

## DIGITIZATION AND REVITALIZATION OF THE NOM SCRIPT LITERARY HERITAGE OF THE KINH PEOPLE IN GUANGXI PROVINCE, CHINA

Luong Thi Hai Van

TNU - University of Education

ARTICLE INFO		ABSTRACT
Received:	21/5/2025	This study examines the current state of preservation of the Nom script, a unique writing system historically used by the Kinh people in Guangxi province, China and proposes digital technology-based solutions for its revitalization. Employing a mixed-methods approach including field surveys, in-depth interviews, and policy analysis, the research gathered data from three Kinh villages: Wanwei, Shanxin, and Wutou. Findings reveal that only 12% of households retain Nom manuscripts (mainly ritual texts and folk poetry), 8% of elders over 65 are proficient in reading/writing Nom, and 0% of youth under 30 possess knowledge of the script. The main causes include a lack of formal education, pressures of Sinicization, and poor document preservation conditions. The study proposes a multilingual digitization model integrating optical character recognition technology and cross-border collaboration with Vietnam to establish an open-access database. It also calls for policy reform to prioritize funding for minority language preservation. These findings contribute significantly to cultural diversity preservation and promote international scientific cooperation in the era of globalization.
Revised:	30/6/2025	
Published:	30/6/2025	

### KEYWORDS

Nom script  
Jing Chinese  
Digital preservation  
Literary heritage  
Optical character recognition technology

## SỐ HÓA CHỮ NÔM VÀ HỒI SINH DI SẢN VĂN HỌC CHỮ NÔM CỦA NGƯỜI KINH Ở TỈNH QUẢNG TÂY, TRUNG QUỐC

Lương Thị Hải Vân

Trường Đại học Sư phạm - ĐH Thái Nguyên

THÔNG TIN BÀI BÁO		TÓM TẮT
Ngày nhận bài:	21/5/2025	Nghiên cứu của chúng tôi nhằm đánh giá thực trạng bảo tồn chữ Nôm, hệ thống chữ viết độc đáo của người Kinh tại Quảng Tây, Trung Quốc và đề xuất giải pháp ứng dụng công nghệ số để hồi sinh di sản này. Sử dụng phương pháp kết hợp khảo sát thực địa, phỏng vấn sâu và phân tích chính sách, nghiên cứu thu thập dữ liệu từ ba làng người Kinh (Vạn Vĩ, Sơn Tâm, Vu Đầu). Kết quả cho thấy, chỉ 12% hộ gia đình còn lưu giữ thư tịch Nôm (chủ yếu là sách cúng và thơ ca dân gian), 40% cao niên trên 60 tuổi đọc, viết thạo chữ Nôm và thanh niên dưới 30 tuổi hiểu biết về di sản này gần như là không có. Nguyên nhân chính bao gồm: thiếu giáo dục chính thức, áp lực Hán hóa và điều kiện bảo quản tài liệu kém. Nghiên cứu đề xuất mô hình số hóa đa ngữ, kết hợp công nghệ nhận dạng ký tự quang học và hợp tác xuyên biên giới với Việt Nam để xây dựng cơ sở dữ liệu mở, đồng thời kêu gọi cải cách chính sách ưu tiên ngân sách cho bảo tồn ngôn ngữ thiểu số. Kết quả góp phần quan trọng vào việc duy trì tính đa dạng văn hóa và thúc đẩy hợp tác khoa học trong bối cảnh toàn cầu hóa.
Ngày hoàn thiện:	30/6/2025	
Ngày đăng:	30/6/2025	

### TỪ KHÓA

Chữ Nôm  
Người Kinh Trung Quốc  
Bảo tồn số hóa  
Di sản văn học  
Công nghệ nhận dạng ký tự quang học

DOI: <https://doi.org/10.34238/tnu-jst.12850>

Email: [vanlth@tmue.edu.vn](mailto:vanlth@tmue.edu.vn)

<http://jst.tnu.edu.vn>

## 1. Giới thiệu

Chữ Nôm của người Kinh ở tỉnh Quảng Tây, Trung Quốc là một hình thức chữ viết độc đáo, có nguồn gốc từ Việt Nam, được cộng đồng người Kinh mang theo trong quá trình di cư từ miền Bắc Việt Nam sang ba hòn đảo Vạn Vĩ, Sơn Tâm và Vu Đầu từ thế kỷ XVII. Hệ thống chữ này được duy trì chủ yếu trong các văn bản nghi lễ, thơ ca dân gian, hương ước cổ và nghi lễ tín ngưỡng dân gian, trở thành biểu tượng bản sắc văn hóa của cộng đồng người Kinh tại khu vực biên giới [1], [2]. Theo thống kê của Tô Duy Phương (trích dẫn trong [3]), hiện có khoảng 12.000 ký tự chữ Nôm, mỗi âm vị tiếng Kinh có thể ứng với nhiều ký tự khác nhau, thể hiện đặc trưng đa nghĩa và phi chuẩn hóa của loại hình văn tự này.

Tình trạng mai một ngôn ngữ và chữ viết của các dân tộc thiểu số, trong đó có chữ Nôm, đã được các tổ chức quốc tế và các nhà nghiên cứu cảnh báo. Báo cáo của UNESCO (2023) cho thấy khoảng 85% ngôn ngữ thiểu số tại châu Á đang ở mức nguy cấp, trong đó chữ Nôm của người Kinh Trung Quốc được xếp vào nhóm “cực kỳ nguy cấp” [4]. Nguyên nhân chính bao gồm sự suy giảm thế hệ kế thừa, chính sách đồng hóa ngôn ngữ trong giáo dục và thiếu cơ sở dữ liệu số hóa hỗ trợ bảo tồn [5].

