

Một cách nhìn về tương lai của chữ Nôm

Ngô Thanh Nhân, Đại học New York, nhan@cs.nyu.edu
Ngô Trung Việt, Viện Công nghệ Thông tin, vietnt@itprog.gov.vn

Hội nghị Quốc tế về chữ Nôm
Ngày 12-13 tháng 11 năm 2004
Thư viện Quốc gia Việt Nam – Hà Nội

Tóm tắt

Bài này đưa ra một số nhận xét về những tiến triển trong việc nghiên cứu, in ấn, phổ cập sử dụng chữ Nôm trong những năm gần đây, cả về số lượng, chất lượng, người và của, trong và ngoài nước. Sự tiến triển này cơ bản không tách rời tiến triển chung của kỹ thuật xử lý đa ngữ, tra cứu mạng, kiến thức ngôn ngữ học, và kiến thức xử lý văn bản, và những chính sách thích hợp. Theo ý chúng tôi, trung tâm của sự tiến triển này là bộ mã chuẩn Unicode và ISO/IEC 10646 mà Việt Nam đã quyết định tham gia từ đầu những năm 1990.

Cơ bản của cách nhìn về tương lai của chữ Nôm là chuẩn mã (*codepoints*), dạng chính tắc (*canonical forms*), cấu phần cơ bản (*basic components*) và bổ sung thêm kho (*repertoire*) chữ Nôm dành cho mọi loại nghiên cứu và sử dụng rộng rãi, trên cơ sở tra cứu (*search*) mạng *internet*. Từ một bộ mã chung qua mạng, thông qua tiêu chuẩn Việt Nam, mà những nhà nghiên cứu, giáo dục, bảo tồn, bảo tàng, thư viện, v.v. có thể nghĩ đến những tiến triển xa hơn, trong tri thức tập thể của cộng đồng Việt Nam và thế giới.

Việc này không thể làm được nếu chúng ta không có một chương trình đưa các văn bản văn hoá quan trọng của Việt Nam vào một kho lưu trữ quốc gia mà ai cũng tham khảo được. Do đó, bài này phác họa một số hoàn cảnh cơ bản và định hướng để tiến tới một tương lai mà chúng tôi cho là mặc nhiên—trong đó, chúng ta không để mất những tài sản tri thức có thể mất, lấy lại những tài sản tri thức đã mất, và làm cho sử và đặc trưng của dân tộc Việt Nam hiển lộ với đầy đủ chứng cứ.



1. Bối cảnh

Việt Nam là một quốc gia có lịch sử và văn hoá lâu dài hàng nghìn năm. Những phát hiện mới đây về di tích Hoàng thành tại Ba Đình, Hà Nội càng làm cho chúng ta có cảm giác tự hào về một nước Việt Nam với truyền thống lâu dài. Nhiều giá trị văn hoá cổ truyền đang được phục hồi và phát triển để trả lại cho dân tộc những chứng cứ, những điều tưởng chừng như bị mai một trong quá khứ và trong những biến chuyển xã hội lớn lao của thế kỷ thứ 20. Chữ Nôm chính là một trong những giá trị đó và hiện đang được khôi phục trả lại vị trí xứng đáng của nó như một mảnh đất lưu giữ nhiều vốn văn hoá cổ, nhiều tri thức, hiểu biết, phong tục, cách nói cổ...

Hơn bao giờ hết, xã hội Việt Nam ngày nay đang quan tâm ngày càng nhiều tới việc phục hồi chữ Nôm, không phải như một vật phẩm khảo cổ được trưng bày bảo tàng mà là một tri thức dân tộc sống động trong con người và cộng đồng. Sự quan tâm của cả cộng đồng nói chung, nhất là của giới trẻ nói riêng đối với việc khôi phục và sử dụng chữ Nôm hiện đã tăng lên rất nhiều, nhất là qua các phương tiện kỹ thuật hiện đại. Sự quan tâm của thế hệ trẻ với chữ Nôm trên nền công nghệ hiện đại hứa hẹn tương lai khác hơn cho chữ Nôm trong tiến trình văn hoá dân tộc. Thanh thiếu niên Việt Nam ở mọi nơi bắt đầu chú ý đến vốn chữ Hán Nôm—coi tiếng Việt có hai cách viết, một cách theo la-tinh, một cách theo chữ biểu ý. Cả hai đều được gọi là chữ quốc ngữ—chữ Hán Nôm được gọi là quốc ngữ trong quá khứ, chữ Việt la-tinh được gọi là chữ quốc ngữ từ đầu thế kỷ thứ 20.

Công nghệ thông tin (CNTT) với những bước tiến phổ cập máy vi tính và phát triển mạng máy tính toàn cầu đã trở thành phương tiện hữu hiệu cho việc khôi phục và phổ cập sử dụng chữ Nôm. Máy tính hiện đại ngày càng có năng lực xử lý nhiều ngôn ngữ và chữ viết phức tạp, nói riêng đã có khả năng xử lý được cho chữ biểu ý, chữ Nôm từ đầu những năm 1990. Các chuẩn đa ngữ quốc tế đã đưa nhiều tiến bộ công nghệ vào phổ cập giúp cho việc khôi phục và gìn giữ các giá trị văn hoá cổ.

Sự thâm nhập của CNTT vào các hoạt động nghiên cứu và bảo tồn di sản Hán Nôm đã và đang phát triển mạnh trong vòng hơn một chục năm qua. Trong hơn mười năm qua đã có nhiều cố gắng và nỗ lực của các cá nhân và tập thể để đưa tiến bộ công nghệ vào việc phục hồi vị trí xứng đáng của chữ Nôm. Nói riêng, mọi việc chuẩn bị để cho các máy tính hiện đại có thể xử lý được chữ Nôm đã được bắt đầu từ năm 1992 với việc Việt Nam yêu cầu tổ chức tiêu chuẩn quốc tế xem xét đưa chữ Nôm của Việt Nam vào bộ kí tự toàn thế giới Unicode và ISO/IEC 10646. Và từ đó, Việt Nam đã tham gia phát triển chuẩn quốc tế Unicode và ISO/IEC 10646, với mục đích đưa chữ quốc ngữ la-tinh, chữ Nôm, và các chữ dân tộc Việt Nam vào bộ kí tự quốc tế.

Những tiến bộ công nghệ mới về phần mềm và mạng máy tính cho phép phổ cập và phục hồi việc sử dụng di sản chữ Hán Nôm đã từng bị quên lãng trong gần 100 năm qua. Chữ Nôm đã xuất hiện trên màn hình mạng toàn cầu trước con mắt đầy ngạc nhiên của nhiều em nhỏ, điều mà các em thậm chí còn chưa hề được nhìn thấy qua sách học chính thức của các lớp phổ thông, mặc dầu vẫn thấy chúng trên các câu đối, hoành phi.

