

TRADITIONAL DRAMA STUDIES



VOLUME 2

1988

THE PROBLEMS OF ANALYSING FOLK PLAY CAST LISTS USING NUMERICAL METHODS

PETER T. MILLINGTON

Introduction

The procedure for analysing collections of folk play casts, whether it be using numerical classification or conventional methods, can be summarised as follows:-

1. Selection of folk play cast lists for analysis.
2. Listing and standardisation of the data.
3. Development of a classification scheme which illustrates the relationships between different groups of plays and/or characters.

In this paper I will work through this procedure in detail, pointing out some of the problems and pitfalls involved, and describing ways of resolving them. I will draw on examples taken from a numerical analysis of some eighty Nottinghamshire plays.

Selection of cast lists for analysis

In order to be able to conduct a meaningful analysis, only complete cast lists should be used. Often, internal evidence within a record indicates whether or not a cast is complete. However there still remain many records where this is not obvious. Fig. 1 shows the frequency of the number of characters in 206 accounts of plays from Nottinghamshire. Two peaks are obvious, the left hand one representing mostly incomplete casts, and the right hand one mostly complete casts. This pattern can be verified by producing separate graphs for known complete and known incomplete casts. The low point between the peaks occurs at three characters per account, and can be regarded as the borderline between complete and incomplete casts. This indicates therefore that if an account lists four or more characters, unless there is evidence to the contrary, the cast can be treated as complete.

THE PROBLEMS OF ANALYSING FOLK PLAY CAST LISTS USING NUMERICAL METHODS

PETER T. MILLINGTON

Introduction

The procedure for analysing collections of folk play casts, whether it be using numerical classification or conventional methods, can be summarised as follows:-

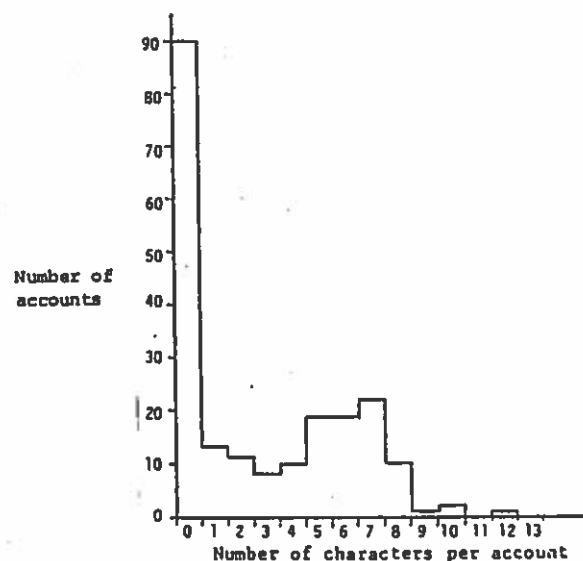
1. Selection of folk play cast lists for analysis.
2. Listing and standardisation of the data.
3. Development of a classification scheme which illustrates the relationships between different groups of plays and/or characters.

In this paper I will work through this procedure in detail, pointing out some of the problems and pitfalls involved, and describing ways of resolving them. I will draw on examples taken from a numerical analysis of some eighty Nottinghamshire plays.

Selection of cast lists for analysis

In order to be able to conduct a meaningful analysis, only complete cast lists should be used. Often, internal evidence within a record indicates whether or not a cast is complete. However there still remain many records where this is not obvious. Fig. 1 shows the frequency of the number of characters in 206 accounts of plays from Nottinghamshire. Two peaks are obvious, the left hand one representing mostly incomplete casts, and the right hand one mostly complete casts. This pattern can be verified by producing separate graphs for known complete and known incomplete casts. The low point between the peaks occurs at three characters per account, and can be regarded as the borderline between complete and incomplete casts. This indicates therefore that if an account lists four or more characters, unless there is evidence to the contrary, the cast can be treated as complete.

