Studies in Stemmatology II

Studies in Stemmatology II

Edited by

Pieter van Reenen

August den Hollander

Margot van Mulken

With the assistance of Annelies Roeleveld

John Benjamins Publishing Company
Amsterdam/Philadelphia

Publication of this volume was financially supported by the Netherlands organization for scientific research (NWO)

Library of Congress Cataloging-in-Publication Data

Studies in Stemmatology II / edited by Pieter van Reenen, August den Hollander and Margot van Mulken.
p. cm.
Includes bibliographical references and indexes.
1. Manuscripts. 2. Manuscripts, Medieval. 3. Transmission of texts. I. Reenen, Pieter Th. van. II. Hollander, A. A. den III. Mulken, Margot van

Z105.S782 2004
091-dc22 20040100159
ISBN 90 272 3222 9 (Eur.) / 1 58811 535 6 (US) (Hb; alk. paper)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Prologue

The publication of this volume of *Studies in Stemmatology* is the second in a series. Its predecessor was published in 1996 and opened the most actual state of the art in stemmatology to a broad audience.[1] That volume not only aimed at giving scholars access to modern stemmatological methods and techniques, but also at illustrating how profitable the application of these methods might be for their future work. The first volume was very well received by stemmatologists all over Europe and also gave an impulse to new research, as several articles in *Studies in Stemmatology II* clearly illustrate.

The contributions to this present book partly proceed from those of the first volume. Most of them are the result of the on-going scholarly debate on stemmatology of recent years. Several of the contributions to this volume were presented on 13–14 April 2000, during the NOSTER-conference at the Netherlands Institute of Advanced Studies (NIAS) at Wassenaar, and on 13 October 2000, during the Stemmatology Conference at the Vrije Universiteit in Amsterdam. Some others are the result of the annual colloquia of stemmatologists, held at the Vrije Universiteit in Amsterdam.[2]

The object of this second volume of *Studies in Stemmatology* is the evaluation of the most recent methods and techniques in the field of stemmatology, as well as the development of new ones. The book is largely interdisciplinary in character: it contains contributions from scholars from classical, historical, biblical (Smelik, Houtman, Den Hollander), medieval and modern language studies, as well as from mathematical and computer scientists (Wattel) and biologists (Howe e.a.; Spencer e.a.). Various manuscript traditions are dealt with here: some of them within one field of language (Van Mulken, Schøsler), some multilingual (Roeleveld/Langbroek/Wattel), the last group of course requiring a special methodological approach to the establishment of variants. Other traditions were very extensive, e.g. the New Testament manuscripts (Mink, Wachtel) and the Old Church Slavonic manuscripts (Mironova). The contributions in the book have been divided into two sections. The first section deals with various stemmatological methods and techniques. The second section focusses more specifically on the various problems concerning textual variation.

Stemmatological methods

Not a bifurcating tree?

Christopher Howe, Adrian Barbrook, Linne Mooney and Peter Robinson present a relatively new stemmatological approach. They explore the similarities between the evolution of DNA sequences and the changes occurring in manuscript traditions. They show how the techniques of evolutionary biology can be applied to stemmatic analysis and how a number of features of manuscript traditions have clear parallels in genetics. They conclude that the process of incorporation into DNA mirrors the incorporation of changes into the manuscripts. It follows that programs for phylogenetic analysis of sequence data can be exploited for stemmatic analysis of manuscript tradition. For this approach the Splits Tree program is used, which has the advantage that it does not presuppose, as many methods do, that the tree is a bifurcating one.

Local stemmata?

Methodologically new, and very promising, is the contribution of Gert Mink. In his contribution he broaches the problem of a text tradition of many hundreds of manuscripts in which hardly any type-2 variants are found and contamination is the rule. This is the case in one of the catholic letters of the New Testament, the Epistle of James. In such a tradition existing methods cannot be applied. Instead, Mink has two working hypotheses on which his approach is based:

a. If more than one exemplar was consulted by a scribe, the exemplars are closely related.
b. Variants are analysed one by one in local trees.

Within the local trees the direction of the changes can be determined: they can be oriented by establishing which variant derives from which other variant. When groups of local trees are oriented in the same direction, parts of global trees can be constructed.

Reduction of witnesses

Klaus Wachtel also deals with this extremely large textual tradition. In his contribution he shows how the number of manuscripts to examine can be reduced before the structuring of a stemma. The number of extant manuscripts of the

New Testament is so large that any reasonable form of reduction of quantities must be accepted before the building of a stemma starts. By distinguishing two groups of manuscripts in the New Testament tradition Wachtel succeeds in doing so. The two groups distinguished are the Majority group and the Byzantine group. If two or more manuscripts are almost alike, there is no need for further analysis. By applying this approach Wachtel succeeds in reducing the numbers considerably, without the risk of excluding manuscripts which contain crucial textual information. The resulting group forms the input of Gert Minks analysis.

Dealing with successive contamination

A illuminating example of how profitable the application of modern stemmatological tools can be is given in the contribution of August den Hollander. One of the complex problems a philologist has to deal with is a contaminated text tradition. In the first volume of *Studies in Stemmatology* Wattel and Van Mulken offered the instrument of the so-called shock waves (cardiograms) as a help to reveal successive contamination in a text tradition which is rather entangled. In his contribution Den Hollander shows how the application of this instrument indisputably revealed successive contamination in the textual tradition of early sixteenth-century printed Dutch Bibles.

In her contribution to the present volume, Margot van Mulken shows that the output of the quire separator developed by Wattel (see first volume) may have serious consequences for the further treatment of the stemmatological process. When the separator indicates successive contamination, as in the case of the *Cligés*, it may be necessary to presuppose a multiple orientation of the stemmata. However, in the case of the *Cligés*, all the archetypes can be found in the neighbourhood of one manuscript, which fortunately reduces the complexity of this operation.