Chữ Nôm cũng bắt đầu bước vào lĩnh vực xuất bản điện tử với việc ra đời ở Mỹ cuốn thơ Hồ Xuân Hương, in bằng chữ Nôm, chữ quốc ngữ và chữ tiếng Anh, vào năm 2000. Và trong năm nay cuốn *Giúp đọc Nôm Hán Việt* của Linh mục Anthony Trần Văn Kiệm để tra cứu chữ Nôm, được in dùng phong chữ Nôm trên máy tính, cũng đã được xuất bản. Trên thế giới một bộ phong cho chữ Hán Nôm của nhóm Đạo Uyên cũng đã được phổ biến. Trong nước, đã xuất hiện nhiều trang web cho phép tra cứu và sử dụng chữ Nôm như của HueSoft, HaNoSoft. Tại Mỹ, *Hội*

Bảo tồn Di sản chữ Nôm thường xuyên duy trì trang web có chương trình tra cứu các chữ Nôm mà Việt Nam đã đăng kí với thế giới. Tất cả các hệ phần mềm đều dùng chữ quốc ngữ la-tinh song song với chữ Hán Nôm.

Đối với những nhà nghiên cứu, việc có khả năng sử dụng, in ấn, truy nhập thông tin chữ Nôm trên máy tính đã mở ra một chân trời mới cho những nghiên cứu mới theo hướng tổng hợp toàn diện mọi khía cạnh đã thu thập được của chữ Nôm. Đây là một điều kiện hết sức lí tưởng mà không phải người nghiên cứu nào trước đây cũng có thể có được. Hiện nay yêu cầu tri thức về chữ Nôm là CNTT phải phản ánh được đặc thù của cấu tạo, nội dung và sử dụng chữ Nôm.

Ví dụ: Trong quá khứ, người Việt đã sáng tạo chữ Nôm để ghi lại tiếng Việt. Trong tiến trình đó, họ đã dùng chữ Nôm để mô tả các đặc tính của tiếng Việt: một thứ tiếng đơn âm (mỗi tiếng là một âm tiết, và được ghi bằng một chữ). Vì số tiếng giới hạn, chỉ khoảng 6.700, người Việt đã dùng loại từ và ngữ cảnh (ta còn gọi là nét nghĩa) để phân biệt cách các tiếng cùng âm khác ngữ cảnh. Các nét nghĩa này được ghép thẳng vào các hình chữ ghi âm. Cách dùng chữ Nôm như thế chẳng những ghi lại được các phương ngữ tiếng Việt mà còn có các nghĩa khác nhau của tiếng. Ví dụ, chữ 多 “đa” được dùng để ghi 44 chữ sau đây theo cơ sở tri thức kèm theo bộ phong Nôm Na, Tổng thể, dạng mảnh, NomNaTongLight (phiên bản 1.0):

𪗇	bê, da [đa + bì] U+63010	𪗇	bộn [đa + bản] U+635b4
𪗈	bộn [đa + bạn] U+215b1	𪗈	cầu [đa + câu] U+05920
𪗉	cầu [câu + đa] U+0591f	𪗉	cười [đa + cái] U+635b1
𪗊	da [nhục + đa] U+043e7	𪗊	day, di, dây, dê, dôi, giay, rời [hoà + đa] U+079fb
𪗋	dám [bấm + đa] U+215cb	𪗋	dám [bấm + đa] U+60134
𪗌	đa [mẽ + đa] U+25e60	𪗌	đa [mộc + đa] U+06818
𪗍	đa [điều + đa] U+2a013	𪗍	đá, đớ, đứ, đừa [thạch + đa] U+254a5
𪗎	đá [túc + đa] U+08de2	𪗎	đe, đớ, đứ, xỉ, xỉ [khẩu + đa] U+054c6
𪗏	đi [khứ + đa] U+20afe	𪗏	đi [đa + khứ] U+634db
𪗐	đừ [thủy + đa] U+23d19	𪗐	đưa [xước + đa] U+08ffb
𪗑	đưa [thủ + đa] U+062f8	𪗑	đưa [đa + tổng] U+635b9
𪗒	đừa, sẩy, xẩy, xẩy, xi, xỉ [nhân + đa] U+04f88	𪗒	đừa [đinh + đa] U+20032
𪗓	đừa [đa + đinh] U+635af	𪗓	đơ [trực + đa] U+63029
𪗔	ghi [cơ/ki + đa] U+635bd	𪗔	lai [lai + đa] U+635b5
𪗕	lắm [lắm + đa] U+635be	𪗕	lắm [thậm + đa] U+05926
𪗖	loã [quả + đa] U+05925	𪗖	nhèo, nhiều [nghiêu + đa] U+215c9
𪗗	nhieu [nhiều + đa] U+635bf	𪗗	nhóm [chiếm + đa] U+215a1
𪗘	quên [đa + mai] U+635bb	𪗘	răng [lăng + đa] U+635b6
𪗙	rườm [đa + rềm] U+635bc	𪗙	sai [sai + đa] U+215c2

𠂇	sánh [á + đa] U+215bc	𠂇	si [mục + đa] U+07735
𠂇	ta [phụ + đa] U+07239	𠂇	trẩy [lễ + đa] U+63052
𠂇	tụm [đa + tạm] U+215ce	𠂇	vai [đa + mai] U+635ba

Các chuyên gia CNTT có thể có những nhận xét sau đây :

- Những chữ như “cấu” 夠够, “đi” 𠂇𠂇 hay “đưa” 𠂇𠂇, có thể viết hai cách khác nhau, và bộ phận “đa” 多 tự do kết hợp trong khung vuông. Sự tự do này cho thấy người Việt xem khung vuông là vùng kết hợp lỏng.
- Những chữ như “bộn” 𠂇𠂇, “đa” 𠂇𠂇, “đá” 𠂇𠂇, “đưa” 𠂇𠂇, “đi” 𠂇𠂇, “lắm” 𠂇𠂇, ... có cùng bộ phận “đa” 多 nhưng khác ở bộ phận kết hợp cho thấy cách giải quyết đặc biệt vấn đề cùng âm khác nghĩa trong chữ Nôm mà chúng ta không thấy trong chữ quốc ngữ la-tinh.

2. Những thành tựu đã đạt được

Để có được những kết quả như ngày hôm nay chúng ta thấy đối với chữ Nôm, chúng ta đã phải mất nhiều công sức, lao động và tiền bạc. Những đầu tư đó vào chữ Nôm không bao giờ phí hoài, trái lại nó đã thúc đẩy sự quan tâm của cả cộng đồng với vốn văn hoá chữ Nôm mà có thời đã từng bị coi là không còn giá trị sử dụng gì nữa. Sau đây xin điểm lại một số kết quả chính đã thu được:

Mã hoá: Trong 12 năm từ 1992-2004 chúng ta đã kiên trì theo đuổi các kì họp quốc tế để yêu cầu đưa chữ Nôm vào bộ kí tự thế giới Unicode và ISO/IEC 10646. Chúng ta đã đưa vào 5.067 chữ trùng hình với chữ CJK Trung-Nhật-Hàn trong mặt phẳng cơ sở BMP, 4.232 chữ trong mặt phẳng 2, và hiện đang đề nghị đưa vào thêm 2.200 chữ nữa. Như vậy số chữ Nôm rút từ nguồn tư liệu thật đã được đăng kí với tổ chức tiêu chuẩn thế giới là khoảng trên 11.500 chữ.

