Ecdotica
9
(2012)

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

C^E_CE
CENTRO PARA LA EDICIÓN DE LOS
CLÁSICOS ESPAÑOLES

FONDAZIONE
CASSA DI RISPARMIO
IN BOLOGNA

AQUALOGY

INDICE

Questioni

Rassegne

Saggi

THE TEXTUAL TRADITION OF DANTE'S «COMMEDIA» AND THE «BARBI *LOCI*»

PETER ROBINSON

The textual problem with the *Commedia* is simply stated. As its most recent editor, Prue Shaw, puts it: «The first problem facing any editor of Dante's *Commedia* is the richness of the manuscript record – around 600 copies if we count only complete texts of the poem, more than 800 if we include partial and fragmentary copies».[1] In terms of size: no editor, no team of editors working with the traditional methods of Italian philology could ever find the resources to look at every reading in every one of the 14233 lines of the *Commedia* in every one of the 800 plus manuscripts and significant early print editions of the poem.

Nor is the problem just the number of manuscripts. It is known that three early and very influential recensions of the whole text were made by Giovanni Boccaccio between 1355 and 1373, and that all three primary Boccaccio copies, though based on the still-extant Vatican manuscript known to *Commedia* editors as «Vat», introduced many readings from many other manuscripts: what editors know as «contamination». From that point on, the manuscripts show such a mixture of readings that standard stemmatic processes, depending on the orderly copying of

[1] «Introduction» in Prue Shaw, ed., *Dante Alighieri. Commedia. A Digital Edition*, Birmingham-Florence, Scholarly Digital Editions-SISMEL-Edizioni del Galluzzo, 2010.

readings from one manuscript into another, appear impossible.[2] Even if one could record every reading in every witness there is common agreement among many scholars that traditional stemmatic analysis, as identified with Lachmann and the followers of the method named for him, cannot deal with cases of contamination and hence it is impossible to construct a useful genetic hypothesis concerning the relations among the many manuscripts of the *Commedia*.[3] It appears, then, that we are faced with an impossible situation. The fundamental importance of the *Commedia* to Italian and world literature makes it imperative that Dante's master-work be edited, and edited to the highest possible standard. Yet, either difficulty – the sheer size of the tradition or the prevalence of contamination – would on its own make it impossible to achieve what one could argue are the two minimal requirements of a scholarly edition: first, that it examine all the evidence in all the witnesses; second, that it derive from all this evidence an understanding of the whole tradition which can then be used by the editor to identify which manuscripts and which readings within them are most likely nearest to Dante's own text.[4]

Over the last centuries of Dante scholarship, scholars have tried several routes past this impossibility. Giorgio Petrocchi, editor of the most

[2] Indeed, manuscripts datable from before 1335 already show clear signs of contamination: that is, the import of readings from manuscripts other than the exemplar. The Landino manuscript (La), dated to 1336, has readings scraped away and others substituted across its whole length. The even earlier manuscript, not now extant, dated to 1330-1331 and used by Luca Martini in his collation of the 1515 Aldine text, also included readings from manuscripts other than the exemplar (Shaw, «Introduction: Overview»).

[3] Among Middle English textual scholars, the view that no useful genetic hypotheses can be created for large manuscript traditions has become so widely accepted as to count as gospel: see George Kane's editions of *Piers Plowman* and his assault on John Manly and Edith Rickert (who did think they could disentangle the 80-plus manuscripts of the *Canterbury Tales*) in Paul Ruggiers, ed., *Editing Chaucer: the Great Tradition*, Norman, Oklahoma, Pilgrim Books, 1984.

[4] This pessimistic view, that no genealogical representation of the relationships among the manuscripts of the *Commedia* is possible, was reached by the nineteenth-century English scholar Edward Moore, who examined hundreds of manuscripts of the *Commedia* only to conclude «My own belief would be that owing to the complicated intermixture of texts, such a genealogy never can be constructed.» (Edward Moore, *Contributions to the textual criticism of the* Divina Commedia, *including the complete collation throughout the Inferno of all the mss. at Oxford and Cambridge*, Cambridge, Cambridge University Press, 1889, p. xxxi). Shaw cites the German scholar Witte as implicitly reaching the same conclusion in the Prolegomena to his 1862 edition of the poem, based on his own independent examination and collation of hundreds of manuscripts («Introduction», n. 22).

significant modern edition of the *Commedia*, addressed both problems with a single strategy: that of basing his edition of the *Commedia* «secondo l'antica vulgata» only on the 27 manuscripts (twenty-four complete, three fragmentary, and counting Martini's collation of the Aldine edition as a manuscript) accepted at the time of his edition as dating securely before 1355.[5] This approach both reduced the number of manuscripts which he needed to examine to a manageable number, and also greatly reduced the problem of contaminated manuscripts by looking only at manuscripts written before Boccaccio started work on his recensions of the *Commedia*. Petrocchi stated very clearly that no edition based on such a small selection of manuscripts could claim to be «un'edizione critica», which must be based on all the manuscript evidence, not just some of it.[6] Accordingly, he claimed only the more modest title of «La Commedia secondo l'antica vulgata» for his edition. Nonetheless, he also argued that even though his text was based only on this small fraction of the manuscripts, all the later manuscripts would not contribute a single reading not already present in or readily deducible from the antica vulgata manuscripts to a critical text. Therefore, he declared, a sound editorial text based only on the *antica vulgata* manuscripts would be identical to that based on the whole tradition, and so his text could stand for the whole tradition, despite the provisional status implied by the «secondo l'antica vulgata» qualification.

Petrocchi's approach evades, rather than satisfies, the need to form a view of the entire manuscript tradition. Indeed, his assertion that a sound edition could be based on these 27 manuscripts alone could only be validated by looking at all the evidence in all the manuscripts. But how can this be done, given over eight hundred manuscripts, more than 14000 lines of text and 100,000 words in a complete manuscript? Before Petrocchi, the Italian Dante scholar Michele Barbi, then at the beginning of a remarkable fifty-year philological career, and at the instigation of three senior Dante scholars associated with the newly-founded Soci-

---

[5] Dante Alighieri, *La Commedia secondo l'antica vulgata*, a cura di Giorgio Petrocchi, in *Le opere di Dante Alighieri. Edizione Nazionale a cura della Società Dantesca Italiana*, VII, Milano, Mondadori, 1966-1967, 4 voll. Note that in this article all references to Mart, in terms of readings found and analysis based on those readings, are to the Martini collation except where otherwise specified.

[6] «Ai fini di un'edizione critica, e perché tale veramente sia, si deve tuttora tener fede alla legge di partenza e di fondo d'ogni ricerca testuale: l'interrogazione integrale della tradizione.» Giorgio Petrocchi, «Proposte per un testo-base della *Divina Commedia*», *Filologia romanza*, II (1955), pp. 337-365 [p. 343], cited by Shaw, «Introduction: Overview».

età Dantesca Italiana, produced a list of lines in the *Commedia* which he judged critical for the establishment of manuscript relations across the whole text of the *Commedia* and all its manuscripts. These are known as the «400 loci»: in fact, there are 396 of them.[7] Barbi explains in a later article that these *loci* were chosen on the basis of considerable experience working with *Commedia* manuscripts in the Florentine libraries, and on consideration of the significance of particular variant readings.[8] As well as reducing the amount of effort required to survey the whole tradition (one need look at only 396 lines, not at 14233), this also offered the promise of a collaborative approach: the work could be divided among many scholars. Accordingly, special forms were printed off, to be distributed to scholars and so enable a complete survey of these key lines in every manuscript everywhere. As Shaw relates, the response was disappointing: some eleven scholars, including Barbi himself, examined a few manuscripts and reported their findings. Among these eleven were Giuseppe Vandelli and Mario Casella: yet when these two scholars came to produce their own editions of the *Commedia* in 1921 and 1923 respectively, neither used the collation of the Barbi *loci* as the base for an analysis of the whole tradition, and hence their edition.

As we have seen, the next great edition after Vandelli, that of Petrocchi in 1965, also made no use of the Barbi *loci*, choosing a different rationale for a selection of manuscripts on which to base the edition. Lanza's 1995 edition took an even more extreme approach to the problem of how to derive a single text from a vast number of manuscripts: he used just one manuscript, the Trivulziano 1080 («Triv»), one of the very earliest surviving Florentine manuscripts and long famous for the quality of its text and the beauty of its decoration. In this choice, Lanza was acting consciously in the tradition of «best-text editing» associated with Joseph Bédier. Following his work on *Le Lai de l'ombre*, Bédier asserted the impossibility of creating a useful genetic hypothesis about any large and complex manuscript tradition.[9] In his view, Lachmannian stemmatics is an impossibil-

---

[7] The three scholars were Adolfo Bartoli, Alessandro D'Ancona, Isidoro Del Lungo, writing in «Per l'edizione critica della *Divina Commedia*», *Bullettino della Società Dantesca Italiana*, 5-6 (1891), pp. 25-27, and followed by Barbi's «Canone di luoghi scelti per lo spoglio dei mss. della *Divina Commedia*» on pages 28-38. All the Barbi *loci* are listed in Appendix A of Shaw's edition, with convenient hypertext links to the edition's collation of each line in the seven manuscripts and two editions included within it.