Textual variation

Accidental variation

Ulrich Schmid explores the phenomenon of accidental variation (parallelism). His contribution is a reaction to the recent study of B. J. P. Salemans, who systematically reviewed various types of variant readings used in genealogical studies, and offered strict text-genealogical rules in order to exclude possible

variants caused by accidental variation.[3] In his contribution, Schmid illustrates the implications of applying Salemans' rules to a text tradition: on the one hand they would exclude too much, leaving out many genealogically 'valid' variants, on the other hand even the variants that would be included on the basis of Salemans' rules, still contain parallelistic readings. Therefore, Schmid concludes, no safe line can be drawn without proper statistical evaluation.

No reduction of variants

Evert Wattel also writes as a reaction to the dissertation of Salemans. Textual scholars do not generally agree on which type of variant readings are suitable for the construction of a stemma, and which are not. In his contribution Evert Wattel argues for the acceptance of as many version formulas as possible, in addition to expressing the reliability of the variants by adding more or less weight to the so-called version formulas. His main focus is on the computational problems of constructing a stemma on the basis of the collective formulas. Along with a more general methodological discussion, he dwells on specific problems such as lacunary version formulas and the computational complexity.

Categories of variants

In his contribution Willem Smelik deals with variant selection and tree construction in the text tradition of Targum Judges. The core of this study consists of observations on the phenomenon of random variation in the manuscript reproduction. To identify random or coincidental variation, he suggests a transparent, verifiable categorisation of variant readings. Secondly, he discusses the possible genealogical information of these various types of variants in great detail. Further, Smelik draws stemmata for each type or group of types of variants. Finally, comparison of these stemmata reveals which types of variants turned out to be genealogical relevant in his textual tradition, and which not.

Lene Schøsler compares the categorised variants of two closely related manuscripts of the *Perceval* and four (five?) manuscripts of the *Charroi de Nîmes*. The first two were copied by the same scribe, with perhaps a difference in time. Assuming that the scribe copied twice from the same exemplar, it is remarkable that the variations found between the two manuscripts and those in the tradition of the *Charroi de Nîmes* are hardly different. In other words, whether the same scribe copies the same manuscript twice or different scribes copy a manuscript may not necessarily result in more variants.

Weighting variants?

In their contribution Matthew Spencer, Linne Mooney, Adrian Barbrook, Barbara Bordalejo, Christopher Howe and Peter Robinson attempt to increase the chance of reconstructing correct stemmas by categorizing variants into ten different kinds, such as: "line changed completely", "word change affecting rhyme", "word variant, changes meaning", "minor word added or omitted, without changing meaning". On the assumption that not all kinds of variants are equally reliable, the more a category of variants is reliable the more weight it is assigned. On comparison between stemmata of the 55 manuscripts and three printed versions of Lydgate's *Kings of England*, the choice of weights appeared relatively unimportant. However, the authors expect that this may be different in larger textual traditions. The method used to reconstruct the stemmata was neighbour-joining: a simple clustering algorithm which sequentially separates pairs of manuscripts from an initially unresolved stemma.

Dina Mironova deals with the problem of a textual tradition of many manuscripts. She compares two different formal genealogical methods in her study of the Gospels in Slavonic: cluster analysis (Alexeev) and the Three Level Method (Wattel). Her research comprises no fewer than 531 manuscripts, still presenting, however, a rather stable text. The large number of witnesses impelled her to work with groups of manuscripts as a way of reduction. Alexeev's method turned out be less accurate, but more economical, since it is still easier to apply to large traditions. Wattel's method is, however, more accurate, and forces the scholar to formulate precise classifications or explicit philological labellings. Despite the difficulties with extremely large textual traditions, according to Mironova his method is to be preferred when variants should be evaluated (weighted).

Exclusion of variants

Dineke Houtman studies the textual history of Tosefta Targum Jonathan, an extended Aramaic Bible commentary. She focuses on the question of how to deal with this type of text in stemmatological research, especially when comparing it with the Hebrew Bible text and its paraphrasing Aramaic Targum text. All three types of text represent different stages in the textual history. The text of the Targum remains close to the Hebrew Bible, the text of Tosefta Targum, however, gives a more free rendering. Houtman concludes that including textual variants from the Tosefta Targum may introduce a lot of bias in the re-

sults of the stemmatological research and should therefore be done with great precaution.

Alternative classification of variants

Annelies Roeleveld and Erika Langbroek, in cooperation with Evert Wattel, deal with the text tradition of 'Valentin and Namelos'. Its extant manuscripts are written in no fewer than four languages/dialects: Middle Dutch, Middle Low German, Middle Mid German and Old Swedish. Other differences between the texts are also considerable: some are in verse, others in prose; some are much more lengthy and elaborate in their descriptions than others; some have been preserved in fragments only. The problem in this tradition is what sort of variant to group in comparable units. The authors develop a classification in which the number of incidents, the order of incidents and the detail in the description of incidents play a part. Although this notion of variant goes beyond the traditional view, their approach shows that satisfactory results can be obtained. The resulting stemma plausibly shows that the Middle Dutch versions are the more original. Rhyme analysis had already pointed in the same direction; the original language was most probably Middle Dutch.

The editors hope that this second volume of *Studies in Stemmatology* will inspire scholars like the first volume did, and stimulate the development of new methods and strategies aiming at further control of variation and contamination in (large) text traditions.

Pieter van Reenen
August den Hollander
Margot van Mulken

Notes

1. *Studies in Stemmatology* (1996). Pieter van Reenen & Margot van Mulken (Eds.). Amsterdam/Philadelphia: John Benjamins Publishing Co.

2. At this place we wish to thank NIAS, NOSTER and the Faculty of Arts of the Free University for their willingness to support the various stemmatological meetings in the past years.

3. B. J. P Salemans (2000). *Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian, way. The Case of Fourteen Text Versions of Lanseloet van Denemerken* (diss. Nijmegen). Nijmegen.

