

## PHÂN TÍCH TẦN SỐ VÀ HIỆU SUẤT SỬ DỤNG CỦA CHỮ NÔM TỰ TẠO TRONG NĂM BỘ TỪ THƯ HÁN NÔM CỔ HƯỚNG ĐẾN VIỆC HÌNH THÀNH BẢNG PHÂN CẤP

HUỠNH ĐỨC DU\*

TRẦN TUYẾT KHANH\*\*

*Tóm tắt:* Nghiên cứu này phân tích tần số và hiệu suất sử dụng của chữ Nôm tự tạo trong năm bộ từ thư Hán Nôm cổ của Việt Nam, bao gồm *Chỉ Nam ngọc âm giải nghĩa*, *Nhật dụng thường đàm*, *Tự Đức thánh chế tự học giải nghĩa ca*, *Đại Nam quốc ngữ* và *Nam phương danh vật bị khảo* nhằm đánh giá số lượng, quy luật sử dụng và đặc điểm hiệu dụng, hướng đến việc đề xuất một bảng phân loại chữ Nôm tự tạo thường dùng và thông dụng. Bằng phương pháp định lượng, nghiên cứu so sánh tần suất, tỷ lệ bao phủ của chữ Nôm tự tạo và chữ Hán, làm rõ sự khác biệt về phạm vi sử dụng và mức độ chuẩn hóa. Kết quả cho thấy, chữ Hán có xu hướng được ưu tiên trong các khái niệm trừu tượng, chuyên môn, trong khi chữ Nôm tự tạo thường được dùng cho từ ngữ thông dụng. Từ đó, nghiên cứu đề xuất bảng phân loại chữ Nôm tự tạo theo cấp độ thường dùng và thông dụng, cung cấp công cụ hữu ích cho việc học tập và nghiên cứu chữ Nôm.

*Từ khóa:* Chữ Nôm tự tạo, tần số chữ, hiệu suất sử dụng, từ thư Hán Nôm cổ, phân loại chữ Nôm

*Abstract:* This study analyzes the frequency and usage efficiency of self-created Nôm script within five classical Sino-Vietnamese dictionaries, including the *Chỉ Nam Ngọc Âm Giải Nghĩa*, *Nhật Dụng Thường Đàm*, *Tự Đức Thánh Chế Tự Học Giải Nghĩa Ca*, *Đại Nam Quốc Ngữ*, and *Nam Phương Danh Vật Bị Khảo*. The investigation assesses the quantity, usage patterns, and effectiveness of these characters, aiming to propose a classification table of frequently used and generally applicable self-created Nôm script. Employing quantitative methods, the research compares the frequency and coverage rate of self-created Nôm script against standard Chinese characters, elucidating differences in usage scope and standardization levels. Results indicate that Chinese characters tend to be favored for abstract and specialized concepts, while self-created Nôm script is more commonly used for everyday vocabulary. Consequently, the study proposes a classification table of self-created Nôm script based on frequency and general applicability, offering a valuable tool for learning and researching Nôm script.

*Keywords:* Self-created Nôm script, character frequency, usage efficiency, Sino-Vietnamese dictionaries, Nôm script Classification

### 1. Dẫn luận

Chữ Nôm là hệ chữ viết độc đáo của nước ta, vừa có quan hệ với chữ Hán, vừa mang những đặc điểm riêng. *Vân Đài loại*

*ngữ: Âm tự ngữ* chép rằng: “Vạn quốc ở hải ngoại; non sông xa tít, khí trời chỗ nhẹ,

\* TS., Trường Đại học Nguyễn Tất Thành

\*\* ThS., Trường Đại học Nguyễn Tất Thành

chỗ nặng, chỗ trong, chỗ đục, khác nhau; nên tiếng nói cũng khó phân biệt; phải sáng tạo ra văn tự riêng (海外萬國，山川僻遠，輕重清濁之氣隨處不同。其語尤為難辨，故各自制文字)”<sup>1</sup>. Chữ Hán từng là văn tự chính thức trong học thuật và hành chính, trong khi chữ Nôm, mặc dù xuất hiện từ sớm, chủ yếu được sử dụng để ghi lại ngôn ngữ dân gian. Đến thế kỷ XVII đến thế kỷ XIX, chữ Nôm đã đạt đến giai đoạn phát triển mạnh mẽ.

So với Nhật Bản và Triều Tiên, chữ viết Việt Nam phát triển theo hướng khác biệt. Nhật Bản tạo chữ Kana, Triều Tiên sáng tạo chữ Hangeul, trong khi Việt Nam sử dụng cấu trúc chữ Hán để sáng tạo chữ Nôm. Điều này giúp chữ Nôm có đặc điểm biểu âm, biểu nghĩa tương đồng chữ Hán, nhưng về hình thức lại phức tạp hơn cả chữ Hán, Kana và Hangeul. Sự phát triển của chữ Nôm kéo theo việc gia tăng số lượng chữ mới và chữ dị thể.

Những năm gần đây, nghiên cứu định lượng đã ngày càng được ứng dụng rộng rãi trong lĩnh vực nghiên cứu chữ Nôm. Một số công trình tiêu biểu bao gồm các nghiên cứu như *Nghiên cứu về hệ thống ngữ âm Hán Việt và chữ Nôm* 漢越語音系與喃字研究 (2012) và *Nghiên cứu định lượng chữ Nôm: Trường hợp Quốc âm thi tập và Đoạn trường tân thanh* 喃字定量研究——以《國音詩集》和《斷腸新聲》為例 (2023) của Nguyễn Đình Hiền; *Thu thập và phân tích Tự học toàn yếu và Chỉ Nam giải âm* 《字學纂要》與《指南解音》整理分析 (2017) của Shi Weikun 史維坤; *Nghiên cứu chữ Nôm trong Tự Đức thánh chế tự học giải nghĩa ca* 《嗣德聖製字學解義歌》喃字研究 (2019) và

*Nghiên cứu về Từ điển cổ Hán Nôm Việt Nam và Các chữ, từ ngữ được thu thập* 越南漢喃古辭書及所收字詞研究 (2023) của Huỳnh Đức Dur; *Nghiên cứu từ điển song ngữ Hán Nôm Nam phương danh vật bị khảo* 漢喃雙語詞典《南方名物備考》研究 (2020) của Ứng Thùy Linh và *Nghiên cứu về cách dùng chữ trong tác phẩm văn học chữ Nôm Lục Vân Tiên diễn ca của Việt Nam* 越南喃字文學作品〈陸雲僊歌演〉用字研究 (2021) của He Meiyu. Những nghiên cứu này thực hiện thống kê toàn diện về chữ Nôm trong các văn bản cổ, giúp hình thành một cái nhìn tổng quan về hệ thống chữ Nôm.

Bài viết này tập trung vào việc nghiên cứu chữ Nôm tự tạo, tức là những chữ do người Việt sáng tạo ra. Thông qua phân tích thống kê và khai thác dữ liệu, chúng tôi đánh giá số lượng, quy luật sử dụng, và hiệu suất sử dụng của loại chữ này. Trước đó, bài viết *Tự Đức thánh chế tự học giải nghĩa ca: Từ điển song ngữ Hán Nôm thánh chế theo môn loại có tính bách khoa thư* (2013) của Phạm Văn Khoái, Hà Đăng Việt và luận án *Nghiên cứu các tự điển, từ điển Hán Nôm thế kỷ XIX* (2014) của Hà Đăng Việt đã tiến hành thống kê mục từ chữ Hán trong năm bộ từ thư Hán Nôm, làm nổi bật vai trò của *Tự Đức Thánh chế tự học giải nghĩa ca* 嗣德聖製字學解義歌.

Nghiên cứu này tiếp tục mở rộng phân tích định lượng về chữ Nôm tự tạo trong năm bộ từ thư quan trọng, bao gồm *Chỉ Nam ngọc âm giải nghĩa* 指南玉音解義 (CN), *Nhật dụng thường đàm* 日用常談 (ND), *Tự Đức Thánh chế tự học giải nghĩa ca* 嗣德聖製字學解義歌 (TĐ), *Đại Nam quốc ngữ* 大南國語 (ĐN) và *Nam phương*

*danh vật bị khảo* 南方名物備考 (BK). Chúng tôi tiến hành đối chiếu theo cả chiều ngang (giữa các từ thư) và chiều dọc (theo thời gian) để làm rõ sự biến đổi của chữ Nôm tự tạo. Bằng cách thống kê tần số xuất hiện của chữ Nôm tự tạo, chúng tôi tổng hợp và phân loại theo từng giai đoạn lịch sử, nhằm làm rõ xu hướng phát triển của nhóm chữ này.

Dựa trên kết quả phân tích, bài viết đề xuất một Bảng chữ Nôm tự tạo thường dùng và thông dụng, giúp người học tiếp cận hệ thống chữ Nôm một cách có hệ thống.