[8] «Fu frutto di lunghi studi, e fissata quindi non a priori, cioè a caso … »; Barbi, «Ancora sul testo della *Divina Commedia*», *Studi danteschi*, XVIII (1934), p. 56.

[9] «La tradition manuscrite du Lai de l'ombre», *Romania*, 54 (1928), pp. 161-296, 321-356; reprinted Paris, E. Champion, 1970.

ity in such circumstances. The only recourse then is to identify the best single witness to the text, by whatever criteria are available, and use that as the base for an edition. As Lanza also argued, such an edition can (if rigorously faithful to the source) present a text which actually existed at a single historic moment, unsullied by editorial impulse or linguistic revision. In fact, Lanza did permit emendations from other manuscripts into his edition, but only under stringent circumstances and from a very limited range of other Florentine manuscripts. Of course, this approach frees the editor completely from the need to survey the entire tradition, whether by using the Barbi *loci* or by any other means.

Around 1996, then, Barbi's *loci* looked like a road-not-taken in *Commedia* textual scholarship: an interesting idea but never implemented. However, in 2001 a new edition of the *Commedia* appeared, edited by Federico Sanguineti, in which the Barbi *loci* moved back to the centre stage of *Commedia* textual scholarship.[10] This edition was startling for several reasons. First, where Petrocchi, Lanza and other editors agreed that it was impossible to form a useful genetic hypothesis of the entire tradition, Sanguineti claimed that it was possible to achieve this for the *Commedia*. In opposition to Bédier's gloomy pessimism about stemmatics, Sanguineti asserted that Lachmannian stemmatics could be practiced rigorously and usefully on the whole *Commedia* tradition. Second, Sanguineti declared that not only could traditional stemmatics be applied to the whole tradition, but he had done it. He had looked at all the Barbi *loci* in every one of the 800 manuscripts (so achieving on his own, with virtually no support, what scholars had failed to achieve in over a hundred years), and from analysis of the readings at these *loci* he had created a comprehensive account of the whole tradition, and isolated just seven manuscripts as necessary and sufficient for the creation of a critical text.

Sanguineti's identification of the «Sanguineti seven» was the starting point of Shaw's edition. It seemed possible to produce a digital edition which would do for these seven manuscripts what Shaw's edition of Dante's *Monarchia* had done for the manuscripts of that work: that is, offer full machine-readable transcripts of them, create an extremely precise collation of the transcripts, and then use various methods (including phylogenetic analysis, derived from evolutionary biology) to create a view of the relationships among the manuscripts.[11] At first, Shaw (and I) planned

[10] Dantis Alagherii *Comedia*, edizione critica per cura di Federico Sanguineti, Firenze, Edizioni del Galluzzo, 2001.

[11] Prue Shaw, ed., Dante Alighieri, *Monarchia. A Digital Edition*, Birmingham-Florence, Scholarly Digital Editions-Società Dantesca Italiana, 2006. See the review article by Paolo

to work with Sanguineti in the making of this digital edition, and there were several substantive discussions towards this aim. However, Sanguineti left the partnership at an early stage, and almost all the work of the edition proceeded without his involvement. In turn, the aim of the edition changed: it became a test of Sanguineti's arguments about the relationships among these seven manuscripts. Shaw examines Sanguineti's claims in considerable detail, and there is no need here to do more than summarize her criticisms of them, both in terms of his conclusions and of his methodology. Briefly: it appears from her analysis, based not on any selection of variants but on the entire body of variation in the seven manuscripts, firstly that Sanguineti's key conclusion is wrong, and secondly that the methodology he used to reach this conclusion was flawed. Sanguineti's key conclusion is that one manuscript, and one manuscript alone – Vatican Library ms. Urbinate latino 366 (Urb) – represents a pure line of descent from Dante's original text independent of every other manuscript. Thus his stemma of the seven manuscripts (FIGURE 1):

FIGURE 1

Sanguineti's stemma of the «Sanguineti seven» manuscripts of the *Commedia*. Note the positioning of Rb as sharing an ancestor α with the five manuscripts LauSC Mart Triv Ash Ham, leaving «U» (Urb) as independent of Rb and all other manuscripts.
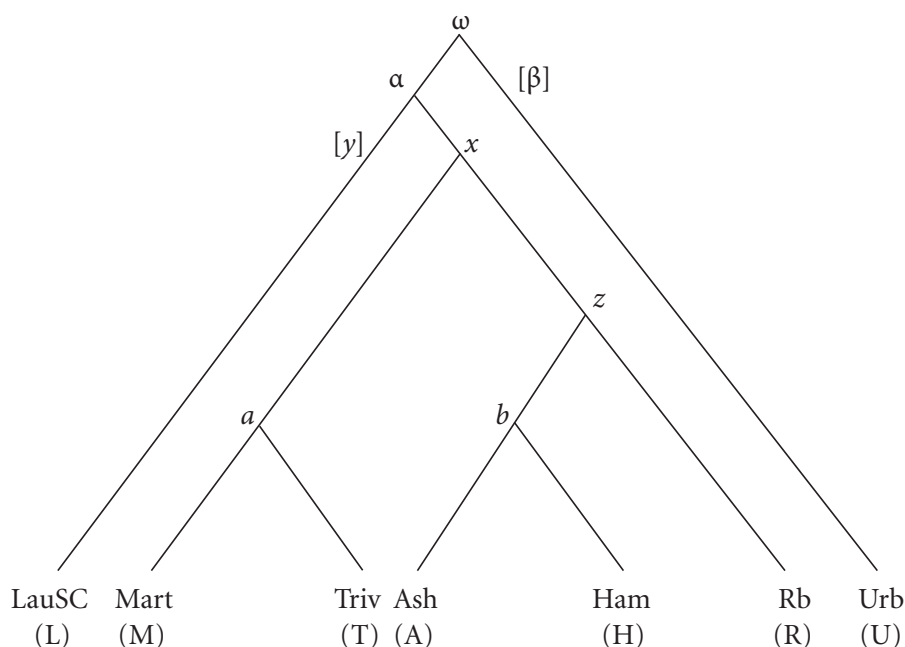
Trovato (with an appendix on the Shaw *Commedia* edition) in «La Doppia "Monarchia" di Prue Shaw (con una postilla sulla *Commedia*)», *Ecdotica*, 7 (2010), pp. 193-207.

Here, Dante's original is ω, and Sanguineti argues that six of the seven descend from a single copy of ω, labelled α, while Urb is on its own, descended independently from ω. Accordingly, the one manuscript Urb has the same authority as the other six combined. Where most editions have centred their text around the Florentine manuscripts in the α branch of this stemma (thus, Lanza used Triv as his base, Casella LauSC; Petrocchi often favoured readings in Triv and Mart ahead of those in Urb) Sanguineti based his entire edition on Urb. Sanguineti's argument depends on Urb sharing no ancestor below ω with any other manuscript, and so contradicts Petrocchi, who argued that Rb and Urb share an ancestor (his *e*) below the original (FIGURE 2). As Shaw asserts, «the position of ms. Rb in the stemma is critical for Sanguineti's argument:

FIGURE 2

Part of Petrocchi's stemma of the *antica vulgata* manuscripts of the *Commedia*. Note the positioning of Rb, as sharing an ancestor with Urb (and Mad).
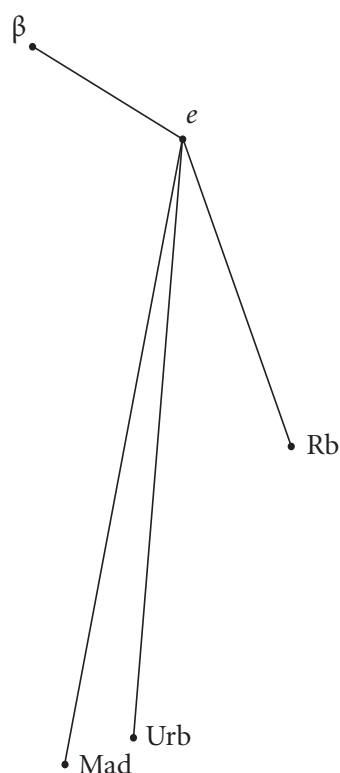
if he is wrong about that, *tutto crolla* – the whole edifice collapses.» Her analysis, supported by phylogenetic analysis (see further below), is decisive: Petrocchi is right, Rb and Urb do share an ancestor, and Sanguineti is wrong. Further, one can trace Sanguineti's error directly to his meth-

odology. Although Sanguineti claimed that he based his analysis on the 396 Barbi *loci*, Shaw points out that he actually uses very few of them to support his conclusions. Sanguineti identifies over four hundred manuscripts as members of a single group («la cosiddetta "tradizione β"») on the basis of just four readings. Further, one of these readings (*Par.* xxiii 103, *spiro* for *giro*) has the character of a classic polygenetic error, which could readily occur independently in unconnected scribal copies. Elsewhere, he justifies discarding the evidence of this large group (expanded to over 600 manuscripts on the basis of a very few readings) because it is contaminated by readings drawn from the distinct traditions which he sees underlining the seven manuscripts: thus, the appearance of a single reading from Rb in some of these 600 manuscripts (*ale* for *aer* at *Purg.* ii 35) is used to argue that this group is contaminated from Rb or a manuscript closely related to it.