Part I

Stemmatological methods and techniques

Parallels between stemmatology and phylogenetics

Christopher Howe*, Adrian Barbrook*, Linne Mooney, and Peter Robinson
University of Cambridge* / University of Maine, Orono / De Montfort University, Leicester

1. Introduction

The work and ideas we discuss here are part of a project entitled "STEMMA – Studies on Textual Evolution of Manuscripts by Mathematical Analysis", funded by the Leverhulme Trust. This project aims to apply the techniques of evolutionary biology to the analysis of manuscript traditions. In particular we are interested in the application of computer programs developed for evolutionary biology to the study of manuscripts. In this paper we explore the similarities between the evolution of DNA sequences and the changes occurring in manuscript traditions. We will show how the techniques of evolutionary biology can be applied to stemmatic analysis and how a number of features of manuscript traditions have clear parallels in genetics. Another paper in this volume (Spencer et al.), to which this chapter should serve as an introduction, discusses more specific issues in this work.

The computer programs we are using were developed for research in biology, so their application to manuscript stemmatics requires some knowledge of the underlying biology. A more detailed discussion can be found in biological textbooks (e.g. Voet et al. 1999). DNA (deoxyribonucleic acid) is comprised of four kinds of unit. These are collectively called nucleotides (more fully deoxyribonucleotides), and the four kinds are adenosine, guanosine, cytidine and thymidine deoxyribonucleotides. They are more conveniently designated A, G, C and T. DNA molecules are composed of chains of such nucleotides, and the order in which individual nucleotides comes carries the information used in

biological systems. It specifies the order of amino acids in proteins, and thus the structure and function of proteins, which are the main functional entities in cells. For example, enzymes, which catalyse the reactions of metabolism, are proteins.

In most systems, the DNA chains come in pairs – forming the double helix made famous by Watson and Crick (Watson & Crick 1953). The sequence of nucleotides in one strand determines the sequence in the other. Thus A in one is always opposite T in the other, G in one opposite C in the other. As cells divide, their genetic information has to be duplicated. This process is termed DNA replication, and the aim is to make an identical copy of the parental molecule. However, errors can occur in the replication process, so a parental molecule with sequence, say,

ACGGTACTAG
TGCCATGATC

might give rise to two daughters, one of which had the same sequence, and the other of which had a different sequence, say

ACGGCACTAG
TGCCGTGATC

Here, the T in the fifth position in the upper strand has been replaced by a C, and in the lower strand an A has been replaced by a G. We say that a 'mutation' has occurred, and the information in the DNA has been altered.

2. Recovering phylogenetic trees

As biological species evolve and give rise to new species, they accumulate mutations in their DNA. The longer it is since two species had a common ancestor, the more different (in general) is their DNA sequence. So we can use the differences in DNA sequence among species as a way of inferring their evolutionary relationships. We can recover something akin to a family tree, showing which species share a common ancestor to the exclusion of others. This is called a phylogenetic tree. Recovering a phylogenetic tree using a given DNA sequence for a group of species requires both a *model* for sequence evolution and a *method* for tree recovery.

Models of sequence evolution are in effect a set of assumptions about how the sequences change. They may be explicit (in which case the computer program being used will require them to be specified or use default settings) or

Suppose we have species W, X, Y and Z, with the following sequences (in reality, much longer sequences would be used):

W	AAAAAAAA
X	GGAAAAAA
Y	CCTTTTAA
Z	CCTTCCAA

The distance matrix would be

	W	X	Y	Z
W	–	2	6	6
X	–	–	6	6
Y	–	–	–	2

and the tree inferred would be:

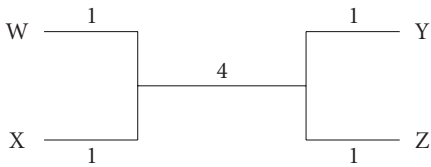

Figure 1. Hypothetical example of a distance matrix analysis.

they may be implicit, in which case one may be unaware that these assumptions are being made. It is particularly important to realize that, even if invalid assumptions are being made, it will still usually be possible to recover a phylogenetic tree. However, the tree may be incorrect as a result of making the wrong assumptions. Examples of the assumptions that may be made as part of a model include the relative frequencies of certain kinds of mutation (often termed the transition/transversion ratio); independence of mutations (i.e. that mutation at one position does not affect the chance of a mutation at another position); or identical distribution (that the same regions of a sequence are potentially able to mutate in all the organisms being considered).

Evolutionary biologists use several methods for tree recovery, including distance matrix, parsimony, maximum likelihood (reviewed by Felsenstein 1988; Beanland & Howe 1992) and split decomposition methods (Huson 1998).

With distance matrix methods, one calculates a matrix showing the number of differences between pairs of sequences, and determines the tree which has the best fit to this matrix. An example is given in Figure 1.

With parsimony methods we prefer the tree which requires the fewest mutations. This is in effect a cladistic analysis. So, for example, with the following DNA sequences (showing only one of the two strands):

W AAGCCAAT
X TAGCCAAT
Y CGCTTGGT
Z GGCTTGGT

Positions 2–7 would require fewest mutations if species W and X shared most recent common ancestry to the exclusion of Y and Z. Then, the most parsimonious tree would be as shown in Figure 2A.

Maximum likelihood attempts to find the tree that has the highest probability of generating the data (i.e. sequences) observed. This approach can be very computationally intensive. With split decomposition, as implemented in the program SplitsTree we consider possible splits between the sequences, e.g. (WX, YZ) or (WY, XZ) etc. and look at the number of nucleotide positions consistent with each split. Split decomposition has a number of potentially desirable features. It does not presuppose, as many methods do, that the tree is a bifurcating one. It allows recovery of trees where one individual can have many descendents. It also allows conflicting information to be shown – where there are some positions in the data which are not consistent with the preferred tree. These appear as boxes in the output, giving a network rather than a tree. So the example given in Figure 2B would arise where there was some signal linking W with Y and X with Z, as well as the signal linking W with X and Y with Z.

Figure 2A. Other tree-recovery methods. 2A shows results of parsimony analysis using the hypothetical data shown in the text.

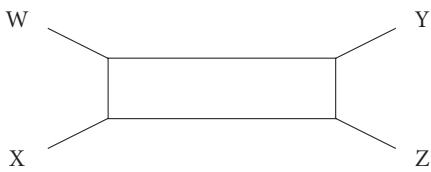

Figure 2B. Other tree-recovery methods. 2B shows the result of a hypothetical split decomposition analysis where there is support both for WX, YZ groupings and for WY, XZ groupings.