## 2. Phân tích và khảo sát tần số chữ

### 2.1 Phân tích tần số chữ

Tần số chữ được định nghĩa là tỷ lệ giữa số lần xuất hiện của một dạng chữ trong một ngữ liệu cụ thể so với tổng số đơn vị văn bản.<sup>2</sup> Trong nghiên cứu này, tần số chữ được hiểu là tỷ lệ số lần xuất hiện của chữ Nôm tự tạo trong tổng số chữ được sử dụng trong năm bộ từ thư Hán Nôm cổ của Việt Nam.

Thống kê tần số chữ đóng vai trò quan trọng trong nghiên cứu văn tự vì nó không chỉ phản ánh tình hình sử dụng các dạng chữ trong toàn bộ văn bản hoặc trong từng giai đoạn lịch sử mà còn giúp tổng hợp hiệu suất sử dụng của từng loại chữ, so sánh sự khác biệt giữa chúng. Ngoài ra, thống kê tần số còn giúp xác định chữ thường dùng, chữ ít dùng và chữ hiếm dùng, từ đó làm rõ xu hướng phát triển và biến đổi của các nhóm chữ qua từng thời kỳ.

So với nghiên cứu về số lượng chữ, nghiên cứu về tần số chữ liên quan đến chữ Nôm trong các văn bản Hán Nôm của Việt Nam còn tương đối hạn chế. Vì vậy, việc

thống kê và phân tích tần số của chữ Nôm tự tạo trong năm bộ từ thư Hán Nôm cổ là cần thiết. Nguyên tắc thống kê tần số được thực hiện tương tự như thống kê số lượng chữ, trong đó các chữ dị thể và chữ dị dụng vẫn được thống kê riêng biệt.

Trong nghiên cứu này, chúng tôi định nghĩa và phân tích ba khái niệm quan trọng liên quan đến việc phân tích tần suất xuất hiện của chữ Nôm tự tạo trong các từ thư cổ: tần suất xuất hiện (TSXH), tần số chữ (TSC), và tỷ lệ bao phủ (TLBP).

Tần suất xuất hiện là số lần một chữ xuất hiện trong từ thư. Do tính chất phức tạp của văn bản từ thư, chúng tôi phân loại tần suất xuất hiện theo các khoảng: 1 lần, 2-5 lần, 6-10 lần, 11-20 lần, 21-50 lần, 51-100 lần và trên 100 lần.

Tần số chữ là tỷ lệ phần trăm giữa tần suất xuất hiện của một chữ và tổng tần suất xuất hiện của toàn bộ nhóm chữ trong từ thư. Công thức tính tần số chữ như sau:

$$TSC (\%) = \frac{\text{TSXH của một chữ}}{\text{Tổng TSXH của nhóm chữ}} \times 100$$

Ví dụ, trong *Chi Nam ngọc âm giải nghĩa*, tổng tần suất xuất hiện của nhóm chữ Nôm tự tạo là 1298, chữ 𠵹 xuất hiện 169 lần, tần số chữ là:

$$TSC (\%) = \frac{169}{1298} \times 100 = 13,02\%$$

Tỷ lệ bao phủ được áp dụng cho từng khoảng tần suất xuất hiện, phản ánh tỷ lệ phần trăm giữa tổng tần suất xuất hiện của tất cả các chữ thuộc một khoảng tần suất nhất định so với tổng tần suất xuất hiện của nhóm chữ đó trong từ thư. Công thức tính tỷ lệ bao phủ như sau:

$$TLBP (\%) = \frac{\text{Tổng TS của nhóm chữ trong một khoảng tần suất}}{\text{Tổng TSXH của nhóm chữ}} \times 100$$

Ví dụ, trong *Đại Nam quốc ngữ*, tổng tần suất xuất hiện của chữ Nôm tự tạo là 2276, trong đó các chữ chỉ xuất hiện một lần có tổng tần suất là 246, suy ra tỷ lệ bao phủ của nhóm này là:

$$TLBP (\%) = \frac{246}{2276} \times 100 = 10,81\%$$

Những khái niệm và công thức này giúp chúng tôi phân tích và diễn giải dữ liệu một cách chính xác và hiệu quả.

## 2.2 Khảo sát tần số chữ

Bảng 1: Tổng quan khảo sát tần suất chữ Nôm tự tạo

TSXH (lần)	Tổng TSXH					TLBP (%)				
	CN	ND	TĐ	BK	ĐN	CN	ND	TĐ	BK	ĐN
1	126	271	583	230	246	9,71	16,01	7,40	15,43	10,81
2-5	255	465	1256	364	505	19,65	27,47	15,94	24,41	22,19
6-10	215	305	922	223	341	16,56	18,02	11,70	14,96	14,98
11-20	287	279	889	191	523	22,11	16,50	11,28	12,81	22,98
21-50	359	317	1631	264	661	27,66	18,72	20,70	18,74	29,04
51-100	56	56	1051	219	—	4,31	3,28	13,34	13,65	—
> 100	—	—	1546	—	—	—	—	19,54	—	—
Tổng	1298	1693	7878	1491	2276	100,00	100,00	100,00	100,00	100,00

Tổng tần suất xuất hiện của chữ Nôm tự tạo trong năm bộ từ thư Hán Nôm lần lượt là 1298 lần, 1693 lần, 7878 lần, 1491 lần, và 2276 lần. Trong *CN* và *ND*, chữ có tần suất xuất hiện thấp nhất đều chỉ 1 lần, trong khi chữ có tần suất cao nhất đều lên đến 56 lần. Điều này cho thấy một hiện tượng mất cân đối: chữ có tần suất cao thường ít về số lượng nhưng có hiệu suất sử dụng cao, trong khi chữ có tần suất thấp lại chiếm số lượng lớn nhưng hiệu suất thấp.

Dựa vào bảng khảo sát tần suất chữ, có thể thấy thông tin chi tiết về các bộ từ thư như sau: Bộ *CN* có tổng cộng 1298 lần xuất hiện, với tỉ lệ bao phủ cao nhất là 27,66% ở khoảng 21-50 lần, tiếp theo là 19,65% ở khoảng 2-5 lần, gồm các chữ như và 22,11% ở khoảng 11-20 lần. Bộ *ND* có tổng cộng

1693 lần xuất hiện, tỉ lệ bao phủ cao nhất là 27,47% ở khoảng 2-5 lần, tiếp theo là 18,72% ở khoảng 21-50 lần và 18,02% ở khoảng 6-10 lần. Bộ *TĐ* xuất hiện tổng cộng 7878 lần, với tỉ lệ bao phủ cao nhất là 20,70% ở khoảng 21-50 lần, tiếp theo là 13,34% ở khoảng 51-100 lần và 19,54% ở khoảng trên 100 lần. Bộ *ĐN* có tổng cộng 1491 lần xuất hiện, với tỉ lệ bao phủ cao nhất là 24,41% ở khoảng 2-5 lần, tiếp theo là 18,74% ở khoảng 21-50 lần và 15,43% ở khoảng 1 lần. Cuối cùng, bộ *BK* xuất hiện tổng cộng 2276 lần, với tỉ lệ bao phủ cao nhất là 29,04% ở khoảng 21-50 lần, tiếp theo là 22,98% ở khoảng 11-20 lần và 22,19% ở khoảng 2-5 lần.

Quan sát kỹ hơn, ta thấy chữ có tần suất cao thường có số nét ít và cấu trúc đơn

giản, trong khi chữ có tần suất thấp lại có số nét nhiều và cấu trúc phức tạp. Ví dụ, nhóm chữ có tần suất 21-50 lần thường có nét đơn giản như 𦵑 (rễ), 𦵑 (đá), 𦵑 (trời), 𦵑 (vóc), trong khi nhóm chữ xuất hiện 1 lần lại có nét phức tạp như 𦵑 (vây), 𦵑 (nách), 𦵑 (sùng), 𦵑 (gió).

Dựa trên tần suất xuất hiện của từng chữ Nôm tự tạo trong năm bộ từ thư, các chữ này có thể được phân loại theo mức độ tần suất cao, trung cao, trung bình, trung thấp và thấp. Sự phân loại này giúp làm rõ xu hướng sử dụng và hiệu suất của từng nhóm chữ trong các văn bản cổ.

