XÂY DỰNG PHẢ HỆ VĂN BẢN *NGÔN CHÍ THI TẬP* CỦA PHÙNG KHẮC KHOAN BẰNG PHẦN MỀM PAUP V4.0

Phùng Diệu Linh*

1. Mở đầu

Xây dựng sơ đồ truyền bản thông qua quy nạp hệ thống văn bản của một thư tịch cổ là việc làm không thể thiếu trong công tác hiệu khám văn bản. Tuy nhiên, thách thức đặt ra với người nghiên cứu là: trong nhiều trường hợp, khi văn bản hiện tồn không còn lưu lại thông tin sao chép như: bản nguồn (bản dùng để chép), bản đích (bản được chép), người sao chép, địa điểm sao chép, thời gian sao chép, thì việc xây dựng sơ đồ truyền bản trở nên vô cùng khó khăn. *Ngôn chí thi tập* của Phùng Khắc Khoan là một trường hợp như vậy.(1) Việc căn cứ vào sự xuất hiện của chữ húy trong trường hợp này tỏ ra kém chắc chắn bởi tính thiếu nhất quán của người sao chép trong việc có hay không sử dụng chữ húy. Cụ thể, ở 80 bài đầu tiên của *Ngôn chí thi tập*, bao gồm cả chú thích dẫn giải có 57 vị trí xuất hiện chữ thời 時/辰. Trong 57 trường hợp có húy nói trên, số lượng chữ húy được sử dụng ở các bản không thống nhất, cụ thể:

	VHv.1951	VHv.1442	VHb.264c	A.555	A.1364	A.431	VHv.2163	R7
時	55	54	53	16	18	53	22	20
辰	2	3	4	41	29	4	35	37

Như vậy, ngoại trừ bản VHv.1951 không húy chữ thì 時 (2 vị trí (6,6,4) (bài 6, câu 6 chữ thứ 4) và (26,6,2) (bài 26, câu 6 chữ thứ 2) chữ *thần* 辰 xuất hiện nhưng không phải là húy, các bản còn lại đều có hiện tượng lúc húy lúc không. Do vậy, việc chỉ dựa vào chữ húy trong văn bản chép tay để minh định niên đại từ đó suy ra sơ đồ truyền bản là chưa thực sự hợp lý.

Nhà văn hiến học Trung Quốc Đổng Hồng Lợi(2) đề xuất: *"Với những văn bản nguồn gốc không rõ ràng, không còn cách nào khác là xếp riêng thành trường hợp đơn lẻ đồng thời thuyết minh thêm"*. Soi vào *Ngôn chí thi tập*, tám bản chép tay hiện tồn chỉ còn 1 bản ghi tên người sao chép, 1 bản ghi thời gian sao chép và không có bản nào chú thích về bản nguồn hay bản đích, vì thế nếu theo cách giải quyết của Đổng Hồng Lợi cả 8 văn bản của *Ngôn chí thi tập* sẽ đều được xếp vào "trường hợp đơn lẻ" và như vậy nhiệm vụ quy nạp hệ thống văn bản gần như bế tắc. Câu hỏi đặt ra là: liệu có cách nào xác định mối quan hệ giữa các văn bản hiện tồn của *Ngôn chí thi tập* không? Có thể xây dựng một phả hệ văn bản chỉ căn cứ vào những thông tin hiện có về văn bản không? Chúng tôi đi tìm đáp án những câu hỏi này trong ngành Phê bình văn bản phương Tây.(3)

2. Nội dung

2.1. Vài nét về phương pháp xây dựng Phả hệ văn bản trong Phê bình văn bản phương Tây

Có một thực tế là giới nghiên cứu thư tịch học phương Tây phải đối diện với hàng vạn bản sao Kinh thánh mà hầu hết trong số đó là bản chép tay do

* Trường Đại học Sư phạm Hà Nội.

những người chuyên chép kinh để lại. Những văn bản này hầu như không ghi tên người sao chép, thời gian và địa điểm sao chép. Trước tình hình đó, việc quy nạp hệ thống văn bản đã được các nhà thư tịch học lưu ý từ lâu, cùng với nó là sự xuất hiện và phát triển của trường phái stemma/stemmatics trong ngành phê bình văn bản

Stemma vốn xuất phát từ tiếng Latin với nghĩa là cành cây, nhánh cây. Sinh học dùng stemma để chỉ một sơ đồ biểu thị mối quan hệ tiến hóa trước sau của loài hay giữa các cá thể trong một gia đình. Stemma được dịch ra tiếng Việt là "cây gia đình" "cây phả hệ" chuyên dùng để nghiên cứu mối quan hệ giữa các loài hay sự phát sinh chủng loại trong sinh vật học. Trong ngành Phê bình văn bản học, "stemma như là một công cụ để phục nguyên văn bản, không thể thiếu một phả hệ hay một stemma của các văn bản hiện tồn, phả hệ đó hiển thị mối quan hệ lẫn nhau giữa các văn bản và cũng được sử dụng để tìm ra dấu vết của bản gốc".(4) **Stemmatics, stemmology hay stemmatology là một cách tiếp cận trong Phê bình văn bản học, chỉ những nỗ lực nhằm tái hiện lại lịch sử sao chép của một văn bản (đặc biệt là bản chép tay) dựa trên mối quan hệ giữa các văn bản hiện tồn**. Karl Lachmann (1793-1851) nhà thư tịch học người Đức được xem là ông tổ của cách tiếp cận này vì thế phương pháp stemmatics còn được gọi là phương pháp Lachmann. Tuy nhiên Lachmann không phát biểu và cũng không xây dựng hệ thống lý thuyết về stemmatics mà ông chỉ nhóm một số văn bản thành các nhóm dựa trên sự tương đồng về lỗi sai giữa chúng, bằng cách này, các nhóm "gia đình" văn bản được hình thành và mối quan hệ của chúng trở nên rõ ràng hơn. Phương pháp tiếp cận văn bản này của ông được lý luận hóa trong công trình Phê bình văn bản (textual criticsm) của Paul Mass (1880-1964) xuất bản năm 1927.(5) Thực chất chính Paul Mass mới là người định hình và phát triển phương pháp Lachmann. Hạt nhân của phương pháp Lachmann được Paul Mass phát triển là:

Những tương đồng về lỗi sai biểu thị tương đồng về nguồn gốc*, nếu văn bản J xuất hiện toàn bộ các lỗi sai của văn bản F và có ít nhất một lỗi sai của J không xuất hiện tại F thì J có nguồn gốc trực tiếp từ F*(6) và sơ đồ phả hệ văn bản được xây dựng thô sơ theo lập luận này. Paul Mass đã cố gắng xây dựng lại một văn bản "gần nhất với bản gốc"(7) thông qua các "lỗi chung". Phương pháp Lachmannn chi phối ngành Phê bình văn bản học phương Tây gần như suốt thế kỉ 19, đây là cách sơ khai nhất để xác định sơ đồ phả hệ văn bản. Trong cách tiếp cận truyền thống này, cây phả hệ được xây dựng từ dưới lên trên, sử dụng "lỗi sai chung" giữa 2 bản để xác định bản nào là bản nguồn nhánh trên trực tiếp của bản còn lại, như vậy những "lỗi sai" này mặc nhiên là đã xảy ra từ trước chứ không phải trong quá trình sao chép.(8) Đây cũng là một hạn chế trong phương pháp truyền thống. Nó không có cơ chế xem xét những "lỗi sai" riêng xảy ra ngay trong quá trình chép.

Henri Quentin (1872-1935)(9) xem xét "dị văn" thay vì các "lỗi sai" là một trong những đóng góp quan trọng nhất của ông. Theo quan sát của Quentin, vấn đề chi phối hầu hết các nhà nghiên cứu văn bản trong suốt thế kỷ trước (thế kỷ 19) đó là: có thể phục hồi được bản gốc hay không. Phương pháp cố gắng để tái thiết một phiên bản mới gần với bản gốc nhất. Quentin chỉ ra rằng mục đích bất khả thi này (việc khôi phục bản gốc) là một khiếm khuyết trong

phương pháp luận của phương pháp Lachmann. Điểm nhìn này của Henre Quentin khiến cho đóng góp của ông trở thành vô cùng có giá trị cho sự phát triển của phả hệ học văn bản.

Gần như cùng thời với Henri Quentin, Joseph Bédier (1864-1938) trong *'Best Text' Editing* chỉ ra rằng, hầu hết các sơ đồ phả hệ áp dụng phương pháp Lachmann đều chia làm 2 nhánh từ gốc (105/110 phả hệ ông quan sát), đây là một điểm bất hợp lý và thiếu cơ sở.(10) Ông đề nghị thay vì nỗ lực khôi phục lại bản gốc thì nhà nghiên cứu nên tìm những chứng cứ tốt nhất cho văn bản và sử dụng nó, văn bản tốt nhất là văn bản có số lượng tối thiểu những điểm phải hiệu chỉnh (đây chính là khởi đầu cho trường phái nghiên cứu bản nền hay văn bản cơ sở - copy text editing).

Trong khi lý thuyết về bản nền của Besdier trở nên phổ biến ở Pháp và Tây Ban Nha thì người Đức và người Ý vẫn tiếp tục phân tích và suy nghĩ về phương pháp Lachmann với những hướng rẽ của phương pháp này.

Gây được tiếng vang và ảnh hưởng sâu rộng nhất gần suốt thế kỷ 20 là Greg (1875-1959) và Dearing (1920-2005). Hai ông đã đưa ra cách tiếp cận tốt hơn đối với văn bản. Greg và Dearing tiếp tục chỉ ra những điểm bất hợp lý của phương pháp Lachmannn, đó là rất khó để xác định "lỗi sai chung" một cách khoa học (khoa học tức là có thể kiểm chứng được đồng thời có thể lặp lại được). Cách mà các nhà khoa học theo trường phái Lachmannn xác định lỗi sai đều dựa trên biện luận chủ quan người chỉnh lý và nằm ngoài tầm kiểm soát của khoa học. Họ đã chứng minh được rằng, stemma có thể xây dựng bằng 2 bước. Bước 1, những nhà phả hệ học văn bản mới này phát triển một "cấu trúc sâu" (deep-structure) của phả hệ, được gọi là các chuỗi. Bước 2, họ xây dựng một cây phả hệ từ các chuỗi đã định hình.(11) Lợi ích của phương pháp mới là: việc biện luận về nguồn gốc của dị bản đã không còn quan trọng nữa, điều đó hạn chế tính chủ quan của nhà nghiên cứu. Tuy vậy ngay cả Greg và Dearing vẫn gặp phải những hạn chế khi xử lý các dị văn phức tạp. Trong quá trình xây dựng phả hệ họ chỉ có thể dùng một mô hình dị văn. Mô hình này được gọi là dị văn loại 2, đó là những dị văn chia các văn bản hiện có thành 2 nhóm mới đủ điều kiện để xây dựng cây. Giả sử có 6 văn bản là A, B, C, D, E, F, những dị văn loại 2 chia 6 bản trên thành 2 nhóm (số lượng văn bản ở mỗi nhóm không hạn định nhưng tối thiểu phải là 2), ví dụ ở vị trí 1, các bản A, B, C, D là chữ *m* các bản E, F là chữ *n*, thì ta có chuỗi ABCD:EF. Hạn chế của phương pháp này rất rõ ràng, đó là, có rất nhiều vị trí thuộc văn bản, dị văn không phải chỉ là 2 mà còn là 3, 4, 5 thậm chí mỗi 1 văn bản chép 1 chữ khác nhau thì Greg và Dearing không xử lý được.

