   7   8  10  15  11
**    9  13
LEVEL 70.0
**    1   2   6   4  16   5   3  14   7   8  10  15  11  12
**    9  13
LEVEL 67.5
**    1   2   6   4  16   5   3  14   7   8  10  15  11  12   9  13
```

M.S.C.

Fig. 5 Cluster diagram based on centroid sorting



From the dendrogram we can identify five groups of texts including one isolated maverick. These comprise:

1 Texts A/F, B, D, E, P
2 " G, H, O, J, K
3 " N, C
4 " I/M
5 " L (Maverick)

Here, then we are presenting the data in a form based on its internal relationships and not subject to an external classification scheme. This method highlights the complexity of the interrelationships in the data which could be disguised if an imposed classification, with its inherent possibility for simplification, were used.

At this juncture, however, one point needs to be made. Although we have derived a quantitative measure of affinity between the texts, neither the existence nor level of affinity calculated can be taken as indicating that any direct relationship in terms of transmission of information, causation, etc., exists between these recorded texts. Rather, this method provides us with a data reduction technique and classification tool which relates the data in a form suitable for the generation of theory and hypothesis testing.

From the initial apparently single class of T'Owd Tup plays, we have now obtained a much finer and more accurate picture of the relationships amongst these texts. Having derived such a classification, the next question is, can we use this information to further our understanding of the nature of the relationship amongst such plays? To demonstrate the possibilities in this area, it was decided to examine some of the basic spatio-temporal factors hypothesised as affecting the distribution of such texts.[19]

The initial stage in this exercise was to map the textual affinities existing amongst the 16 plays (see Map 1). We are able, therefore, with this approach, to map the quantitative, as opposed to the inferred relationships in the data.

It is immediately apparent from the map that two major groups of texts, A/F, B, D, E, P and G, H, O, J, K, form distinct spatial clusters. It is however also apparent that more than one type of relationship may exist between the texts plotted. How then can we explain and test the observed spatial patterns? In attempting to interpret the mapped configurations of the isolines for the five groups above, the following initial interpretation appears feasible:

Group 1 The high levels of affinity amongst plays at adjacent locations suggests diffusion of information from two centres, A/F-B and D-P

Group 2 — The high levels of affinity amongst plays at adjacent locations suggests diffusion of information from the single centre, J-O

Group 3 — The similarity between these two distant locations, N and C, suggests a migration of information