Bảng 2: Phân loại tần suất chữ Nôm tự tạo trong năm bộ từ thư Hán Nôm cổ

		Cao tần	Trung cao tần	Trung tần	Trung thấp tần	Thấp tần
CN	Tổng số lượng	56	359	502	255	126
	Phạm vi TSC (%)	3,86-4,31	1,55-3,85	0,40-1,54	0,15-0,39	0,08-0,14
	Tổng TLBP (%)	4,31	27,66	38,67	19,65	9,71
	Tổng lượng chữ đơn	1	11	51	88	126
ND	Tổng số lượng	56	317	584	465	271
	Phạm vi TSC (%)	2,60-2,85	1,07-2,59	0,31-1,06	0,10-0,30	0,05-0,09
	Tổng TLBP (%)	3,28	18,72	34,52	27,47	16,01
	Tổng lượng chữ đơn	1	11	59	162	271
TD	Tổng số lượng	2597	1631	1811	1256	583
	Phạm vi TSC (%)	0,65-2,36	0,27-0,64	0,08-0,26	0,03-0,07	0,01-0,02
	Tổng TLBP (%)	32,88	20,70	22,98	15,94	7,40
	Tổng lượng chữ đơn	26	51	183	445	583
BK	Tổng số lượng	219	264	414	364	230
	Phạm vi TSC (%)	3,42-4,96	1,41-3,41	0,40-1,40	0,13-0,39	0,01-0,12
	Tổng TLBP (%)	13,65	18,74	27,77	24,41	15,43
	Tổng lượng chữ đơn	3	8	43	131	230
ĐN	Tổng số lượng	611	523	341	505	246
	Phạm vi TSC (%)	0,92-2,24	0,48-0,91	0,26-0,47	0,09-0,25	0,01-0,08
	Tổng TLBP (%)	29,04	22,98	14,98	22,19	10,81
	Tổng lượng chữ đơn	22	35	46	170	246

Dựa trên kết quả khảo sát, có thể thấy rằng các chữ có tần suất cao (từ 51-100 lần) thường là những từ có ý nghĩa khái quát, bao quát các khái niệm rộng. Ví dụ, các từ chỉ động vật như 𦵑 (trâu), các khái niệm về thiên nhiên như 𦵑 (trời), hay hành động như 𦵑 (làm). Những chữ này thường có số nét ít và cấu trúc đơn giản, điều này góp phần vào việc minh chứng chúng được sử dụng rộng rãi trong văn bản cổ. Các chữ có

tần suất trung bình (từ 6-50 lần) thường liên quan đến các từ có ý nghĩa chi tiết hơn, đặc trưng hoặc có sự phân biệt rõ ràng, như 𦵑 (vây), 𦵑 (giặt). Những chữ này đóng vai trò quan trọng trong việc thể hiện sự đa dạng và phong phú của ngôn ngữ trong văn bản cổ. Các chữ có tần suất thấp (dưới 5 lần) chủ yếu là những từ ít xuất hiện, có nghĩa chi tiết hoặc mang tính đặc trưng, phân biệt rõ ràng. Ví dụ như các từ có đặc điểm khu biệt

hoặc từ lặp lại, như 淋 (rằm), 穰 (riêng). Những chữ này thường có số nét nhiều và cấu trúc phức tạp, điều này có thể giải thích lý do chúng ít được sử dụng hơn.

Sự phân loại này không chỉ giúp hiểu rõ hơn về mức độ phổ biến của các chữ Nôm tự tạo trong các văn bản cổ mà còn thể hiện rõ sự đồng nhất trong việc áp dụng tỷ lệ phân loại giữa các bộ từ thư. Các chữ cao tần thường có tính khái quát rộng rãi, trong khi các chữ thấp tần lại chi tiết và đặc trưng hơn. Việc phân loại tần suất này là một công cụ hữu ích để nghiên cứu về sự phân bố và hiệu suất sử dụng của các chữ Nôm tự tạo trong văn bản cổ.

### 3. Hiệu suất sử dụng của các nhóm chữ và phân tích nguyên nhân

#### 3.1 Hiệu suất sử dụng

Để làm rõ hiệu suất sử dụng của các nhóm chữ trong các từ thư Hán Nôm cổ Việt Nam, mục này sẽ tiến hành phân tích và so sánh thêm về tần suất xuất hiện của chữ Hán và chữ Nôm đã được đề cập trước đó. Hiệu suất sử dụng của chữ Hán trong mục này chỉ ra việc sử dụng chữ Hán trong tiếng Việt, chứ không phải trong tiếng Hán,

và các dữ liệu phân tích đều lấy từ tần suất chữ Hán và chữ Nôm trong năm bộ từ thư cổ Việt Nam đã được nhắc đến trước đó. Các chữ Nôm được đề cập chỉ là các chữ Nôm tự tạo, còn chữ Hán là các chữ thuần Hán như đã được nói đến trước đó.

Trong phần trên, chúng ta đã phân loại tần suất chữ vào 7 nhóm. Mặc dù không phải tất cả các từ thư đều có đầy đủ 7 nhóm này, nhưng để phục vụ cho việc phân tích so sánh, chúng tôi vẫn duy trì sử dụng 7 nhóm này. Vì mục tiêu của phần này là thảo luận về hiệu suất sử dụng của chữ Hán và chữ Nôm trong các từ thư Hán Nôm cổ Việt Nam, cần phải xem các từ thư cổ này như một thể thống nhất. Do đó, cần tính toán tỷ lệ bao phủ trung bình.

Nếu chúng ta xem phạm vi tần số chữ (TSC) là 1 của các từ thư tương ứng là A1, A2, A3, A4, A5, thì tỷ lệ bao phủ trung bình (TLBPTB) của các từ thư cổ đối với phạm vi TSC sẽ là:

$$TLBPTB (\%) = \frac{A1+A2+A3+A4+A5}{5}$$

Thông tin chi tiết xin tham khảo bảng dưới đây.

Bảng 3: Tỷ lệ bao phủ trung bình của chữ Nôm và chữ Hán trong năm bộ từ thư Hán Nôm cổ

TSXH	Chữ Nôm (%)							Chữ Hán (%)						
	1	2-5	6-10	11-20	21-50	51-100	>100	1	2-5	6-10	11-20	21-50	51-100	>100
CN	9,71	19,65	16,56	22,11	27,66	4,31	0,00	13,01	26,68	19,15	10,60	14,61	6,76	9,19
ND	16,01	27,47	18,02	16,50	18,72	3,28	0,00	17,36	29,14	16,51	16,03	16,72	4,24	0,00
TĐ	7,40	15,94	11,70	11,28	20,70	13,34	19,54	8,79	19,09	10,93	11,04	15,97	11,25	22,93
BK	15,43	24,41	14,96	12,81	20,25	12,14	0,00	28,23	37,22	12,34	7,08	15,13	0,00	0,00
ĐN	10,81	22,19	14,98	22,98	29,04	0,00	0,00	19,11	32,77	15,46	16,23	8,29	8,14	0,00
TLBPTB	11,90	21,93	15,24	17,14	23,27	6,61	3,91	17,30	28,98	14,88	12,20	14,14	6,08	6,42

Phân tích dữ liệu từ Bảng 3 cho thấy sự khác biệt rõ rệt về tần suất xuất hiện và mức độ bao phủ trung bình giữa chữ Nôm và chữ Hán trong năm bộ từ thư Hán Nôm cổ. Nhìn chung, chữ Nôm tập trung chủ yếu ở các nhóm tần suất trung bình (11-50 lần), với tỷ lệ trung bình đạt 40,41%, trong khi đó, chữ Hán lại có xu hướng phân bố rộng hơn ở cả các nhóm tần suất cao, đặc biệt nổi bật ở nhóm >100 lần với tỷ lệ 6,42%. Ví dụ, trong *CN*, chữ Nôm đạt tỉ lệ cao nhất ở nhóm 21-50 lần (27,66%), phản ánh chức năng ghi âm phổ biến của chữ Nôm trong phần chú giải âm Hán, như các từ 天 (trời), 襖 (áo) xuất hiện lặp lại trong nhiều mục từ. Tương tự, *ND* cũng cho thấy chữ Nôm xuất hiện nhiều ở nhóm 2-5 và 6-10 lần (tổng cộng 45,49%), phản ánh tính đối thoại phong phú của tác phẩm. Trái lại, trong *TD*, tỷ lệ chữ Hán ở nhóm >100 lần lên đến 22,93%, cao nhất trong năm bộ, cho thấy vai trò trung tâm và độ lặp lại lớn của các từ Hán trong văn bản giáo huấn như 學 (học), 國 (quốc), 民 (dân), thường xuyên được sử dụng theo cấu trúc vắn điệu. Ngoài ra, *BK* có tỷ lệ chữ Hán ở nhóm 2-5 lần rất cao (37,22%), phù hợp với đặc điểm từ điển học định danh sự vật, trong khi chữ Nôm ở đây chủ yếu ghi chú giải âm hoặc đối ứng nghĩa cho các từ loại động - thực vật dân gian như 豕 (heo), 鷄 (gà), 鞞 (cỏ). Những số liệu này cho thấy rõ sự phân công chức năng giữa chữ Hán và chữ Nôm: chữ Hán giữ vai trò trục chính về từ nguyên và ngữ nghĩa, còn chữ Nôm đảm nhiệm chức năng diễn giải, ghi âm và phản ánh lớp từ vựng bản địa, thường xuất hiện với

tần suất vừa phải và có tính bổ trợ trong văn bản.