Năm 1973, thành công vang dội của lý thuyết phân loại số trong ngành sinh học(12) đã ảnh hưởng không nhỏ tới Phê bình văn bản học.(13) Các nhà phê bình văn bản đã ứng dụng thành công cách tiếp cận của ngành sinh học, sử dụng các phần mềm xây dựng cây phát sinh chủng loại(14) để tái hiện lịch sử sao chép văn bản tạo nên một giai đoạn mới trong nghiên cứu Phả hệ văn bản gọi là New-stemmatics - Tân phả hệ văn bản. Đây được xem là giai đoạn "bùng nổ" những thành tựu xuất sắc trong nghiên cứu Phả hệ văn bản với những đóng góp tiêu biểu của Ben Salemans, Robinson và O'Hare.

Năm 1987, Ben Salemans công bố nghiên cứu về phương pháp mà ông gọi là Phê bình văn bản phân nhánh (cladistic textual criticism). Theo ông, có một sự tương đồng lớn giữa công việc của nhà nghiên cứu sinh vật học và các nhà nghiên cứu văn bản. Trong khi nhà sinh vật học theo trường phái Phân loại học phân tích nhánh (cladistic) căn cứ vào đặc điểm của các loài để phân loại và xếp chúng trên cây tiến hóa thì các nhà phê bình văn bản cũng nỗ lực tìm ra sơ đồ phả hệ của các văn bản cổ. Vì sự tương đồng này, người nghiên cứu văn bản hoàn toàn có thể tận dụng thành tựu nghiên cứu của ngành sinh vật học mà gần gũi hơn cả là Phân loại học phân tích nhánh (cladistic) trong việc tái hiện lịch sử sao chép của văn bản. Theo Salemans: dị bản văn bản có thể được xem như một đơn vị phân loại (taxa) và dị văn đó phải bộc lộ được các mối quan hệ gần gũi giữa các văn bản.(15)

Cùng giai đoạn này, vào năm 1992 một nhóm nhiên cứu khác là Robinson và O'Hara công bố công trình ứng dụng phần mềm PAUP và thành tựu ngành Phân loại học phân tích nhánh trong phân loại tiến hóa sinh học để xây dựng sơ đồ phả hệ 44 văn bản tác phẩm *Svipdagsmal*. Kết quả công trình này được trình bày trong *Báo cáo về lời thách thức nghiên cứu phê bình văn bản năm 1991*.(16) Sở dĩ có tên gọi này vì đây là thành tựu xuất phát từ lời thách thức của nhà thơ, nhà nghiên cứu Housman rằng, thống kê và toán học không có chỗ trong nghiên cứu văn bản:

Một nhà phê bình văn bản thực hiện công việc của anh ta không giống như Newton khảo sát về chuyển động của các hành tinh, anh ta giống như con chó đang săn bọ chét. Nếu con chó săn bọ chét trên các nguyên tắc toán học hoặc, dựa vào các thống kê về diện tích và dân số thì nó không bao giờ bắt được con nào ngoại trừ nhờ vào ngẫu nhiên.(17)

Robinson, O'Hara và 9 nhà nghiên cứu chia làm 3 nhóm nhỏ cùng khảo sát văn bản *Svipdagsmal* (bao gồm 2 phần "Grougaldr" và "Fjolsvinnsmal") dài 1.500 từ, 47 dị bản. Tác phẩm này được viết ở Iceland, Đan Mạch và Thụy Điển. Ban đầu Robinson dùng cách truyền thống của nghiên cứu văn bản học để phân tích 44 văn bản *Svipdagsmal* (ông cho rằng 3 bản còn lại không có giá trị, không nói rõ lý do) và xây dựng nên 1 cây phả hệ theo phương pháp biện luận truyền thống. Nhóm thứ hai sử dụng phần mềm thống kê, nhóm này mặc dù ra được các số liệu nhanh chóng nhưng không biểu thị được mối quan hệ giữa các văn bản. Nhóm thứ 3, O'Hara sử dụng phần mềm PAUP theo phương pháp phân tích Parsimony,(18) chỉ trong 5 phút cây phả hệ văn bản *Svipdagsmal* đã hoàn thành, kết quả trùng khớp với kết quả kiểm tra của nhóm 1 đưa ra sau đó vài tháng. Báo cáo này nhanh chóng được xuất bản liên tục trên internet (57 lần xuất bản) và trở thành hiện tượng đột phá trong nghiên cứu phả hệ văn bản. Lợi ích lớn nhất của việc sử dụng máy tính trong phân tích và xây dựng phả hệ văn bản là tiết kiệm tối đa thời gian, chính xác tối ưu kết quả trong xử lý số lượng văn bản lớn. Tiếp nối thành công này Robinson và các cộng sự đã công bố nhiều thành quả nghiên cứu ứng dụng thành tựu ngành phân loại học phân tích nhánh cùng với sự hỗ trợ của phần mềm xây dựng cây phát sinh chủng loại PAUP.(19)