Manifestly, arguments based on just a few readings in a text of over 100,000 words will lack conviction: a reason why Barbi suggested analysis of some 400 lines of text, not just twenty or so. Ideally, one would base analysis on the whole text, on every variant at every word in every line. To do this for all 800 manuscripts would be impossible with the resources available to any current project; but one could do it for a smaller number: indeed, for the seven manuscripts identified by Sanguineti, and this was what Shaw did. The base of her edition was a complete transcription of the seven manuscripts, and creation of a very precise collation of every word of every one of these seven with each other, and with two major edited texts: those of Petrocchi and Sanguineti. While the edition's eventual aim was to use this collation and the analysis built on it to explore Sanguineti's hypothesis, as explained in the last paragraphs, this same collation can be used for another purpose: to test the efficacy of the Barbi *loci* for their declared aim (as accepted by Sanguineti), as a base for an analysis of the entire tradition. Of course, in this edition we have only seven manuscripts. However, the key place in the tradition of at least six of these, accepted by every scholar in the last century, makes these manuscripts a good place to start. Moreover, as we have seen, Sanguineti based his identification of these seven manuscripts as critical to the understanding of the whole tradition of the *Commedia* on his collation of these 396 lines. Accordingly, one should expect that these *loci* should be pre-eminently informative in the analysis of these seven manuscripts.

Shaw's edition proceeded as follows. First, an exact word-by-word collation of the whole text in the seven manuscripts was created as follows:

1. All seven manuscripts were transcribed in full, word by word and line by line, into machine-readable form.

2. The transcription recorded layers of scribal revision in each manuscript, so that one could distinguish at any point what was written by the original scribe, changed by him or her, then changed again by later scribes. This is particularly useful with respect to Mart, where it is actually the readings recorded by Luca Martini from a now-lost manuscript as alterations to the Aldine edition which are of primary interest. Thus, in *Purg.* vii. 51, where the Aldine edition prints «o pur sarria», Martini writes «o non saria» in the right margin (FIGURE 3), both readings are recorded in our transcription. Further, the layers of readings are marked so they can be compared separately.

FIGURE 3
Martini's annotations to the Aldine edition at *Purg.* vii. 51.
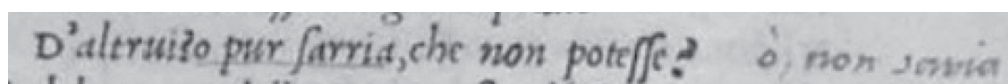
FIGURE 4
The «Literal» view of the transcription, showing both the original text and Martini's annotations as marked in the right margin.
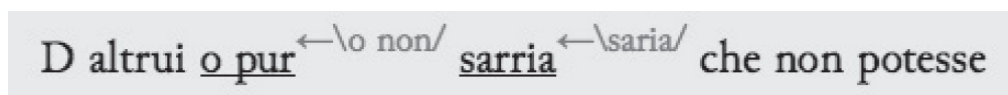
FIGURE 5
The «Aldine Original» view of the transcript, showing the text as originally printed.

FIGURE 6
The «Martini's Collation» view, showing the text with Martini's annotations applied.

3. The transcripts were then compared with one another, and with the texts of the Petrocchi and Sanguineti, by Shaw and her team, using the computer program *Collate*. The collation removed orthographic and spelling variants, as can be seen in this collation of the second word of the first line of the Inferno canto 1 shown in FIGURE 3:

FIGURE 7
Collation of mezzo, *Inf*. i. 1.

| mezzo | Mart PET (mezzo), LauSC (MEZZO), Rb (meço), Ham (mecço), Ash (meço), FS (mezzo), Urb (meggio), Triv (MEÇ\|ÇO) |

Here, the eight different spellings of *mezzo* in the nine witnesses (seven manuscripts plus Sanguineti [FS] and Petrocchi [PET]) are regularized to *mezzo*, so that no variant shows in this line.

The collation also preserved information about the layers of scribal revision, so that one could compare the different states of the text within a manuscript, and between that manuscript and others. Thus, the collation at *Purg*. vii. 51 in FIGURE 8 distinguishes the Aldine text («Mart-orig») and Martini's collation («Mart-c2»), and shows too that the reading «o pur» is found in Urb, and hence in Sanguineti («FS»), while four other manuscripts (Ash Ham Rb Triv) have the same reading as Martini (this is also the reading of Petrocchi) and LauSC alone reads «o uer.»

FIGURE 8
The collation of «o non» in *Purg*. vii. 51.

o non	Ash, Ham, Mart-c2, Rb, Triv, PET
o uer	LauSC
o pur	Mart-orig, Urb, FS

At the end of this process, a complete record of every significant variant at every word of the *Commedia* at every layer of writing in these seven manuscripts was available.

The next step was to analyze this record of variation to see what could be deduced about the manuscript relations. Our analysis followed two paths. First, Shaw used the traditional means of philology, scrutinizing all the variants bearing on her argument, categorizing them, and seeking explanations which take account of the full range of variants. Thus, in the key section «The position of Rb» she looks at twenty-eight variants

discussed by Petrocchi, and then supplements her discussion of these by examination of a further 76 variants. Her argument hinges not only on the nature of the variants in themselves, but also in the consistent agreement of Urb and Rb against all, or nearly all, the other manuscripts:

Some of these categories are not especially significant in themselves (many of them are included in Brandoli's categories of polygenetic error). Singly, they mean next to nothing. But it is the presence of a long series of them uniformly right across the text in a very small number of manuscripts which is striking (and this is surely what Petrocchi's phrase «foltezza di statistica» refers to at least in part). It is simply impossible to imagine that copyists working independently would make precisely these small changes at precisely these same points right across a text of this length. The most economical hypothesis is that they are working from a common exemplar and inherit these readings from that exemplar.

FIGURE 9

The unrooted phylogram for the seven manuscripts, from the Shaw edition. The arrow pointing to the node linking all of Ash/Ham, Urb/Triv, LauSC and Mart/ Triv has been added to this figure.
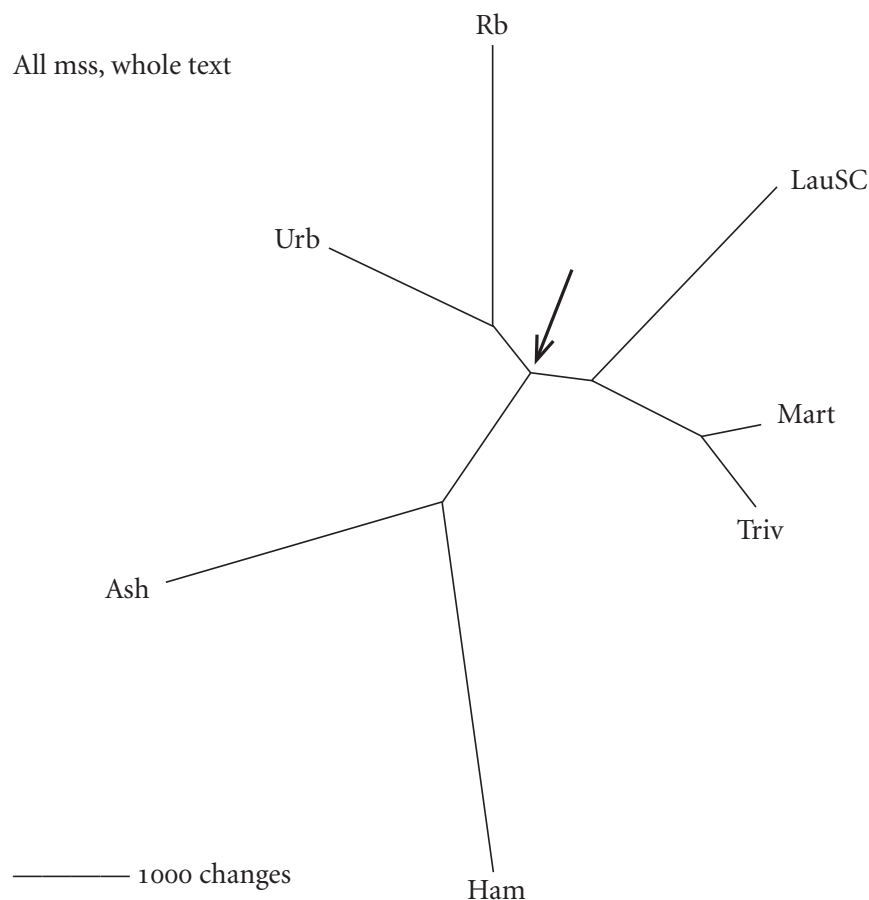
In the second path of analysis, I took the complete record of variation created by the collation and submitted that to analysis by the program PAUP («Phylogenetic Analysis Using Parsimony»), used by evolutionary biologists to create hypotheses about the genetic relationships of organisms based on the characteristics they share and do not share. A full account of how phylogenetic analysis works, and why and how it is suited to the analysis of textual traditions, with evidence of its successful use in many contexts (including on artificial traditions specially devised to test its adequacy) may be found elsewhere.[12] FIGURE 9 gives the «phylogram» for the whole *Commedia* generated by PAUP based on some 94,000 places of variation (typically a single word; but also phrases) in the seven manuscripts.

