

KHẢ NĂNG VÀ MỨC ĐỘ THAM GIA CẤU TẠO TỪ CỦA YẾU TỐ HÁN VIỆT

ĐẶNG THI THU HIỀN¹ - LÊ THI MAI ANH² - NGUYỄN THI PHƯƠNG ANH³
NGUYỄN THI HỒNG NGỌC⁴ - VI THI NHƯ QUỲNH⁵

Abstract: This study examines Sino Vietnamese elements as fundamental units in Vietnamese word formation, highlighting their productivity and the broader influence of Chinese language and culture. By collecting, classifying, and systematizing these elements, the research establishes a reliable corpus for linguistic analysis. Statistical data and surveys are employed to assess the extent and capacity of Sino-Vietnamese participation in word formation, providing a foundation for both theoretical insights and practical applications. The findings are expected to contribute to research in lexicology and applied linguistics, while also informing efforts toward language standardization and development.

Keywords: *Sino Vietnamese elements, productivity, word formation, lexicology, culture and language*

1. Mở đầu

Yếu tố Hán Việt và từ Hán Việt là một bộ phận quan trọng trong từ vựng tiếng Việt, điều này đã được khẳng định và xác minh từ lâu trong nghiên cứu Việt ngữ học. Các nghiên cứu này tập trung vào nhiều khía cạnh của từ Hán - Việt, như: nguồn gốc lịch sử, ngữ nghĩa, phong cách chức năng, văn hoá và ứng dụng thực tiễn,... tạo nên một cơ sở lý luận và thực tiễn phong phú cho nghiên cứu về đối tượng này.

Phương diện đầu tiên mà các nhà Việt ngữ học quan tâm là khía cạnh âm đọc của yếu tố Hán - Việt. Nguyễn Tài Cẩn [3] đã phân tích chi tiết quá trình du nhập từ Hán vào tiếng Việt, làm rõ sự biến đổi âm đọc từ tiếng Hán cổ sang cách đọc Hán Việt qua các giai đoạn lịch sử, như sự chuyển đổi từ âm [k-] thành [χ -].

Phan Ngọc [10], với góc nhìn văn hoá và ngôn ngữ học xã hội, lại đề cập đến yếu tố Hán Việt trong việc phản ánh ảnh hưởng của văn hoá Trung Hoa lên văn học và tư tưởng người Việt, đặc biệt trong thơ Nôm và văn bia thời Lý - Trần.

Nguyễn Thị Thanh Chung trong *Khảo luận tầm nguyên từ ghép Hán Việt có cơ chế ngữ âm tương đồng với từ láy* [6] đã tiến hành khảo sát các từ ghép Hán Việt như *phồn vinh*, *hiển hách*, phân tích sự tương đồng về ngữ âm với từ láy tiếng Việt như *lấp lánh*. Nghiên cứu này cho thấy khả năng kết hợp của yếu tố Hán Việt với đặc trưng tiếng Việt, nhưng chưa mở rộng sang tính năng sản tổng quát. Hay trong bài viết *Định hướng khảo luận tầm nguyên hệ thống yếu tố Hán Việt bằng phương pháp nghiên cứu lịch sử - so sánh* [4], Nguyễn Thị Thanh Chung đã đề xuất cách tiếp cận lịch sử để tìm hiểu nguồn gốc của từ Hán Việt như *học*, *văn*, qua sự so sánh với tiếng Hán cổ. Ngoài ra, trong bài nghiên cứu *A discussion on the technique of etymological analysis and its applicability in tracing the origin of Sino-Vietnamese elements* [5], tác giả cũng đã cung cấp góc nhìn quốc tế về phân tích tầm nguyên yếu tố Hán Việt. Các công trình của Nguyễn Thị Thanh Chung mang tính lý thuyết cao, làm sáng tỏ nguồn gốc và ngữ nghĩa, nhưng chưa chú trọng đến khả năng sinh từ mới của yếu tố Hán Việt trong tiếng Việt.