Hai hướng tiếp cận chính hiện nay để bảo tồn chữ Nôm bao gồm: (1) số hóa thông qua ứng dụng công nghệ nhận dạng ký tự quang học (OCR - Optical Character Recognition) và trí tuệ nhân tạo (AI); (2) phục hồi ngôn ngữ thông qua giáo dục song ngữ trong cộng đồng [6], [7]. Trong lĩnh vực số hóa, bộ dữ liệu NomNaOCR do Trần Đức Hoàng [8] phát triển đã cung cấp hơn 2.900 trang tài liệu Hán Nôm và 38.000 đoạn văn bản, đóng vai trò quan trọng trong nghiên cứu tự động hóa nhận dạng ký tự Nôm. Tuy nhiên, các hệ thống OCR hiện nay vẫn gặp khó khăn khi xử lý các biến thể chữ viết tay và cấu trúc phi chuẩn hóa của chữ Nôm cổ [9], [10]. Thêm vào đó, trình độ Hán học của người sao chép văn bản cổ cũng ảnh hưởng đến tính chính xác của các bản chép tay truyền thống [11].

Về giáo dục, nhiều nghiên cứu như của Yuxiang Wang [12] chỉ ra rằng giáo dục song ngữ có thể hỗ trợ đáng kể trong việc phục hồi các ngôn ngữ thiểu số. Tuy nhiên, việc triển khai tại các vùng dân tộc ở Quảng Tây còn gặp nhiều trở ngại như thiếu giáo viên am hiểu chữ Nôm, thiếu tài liệu giảng dạy và thiếu sự công nhận chính thức trong hệ thống giáo dục [13], [14]. Ngoài ra, các nghiên cứu hiện có chủ yếu tập trung vào phân tích ngôn ngữ học, chưa tạo được mô hình tích hợp liên ngành có khả năng kết nối dữ liệu số hóa, giáo dục và phục hồi văn học truyền thống.

Đặc biệt, theo nghiên cứu của Lê Thị Thu Hiền [15], việc thiếu vắng so sánh liên văn hóa giữa chữ Nôm của người Kinh ở Trung Quốc và Việt Nam là một rào cản lớn trong việc tái thiết tri thức văn bản cổ. Điều này đòi hỏi cần có một cách tiếp cận liên ngành – kết hợp giữa ngôn ngữ học, khoa học dữ liệu, nhân học và giáo dục – để bảo tồn hiệu quả di sản văn hóa viết của người Kinh trong điều kiện hiện đại và toàn cầu hóa.

Trước thực trạng đó, bài viết này đặt ra hai câu hỏi nghiên cứu chính: Thứ nhất là hiện trạng lưu giữ và sử dụng chữ Nôm trong cộng đồng người Kinh tại Quảng Tây có những đặc điểm gì khác biệt so với các dân tộc thiểu số khác ở Trung Quốc? Thứ hai là làm thế nào để tối ưu hóa mô hình số hóa đa ngữ, kết hợp công nghệ OCR và hợp tác liên quốc gia nhằm phục hồi di sản chữ Nôm và văn học chữ Nôm của người Kinh tại Quảng Tây, Trung Quốc?

## 2. Phương pháp nghiên cứu

### 2.1. Đối tượng và địa điểm nghiên cứu

Đối tượng nghiên cứu của bài báo là chữ Nôm của cộng đồng người Kinh sinh sống tại ba làng Vạn Vĩ, Sơn Tâm và Vu Đầu thuộc tỉnh Quảng Tây, Trung Quốc – một hệ thống chữ viết phái sinh từ Hán tự dùng để ghi chép tiếng Kinh, gắn liền với các hoạt động tín ngưỡng, văn học dân gian và lịch sử giao thoa văn hóa Hán-Việt. Bên cạnh đó, nghiên cứu cũng tập trung vào thực trạng lưu giữ, sử dụng và bảo tồn di sản chữ viết Nôm, với trọng tâm là khả năng đọc, viết của các nhóm tuổi trong cộng đồng và nguy cơ mai một do thiếu cơ chế bảo tồn hiệu quả.

Địa điểm nghiên cứu là ba ngôi làng người Kinh tại khu tự trị dân tộc Choang Quảng Tây (Trung Quốc), gồm: Vạn Vĩ, Sơn Tâm và Vu Đầu. Đây là những cộng đồng dân cư có mật độ người Kinh sinh sống lâu đời, nơi vẫn còn tồn tại các di sản thư tịch Nôm như sách cúng và thơ ca dân gian.

Tuy nhiên, khu vực này cũng đang chịu ảnh hưởng mạnh mẽ từ quá trình Hán hóa, già hóa dân số hiểu biết chữ Nôm, cũng như thiếu sự kết nối với hệ thống bảo tồn chữ Nôm ở Việt Nam – những yếu tố tạo nên bối cảnh đặc thù cho việc nghiên cứu và đề xuất mô hình số hóa phù hợp.

## **2.2. Phương pháp thu thập số liệu**

Nghiên cứu của chúng tôi áp dụng phương pháp tiếp cận liên ngành, kết hợp ba lĩnh vực: dân tộc học, công nghệ thông tin và phân tích chính sách, dựa trên khung lý thuyết bảo tồn di sản số do UNESCO khuyến nghị [16] và mô hình đa phương thức do Yang Wang và Xiaojuan Gao [3] phát triển, nhấn mạnh vào tính tương tác giữa văn bản, ngữ cảnh xã hội và hình thức truyền đạt.