This representation of manuscript relations differs from a traditional stemma (as for example, that given by Petrocchi). Firstly, it is «unrooted»: that is, it represents the groupings of the manuscripts without any presumption as to originality, or the direction of variation. Secondly, the manuscripts are shown as related to one another through shared nodes, with three pairs each sharing a common node (Mart/Triv; Urb/Rb; Ash/Ham, and the seventh manuscript (LauSC) descending from a node on the line between Mart/Triv and the arrowed node linking the other four manuscripts. The relative lengths of the lines between nodes, and between nodes and manuscripts, are significant, and one can measure approximately the differences between the manuscripts using the scale «1000 changes» (or, ten changes per canto, one every 14.23 lines) on the bottom left. Thus: the line between the node joining Ash/Ham up to the arrowed node which links to the other five manuscripts is about the same length as the scale line (thus, c.1000 changes), and approximately double the length of the line from the arrowed node to the shared ancestor of Urb/Rb (thus, c.500 changes), and approximately the same as from the arrowed node to the ancestor of Mart/Triv (thus, c.1000 changes again).

Experience using these phylograms has taught us to be careful. This phylogenetic analysis appears to suggest that Sanguineti is wrong: Urb and Rb do indeed share a common ancestor, as both Petrocchi and Shaw assert. However, the software, left to itself, will always place a manuscript somewhere on the phylogram: sometimes this can suggest a rela-

---

[12] See, for example, the publications listed at http://www.textualscholarship.org/newstemmatics/bibliography/index.html; notably the collection of articles in Pieter van Reenen, August den Hollander and Margot van Mulken, eds., *Studies in Stemmatology II*, Amsterdam, John Benjamins, 2004.

tionship which in fact does not exist. Also, these phylograms cannot represent contamination. Therefore, we have to turn to other methods to confirm, deny, or qualify the relationships suggested by the phylogenetic analysis.

In our analysis of the tradition of the *Commedia*, one of the «other methods» Shaw and I used was database analysis of the distribution of variants across the manuscripts and across the whole length of the tradition.[13] Our hypothesis is this: consider the manuscripts descending directly from a single node (for example, Urb and Rb). If the manuscripts descended directly from this single node really share a common ancestor below the archetype, then one should be able to identify a set of variants likely to have been introduced into that shared ancestor, and then descending to those manuscripts. Thus, there should be variants which satisfy the following four conditions:

1. They should be present in the manuscripts descended directly from the shared ancestor (thus, in both Urb and Rb);

2. They should be likely not to have been present in the archetype;

3. They should be found rarely in other manuscripts outside those descended from the shared ancestor (thus, rarely in any of Ash Ham LauSC Mart Triv);

4. There should be a significant number of such variants.

We used a database-like search tool, «VBase», to find the variants which satisfied these conditions. This tool is available on the DVD publication, where it is pre-supplied with key searches we found useful in the course of our analysis. Here is our search for variants likely to have been introduced by the joint ancestor of Urb/Rb, and whose existence therefore supports Petrocchi's view that Urb and Rb share an ancestor below the archetype, and contradicts Sanguineti's view that they do not (FIGURE 10).

---

[13] The database analysis was carried out by myself, employing all the data generated by the transcription and collation directed by Shaw and using the criteria based on the Shaw analysis. Shaw and I discussed and reviewed the results, and jointly wrote an article «Phylogenetic Analysis», included in the Shaw edition, drawing heavily on the database analysis. Henceforth, this database analysis is described as the «Shaw/Robinson» analysis, and the variant lists produced by it of the distinct groups as the «Shaw/Robinson» lists.

FIGURE 10

VBase search for variants evidencing a common ancestor below the archetype for Urb Rb.

The first line of this search requests variants only present in both Rb and Urb, and hence likely to have been present in their common ancestor (condition 1, above). The next line asks only to see those variants present in fewer than five (that is, in Urb Rb and no more than two other) witnesses, and hence not likely to have been present in the ancestor of the whole tradition (condition 2). The likelihood is that if a variant is present in three or more other witnesses besides Urb and Rb, then that variant would very probably have been present in the archetype. Indeed, if one increases the number from <5 to <6, or more, we see the search returning more and more variants, as it catches readings likely present in the archetype (thus: 488 readings <6; 806 readings <7; 1881 readings <8; 6109 <9). Accordingly, this restriction (returning only read-

ings in fewer than five witnesses) satisfies the second condition given above. Note that because we include the Sanguineti edition («FS») in our collation, and Sanguineti frequently chooses the Urb reading over that of other manuscripts, as in all three cases from *Inferno* 1 given here, the restriction to five is effectively Urb + RB + FS plus one other manuscript only (in fact, fully 247 of these 308 readings returned by this search are present in FS).

The next condition (condition 3) is that the readings evidencing an exclusive common ancestor for Urb and Rb should be rarely found in any other manuscript. This table gives the number of occurrences of any of these 308 readings across all the witnesses here collated:

TABLE 1
Occurrence of Urb/Rb readings in all witnesses.

Manuscript	Number of Urb/Rb readings (max. 308)	Percent of Urb/Rb readings present
Ash	40	13
Ham	46	15
LauSC	11	4
Mart	4	1
Rb	308	100
Triv	9	3
Urb	308	100
FS	247	80
PET	19	6

Because of the operation of co-incident (or convergent) variation, it is to be expected that by simple chance, a proportion of readings introduced by one scribe in one copying might be found in a quite unrelated copy made by a different scribe. One might hypothesize that in any two acts of copying, the different scribes might introduce the same mistake some ten percent of the time, or in one in ten variants. On this calculation, the incidence of the 308 Urb/Rb in all except FS Ash Ham can be explained as simple coincidence – and the high number of Urb/Rb variants present in FS results from Sanguineti's decision to base his edition

on Urb. The number of Urb/Rb variants in Ash and Ham is marginally higher than one might expect, and this may be the result of contamination between the Urb/Rb and Ash/Ham groups (see below). Accordingly, this group of variants satisfies the third condition given above: readings from this group are rarely found in other manuscripts.

The last condition is that the number of variants should be «significant»: that is, not just a random agreement between these two witnesses but the likely result of shared descent from a single ancestor containing these readings. Again, one can use VBase searches to explore what levels of agreement between any two witnesses might be the result of co-incident variation. From the phylogram given in FIGURE 9, one may hypothesize that the following pairs of manuscripts do not share a common ancestor below the archetype: Rb/Mart, Rb/Triv, Urb/Mart, Urb/Triv, Urb/LauSC, Rb/LauSC, Rb/Ham, Urb/Ham, Rb/Ash, Urb/Ash, Ash/LauSC, Ham/LauSC, Mart/Ash, Mart/Ham, Triv/Ash, Triv/Ham, Mart/LauSC, Triv/LauSc. Here are the results of VBase searches on these pairs of manuscripts, using the same criteria employed for the search given in FIGURE 10:

TABLE 2

Number of readings found in hypothetically unrelated witness pairs, likely to be the result of chance agreement. All searches were for: witness pair + <4 of all witnesses, i.e. witness pair + one other, except for pairs including Urb (marked *), where searches were for witness pair + <5 of all witnesses, i.e. witness pair + FS + one other.