Dù các nghiên cứu trên đã đạt được nhiều thành tựu, tuy nhiên hầu hết chúng chủ yếu tập trung vào khía cạnh lịch sử và âm vị học (Nguyễn Tài Cẩn, Lê Đình Khẩn, Nguyễn Thị Thanh Chung), văn hoá và văn học cổ (Phan Ngọc), hoặc ứng dụng thực tiễn và khảo cứu tư liệu (Nguyễn Văn Khang, Dương Tuấn Anh...). Một vấn đề chưa được khai thác triệt để là khả năng sinh ra từ mới và mở rộng hệ thống từ vựng tiếng Việt từ đó cho thấy tiềm năng sáng tạo từ vựng. Khoảng trống này càng rõ ràng trong bối cảnh tiếng Việt hiện đại, khi nhu cầu tạo từ mới ngày càng tăng để đáp ứng sự phát triển xã hội và giao thoa ngôn ngữ.

1, 2, 3, 4, 5 Trường Đại học Sư phạm Hà Nội; Email liên hệ: hiendt@hnue.edu.vn

Gần đây nhất (2025), chúng tôi cộng tác cùng với tác giả Dương Tuấn Anh và đã công bố kết quả nghiên cứu bước đầu của việc khảo sát, nhận diện số lượng yếu tố Hán Việt và đề cập đến khả năng độc lập tạo từ của các yếu tố này trong tiếng Việt [2]. Tuy nhiên, vì đây mới chỉ là bước đầu trong quá trình nghiên cứu về tính năng sản của yếu tố Hán Việt nên chúng tôi chưa làm rõ được khả năng cụ thể và mức độ khác nhau của các yếu tố Hán Việt khi tham gia vào cấu tạo từ tiếng Việt. Bài viết này là kết quả của bước nghiên cứu tiếp theo của chúng tôi và sẽ tập trung vào thực hiện nhiệm vụ đó.

2. Phương pháp và quy trình nghiên cứu

Thực hiện nghiên cứu này, chúng tôi đã sử dụng phối hợp hai phương pháp nghiên cứu chủ yếu sau:

- Phương pháp thống kê định lượng: Phương pháp này được sử dụng để thống kê, hệ thống hoá số lượng yếu tố Hán Việt và số lượng các từ, kiểu từ mà chúng tham gia cấu tạo trong tiếng Việt. Việc làm này sẽ cung cấp một dữ liệu chính xác về số lượng các yếu tố Hán Việt có tiềm năng du nhập vào tiếng Việt và số lượng các từ tiếng Việt đã hiện thực hoá tiềm năng đó. Cũng từ dữ liệu nguyên cấp đó, chúng tôi tiến hành thống kê mô tả bằng một phép đo trung vị để tìm ra mức độ tạo từ trung bình của yếu tố Hán Việt, lấy đó làm cơ sở để phân hoá các nhóm yếu tố Hán Việt có mức độ tạo từ khác nhau.

- Phương pháp phân tích định tính: Phương pháp này được sử dụng trong việc mô tả đặc điểm của các yếu tố Hán Việt đã khảo sát. Dựa trên kết quả định lượng và những công cụ lí thuyết từ vựng học, chúng tôi thực hiện việc các mô tả đặc điểm của yếu tố Hán Việt và tương quan giữa các nhóm yếu tố về khả năng và mức độ tham gia cấu tạo từ tiếng Việt. Các dữ liệu định lượng đã giúp cho các phân tích định tính của chúng tôi có cơ sở khoa học và thực tiễn.

Quy trình nghiên cứu được chúng tôi thực hiện theo các bước như sau:

Bước 1: Khảo sát số lượng yếu tố Hán Việt để làm căn cứ tìm hiểu khả năng tham gia cấu tạo từ tiếng Việt của chúng. Chúng tôi đã tiến hành khảo sát các yếu tố Hán Việt thông dụng trong tiếng Việt dựa vào *Từ điển yếu tố Hán Việt thông dụng* [7].

Bước 2: Khảo sát số lượng từ tiếng Việt và từ đó làm rõ số lượng từ có sự tham gia của các yếu tố Hán Việt. Ở bước này, chúng tôi lựa chọn *Từ điển tiếng Việt* của Hoàng Phê [11] làm nguồn dữ liệu để khảo sát. Đây là công trình từ điển học được đánh giá cao, cập nhật đến năm 2023, bao quát được từ vựng tiếng Việt hiện đại, bao gồm cả yếu tố Hán Việt thông dụng và những biến đổi từ vựng trong những năm gần đây, đảm bảo độ tin cậy và phong phú của dữ liệu.