Group 4 — Apparent spatially static diffusion of information over time

Group 5 — Apparent low relationship to other locations

This implies that in the interpretation of the data we have to take two sets of factors into account. Firstly, the nature of transmission of information between locations in terms of distance - a distinction being made between diffused traditions having continuous transmission from one adjacent point to the next, and migratory traditions with transmission taking place between remote points and not through adjacent sectors. Secondly, the two dimensions of time and space through which the information has been transmitted.[20]

The resulting combinations of possible transmission pathways can be set out thus:

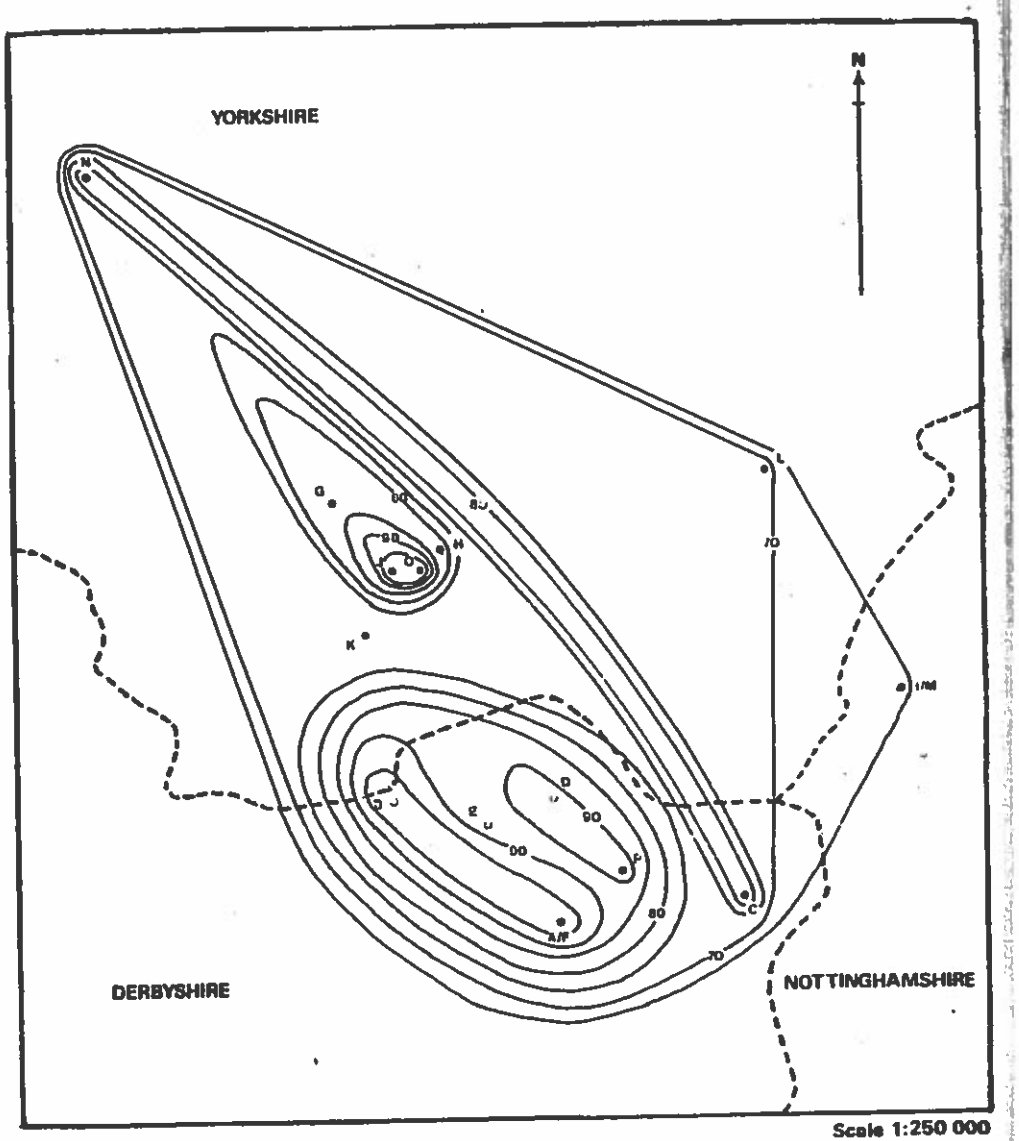
Fig. 6 Varieties of Transmission Pathways

Dimensions of Transmission	Distance of Transmission: Diffusion	Distance of Transmission: Migration
Space	A	B
Time	C	D

Having outlined the possible transmission pathways, it must however be stated at this point that the level of affinity between plays recorded at any two locations or time periods is a product, not solely of the distance between the points, but rather the number of intervening transmissions of information which have taken place between the two points. Therefore, the greater the number of intervening transmissions between two points, the greater the possibility of the level of affinity being lower.

Unfortunately it is apparent that because of inconsistencies in data collection, and the problematic nature of tracing paths of information transmission, it is impossible to know the number of intervening transmissions which have taken place between any pair of performances of a play. Consequently, distance in terms of miles or years between points will have to be used as an approximate measure. It is also important to remember that the data we are dealing with, regardless of how it reached the performers, is at least three intervening transmissions from the point of transfer to the performers, having at least one transmission from that point to the performance of the play, at least one

Map 1 Centroid Sorted Clusters of Play Texts



Key to Locations

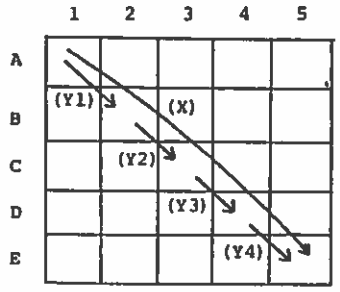
A/F	- Staveley	I/M	- Carlton-in-Lindrick
B	- Coal Aston	J	- Pitsmoor
C	- Whitwell	K	- Netherthorpe
D	- Eckington	L	- Braithwell
E	- Handley	N	- Denby Dale
G	- Ecclesfield	O	- Page Hall
H	- High Wincobank	P	- Clowne/Eckington

transmission from the performance to the recorder and at least one transmission from the recorder to us. Therefore we must acknowledge that the number of transmissions which have taken place between the point of information transfer and this analysis must in themselves increase the opportunity for modifications to occur which in turn will affect the levels of affinity amongst the texts.