Năm 2000, Ben Salemans tiếp tục hướng nghiên cứu của mình và kế thừa các thành công của nhóm Robinson, O'Hara trong luận án tiến sĩ có tên

Xây dựng phả hệ bằng máy tính trong Phân loại học phân tích nhánh, phương pháp Tân Lachmann, trường hợp 14 văn bản Lanseloet van Denemerken,[20] ông đã hoàn thiện cơ sở lý luận ngành Tân phả hệ văn bản (New Stemmatics, ông gọi là Tân Lachmann) dựa trên sự kết hợp thành tựu Phê bình văn bản học và Phân loại học phân tích nhánh.

Salemans đề ra 5 bước[21] để xây dựng một stemma cho văn bản. Lấy 14 văn bản tác phẩm kịch Hà Lan *Lanseloet van Denemerken* như là trường hợp nghiên cứu, sử dụng phương pháp Maximum Parsimony với sự hỗ trợ của phần mềm PAUP để dựng phả hệ 14 văn bản này. Với 6 nguyên tắc lựa chọn dị văn, Salemans tạm thời khu trú được những rắc rối của người nghiên cứu phả hệ văn bản khi đối diện với các trường hợp dị văn phức tạp (là dị văn được tạo ra khi người sao chép sử dụng nhiều hơn 1 bản nguồn, dị văn xuất hiện trong bản của anh ta có thể được tổng hợp từ dị văn của các bản nguồn cũng có thể là được tạo mới một cách vô tình hay cố ý trong quá trình sao chép). Những người như Robinson, Salemans thực sự đã tạo nên một bước ngoặt mới trong nghiên cứu lịch sử sao chép văn bản mà ta gọi là Tân phả hệ văn bản (New Stemmatics).

Tân phả hệ văn bản mã hóa các dị văn giữa các văn bản theo nguyên tắc nhị phân: đối với dị văn có trong văn bản được ghi là (1), không có được ghi là (0) và được coi như là một chuỗi đặc điểm của văn bản. Trên cơ sở dữ liệu này, các văn bản sẽ được hệ thống hóa theo phương pháp Maximum Parsimony (MP) với sự hỗ trợ của phần mềm phân tích dữ liệu PAUP (O'Hara, Ben Salemans) hay tổng hợp các phương pháp Maximum Parsimony, Distance Matrix hay Likelihood Methods[22] trong Mac Clade hoặc PHYLIP (Lee AR),[23] qua đó các văn bản sẽ được hệ thống hóa trên cây phả hệ biểu thị mối quan hệ giữa các văn bản một cách nhanh chóng rút gọn tối đa về thời gian và cho kết quả chính xác.

Từ năm 2000 tới nay, cùng với tiến bộ của công nghệ gen và sự phát triển các phương pháp, các phần mềm phân tích phát sinh loài, những nhà nghiên cứu văn bản theo trường phái Tân phả hệ vẫn tiếp tục đi tìm lời giải đáp tối ưu hơn cho phả hệ văn bản đặc biệt là xử lý dị văn phức tạp mà Salemans mới chỉ ra biện pháp khu trú mà chúng tôi đã đề cập ở trên. Dị văn phức tạp thực chất cũng tương tự như quá trình chuyển gen ngang (horizontal gene) trong công nghệ gen. Thay vì chuyển gen dọc tức di truyền từ bố mẹ sang con thì gen hoặc nguyên liệu di truyền được chuyển từ cá thể này sang cá thể khác nhờ quá trình tương tự sự gây nhiễm, quá trình này được gọi là công nghệ DNA tái tổ hợp. Phương pháp mới trong phân tích phát sinh loài giúp hình thành mạng lưới phát sinh loài (phylogenetic networks) thay vì cây phát sinh loài (Neighbornet hoặc Splitstree) đã được giới Tân phả hệ văn bản ứng dụng thành công.[24] Cùng thời gian này nhiều công trình kiểm tra kết quả của Tân phả hệ văn bản cũng đã ra đời. Nhà nghiên cứu thiết lập quá trình truyền bản giả bằng cách đưa ra một văn bản cho nhiều người sao chép, hoặc một người chép nhiều lần khác nhau, hoặc bản sao lại được dùng làm bản nguồn để sao lần hai... sau đó nhập dữ liệu dị bản, dị văn và dùng các phương pháp khác nhau xây dựng cây/mạng lưới phả hệ văn bản. Kết quả này được so sánh với sơ đồ sao chép thực tế để kiểm nghiệm xem kết quả nào giống với thực tế nhất. Cách làm này đã chỉ rõ được ưu nhược điểm của từng phương pháp, nó giúp

cho giới nghiên cứu văn bản có thêm dữ liệu khi quyết định sử dụng phương pháp nào trong nghiên cứu.(25)

