**Vài suy nghĩ về giữ gìn di sản Hán Nôm**

Nguyễn Phước Lộc

Nghiên cứu tự do, Việt Nam

**Tóm tắt**

Di sản Hán Nôm là tài sản quý giá của dân tộc; về văn hóa, đó là nét độc đáo giàu bản sắc; về xã hội, thể hiện sự gắn kết của các cộng đồng người Việt; về chính trị, thể hiện tinh thần độc lập tự chủ; về khoa học và văn chương, có những trước tác vô giá; về ngôn ngữ học, đó là gạch nối giữa chữ Hán và chữ quốc ngữ. Giữ gìn di sản Hán Nôm là trách nhiệm và là niềm tự hào của chúng ta, ngoài ra còn góp phần tiếp thu, sàng lọc hay cải biến những ưu khuyết trong quá trình hội nhập và phát triển văn hóa sao cho văn hóa nội sinh ít nhất không thể lùi vì sự hấp thụ và đề kháng quan trọng như nhau. Trong bài viết này, tôi nêu một số suy nghĩ và đề xuất để giữ gìn di sản Hán Nôm.

**Từ khóa:** chữ Nôm, di sản Hán Nôm, số hóa tài liệu Hán Nôm, biên mục.

**Abstract**

Han Nom legacy (Sinitic Vietnamese and Nom legacy) is the precious property of our nation, which is an original cultural character. It indicates the community connection in Vietnamese society and represents the independent spirit about political sovereignty. It bequeaths invaluable literature works which are immortal in flow of time. Especially, Han Nom legacy is the connector between Han language (Sinitic Vietnamese) and national language (Vietnamese). Preserving Han Nom legacy is our responsibility and proudness so that we can acquire, select, and adapt new ideas and outer cultural trends along with their weak and strong points for developing our culture. The most essential mission is that the inner national culture must not be degraded when coping with the alien culture because the absorption and resistance in cultural development process have the same important level. In this report, I raise some thoughts and proposals for preserving Han Nom legacy.

**Keywords:** Sinitic Vietnamese, Han Nom legacy, digitalizing Han Nom documents, bibliography, catalogue.

**1. Mở đầu**

Di sản Hán Nôm là tài sản quý giá của dân tộc; về văn hóa, đó là nét độc đáo giàu bản sắc; về xã hội, thể hiện sự gắn kết của các cộng đồng người Việt; về chính trị, thể hiện tinh thần độc lập tự chủ; về khoa học và văn chương, có những trước tác vô giá; về ngôn ngữ học, đó là gạch nối giữa chữ Hán và chữ quốc ngữ vì chữ quốc ngữ ghi âm mà chữ Nôm cũng có phép biểu âm. Dùng chữ Hán để biểu âm nhưng bỏ ý (Wikipedia, 2020) (Tran, 2015) hoặc phép hình thanh là cách thức gần với dùng bảng chữ cái khi một chữ Hán hoặc bộ chữ Hán có thể được xem như “chữ cái”. Bất luận thế nào, tiến trình từ tượng hình đến tượng thanh là sự phát triển; ngay trong chữ Hán, phép phiên thiết, hình thanh, giả tá gần với ngôn ngữ tượng thanh, dùng những chữ tượng hình cơ bản như những “chữ cái”. Vì âm Việt nhiều hơn âm Hán nên chữ Nôm gần với chữ quốc ngữ hơn chữ Hán gần với chữ quốc ngữ. Vì đơn âm nên chữ quốc ngữ gần với Nôm hơn ngôn ngữ la-tinh gần với Nôm. Bỏ qua chính tả và cách hành văn cũ, văn phạm chữ Nôm rất gần văn phạm chữ quốc ngữ. Sự mai một gần đây của chữ Nôm là do cách viết phức tạp của chữ Nôm và sự phổ biến của tiếng Hoa hiện đại ngoài việc chữ quốc ngữ là ngôn ngữ chính thức.