Witness Pair	Number of readings found by Vbase
Rb/Mart	37
Rb/Triv	91
Urb/Mart*	28
Urb/Triv*	58
Urb/LauSC*	99
Rb/LauSC	70
Rb/Ham	219
Urb/Ham*	203
Rb/Ash	201

Urb/Ash*	167
Ash/LauSC	103
Ham/LauSC	91
Mart/Ash	45
Mart/Ham	57
Triv/Ash	102
Triv/Ham	123
Mart/LauSC	35
Triv/LauSc	107

For all except the pairings involving Urb/Rb/Ash/Ham, these searches typically return fewer than one hundred variants (ranging from 35 for Mart/LauSC up to 123 for Triv/Ham). One may infer from this that for this text, any two witnesses may happen to share up to around 100 readings (or one a canto) and it will mean nothing whatever: the agreement is just chance. Conversely, one may deduce that chance cannot explain two witnesses sharing 308 variants (or three a canto): this is likely to be the result of shared descent from a common ancestor. Thus, these 308 readings may satisfy the fourth condition: their number is significant of more than chance agreement.

I except here the pairings involving Urb/Rb with Ash/Ham. The number of agreements between any two of these ranges from 167 (Urb/Ash) to 219 (Rb/Ham). This seems higher than one might expect from random agreement, and we recall that the number of Urb/Rb readings found in Ash and Ham given in Table 1 was also rather higher than one might expect from random agreement. Shaw and I examine this question in our «Phyologenetic analysis» argument and suggest that this agreement across the pairs Ash/Ham and Urb/Rb is likely to be the result of contamination, particularly in *Inferno* 1-7 and the last thirteen *canti* of *Paradiso*.

To summarize: by this process of VBase searching, built upon the results of the phylogenetic analysis, we isolated a list of 308 variants which we believe are evidence of a family relationship between the two manuscripts Rb and Urb. The phylogram (FIGURE 9) suggests that there are two other manuscript pairs among the seven: Ash/Ham and Mart/ Triv. Further, it may be that the five witnesses Ash/Ham/Mart/Triv/

LauSC (that is, the two pairs Ash/Ham and Mart/Triv plus LauSC) may also descend from a single copy below the archetype, and therefore also form a family grouping within the tradition. Following Petrocchi, we call this group α. For all these groupings, we carried out similar VBase searches to that we described above for Urb/Rb, identifying for each a set of variants characteristic of each group and likely to have been introduced by the exclusive common ancestor of each. The groups (hereafter, the Shaw/Robinson lists), the number of variants we found for each group, and the number of lines in which these variants occur (rather fewer than the number of variants, as some lines have more than one variant) are given in TABLE 3.

TABLE 3
Numbers of variants evidencing hypothetical manuscript groups, and numbers of lines in which they occur.

Group	No. Variants	No. Lines
Urb/Rb	308	297
Mart/Triv	874	842
Ash/Ham	770	722
α	327	314
TOTAL		1885

Taken together, this gives a total number of 1885 lines containing variants which our analysis indicates are significant for establishing family relationships among these seven witnesses.[14] That is: of the 14233 lines in the whole *Commedia*, it appears that 1885 give evidence of family relationships. Earlier in this essay I described the Barbi *loci*: his list of 396 lines which he chose as indicative for the establishment of manuscript relations across the whole 14223 lines of the *Commedia*. We have 1885 lines which we find are indicative; Barbi suggests 396 which he believed to be indicative.

This gives us a standpoint from which we can assess Barbi's choice of lines. Both Barbi and ourselves (Shaw and myself) were seeking the same thing. Accordingly, one would expect a high degree of correspond-

---

[14] If one adds the numbers of lines (297+842+722+314) the sum is 2175, not 1885. This is because 290 lines contain variants from more than one group.

ence between the two lists of lines: say, with very many of Barbi's 396
lines appearing in our larger list of 1885 lines. Of course, the same lines
might appear on both lists and mean nothing more than chance agree-
ment. In fact, there are well-known probabilistic methods for estimat-
ing in a case such as this what proportion of agreement between the lists
might be simple chance, and what more than chance.

Consider that we have an urn containing 14233 balls, each with a
number from 1 to 14233 drawn on it: our 14233 lines of the *Commedia* .[15]
We choose, purely at random, 1885 balls from the urn: this is the Shaw/
Robinson list. We put all 1885 balls back, and we choose, purely at random,
396 balls from the urn: this is the Barbi list. Now, for any one ball that
we happen to choose, the chance that it is on the Shaw/Robinson list is:

$$1885 / 14333 (=0.13244 \text{ approx})$$

So if we choose 396 balls from the urn, then by simple chance the number
of balls we are likely to get which are on the Shaw/Robinson list will be:

$$396 \times 1885 / 14233 (=396 \times 0.13244) = 52.446$$

That is: by simple chance agreement, if we chose 396 lines at random
several times over (taking care to put back all the balls back in the urn
between choices) we would expect the average number of lines chosen
to approximate to 52.446. Say we did this exercise of choosing two sets of
balls (one, 1885 balls, another 396 balls) a hundred times, two hundred
times, each time noting down the number of balls in common between
the two choices. Sometimes there would be 45 in common, sometimes
59, sometimes 49, sometimes 55 etc.: but the more times we chose the
two sets of balls, the more the average number of lines in both lists would
approach 52.446. Hence, 52.466 is the «mean number».

Probability can do even more for us. The «mean number» is useful,
as we can presume that if we find that the number of lines in common
on the two lists is close to this number, it is mere random chance. Thus,

if we find 52 lines in common between the two lists, this is just chance. But what if we find 51 lines, or 53? Instinctively, we will think this is likely to be chance too. But what of 70 lines, or 100, or 30? At what point does it become highly unlikely that a given number of lines will occur on both lists by chance? Probability can give us answers to these questions too. This situation, where we can assign statistical probabilities to the range of numbers of possible lines in common between the two lists (between 0, for no numbers in common, and 396, with every number on the Barbi list appearing on our list) is an instance of what statisticians call a «hypergeometric distribution», and it is possible to calculate various probabilities for each number of lines between 0 and 396, thus:

1. The probability that exactly this number of lines should be in common in the two lists. For 52 lines, that is 0.0599, or slightly better than one chance in twenty (for 51 it is 0.0592, for 53 0.0594)

2. The probability that there should be this number of lines or fewer in common in the two lists. For 52 lines, that is 0.5102, or just better than one chance in two – the equivalent of tossing a coin and getting heads.

3. The probability that there should be this number of lines or more in common in the two lists. For 52 lines, that is 0.5498, or just better than one chance in two – again, the equivalent of tossing a coin and getting heads.

We can use these figures to tell us, for any given number of lines, not just what the chances are of exactly that number of lines occurring in both lists, but also what are the chances of a lesser or greater number of lines occurring. That is: we can use the hypergeometric distribution table to tell us (say) what are the chances of there being 30 or less lines in common in the two lists (the number is 0.0002, or around 2 chances in 10,000), or of 70 lines or more in the two lists (the number is 0.0066, or six chances in 1000).

Even more usefully: we can examine the hypergeometric distribution to assign «upper» and «lower» bounds. In statistical terms, anything with a less than 0.05 probability, that is less than one in twenty, may be regarded as unlikely. The «upper» bound will be the lowest value with a probability of less than 0.025 that there will be so many lines in common between the two lists; the «lower» bound will be the highest value with a probability of less than 0.025 that there will be so many lines in common between the two lists. For the Barbi list of 396 lines, the probability of there being 66 or more lines in common with the Shaw/Robinson list of 1885 lines is 0.0275 (for 67 lines or more, it is 0.020): this gives an «upper bound» of 66. The probability of there being 39 or less lines in common with the Shaw/Robinson list is 0.023 (for 40 lines or less, it is 0.039): this gives a «lower bound» of 39. Hence one can declare that the

chances are around 19 in 20 that by simple chance, there will be between 39 and 66 lines in common on the two lists. Correspondingly, if we find the number of lines in common is outside the range 39 to 66, then it is unlikely that this is simple chance; and the further outside this range is the number of lines in common, the more unlikely it is that this is simple chance.

Armed with these formulae, we may look at the correspondences between the Barbi *loci* and the Shaw/Robinson lists of lines we found significant for the stemmatic analysis of the *Commedia* (Table 4).

TABLE 4

Numbers of lines on both the Shaw/Robinson and Barbi lists (second last column), compared to the number expected by simple chance (third column), the upper and lower numbers consistent with chance (<0.05, or one in twenty), and the probability of that number of lines or greater in common (last column).