Có thể nói, Tân phả hệ văn bản là thành tựu của quá trình phát triển không ngừng nghỉ trong ngành Phê bình văn bản học phương Tây. Từ phương pháp biện luận thuần logic của Lachmann, Paul, Henri Quentin, Bédier trải qua giai đoạn bắt đầu sử dụng thuật toán trong tính toán (Greg, Dearing) và cuối cùng Tân phả hệ văn bản sử dụng thành tựu nghiên cứu của phương pháp Phân loại học phân tích nhánh (Cladistic) với sự hỗ trợ của các phần mềm xây dựng phả hệ (Lee, Robinson, Salemans, Windram, Spencer...) là bước phát triển vượt bậc với những đổi thay căn bản trong nghiên cứu phả hệ văn bản. Cho tới thời điểm hiện tại, khi các dự án nghiên cứu văn bản được thực hiện rầm rộ bởi nhiều nhóm học giả uy tín thì Tân phả hệ văn bản với những ứng dụng thành tựu từ Phân loại học, công nghệ gen vẫn là xu hướng nghiên cứu chiếm ưu thế và gặt hái nhiều thành công.(26)

2.2. Xây dựng sơ đồ phả hệ văn bản Ngôn chí thi tập của Phùng Khắc Khoan bằng phần mềm PAUP V4.0

2.2.1. Đối tượng và phạm vi nghiên cứu

Đối tượng: 8 văn bản: VHv.1951, VHv.1442, VHb.264c, A.555, A.1364, A.431, VHv.2163 và R7.

Phạm vi: Khảo sát dị văn 80 bài đầu tiên trong 2 quyển 1, 2 của *Ngôn chí thi tập*. Nguyên nhân của việc chỉ lựa chọn 80 bài đầu tiên là do trong số 8 văn bản chép thay *Ngôn chí thi tập* mà chúng tôi sưu tầm và khảo sát thì bản A.431 chỉ chép 80 bài, từ bài số 81 chép thơ đi sứ thuộc *Mai Lĩnh sứ hoa thi tập*. Do đó chúng tôi tạm coi đây là giới hạn cho phạm vi khảo sát.

2.2.2. Phương pháp nghiên cứu

- Cách mã hóa dữ liệu: Dữ liệu được tính toán dựa trên 1879 dị văn xuất hiện ở 8 văn bản theo nguyên tắc nhị phân, đối với dị văn xuất hiện ở văn bản này mà không có ở văn bản khác được mã hóa là (1), ngược lại đối với dị văn không xuất hiện ở văn bản này mà có ở văn bản khác được mã hóa là (0); các văn bản có cùng dị văn xuất hiện hoặc không xuất hiện được mã hóa giống nhau. Do dữ liệu nhập theo nguyên tắc nhị phân nên các dị văn được tách nhỏ:

Bảng 1. Ví dụ về cách thức nhập dữ liệu đầu vào

1	2	3	4	5	6	7	8	9	10
Địa chỉ	**chữ**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
1, 1, 7	餘	1	1	0	0	0	1	1	0
1, 1, 7	初	0	0	0	1	0	0	0	0
1, 1, 7	秋	0	0	1	0	1	0	0	1

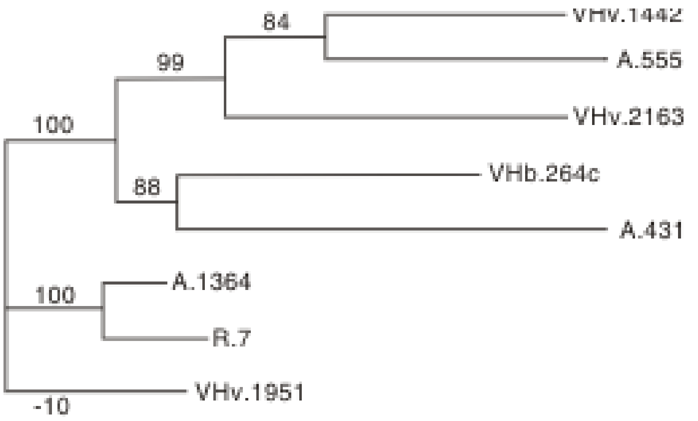
Cột 1: Địa chỉ dị văn, 1, 1, 7 nghĩa là bài 1, câu 1 chữ thứ 7.

Cột 2: Ở địa chỉ này, dị văn xuất hiện 3 chữ, chữ 餘 xuất hiện tại các văn bản A, B, F, G các vị trí này mã hóa là "1", các bản còn lại không có 餘 mã hóa là "0". Tương tự 初 xuất hiện tại D, 秋 xuất hiện tại C, E, H.

Theo các thức nhập dữ liệu dị văn này thì 1.879 dị văn trong 8 bản ở 80 bài *đầu Ngôn chí thi tập* được tách nhỏ thành 2522 lượt văn tự (tương ứng 2.522 dòng giá trị excel).(28)

- Phương pháp phân tích số liệu: dị văn trong các văn bản được phân tích bằng phương pháp Maximum Parsimony (Robinson và O'Hara, 1996) bằng phần mềm PAUP 4.0 (Swofford, 1998)[29] trong đó giá trị của các dị văn là như nhau trong các văn bản. Tần suất lặp lại của mối liên hệ giữa các văn bản (Bootstrap) được tính toán nhắc lại 100 lần.

2.2.3. Kết quả và thảo luận



Hình 1: Sơ đồ phả hệ biểu thị mối liên hệ giữa các văn bản.

Kết quả phân tích dị bản bằng phương pháp Maximum Parsimony cho thấy trên tổng số 2.522 lượt văn tự phân tích, 1.551 lượt không có giá trị phân tích, 971 lượt có giá trị phân tích và đưa ra một mô hình liên hệ giữa các văn bản duy nhất (Hình 1):

Bảng 2. Tỷ lệ dị bản sai khác (%) (dòng trên) và số lượng dị bản sai khác (dòng dưới) giữa các văn bản nghiên cứu.