Ngoài tỷ lệ bao phủ trung bình (TLBPTB), cần phải tính toán tỷ lệ bao phủ tích lũy (TLBPTL). TLBPTL là tỷ lệ phần trăm tổng số tần suất xuất hiện (TSXH) của nhiều chữ đơn so với tổng tần suất của tài liệu, và có thể được sử dụng để phản ánh tình hình sử dụng chữ Hán và chữ Nôm tại Việt Nam. Việc tính toán TLBPTL dựa trên TLBPTB, là một khái niệm bao quát.

Lấy ví dụ với các chữ Nôm tự tạo, TLBPTB đối với các chữ có TSXH lớn hơn 100 là 3,91%, tức là chỉ có 3,91% các chữ tự tạo có TSXH lớn hơn 100. TLBPTB đối với các chữ có tần suất từ 51 đến 100 là 6,61%, nhưng nếu tính tích lũy, số lần xuất hiện của các chữ có tần suất từ 51 đến 100 không chỉ giới hạn trong 6,61% chữ tự tạo này mà còn bao gồm cả 3,91% các chữ tự tạo có tần suất lớn hơn 100 đã đề cập trước đó. Do đó, TLBPTL đối với các chữ có tần suất từ 51 đến 100 là 10,52%.

Tương tự như vậy, TLBPTL đối với các chữ có tần suất từ 1 lần trở lên là 100%, vì tất cả các chữ tự tạo đã được thống kê trong phần này đều xuất hiện trong các từ thư Hán Nôm cổ.

Công thức tính TLBPTL có thể được biểu diễn như sau:

$$TLBPTL(\%) = \sum_{i=1}^n TLBPTB_i$$

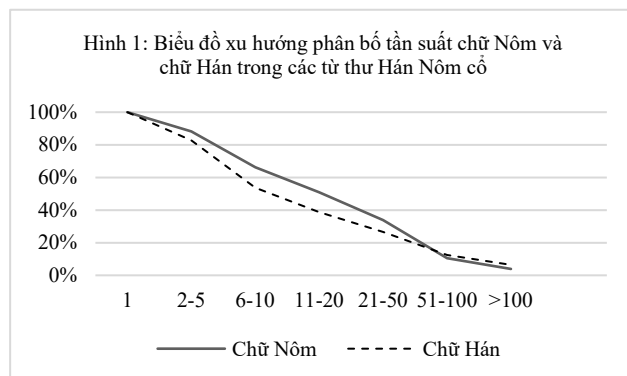
Trong đó  $TLBPTB_i$  là tỷ lệ bao phủ trung bình của nhóm chữ thứ  $i$  và  $n$  là số nhóm chữ được tích lũy.

Dựa trên bảng dữ liệu, có thể tính được TLBPTL của chữ Hán và chữ Nôm trong

các từ thư Hán Nôm cổ, tham khảo chi tiết trong bảng thống kê cụ thể.

Bảng 4: TLBPTL của chữ Nôm và chữ Hán trong các từ thư Hán Nôm cổ

TSC (lần)	Chữ Nôm		Chữ Hán	
	TSTL	TLBPTL	TSTL	TLBPTL
>100	1546	3,91%	900	6,42%
51-100	2890	10,52%	1874	12,50%
21-50	6160	33,79%	3853	26,64%
11-20	8329	50,93%	5764	38,84%
6-10	10335	66,17%	8246	53,72%
2-5	13180	88,10%	12460	82,70%
1	14636	100,00%	15695	100,00%



Để hiển thị trực quan hơn về tình hình sử dụng chữ Hán và chữ Nôm trong năm bộ từ thư Hán Nôm cổ, dữ liệu từ bảng 4 đã được minh họa thành hình 1. Trục hoành đại diện cho tần số chữ (TSC) và trục tung đại diện cho tỷ lệ bao phủ tích lũy (TLBPTL), với phạm vi từ 0% đến 100%, được chia thành 5 mức độ.

Dựa trên bảng 4 và hình 1, có thể thấy rằng xu hướng sử dụng chữ Hán và chữ Nôm tự tạo trong từ thư Hán Nôm cổ nhìn chung đều theo xu hướng giảm dần. Xu hướng sử dụng có thể chia thành ba giai đoạn chính:

- Từ chỉ xuất hiện 1 lần đến 6-10 lần: TLBPTL của chữ Hán và chữ Nôm tự tạo giảm mạnh.
- Từ 6-10 lần đến 21-50 lần: TLBPTL của chữ Hán và chữ Nôm tự tạo giảm dần và chậm lại.
- Từ 51-100 lần và lớn hơn 100 lần: TLBPTL của chữ Hán và chữ Nôm tự tạo giảm chậm, ổn định hơn và gần như đạt mức 0%.

Quan sát kỹ hơn, có thể thấy rằng đường biểu diễn TLBPTL của chữ Hán nằm

phía trên đường của chữ Nôm tự tạo, cho thấy tình hình sử dụng chữ Hán tại Việt Nam cao hơn so với chữ Nôm tự tạo. Phân tích chi tiết, khi phạm vi tần suất là từ 2-5 lần, TLBP của chữ Nôm tự tạo là 88,10%, chỉ giảm 11,9%, trong khi TLBP của chữ Hán là 82,70%, giảm 17,3%. Điều này cho thấy, đối với các chữ có tần suất thấp và thấp thứ hai, chữ Nôm tự tạo ổn định hơn và có xu hướng ghi lại những sự vật đặc thù.

- Khi phạm vi tần suất là từ 6-10 lần, TLBP của chữ Nôm tự tạo là 66,17%, giảm 21,93%, trong khi TLBP của chữ Hán là 53,72%, giảm 28,98%. Điều này cho thấy, đối với các chữ có tần suất trung bình, chữ Nôm tự tạo vẫn ổn định hơn.
- Khi phạm vi tần suất là từ 11-20 lần, TLBP của chữ Nôm tự tạo là 50,93%, giảm 15,24%, trong khi TLBP của chữ Hán là 38,84%, giảm 14,88%. Đối với các chữ có tần suất cao thứ ba, mức giảm của chữ Hán bắt đầu thấp hơn chữ Nôm tự tạo, tức là sự ổn định của chữ Hán dần được thể hiện.
- Khi phạm vi tần suất là từ 21-50 lần, TLBP của chữ Nôm tự tạo là 33,79%, giảm 17,14%, trong khi TLBP của chữ Hán là

26,64%, chỉ giảm 12,20%. Khi phạm vi tần suất trong nhóm chữ này càng lớn, mức giảm của chữ Hán càng thấp hơn chữ Nôm tự tạo, sự ổn định của chữ Hán càng rõ rệt.

- Khi phạm vi tần suất là từ 51-100 lần, TLBP của chữ Nôm tự tạo là 10,52%, giảm 23,27%, trong khi TLBP của chữ Hán là 12,50%, giảm 14,14%. Đối với các chữ có tần suất cao, mức giảm của chữ Hán rõ rệt thấp hơn chữ Nôm tự tạo, cho thấy khi ghi lại các từ ngữ có ý nghĩa rộng, người ta thường ưu tiên sử dụng chữ Hán.

- Khi phạm vi tần suất lớn hơn 100 lần, TLBP của chữ Nôm tự tạo là 3,91%, giảm 6,61%, trong khi TLBP của chữ Hán là 6,42%, giảm 6,08%. Trong phạm vi này, TLBPTL của cả chữ Hán và chữ Nôm tự tạo đều gần bằng 0%, tuy nhiên mức suy giảm ở chữ Hán thấp hơn so với chữ Nôm tự tạo. Điều này phản ánh rằng, trong việc ghi lại các khái niệm có tần suất sử dụng cao trong đời sống, chữ Hán vẫn chiếm ưu thế nhất định. Nguyên nhân có thể bắt nguồn từ việc một bộ phận lớn từ vựng Hán Việt - vốn đã được tích hợp sâu vào tiếng Việt qua thời gian - có sẵn hình thức biểu đạt bằng chữ Hán, với nét nghĩa ổn định và dễ hiểu. Do đó, chữ Hán thường được vay mượn nguyên dạng để sử dụng trong văn bản Nôm, nhất là với những từ đồng âm với tiếng Việt hoặc có thể đọc lệch âm theo thói quen ngôn ngữ bản địa.