Group	Shaw	Barbi and Shaw/Rob. Mean	Lower bound	Upper bound	Barbi and Shaw/Rob. actual[16]	Probability (>=)
Urb/Rb	297	8	3	14	15 (6)	0.019 (0.838)
Mart/Triv	842	23	14	34	64 (64)	$<1\times10^{-12}$
Ash/Ham	722	20	11	30	49 (37)	$<1\times10^{-10}$ (0.00025)
α	314	9	3	16	19 (14)	0.0013 (0.0566)
TOTALS	1885	52	39	66	121	$<1\times10^{-13}$

[16] Many of the Barbi lines correspond with more than one group in the Shaw/Robinson analysis. That is: while a total of 121 lines from the Barbi lists appear among the 1885 lines identified in the Shaw analysis, 26 of the 121 lines appear in more than one of the Shaw/Robinson lists. Accordingly, one should adjust the count of corresponding lines for each group to take account of this overlap: the adjusted line counts are given in parentheses in this column, and the corresponding probability in parentheses in the last column. To explain: presuming, as I argue, that Barbi's choice of lines was conditioned by his awareness of Triv, then lines which evidence both Mart/Triv and another grouping should not be counted among lines found by Barbi for that other grouping. Therefore, one should remove from the Urb/Rb count of 15 the four lines also found in the Mart/Triv grouping, hence 11. Triv is also a prime witness for variants found in the hypothetical α group, and so one should also remove the five lines shared by Rb/Urb with α: hence, six variants (15-4-5), as given in parentheses. Similarly, one

Look first at the bottom «Totals» line in TABLE 4. This suggests that given the Shaw/Robinson list of 1885 lines and the Barbi list of 396 lines, both drawn from the 14233 lines of the *Commedia*, one would expect that by simple chance alone between 39 and 66 lines might appear on both lists. The last column tells us that Barbi in fact has 121 of the Shaw/Robinson list, considerably higher than one might expect from simple chance. Indeed, the chance of there being 121 (or more) lines in common is 0 to twelve decimal places; or, to put it another way, roughly equal to the chance of throwing a coin forty times and getting heads each time. It cannot be chance that there are so many lines appearing on both lists. It must be because both lists seek to represent the same phenomena, in this case: lines useful for genetic reconstruction of the tradition. For these 121 lines, then, Barbi's hypothesis that these lines would be useful for genetic reconstruction appears justified by our analysis. Accordingly, an analysis based on the 396 lines selected by Barbi will give a better picture of the manuscript relations than (say) just choosing any 396 lines. To that extent, one might indeed use the Barbi *loci* towards a genetic reconstruction of the history of the tradition, as Barbi planned and as Sanguineti attempted.

However, not all genetic reconstructions are equal. A closer look at Table 4 shows some problems. Consider the numbers of variants found in the Barbi list which also appear on the Shaw/Robinson lists for each of the four groups. The number of lines for the Mart/Triv group found on the Barbi list is much higher than one would expect by simple chance: 64, over double the number which could be reasonably explained by chance alone (taking the upper bound of 34 as the most chance might reasonably give). For Ash/Ham, the number is significantly higher than chance might give (though less so than for the Mart/Triv). But for α the number is only marginally higher than chance might suggest (19, compared to an upper limit of 16) and for the crucial pair Urb/Rb the number of 15 lines is only just higher than chance would explain.

It seems then that analysis based on the Barbi *loci* will be very effective at revealing the Mart/Triv and Ash/Ham pairs, but much less effective for α and for Urb/Rb. The prominence of Mart/Triv in the Barbi *loci* is likely to be no accident. The Trivulziano 1080 manuscript – exceptional for the extraordinary beauty of its script, with page after page free from

---

should discount the 49 Ash/Ham lines by the 10 lines found in Mart/Triv and the two found in α, hence 12; and discount also the 19 α lines by the three found in Mart/Triv and the two found in Ash/Ham, hence 14.

any blemish, and famous for its scribe Francesco di Ser Nardo, its age, the consistency of its language (see the description in Shaw) – was already well-known to Barbi. It is reasonable to assume that Barbi's knowledge of Triv was a factor in his choice of *loci*; hence the high number of Mart/Triv lines found in the Barbi list. Accordingly, we should discount the numbers of lines given in the second-last column of TABLE 4 as suggested in the footnote to the table, and use the numbers given in brackets as the base for our assessment. This (of course) has no effect on the efficacy of the Barbi list as far as Mart/Triv is concerned, and the Ash/Ham count (37 adjusted variants against an upper chance bound of 30) is also still beyond what could be explained by chance. But the number of Urb/Rb lines on the Barbi list now falls to six: lower even than the expected number which simple chance would give.

Sanguineti, as we have seen, based his entire stemmatic analysis on the Barbi *loci*. Accordingly, he would have collated just six lines which (in the Shaw analysis) alone contain readings likely to have been introduced by the common ancestor of Urb/Rb. Indeed, in his analysis he refers to only four of these lines.[17] Shaw is able to base her assertion that Rb and Urb share a common ancestor below the archetype on the 308 variant readings, found in 297 lines, identified by VBase working on the results of the phylogenetic analysis. Sanguineti, it appears, following Barbi, examined only a handful of lines likely to contain readings which evidence common descent of Urb and Rb from an ancestor below the archetype (Petrocchi's β). Hence, his argument that Urb and Rb do not share any such ancestor, that Urb descends directly from the archetype, sharing no intermediate ancestor with any other extant manuscript. This gives Urb, in his analysis, extraordinary and unique status:

[17] The four are: *Inf.* xi 84 men Dio offende e men biasimo accatta; *Inf.* xiii 63 tanto ch' i' ne perde' li sonni e' polsi; *Purg.* iv 72 che mal non seppe carreggiar Fetòn; *Purg.* xvi 145. Così tornò, e più non volle udirmi. The two he appears not to have used are *Inf.* xxvi 15 rimontò 'l duca mio e trasse mee and *Purg.* v 88 Io fui di Montefeltro, io son Bonconte. Another two appear in Mart/Triv as well as in Urb/Rb, and are used by Sanguineti: *Inf.* i 28 Poi ch' èi posato un poco il corpo lasso; *Purg.* xviii 57 e de' primi appetibili l'affetto; two appear in Mart/Triv but are not used by Sanguineti: *Purg.* xviii 58 che sono in voi sì come studio in ape; *Purg.* xxiv 61 e qual più a gradire oltre si mette; two appear in α and are used by Sanguineti: *Inf.* xvii 50 or col ceffo or col piè, quando son morsi; *Purg.* xiii 121 tanto ch' io volsi in sù l' ardita faccia; three appear in α and are not used by Sanguineti: *Inf.* i 116 vedrai li antichi spiriti dolenti; *Purg.* xxiii 2 ficcava ïo sì come far suole; *Par.* i 54 e fissi li occhi al sole oltre nostr' uso. It should be noted that Sanguineti also does not use another line on the Barbi list, *Par.* I 25, which both Petrocchi and Shaw regard as significant for analysis of Urb/Rb.

indeed, the authority of this one manuscript Urb is equivalent to that of all the other 800-plus manuscripts of the *Commedia* put together. Accordingly, Sanguineti consistently prefers the readings of Urb to those of other manuscripts, accepting the Urb reading in every case except where there is clear error.

Sanguineti's assertion of the unique authority of Urb affects his text significantly. In some 1406 places, he accepts the reading of Urb over that found in other manuscripts and accepted by Petrocchi. If you read the *Commedia* in Sanguineti's edition, about one line in ten will be different from that given in the widely-accepted Petrocchi text (or, indeed, from any edition based on Triv or other manuscripts from the Florentine tradition). All this follows from his assertion that Urb is not genetically related to Rb, which itself can be traced to his apparent examination of just a handful of the many variants which in Shaw's analysis (as in Petrocchi's) evidence this relationship.

How did this situation arise? It appears that Sanguineti put all his trust in the Barbi *loci*, as the basis for a complete account of the entire tradition. However, the Barbi *loci* include very few of the variants identified by the VBase analysis as evidence for the genetic relationship of Urb and Rb. Accordingly, he failed to see this relationship, and this led directly to his hypothesis of the unique authority of Urb. It can be argued that this failure is exacerbated because it appears Sanguineti did not use all the Barbi *loci*: he used only four of the six lines given by Barbi which in our analysis evidence the Urb/Rb relationship alone, and also used only four of the nine lines given by Barbi which evidence Urb/Rb and also either or both of Mart/Triv and α. The Barbi *loci* contain a total of fifteen lines which evidence Urb/Rb, either on its own (six lines) or in combination with other groups (nine lines); Sanguineti uses only eight of these. One is reminded that key parts of Sanguineti's argument rest on a very few readings: he uses just four to eliminate more than four hundred manuscripts from consideration (see Shaw's discussion, «Introduction – Sanguineti: The Tradition»). Yet, even if Sanguineti had used all the Urb/Rb variants included in the Barbi *loci*, it is still likely that he would have reached the same conclusion.