		1	2	3	4	5	6	7	8
1	VHv.1951	-	26.1	29.3	37.8	15.4	36.1	37.4	16.3
2	VHv.1442	657	-	33.3	30	28	41.7	40	29
3	VHb.264c	737	839	-	41.6	30.5	42.5	45.7	31.6
4	A.555	952	757	1049	-	34.1	46.7	42.4	35.2
5	A.1364	388	707	768	860	-	35.4	36.6	6.1
6	A.431	910	1050	1071	1179	893	-	52.5	36.2
7	VHv.2163	941	1008	1152	1070	922	1323	-	36.5
8	R7	411	732	797	887	153	914	919	-

Mối liên hệ giữa các văn bản dựa trên phân tích dị bản bằng phương pháp Maximum Parsimony. Chiều dài cây = 3399, CI = 0.74, RI = 0.45. Số ở gốc là giá trị Bootstrap

Theo đó các văn bản được chia thành 04 nhóm:

- Nhóm 1: VHv.1442, A555, VHv.2163

Văn bản VHv.1442 chung nhánh với văn bản A.555 nhưng giữa chúng có sự sai khác về các dị văn là 30% (tương ứng 757 dị văn) (Bảng 2) và chúng có mối quan hệ gần gũi với văn bản VHv.2163.

- Nhóm 2: VHb.264c, A.431. Quan sát bảng 2 chúng ta nhận thấy nhóm 2 là nhóm có số lượng dị văn sai khác và tỷ lệ dị văn sai khác cao hơn hẳn các nhóm khác.
- Nhóm 3: A.1364, R7.
- Nhóm 4: VHv.1951.

3. Kết luận

Kết quả phân tích thông qua phần mềm PAUP V4.0 chỉ ra rằng 8 văn bản *Ngôn chí thi tập* thuộc 4 nhóm, trong đó VHv.1951 đứng riêng, VHv.1442,

A.555 cùng nhánh và có quan hệ gần gũi với VHv.2163, VHb.246c và A.431 cùng nhánh, A.1364 và R.7 cùng nhánh. Việc chỉ ra mối quan hệ giữa các văn bản sẽ quyết định bản nào trở thành bản nền, bản nào là bản đối hiệu trong công tác hiệu khám văn bản.

Văn bản dùng để hiệu khám là các bản đại diện cho các nhóm khác nhau, bản đó có số lượng dị văn sai khác và tỷ lệ dị văn sai ít hơn các bản khác trong cùng nhóm. Theo nguyên tắc này các bản dùng trong công tác hiệu khám văn bản sẽ là VHv.1951, A.1364, VHb.264c, VHv.1442. Tuy nhiên, như trên đã nói, nhóm 2 là nhóm có khoảng cách sai khác dị văn lớn hơn hẳn các nhóm còn lại, bản VHb.264c cũng bỏ qua hầu hết các chú thích dẫn giải khi sao chép nên độ khả tín của văn bản này không cao, do đó chúng tôi không dùng bản này trong công tác hiệu khám. Như vậy 3 bản được dùng làm bản hiệu khám sẽ là VHv.1951, A.1364 và VHv.1442.[*]

P D L

CHÚ THÍCH

(1) Xem thêm: Phùng Diệu Linh (2014), "Lược khảo văn bản *Ngôn chí thi tập* của Phùng Khắc Khoan", tạp chí *Khoa học,* Trường Đại học Sư Phạm Hà Nội, Quyển 59, Số 6BC, tr. 13-19.

(2) Đổng Hồng Lợi (2013), *Cổ điển văn hiến học cơ sở,* Bắc Kinh đại học xuất bản xã, tr. 126.

(3) Phê bình văn bản dịch từ cụm "textual criticism". Cụm này không hoàn toàn tương ứng với "văn bản học" mà chúng ta dùng trong tiếng Việt. Thực tế nó gần với "hiệu khám học" hơn. Trong Tô Kiệt (2009), *Tây phương hiệu khám học luận trứ*, Thượng Hải nhân dân xuất bản xã, cũng dịch textual criticism thành "hiệu khám học".

(4) Salemans, B. J. P, (2000), *Building stemmas with the Computer in a Claditics Neo-Lachmannnian, Way the case of Fourteen Text Version of* Lanseloet van Denemerken*,* ISBN 90-9013480-8 (Ben Salemans), tr. 10.

(5) Paul Mass, *Textual Criticism*, (Chúng tôi dùng bản tiếng Anh của Barbara Flower (1958), dịch từ tiếng Đức, Nxb Oxford at the Clarendon Press).

(6) Paul Mass, đã dẫn, tr. 4.

(7) Paul Mass, đã dẫn, tr. 1.

(8) Ph. V. Baet, C.Macé, P.Robinson (2004), *Testing methods on an artificially createt textual tradition,* The Evolution of Texts: Confronting stemmatological and Genetical Methods, Proceedings of the International Workshop held in Louvain-la-Neuve on September 1-2, tr. 264.

(9) William P. Shepard (1930), *Recent Theories of Textual Criticism*, Modern Philology, Vol. 28, No. 2, pp. 129-141.

(10) Dẫn theo Christopher J. Howe, Heathes F. Windram (2011), *Phylomenetics-Evolutionary Analysis beyond the Gene,* PloS Biol 9 (5), Issue 5, e1001069.