Nôm và quốc ngữ đều phong phú vì có nhiều âm hơn chữ Hán; nếu ghi được tất cả chữ thì Nôm bao gồm vốn từ vựng Hán Việt cùng thuần Việt và phân biệt được những chữ đồng âm mà quốc ngữ chỉ có thể phân biệt dựa vào ngữ cảnh hoặc từ ghép; ví dụ chữ “thanh 青” (màu xanh) khác với “thanh 清” (trong sạch) có bộ thủy. Nếu tập trung vào âm tiết, một chữ có thể vừa đa nghĩa Hán Việt lại vừa thuần Việt cùng với láy từ, đảo từ khiến tiếng Việt có cách diễn đạt dí dỏm tinh vi, biến ảo kỳ diệu đặc biệt trong thơ ca. Ngoài ra, ngữ pháp tiếng Việt linh động hơn tiếng Hán và la-tinh, ví dụ tính từ có thể đứng trước và sau danh từ, trạng từ có thể đứng trước và sau động từ, không chia thì động từ,… giúp lối hành văn thêm mềm dẻo, nhiều khi có vẻ sai ngữ pháp nhưng thực ra chưa hẳn sai. Ví dụ câu Hán tự “我今天到你的家吃饭” (ngã kim thiên đáo nhĩ đích gia ngật phạn) có thể viết thành nhiều cách theo tiếng Nôm “tôi hôm nay đến nhà bạn ăn cơm”, “tôi đến nhà bạn ăn cơm hôm nay”, “nhà bạn tôi đến ăn cơm hôm nay”; đặc biệt câu “nhà bạn tôi đến ăn cơm hôm nay” chệch cấu trúc chủ ngữ - tân ngữ nhưng không hẳn sai vì “(tại) nhà bạn” có thể hiểu là trạng ngữ chỉ nơi chốn. Một cách tương tự (Ham Chau, 2011), câu thơ của Nguyễn Du trong Truyện Kiều “giếng vàng đã rụng một vài lá ngô” gồm trạng ngữ chỉ nơi chốn “(tại) giếng vàng” và đảo ngữ của “một vài lá ngô đã rụng”. Nếu xem tiếng Việt gồm 2 cách viết: Nôm và quốc ngữ thì có sự bổ túc qua lại; chẳng hạn, Nôm giúp phân biệt những từ đồng âm và quốc ngữ có thể giúp phiên dịch tiếng nước ngoài. Ví dụ, thay vì phiên dịch một từ nước ngoài sang chữ Hán rồi một lần nữa chuyển sang âm Hán Việt, ta lại dùng quốc ngữ ghi âm chữ nước ngoài rồi sau đó có thể dùng chữ Nôm tượng hóa các âm này. Ví dụ, chúng ta có thể gặp cách viết “tiếng Phờ Răng” thay vì “tiếng Pháp” và hoàn toàn có thể viết Nôm tượng hình từ “Phờ Răng” như 𠱀𦝄. Sự bổ túc hay song hành Nôm – quốc ngữ mang đến nhiều điều bổ ích và thú vị, về ngôn ngữ học có thể giúp ta hiểu rõ tiếng Việt hơn. Ngoài ra, chữ Nôm vốn là ngôn ngữ tượng hình giúp bảo tồn phép chiết tự và thư pháp quý giá. Tóm lại, chữ Nôm vừa tượng hình vừa có chất tượng thanh, chính xác là gạch nối giữa chữ Hán và quốc ngữ nên cần khẳng định rằng giữ gìn chữ Nôm rất cần thiết.

**2. Vài suy nghĩ về giữ gìn di sản Hán Nôm**

Tôi nghĩ rằng trước tiên chúng ta cần chuẩn hóa chữ Nôm vì sự chuẩn hóa xác lập vị trí của chữ Nôm như một ngôn ngữ viết có tính song hành, vượt ra ngoài sự hoài niệm hoặc chỉ dành cho nghiên cứu. Việc chuẩn hóa chữ Nôm bao gồm:

1. Chúng ta quy định một cách viết cho những chữ đồng nghĩa và nên nghiên cứu kĩ những chữ nhập nhằng. Ngoài ra có thể quy định những cách viết khác nhau cho những âm khác nhau với trường hợp một chữ tượng hình cho nhiều âm sao cho hạn chế đoán nghĩa dựa vào từ ghép hoặc ngữ cảnh – trường hợp phổ biến của chữ quốc ngữ. Nhìn chung, cần đơn giản hóa chữ Nôm bao gồm việc thống nhất các quy tắc.
2. Chúng ta có thể tạo ra những chữ mới (phổ biến nhưng chưa có) theo các cách của tiền nhân sao cho càng nhiều như chữ quốc ngữ càng tốt. Tuy nhiên có lẽ không nên gượng tạo ra những chữ không cần thiết vì chúng ta không cần sự chuyển đổi qua lại giữa chữ Nôm và chữ quốc ngữ. Vấn đề quan trọng là giữ gìn di sản chữ Nôm cùng những trước tác và tìm hiểu tâm tư, nguyện vọng, đời sống tinh thần của tiền nhân.
3. Chúng ta có thể lập một bộ từ điển Hán Nôm chuẩn trên nền những bộ từ điển đã có.