Phông Nôm: Thực tế trong sử dụng hiện nay trên các máy tính, mới chỉ có bộ phông Arial Unicode MS chứa khoảng hơn 5.000 chữ Nôm trùng hình chữ Hán. Viện Văn Tự Kính Mojikyo tại Nhật Bản đã làm ra phông chữ TrueType cho 9.299 chữ Nôm mà Việt Nam đã có trong bộ mã quốc tế. Công ti DynaLab Đài Loan có trụ sở tại Thượng Hải và Hồng Kông đã xây dựng bộ phông Tổng thể DFSongLight_Vietnam2.ttf cũng cho 9.299 chữ Nôm này. Nhóm Đạo Uyển (Đỗ Quốc Bảo tại Đức và Thiên viện Viên Chiếu) đã phát triển bộ phông HanNom (trên 30.000 chữ) có thể sử dụng trên mạng. Nhóm Nôm Na đã phát triển bộ phông TrueType NomNaTongLight.ttf (trên 17.700 chữ)

Phần mềm: Đã phát triển phần mềm tra cứu chữ Nôm gọi là Nôm Lookup Tool (NLT) được sử dụng rộng rãi trên mạng cả trong nước và trên thế giới. Các phần mềm gõ chữ Nôm và phần mềm từ điển đã được một số nhóm chuyên gia tin học trong nước phát triển: các nhóm của Phan Anh Dũng (Huế) và Tổng Phước Khải-Lê Anh Minh (tp Hồ Chí Minh).

In ấn: Đã thực hiện việc in ấn chữ Nôm từ máy tính cho một số bộ từ điển chữ Nôm. Nhiều tác phẩm chữ Nôm đã và đang được in ấn trực tiếp từ máy tính và tra cứu trên mạng.

3. Các xu hướng phát triển tương lai cho chữ Nôm

Trên cơ sở những kết quả đã đạt được nhờ sự kết hợp của học giả chữ Nôm và các chuyên gia CNTT, chúng ta có thể phóng tầm nhìn vào tương lai để hình dung một số xu hướng phát triển sắp tới và đề định hướng cho các hoạt động nghiên cứu, triển khai sắp tới. Các góc nhìn sẽ đề cập tới việc phát triển sử dụng chữ Nôm, việc thúc đẩy các nghiên cứu chữ Nôm với sự hỗ trợ của công nghệ, việc mở rộng các ứng dụng xử lý có liên quan tới chữ Nôm cho các lĩnh vực thư viện, kho tư liệu và bảo tàng.

3.1 Phổ cập chữ Nôm qua máy tính

Người ta thường cho rằng chữ Nôm từ xưa tới nay không phải là thứ chữ được phổ cập rộng rãi, vì nó khó học, phải nhớ nhiều, và cần có vốn chữ Hán nào đó. Thời xưa, làng nào có truyền thống hiếu học thì mới có thầy đồ dạy chữ Hán-Nôm cho trẻ trong làng. Thầy dạy chữ Nôm không có tiêu chuẩn đi học làm thầy hay kiến thức giáo dục. Thời nay, số người biết và có thể dạy được chữ Nôm ngày càng ít đi, nhưng bù lại, nhờ sự hoàn chỉnh của chữ quốc ngữ latin, công nghệ phát triển đã tạo khả năng có được các ông thầy ảo trên máy tính, có thể giúp cho mọi người học được chữ Nôm qua máy tính có phương pháp. Và giới trẻ rất nhanh nhẹn, tháo vát trong việc tiếp thu những tri thức, kỹ năng mới, cũng không thấy khó khăn gì khi cần học chữ Nôm trên máy tính. Do đó song song với quá trình phổ cập tin học, phổ cập máy tính trong cả nước, việc tạo ra khả năng để các em nhỏ được tiếp xúc và học hỏi chữ Nôm trên máy tính và trong giáo trình cơ sở sẽ tạo điều kiện làm cho chữ Nôm sống lại trong tâm trí các thế hệ trẻ.

Chữ Nôm sẽ trường tồn với lịch sử khi các thế hệ trẻ vẫn còn say mê tìm hiểu và học tập về chữ Nôm. Xã hội và nền giáo dục cần trang bị cho mọi học sinh niềm tự hào về di sản văn hóa chữ Nôm, các em cần được thấy chữ Nôm trong các sách giáo khoa phổ thông. Như vậy việc đào tạo những thế hệ trẻ biết sử dụng chữ Nôm và CNTT trong thực tế là điều quan trọng và cần thiết. Và việc bảo tồn, duy trì vốn văn hoá chữ Nôm này ngay trong những người sống, những người trẻ, chính là cách thức bảo tồn hữu hiệu nhất và không mất đi được.

Khi mà xã hội phát triển, đời sống được nâng cao thì nhu cầu học và dùng chữ Nôm trong các hoạt động văn hóa sẽ càng lớn. Sẽ tới lúc tiêu chuẩn này trở thành tiêu chuẩn cho con người có văn hóa và am hiểu văn hóa dân tộc: người biết và sử dụng được chữ Nôm trên máy tính.

Để phổ biến các tư liệu chữ Nôm cho mọi người có nhu cầu, cần phát triển các kho lưu trữ các văn bản Hán-Nôm dạng điện tử ở cấp quốc gia, địa phương và gia đình, phổ biến miễn phí qua mạng. Chính việc nhiều người có khả năng truy nhập và sử dụng trực tiếp các nguồn tư liệu cổ chữ Hán-Nôm là một đảm bảo chắc chắn cho sự trường tồn của những vốn văn hoá cổ này. Việc cất giữ chu đáo các văn bản cổ trong các bảo tàng chắc chắn để mọi người chiêm ngưỡng cũng là một hình thức bảo tàng, nhưng cách gìn giữ chủ động hơn cả là làm cho vốn văn hoá đó sống ngay trong tâm khảm người Việt. Và đó là một trong những mục đích, chủ yếu của việc phổ biến mọi tài liệu chữ Nôm cho mọi người có nhu cầu.

Để thực hiện được tầm nhìn phổ biến chữ Nôm trên máy tính này, ngoài những cố gắng về mặt công nghệ, những cố gắng phát triển phần mềm khai thác và xử lý chữ Nôm trên máy tính, cần thúc đẩy triển khai các ứng dụng CNTT liên quan tới thư viện và lĩnh vực bảo tồn di sản chữ Nôm. Khi xuất hiện thêm các nhu cầu triển khai ứng dụng cho chữ Nôm trên máy tính và mạng, thì đó sẽ lại là động lực thúc đẩy thêm nữa các nhà sản xuất phần mềm đưa ra thị trường các sản phẩm phần mềm xử lý tốt hơn cho các ứng dụng đặc thù chữ Nôm.