3. Stemmatic analysis

The techniques which are used to recover evolutionary trees can be applied in principle to datasets derived from manuscript traditions, using changes between texts in the same way as evolutionary biologists use changes between the DNA sequence of different organisms (Platnick & Cameron 1977; Cameron 1987; Lee 1989; O'Hara & Robinson 1989; Barbrook et al. 1998). We are applying this to a range of texts, such as John Lydgate's 15th century poem, *Kings of England*, which exists in over 30 manuscript versions comprising a set of stanzas describing the Kings of England from William the Conqueror onwards. Thus we have variants such as:

> Worthy to stand among the worthy nyne
> Able to stand among the worthi nyne
> Able to stande among the worthyes nyne

for the same line in different texts. (These are variants of line 96 of the poem in Oxford, Bodleian, Rawlinson C.316, fol. 122v; Cambridge, Jesus College Q.G.8 (56), fol. 47v; and Ipswich, County Hall Deposit C4/4, Percyvale's Great Doomsday Book, Bk. VI, fol. 239v.) We can encode this information to produce a dataset resembling a nucleotide sequence, (but using any letters or numbers rather than just A, C, G or T) to represent different readings at a given position, and use this directly as an input to a phylogenetic program such as SplitsTree. Thus we recently produced a stemmatic analysis of the extant manuscript versions of Lydgate's *Kings of England* based on evolutionary biological techniques, which replicated stemmatic analysis done by traditional methods (Figure 3).

Many of the manuscripts grouped together in the traditional stemmatic analysis were also grouped together in the phylogenetic analysis. Some were grouped together in the former, but not in the latter. This might simply reflect a need for more information in the dataset used. All groupings in the phylogenetic analysis were consistent with those in the traditional analysis. This study is described in more detail elsewhere (Mooney et al. 2001).

It is remarkable how many parallels there are between the evolution of genetic material and the changes occurring in manuscripts (Howe et al. 2001). These include recombination, convergent evolution and transposition. We will look at these parallels in turn.

Genetic recombination is the process whereby two different copies of a gene come together to produce a hybrid called a recombinant. The first part of the recombinant comes from one precursor gene, and the second from the other. The parallel in manuscript traditions is the change of exemplar de-

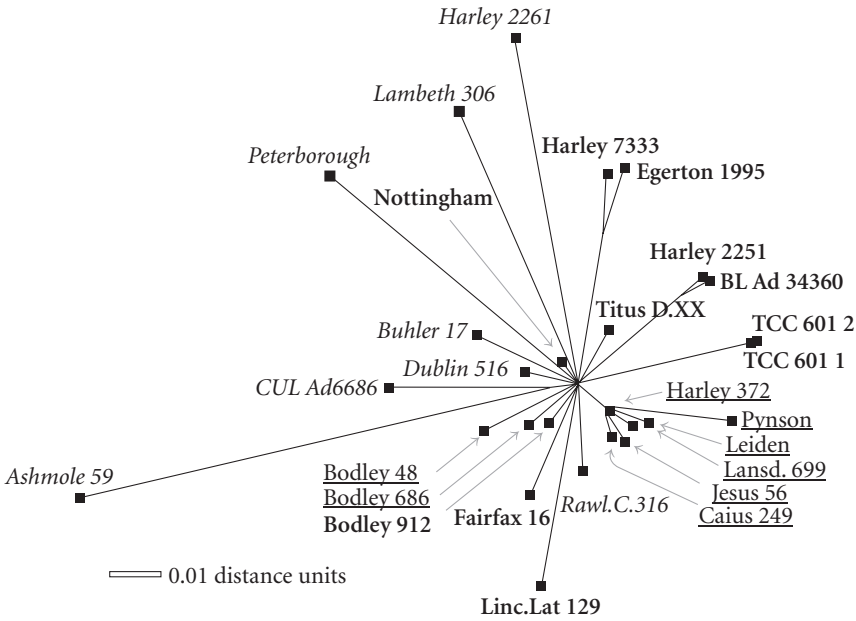

Figure 3. SplitsTree diagram of relationships among 27 most complete manuscripts and early printed copies of Lydgate's *Kings of England*. Bold indicates group A, italics B, underlined C, suggested by manual analysis (Mooney et al. 2001).

scribed as "successive contamination" (Wattel & van Mulken 1996), when a scribe changed the text from which he was copying part of the way through. We found examples of this in the analysis of the *Prologue* to the *Wife of Bath's Tale*, in that the position of some texts on the stemmatic tree changed when an analysis performed on the first part of the data was compared with one done on the second (Barbrook et al. 1998). Programmes are being developed to identify recombination breakpoints in genes, and these may be useful in stemmatic analysis (e.g. Holmes et al. 1999). As well as generating hybrids where there is a clear breakpoint, recombination can give rise to more complex products, where the resulting DNA sequence is a mosaic of the two parental versions of the sequence (e.g. Medgyesy et al. 1985). This process is analogous to simultaneous (where the scribe used several exemplars simultaneously) and incidental contamination (where the scribe used a single exemplar first and then modified the text afterwards with other exemplars) (Wattel & van Mulken 1996).