Indeed, Petrocchi warned explicitly against use of the Barbi *loci* as a base for the textual reconstruction of the whole tradition. In his «Proposte per un testo-base della *Divina Commedia*», *Filologia Romanza*, II (1955), pp. 337-365, he argues that while Barbi's hypothesis (that a genetic reconstruction of the tradition is possible on the basis of a small number of chosen lines) was plausible in 1890, research into the tradi-

tion in the years since Barbi's article showed that the complexities in the tradition are of such a nature that any such partial analysis can yield only «un primo e generico orientamento nel fitto dei rapporti tra i codici»: «a preliminary and general sense of direction within the thicket of relationships among the manuscripts». The analysis in this article suggests that Petrocchi is exactly right. Analysis based on the Barbi *loci* shows the broadest features of the tradition: thus the clearly marked pairs Ash/Ham and Mart/Triv. But it fails to show more elusive, yet real, relationships, as we have seen for Urb/Rb. Such maps are dangerous. As we noted earlier , Petrocchi accordingly took a quite different direction. Rather than sample a few readings across very many manuscripts, as Barbi suggests, he chose to limit the number of manuscripts surveyed and examined all the readings in those manuscripts. The Shaw edition follows this path, narrowing the number of manuscripts still further to just seven (including one, LauSC, not among the *antica vulgata* manuscripts used by Petrocchi), and using an array of analytic techniques not available to Petrocchi. The close agreement between Shaw and Petrocchi on almost all points may be seen as a vindication of Petrocchi's method.[18]

It could be argued that it is no surprise that Sanguineti failed to create an adequate account of the manuscript relations on the basis of the Barbi *loci*. One might question (as Petrocchi did) the fundamental methodology, of trying to create a manuscript stemma of so vast a tradition, and so long a text, on a sample of less than 3% of the material. One might also point to weaknesses in Sanguineti's use of the *loci*: his dependence on a very few readings to classify very many manuscripts; his apparent failure to use all the *loci*. But there is one aspect of this comparison of the Shaw results with the Barbi *loci* which might be surprising. I noted above that Barbi designated 396 of the 14233 lines of the *Commedia* as most likely to be useful for tracing the textual families to the original («rintracciare le famiglie dei testi a penna»). The Shaw/Robinson analysis identified 1885 lines as containing variants distinctive of family groupings. One might expect that a high proportion of Barbi's 396 lines would be among the 1885 identified in the Shaw analysis: in fact, there are 121 lines in common. While this is far more than simple chance might predict (as explained above), it seems lower than one might expect. Barbi has a near-legendary status among Italian philologists. He was in his early 20s at the

---

[18] But not on all points: Petrocchi argues that Ham is a descendant of Ash; Shaw shows that Ash and Ham are both independently descended from the same ancestor within the tradition.

time of publication of the *loci* article, at the beginning of an illustrious career as editor, philologist and Dante scholar, in which he came to know the business of editing and the manuscripts and text of the *Commedia* as few have ever done or will do. If one were to choose any one person to select a group of lines most likely to be useful for stemmatic analysis, Barbi would be the choice of many scholars. Yet, less than a third of the lines he selected appear in the list of 1185 lines identified by the Shaw analysis. That is: some two thirds of the lines he chose (275 of 396) turned out not to be productive for Shaw's analysis. One could make some allowance, that some of these 275 lines might be revelatory of relations among manuscripts not studied by Shaw, but the fundamental position of six of the seven Shaw manuscripts in Petrocchi's account of the early history of the tradition (the area of most concern to Barbi, as for any editor) make this a difficult argument to sustain.[19]

Why, then, does Barbi manage to identify so few lines which are actually productive for phylogenetic analysis? Barbi's account of the principles on which he chose these lines is highly compressed, just a few sentences. He remarks that it would be «inutile affatto» to set out the reasons for the choice verse by verse, and states the general principle

---

[19] Petrocchi divides the manuscripts of the *antica vulgata* into two major groups: α and β. He divides β into two groups, *d* and *e*, with Rb and Urb as two of the three representatives of *e*. He divides α into three groupings: *a b c*. Mart and Triv are the sole representatives of *a*; Ash and Ham represent two (in Shaw's analysis; Petrocchi sees Ham as descended from Ash) of some five branches of *b*; the original text of Mart («Mo» in Shaw) is close to Vat, a primary representative of Petrocchi's *c*. Thus, both primary and all three secondary branches of the tradition defined by Petrocchi are represented in the six manuscripts Ash Ham Mart Triv Urb Rb studied by Shaw. The only branch of Petrocchi's stemma (if it is a branch) not represented is that of the anomalous Landino manuscript, which Petrocchi places as a blend of readings drawn from *c* (via «*c1*») and β (via «*d*»). It should be noted that Shaw's study also shows that the Laurenziano Santa Croce manuscript, much favoured by early editors of the *Commedia* (Witte thought it the best surviving manuscript and usually chose its readings ahead of those of other manuscripts for his 1862 edition; Moore and Casella also rated it highly; see Shaw «Introduction», notes 31 and 47) and chosen by Sanguineti as one of his seven, is actually a classically contaminated manuscript, combining readings in its original text drawn from four of the five sub-archetypes identified by Petrocchi (that is: *a b c e*), and the contamination is compounded by the corrections introduced into the manuscript, with the «c2» corrector apparently introducing readings from another witness which appears to combine readings from Petrocchi's *c* and *e* branches (Robinson and Shaw «Phylogenetic Analysis», in Shaw *Commedia*.) This analysis of LauSC accordingly confirms Petrocchi's argument: that manuscripts written after 1355 are likely to contain such mixes of readings drawn from multiple branches as to be of no value towards reconstruction of the earliest phases of the tradition.

guiding the choice: there should be two or more variants present, each in several manuscripts. Beyond this, he sets out the readings which he excluded: readings which show transcribers and commentators exhibiting critical judgement; and those where separate scribes are highly likely to introduce the same reading for reasons of «eufonia, sinonimia, correlazione grammaticale», and notes that he paid special attention to the help which major errors can afford towards the classification of manuscripts («Abbiamo anche pensato all'aiuto che porgono per la classificazione dei manoscritti, gli errori grossolani di lezione»).

We may compare these criteria with some of the readings identified by the Shaw/Robinson analysis as indicative of family relations. Here are five of the 308 readings identified as likely to have been introduced by the common ancestor of Urb/Rb; none of these five lines appear among the Barbi *loci*.

Inf. i 89: aiutami da lei, **famoso saggio,**

famoso e saggio	LauSC Rb Urb FS
famoso saggio	Ash Ham Mart Triv PET

Inf. ii 71: vegno **del** loco ove tornar disio;

di	LauSC Rb Urb FS
del	Ash Mart Triv PET
dal	Ham

Inf. ii 110: a far lor **pro o a** fuggir lor danno,

pro e a	Mart-orig Rb Urb FS
pro ne a	Ash LauSC-c2 Triv
p*ro*de *et* a	Ham
pro ... a	LauSC-orig
pro o a	Mart-c2 PET

Inf. iii 3: per me si va **tra la** perduta gente.

ne la	Rb Urb FS
tra la	Ash Ham LauSC Mart Triv-c1 PET
tra	Triv-orig

Inf. iii 22: Quivi sospiri, pianti e **alti** guai

altri	Ash-orig Rb Urb
alti	Ash-c2 Ham LauSC Mart Triv FS PET

We can see why Barbi would not have chosen these readings. All have the slight character of variants which might have appeared readily at any point of the tradition, and hence are likely (to use the term employed by Petrocchi and Shaw) to be «polygenetic»; in Shaw's definition «an error liable to arise independently in independent manuscripts». One can imagine that no textual scholar, educated to reject errors likely to be polygenetic as indicative of anything, would select these as the basis for any kind of argument: indeed, only one, tra la/ne la *Inf.* iii 3, is used by Shaw in her account of Rb/Urb. Yet the Shaw/Robinson analysis of the distribution of the variants across 94779 distinct sites of readings (usually, one word, but often a phrase) over the 14233 lines of the *Commedia* suggests that the appearance of these five readings in Urb/Rb results from their introduction into an ancestor copy shared by Urb and Rb. That two of these five (*Inf.* i 89 and ii 71) appear also in LauSC is likely to be the result of contamination in that manuscript (see footnote 20). This leaves only *Inf.* ii 110 «pro e a» in the Aldine text which is the base of the Martini collation (itself derived from Vat) and Inf. iii 22 «altri» in the original text of Ash as likely polygenetic errors, if not deriving from contamination.