Để dùng được chữ Nôm từ xưa đều phải có những người am hiểu khá kỹ về chữ Hán. Ngày nay một tình huống ngược lại đang phát sinh: nhiều người trẻ rất am hiểu máy tính và mong muốn đi vào tìm hiểu chữ Nôm qua máy tính. Những người già biết về chữ Hán-Nôm lại thường không đủ hiểu biết làm chủ và sử dụng máy tính. Vì vậy một tầm nhìn mới đối với sự phát triển của chữ Nôm trong thời hiện đại là thúc đẩy việc làm chủ chữ Nôm của thế hệ trẻ, những người đã được công nghệ trang bị cho nhiều tri thức tổng hợp về chữ Nôm, thông qua các cơ sở dữ liệu chữ Nôm được thiết lập trên máy tính, để họ sẽ trở thành những người nghiên cứu và bảo tồn chữ Nôm trong tương lai. Do đó việc phát triển và xây dựng các giáo trình về chữ Nôm và CNTT trở thành một đòi hỏi của thực tế hiện nay. Thiếu sự đầu tư công sức và hiểu biết cho lĩnh vực này sẽ làm chậm lại việc phổ cập chữ Nôm cho lớp trẻ.

3.2 Nghiên cứu về chữ Nôm

Bên cạnh việc phổ cập khả năng tiếp xúc việc học chữ Nôm, việc truy nhập vào các kho tư liệu chữ Nôm, thì việc hình thành cách nghiên cứu mới về chữ Nôm dựa trên những tiến bộ công nghệ cũng là một tầm nhìn mới cần được thiết lập và phổ biến trong giới nghiên cứu.

3.2.1 Phát triển phương pháp luận làm việc mới với chữ Nôm dựa trên CNTT

Vấn đề lớn nhất là làm thế nào các nhóm nghiên cứu có thể trao đổi thảo luận các chủ đề về chữ Nôm trên mạng. Muốn làm được việc này phải làm gì? Cần có sự thống nhất về biểu diễn chữ Nôm trên máy tính để sao cho khi thông tin về chữ Nôm truyền qua mạng máy tính, người gửi và nhận vẫn có thể đọc được đúng thông tin chữ Nôm đã được gửi. Trong công nghệ thông tin, không thể có tiến bộ với sự tham gia của nhiều người ở nhiều nơi, nếu không có chuẩn để trao đổi tri thức tập thể. Chuẩn là mốc đánh dấu tiến bộ trong làm việc tập thể. Do đó các nghiên cứu về chữ Nôm cần dựa trên việc thống nhất về một số chuẩn trao đổi thông tin. Trong trường hợp có những bất đồng khi xem xét chữ Nôm, mọi khả năng biểu diễn bất đồng đều cần được bảo lưu và thể hiện trên máy để cho phép nhiều người khác cùng tham khảo và góp ý kiến.

Rõ ràng cho tới nay chuẩn mã quốc tế và tiến bộ công nghệ mới đang trở thành con đường duy nhất để đưa nghiên cứu tiếng Việt vào thế kỷ thứ 21. Vấn đề cần xác định rõ các chuẩn trong nghiên cứu Hán Nôm và CNTT là gì. Trước mắt đó là vấn đề chuẩn về mã hoá, ứng mỗi chữ với một con số để máy có thể xử lý được, và con số độc nhất cho chữ Nôm này phải được thống nhất trên toàn thế giới. Tiếp đó sẽ là một số chuẩn cách viết, chuẩn cách đặt tương ứng giữa hình chữ Nôm và âm đọc. Cũng có thể nhận diện một số qui tắc phán đoán, một số hoàn cảnh, tình huống để thiết lập mối tương quan giữa hình chữ Nôm và âm đọc. Từ xưa tới nay, và ngay cả trong Hội nghị này, việc không thống nhất trong đoán định các âm đọc cho một số chữ Nôm đã bị thay đổi đã từng là đề tài cho nhiều tranh luận, thảo luận. Ví dụ chữ 里 “lí” có khoảng 18 cách viết, cách viết nào xuất hiện lúc nào, văn bản nào, và phương ngữ nào còn là vấn đề để mở, v.v.

李 [U+0674e], 里 [U+091cc], 哩 [U+054e9], 裡 [U+088e1], 裏 [U+088cf],
 俚 [U+04fda], 理 [U+07406], 鋸 [U+092f0], 鋸 [U+09502], 鯉 [U+09bc9],
 鯉 [U+09ca4], 鱣 [U+09c7a], 鯪 [U+09ca1], 漣 [U+09090], 漣 [U+09026],
 履 [U+05c65], 莖 [U+08372], 媿 [U+05a0c],...

Như thế vẫn chưa đủ, mỗi chữ trên lại có nhiều cách đọc khác nhau...

- lí 里 [U+091cc] còn đọc là **li**,
- lí 哩 [U+054e9] còn đọc là **ré, rí**,
- lí 理 [U+07406] còn đọc là **lẽ**,
- lí 履 [U+05c65] còn đọc là **giày**...

Cách nhìn cũ coi đây là nhược điểm của chữ Nôm. Cách nhìn mới ghi nhận rằng chữ viết ghi lại tiếng nói trong lịch sử, ví dụ tiếng Anh ghi lại cách phát âm cũ trước khi các nguyên âm thay vị trí nhau, gọi là chuyển dời nguyên âm *the great vowel shift*. Đây chính là những chứng cứ để chúng ta nhìn vào lịch sử phát triển của ngôn ngữ tiếng Việt.

Một nhu cầu lớn hiện nay là cần xác định phương pháp luận nghiên cứu mới đối với các vấn đề liên quan tới chữ Nôm. Phương pháp luận nghiên cứu này dựa trên việc sử dụng tiến bộ CNTT giúp cho nghiên cứu về chính bản thân chữ Nôm, các văn bản cổ và sự phát triển của tiếng Việt qua các văn tự. Với khả năng của máy tính có thể tích trữ vào trong nó rất nhiều tư liệu, gần như có thể vét cạn cả kho tư liệu sách vở thông thường, người nghiên cứu giờ đây có thể có con mắt bao quát, xuyên suốt nhiều tư liệu để rút ra các đặc trưng của ngôn ngữ, của cách viết. CNTT hiện nay có thể làm được nhu cầu này. Nếu như các học giả trước đây thường phải tự mình lưu trữ các kho tư liệu riêng để soạn sách, viết bài, thì ngày nay, với phương pháp làm việc mới, máy tính có thể hỗ trợ trong việc quản lí kho tư liệu chung này. Do đó các phương pháp luận lưu giữ và trích rút thông tin tổng quát từ kho tư liệu số thức hoá trên máy tính sẽ trở thành cực kì có ích cho giới nghiên cứu. Nói riêng, với những người ít có khả năng truy nhập trực tiếp vào các kho tư liệu này, có thể cần tới một đội ngũ các kĩ thuật viên giúp cho họ truy nhập và lấy ra thông tin cần thiết.