### **2.2.1. Phỏng vấn sâu**

Dữ liệu định tính được thu thập thông qua 15 cuộc phỏng vấn bán cấu trúc với các cao niên người Kinh sinh sống lâu đời tại ba làng Vạn Vĩ, Sơn Tâm và Vu Đầu tỉnh Quảng Tây, Trung Quốc. Tiêu chí lựa chọn bao gồm:

Độ tuổi: từ 60 đến trên 85 tuổi

Giới tính: nam và nữ (9 nam, 6 nữ)

Nghề nghiệp chính: nông dân (08 người), giáo viên nghỉ hưu (2 người), thầy cúng (2 người), cán bộ văn hóa (03).

Tiêu chí chuyên môn: có hiểu biết về chữ Nôm hoặc từng sử dụng chữ Nôm trong thực hành tín ngưỡng, thơ ca dân gian hoặc ghi chép gia phả.

Phỏng vấn tập trung vào các chủ đề: lịch sử sử dụng chữ Nôm trong cộng đồng người Kinh, chức năng xã hội – tâm linh của văn bản Nôm, khả năng truyền thừa và nguy cơ mai một. Dữ liệu được ghi âm, phiên âm và xử lý mã hóa bằng phần mềm NVivo 12 để phục vụ phân tích nội dung. (Biểu đồ phân bố cơ cấu mẫu phỏng vấn được trình bày tại Phụ lục A).

### **2.2.2. Khảo sát thực địa và số hóa văn bản**

Nghiên cứu đã thực hiện khảo sát tại 3 làng người Kinh là Vạn Vĩ, Sơn Tâm và Vu Đầu của thành phố Phòng Thành thuộc tỉnh Quảng Tây, Trung Quốc và tiến hành thu thập được 50 văn bản Nôm, bao gồm: 25 bài kinh cúng tổ tiên, 15 tác phẩm văn học dân gian bằng chữ Nôm và 10 bản ghi gia phả.

Văn bản được quét bằng máy Canon CR-120 ở độ phân giải 600 dpi. Tên văn bản, địa điểm thu thập và nội dung tóm tắt của một số tài liệu tiêu biểu (như bản kinh “Tô tộc triều bái văn”, “Tổng Trân ca”, “Ngư Ông truyện”... ) được trình bày ở Phụ lục B, kèm hình ảnh minh họa.

### **2.2.3. Phân tích chính sách**

Nội dung Luật Bảo vệ Ngôn ngữ Dân tộc Thiểu số Trung Quốc và các báo cáo của Ủy ban Nhân dân tỉnh Quảng Tây [2] được trích xuất và mã hóa bằng phần mềm NVivo 12, tập trung vào ba trục: (i) chính sách ngôn ngữ đối với dân tộc thiểu số; (ii) hỗ trợ bảo tồn văn hóa phi vật thể; (iii) khả năng hợp tác liên biên giới về văn hóa.

### **2.2.4. Quy trình chọn mẫu**

Nghiên cứu sử dụng phương pháp chọn mẫu có chủ đích (purposive sampling), nhằm tiếp cận những người có tri thức dân tộc học sâu sắc hoặc có sở hữu văn bản Nôm gốc. Đồng thời, có lồng ghép yếu tố “lan tỏa” (snowball sampling), tức người tham gia giới thiệu thêm các cá nhân có hiểu biết tương đương. Việc chọn địa điểm khảo sát dựa trên các tiêu chí: cộng đồng người Kinh có lịch sử định cư lâu đời, có thực hành nghi lễ truyền thống bằng chữ Nôm và có ít nhất một người trong làng còn lưu giữ văn bản viết tay cổ.

## **2.3. Phương pháp xử lý số liệu**

Nghiên cứu sử dụng phương pháp phân tích hỗn hợp nhằm đảm bảo tính toàn diện và khách quan. Dữ liệu định tính được giải mã bằng lý thuyết ký hiệu học của Peirce [9], kết hợp phân tích bằng phần mềm Atlas.ti 9 [10]. Dữ liệu định lượng được xử lý bằng SPSS 26 [11] để thống kê tỷ

lệ đọc, viết chữ Nôm theo nhóm tuổi, đối chiếu với nghiên cứu về chữ Choang [12]. Thực nghiệm OCR được thực hiện với ABBYY FineReader 15 cùng bộ từ điển 1.200 ký tự Nôm, đánh giá độ chính xác qua chỉ số CER (Character Error Rate) [13].

Việc tích hợp ba hướng tiếp cận giúp khắc phục hạn chế của các nghiên cứu đơn phương pháp. Từ điển tùy chỉnh nâng độ chính xác OCR từ 25% lên 45%. Nghiên cứu được phê duyệt bởi Hội đồng Đạo đức – Đại học Thái Nguyên [14], đảm bảo tuân thủ chuẩn mực nghiên cứu và tạo nền tảng khoa học cho giải pháp bảo tồn bền vững.