To sum up: a high proportion of the 1885 lines identified by the Shaw/Robinson analysis as indicative of manuscript relations contain variants which would not satisfy Barbi's criteria as being unlikely *a priori* to have arisen in independent copies. Yet, these same variants are useful for our analysis, precisely because as a group they did not arise in independent copies. The key phrase here is «as a group»: individual variants among them (say, among the 308 Urb/Rb variants) did appear in other manuscripts, sometimes perhaps by contamination, sometimes by polygenesis. But other manuscripts outside each group appear alongside the members of the group in a scattered, random fashion, as we see in these five variants: Urb/Rb agree in all five, once with no other manuscript (iii 3), twice with LauSC (I 89, ii 71), once with the original text of Mart (ii 110), once with the original text of Ash (iii 22). This is exactly as we should expect in a real manuscript tradition: that the scribe of the common ancestor of Urb/Rb introduced some 308 readings into the copy, which then descended into Urb/Rb. Of those readings, we should expect a few of them to appear in other manuscripts, by contamination or by polygenesis: thus, we find 40 of the 308 in Ash (also, 2 Ash-orig), 11 in LauSC (11 LauSC-orig, 4 LauSC-c1, 5 LauSC-c2, 4 in Mart (Mart-orig 13), 9 in Triv, 46 in Ham, while all 308 appear in both Urb and Rb.

This suggests that one cannot determine whether a particular variant is monogenetic (as, ideally, we would want all variants to be) or polygenetic purely on the basis of the variant itself, as Barbi suggests we do. One must also take into account both the actual distribution of manuscripts which have the variant, and the consistency of this manuscript distribution across many other variants. When we see the pair Urb/Rb appear over and over, sometimes on their own, sometimes with a varying cast of other manuscripts, then we are entitled to presume that this consistency of distribution itself may suggest a genetic relationship between the manuscripts. Although both Petrocchi and Shaw pay close attention to variants which they regard as likely to be monogenetic, on the basis of the variant itself (as Barbi does), they also are aware of the significance of a consistent pattern of agreements, found in variant after variant. Shaw cites Petrocchi's argument that the trio Urb/Rb/Mad have a closeness «nettamente superiore per foltezza di statistica e natura dei casi a qualsiasi affinità tra il testimone urbinate e gli altri» (Introduction, citing Petrocchi *Introduzione* 334-389), and points out the significance of the phrase «foltezza di statistica»: «statistical density». As I noted earlier, Shaw expands on Petrocchi's perception, as she discusses a number of readings supporting the genetic relationship of Urb/Rb which, she concedes, are each individually possibly polygenetic, but which taken together evidence a genetic relationship. Thus her conclusion:

These are of course small errors, possibly polygenetic. Each one in itself would count for very little. But it is the cumulative picture built up of a whole series of such small errors right across the text and shared with no or very few other manuscripts which is significant and suggests a common exemplar.

Seen this way, Petrocchi and Shaw's use of «foltezza di statistica» alongside consideration of the monogenetic or polygenetic character of each variant is an extension of Barbi's method, not a repudiation of it. Barbi, Petrocchi and Shaw all agree that it is unsafe to rely on certain individual variants, considered on their own, as indicative of relations across the whole tradition. Indeed, Shaw criticizes Sanguineti (as I note above) for basing his analysis on a few variants. However, both Petrocchi and Shaw are able to use «foltezza di statistica» because they do take into account every variant, every reading, in the manuscripts they survey – an approach only possible because they choose not to base their analysis on a sample of readings from a great many manuscripts (the Barbi *loci* methodology, followed by Sanguineti) but on all the readings from a small number of manuscripts.

Barbi; Petrocchi; Shaw: more than a century of Dante textual schol-
arship, a century which saw Italian textual scholarship absorb, and then
reshape itself in reaction to, the impact of Lachmannian methods and
the opposition to these methods led by Bédier and others; a scholarly tra-
dition which is now reckoning with the first impact of another revo-
lution, as digital methods open up new ways of approaching large tex-
tual traditions. We should mention two more names: Giorgio Pasquali,
whose *Storia della tradizione e critica del testo* (Florence, Le Monnier, 1934)
repudiated the rigid methodology outlined in Maas's *Textkritik* (Leipzig,
Teubner, 1927), without rejecting, as Bédier did, the possibility of making
sense of large and complex traditions; and Gianfranco Contini. In 1935,
when Contini reviewed Pasquali's book, he was around the same age as
Barbi when he was compiling the list of *loci*. Barbi's own *La nuova filo-
logia e l'edizione dei nostri scrittori da Dante al Manzoni* (Florence, San-
soni, 1938) follows Pasquali in advocating what Pugliatti describes as a
«virtuous middle way» between Lachmann and Bédier.[20] Contini follows
this «middle way», asserting throughout his career that intelligent use of
every available method, including stemmatics where appropriate, joined
with deep knowledge of the manuscripts and the texts they carry, would
permit a useful understanding of whole manuscript traditions. Shaw
studied with Contini, first as a post-graduate from 1962 to 1964, and then
from 1966 as she wrote her Dott. Lett. dissertation, an edition of the ear-
liest Italian translation of the *Monarchia*, with Contini as her supervisor,
thus bracketing the appearance of Petrocchi's great edition in 1966. With
Contini's encouragement, she then edited Marsilio Ficino's translation of
the *Monarchia*. Again with Contini's encouragement, in 1981 she wrote
an article critical of some aspects of Ricci's edition of the *Monarchia*.[21]
After the article was published, Contini suggested to Shaw that she under-
take an edition of the *Monarchia*. Shaw hesitated at first: she was aware
that a factor in the problems of Ricci's edition of the *Monarchia* was the
amount of data generated by a text in many versions, and Ricci's inability
to manage all this data. It was only when she undertook an introductory
course in computing at the Cambridge University Computing Centre in
October 1985 and saw the possibility of using a computer to record and

---

[20] Paola Pugliatti, «Textual Perspectives in Italy: from Pasquali's Historicism to the
Challenge of "Variantistica" (and Beyond)», *Text*, 11 (1998), pp. 155-188 [p. 162].

[21] Shaw's doctoral dissertation was published as «Il volgarizzamento inedito della
Monarchia», *Studi danteschi*, XLVII (1970), pp. 59-224; the edition of the Ficino transla-
tion as «La versione ficiniana della *Monarchia*», *Studi danteschi*, LI (1978), pp. 289-407; the
article on Ricci in «Sul testo della *Monarchia*», *Studi danteschi*, LIII (1981), pp. 187-217.

explore all the data generated by full transcripts of each manuscript that she committed herself to making the edition – and, to using computer methods to do so.[22]

Shaw discussed this new methodology with Contini, and he encouraged her in this path. She described her work in a paper at the conference in Florence in 1988, convened to celebrate the centenary of the founding of the Società Dantesca Italiana in 1888.[23] Her decision, to base the edition on full-text transcripts of the manuscripts looks back to Petrocchi, just as her decision that the transcripts be made in machine-readable form looks forward to the development of computer-based systems for comparison and analysis. In 1991, Shaw had completed the first transcripts of the *Monarchia* manuscripts and was exploring how these transcripts might be used to make an edition.[24] I was starting work on what became the *Collate* suite of software tools for creating editions from multiple transcripts of works, and also discovering how tools from evolutionary biology could be used to advance understanding of large textual traditions. Shaw came to see me (it is a fortunate chance that her dentist was based in Oxford, where I was then working), and this began a two-decade collaboration which issued in her editions of the *Monarchia* and *Commedia*.

Barbi, Pasquali, Petrocchi, Contini, Shaw. These scholars are joined in a common belief: sense can be made of large manuscript traditions using good method and expert philological knowledge, and useful edi-

---

[22] As an index of the difficulty of this work, in those days before computers became portable and personal: Shaw transcribed the manuscripts onto paper from a microfilm reader at the University library; then carried the transcripts across the Cam to the computing centre, when she entered them into the Cambridge mainframe using the «Zed» line editor.

[23] Published, including some printouts of her transcripts, in «Per un nuovo testo critico della *Monarchia*», in *La Società Dantesca Italiana 1888-1988. Atti del Convegno Internazionale di Studi Danteschi tenuto a Firenze nel novembre 1988*, Milano-Napoli, Ricciardi, 1995, pp. 435-444.

[24] Other scholars at this time were coming to the view that full-text computer-readable transcripts were the way forward for textual scholarship applied to large manuscript traditions. In particular, Francesco Mazzoni, Contini's successor as President of the Società Dantesca Italiana, encouraged his students to make full-text transcripts of *Commedia* manuscripts, and developed a set of transcription guidelines for this work. These «principi di trascrizione» may be found at http://www.danteonline.it/italiano/risorse.htm (with an English translation by Shaw). Transcriptions of several manuscripts are available on the danteonline site. Mazzoni himself, and other associates of the SDI (particularly Paula Laurella), were extremely supportive of Shaw's continuing work on the *Monarchia* (her print edition was published in the SDI's Edizione Nazionale series; the SDI co-published the DVD version) and on the *Commedia*.

tions of works can be founded on that sense. Full-text machine read-
able transcripts of whole manuscripts, passed through an array of com-
puter-based methods, are now part of a philologist's armory, to be used
in conjunction with (and not in opposition to) knowledge of the man-
uscripts, of the text itself and its contexts, and with editorial intuition.
This seems a useful starting-point for further explorations into the *Com-
media* tradition.

University of Saskatchewan