Những nỗ lực tâm huyết và thành tựu đưa ký tự Nôm vào bảng mã Unicode của VNPF (http://www.nomfoundation.org) cùng sưu tầm, lưu trữ, số hóa, và lập thư mục sách Hán Nôm của Viện Nghiên cứu Hán Nôm (http://www.hannom.org.vn) hết sức quan trọng; trên nền tảng đó việc giữ gìn di sản Hán Nôm gồm 2 tác vụ chính:

1. Tiếp tục sưu tầm và số hóa tài liệu Hán Nôm trên khắp nước. Có lẽ chúng ta nên dùng nguyên tắc biên mục mới (MARC21 theo nguyên tắc FRBR) để biên mục tài liệu Hán Nôm đã số hóa.
2. Tiếp tục đưa những ký tự Nôm mới vào bảng mã Unicode. Những bộ gõ chữ Nôm có nhận dạng ký tự vẽ tay rất cần thiết.

Riêng đối với số hóa tài liệu Hán Nôm (Chu, 2004) (Nguyen, 2019), tôi nghĩ rằng việc quét chụp và xuất tài liệu thành tập tin ảnh rồi sau đó biên mục thì đơn giản và hiệu quả nhất. Việc nhận dạng tài liệu vật lý (như sách giấy dó, sắc phong, thác bản văn bia) sang tập tin văn bản bằng phần mềm nhận dạng quang học (OCR) sẽ không hiệu quả vì độ mờ cùng những hư hại của tài liệu vật lý. Một giải pháp khác bằng kết hợp giữa phần mềm nhận dạng tiếng nói (SR) cùng sự đọc lại tài liệu để chuyển tài liệu vật lý thành tập tin văn bản cũng không tối ưu vì giá trị của một số tài liệu như sắc phong còn nằm ở bố cục vật lý mà hình ảnh còn lưu lại được. Tuy nhiên chúng ta hãy khuyến khích sự nghiên cứu về OCR và SR vì 3 lý do sau:

1. Kêu gọi sự quan tâm từ trong và ngoài ngành Hán Nôm. Các chuyên gia đến từ những ngành khoa học khác góp phần mở rộng và phổ biến ngành Hán Nôm.
2. Làm phong phú và nhận thêm nguồn tài trợ với những đề tài khoa học.
3. Văn bản chú thích đính kèm một tập tin ảnh rất có giá trị, giúp người đọc hiểu nhanh nội dung tài liệu. Việc ghi lại văn bản sẽ chính xác nếu được chuyên gia thực hiện nhưng nhận dạng tự động (OCR và SR) mới có thể thực hiện với lượng tài liệu lớn.

Sau khi chuẩn hóa chữ Nôm, tôi đề xuất một cách tiếp cận mở rộng để số hóa, nhận dạng và biên mục tài liệu Hán Nôm.

1. Quét, chụp và xuất tài liệu thành tập tin ảnh (Nguyen, 2019). Trong bài nghiên cứu này, tập tin PDF được tạo từ tập tin ảnh được xem như tập tin ảnh và tập tin ảnh gồm một đến nhiều ảnh (Nguyen, 2019).
2. Tiền xử lý tập tin ảnh như nâng cao độ sắc nét, chuẩn khổ, đánh số trang. Tiền xử lý được xem như một bước riêng biệt vì chất lượng ảnh quan trọng đối với việc nhận dạng chữ viết.
3. Tập tin ảnh được xem như tài liệu đã số hóa (TLSH), được đưa vào cơ sở dữ liệu (CSDL) (Chu, 2004) và lập biên mục có gắn chú thích (CT).
4. Phần mềm OCR tự động truy xuất CSDL và nhận dạng TLSH thành tập tin văn bản được gọi là văn bản chú thích (VBCT). Một cách tiếp cận khác, chuyên gia sẽ đọc TLSH để phần mềm SR nhận dạng tiếng nói và chuyển TLSH thành VBCT. Phần mềm OCR hoặc SR cũng thực hiện liên kết CT với VBCT.
5. Chuyên gia theo định kỳ sẽ chỉnh sửa lại VBCT, điều này có nghĩa chấp nhận sự tồn tại những VBCT thiếu chính xác. Tất nhiên VBCT là văn bản Hán Nôm và được phân trang đồng nhất với TLSH.