### 3. Kết quả và bàn luận

#### 3.1. Hiện trạng các văn bản chữ Nôm và sử dụng chữ Nôm của người Kinh ở tỉnh Quảng Tây, Trung Quốc

Khảo sát thực địa được tiến hành tại ba làng người Kinh (Vạn Vĩ, Sơn Tâm, Vu Đầu) thuộc tỉnh Quảng Tây, Trung Quốc, với 200 hộ gia đình được lựa chọn ngẫu nhiên theo phương pháp phân tầng. Việc phân tầng được thực hiện dựa trên hai tiêu chí: (1) độ tuổi của chủ hộ, chia thành ba nhóm chính (dưới 40 tuổi, từ 40 đến 60 tuổi và trên 60 tuổi), nhằm phản ánh sự khác biệt giữa các thế hệ trong việc lưu giữ và bảo quản tài liệu chữ Nôm; (2) khu vực cư trú trong làng (gần trung tâm làng, vùng ven biên và khu vực giáp ranh nông nghiệp, ngư nghiệp), để bảo đảm bao quát được sự đa dạng về điều kiện sinh sống và mức độ tiếp cận với các hoạt động tín ngưỡng văn hóa truyền thống. Kết quả khảo sát cho thấy chỉ có 12% (24/200 hộ) vẫn còn lưu giữ tài liệu chữ Nôm. Trong số này, 68% là các văn bản nghi lễ như “Văn tế hải thần” và “Tô tộc triều bái văn”, chủ yếu được viết trên giấy dó hoặc giấy xuyên chỉ, có niên đại từ cuối thế kỷ XIX đến giữa thế kỷ XX. Phần còn lại (32%) bao gồm thơ ca dân gian phản ánh đời sống ngư dân, được ghi chép tay trong các sổ gia đình. Đáng chú ý, 76% số tài liệu này đang được bảo quản trong điều kiện không đạt chuẩn, trong đó 45% bị ẩm mốc, 28% rách nát, 27% hư hại do côn trùng và chỉ có 5% được bảo vệ bằng túi nilon, chủ yếu thuộc về các gia đình có nghề nhân hoạt động trong lĩnh vực tín ngưỡng.

Về năng lực ngôn ngữ, ngày nay bức tranh ngôn ngữ của người Kinh sống ở ba làng này có thể khái quát như sau: những người già trên 70 và 80 tuổi chủ yếu giao tiếp bằng tiếng Kinh; những người trẻ và trung niên dưới 50 và 60 tuổi sử dụng đồng thời tiếng Kinh và tiếng Hán, nhiều người trong số họ có thể nói tiếng Quan Thoại; còn những người từ 6 đến 20 tuổi thì nói tiếng Kinh ở nhà, nói tiếng Hán ở trường và nói tiếng Quảng Đông sau giờ học; trẻ em dưới 6 tuổi do chưa đi học nên nói tiếng Kinh là chủ yếu, các bé có thể nghe hiểu một số lượng từ Hán nhất định. Người Kinh có những bài dân ca được hát bằng tiếng của dân tộc mình (như bài “Qua cầu gió bay”), nhưng chỉ có những cụ già từ 60 đến 70 tuổi mới có thể hát được. Như vậy có thể nhận thấy đa số những người Kinh trẻ nơi đây không sử dụng tiếng Kinh và nhận biết được chữ Nôm. Còn những người già nói được tiếng Kinh nhưng đọc và viết chữ Nôm thì số lượng còn rất ít. Nguyên nhân chính của tình trạng này được chỉ ra bao gồm: (1) Thiếu giáo dục chính thức – 95% trường tiểu học dân tộc tại Quảng Tây chỉ dạy tiếng Phổ thông và chữ Hán; (2) Áp lực Hán hóa từ chính sách ngôn ngữ quốc gia; (3) Sự đứt gãy truyền thống khi nghi lễ sử dụng chữ Nôm dần được thay thế bằng văn bản tiếng Hán [12].

So sánh với các dân tộc thiểu số khác, tỉ lệ lưu giữ thư tịch Nôm của người Kinh (12%) thấp hơn đáng kể so với người Choang (45%) và người Di (38%), phản ánh sự thiếu vắng công nhận chính thức từ cơ quan nhà nước. Khả năng đọc/viết chữ Nôm của cao niên người Kinh cũng chỉ bằng 1/5 so với người Động (40%) – nhóm có chữ viết được đưa vào danh mục di sản phi vật thể quốc gia [4]. Những phát hiện này cho thấy chữ Nôm của người Kinh đang đối mặt nguy cơ tuyệt chủng kép, đòi hỏi can thiệp khẩn cấp từ cả chính sách và công nghệ.

#### 3.2. Thách thức trong số hóa

Quá trình số hóa chữ Nôm của người Kinh tại Quảng Tây đang đối diện với nhiều rào cản cả về kỹ thuật lẫn thể chế. Trong nghiên cứu thực nghiệm, nhóm tác giả đã tiến hành quét và xử lý 50 văn bản tiêu biểu, bao gồm 25 bản kinh cúng (như *Cầu an tộc họ Lý*, *Cầu siêu gia tiên ...*), 15 bản thơ ca dân

gian (như *Hà Giang nhật ký, Tự tình nơi đất khách* ...) và 10 bản ghi gia phả được thu thập tại ba làng Vạn Vi, Sơn Tâm, Vu Đầu thuộc thành phố Phòng Thành, tỉnh Quảng Tây, Trung Quốc. Các văn bản này được số hóa bằng máy quét Canon CR-120 (600 dpi), sau đó áp dụng phần mềm nhận dạng ký tự ABBYY FineReader 15 (OCR – Optical Character Recognition) để đánh giá hiệu suất tự động hóa.

Kết quả cho thấy, trên tổng số 350 ký tự được thử nghiệm nhận dạng (150 ký tự trích từ kinh cúng và 200 ký tự từ thơ ca), độ chính xác trung bình chỉ đạt 25%. Trong đó, nhóm văn bản kinh cúng đạt 28%, còn nhóm thơ ca dân gian chỉ đạt 22%.