(11) W. W. Greg (1927), *The calculus of variants, an essay on textual criticism*, Oxford at the Clarendon Press.

(12) Sokal & Sneath (1973), *Principles of Numerical Taxonomy*, San Francisco: W.H. Freeman. Phân loại số (Numerical Taxonomy) là một hệ thống phân loại trong sinh học nhằm xử lý các nhóm sinh vật dựa trên sự tương đồng về đặc điểm của chúng. Công bố này tính tới năm 1987 đã được trích dẫn ở hơn 2.015 bài báo, tạo nên một "cơn sốt" về ứng dụng công nghệ thông tin, máy tính trong nghiên cứu khoa học.

(13) Christopher J. Howe, Heathes F. Windram (2011), xem chú thích (10).

(14) Cây phát sinh chủng loại (Phylogenetic tree) hay còn gọi là Cây phả hệ; Cây tiến hóa, là sơ đồ hình cây được dùng để mô hình hóa lịch sử tiến hóa thực tế của một nhóm các trình tự

* **Lời cảm ơn**: Chân thành cảm ơn TS Phan Kế Long, PGĐ Bảo tàng Thiên nhiên Việt Nam, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã nhiệt tình chỉ dẫn cho tôi những kiến thức ngành Sinh học và hỗ trợ phần mềm PAUP để tôi có thể hoàn thiện tốt bài báo này. P D L.

hay các sinh vật. (Xem: *http://encyclopedia.thefreedictionary.com/Phylogenetic+tree,* truy cập ngày 01/5/2015).

(15) Ben Salemans (2000), xem chú thích (4), Appendix E., page 601.

(16) Robinson, P. M. W., & O'Hara, R. J (1991), *Report on the textual criticism challenge 1991*, Humanist Discussion Group, Vol. 5, No. 0262. Wednesday, 31 Jul.

(17) A.E. Housman (1921), *The Application of Thought to Textual Criticism*, Proceedings of the Classical Association, Vol XVIII. The meeting of the Classical Society in Cambridge England, August.

(18) M. Witzet 2014, *Textual criticism in Indology and in European philology during the 19th and 10th century,* Electronic Journal of Vedic Studies (EJVS),Vol. 21, 2014, Issue 3, pp. 9-91. Parsimony là một trong ba phương pháp trong nghiên cứu phân tích phát sinh loài (phylogenetic analysis) bao gồm: Distance Methods, Maximum Parsimony, Maximum Likelihood. Nguyên tắc cốt lõi của phương pháp Parsimony trong Phê bình văn bản phân nhánh học miêu tả là: cây phả hệ tốt nhất là cây có số lượng dị văn biến đổi thấp nhất. Trang 70.

(19) Xem: - Robin C. Cover and Peter M. W. Robinson (1995), *Encoding Textual Criticism,* Computers and Humanities, Vol 29, No 2, The Text Encoding Initiative: Background and Source, pp. 123-136.

- Robinson, P. M. W., & O'Hara, R. J (1996), *Cladistic analysis of an Old Norse manuscript tradition.* In S. Hockey, & N. Ide (Eds.), Research in humanities computing 4 (pp. 115-137), Oxford University Press.

(20) Ben Salemans, (2000), xem chú thích (4).

(21) Ben Salemans (2000), tr. 63.

(22) M. Witzet (2014), xem chú thích (16).

(23) Christopher J. Howe, Heathes F. Windram (2011), xem chú thích (10).

(24) Windram HF, Howe CJ, Spencer M (2005), *The identification of exemplar change in the Wife of Bath's Prologue using the maxium chi-square method*, Literay and Linguistic Computing 20: 189-204.

(25) Ph. V. Baet, C.Macé, P.Robinson (2004), xem chú thích (8).

(26) Xem: *http://www.tei-c.org/About/Archive_new/ETE/Preview/robinson.xml; http://www.textual scholar ship.org/gencrit/index.html*

(27) Dị văn (dịch từ chữ *variant*) là những chữ chép sai khác làm ảnh hưởng tới nội dung văn bản. Những chữ giản thể, dị thể, tá tự không được liệt kê ở đây, theo Ben Salemans (2000) những chữ có cùng nội dung nhưng khác nhau về chính tả là những chữ không có thông tin trong phân tích phả hệ. Variant trong một số trường hợp cũng được dùng để chỉ dị bản.

(28) Swofford, D.L. PAUP* Phylogenetic analysis using parsimony. Version 4. Sinauer, Sunderland, MA, 128 pp, (1998). Do bảng dữ liệu đầu vào này rất dài (71 trang A4 cỡ chữ 12) nên chúng tôi không đưa vào đây mà chỉ lấy ví dụ như Bảng 2.

TÀI LIỆU THAM KHẢO

1. Christopher J. Howe, Heather F. Windram (2011), *Phylomemetics—Evolutionary Analysis beyond the Gene,* PLoS Biology, *www.plosbiology.org*, 1 May 2011, Volume 9, Issue 5, e1001069.
2. Michael Witzel (2014), *Textual criticism in Indology and in European philology during the 19th and 20th centuries,* Electronic Journal of Vedic Studies (EJVS), Vol. 21, Issue 3.
3. Ph. V. Baet, C. Macé, P. Robinson (2004), *Testing methods on an artificially created textual tradition,* The Evolution of Texts: Confronting stemmatological and Genetical Methods, Proceedings of the International Workshop held in Louvain-la-Neuve on September 1-2.
4. Paul Mass (1927), *Textual Criticism by Paul Mass*, translated from the Germany by Barbara Flower 1958, Oxford at the Clarendon Press.
5. Phan Kế Long, Vũ Đình Duy, Phan Kế Lộc, Nguyễn Giang Sơn, Nguyễn Thị Phương Trang, Lê Thị Mai Linh, Lê Thanh Sơn (2014), "Mối quan hệ di truyền của các mẫu sâm thu ở Lai Châu trên cơ sở phân tích trình tự Nucleotide vùng MATK và ITS- rDNA", tạp chí *Công nghệ Sinh học,* 12(2): 327-337.