Fig. 1 - Frequency of number of characters in 206 Nottinghamshire Plays



Listing and standardisation of the data

The raw, i.e. unstandardised, casts can initially be listed as in Fig. 2. Note the use of serial numbers and the inclusion of all the alternative names and spellings for each character. As an indication of authenticity, the letters in brackets indicate where each variant name occurs in the record:

- (D) = Used in dialogue
- (L) = Not used in dialogue but known to have been used by the performers
- (N) = Neither used in dialogue nor known to have been used by the performers

Names used trivially can be ignored. The most common example of this is the name "Jack" which the "Doctor" often used trivially when curing his patients. Also, in some scripts, characters are only referred to in the line tags by numbers or number-like names (e.g. "5th Boy", "No. 3 Actor" etc.) Although this is a practice worth noting, the individual designations themselves do not contribute to analysis since a character's number will change from record to record. I only include these individual names when there is no other name available.

Fig. 2 - Part of a listing of Nottinghamshire folk play casts

- PO52 BRINSLEY (SK4649) 1930
 P. T. Millington Collection ("Old Brunsly-ite")
 Christmas, Guysers (Guysering) play
 Master of Ceremonies (L); Saint George (D); Slasher (D);
 Old Man (L)/Slasher's Father (L); Doctor (D); Beelzebub (D);
 Last Character (L)/Devil Doubt (D)
- PO61 BULCOTE (SK6644) Revived 1952 & 1953
 M. W. Barley Collection (Mrs. P. Underwood)
 Plough Monday play
 Tom Fool (L)/Bold Tom (D); Farmers Man (D); Lady (L)/
 Lady Bright and Gay (D); Recruiting Sergeant (D);
 Dame Jane (D); Beelzebub (D); Doctor (D)
- PO71 BULWELL KILNYARDS (SK5345) 1872
 F.M.E.W. (1923)
 Christmas, Guysers play
 St. George (N); Doctor (N); Beelzebub (N); Bess (N);
 Jack (N)
- PO91 CAUNTON (SK7460) Mid 19th Century
 D. Hole (1902)
 Plough Monday, Morris Dancers play and dance
 Robin Hood (D); Little John (D); Maid Marion (D); Fool (N)/
 Clown (N)
 (numerical designation used)

Variation of character names

Before describing the techniques which can be used in standardising character names, it is appropriate at this stage to examine how and why these names vary as well as how characters evolve. Characters can be divided into two broad categories: those which are named in dialogue and those which are not. The names of characters which are given in dialogue are not prone to vary because they have to fit in with the metre and rhyme of the play. Consequently, what name variations there are, apart from minor orthographical variations, tend to be significant. Usually the same names are used in line tags as are used in dialogue, although often in abbreviated form. On the other hand, some collectors and informants chose to supplement names given in dialogue by using descriptive names in line tags (e.g. "Beelzebub" becomes "Devil"). These supplementary names do not contribute to analysis, but they can be helpful in identifying characters listed in accounts which lack recorded dialogue.

Characters which are not named in dialogue can have a wide variety of names, lacking as they do the controlling influence of metre and rhyme. The name which is recorded in a given account seems to depend primarily on the informant

or collector, and it is easy to find cases where co-performers have given different names for the same character, and even cases where different names have been recorded from the same informant at different times. With these characters therefore, roles and typical speeches are more reliable identification than names, although there are a few exceptions (e.g. "Tom Fool") where tradition keeps the name constant.

Evolution of characters' names

Characters' names can evolve either qualitatively and/or quantitatively. Qualitative evolution implies the alteration of words, lines, etc. in the play. In the case of names this involves such changes as "St. George" to "King George", "Mickey Bent" to "Neddy Gent", etc.