### 3.2 Phân tích nguyên nhân

Sau khi chữ Hán được du nhập vào Việt Nam, nó đã trở thành văn tự chính thức. Dưới ảnh hưởng của các yếu tố chính trị, văn hóa và giáo dục, một số từ vựng ghi bằng chữ Hán đã được tiếp nhận và sử

dụng trong tiếng Việt. Tuy vậy, cũng cần nhận thức rằng, chữ Hán vốn được sáng tạo ra để ghi lại tiếng Hán. Sự khác biệt về ngữ âm và cấu trúc giữa tiếng Hán và tiếng Việt dẫn đến việc trong nhiều trường hợp, chữ Hán không đủ khả năng ghi âm chính xác tiếng Việt. Điều này là một trong những nguyên nhân thúc đẩy sự xuất hiện và phát triển của các chữ Nôm tự tạo - như một giải pháp bổ sung nhằm đáp ứng nhu cầu biểu đạt các lớp từ thuần Việt hoặc cấu trúc ngữ pháp mang tính đặc thù bản địa. Khi tiếng Việt phát triển, sự khác biệt giữa âm Hán Việt và âm Nôm ngày càng rõ rệt. Do đó, sau khi chữ Nôm mượn Hán xuất hiện, đã dần dần có sự xuất hiện của các chữ Nôm tự tạo được xây dựng dựa trên nguyên lý Lục thư, sử dụng chữ Hán hoặc bộ phận của chữ Hán. Chữ Hán là văn tự chính thức, trong khi chữ Nôm, đặc biệt là các chữ Nôm tự tạo, lại phù hợp hơn với âm Nôm. Vì vậy, cả hai loại chữ này đã cạnh tranh với nhau trong cuộc sống hàng ngày, dẫn đến việc một số sự vật có thể được ghi lại bằng chữ Hán hoặc chữ Nôm, điều này không phù hợp với nguyên lý tiết kiệm ngôn ngữ (linguistic economy principle).

Để sử dụng chữ viết một cách hiệu quả, người dùng thường ưu tiên những chữ có khả năng thể hiện nhiều khái niệm với số lượng tối thiểu, giúp tiết kiệm trong việc ghi chép ngôn ngữ. Điều này khiến một số chữ Hán và chữ Nôm tự tạo được sử dụng thường xuyên hơn do khả năng thể hiện nhiều khái niệm với tần suất cao. Tuy nhiên, những chữ Hán và chữ Nôm tự tạo biểu thị khái niệm đặc thù thường ít được sử dụng hơn, dẫn đến tần suất sử dụng thấp. Mặc dù

chữ Hán từng là văn tự chính thức và được sử dụng rộng rãi trong các sách cổ và kỳ thi khoa cử, nhưng trong quá trình cạnh tranh với chữ Nôm, vị thế của chữ Hán đã bị ảnh hưởng. Tuy nhiên, chữ Hán vẫn được sử dụng nhiều hơn chữ Nôm do hai nguyên nhân chính:

Thứ nhất, mức độ chuẩn hóa của chữ Hán cao hơn chữ Nôm. Chữ Nôm được sáng tạo bởi người Việt Nam, ban đầu chưa được công nhận chính thức, dẫn đến sự đa dạng trong hình thức và mức độ chuẩn hóa thấp. Điều này khiến một chữ Nôm có thể có nhiều dạng khác nhau, dẫn đến tàn suất bị phân tán. Ngược lại, chữ Hán có mức độ chuẩn hóa cao hơn, giúp giảm thiểu số lượng chữ dị thể và tăng tàn suất sử dụng.

Thứ hai, phạm vi sử dụng của chữ Hán và chữ Nôm khác nhau. Chữ Hán chủ yếu dùng để ghi tiếng Hán và các từ vựng chuyên dụng, trong khi chữ Nôm dùng để ghi tiếng Việt và từ vựng hàng ngày. Do phạm vi sử dụng của chữ Nôm nhỏ hơn, hiệu dụng của nó không thể bằng chữ Hán. Tổng tàn suất của chữ Hán là 15695 với 5300 chữ, trong khi chữ Nôm tự tạo là 14636 với 3004 chữ, đều thấp hơn chữ Hán.

#### 4. Khảo sát chữ Nôm thường dùng và thông dụng

Giống như chữ Hán, chữ Nôm cũng xuất hiện nhiều dị thể, cổ tự, và chữ khó nhận biết do nhiều nguyên nhân. Vì vậy, cần thiết phải tổng kết và phân loại chữ Nôm thành thường dùng, ít dùng, và hiếm dùng dựa trên tàn suất sử dụng. Trên thực tế, việc sắp xếp chữ Nôm thường dùng đã được thực hiện từ lâu, với các văn bản học chữ cơ bản như *Thiên tự văn giải âm* 千字

文解音, *Tam thiên tự giải âm* 三千字解音, và *Ngũ thiên tự dịch quốc ngữ* 五千字譯國語. Những văn bản này lần lượt thu thập khoảng 1000, 3000 và 5000 chữ Hán, mỗi chữ đều có cách đọc (giải thích bằng tiếng Việt, một phần sử dụng chữ Nôm tự tạo để ghi lại). Mặc dù mục đích chính của những văn bản này là để học chữ Hán, chữ Nôm không phải là đối tượng học tập chính, nhưng những chữ Nôm được thu thập trong đó cơ bản phản ánh nhu cầu cơ bản của xã hội đối với chữ Nôm vào thời điểm đó.

Trong những thập kỷ gần đây, việc thiết lập và nghiên cứu chữ Nôm thường dùng đã được triển khai và phát triển sâu rộng, trở thành một phần quan trọng trong nghiên cứu chữ Hán, chữ Nôm của Việt Nam. Năm 1991, Hội Ngôn ngữ học Thành phố Hồ Chí Minh đã xuất bản *Sách tra chữ Nôm thường dùng* (STCNTD) do Lạc Thiện biên soạn, đây là tác phẩm mở đầu cho việc thiết lập bảng chữ Nôm thường dùng, thu thập khoảng 5000 chữ Nôm thông dụng, chủ yếu lấy từ các tác phẩm văn học chữ Nôm như *Nhị Độ Mai diễn ca* 二度梅演歌, *Cung oán ngâm khúc* 宮怨吟曲, *Hồng Đức Quốc âm thi tập* 洪德國音詩集, và *Đại Nam quốc sử diễn ca* 大南國史演歌. Năm 2022, Ủy ban Phục sinh Hán Nôm Việt Nam đã biên soạn *Bảng chữ Hán Nôm chuẩn thường dùng* (BCHNCTD), đây là kết quả mới nhất về bảng chữ thường dùng. Cuốn sách này phân chia hệ thống chữ Nôm thành bảng chữ cơ bản thường dùng và bảng chữ cao cấp thường dùng, cũng lấy từ các tác phẩm văn học chữ Nôm như *Truyện Kiều* 翹傳, *Chinh phụ ngâm khúc* 征婦吟曲 và *Xuân Hương thi tập* 春香詩集. Đáng chú ý là BCHNCTD còn tham

khảo nhiều từ điển quan trọng như *Khang Hy tự điển* 康熙字典, *Tự Đức thánh chế tự học giải nghĩa ca* 嗣德聖製字學解義歌 và *Đại Nam Quốc âm tự vị* 大南國音字彙. Mặc dù BCHNCTD đã có những tiến bộ đáng kể, nhưng cần tham khảo thêm nhiều từ thư để biên soạn một bảng tra chữ Nôm toàn diện hơn.

#### 4.1 Phân loại chữ Nôm tự tạo theo tần suất và mức độ phổ biến

Dựa trên dữ liệu về tần suất sử dụng của chữ Nôm tự tạo trong năm bộ từ thư Hán Nôm cổ của Việt Nam, phần này phân chia chữ Nôm tự tạo thành hai bảng: Bảng chữ Nôm tự tạo thường dùng (bảng thường dùng) và bảng chữ Nôm tự tạo thông dụng (bảng thông dụng).

Bảng thường dùng được phân thành ba cấp độ dựa trên tần suất sử dụng:

- Loại A: Bao gồm các chữ Nôm tự tạo có tần suất sử dụng cao, hiệu dụng cao, và xuất hiện trong từ ba đến năm bộ từ thư Hán Nôm cổ. Hầu hết các chữ thuộc nhóm này nằm trong các dải cao tần, trung cao tần hoặc trung tần, phản ánh mức độ lặp lại và phổ biến cao trong kho từ vựng văn bản. Tuy nhiên, không phải tất cả các chữ có tần suất cao đều được xếp vào loại A, vì tiêu chí phân loại còn xét đến mức độ phân bố: những chữ chỉ xuất hiện trong một hoặc hai bộ từ thư sẽ không được đưa vào nhóm này, dù tần suất nội tại cao. Chẳng hạn, chữ 𠂔 (trời) xuất hiện trong cả năm bộ từ thư khảo sát và nằm trong nhóm chữ cao tần của ba bộ: *CN*, *TĐ* và *BK*, đáp ứng đầy đủ cả hai tiêu chí về tần suất và độ phổ cập, nên được xếp vào nhóm A. Tương tự, chữ 𠂔 (trăng) cũng có mặt trong cả năm bộ từ

thư và thuộc nhóm cao tần ở bốn bộ: *CN*, *TĐ*, *BK* và *ĐN*, do đó cũng được phân loại vào nhóm thường dùng loại A. Tổng số lượng chữ trong nhóm này là 45 chữ.