Các bước 1, 2, 3 nhìn chung đều đã được thực hiện cùng với CSDL có sẵn. Như vậy, thao tác ở bước 4 được “gắn” thêm vào như một phần mềm tự động sẽ không ảnh hưởng đến tác vụ và kiến trúc hiện hành. Bước 5 tùy chọn và trong suốt nên cũng không gây bất kỳ ảnh hưởng tiêu cực. Khi CSDL được mở, nhiều chuyên gia khắp nơi có thể tham gia chỉnh sửa VBCT thông qua một giao diện web. Vấn đề là làm cách nào để gắn CT vào biểu mẫu biên mục. Giải pháp đơn giản là dùng 2 biểu mẫu MARC21 (LOC, 2020) cho TLSH và VBCT. Biểu mẫu của TLSH có trường 785 (succeeding entry) trỏ đến trường 130 (uniform title) của biểu mẫu VBCT trong khi biểu mẫu VBCT có trường 780 (preceding entry) trỏ đến trường 130 (uniform title) của biểu mẫu TLSH. Nói cách khác, ở đây phối hợp hai trường 785 và 780 tham chiếu lẫn nhau giữa hai biểu mẫu. Giải pháp này còn hỗ trợ sao lưu dự phòng vì thay đổi một trong hai biểu mẫu không ảnh hưởng đến biểu mẫu kia và mất một trong hai biểu mẫu không làm mất toàn bộ tài liệu.

Nhìn chung, giữ gìn di sản Hán Nôm rất cần nhiệt huyết và sự đồng lòng hợp tác của nhiều người; từ nhiệt huyết và đồng lòng sẽ có năng lượng và giải pháp sáng suốt và khi đó, chỉ còn một khoảng cách phụ thuộc vào thời gian và nguồn lực dẫn đến thành tựu mới – một CSDL Hán Nôm mở, chuẩn hóa và đầy đủ từ nguồn tài liệu khổng lồ hiện có (nguồn tài liệu này là thành tựu rất lớn mà chúng ta đã đạt được). Một giải pháp là xây dựng CSDL phân tán theo 5 bước được đề cập ở trên – gồm một CSDL chính và mỗi tổ chức trong và ngoài nước đều có một bản sao; cập nhật CSDL chính sẽ cập nhật các bản sao và ngược lại, cập nhật từng bản sao sẽ cập nhật CSDL chính. Mỗi tổ chức thu thập và số hóa tài liệu theo cách riêng với CSDL riêng nhưng vẫn bảo đảm người dùng truy cập vào nguồn tài liệu rộng lớn và nhất quán.

**Tham khảo**

Chu, T.-L. (2004). Số hoá để bảo quản và khai thác di sản Hán Nôm Việt nam: triển vọng và thách thức. *Hội nghị Quốc tế về chữ Nôm* (pp. 1-6). Hà Nội: Thư viện Quốc gia Việt Nam. Retrieved June 3, 2020, from https://bit.ly/2UOzmO2

Ham Chau. (2011, March 2). GS Nguyễn Tài Cẩn – Người soi rọi ngọn nguồn tiếng Việt. *Nhà nghiên cứu Hán Nôm*. Retrieved June 16, 2020, from https://ngnnghc.wordpress.com/tag/nha-nghien-c%E1%BB%A9u-han-nom/

LOC. (2020, May). *MARC 21 Format for Bibliographic Data*. (The Library of Congress) Retrieved June 14, 2020, from MARC Standard: http://www.loc.gov/marc/bibliographic

Nguyen, S. V. (2019, January 08). Quy trình và Kinh nghiệm số hóa nguồn tài liệu Hán Nôm tại Thư viện KHXH. *Tài liệu mở*. Retrieved from https://www.vansuforkid.com/2019/01/quy-trinh-va-kinh-nghiem-so-hoa-nguon.html

Tran, D. T. (2015, September). Nguồn gốc, lịch sử và cấu trúc của chữ Nôm từ bối cảnh văn hóa Đông Á. pp. 1-15. Retrieved from https://sealinguist.files.wordpress.com/2015/09/tran-2015.pdf

Wikipedia. (2020, May 19). *Chữ Nôm*. (Wikimedia Foundation) Retrieved June 16, 2020, from Wikipedia website: https://en.wikipedia.org/wiki/Ch%E1%BB%AF\_N%C3%B4m