Quantitative evolution on the other hand involves the addition or deletion of lines, characters, etc., without altering the original material. As regards character analysis we are principally concerned with (a) the division of a character into two, (e.g. the creation of a special character to interrogate the Doctor), and (b) the fusing of characters to form a hybrid with two or more names (e.g. "In comes I the Recruiting Sergeant, Bold Slasher is my name"). One has to be careful not to confuse hybrids with characters that normally have two names. Likewise one needs to watch for cases of erroneous hybrids where either the informant or collector has confused the record.

Many of the changes described above tend to be single occurrences, in which case there may not be a lot of point in giving them special treatment (i.e. it may be more practical to treat hybrids as two characters, or to group a unique variant, e.g. Neddy Gent, in with the main form, e.g. Mickey Bent). On the other hand, if a particular hybrid or qualitative variant occurs more than once it should be treated as a distinct character, at least until preliminary analysis shows whether or not the variation is significant. Consequently we end up with pairs or groups of related names. It is usually not possible to say which is the earlier form, and it is therefore prudent to speak of them as equal subtypes. Examples are "King George" and "St. George", "Beelzebub" and "Betsy Beelzebub", "Doctor" and "Dr. Brown", etc.

Producing a standardised list of characters

A standardised list of characters needs to be produced methodically, starting with names which occur in dialogue and coalescing spelling variations as well as other names used as synonyms within individual accounts. Characters which are not named in dialogue are grouped according to their roles and/or speeches. It may be possible to link some of these characters with those names in dialogue. The detailed procedure used for the Nottinghamshire plays is given in Appendix A. A standardised list will normally be presented in a basic form as in fig. 3.

Fig. 3 - List of Standardised Character Names

Key to numbering system

'C' indicates the number is for a character. The first two digits number each group of names, and the final digits distinguish names within a group. Asterisks indicate artificial names created to group related names.

CO10	Introducer (not named in dialogue)	C240	Sankey Benny
CO11	Tom Fool	C250	Jenny Wibble
CO12	Bold Anthony	*C260	Mickey Bent/Neddy Gent
*CO20	Disputant	C261	Mickey Bent
CO30	George	C262	Neddy Gent
CO31	King George	*C270	Molly/Sally Mop
CO32	Prince George	C271	Molly Mop
CO33	Saint George	C272	Sally Mop
CO40	Soldier	C280	Polly Flinders
CO41	Recruiting Sergeant	*C290	Rake/Snake
CO42	Slasher	C291	Rake
CO50	Turkish Knight	C292	Snake
CO60	Bull Guy	C300	Tommy Tup
*CO70	Prince/Princess of Paradise	C310	Tup
CO71	Black Prince of Paradise	C320	Sally
CO72	Princess of Paradise	C330	Rolling-Tolling-Tippling-Tom
CO80	Prince of Orange	C340	Me and our owd lass
*CO90	Jack/Tom Steel	C350	Leader (of animal characters)
CO91	Jack Steel	C360	Butcher
CO92	Tom Steel	C370	Robin Hood
C100	Threshing Blade	C380	Maid Marion
*C110	Farmer's Man/Recruit	C390	Little John
C111	Farmer's Man (... whip in hand)	C400	Father Christmas
C112	Recruit	C410	Cut the Dash
C120	Salina	C420	Little Johnny Jack
C130	Lady	C430	Oliver Cromwell
C140	Dame Jane	C440	Beast
*C150	Victim's Parent	C450	Forman
*C160	Interrogator	C460	Vintner
C170	Doctor	C470	Old Squire
C171	Doctor Brown	C480	Plough Bullock
C172	Mr. Killy-me-I-do	C490	Harmonica Player
C180	Eezum Squeezum	C500	Little Clown
C190	Beelzebub	C510	Mr. Miller
*C191	Betsy Beelzebub	C520	King of England
C192	Dick-Beelzebub	C530	King Sam
C193	Hub Bub	C540	Sam
*C210	Collector	C550	Bess
C220	Devil Doubt	C560	Jack
C230	Hopper Joe	C570	Ancient

It is also useful however to supplement this with a further list showing spelling variations, synonyms and related names, since this aids subsequent processing of the cast data. A suitable format to use is that employed by a standard technical thesaurus¹ (See Fig. 4)