Dựa trên công nghệ hiện đại, cần xác định ra các qui trình nghiên cứu và làm việc mới có liên quan tới chữ Nôm và CNTT để tạo ra nề nếp và thói quen nghiên cứu mới. Các nghiên cứu này thường bao giờ cũng phải bắt đầu bằng việc tổng hợp mọi thông tin có được trong kho theo các chủ đề nghiên cứu, để rồi từ đó người nghiên cứu rút ra các nhận định, các ý kiến riêng của mình. Những thông tin mới về chữ Nôm, những phát kiến mới về việc thiết lập tương ứng giữa hình chữ và âm đọc cần được đưa vào máy tính lưu giữ theo những qui trình thống nhất để có thể phục vụ được cho nhiều người khác cùng tham khảo và nghiên cứu. Do đó, việc xây dựng các chuẩn về tư liệu, về biểu diễn văn bản chữ Nôm trở thành quan trọng và cần thiết.

Chúng ta có thể nói:

- Chuẩn mã chữ Hán-Nôm giúp cho việc trao đổi các thông tin có độ tin cậy tuyệt đối. Ví dụ, hình chữ 里 có mã U+091cc sẽ không thay đổi khi gửi qua mạng hay thư điện tử ở bất cứ nơi nào trên thế giới, nhờ tuân thủ theo chuẩn quốc tế;
- Kho cơ sở dữ liệu cho phép chúng ta rút những thông tin khác của mã U+091cc, mã có hình chữ 里 cho ta cách đọc “lí” tiếng Hán-Việt, “li” trong chữ Nôm, và những thông tin khác, như U+091cc thuộc bộ nào, có bao nhiêu nét, ở trong ngữ cảnh nào, là danh hay động, xuất hiện ở các câu như thế nào trong các văn bản, cách đọc ra sao trong các văn bản ấy, ở thế kỷ nào, v.v. Chữ này có cách đọc nào khác trong các thứ tiếng như Trung Quốc, Hàn Quốc, Nhật, v.v.

3.2.2 Giải quyết vấn đề tư liệu chữ Nôm qua hệ thống lưu trữ tài liệu chữ Nôm

Nhà ngôn ngữ học Firth nói: *Know a word by the company it keeps*—ta biết nghĩa của chữ bằng các chữ khác nó giữ làm bạn đồng hành. Vấn đề tư liệu và văn bản học đối với chữ Nôm từ xưa tới nay đã là vấn đề khó giải quyết vì chúng ta phải chấp nhận một quá khứ không có khái niệm bản quyền, một quá khứ mà mọi tác phẩm kinh điển đều có sự tham gia nhuận sắc của nhiều người, kể cả của chính người tổ chức việc in khắc. Giải quyết triệt để vấn đề tư liệu và văn bản học chỉ có thể được thực hiện nếu chúng ta có công cụ lưu giữ tất cả các dị bản và có công cụ trích rút ra những thông tin từ mọi dị bản đó. Các nghiên cứu sẽ được chính xác hơn khi mà mọi ý kiến và suy luận được xem xét một cách tổng thể trên mọi dị bản chứ không bị phụ thuộc vào mỗi một bản. Việc xây dựng kho dữ liệu điện tử về tất cả các tư liệu hiện có, từ tận gốc, và hình thành kho tư liệu nghiên cứu chung cho giới Nôm học thực sự là một nhu cầu cấp thiết và hữu ích. Đây sẽ là nền tảng để giới Nôm học đi tới những nhận định thống nhất, toàn diện về các vấn đề tranh cãi, không bị buộc phải dựa vào những nhận định bắt nguồn từ một số tư liệu riêng lẻ.

Kho dữ liệu điện tử này sẽ có hai tầng: tầng chữ và tầng tác phẩm. Kho dữ liệu chữ Nôm sẽ chứa tất cả các dạng chữ Nôm đã xuất hiện trong bất kì tài liệu nào, và ngoài những thông tin nội tại của nó, còn có các thông tin liên quan tới sự xuất hiện của chúng trong mọi tài liệu khác. Kho dữ liệu các tác phẩm chữ Nôm lưu giữ tất cả các tác phẩm chữ Nôm hiện có (theo thời gian điều này có thể làm được) dưới dạng điện tử. Nó cho phép người nghiên cứu có thể trích đoạn, rút ra các câu chữ cần thiết minh hoạ cho nghiên cứu của mình.

Để tiến kịp trình độ thế giới, rõ ràng và cụ thể nhất là tiến đến việc sử dụng chữ Nôm trong *Google*, một công cụ tra cứu đa ngữ toàn thế giới. Ví dụ, Thư viện X ở một nước Y đưa quyền *Truyện Kiều* lên mạng Thư viện X. Tại Việt Nam, khi tra thuần túy bằng quốc ngữ, không cần quan tâm đến nghĩa, sự kết hợp ngữ pháp và ngữ nghĩa của các chữ, như cách tra của mạng *Google*, ta có thể hình dung như sau:

trăm	(3)	𣪠 𣪡 𣪢
năm	(6)	𣪣 𣪤 𣪥 𣪦 𣪧 𣪨
trong	(16)	𣪩 𣪪 𣪫 𣪬 𣪭 𣪮 𣪯 𣪰 𣪱 𣪲 𣪳 𣪴 𣪵 𣪶 𣪷 𣪸
cõi	(3)	𣪹 𣺁 𣺂
người	(3)	𣺃 𣺄 𣺅
ta	(10)	𣺆 𣺇 𣺈 𣺉 𣺊 𣺋 𣺌 𣺍 𣺎 𣺏

Theo phương pháp thuần toán, phi ngôn ngữ học, ta nhân 3 chữ “trăm”, cho 6 chữ “năm”, 16 chữ “trong”, 3 chữ “cõi”, 3 chữ “người” và 10 chữ “ta”, ta có 25.920 cách xếp sáu chữ “trăm năm trong cõi người ta”, 1 trong số 25.920 “chuỗi sáu chữ” đó là đúng văn phạm và ngữ nghĩa.

Google vẫn tìm ra đúng văn bản Kiều trong Thư viện X, bởi vì hàng chữ Nôm trong văn bản Kiều là đúng văn phạm và ngữ nghĩa.

Dĩ nhiên, nếu chúng ta có một máy phân tích ngữ pháp câu thì chỉ có 9 câu trong số 25.920 câu “trăm năm trong cõi người ta” với những chữ cùng nghĩa khác hình (như 3 chữ “năm” 年 輛 輛, 3 chữ “trong” 中 中 中) là đúng ngữ pháp và ngữ nghĩa, rút ngắn thời gian tìm của Google.