Transposition of genetic information is when material from one source is inserted somewhere quite different. This is a feature of the life cycle of some viruses, for example. If the source of inserted DNA is a different species it would

be regarded as 'lateral gene transfer'. We have found parallels in the *Kings of England* tradition where there is contamination of one tradition with material from another. Thus, the verse relating to William I begins typically:

> This myghti William Duk of Normandye,
> As bokes old makith mencioun,
> By just title and by his cheualrye,
> Made kyng by conquest of brutes Albyoun... (British Library, Harley 2251)

There exists another fifteenth-century poem on the Kings of England, not by Lydgate, which typically has this for William I:

> At Westmyster William icrowned was
> The furst day of Cristemas;
> A gret thyng after he dude thanne;
> Made the kyng of Skottys his legeman... (Bodleian, Ashmole Rolls 21)

However, within this second tradition is a text with a clear example of transposition from the first:

> This myghtty William duke of Northmandy,
> That by juste tytill and also by chyualery
> Conquered this land and king bycome
> And the kyng of Scotts he made his legeman... (Bodleian, Bodley 131)

Convergent evolution is when the same change occurs independently in different lineages. Thus, for example an AT base pair might replace a GC base pair at the same position independently in two different species. If this occurs frequently enough, evolutionary tree-building may be misled and species with a large number of convergent changes will be grouped artefactually closely. Convergent evolution is comparable to convergence or parallelism in manuscript traditions (Salemans 1996). So, for example, scribes working independently but in the same geographical area might alter words to fit their own dialect. An example of this might be the substitution of 'kirk' for 'church' in northern England and Scotland. Thus the same change may happen in two or more manuscripts not as a result of common ancestry, but as a result of having been produced in the same part of the country.

4. Conclusions

The process of incorporation of changes into DNA mirrors the incorporation of changes into manuscripts. For this reason, programs for phylogenetic analysis of sequence data can be exploited for stemmatic analysis of manuscript traditions, and we believe the SplitsTree program has particular advantages. Just as phenomena such as recombination, transposition and convergent evolution may pose problems for the evolutionary biologist, there are closely parallel problems in stemmatic analysis. We hope that the development of techniques in one discipline to deal with these problems will help in their solution elsewhere.

Acknowledgements

We thank Matthew Spencer and Barbara Bordalejo for helpful discussions, and the Leverhulme Trust for financial support.

References

Barbrook, A. C., C. J. Howe, N. Blake, & P. Robinson (1998). "The phylogeny of the Canterbury Tales". *Nature, 394*, 839.

Beanland, T. & C. J. Howe (1992). "The inference of evolutionary trees from molecular data". *Comp. Biochem. Physiol., 102B*, 643–659.

Cameron, H. D. (1987). "The upside-down cladogram: problems in manuscript affiliation". In H. M. Hoenigswald & L. F. Wiener (Eds.), *Biological metaphor and cladistic classification: an interdisciplinary perspective* (pp. 227–242). London: Frances Pinter.

Felsenstein, J. (1988). "Phylogenies from molecular sequences: inference and reliability". *Ann. Rev. Genet., 22*, 521–565.

Holmes, E. C., M. Worobey, & A. Rambaut (1999). "Phylogenetic evidence for recombination in dengue virus". *Mol. Biol. Evol., 16*, 405–409.

Howe, C. J., A. C. Barbrook, M. Spencer, P. Robinson, B. Bordalejo, & L. R. Mooney (2001). "Manuscript evolution". *Trends Genet., 17*, 147–152.

Huson, D. H. (1998). "Splitstree: a program for analyzing and visualizing evolutionary data". *Bioinformatics, 14*, 68–73.

Lee, A. R. (1989). "Numerical taxonomy revisited: John Griffith, cladistic analysis and St. Augstine's *Quaestiones in Heptateuchem*". *Studia Patristica, 20*, 24–32.

Medgyesy, P., E. Fejes, & P. Maliga (1985). "Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid". *Proc. Natl Acad. Sci. USA, 82*, 6960–6964.

Mooney L. R., A. C. Barbrook, C. J. Howe, & M. Spencer (2001). "Stemmatic analysis of Lydgate's *Kings of England:* a test case for the application of software developed for evolutionary biology to manuscript stemmatics". *Revue d'Histoire des Textes, 31*, 275–297.

O'Hara, R. & P. Robinson (1993). "Computer-assisted methods of stemmatic analysis". In N. Blake & P. Robinson (Eds.), *The Canterbury Tales project occasional papers*, Vol. 1 (pp. 53–74). London: Office for Humanities Communication Publications.

Platnick, N. I. & H. D. Cameron (1977). "Cladistic methods in textual, linguistic, and phylogenetic analysis". *Syst. Zool., 26*, 380–385.

Salemans, B. J. P. (1996). "Cladistics or the Resurrection of the Method of Lachmann". In P. van Reenen & M. van Mulken (Eds.), *Studies in Stemmatology* (pp. 3–55). Amsterdam: John Benjamins.

Voet, D., J. G. Voet, & C. W. Pratt (1999). *Fundamentals of Biochemistry*. New York: John Wiley & Sons.

Wattel, E. & M. J. P. van Mulken (1996). "Shock waves in text traditions". In P. van Reenen & M. van Mulken (Eds.), *Studies in Stemmatology* (pp. 105–121). Amsterdam: John Benjamins.

Watson, J. D. & F. H. C. Crick (1953). "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid". *Nature, 171*, 737–738.

Problems of a highly contaminated tradition: the New Testament

Stemmata of variants as a source of a genealogy for witnesses

Gerd Mink
Westfälische Wilhelms-Universität, Münster

0. Introduction

It is a well-known fact that the textual tradition of the New Testament poses a formidable challenge in the way of textual criticism and edition technique. The number of manuscripts that have come down is large; some 5600 known copies so far, although most of the older ones, in particular, have been lost. The real cause of the problems, however, is the vast degree of contamination. For such a textual tradition, existing methods of reconstructing the tradition are not sufficient and other approaches have to be developed. This study attempts to show a new way (i) of finding and evaluating the genealogical data that can be used to construct a stemma of such texts, and (ii) of constructing a stemma that reflects all genealogical data.[1] The concept of coherence (cf. paragraph 4.11) will be essential for the analysis of genealogical relationships. The method is based on two design choices:

(1) *Instead of trying to start with the construction of overall structures of the relations between witnesses, the first step is to construct local stemmata of variants. Local stemmata consist of trees based upon just one place of variation, and not more than one. If possible, a local tree is constructed for each place of variation.*

These local stemmata are oriented as far as possible, before being used as a basis of (sub)stemmata of witnesses or textual states (not of manuscripts). The following example illustrates how a local stemma can be constructed.