To turn to spatial transmission, whilst in our simple model (see Fig. 6) we have discriminated between diffused and migratory transfers of information, this does not imply that they operate independently of each other. In reality, both systems may operate simultaneously. A single cultural tradition may therefore be transmitted from a single location (A1) to another (E5) by two separate routes, (X) by migration and (Y) by diffusion. (see Fig. 5)

Diffused transmission (Y_1, Y_2, Y_3, Y_4) through all the diagonal locations in this instance gives us three intervening transmissions, whilst migratory transmission (X) gives us one. Consequently, of the two states of the tradition

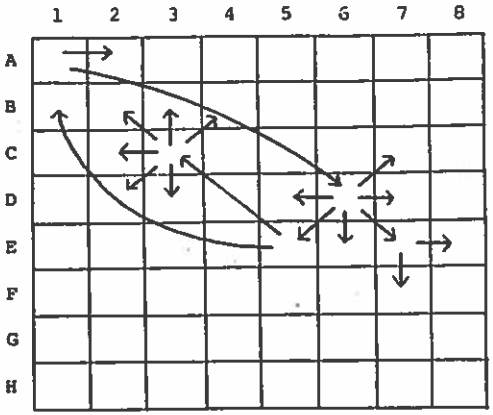
Fig. 7 Multiple tramsission of information through space by alternative routes



apparent at E5, one could have a higher affinity with A1 than the other. If this situation does occur, depending on the extent of the similarity affinity between the two traditions at E5, one of two things could happen. Either the traditions combine, and become a hybrid form, or they co-exist as distinct traditions. A high degree of variation between two such traditions at one location may tend to favour co-existence rather than amalgamation.

To further complicate the matter, unlike the model of diffusion adopted for studying settlement patterns, etc.,[21] where once a plot of land has been settled it cannot be resettled, the converse is true of the transmission of information. Thus, where Bylund's model indicates a multiple outward movement from a point of origin, in the case of information flow the direction of movement can be both forward and backward. When this is combined with simultaneous migratory and diffused transmissions, a complex situation of transmission can evolve, (see Fig. 8).

Fig. 8 Multidirectional migration and diffusion of information through space



Here, a migratory spatial transmission from A1 brings a tradition to D6; localised spatial diffusion then spreads the tradition through adjacent areas, however, migratory transmissions from E5 return a form of the tradition almost back to A1 and also to C3, where localised diffusion again sends a further form back towards A1 (i.e. B2). We can therefore expect that the traditions found at A1, A2 and D6 would exhibit least variation (all having only one transmission); conversely, A1 and B1, although adjacent, would exhibit the greatest variation, as B1 is three intervening transmissions from A1. Such complex multidirectional spatial transmissions of information could therefore easily produce spatial patterns between play texts that defy analysis.

Having plotted the spatial distribution of the calculated textual affinities and made some subjective observations concerning their spatio-temporal distribution, in order to unravel this matter further, the following hypothesis was set out:

> As physical distance between locations of performances increases, so the affinity between play texts will decrease

This was tested using Spearman's Rank Order Correlation to compare the rank order of the distance between locations measured as a straight line against the similarity coefficients acting as a measure of affinity. The correlation was calculated using each of the texts in turn as the point of origin, where distance equals nought and similarity equals a hundred. From the Rank Order Correlation Coefficient, t scores and levels of significance set out in Table 1, it is apparent that in 14 out of the 16 cases, the hypothesis was confirmed and affinity does decline with distance. This then, argues for the case of spatial diffusion of information in 14 out of 16 cases.