Ví dụ, câu sáu “Trăm năm trong cõi người ta...” có hình cây ngữ pháp sau, theo nhà ngôn ngữ học Zellig Harris và nhóm luận điệu “chuỗi ngôn ngữ” *Linguistic String Project*:

```

<SENTENCE>
  <SA> <===== cùm phụ của câu
  <NSTGT> <===== mang nghĩa thời gian
  <NSTG> <===== cùm danh
  <LNR>
  <LN>
  <TPOS/>
  <QPOS> <===== cùm lượng
  <LQR>
  <LQ></LQ>
  <QVAR><Q> 冪 <Q></QVAR> <===== “trăm”
  <RQ></RQ>
  </LQR>
  <APOS/>
  <NPOS/>
  </LN>
  <NVAR> <N> 輛 </N></NVAR> <===== danh từ “năm”
  <RN>
  <PN>
  <P> 中 </P> <===== giới từ “trong”
  <NSTGO>
  <NSTG>
  <LNR>
  <LN></LN>
  <NVAR><N> 增 <N></NVAR> <===== danh từ “cõi”
  <RN>
  <NSTGO>
  <NSTG>
  <LNR>
  <LN></LN>
  <NVAR><N> 得些 </N> </NVAR> <===== “người ta”
  <RN></RN>
  </LNR>
  </NSTG>
  
```

```

        </NSTGO>
        </RN>
        </LNR>
        </NSTG>
        </NSTGO>
        </PN>
        </RN>
        </LNR>
        <NSTG>
        </NSTGT>          <===== mang nghĩa thời gian
</SA>
<SUBJECT> ...         <===== câu chưa đủ... thiếu chủ...
<VERB> ...
<OBJECT> ...
<SA> ... </SA>
</SENTENCE>
    
```

Phân tích ngữ pháp này cho phép chúng ta đặt cả câu 6 “*trăm năm trong cõi người ta*” mang nghĩa thời gian cho câu chính sắp xảy ra.

Muốn làm được điều này, các tự/từ điển (và cơ sở dữ liệu) phải cho thêm thông tin, như “*trăm*”: lượng, “*năm*”: danh, “*trong*”: giới từ, “*cõi*”: danh, “*người ta*”: danh, v.v. Thông tin ngữ pháp ngày nay có thể rút ra từ vị trí của từ trong câu.

Phân tích ngữ pháp cho phép chúng ta gắn thời gian, nhân vật (chủ), tên người, cây cỏ, địa lý, v.v. vào một văn bản, để có thể rút ra thông tin ấy khi cần.

3.2.3 Cách tiếp cận toàn diện tới nghiên cứu về chữ Nôm và tiếng Việt

Có thể coi chữ Nôm như nơi chôn giấu của nhiều chứng tích cổ của tiếng Việt: những cách nói cổ, những cách diễn đạt ý, và tri thức cộng đồng ngày xưa trong các tác phẩm cổ. Những tác phẩm đó đã được cố định lại trong các hình chữ Nôm đó, gần như một dạng cô đọng cho các quan niệm, tư tưởng, cách nghĩ, văn hoá của người Việt. Nay với công cụ xử lý thông tin hiện đại, với việc bỏ công sức đưa tất cả vốn văn hoá cổ của cha ông vào lưu trữ dưới dạng điện tử, các nhà nghiên cứu về sau sẽ có khả năng vẽ lại, hình dung lại các thời lịch sử ngày xưa thông qua ngôn ngữ.

Mặt khác, với việc đã mã hoá được hầu hết các âm tiết của tiếng Việt qua chữ Nôm, có thể đặt vấn đề nghiên cứu tiếng Việt không chỉ dưới dạng chữ viết mà còn là dưới dạng mã hoá cho các âm tiết tiếng nói. Ví dụ, vì “lí” và 里 là một chữ, mỗi chữ có một mã duy nhất, mỗi mã là một tiếng, mỗi tiếng là một âm tiết, nên mã U+091cc là mã quốc tế độc nhất cho phép ta kèm theo âm đọc chuẩn, âm đọc Hà Nội, âm đọc địa phương, âm đọc cổ, v.v.

Một ví dụ nữa để cụ thể hoá cách giải quyết văn bản học của những bản Kiều khác nhau bằng phương pháp CNTT.

Chúng ta có thể nghĩ một bản Kiều trên mạng có cấu trúc như sau theo chuẩn eXtensible Markup Language hay XML (một biến thể của SGML) của Tổ chức Chuẩn Quốc tế ISO:

```

<sách>
  <tựa>Truyện Kiều
    <thưviện> ... các thông tin thư viện ... </thưviện>
    <baotang> ... các thông tin thư viện ... </baotang>
    <soạngiả> ... </soạngiả>
    ... các thông tin khác ...
  </tựa>
  <nộidung>
    <trang số=1>
      <ảnh nguồn=ảnhtrang1.jpg>
      <dòng số=1>
        <Nôm>
          <chữ số=1> 羸 </chữ>
          <chữ số=2> 輜 </chữ>
          <chữ số=3> 蝨 </chữ>
          <chữ số=4> 增 </chữ>
          <từ><chữ số=5> 馱 </chữ>
          <chữ số=6> 些 </chữ></từ>
        </Nôm>
        <quốcngữ>
          <chữ số=1>trăm</chữ>
          <chữ số=2>năm</chữ>
          <chữ số=3>trong</chữ>
          <chữ số=4>cõi</chữ>
          <từ><chữ số=5>người</chữ>
          <chữ số=6>ta</chữ></từ>
        </quốcngữ>
        <tiếngAnh>
          <chữ số=1>hundred</chữ>
          <chữ số=2>year</chữ>
          <chữ số=3>in</chữ>
          <chữ số=4>realm, world,...</chữ>
          <chữ số=5-6>human; human being;...</chữ>
        </tiếngAnh>
      </dòng>
      <dòng số=2>... vân vân ... </dòng>
    </trang>
    <trang số=2> ... vân vân ... </trang>
  </nộidung>
</sách>

```

Loại tổ chức mới này có thể làm tự động, khi dùng, người xem tùy ý trình bày theo hàng dọc hay hàng ngang, đối chứng từng chữ một, hoặc từng câu, hoặc từng trang, v.v. cũng như theo

dõi tất cả hoạt động của một nhân vật, ví dụ như Mục Tú bà. Hoặc so sánh những bản khác nhau của Kiều cùng trang, cùng dòng, cùng chủ đề, v.v.

Cách tổ chức mới như trên khi áp dụng vào văn bản sử, chúng ta có thể từ văn bản sử ấy tìm được tất cả những gì xảy ra trong một năm nào đó của lịch sử (nếu quyển sử có nói), hoặc tất cả các hoạt động của một nhân vật, ví dụ như Lê Thánh Tông, ... hoặc những điều đã ghi về một địa điểm lịch sử như Bạch Long Vĩ, v.v.

4. Mở rộng phạm vi ứng dụng CNTT và chữ Nôm

Việc công nghệ thông tin cho phép thể hiện chữ Nôm và các văn bản chữ Nôm trên máy tính và trên mạng toàn thế giới đã mở rộng thêm các lĩnh vực ứng dụng có liên quan tới CNTT và chữ Nôm. Trước hết phải kể đến các cơ sở bảo tồn bảo tàng hiện đang lưu giữ nhiều hiện vật quý báu có ghi lại nhiều thông tin bằng chữ Hán Nôm. Những cơ sở này cần triển khai các ứng dụng CNTT giúp cho việc lưu trữ và tìm kiếm thông tin viết bằng chữ Hán Nôm trên các hiện vật đó. Các dự án sử dụng CNTT trong bảo quản, khai thác và phổ biến di sản phi vật thể chữ Nôm cần được lập ra và triển khai dựa trên những thành tựu mà các phần mềm và CNTT cung cấp.