- Loại B: bao gồm các chữ Nôm tự tạo có tần suất sử dụng và mức độ hiệu dụng trung bình. Nhóm này chủ yếu bao gồm các chữ thuộc dải cao tần hoặc trung tần nhưng không đáp ứng đầy đủ các tiêu chí để được xếp vào loại A, thường do giới hạn về phạm vi phân bố trong các bộ từ thư. Ngoài ra, một số chữ có tần suất trung thấp hoặc thấp tần, nhưng lại xuất hiện đồng thời trong cả năm bộ từ thư, cũng được xếp vào loại này vì thể hiện mức độ phổ biến về mặt không gian văn bản, dù tần suất lặp lại không cao. Tổng số lượng chữ trong nhóm B là 295 chữ. Ví dụ, chữ 𠂔 (đèn) xuất hiện trong cả năm bộ từ thư khảo sát, thể hiện tính phân bố rộng. Tuy nhiên, về tần suất, chữ này chỉ nằm trong nhóm cao tần ở hai bộ là *BK* và *ĐN*, còn lại trong ba bộ kia, chữ này chỉ thuộc nhóm thấp tần. Do đó, mặc dù có độ phủ cao về mặt số lượng văn bản, nhưng mức độ sử dụng không đủ nổi bật để xếp vào nhóm A, và được phân loại hợp lý vào nhóm B.

- Loại C: bao gồm các chữ Nôm tự tạo có tần suất sử dụng và hiệu dụng ở mức thấp nhất trong toàn bộ kho từ liệu khảo sát. Nhóm này chủ yếu bao gồm các chữ thuộc dải trung thấp tần và thấp tần, đồng thời không đáp ứng các tiêu chí để được xếp vào nhóm B, như không xuất hiện đủ rộng trong nhiều bộ từ thư hoặc không có tần suất đủ cao trong bất kỳ văn bản nào. Số lượng chữ thuộc nhóm C là 1693 chữ, chiếm tỉ lệ lớn nhất trong toàn bộ hệ thống phân loại. Một ví dụ điển hình là chữ 𠂔

(ngựa), chỉ xuất hiện trong ba bộ từ thư gồm *ND*, *TD* và *DN*. Trong cả ba trường hợp, chữ này đều thuộc nhóm thấp tần, cho thấy mức độ xuất hiện hạn chế và phạm vi sử dụng không rộng. Do đó, chữ 𨾏 (ngựa) được xếp hợp lý vào nhóm C - nhóm các đơn vị có tính phổ dụng thấp và phạm vi lưu hành hẹp.

Bảng thông dụng được thiết lập dựa trên mức độ phổ biến của chữ Nôm tự tạo trong năm bộ từ thư Hán Nôm cổ, không xét đến tần suất sử dụng. Tiêu chí phân loại tập trung vào phạm vi phân bố của mỗi chữ trong hệ thống văn bản khảo sát, từ đó phản ánh tính kế thừa và mức độ ổn định trong quá trình sử dụng. Hệ thống chia thành năm cấp độ, cụ thể như sau:

- Cấp 1: Bao gồm các chữ xuất hiện trong cả năm bộ từ thư, thể hiện mức độ phổ biến và tính kế thừa cao nhất. Nhóm này có 31 chữ, tiêu biểu như chữ 澗 (sông) và 𨾏 (say), đều có mặt trong toàn bộ văn bản khảo sát, phản ánh rõ rệt vai trò cốt lõi trong từ vựng Hán Nôm.

- Cấp 2: Gồm các chữ xuất hiện trong bốn bộ từ thư, phản ánh mức độ kế thừa tương đối. Tổng số chữ trong nhóm này là 93 chữ. Ví dụ, chữ 𨾏 (lợn) và 𨾏 (nữ) đều vắng mặt trong *CN*, nhưng có mặt trong bốn văn bản còn lại.

- Cấp 3: Gồm các chữ xuất hiện trong ba bộ từ thư, với tính kế thừa yếu hơn. Nhóm này có 132 chữ. Chẳng hạn, chữ 𨾏 (rêu) chỉ được ghi nhận trong ba bộ, không xuất hiện trong *CN* và *ND*.

- Cấp 4: Bao gồm các chữ xuất hiện trong hai bộ từ thư, mang tính cá nhân hóa cao hơn và mức độ phổ biến thấp. Có 254 chữ thuộc nhóm này. Ví dụ, chữ 𨾏 (cày) chỉ được tìm thấy trong *CN* và *ND*.

- Cấp 5: Là nhóm chữ chỉ xuất hiện trong một bộ từ thư duy nhất, phản ánh mức độ lưu truyền hạn chế và tính địa phương rõ nét. Đây là nhóm chiếm tỷ lệ lớn nhất với 1517 chữ. Tiêu biểu là chữ 𨾏 (mắt), chỉ xuất hiện trong *CN*, hoàn toàn vắng mặt trong các bộ còn lại.

#### 4.2 Bảng thường dùng và bảng thông dụng

Theo các tiêu chuẩn phân loại chữ thường dùng kết hợp với cấp độ thông dụng như trên, ta có thể xác định các nhóm chữ sau:

Nhóm A1: Gồm chữ vừa thuộc loại A (thường dùng) vừa thuộc cấp 1 (thông dụng cao). Số lượng: 14 chữ; Nhóm A2: Chữ loại A kết hợp cấp 2. Số lượng: 21 chữ. Nhóm A3: Chữ loại A kết hợp cấp 3. Số lượng: 9 chữ. Nhóm B1: Chữ loại B kết hợp cấp 1. Số lượng: 17 chữ. Nhóm B2: Chữ loại B kết hợp cấp 2. Số lượng: 36 chữ. Nhóm B3: Chữ loại B kết hợp cấp 3. Số lượng: 48 chữ. Nhóm B4: Chữ loại B kết hợp cấp 4. Số lượng: 52 chữ. Nhóm C2: Chữ loại C kết hợp cấp 2. Số lượng: 36 chữ. Nhóm C3: Chữ loại C kết hợp cấp 3. Số lượng: 75 chữ. Nhóm C4: Chữ loại C kết hợp cấp 4. Số lượng: 202 chữ.

Bảng 5: Bảng Chữ Nôm tự tạo thường dùng và thông dụng (cấp 1 đến cấp 4)<sup>4</sup>

Chữ	Âm đọc	CN	ND	TĐ	BK	ĐN	Cấp thường dùng	Cấp thông dụng	Chữ	Âm đọc	CN	ND	TĐ	NP	ĐN	Cấp thường dùng	Cấp thông dụng
奎	trời	○	8	○	○	8	A	1	淹	sông	○	8	○	8	○	A	1
肢	trăng	○	8	○	○	○	A	1	馮	mỡ	○	8	○	8	○	A	1
猓	lợn	○	○	Í	○	○	A	2	炤	lửa	○	○	Í	○	○	A	2
枷	tre	○	8	Í	○	○	A	2	蒼	thom	○	8	○	Í	○	A	2
媧	vợ	○	○	Í	Í	○	A	3	茈	thước	Í	○	○	Í	○	A	3
蛟	sùng	○	○	Í	Í	○	A	3	翺	cánh	Í	○	Í	○	○	A	3
畑	đèn	8	8	8	○	○	B	1	醜	say	8	8	○	8	8	B	1
炆	bếp	8	8	8	8	○	B	1	晦	mối	8	8	8	8	8	B	1
捕	bò	○	8	Í	○	8	B	2	狝	chuột	○	8	Í	8	○	B	2
蠅	rồng	○	Í	8	8	8	B	2	梔	dầu	Í	○	○	8	8	B	2
紉	bó	Í	8	Í	8	○	B	3	螻	lươn	8	8	Í	Í	○	B	3
蓀	rêu	Í	Í	8	8	○	B	3	噉	chén	Í	8	Í	8	○	B	3
棋	cây	○	8	Í	Í	Í	B	4	逄	đưa	○	Í	○	Í	Í	B	4
扇	rày	○	Í	Í	Í	8	B	4	蝮	rắn	○	Í	○	Í	Í	B	4
苒	nờ	Í	8	8	8	8	C	2	鯤	vây	8	8	8	8	Í	C	2
蛛	traí	8	8	Í	8	8	C	2	螭	chuồn	8	Í	8	8	8	C	2
罍	ngừa	Í	8	8	Í	8	C	3	躄	đứng	Í	8	Í	8	8	C	3
啞	nức	Í	8	Í	8	8	C	3	黉	sau	Í	8	Í	8	8	C	3
抹	chùi	8	Í	8	Í	Í	C	4	禡	mớ	8	Í	Í	Í	8	C	4
焮	chóp	8	Í	8	Í	Í	C	4	猗	gà	8	Í	Í	Í	8	C	4

Ngoài ra, còn có một số lượng chữ Nôm không thông dụng (cấp 5), bao gồm 139 chữ loại B không thông dụng và 1378

chữ loại C không thông dụng. Những chữ không thông dụng này chỉ xuất hiện trong một bộ từ thư duy nhất.