Table 1 Table of Spearman Rank Order Correlation derived from analysis of distance/similarity data

Text	r	t Score	Level of significance
A	-.56	2.52	$p < 0.01$
B	-.53	2.33	$p < 0.01$
C	-.13	.829	$p < 0.01$
D	-.72	3.88	$p < 0.01$
E	-.59	2.73	$p < 0.01$
F	-.54	2.39	$p < 0.01$
G	-.85	6.01	$p < 0.01$
H	-.70	3.66	$p < 0.01$
I	-.34	1.35	$p < 0.01$
J	-.88	6.92	$p < 0.01$
K	-.44	1.83	$p < 0.01$
L	+.005	.019	Not significant
M	-.26	1.00	$p < 0.01$
N	+.064	.239	Not significant
O	-.72	3.88	$p < 0.01$
P	-.53	2.33	$p < 0.01$

Closer examination of the r scores confirms the suggestion that group 1, comprising locations A/F, B, D, P & E and group 2, comprising locations J, O, H, G & K are the product of spatial diffusion processes, as they all have high negative correlation scores. Group 3, comprising locations N & C, on the other hand, have much lower scores, one being a low negative correlation significance at p 0.1 level and the other a low positive, but not significant correlation. This could be taken as indicative that spatial migration of information linked these two distant points. Group 4, on the other hand, comprising texts I & M at one location, althoguh confirming a diffusion hypothesis, must still be thought of as an isolate, the diffusion distance/similarity r score being at a lower level than groups 1 and 2. In addition, as both these texts were recorded at the same location and, if the 75 percent cut-off level is enforced, only cluster with each other, it is possible that they represent a spatially static play text which has been modified over time.

Group 5, being the single "maverick" text at location L, is similarly confirmed as an isolate - the rs being very low positive and not significant. The existence of these two isolates can be explained in three ways: firstly, intervening data could be lacking to connect them to adjacent plotted locations; alternatively, the processes acting on information during transmission has produced drastic variation in these texts.[22] However, it would seem that such variation is more likely to be the product of the number of intervening transmissions in a system of information diffusion than as the product of a single migratory transmission. This therefore suggests again the lack of intervening data for analysis. Finally, the possibility exists that these two texts are the product of what is initially a spatially static process of independent evolution.[23] However, as both these isolates link to the main body of the locations at a 70 percent similarity level, and the fact that they are less than eight miles to the nearest plotted play, would argue against this latter interpretation. Consequently it can be suggested that locations L and I/M are isolates only because of the lack of intervening data.

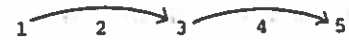
Having shown that it is possible to test the distance/affinity hypothesis and at the same time make interpretations as to the mechanisms of spatial transmission utilised, the problematic nature of such an interpretation can be highlighted by the introduction of the temporal dimension into the analysis.

Temporal diffusion of information, unlike its spatial counterpart, moves in a single direction. This implies that the information being transmitted is always present, in that it is passed from one time period to another, therefore exhibiting a continuous temporal distribution.

Fig. 9 Temporal diffusion of information

Temporal migration of information, conversely, is taken to imply that the information being transmitted jumps some intervening time periods - see Fig. 10.

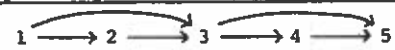
Fig. 10 Temporal migration of information



This implies that some information is available to be used by individuals in certain time periods and not in others. Such temporal migratory transmission may for instance occur because information is taken out of circulation because of the presence of an unfavourable environment for the use of that information. For example, the majority of the plays in this study were performed by boys and youths under twenty for the purpose of collecting money. It is apparent, therefore, that if the economic factors in their environment improve, they may discontinue performances. Nevertheless, at times of economic crisis, the revival of such plays by the original performers has been documented.[24] Similarly, elderly ex-performers are known to have taught the plays to their grandchildren. This therefore sets up a situation in which information regarding cultural traditions is not perhaps always available, but rather has a sporadic temporal distribution, depending on the prevailing social/economic environment.

The temporal diffusion and migration of information, as was pointed out with their spatial counterparts, are not separate systems, but rather can work simultaneously. Therefore, one item of information can be transmitted from a single point and reach a common destination by different paths (see Fig. 11), in two different forms, depending on the number of intervening transmissions which have taken place.

Fig. 11 Multipath temporal transmission of information



As there is such a high similarity between spatial temporal transmission models We could perhaps expect that if we reformulate our original hypothesis to substitute temporal for spatial distance, the expectation would be that:

As the length of time increases amongst any group of recorded performances of plays, the textual affinity amongst them will decrease

To test this hypothesis, the Spearman Rank Order Correlation Coefficient was again applied, this time utilising the temporal distance, in years, of all texts from A (the earliest play) as the measure of time, and similarity coefficients of all plays to A as a measure of affinity. This produced a correlation coefficient of +.48 which was significant at the $p < .01$ level. Consequently, the hypothesis of a negative relationship between time and distance had to be rejected.