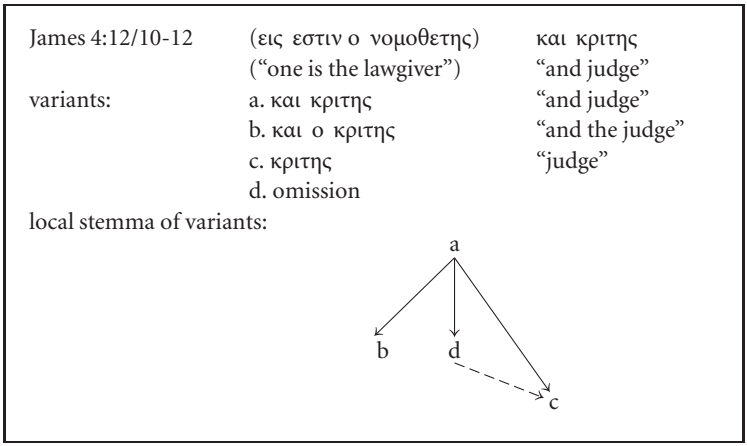

Figure 1. A local stemma of variants

Following the custom of the *Editio Critica Maior* (*ECM*, cf. the next paragraph) the places of variation will be referred to by chapter and verse as well as word address.[2] Lower-case letters indicate the variants. In the example from James 4:12/10-12 the best hypothesis is that variant *a* represents the original text, since the word κριτής ('judge') is very important for the author's argument in the context. The omission in variant *d*, however, is easily explained as caused by *homoioteleuton*.[3] Variant *b* occurs in rather unimportant witnesses and only adds the article to variant *a*.[4] Variant *c* occurs in a single witness only;[5] this witness is quite distant from even its closest relatives,[6] which are part of the *d* attestation.[7] Even more distant are the relatives that belong to the *a* attestation. Apparently, variant *c* is flawed, as the word καί ('and'), which is in fact indispensable in the context, is missing. The stemma rests on the assumption that variant *c* is based upon variant *d* and corrected the omission in accordance with variant *a*. In that process, καί would have been overlooked, a typical error during revisions. However, it is also possible (hence the interrupted edge from *d*) that the only basis was variant *a*, and καί was omitted on account of the similarity of its initial letter with that in κριτής (*homoioarkton*); yet the closest relatives of the variant *c* witness (631) are all in the *d* attestation.[8]

(2) *The contamination in the tradition is viewed as a process. The assumption is that, if contamination occurs, it emerges from those texts which were at the disposal of the scribe, i.e. texts in his direct environment, i.e. texts which are, for the most part, closely related with each other.*

James 5:7/45 "The farmer waits for the precious fruit being patient about it"
(εως λαβη προιμον και οψιμον)
("until he receives an early and late one" /
"until he/it receives something early and something late")

variants:
a. (εως λαβη) *omission* (προιμον και οψιμον)
("until he receives an early and late one" /
"until he/it receives something early and something late")
b. (εως λαβη) υετον (προιμον και οψιμον)
"(until he/it receives early and late) rain"
c. (εως) υετον (λαβη προιμον και οψιμον)
"(until he/it receives early and late) rain"
d. (εως λαβη) καρπον (προιμον και οψιμον)
"(until he receives early and late) fruit"
e. (εως λαβη) καρπον τον (προιμον και οψιμον)
"(until he receives the early and late) fruit"
f. (εως λαβη) και τον (προιμον και οψιμον)
"(until he) also (receives) the (early and late one)

local stemma of variants:

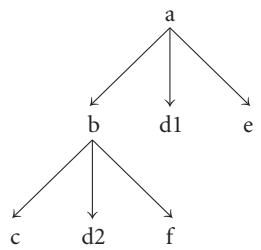

Figure 2. Local stemma of variants: a more complex case

First of all, another example will demonstrate the importance of a close relationship between witnesses for the construction of a local stemma. After that, the two examples will be used to highlight the different textual developments, so typical of a contaminated tradition, in a given witness and its closest relatives.

In Figure 2 the interpretation of variant *a* is not entirely clear. Is the meaning 'the early and late fruit', or something different? The vast majority of witnesses has variant *b*, resulting in a text of clear meaning. Because so many witnesses have this reading, relationships with witnesses of all variants can be found. Yet, it can be presumed *a priori* that the oldest layer in the attestation of variant *b* is to be found in some of its best-known witnesses, 02, 33, 81, 1852. These are more closely related to the witnesses of variant *a* than to others. The witnesses of variant *c* are particularly closely interrelated (with 96 to

99% agreements). As could be expected, further relatives are to be found in the *b* attestation. Variant *c* is not represented in an early layer of tradition. For variant *d* there is just one single closely related pair of witnesses (996, 1661). The remaining witnesses of *d* (398, 1175, a corrector of 01) are not closely related with each other; nor is there any special connection with the only witness of variant *e* (01, first hand). Variant *d* seems to have emerged on several occasions, by introducing the 'fruit' from the preceding context into the variant. 1175 has its closest relatives in the *a* attestation (03, 1739) and the *b* attestation (1243, 025; 01 of the *e* attestation is yet more remote); the other *d* witnesses have their closest relatives in the *b* attestation. One would, therefore, trace back their variant to variant *b* (cf. *d2* in the local stemma). The variant of 1175 (*d1*) could be considered to derive from either *a* or *b*. For the time being, the question remains open. The witness of *e* (01) has its closest relative (03) in the *a* attestation (the relatives in the *b* attestation are quite remote). And even if variant *f* is not grammatically impossible, it is probably an error, which could theoretically have arisen from variants *b*, *d*, or *e*. The closest relatives of the single witness of *f* (69) are in the *b* attestation, from which, accordingly, it is being derived.