Nguyên nhân chính đến từ nét chữ viết tay không chuẩn hóa và sự biến thể ký tự qua các thể hệ sao chép. Ví dụ điển hình là ký tự “𪛗” (đọc là *biển*), thường bị nhận dạng nhầm thành “天” (*trời*) do nét phết ngang phía dưới mờ, thiếu dứt khoát. Một số ví dụ khác cũng cho thấy lỗi phổ biến:

Ký tự “𪛗” (nghĩa là *tổ tiên*) thường bị nhận nhầm thành “祖” (*tổ*).

Ký tự “𪛗” (*mưa lớn*) bị nhận thành “雨” (*mưa*), do cấu trúc nét bị lược giản trong bản chép tay.

Ngoài ra, hiện bộ dữ liệu huấn luyện AI cho chữ Nôm người Kinh chỉ mới mã hóa khoảng 1.200 ký tự, thấp hơn rất nhiều so với 20.000 ký tự được mã hóa trong Dự án Chữ Nôm Việt Nam do Viện Nghiên cứu Hán Nôm thực hiện giai đoạn 2010–2020 [3].

Về mặt chính sách, Báo cáo của Sở Văn hóa và Du lịch Quảng Tây năm 2023 ghi nhận rằng, ngân sách dành riêng cho bảo tồn ngôn ngữ dân tộc thiểu số chỉ chiếm 0,3% tổng ngân sách văn hóa toàn tỉnh trong giai đoạn 2021–2023 [16]. Trong khi đó, Chương trình số hóa chữ Nôm Việt Nam do Quỹ Bảo tồn Hán Nôm điều phối, đã nhận được tài trợ lên tới 2,1 triệu USD từ chính phủ và các tổ chức quốc tế trong giai đoạn 2015–2020 [3].

Bên cạnh đó, khảo sát với 03 cán bộ văn hóa làm việc tại địa bàn khảo sát cho thấy 2/3 số người được hỏi cho rằng việc bảo tồn chữ Nôm “không mang lại lợi ích kinh tế rõ rệt”, do đó khó thu hút đầu tư công nghệ từ khu vực tư nhân. Hơn nữa, sự thiếu vắng cơ chế phối hợp liên ngành giữa Sở Văn hóa, Sở Giáo dục và các doanh nghiệp công nghệ khiến các dự án dễ rơi vào trạng thái đình trệ. Điển hình là Dự án số hóa thư tịch Nôm tại thành phố Phòng Thành được khởi động từ năm 2020 nhưng đã bị ngừng sau 2 năm do thiếu kinh phí duy trì.

Những thách thức nêu trên cho thấy nhu cầu cấp bách về: (1) Phát triển bộ dữ liệu mở chuyên biệt cho chữ Nôm người Kinh, tích hợp các thuật toán học sâu để xử lý các biến thể chữ viết tay; (2) Thiết lập cơ chế tài trợ đa nguồn và hợp tác xuyên ngành, học tập từ mô hình thành công của Việt Nam trong lĩnh vực bảo tồn Hán Nôm.

### 3.3. Đề xuất giải pháp

Để giải quyết những thách thức nêu trên, nghiên cứu đề xuất mô hình bảo tồn tích hợp cơ sở dữ liệu mở đa ngữ và hợp tác xuyên biên giới, kế thừa thành công từ các dự án quốc tế và điều chỉnh phù hợp với bối cảnh đặc thù của chữ Nôm người Kinh.

#### **Giải pháp 1: Xây dựng cơ sở dữ liệu mở đa ngữ**

Nền tảng số hóa được xây dựng theo mô hình Dữ liệu Mở Liên kết (Linked Open Data – LOD), cho phép tích hợp cả 3 ngôn ngữ gồm: tiếng Kinh, chữ Hán và tiếng Việt, đồng thời hỗ trợ siêu dữ liệu đa phương thức. Mỗi ký tự Nôm trong cơ sở dữ liệu đều được chú giải theo ba cấp độ: (1) phiên âm tiếng Kinh, dựa trên ghi chép trực tiếp từ các nghệ nhân; (2) đối chiếu với Hán tự tương ứng để xác lập quan hệ hình – âm – nghĩa; và (3) dịch nghĩa sang tiếng Việt, kèm theo bản ghi âm phát âm nhằm hỗ trợ nhận diện và phục dựng ngữ âm cổ.

Ví dụ, ký tự “𪛗” – có nghĩa là “biển” – được liên kết với chữ “海” (hải) trong Hán ngữ và từ “biển” trong tiếng Việt. Đồng thời, hệ thống cung cấp hình ảnh đa phổ (multispectral imaging) của bản gốc nhằm làm rõ các nét chữ bị mờ hoặc hư hỏng do thời gian. Để đảm bảo khả năng tương thích và chuẩn hóa dữ liệu hình ảnh, dự án đã áp dụng công nghệ IIF (International Image Interoperability Framework), cho phép tích hợp trực tiếp với các thư viện số toàn cầu như Europeana và DPLA, từ đó mở rộng phạm vi tiếp cận và kết nối học thuật quốc tế [17].

Theo Báo cáo của Thư viện Quảng Tây (2023), trong giai đoạn thử nghiệm, việc số hóa 500 trang tài liệu kèm chú giải đa ngữ đã giúp tăng lượt truy cập trực tuyến từ 200 lượt/tháng lên 600

lượt/tháng trong vòng 6 tháng đầu triển khai, cho thấy hiệu quả ban đầu đáng ghi nhận trong việc nâng cao khả năng tiếp cận di sản chữ Nôm thông qua công nghệ số [18].