Fig. 4 - Part of a thesaurus of Nottinghamshire character names

Key to abbreviations

USE = Use preferred term NT = Narrower term
 UF = Use for nonpreferred term RT = Related term
 BT = Broader term

Dame Jane

UF Buxom Joan
 Deb Jane
 Jane
 Jenny
 Jinny
 Old Dame Jane
 Old Jenny
 Old Woman (Disputant)

Deb Jane

USE Dame Jane

Devil

USE Beelzebub

Devil Doubt

UF Devil-Doubt
 Devil Dout
 Last Character
 Little Devil Doubt

Devil-Doubt

USE Devil Doubt

Devil Dout

USE Devil Doubt

Development of a classification scheme

Conventional manual analysis of plays is generally a question of pattern spotting. This requires a degree of intuition and is prone to subjectivity. On the other hand, it may parallel the more objective procedures used in numerical classification known as cluster analysis. Here plays are successively grouped into interconnected clusters on the basis of the measured similarity between every pair of plays in the data set. Full details of these procedures are given in Sokal and Sneath², Cole³ and Clifford and Stephenson⁴. There are also computer programmes available, including CLASP⁵ and CLUSTAN⁶. Whichever technique is used, it is usually helpful to draw up the standardised data in tabular form in order to aid comparison. In numerical classification this table is called a data matrix (fig. 5).