It is a prerequisite for constructing a local stemma like the one in Figure 2 that the relationships between the witnesses as based on their degrees of agreement are known. In addition, one's provisional assessments of the age and quality of a number of well-known witnesses is taken into account. Close relationships between witnesses alone do not say anything about the direction of textual development. Yet, if in a relationship the ancestor and the descendant can be successfully determined, further possibilities arise to ascertain a local stemma. Thus, in the example the first assumptions were confirmed; additionally, the ancestors of 1175 (variant *d1*) were successfully located in the *a* attestation,[9] and *d1* was accordingly derived from it (see Figure 2). Arguments for the genealogical connections between variants are based partly on their content and linguistic form, and partly on the relationship between their witnesses. The example of Figure 2 makes it clear how these relationships can explain why some witnesses share the same variant and why one variant could arise from another. Yet, a lack of relationship between witnesses can also reveal coincidental correspondences (cf. variant *d*). Knowledge of the correspondences between witnesses provides a first overview, which must then be examined and supplemented against information about their genealogical relationships.

A closer look at the two instances under consideration reveals, in the different attestations of the variants, the traces of contamination among related

witnesses. Witness P74 has 8 close relatives, which contain both instances: 02, 81, 03, 2344, 1735, 218, 01, 1718 (in order of decreasing degree of agreement).[10] The distribution of the variants in the two variant places is as follows:

James 4:12/10-12	*a*	01. 02. 03. 81. 1735. 2344.
	d	P74. 218. 1718.
James 5:7/45	*a*	P74. 03.
	b	02. 81. 218. 1718. 1735. 2344.
	e	01.

Incidentally, it has emerged that all the close relatives mentioned have more older variants than P74.[11] Therefore, P74 is probably not the 'inventor' of one of the variants at the two places of variation; on the contrary, its text is based on variants which are found in different close relatives with more older variants, and, consequently, we see the traces of contamination.

1. The *Editio Critica Maior* of the New Testament

The Institute for New Testament Textual Research in Münster, founded in 1959 by Kurt Aland, has by now accomplished – apart from a few exceptions – the task of collecting the basic material of the textual tradition of the New Testament. Nearly all known manuscripts of the Greek New Testament are now available on microfilm. The next aim was to sift the material in an intelligent way, to let new views about important manuscripts find their way into the minor editions of the institute, and finally to present an *Editio Critica Maior*, which does justice to present-day standards of knowledge and methodology. A first stage of work on the *ECM* has been the research, published since 1987, on the texts of the Greek manuscripts and their relevance ('*Text und Textwert*').[12] This was intended to separate the majority of manuscripts containing the relatively uniform text which was standard at the end of the Byzantine tradition from the still large number of manuscripts which must be considered relevant on account of their deviations from the majority text. The basis was a collation of all available manuscripts according to a system of places of variation. These so-called test passages ('*Teststellen*') were known for the fact that they were different in the newer and older text forms. Research into '*Text und Textwert*' made it possible to draw on particularly those manuscripts for the *ECM*, which do not contain the uniform text from the end of the textual tradition. Nevertheless, a number of witnesses of this late text are also represented in the *ECM*.

According to text tradition, the New Testament can be divided into five sections: the Gospels, the Acts of the Apostles and the Catholic Letters (which nearly always follow Acts in the manuscripts, making up, together with Acts, the corpus of the so-called *Apostolos*), the Letters of Paul and the Revelation of John. The Catholic Letters were the first field of research in the *ECM* project.[13]

2. The Letter of James: Manuscripts and variants

The Catholic Letters, starting with the Letter of James, appeared to be especially suitable to begin work on the *ECM* with. The number of manuscripts examined in the '*Text und Textwert*' project (552) is of the same order as that for Acts, which have normally been copied together with the Catholic Letters. It is smaller than that for the Letters of Paul (about 750), and a great deal smaller than that for the Gospels (e.g. 1787 for the Gospel of Luke). Nevertheless, the number of witnesses relevant for the textual tradition is in no way smaller. The degree of contamination made the Catholic Letters very attractive for methodological investigation. As the material for the Letter of James was the first to be completely available, it has been researched most extensively. Unless otherwise indicated, the following refers to the Letter of James.

As the *ECM* is mainly interested in the text of the first millennium, we excluded nearly all the uniform witnesses representing the final state of the Byzantine tradition from our study of James, i.e. 371 manuscripts of a total of 535. In addition, 19 lectionaries were selected for the *ECM*. Yet, I did not use them for my genealogical study because they cannot really be compared with the other manuscripts; they do not contain the full, continuous text, but lessons selected to be read during services. The remaining 164 Greek manuscripts contain the continuous text, with minor omissions occurring repeatedly. Some manuscripts have also suffered more substantial damage, resulting in *lacunae* occurring throughout the text. Yet, a number of manuscripts is quite badly fragmented: 10 contain less than 150 out of the 761 places of variation, in some of them no more than a couple of these places has survived. The important manuscript 04, at least, contains more than half of the text.

The text of James contains about 1740 words; the exact number depends on textual decisions. The selected 164 Greek manuscripts, including the fragmentary ones, present 2132 genuine variants at 761 places of variation.[14] Only variants were counted that appear in the first hands of manuscripts containing the continuous text.[15] If the variants from Church Fathers and lectionaries and from early versions (Latin, Coptic, Syriac etc.) of the New Testament are

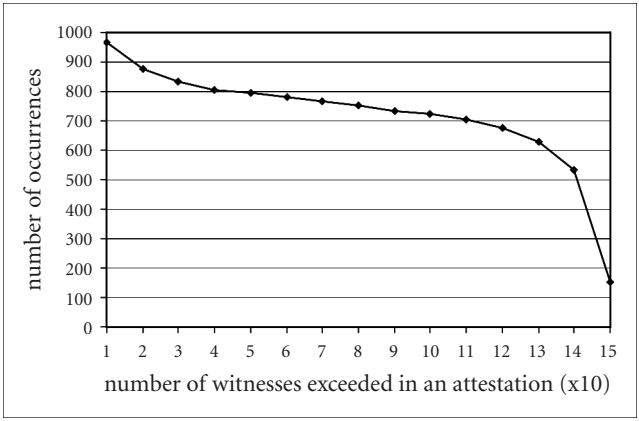

Figure 3. Distribution of size of the attestations

included, the total number for these 761 places is 2349.[16] Since many of these 761 places comprise more than one word and since the text consists of about 1740 words, it follows that about half the text is subject to variation.