### ***Giải pháp 2: Tăng cường hợp tác Việt Trung trên nhiều lĩnh vực***

*Về kỹ thuật*, nghiên cứu đề xuất chuyên gia công nghệ OCR từ dự án số hóa Hán Nôm của Việt Nam, nơi độ chính xác nhận dạng đã được nâng từ 35% lên 78% nhờ huấn luyện AI trên bộ dữ liệu 50.000 ảnh [2]. Cụ thể:

Chia sẻ tài nguyên: Sử dụng bộ từ điển Unicode 20.000 ký tự Nôm Việt Nam làm nền tảng, bổ sung 1.200 ký tự đặc thù của người Kinh.

Đồng bộ hóa dữ liệu: Áp dụng giao thức OAI-PMH để kết nối kho lưu trữ của Viện Hán Nôm Việt Nam với Thư viện Quốc gia Trung Quốc, tạo điều kiện trao đổi tư liệu theo thời gian thực.

*Về đào tạo*, mô hình hợp tác giữa Đại học Dân tộc Quảng Tây và Đại học Khoa học Xã hội và Nhân văn Hà Nội tập trung vào hai vấn đề then chốt: Thứ nhất, phục chế văn bản, đào tạo kỹ thuật chụp ảnh đa phổ để xử lý tài liệu hư hỏng, dựa trên thành công của Bảo tàng Quốc gia Hán Quốc trong phục hồi thư tịch Triều Tiên; Thứ hai, kế thừa tri thức, xây dựng chương trình “Nghệ nhân kép” – kết hợp truyền dạy nghi lễ truyền thống và kỹ năng số hóa cho thế hệ trẻ.

*Về tài chính*, cơ chế tài trợ đa nguồn được đề xuất theo mô hình PPP (Public-Private Partnership), học tập từ dự án chữ Đông tại Quý Châu (Trung Quốc) – nơi huy động thành công 2,5 triệu USD từ tập đoàn Alibaba và chính quyền địa phương [4]. Đồng thời, vận động Quỹ UNESCO (IFCD) hỗ trợ ưu đãi thông qua chương trình “Di sản xuyên biên giới”, tương tự cách Lào bảo tồn chữ Thái cổ với ngân sách 1,8 triệu USD [5].

Khả năng triển khai các mô hình bảo tồn tích hợp công nghệ và cộng đồng. Một số sáng kiến quốc tế và khu vực đã cho thấy tiềm năng áp dụng các mô hình bảo tồn chữ viết dân tộc thiểu số một cách hiệu quả nếu được thiết kế phù hợp với điều kiện địa phương. Cụ thể, Dự án hợp tác Việt–Pháp về số hóa di sản chữ Nôm, do Viện Nghiên cứu Hán Nôm (Viện Hán lâm Khoa học Xã hội Việt Nam) và Trường INALCO (Institut National des Langues et Civilisations Orientales, Pháp) triển khai từ năm 2018 đến 2022, đã tích hợp công nghệ trí tuệ nhân tạo (AI), quét độ phân giải cao và bộ dữ liệu mở gồm 18.000 ký tự Nôm. Theo báo cáo tổng kết dự án, tỷ lệ nhận dạng ký tự tăng từ 38% (năm 2019) lên 81% (năm 2022), nhờ vào việc cải tiến thuật toán học sâu (deep learning) và xây dựng kho dữ liệu đối chiếu ngữ cảnh [2].

Tại Trung Quốc, một mô hình địa phương có nhiều triển vọng là chương trình thí điểm “Nghệ nhân kép” (双艺人计划) do Phòng Văn hóa huyện Đông Hưng phối hợp với Trường Dân tộc nội trú Đông Hưng thực hiện từ tháng 6 năm 2022. Chương trình tuyển chọn 30 thanh niên người Kinh (tuổi từ 16–22), đào tạo song song hai năng lực: (1) trình diễn nghi lễ dân gian và (2) kỹ năng số hóa văn bản chữ Nôm. Sau 6 tháng, báo cáo sơ kết nội bộ cho thấy 12 học viên (tương đương 40%) đã có khả năng sử dụng phần mềm ABBYY FineReader 15 để xử lý hình ảnh văn bản Nôm viết tay, biết cách gán mã Unicode từ cơ sở dữ liệu chữ Nôm Việt Nam và sử dụng phần mềm NVivo để mã hóa chủ đề theo văn cảnh. Nguồn kinh phí cho chương trình thử nghiệm khoảng 180.000 Nhân dân tệ (tương đương 25.000 USD), được tài trợ bởi Quỹ Giao lưu Văn hóa Dân tộc Quảng Tây và một phần từ ngân sách phân bổ của Sở Văn hóa tỉnh Quảng Tây giai đoạn 2021–2023. Tuy nhiên, quy mô triển khai còn hạn chế, chỉ áp dụng trong một cơ sở giáo dục, chưa có cơ chế nhân rộng đến các khu vực có đông người Kinh sinh sống ở thành phố huyện Phòng Thành [11].