Các cơ sở thư viện có sách báo, tư liệu chữ Nôm sẽ cần triển khai các ứng dụng xử lý sách báo có chữ Hán Nôm, lập thư mục, làm tóm tắt, cung cấp dạng số thức của các tác phẩm này. Nhiều ứng dụng CNTT liên quan tới thư viện và chữ Nôm cần được phát triển. Các công ti CNTT cần được khuyến khích và đặt hàng triển khai các phần mềm giúp cho công tác quản lý và phục vụ dịch vụ thư viện với chữ Nôm.

Các cơ sở nghiên cứu liên quan tới chữ Nôm cần đẩy mạnh việc sử dụng thông tin chung về chữ Nôm đã được lưu trữ trên máy tính. Do việc nhiều người nghiên cứu chữ Nôm có thể còn chưa quen thuộc với cách làm việc trên máy tính, có thể cần thành lập một số công ti dịch vụ để hỗ trợ họ trong việc truy nhập vào nguồn tài nguyên thông tin chữ Nôm đã sẵn có trên máy tính.

Để thúc đẩy các nghiên cứu CNTT về chữ Nôm, nên tổ chức trao giải thưởng cho các nghiên cứu chữ Nôm và CNTT của giới trẻ. Một ví dụ, giải nhận dạng chữ Nôm.

Các hoạt động tôn giáo cũng có nhu cầu cần tới việc xử lý chữ Nôm trong các văn bản kinh sách cổ. Việc in lại các kinh sách cổ đó cũng là một trong những nhu cầu mà các nhóm về CNTT và chữ Nôm có thể đáp ứng được. Các bản kinh sách cổ cho phép chúng ta hình dung được các cách đọc chữ Nôm trong các thời đại, ví dụ, việc phiên âm các chữ Phạn trong các bản dịch Nôm trong đạo Phật hoặc chữ La-tinh hay Âu châu trong các bản dịch Nôm kinh thánh và truyện các thánh Ki-tô.

5. Thành lập nhóm kho chuẩn Hán Nôm

Kho tàng chữ Hán Nôm của dân tộc bao gồm nhiều tác phẩm, văn, bia, kí... vẫn chưa được phiên chuyển hết sang quốc ngữ. Có khả năng nhiều chữ Hán Nôm xuất hiện trong các văn bản đó vẫn còn chưa được phát hiện ra và chưa được đưa vào kho chữ Hán Nôm. Vì vậy một trong những nhiệm vụ cần được tiếp tục trong thời gian dài nữa là phát hiện chữ mới đưa vào kho chữ Hán Nôm và tiến hành các hoạt động chuẩn hoá đối với chúng để đưa vào máy tính. Việc này

cần một nhóm chuyên gia am hiểu cả về Hán-Nôm và CNTT thực hiện. Sự tham gia của mọi cá nhân và tập thể vào nỗ lực này là điều rất cần thiết và quan trọng. Đính kèm bài báo này là các mẫu đăng kí tham gia nhóm và mẫu thu thập các dạng chữ Nôm còn chưa có trong kho chữ.

6. Phát triển phần mềm nguồn mở cho chữ Nôm

Vì chữ Nôm là chữ của dân tộc ta, cho nên những người am hiểu chữ Nôm nhất vẫn sẽ là những người Việt. Do đó việc phát triển các phần mềm giúp cho khai thác và sử dụng chữ Nôm sẽ là một nhu cầu tất yếu của nước ta mà không có ở các nước khác. Cần phải có chiến lược phát triển phần mềm nguồn mở cho chữ Nôm bởi lí do đơn giản là không có các phần mềm sở hữu riêng của các công ti dành cho chữ Nôm. Thực tế người Việt sẽ cần phải phát triển các phần mềm đặc thù cho chữ Nôm.

Tuy thế, ngay cả việc chúng ta phải phát triển phần mềm đặc thù cho chữ Nôm, dùng phần mềm nguồn mở, thì điều đó vẫn không có nghĩa là chúng ta tự mình đóng cửa để xử lí lấy vấn đề của mình. Cần phải luôn giữ chiến lược tích hợp phần mềm chữ Nôm vào các phần mềm quốc tế, tận dụng những thành tựu phát triển của giới phần mềm quốc tế để giải quyết thêm cho các vấn đề của chữ Nôm.

7. Một số nhiệm vụ trước mắt

Trên cơ sở xác định những tầm nhìn như vậy cho chữ Nôm, một số nhiệm vụ trước mắt có thể cần được hoàn thành gấp.

7.1. Hoàn chỉnh và bổ sung bộ font chữ Nôm, hoàn chỉnh cơ sở dữ liệu chữ Nôm

Cần có những khuyến khích từ phía nhà nước và cá nhân để hoàn chỉnh, trao đổi, và bổ sung các bộ phông chữ Hán-Nôm trong cả nước và các nơi có nhu cầu nghiên cứu trên thế giới. Hoàn chỉnh cơ sở dữ liệu trên mạng để hình thành một cơ sở dữ liệu lớn về chữ Nôm qua các thời đại, theo từng phương ngữ, tác phẩm, tư liệu có thật trong lịch sử.

7.2. Nhận dạng chữ Nôm

So với kho chữ Nôm khổng lồ chưa khai thác, đã bị chiến tranh huỷ hoại, và đang bị thời gian huỷ hoại, trong khi các chuyên gia chữ Nôm ngày càng mai một, việc khuyến khích sử dụng CNTT để nhận dạng chữ Nôm tự động là cấp bách. Nhận dạng chữ Nôm có nghĩa là từ một hình chữ đưa vào máy quét ảnh, các chương trình ấy biến các ảnh khung vuông thành các mã chữ theo thứ tự nguyên bản. Kỹ thuật này cho đến nay tuy chưa hoàn hảo, nhưng nó rất có ích trong việc hỗ trợ các chuyên gia đưa văn bản ấy thành văn bản xử lý ngôn ngữ học và tri thức được.

7.3. Xây dựng hệ thống đăng kí và thu thập các chữ Nôm mới

Tại Hội nghị này, chúng tôi kêu gọi sự tham gia của các học giả Hán-Nôm vào Nhóm Kho Chuẩn Hán-Nôm, mà nhiệm vụ là để nhanh chóng làm kho chữ Hán-Nôm chuẩn đủ để đáp ứng yêu cầu nghiên cứu Hán-Nôm ở bất cứ nơi nào trên thế giới.

7.4. Phổ biến chương trình bàn phím thuận tiện để đưa chữ Nôm vào các tài liệu

Việc lập trình cho bàn phím chữ Hán-Nôm đương nhiên thừa hưởng kết quả nghiên cứu của các nước sử dụng chữ biểu ý. Ở đây, việc nghiên cứu Hán-Nôm cũng cần có những chương trình nhằm thúc đẩy kỹ thuật để chúng ta theo kịp các nước và sử dụng được lợi thế của chữ quốc ngữ la-tinh.