Dick-Beelzebub

BT Beelzebub
 RT Betsy Beelzebub
 Hub Bub

Disputant

UF 3rd Boy
 King George's Son

Doctor

NT Doctor Brown
 Mr. Killy-me-I-do

Doctor Brown

UF Doctor M. D. Brown
 BT Doctor
 RT Mr. Killy-me-I-do

Doctor M. D. Brown

USE Doctor Brown

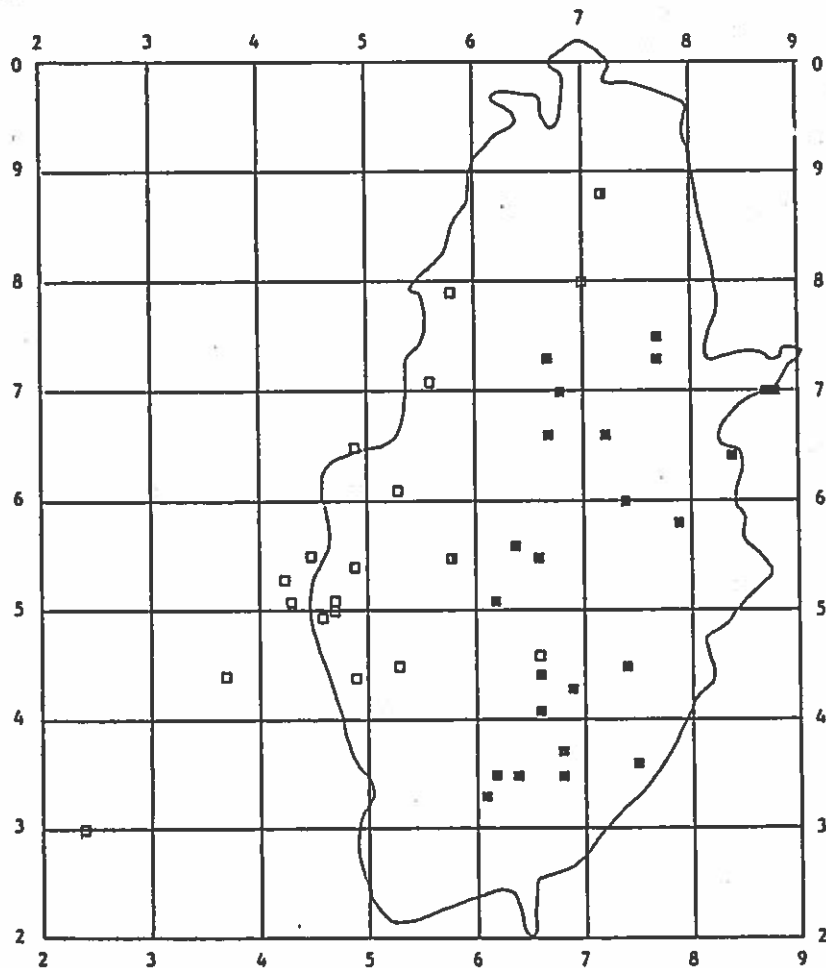
Doorman

USE Introducer

Door Opener

USE Introducer

Fig. 8 - Distribution of the two main types of Nottinghamshire folk play identified using numerical classification



These points show that numerical classification produces results at least as acceptable as conventional methods. In fact numerical classification is probably better in that it is highly objective, the classification being generated by the data whereas conventional methods are prone to subjectivity - the mind observes the data, formulates a classification, and sees if the data will fit it.

Further considerations

The examples used in this paper have been drawn from an analysis of eighty Nottinghamshire plays. Unfortunately, because of limitations in the computer facilities available to me it has not been possible to complete some of the analyses. The next step in this particular classification exercise would be to determine which characters are typical of each group. Although it is possible to make general observations (e.g. "Slasher" is mainly found in the western group, and the "Recruiting Sergeant" is mainly found in the east), some characters cannot be meaningfully apportioned to particular groups or subgroups. This is because the subdivision of some characters in the standardised list into variants (as described under "Evolution of characters") is inappropriate. Therefore, to improve the classification of the plays, and to enable characters to be properly linked with different groups, this analysis needs to be rerun using a refined list of standardised characters. This can be done in two ways. Firstly, the distribution of characters between groups and subgroups in fig. 6 can be analysed. If the subdivision of characters is correct, then they will occur in distinct and separate groups (e.g. "Slasher" and "Recruiting Sergeant"), but if the subdivision is artificial the characters will occur indiscriminately within a group or groups (e.g. "Rate" and "Snake"). Alternatively, the numerical analysis can be reversed so that a tree diagram is produced showing relationships between character names rather than relationships between plays. Invalid character subdivisions would appear closely related in such a dendrogram, whereas valid subdivisions would appear distantly related.

Conclusion

In this paper I have described the main problems associated with the analysis of character names, and have tried to demonstrate that high standards can be obtained using numerical techniques. Numerical classification is not a substitute for conventional classificatory methods. Rather it is an extremely useful analytical tool which enables better pattern recognition and enhances objectivity. Although I have performed this analysis using cast lists, numerical analysis can also be used to examine other features of folk plays. P. S. Smith⁸ for instance has used the texts of the plays. By adopting slightly more complicated techniques one may also look simultaneously at play type, time of occurrence, year of performance, collective name for actors, etc.

NOTES

1. For details of technical thesaurus construction see J. Aitchison and A. Gilchrist, Thesaurus construction: a practical manual (London: Aslib, 1972).
2. R. R. Sokal and P. H. A. Sneath, Principles of numerical taxonomy (San Francisco: W. H. Freeman, 1963).
3. A. J. Cole, Numerical taxonomy: Proceedings of the Colloquium in numerical taxonomy, University of St. Andrews, Sept. 1968 (London: Academic Press, 1969).
4. H. T. Clifford and W. Stephenson, Introduction to numerical classification (New York: Academic Press, 1975).
5. G. J. S. Ross and F. B. Lauckner, CLASP: a classification program. Document: APPLINFO.CLASPMANUAL (Manchester: University of Manchester Regional Computer Centre, June 1974).
6. D. Wishart, CLUSTAN1A user manual (Corrected edn.) (Edinburgh: Jan. 1978).
7. E. C. Cawte, A. Helm and N. Peacock, English ritual drama: a geographical index (London: Folk-lore Society, 1967).
8. Paul S. Smith, "The Problems of Analysis of Traditional Play Texts: A Taxonomic Approach", Traditional Drama Studies, I (1985), 43-65.