The question which now presents itself is, if what we are really dealing with is a situation in which the temporal and spatial mechanisms of transmission affect the affinity between texts, can a combined measure of distance in terms of space/time be evolved which can be used to correlate against similarity? Possibly such a solution will compensate for the positive correlation of the time/affinity hypothesis.

The hypothesis to be tested was therefore formulated:

As distance in terms of time/miles increases amongst locations then the affinity among texts recorded at those locations would decrease

This hypothesis was again tested using the Spearman Rank Order Correlation Coefficient.

The affinity of the plays was again taken as the similarity coefficient of all texts to A (the earliest text). The measure of distance was calculated incorporating both the physical distance from A to all locations in miles and the temporal distance in terms of the number of years elapsed from performance A to each other date of recorded performances. To make both these measures compatible, each distance (miles and years) was converted into a percentage using the equation:

$$\text{Percentage miles/year x from A} = \frac{\text{Distance x to A}}{\text{Distance from A to furthest point}} \times 100$$

The individual percentage miles and years figures for each location were then added and this gave us a time/distance measure of each location from A.

The calculated correlation coefficient was +.25 with a t score of .92 which was significant at the $p < .01$ level. Therefore, this hypothesis had also to be rejected.

The results produced in the compilation of the last two hypotheses, both of which include measures of temporal distance, were positive and significant at the p 0.01 level. This high level of significance indicates that in terms of temporal and spatio-temporal transmission of information there is a positive relationship between time and affinity, in that affinity does not decrease over time, but rather increases.

One speculative interpretation which can perhaps be put forward to explain this finding is the theory of independent convergence.[25] The implication here is that in theory, as time goes on, a more uniform social/economic environment may be evolving over the whole area. Consequently, cultural traditions in adapting to suit this homogeneous environment, independently tend to become more alike. Thus, over time, they would exhibit a higher level of affinity. It must however be stressed that this is a speculative interpretation, and to test it is outside the scope of this study, in that it would require that factors concerning the environmental context of the area, performers and performance be incorporated in the analysis.

The question now remains, even after having tested the hypothesis to a significant level, can we be sure of the validity of the conclusions? Unfortunately, the answer must be no, for without information concerning the direction of the information transfer between locations, subsequent hypotheses and analysis must be considered no more than superficial. This is primarily because in lieu of sampled data, we have been making an assumption that a direct relationship exists between the recorded play texts comprising our data. It is only necessary to compare the total number of recorded performances of T'Owd Tup plays with our sample of 16 plays to see the number of intervening locations having plays between any pair in our data set. Each recorded performance of course presents the possibility of being the actual link between any pair in our sample.

In addition, because of the temporal and spatial biases of the recorders, already commented on,[26] the possibility exists that within the total area we are dealing with, far more locations had plays than were recorded. Thus, the possible paths and number of intervening transmissions existing between any two points are considerable.

These unknown intervening locations therefore make any assumption regarding the existence of an actual physical relationship between clustered texts unacceptable, as the relationship between any pair of texts could possibly be through an unlocated third point:

A ⟶ [x] ⟶ B

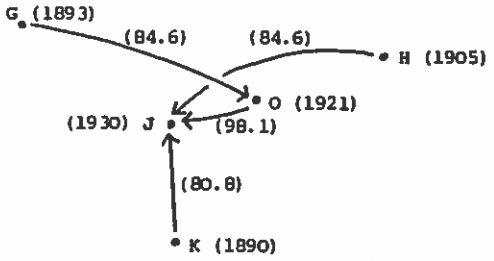
or conversely, both texts could have been derived from a common, unlocated origin:

A ⟵ [x] ⟶ B

It could possibly be considered that the incorporation of the dates of the texts into the analysis would reduce the problem by indicating the sequence of the performance of the play and thus suggest more positive linkages. In practice, however, it appears only to confuse matters further. This can be seen by attempting to interpret the group 2 cluster using both measures of affinity and temporal data (see Fig. 12).