The variants include all types, except for type 0 variations.[17] Type 3 variations hardly play a significant role. Typical variations are

A / BCDEFGHIJKLMNOPQRSTUVWXYZ	(type 1)
ABCDEFGH / IJKLMNOPQRSTUVWXYZ	(type 2)
ABC / DEFGH / IJKLMNO / PQRS / TUV / WXYZ	(type 4)

Type 4 variations are the most frequent. In Chapter 2 of James, spot checks revealed about 40% type 4 variations, 30% type 1 variations and 25% type 2 variations. The remainder was made up of type 3 variations. The distribution was atypical in some passages. A single witness is often responsible for the fact that type 1 variations occur more frequently in a major section of the text. On the other hand, longer units of variation in which several changes are mutually interdependent in all probability lead to type 4 variations.

The values in Figures 3 and 4 are based on the variants of the 164 witnesses at the 761 places of variation, with each witness being represented by the first hand in a manuscript, and by the text (lemma) in a commentary.[18] In Figure 3 the horizontal axis gives the number of witnesses (to be multiplied by 10) which is exceeded in an attestation. The vertical axis tells how often such an attestation occurs (e.g. about 700 attestations cover more than 110 witnesses).[19] Figure 3 gives an indication of how the quantity of the attestations is distributed over their total number. The total number of variants and there-

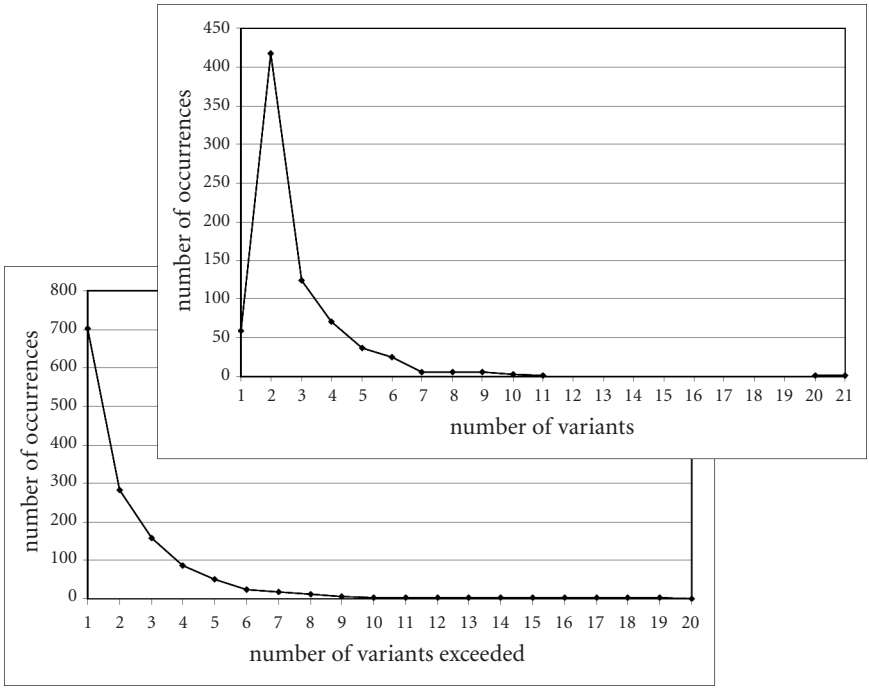

Figure 4. Distribution of number of variants per place of variation

fore of attestations is 2132 (see above). It is apparent that small attestations (of up to 10 witnesses) are the most frequent, for there are only 965 instances of more than 10 witnesses. It is not surprising that this result corresponds to a high frequency of very large complementary attestations (more than 120 witnesses, 676 instances). In comparison with that, medium and large attestations make up only a relatively small part (e.g. there are only 128 instances of 41–120 witnesses).

Figure 4 (top) shows how many variants (horizontal axis) occur in how many places of variation (vertical axis). In Figure 4 (bottom) the horizontal axis shows the number of variants which is exceeded in one place of variation, and the vertical axis gives the number of places of variation with this number of variants. Thus, while at the top it is shown that there are about 125 places with 3 variants,[20] the lower diagram demonstrates that in about 150 places there are more than 3 variants.[21] More than half of the places of variation have only 2 variants. More than 6 variants occur in 25 instances, more than 10 in only 4. A comparatively large part is made up of places of variation with 3 to 6 variants.

Letter of James

about 1740 words

535 Greek manuscripts available

164 manuscripts selected
761 places of variation
2132 variants of the 164 manuscripts

143 cases identical attestation
800 occurrences

104 cases identical one witness attestation
685 occurrences

39 cases of identical more-than-one-witness-attestation
115 ocurrences

Figure 5. The Letter of James; some numbers

The existence of 59 places of variation with only one variant is caused by deviating variants stemming from lectionaries, Church Fathers, or early versions which are assumed to be based on Greek exemplars at these places, but cannot be traced back to the first hands in the 164 manuscripts.

The following numbers are based on the same data as Figure 3 and Figure 4: the variants of the 164 witnesses in 761 places of variation. If corrections, marginal variants, variants in commentaries, and the evidence from lectionaries, Church Fathers and early versions were included in the analysis, not only the total number of variants would be much larger, but the number of occurrences of identical attestations would also be smaller.

The 2132 variants represent 2132 attestations of which 1332 are unique, i.e. there is no other attestation containing only the same witness or the same combination of witnesses. These unique attestations include those in the 59 places mentioned above.[22] The chances of finding identical attestations among these are slight, as in each of the 59 places a number of witnesses is lacking from the 164 (due to fragmentation, unmotivated omission of larger passages or the like).

There are 143 different cases in which identical attestations occur more than once. The sum of all the occurrences of these 143 cases is 800. Out of these 143 cases of identical attestations 104 concern attestations comprising only one witness. The sum of all their occurrences is 685. The 39 remaining cases relate to multiple-witness-attestations. The sum of their occurrences is 115. Of these 39 cases, 12 concern large complementary attestations in two neighbouring