Bảng 6: Bảng chữ Nôm không thông dụng (cấp 5)

B5	𪛗(trông), 糲(vai), 𪛘(mắt), 𪛙(mây), 𪛚(ngươi), 𪛛(răng), 𪛜(lời), 𪛝(cùi), 𪛞(mạ), 𪛟(chất), 𪛠(gạo), 𪛡(đỏ), 𪛢(sợi), 𪛣(chuối), 𪛤(gà), 𪛥(dép), 𪛦(sáng), 𪛧(mũi), 𪛨(sét), 𪛩(ruộng), 𪛪(bánh), 𪛫(chèo), 𪛬(mọt), 𪛭(chim), 𪛮(cháo), 𪛯(cơm), 𪛰(tớ), 𪛱(rượu), 𪛲(mâm), 𪛳(tôm), 𪛴(lụa), 𪛵(tói), 𪛶(vòng), 𪛷(ngòi), 𪛸(buổi), 𪛹(trau), 𪛺(vườn), 𪛻(thẻ), 𪛼(bè), 𪛽(mọi)...
C5	𪛾(vây), 𪛿(rồng), 𪜀(đèn), 𪜁(gió), 𪜂(bếp), 𪜃(sấm), 𪜄(đêm), 𪜅(thằng), 𪜆(dệt), 𪜇(trập), 𪜈(váy), 𪜉(nách), 𪜊(măng), 𪜋(sứa), 𪜌(sốt), 𪜍(oản), 𪜎(khoảng), 𪜏(lay), 𪜐(rồng), 𪜑(đỡ), 𪜒(chín), 𪜓(rói), 𪜔(quạt), 𪜕(khô), 𪜖(dạ), 𪜗(sùng), 𪜘(nằm), 𪜙(cong), 𪜚(dằm), 𪜛(bò), 𪜜(ngỗng), 𪜝(may), 𪜞(uống), 𪜟(rận), 𪜠(nhạn)...

**4.3 So sánh với Sách tra chữ Nôm thường dùng (STCNTD) và Bảng chữ Hán Nôm chuẩn thường dùng (BCHNCTD)**

Để thuận tiện cho việc trình bày, chúng tôi sẽ kết hợp bảng thông dụng và bảng

thường dùng thành bảng thông dụng và thường dùng (sau đây gọi tắt là bảng chữ). Sau đó, chúng tôi sẽ so sánh bảng này với STCNTD và BCHNCTD để chỉ ra những điểm giống và khác nhau giữa chúng.

#### 4.3.1 Điểm giống nhau

Về phương diện ứng dụng, cả hai bảng phân loại đều xuất phát từ thực tế rằng chữ Nôm hiện nay không còn là hệ chữ viết phổ biến trong đời sống đương đại, điều này gây không ít trở ngại cho quá trình tiếp cận và nghiên cứu của các học giả trong và ngoài Việt Nam. Tuy nhiên, do một số lượng lớn văn bản cổ Việt Nam vẫn sử dụng chữ Nôm để ghi chép các nội dung liên quan đến lịch sử, văn hóa, tín ngưỡng và tri thức dân gian, nên việc xây dựng hệ thống phân loại và nhận diện chữ Nôm có ý nghĩa thiết thực trong công tác giải mã, bảo tồn và khai thác nguồn tư liệu văn bản cổ. Chữ Nôm chưa được công nhận chính thức và thường được sử dụng trong các trường hợp không chính thức, dẫn đến sự tồn tại của nhiều dị thể, làm tăng độ khó khi hiểu và sử dụng. Vì vậy, việc xuất hiện các bảng chữ Nôm giúp các học giả nhận biết, nghiên cứu chữ Nôm và so với từ điển, bảng chữ có kích thước nhỏ hơn, dễ tra cứu hơn. Bảng chữ mà chúng tôi biên soạn và STCNTD, BCHNCTD đều có mục đích chung: cung cấp toàn diện hình dạng chữ Nôm tự tạo, phân cấp chúng để giúp người dùng hiểu và giải đọc văn bản chữ Nôm, từ đó mở rộng nghiên cứu liên quan.

Phương pháp biên soạn giống nhau: Bảng chữ của chúng tôi và STCNTD, BCHNCTD đều được biên soạn bằng cách kết hợp kinh nghiệm và phương pháp khoa học. Cụ thể, chúng tôi đều chọn các văn bản cổ làm tài liệu, tiến hành phân tích thống kê, và cuối cùng sắp xếp thành bảng theo mục đích biên soạn.

#### 4.3.2 Điểm khác nhau

Sự khác biệt về nguồn tài liệu: Bảng chữ mà chúng tôi biên soạn và STCNTD, BCHNCTD có sự khác biệt về nguồn tài liệu sử dụng. Về loại tài liệu, STCNTD hoàn toàn dựa vào các tác phẩm văn học chữ Nôm truyền thống của Việt Nam. BCHNCTD chủ yếu dựa vào các tác phẩm văn học chữ Nôm truyền thống, đồng thời tham khảo các từ điển như *TD* và *ĐN*. Trong khi đó, Bảng chữ của chúng tôi hoàn toàn dựa vào năm bộ từ thư cổ tiêu biểu của Việt Nam làm nguồn thống kê. Các tác phẩm văn học chữ Nôm chứa nhiều chữ Nôm quan trọng, nhưng chúng thiếu tính chuẩn hóa và phổ biến, tồn tại nhiều dị thể và hình dạng hiếm gặp, không phù hợp để phân loại là thường dùng. Bảng chữ của chúng tôi sử dụng các từ thư cổ của Việt Nam làm đối tượng thống kê, mang tính quy phạm, khoa học, và hệ thống hơn. Từ góc độ thời gian của tài liệu, STCNTD có thời gian trải rộng hơn, từ cuối thế kỷ XV đến cuối thế kỷ XIX, kéo dài khoảng 4 thế kỷ. BCHNCTD có thời gian trải rộng ít hơn, từ giữa thế kỷ XIX đến cuối thế kỷ XIX, chỉ khoảng 1,5 thế kỷ. Bảng chữ mà chúng tôi biên soạn có thời gian trải rộng vừa phải, từ thế kỷ XVII đến cuối thế kỷ XIX, trải qua 2 thế kỷ. Khoảng thời gian 4 thế kỷ là một khoảng thời gian dài, và việc gộp các chữ Nôm trong cùng một bảng có thể không phù hợp. Trong khi đó, khoảng thời gian từ 1,5 đến 2 thế kỷ có thể được xem là một giai đoạn phát triển hợp lý cho các chữ Nôm.

Sự khác biệt trong quy trình biên soạn: STCNTD và BCHNCTD tập trung vào việc thu thập và sắp xếp các chữ thường

dùng, nhưng cả hai đều không đề cập đến cách định nghĩa thường dùng. Trên thực tế, bất kỳ hình dạng chữ nào xuất hiện trong văn bản đều được thu thập mà không xem xét tình trạng sử dụng hoặc tần suất của chúng. Quy trình biên soạn của hai bảng này bao gồm: (1) Xác định nguồn tài liệu, với STCNTD sử dụng hoàn toàn các tác phẩm văn học, và BCHNCTD sử dụng các tác phẩm văn học cùng một số từ thư. (2) Thu thập và sắp xếp: STCNTD thu thập 5226 đơn vị, sắp xếp theo số nét; BCHNCTD thu thập 5524 đơn vị, phân chia thành chữ cơ bản và nâng cao, sắp xếp theo thứ tự chữ cái Latinh. Bảng chữ mà chúng tôi biên soạn đã kế thừa thành quả nghiên cứu trước đây, tận dụng công nghệ máy tính để tính toán tỷ lệ bao phủ và tần suất sử dụng của mỗi chữ thường dùng và chữ thông dụng. Quy trình biên soạn như sau: (1) Thiết lập cơ sở dữ liệu, nhập tất cả chữ Nôm từ năm bộ từ thư cổ của Việt Nam vào máy tính, tạo một tập dữ liệu toàn diện về chữ Nôm. (2) Thống kê và kiểm định, thực hiện thống kê 1988 chữ Nôm, xử lý định lượng từng chữ hoặc nhóm chữ, và kiểm định dựa trên dữ liệu về tỷ lệ bao phủ và số lần xuất hiện. (3) Phân loại và lập bảng, tổng hợp tất cả dữ liệu, phân loại chữ thường dùng dựa trên tỷ lệ bao phủ, chia thành năm loại: chữ tần suất cao, trung bình cao, trung bình, trung bình thấp, và thấp. Sau đó, tiếp tục phân chia thành ba

loại: chữ loại A, B, và C. Đối với chữ thông dụng, phân loại dựa trên số lần xuất hiện, từ cấp cao nhất đến cấp thấp nhất. Như vậy, ngoài việc thống kê và phân loại, bảng chữ của chúng tôi còn thêm bước kiểm định dữ liệu, giúp quy trình biên soạn trở nên nghiêm ngặt hơn, có thể chịu được kiểm nghiệm, và mang tính khoa học cao hơn.