Fig. 12 Spatio-temporal linkages of group 2 cluster

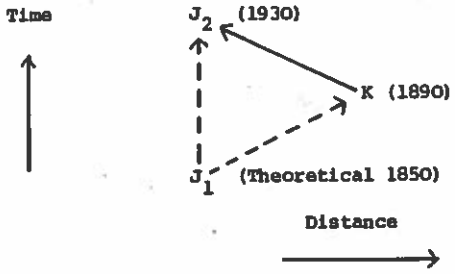
The linkages indicate the highest level of similarity amongst any pair in the cluster (e.g. J has the highest similarity with H).



Here, when dating is taken into account, all the nearest neighbour similarity coefficients converge on J, the most recent performance. Although each of the links plotted can be thought of as acceptable within the theory of transmission of information, this spatio-temporal presentation of the data argues that cluster 2, with nucleus J-O is in fact suspect and should possibly be considered as the point of coincidence of three relict independent adjacent clusters (nuclei G, H & K). However, within the theory of information transmission, we should expect that if the latter were true, J would exhibit least similarity with all the other locations in the group, being a hybrid text. In reality the converse is true.

Alternatively, we can possibly attempt to explain this anomaly in terms of the fact that although the apparent sequence of linkage is from the earlier texts (i.e. all but J), to the later (J), perhaps a form of play has always existed at J. Therefore, instead of the similarity coefficients being a measure of the observed affinity between K and J, it is perhaps a measure of an earlier unrecorded transmission of information from J to K, (see Fig. 13).

Fig. 13 Theoretical transmission pattern



This possible solution highlights the extremely complex linkages which form the underlying structures of such spatial distributions as shown in Map 1. How then can this problem be overcome and the spatio-temporal analysis of cultural traditions be furthered? The most obvious necessity is to review our methods of data collection, perhaps by aiming to obtain sampled data. This can be done in two ways, either by random sampling or by the systematic search of locations within a specified area to gather spatio-temporal profiles of past and present dramatic traditions.

The first approach would enable such data to be analysed and interpreted using only the two measures of affinity and distance. The second approach, using systematic collection of information on both the spatial and temporal aspects of a tradition, would provide suitable information for the analysis of the data in terms of time/distance/ and affinity. It would also seem reasonable to ask performers the source of their information, in that if we can document the route of transmission of a tradition between any two points, a variety of other analytical techniques such as network analysis could be adopted.[27]

To a certain extent, the implementation of sampling techniques improves the quantity of the information obtained. However, consideration should also be given to improving the quality. This would necessitate not only the more accurate documentation of the traditions themselves, in terms of texts, performance, etc., but also the social and economic environments of both the "performers" and the context in which the tradition was practiced. Such information would then enable us to produce more accurate taxonomic classifications based on performance rather than merely on texts whilst at the same time assisting us to better interpret the observed relationships.

At this point, further consideration would need to be given to the units of matching and matching criteria employed and also the selection of appropriate similarity coefficients to deal with the specific problem in hand.

In the final assessment of this exercise, we can say that our original objective has been attained in that it has proved comparatively easy to use cluster analysis techniques to develop a finer taxonomic classification of texts. Even so, subsequent analysis of what on the surface appeared to be fairly simple spatial patterns, did not prove so straightforward. This situation was caused primarily by the complexity of the unknown transmission network underlying the recorded data.

NOTES

1. See for example, E. C. Cawte, Alex Helm and N. Peacock, English Ritual Drama: A Geographical Index (London: 1967), pp. 14-15.
2. B. G. Glaser and A. L. Strauss, The Discovery of Grounded Theory and Strategies for Qualitative Research (London: 1968).
3. E. C. Cawte, Ritual Animal Disguise (Cambridge: 1978), p. 111.
4. T. Ratcliffe, "Christmas Tup", Notes and Queries, Ser. 9, II (1898), 511.
5. P. S. Smith, Information and Context: An examination of factors influencing the performance of traditional plays in the British Isles. (Paper presented at the Popular Drama and Film Conference, University of Kent at Canterbury, September 1977).