Appendix A - Procedure used for compiling the standardised list of Nottinghamshire characters

- (1) A Card Index of character names was compiled, allowing one card for each spelling variation and the like. Under each name, using three separate columns, were listed the numbers of the plays in which the name occurred.

e.g.

<u>KING GEORGE</u>				
<u>D</u>			<u>L</u>	<u>N</u>
PO21	P301		PO42	PO55
P102	P302		P221	
P171	P361			
P172	P431			
P241				
P271				
P281				
P291				

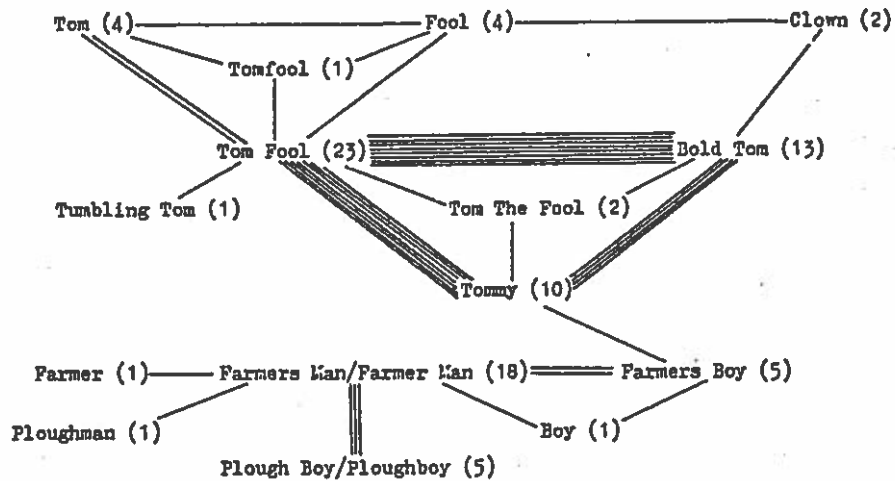
Key D = Name occurs in dialogue
 L = Name does not occur in speeches but is known to be the name used by the performers
 N = Name neither occurs in dialogue, nor is known to be used by performers

- (2) The cards were divided into two sets (i) Names mostly occurring in dialogue, and (ii) names mostly not occurring in dialogue.
- (3) All spelling and other minor orthographical variations of names occurring in dialogue were grouped together.
- (4) To these were added other obvious orthographical variations not occurring in dialogue.
- (5) Diagrams linking synonymous names (see Appendix B) were constructed for all names occurring in dialogue, and used to identify other synonyms and related names.
- (6) The roles or typical speeches were determined for each of the characters on the remaining cards (e.g. Introducer, Victim's parent, Interrogator, Collector, etc.) and the names grouped together accordingly.
- (7) The roles/speeches of characters not named in dialogue were compared with the roles/speeches of those that were, in order to identify any other possible links.

Appendix B - Diagrams linking synonymous names

To help identify synonyms, related names and names capable of dual interpretation, diagrams can be drawn linking those names which occur as alternatives in individual records (see below). These diagrams have to be used carefully however because of two main inherent defects. Firstly, where some names are hybrids, the diagrams may show spurious links between unrelated names (e.g. between "Tommy" and "Farmers Boy" in the diagram below). Such links can however usually be spotted by the fact that the two names normally occur as distinct characters in other plays. Secondly, because some variants are not given alternative names in any record, they will not feature in the diagram (e.g. in the example below "Thomas" does not appear). Therefore their relationships have to be determined by comparing speeches or roles or both.

Fig.2 - Examples of diagram linking synonymous names



Lines linking names indicate that they are used as synonyms for the same character in at least one play. The number of lines equals the number of plays where this happens. The numbers in brackets indicate the number of times each name occurs.