Sự khác biệt trong nguyên tắc chọn chữ: STCNTD và BCHNCTD đều thu thập hơn 5000 đơn vị, trong khi bảng chữ của chúng tôi chỉ thu thập 1988 chữ. Dường như số lượng này quá thấp và không đầy đủ, nhưng thực tế không phải vậy. STCNTD và BCHNCTD không chỉ bao gồm chữ Nôm tự tạo mà còn bao gồm cả chữ Nôm mượn từ chữ Hán và các chữ Hán thuần túy. Nếu loại bỏ những chữ này, số lượng còn lại sẽ không chênh lệch nhiều so với bảng chữ của chúng tôi.

Như vậy, STCNTD và BCHNCTD, và Bảng chữ của chúng tôi có sự khác biệt lớn về nguyên tắc chọn chữ. STCNTD và BCHNCTD không phân biệt rõ ràng giữa chữ Nôm và chữ Hán, mà coi tất cả các chữ xuất hiện trong văn bản chữ Nôm và từ điển giải thích là chữ Nôm. Điều này là không chính xác và không phù hợp với tên gọi chữ Nôm.

Để minh họa, chúng tôi chọn một số chữ để giải thích tiêu chuẩn thu thập của ba bảng. Chi tiết được trình bày trong Bảng 7.

Bảng 7: So sánh tiêu chuẩn thu thập chữ của ba bảng

Chữ	STCNTD	BCHNCTD	Bảng chữ
离	lìa	âm li: li hôn, li khai, phân li, cự li âm lia: xa lia	không thu thập
打	đánh: đánh ghen	âm đả: ấu đả, đả kích	không thu thập

		âm đánh: đánh ghen, đánh chén, lính thủy đánh bộ	
燕	én/ yén	âm yén: chim yén, hải yén, yén sào âm én: chim én	không thu thập

Trong bảng này, các chữ 离 (li/lia), 打 (đả/đánh), 燕 (yén/én) đều là chữ Hán thuần túy. Hầu hết các chữ trong BCHNCTD đều là từ Hán. Theo định nghĩa trước đó, những chữ này không thể được thu thập vào Bảng chữ Nôm tự tạo thông dụng và thường dùng của chúng tôi.

### 5. Kết luận

Nghiên cứu đã góp phần phác họa một cách hệ thống bức tranh tổng thể về tần suất xuất hiện và mức độ hiệu dụng của chữ Nôm tự tạo trong bối cảnh văn hóa và ngôn ngữ Việt Nam giai đoạn sử dụng hệ thống chữ Hán Nôm. Thông qua phân tích định lượng, kết quả cho thấy chữ Hán giữ vai trò cốt lõi trong việc đảm bảo tính ổn định hình thức và hiệu quả truyền đạt, đặc biệt ở những lớp từ vựng có nguồn gốc Hán Việt đã được bản địa hóa lâu dài. Bên cạnh đó, chữ Nôm tự tạo được xác lập là công cụ hỗ trợ thiết yếu, đảm nhiệm vai trò ghi âm, biểu đạt các yếu tố ngữ nghĩa đặc thù của tiếng Việt mà chữ Hán không thể phản ánh đầy đủ, nhất là trong việc ghi chép các khái niệm, hiện tượng mang tính bản địa trong đời sống văn hóa - xã hội truyền thống của người Việt.

Đề xuất về bảng phân loại chữ Nôm tự tạo thường dùng và thông dụng là một đóng góp thiết thực, cung cấp một công cụ hữu ích cho việc nghiên cứu và giảng dạy chữ Nôm. Tuy nhiên, nghiên cứu cũng chỉ ra rằng việc chuẩn hóa chữ Nôm là một

thách thức lớn, do sự đa dạng trong hình thức và cách sử dụng.

Trong thời gian tới, việc tiếp tục nghiên cứu sâu hơn về các yếu tố văn hóa, xã hội và ngôn ngữ chi phối quá trình hình thành và sử dụng chữ Nôm tự tạo là hết sức cần thiết nhằm làm sáng tỏ hơn bối cảnh vận hành và chức năng của hệ thống văn tự này. Bên cạnh đó, các nghiên cứu định lượng bổ sung dựa trên những nguồn tư liệu đa dạng hơn sẽ góp phần kiểm chứng, hiệu chỉnh và mở rộng những kết luận đã đạt được trong nghiên cứu hiện tại. Đặc biệt, việc thiết lập một cơ sở dữ liệu chữ Nôm toàn diện, chuẩn hóa và thân thiện với người dùng không chỉ là công cụ hỗ trợ nghiên cứu hữu hiệu mà còn là bước tiến quan trọng trong chiến lược bảo tồn, số hóa và phát huy giá trị của di sản Hán Nôm trong hệ sinh thái văn hóa Việt Nam./.\*

H.Đ.D - T.T.K

### Chú thích và tài liệu tham khảo

\* Nghiên cứu được tài trợ bởi Trường Đại học Nguyễn Tất Thành, Thành phố Hồ Chí Minh trong khuôn khổ đề tài mã số 2025.01.11/HĐ-KHCN.

1. Lê Quý Đôn soạn, Trần Văn Giáp dịch và khảo thích (2006), *Vân đài loại ngữ*, NXB. Văn hóa-thông tin, Hà Nội, tr.253.

2. Sha Zongyuan 沙宗元 (2008), *Nghiên cứu về chuẩn hóa thuật ngữ văn tự học* 文字學術語規範研究, Hợp Phi (合肥): NXB. Đại học An Huy (安徽大學出版社), tr.402.

3. Wang Fengyang 王鳳陽 (1989), *Hán tự học* 漢字學, Trường Xuân (長春): NXB. Văn học- Lịch sử Cát Lâm (吉林文史出版社), tr.603.
4. Ký hiệu (O) biểu thị chữ xuất hiện trong tài liệu được chỉ định và thuộc nhóm tần suất cao đến trung bình cao. Ký hiệu 8 biểu thị chữ xuất hiện trong tài liệu được chỉ định và thuộc nhóm tần suất thấp. Ký hiệu (Í) biểu thị chữ không xuất hiện trong tài liệu được chỉ định.
5. Nguyễn Đình Hiền (2012), *Nghiên cứu về hệ thống ngữ âm Hán Việt và chữ Nôm* 漢越語音系與喃字研究, Luận án Tiến sĩ, Đại học Phúc Đán, Trung Quốc.
6. Nguyễn Đình Hiền (2023), *Nghiên cứu định lượng chữ Nôm: Trường hợp Quốc âm thi tập và Đoạn trường tân thanh* 喃字定量研究——以〈國音詩集〉和〈斷腸新聲〉為例, *Tạp chí Đại học Sư phạm Dân tộc Quảng Tây* (廣西民族師範學院學報), số 6.
7. Shi Weikun 史維坤 (2017), *Thu thập và phân tích Tự học toàn yếu và Chỉ Nam giải âm* 〈字學纂要〉與〈指南解音〉整理分析, Luận án Thạc sĩ, Đại học Quảng Tây, Trung Quốc.
8. Huỳnh Đức Dư (2019), *Nghiên cứu chữ Nôm trong Tự Đức thánh chế tự học giải nghĩa ca* 〈嗣德聖制字學解義歌〉喃字研究, Luận án Thạc sĩ, Đại học Dân tộc Trung ương, Trung Quốc.
9. Huỳnh Đức Dư (2023), *Nghiên cứu về Từ điển cổ Hán Nôm Việt Nam và Các chữ, từ ngữ được thu thập* 越南漢喃古辭書及所收字詞研究, Luận án Tiến sĩ, Đại học Trịnh Châu, Trung Quốc.
10. Ứng Thùy Linh (2020), *Nghiên cứu từ điển song ngữ Hán Nôm Nam phương danh vật bị khảo* 漢喃雙語詞典〈南方名物備考〉研究, Luận án Tiến sĩ, Đại học Sư phạm Quý Châu, Trung Quốc.
11. He Meiyu (2021), *Nghiên cứu về cách dụng chữ trong tác phẩm văn học chữ Nôm Lục Vân Tiên diễn ca của Việt Nam* 越南喃字文學作品〈蓼雲仙歌演〉用字研究, Luận án Thạc sĩ, Đại học Dân tộc Quảng Tây, Trung Quốc.
12. Phạm Văn Khoái, Hà Đăng Việt (2013), *Tự Đức thánh chế tự học giải nghĩa ca: Từ điển song ngữ Hán Nôm thánh chế theo môn loại có tính bách khoa thư*, *Tạp chí Hán Nôm*, số 1/2013, tr.17-28.
13. Hà Đăng Việt (2014), *Nghiên cứu các tự điển, từ điển Hán Nôm thế kỉ XIX*, Luận án Tiến sĩ, Trường Đại học Sư phạm Hà Nội.
14. Lạc Thiện (1991), *Sách tra chữ Nôm thường dùng*, Hội ngôn ngữ học, Thành phố Hồ Chí Minh.
15. Ủy ban phục sinh Hán Nôm (2022), *Bảng chữ Hán Nôm chuẩn thường dùng*, <https://www.hannomrcv.org/bchnctd/postscript.html>.