6. R. Greig, "Poor Old Horse", Lore and Language, I:9 (1974), 7-10.
7. (A) - Ivor Gatty, "The Old Tup and its Ritual", Journal of the English Folk Dance and Song Society, V (1946), 28-29;
   (B) - P. S. and M. G. Smith Collection (Sutton 25.3.75);
   (C) - P. S. and M. G. Smith Field Collection (Tape 1: 1966);
   (D) - Centre for English Cultural Tradition and Language - Long item File No. 206;
   (E) - R. H. Williams, "The Derby Tup", Derbyshire Countryside, XIX (1935) 43-44;
   (F) - Arthur Court, Staveley, My Native Town - Some Historical Notes of the Parish (Sheffield: 1948), pp. 85-87;
   (G) - P. S. and M. G. Smith Field Collection (Tape 10: 1968);
   (H) - P. S. and M. G. Smith Field Collection (Tape 10: 1968);
   (I) - Mike Howley, "The Little Tup", Folk, II (Oct. 1962), 9-10;
   (J) - P. S. and M. G. Smith Field Collection (Tape 11: 1968);
   (K) - Centre for English Cultural Tradition and Language - Long item File No. 194;
   (L) - Ivor Gatty, "The Old Tup and its Ritual", pp. 24-26;
   (M) - A. S. Buxton Collection, Mansfield Public Library, pp. 60-67;
   (N) - P. S. and M. G. Smith Field Collection (Tapes 1 and 2: 1963);
   (O) - Centre for English Cultural Tradition and Language - Long item File No. 151;
   (P) - R. Greig, Traditional Drama in the Sheffield Area (Unpublished B.Ed. Dissertation, Sheffield City College of Education, 1972), p. 23-25.
8. It must be remembered that the spatio-temporal distribution of cultural traditions is as often as not a function of the spatio-temporal distribution of collectors rather than the performance or usage of the tradition.
9. P. S. Smith, "Tradition - A Perspective: Part II, Transmission", Lore and Language, II:3 (1975), 6.
10. Smith, "Tradition - A Perspective", 10.
11. L. A. Ule, "The Use of Constat in Authorship Investigation", ALLC Bulletin, I:3 (1975), 211-225; A. L. Kroeber, "Statistics...", Language, XXXVI:1 (1960), 1-21; J. G. Griffith, "A Tamonomic Study of the Manuscript Tradition of Juvenal", Museum Helveticum, XXV (1968), 101-138.
12. Byron J. T. Morgan and David J. Shaw, "Graphical Methods of Illustrating Data in the Survey of English Dialects", Lore and Language, 3:7 (1982), 14-29.
13. B. Everett, Cluster Analysis (London: 1974), p. 3.
14. R. R. Sokal and P. H. A. Sneath, Principles of Numerical Taxonomy (San Francisco: 1963).
15. Sokal and Sneath, Principles of Numerical Taxonomy, pp. 128-30.
16. G. J. S. Ross and F. B. Lauckner, CLASP: A Classification Program (Rothamsted Experimental Station, implemented at Manchester Regional Computing Centre, 1974).
17. Everett, Cluster Analysis, p. 61.
18. L. A. Ule, "Cluster Analysis", ALLC Bulletin, II:3 (1974), 19.
19. Paul S. Smith, "Communication and Performance; A Model of the Development of Variant Forms of Cultural Traditions" in A. E. Green and J. D. A. Widdowson, (eds.), Language, Culture and Tradition: Papers on Language and Folklore Presented at the Annual Conference of the British Sociological Association, April 1978 (Leeds and Sheffield: Institute of Dialect and Folklife Studies and Centre for Englsih Cultural Tradition and Language, 1981), pp. 16-32.
20. P. S. Smith, "Tradition - A Perspective: Diffusion, Migration and Independent Evolution", Lore and Language (forthcoming).
21. For a discussion of types of diffusion patterns, see E. Bylund, "Theoretical Considerations regarding the Distribution of Settlement in Inner Northern Sweden", Geografiska Annaler, XLII (1960), 1-4 and 225-231.
22. Smith, Communication and Performance, pp. 5-8.
23. H. E. Driver, "Statistical Studies of Continuous Geographical Distribution", in R. Naroll and R. Cohen (ed.), A Handbook of Method in Cultural Anthropology (New York: 1973), pp. 620-639.
24. P. S. and M. G. Smith Field Collection (Tape 1: 1966).
25. Driver, "Statistical Studies..", p. 629.
26. See note 8.
27. P. S. Smith, M. G. Smith and J. D. A. Widdowson, Traditional Drama (Sheffield: Survey of Language and Folklore, 1972).