7.5. Xây dựng các chương trình giảng dạy về sử dụng CNTT trong các ứng dụng có chữ Nôm

Những chương trình giảng dạy về ứng dụng CNTT này có tính chiến lược vì mọi hoạt động CNTT trên thế giới đều tuân thủ chuẩn quốc tế. Ở đây, chuẩn đa ngữ tiến xa hơn một bước vì nó đòi hỏi một cái nhìn toàn diện và toàn cầu hoá. Ví dụ, ngày nay lập trình CNTT, ngoài việc theo dõi chuẩn, người lập trình sử dụng chuẩn để mang người sử dụng tham gia làm vững mạnh phần mềm của mình.

7.6. Phát triển chương trình dạy chữ Nôm trên máy tính và mạng

Giảng dạy chữ Hán-Nôm ngày nay không còn chỉ nằm trên sách và giấy mực, mặc dù điều này không thể thiếu. Giảng dạy Hán-Nôm trên máy tính và mạng mới lôi cuốn được người trẻ và làm cho việc học Hán-Nôm có tương lai. Ý nghĩa kinh tế của việc học Hán-Nôm đi đôi chiến lược với việc truy cập, cài đặt, nghiên cứu các tư liệu Hán-Nôm trên mạng.

7.7. Xây dựng hệ thống đăng kí tác phẩm chữ Nôm đang có

Do những đặc trưng lịch sử của các triều đại phong kiến trước đây, những tác phẩm Hán-Nôm ít tập trung, mà tản mạn, giấu kín trong các dòng họ gia đình. Việc đăng ký tác quyền các tác phẩm ấy về mặt kinh tế khuyến khích việc sử dụng các tác phẩm ấy rộng rãi, tri thức ở trong các tác phẩm ấy nhờ đó làm giàu thêm cho xã hội.

7.8. Xây dựng kho các văn bản cổ dạng điện tử dùng chung trên mạng cho cộng đồng

Công tác này cấp bách và đi song song với việc tổ chức thành công quy trình chuẩn về nhận dạng và bàn phím, để đưa các văn bản cổ lên mạng cộng đồng, cứu nguy được sự hư hại vĩnh viễn của thời gian. Ta có thể gọi đây là một chương trình quốc gia đưa các văn bản quý giá lên mạng.

7.9. Lập diễn đàn, hội nghị, tiếp tục trao đổi và thảo luận về chữ Nôm cổ

Những vấn đề chữ Nôm về lý thuyết và văn bản học mãi mãi còn tồn tại cho đến khi chúng ta có được một phương án tự động phân tích cú pháp chữ Nôm. Sự trao đổi này nhằm tiến đến khả năng xử lý văn bản ở một số lượng lớn. Việc tổ chức thường niên các hội nghị quốc tế về chữ Nôm là một biện pháp thúc đẩy các nghiên cứu và ứng dụng CNTT vào chữ Nôm phục vụ cho công tác bảo tồn di sản cũng như phổ cập.

8. Kết luận

Nhu cầu trở về với văn hoá dân tộc là động lực chính cho chữ Nôm trường tồn với sự phát triển đi lên của dân tộc. Các tiến bộ của CNTT thực sự đã mở ra những hướng hỗ trợ cho việc bảo tồn và phổ biến chữ Nôm.

Nhu cầu của các cơ sở văn hoá trong nước và quốc tế có tài liệu liên quan tới chữ Nôm là động lực thúc đẩy các nghiên cứu hiện đại về chữ Nôm.

Việc thay đổi phương pháp luận làm việc với các vấn đề liên quan tới chữ Nôm và CNTT là điều rất quan trọng và cần thiết để mở ra một thời kì mới cho các nghiên cứu về chữ Nôm. Các hạ tầng cơ sở cho chữ Nôm đã, đang và sẽ được phát triển.



TÀI LIỆU THAM KHẢO

1. Các tập mã chữ Nôm do Việt Nam cung cấp cho nhóm ISO 10646 JTC1/IRG từ năm 1994 đến nay: NPCT 2.1, TCVN 5712: 1993, TCVN 5773: 1993, TCVN 6056: 1995, VHN1: 1998, VHN2: 1998. Đề nghị CJK Extension C1 của Việt Nam
2. ISO 10646 JTC1/IRG từ năm 1994 đến nay, Unihan 3.1 Radical-Stroke Index.
3. Đỗ Quốc Bảo, các bộ phong HanNom.
4. Linh mục Trần Văn Kiệm. *Giúp đọc Nôm và Hán Việt*. Nhà xuất bản Thuận Hóa, 1999.
5. Lê Mai Phương. *Học chữ Nôm theo Tam thiên tự*. Trình bày tại Hội thảo Hè 2002, Đại học Maine.
6. Ngô Thanh Nhân, Ngô Trung Việt và Nhóm Nôm Na. *Qui trình Nôm Na*, trình bày tại Hội thảo hè 2002, Đại học Maine.
7. Ngô Thanh Nhân. *Đơn vị chính tả và các đặc điểm của tiếng Việt: chữ quốc ngữ, hệ latin, chữ Nôm, hệ biểu ý và Unicode/ISO IEC 10646*, Ủy ban Unicode/ISO 10646 Việt Nam. 2001.
8. Ngô Trung Việt, Trần Lưu Chương, Ngô Thanh Nhân, Đỗ Bá Phước, Nguyễn Quang Hồng, Nguyễn Hoàng, Lê Phạm Ngung Hương, Hà Dương Tuấn, Đỗ Tuyết Khanh & Hồ Văn Tiên: *Phát triển phần mềm chữ Nôm*. Hội thảo Hè 2001 tại Aix-en-Provence, Pháp, 2001.
9. Nguyễn Quang Hồng, Ngô Thanh Nhân, Đỗ Bá Phước and Ngô Trung Việt: Chữ Nôm: Văn hoá cổ truyền và thông tin hiện đại. *Tạp chí Ngôn Ngữ* 4(1999): 11-22. Hà Nội.
10. Viện Ngôn ngữ học. *Bảng tra chữ Nôm*. Nhà xuất bản Khoa học Xã hội Hà Nội 1976.
11. Vũ Văn Kính & Nguyễn Quang Xỷ. *Tự Điển Chữ Nôm*. Trung tâm Học liệu. Sài Gòn 1971.
12. Vũ Văn Kính. *Đại tự điển chữ Nôm*. Nhà xuất bản Văn Nghệ Tp Hồ Chí Minh & Trung tâm nghiên cứu Quốc học Huế. 1999.
13. Nguyễn Quang Hồng (chủ biên). *Lời dẫn Tự Điển Chữ Nôm Tiếng Việt*. Viện nghiên cứu Hán Nôm. Đang in.
14. Trương Đình Tín. *Bảng Phiên âm Nôm Việt*. Nhà xuất bản Thuận Hóa. 2003.
15. Sager, N. *Natural Language Information Processing: A Computer Grammar of English and Its Applications*. Addison-Wesley, Reading, MA. 1981.