6. Phùng Diệu Linh (2014), "Lược khảo văn bản *Ngôn chí thi tập* của Phùng Khắc Khoan", tạp chí *Khoa học,* Trường Đại học Sư phạm Hà Nội, Quyển 59, Số 6BC, tr. 13-19.
7. Đổng Hồng Lợi (2013), *Cổ điển văn hiến học cơ sở,* Bắc Kinh đại học xuất bản xã.
8. Peter Robinson and O'Hara (1991), "Report on the textual criticism challenge 1991", Bryn Mawr Classical Review 3.4.
9. Robin C. Cover and Peter M. W. Robinson (1995), *Encoding Textual Criticism,* Computers and Humanities, Vol 29, No 2, The Text Encoding Initiative: Background and Source, pp. 123-136.
10. Robinson, P. M. W, & O'Hara, R. J. (1996), *Cladistic analysis of an Old Norse manuscript tradition.* In S. Hockey, & N. Ide (Eds.), Research in humanities computing 4 (pp. 115–137), Oxford: Oxford University Press.
11. Salemans, B. J. P. (2000), *Building stemmas with the Computer in a Claditics Neo-Lachmannian, Way the case of Fourteen Text Version of* Lanseloet van Denemerken*,* ISBN 90-3730505-9 (Nijmegen University Press).
12. Swofford, D. L. (1998) PAUP* Phylogenetic analysis using parsimony. Version 4. Sinauer, Sunderland, MA, 128 pp.
13. William P. Shepard (1930), *Recent Theories of Textual Criticism*, Modern Philology, Vol. 28, No. 2 (Nov., 1930), pp. 129-141.
14. Trên trang *http://encyclopedia.thefreedictionary.com/Phylogenetic+tree,* truy cập ngày 01/5/2015.
15. Trên trang *http://www.tei-c.org/About/Archive_new/ETE/Preview/robinson.xml,* truy cập ngày 28/4/2015.
16. Trên trang *http://www.textualscholarship.org/gencrit/index.html,* truy cập ngày 28/4/2015.

TÓM TẮT

Tình trạng thiếu thống nhất giữa tám văn bản chép thơ *Ngôn chí thi tập* của Phùng Khắc Khoan, (gồm VHv.1951, VHv.1442, VHb.264c, A.555, A.1364, A.431, VHv.2163 và R7) cùng với sự khuyết thiếu thông tin sao chép cản trở quá trình xây dựng một sơ đồ truyền bản của các văn bản. Câu hỏi đặt ra là: Chỉ dựa trên các văn bản hiện tồn, liệu có thể xây dựng được một sơ đồ phả hệ biểu thị mối quan hệ giữa các dị bản của *Ngôn chí thi tập* không? Cơ sở lý luận ngành Phê bình văn bản học (textual criticism) cụ thể là trường phái Tân phả hệ văn bản (New- Stemmatics) cùng với sự hỗ trợ của phần mềm PAUP giúp chúng tôi trả lời câu hỏi này. Phần mềm PAUP V4.0 được dùng để phân tích khoảng cách sai khác dị văn giữa các dị bản và xây dựng cây phả hệ văn bản theo phương pháp Maximum Parsimony (MP). Cụ thể, 8 văn bản *Ngôn chí thi tập* chia thành 4 nhóm, Nhóm 1 gồm VHv.1442, A.555 và VHv.2163; Nhóm 2: VHb.264c và A.431; Nhóm 3: A.1364 và R7; Nhóm 4 (nhóm đặc biệt chỉ có 1 văn bản): VHv.1951. Căn cứ vào sơ đồ phả hệ, khoảng cách sai khác dị văn, độ khả tín của văn bản chúng tôi lựa chọn được 3 bản sẽ tham gia vào công tác hiệu khám văn bản là VHv.1951, A.1364 và VHv.1442.

ABSTRACT

BUILDING A STEMMA WITH PAUP V4.0 SOFTWARE: THE CASE OF EIGHT TEXT VARIANTS OF *NGÔN CHÍ THI TẬP* BY PHÙNG KHẮC KHOAN

The inconsistency of the eight variants of *Ngôn chí thi tập* by Phùng Khắc Khoan, (namely VHv.1951, VHv.1442, VHb.264c, A.555, A.1364, A.431, VHv.2163 and R7) and the lack of copying information do not provide us with knowledge to construct a transmission diagram of variants. This study investigates the following research question: Could we construct a stemma which represents close relations among its variants? The theoretical foundation of New-Stemmatics and the support of PAUP software help us address this question; namely, 8 variants could be categorised into 4 groups: Group 1 including VHv.1442, A.555 and VHv.2163, Group 2 including VHb.264c and A.431, Group 3 including A.1364 and R7, and Group 4 (special group having only one variant of VHv.1951). Based on this stemma and the distance among the variants we selected three variants to be used in collation of textual criticism: VHv.1951, VHv.1442 and A.1364.