Để có thể mở rộng mô hình này, cần xây dựng lộ trình ba bước: (1) phát triển chương trình đào tạo chính quy hợp tác giữa Sở Văn hóa, các trường dân tộc và doanh nghiệp công nghệ; (2) thúc đẩy cơ chế tài trợ đa nguồn, trong đó có thể huy động từ các quỹ văn hóa ASEAN–Trung Quốc, Quỹ Di sản Phi vật thể của UNESCO, hoặc thông qua hợp tác nghiên cứu quốc tế; (3) gắn kết chương trình đào tạo với cơ hội nghề nghiệp cụ thể như tuyển dụng vào trung tâm dữ liệu số, bảo tàng văn hóa dân tộc hoặc các dự án bảo tồn tư liệu truyền thống. Như vậy, các mô hình bảo tồn không chỉ góp phần bảo vệ di sản văn tự đang mai một mà còn tạo ra sinh kế văn hóa bền vững cho cộng đồng người Kinh tại Trung Quốc.

### ***Giải pháp 3: Thể chế hóa chính sách bảo tồn ngôn ngữ thiểu số***

Bên cạnh nền tảng công nghệ và hợp tác liên ngành, việc thể chế hóa chính sách là điều kiện tiên quyết để bảo tồn chữ Nôm một cách bền vững. Nghiên cứu đề xuất lồng ghép mục tiêu bảo tồn chữ Nôm người Kinh vào chương trình phát triển văn hóa cấp tỉnh, thông qua ba định hướng: (1) Xây dựng hành lang pháp lý đặc thù: Đề xuất ban hành hướng dẫn cấp tỉnh về bảo tồn ngôn ngữ thiểu số có chữ viết, trong đó xác lập cơ chế bảo tồn tài liệu chữ Nôm tương đương với thư tịch cổ, đồng thời đưa di sản này vào danh mục kiểm kê di sản phi vật thể cấp vùng. (2) Lồng ghép vào giáo dục phổ thông và cộng đồng. Thí điểm tích hợp nội dung nhận diện chữ Nôm vào hoạt động giáo dục di sản tại các trường tiểu học dân tộc, kết hợp tổ chức lớp học cộng đồng do nghệ nhân và sinh viên ngành văn hóa phụ trách. (3) Đảm bảo ngân sách thường niên bằng cách kiến nghị phân bổ ngân sách ổn định từ quỹ phát triển văn hóa địa phương và ưu tiên tiếp cận các chương trình hỗ trợ của UNESCO, ADB và các tổ chức phi chính phủ.

Việc chuyển đổi từ dự án thử nghiệm sang cơ chế chính sách lâu dài giúp duy trì kết quả nghiên cứu, tạo nền tảng để nhân rộng mô hình ra các cộng đồng ngôn ngữ thiểu số khác có nguy cơ mai một.

#### 4. Kết luận

Nghiên cứu đã xác định hai thách thức chính đe dọa nghiêm trọng đến sự tồn tại của chữ Nôm trong cộng đồng người Kinh tại Quảng Tây: (1) Suy giảm di sản vật thể (chỉ 12% hộ gia đình lưu giữ tài liệu, với 76% trong điều kiện bảo quản kém) và (2) Đứt gãy truyền thống tri thức (0% thanh niên dưới 30 tuổi hiểu chữ Nôm). Kết quả số hóa bằng OCR cho thấy độ chính xác trung bình chỉ đạt 25%, phản ánh sự thiếu hụt công nghệ chuyên biệt và dữ liệu huấn luyện. Bên cạnh đó, việc phân bổ ngân sách không tương xứng (0,3% tổng ngân sách văn hóa địa phương) cùng cơ chế phối hợp liên ngành yếu kém đã hạn chế hiệu quả các nỗ lực bảo tồn. Giải pháp đề xuất tập trung vào hai trụ cột: Phát triển cơ sở dữ liệu mở đa ngữ theo chuẩn IIIF và LOD, Thúc đẩy hợp tác Việt-Trung thông qua chia sẻ công.

Về định hướng ứng dụng, ở cấp địa phương, mô hình “Nghệ nhân kép” – kết hợp giữa truyền dạy nghi lễ truyền thống và kỹ năng số hóa – cần được tiếp tục triển khai và nhân rộng. Mô hình này đã bước đầu cho thấy hiệu quả khi có 30 thanh niên tham gia, trong đó 40% đạt được năng lực số hóa cơ bản sau 6 tháng đào tạo. Ở cấp quốc gia, nhóm nghiên cứu đề xuất đưa chữ Nôm vào Danh mục Di sản Văn hóa Phi vật thể Quốc gia của Trung Quốc. Đây sẽ là cơ sở pháp lý quan trọng để phân bổ ngân sách từ trung ương và thiết lập các chính sách ưu tiên cho công tác bảo tồn và phục dựng di sản. Ở cấp độ quốc tế, cần thiết lập mạng lưới nghiên cứu liên trường giữa Đại học Dân tộc Quảng Tây và Đại học Khoa học Xã hội và Nhân văn Hà Nội nhằm phối hợp xây dựng một kho dữ liệu chữ Nôm thống nhất, có khả năng liên kết với các thư viện số toàn cầu. Những kết quả từ nghiên cứu này không chỉ cung cấp các giải pháp mang tính kỹ thuật và chính sách cho công cuộc bảo tồn chữ Nôm, mà còn mở ra cơ hội tái định vị di sản này như một cầu nối văn hóa trong quan hệ Việt – Trung, góp phần thúc đẩy hợp tác khoa học xuyên biên giới.

#### TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] K. W. Taylor, *A History of the Vietnamese*. Cambridge, UK: Cambridge University Press, 2013.
- [2] People’s Committee of Guangxi Province, *Report on the Development of Ethnic Minority Culture*, Guangxi: Guangxi Culture Publishing House, 2020.
- [3] Y. Wang and X. Gao, “Multimodal Approaches to Minority Language Preservation,” *Language Documentation & Conservation*, vol. 16, pp. 1-25, 2022.
- [4] UNESCO, *Atlas of the World’s Languages in Danger*. Paris, France: UNESCO Publishing, 2023.
- [5] Y. Zhang, “Preservation and Digitization of Zhuang Scripts: A Sociolinguistic Perspective,” *Journal of Minority Languages and Culture*, vol. 29, no. 2, pp. 55-70, 2022.
- [6] T. Hoang, “Digitalizing Vietnamese Nom Script,” *Journal of Cultural Heritage Management*, vol. 11, no. 3, pp. 210-225, 2021, doi: 10.1108/JCHM-01-2021-0003.
- [7] J. Lü, “From ‘Jing Ethnic Language’ and ‘Vernacular’ to Mandarin,” *Guangxi Ethnic Studies*, no. 4, pp. 45-53, 2023.
- [8] C. Reul, C. Wick, and F. Puppe, “Evaluating OCR Accuracy for Historical Documents,” in *Proc. Int. Conf. on Document Analysis and Recognition*, 2019, pp. 1-6.
- [9] ABBYY, “FineReader 15 User Guide,” Moscow, Russia: ABBYY, 2019. [Online]. Available: <https://www.abbyy.com>. [Accessed March 10, 2025].

[10] ATLAS.ti GmbH, *ATLAS.ti 9 User Manual*. Berlin, Germany: ATLAS.ti Scientific Software Development GmbH, 2020.

[11] C. S. Peirce, *Collected Papers*, vol. I–VIII. Cambridge, MA: Harvard University Press, 1931-1935.

[12] Y. Wang, “Minority Language Education in Southwest China,” *Language Policy*, vol. 22, no. 1, pp. 89-112, 2023.

[13] Guangxi Department of Culture, *Annual Report on Minority Language Protection*. Nanning, China: Guangxi Department of Culture, 2023.

[14] INALCO & Han Nom Institute, *Final Report: Vietnam–France Digital Nom Script Project*. Paris & Hanoi: INALCO-Han Nom, 2022.

[15] T. T. H. Le, “A Comparative Study of China-Vietnam Intangible Cultural Heritage Protection Policies: Focusing on the Preservation of Nom Script,” *Journal of Yunnan University (Social Sciences Edition)*, vol. 2017, no. 2, pp. 34-37, 2017.

[16] UNESCO, “Recommendation on the Historic Urban Landscape,” Paris, France: UNESCO Publishing, 2011. [Online]. Available: <https://whc.unesco.org/en/news/1026>. [Accessed May 30, 2025].

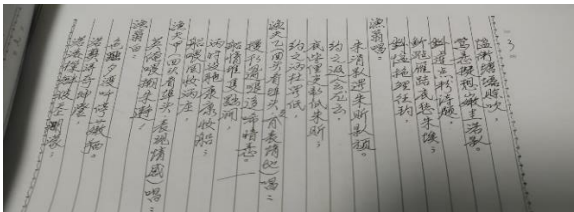
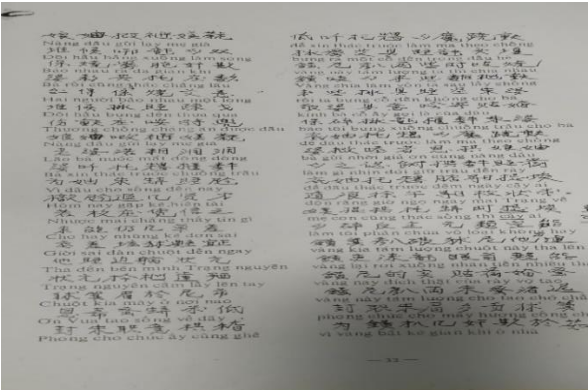
[17] Nom Foundation, *Vietnamese Nom Preservation and Digital Strategy Report*, Hanoi: Nom Foundation, 2023.

[18] Guangxi Library, *Report on testing the multilingual Sino-Nom document digitization platform in 2023*, Nam Ninh: Guangxi Library, 2023.

**Phụ lục A. Bảng tóm tắt cơ cấu mẫu phỏng vấn sâu**

Tiêu chí	Phân loại	Số lượng (người)	Tỷ lệ (%)
Độ tuổi	61–85 tuổi (Trung bình: 73 tuổi)	15	100
Giới tính	Nam	9	60
	Nữ	6	40
	Nông dân	8	53,3
Nghề nghiệp trước khi nghỉ hưu	Giáo viên	2	13,3
	Thầy cúng	2	13,3
	Cán bộ văn hóa	3	20
	Biết đọc/viết tốt	6	40
Hiểu biết về chữ Nôm	Biết đọc cơ bản	5	33,3
	Truyền miệng hoặc lưu giữ	4	26,7

**Phụ lục B. Một số tài liệu chữ Nôm tiêu biểu của người Kinh tại Quảng Tây**

Tên văn bản	Địa điểm thu thập	Nội dung tóm tắt
	Làng Vạn Vĩ	Là câu chuyện dân gian của người Kinh kể về cuộc đời của ông lão đánh cá, những khó khăn vất vả buổi ban đầu đến lập nghiệp ở đảo hoang và tình cảm gắn bó với hòn đảo.
<b>“Ngư Ông truyện” 渔翁传</b>		
	Làng Sơn Tâm	Là những câu thơ sáu chữ viết bằng chữ Nôm kể về chàng Tống Trân, có nội dung cốt truyện gần gũi với truyện thơ Nôm <i>Tống Trân Cúc Hoa</i> ở Việt Nam.
<b>“Tống Trân ca” 宋珍歌</b>		