Bước 3: Khảo sát số lượng từ cụ thể mà mỗi yếu tố Hán Việt tham gia cấu tạo trong tiếng Việt (sử dụng kết hợp kết quả ở bước 1 và bước 2). Kết quả của bước này sẽ làm cơ sở định lượng để chúng tôi đưa ra các nhận định về khả năng tham gia tạo từ của yếu tố Hán Việt.

Bước 4: Tính toán mức độ tạo từ trung bình và phân hoá các nhóm yếu tố Hán Việt theo mức độ tạo từ của chúng (cao, trung bình, thấp), từ đó mô tả đặc điểm của từng nhóm trong mối tương quan lẫn nhau giữa chúng.

3. Kết quả nghiên cứu

3.1. Số lượng yếu tố Hán Việt trong tiếng Việt

Chúng tôi đã tiến hành khảo sát các yếu tố Hán Việt thông dụng trong tiếng Việt dựa vào nguồn ngữ liệu [7]. Kết quả thu được là: Có 2.944 yếu tố gốc Hán được tiếp thu và sử dụng trong tiếng Việt - được gọi là yếu tố Hán Việt - chiếm khoảng 12,8% tổng số yếu tố gốc Hán⁶.

⁶Theo kết quả khảo sát được công bố trong một công trình trước đây của nhóm nghiên cứu [2], tiếng Hán có khoảng 23.000 yếu tố (đều có thể đọc bằng âm đọc tiếng Việt); tuy nhiên, trong số đó, chỉ có “2942 yếu tố gốc Hán người Việt đã tiếp nhận và sử dụng để tạo từ, tương đương 12,8% tổng số yếu tố gốc Hán”. Mặt khác, sau một thời gian tiếp tục nghiên cứu, chúng tôi đã nhận diện thêm 2 yếu tố Hán Việt nữa và nâng tổng số yếu tố Hán Việt được sử dụng để tạo từ tiếng Việt thành 2.944 yếu tố.

Sau khi tiến hành rà soát và đối chiếu kết quả khảo sát từ hai nguồn dữ liệu, đồng thời loại trừ một số trường hợp nhất định, chúng tôi nhận thấy số lượng yếu tố Hán Việt thực tế sẽ giảm xuống còn hơn 2.700 đơn vị, tương đương khoảng 11,8% trên tổng số yếu tố gốc Hán. Các trường hợp loại trừ bao gồm:

(1) Những yếu tố tượng thanh (như: 嗚 ô, 嗟 ta...);

(2) Những yếu tố trùng lặp do biến âm (ví dụ: 堂 *đàng/đường*, 局 *cuộc/cục*, 日 *nhật/nhựt*, 進 *tiền/tấn*,...);

(3) Những yếu tố không phải là yếu tố Hán Việt (như: 汀 *đình* - không có âm đọc Hán Việt và đã biến đổi thành *đênh* trong từ *lênh đênh*).

Đồng thời, chúng tôi cũng bổ sung thêm một số yếu tố chưa được thu thập trước đó trong từ điển (như: 揚 *duyong*, 爺 *gia*, 氫 *khinh*...) và làm cho số lượng yếu tố Hán Việt thống kê được có sự chênh lệch nhẹ nhưng không đáng kể.

Như vậy, chúng tôi cho rằng chỉ những yếu tố gốc Hán đã được tiếp nhận và sử dụng chính thức trong tiếng Việt mới được xem là yếu tố Hán Việt. Phần lớn các yếu tố gốc Hán có âm đọc Hán Việt nhưng không được người Việt tiếp thu và sử dụng trong thực tiễn nên không được coi là yếu tố Hán Việt. Một số ví dụ về những yếu tố không được tiếp nhận như: 考 *cầu* (già cả), 齧 *củ* (sâu răng, sún răng), 懼 *cụ* (sợ hãi), 颶 *cụ* (gió bão), 誑 *cuống* (nói dối, lừa đảo), 溜 *lưu* (trượt đi, lướt đi, lạng lẽ rời đi), 碣 *kiệt* (hòn đá đứng riêng lẻ), 櫟 *kiệt* (cột buộc súc vật)...

3.2. Số lượng từ có yếu tố Hán Việt trong tiếng Việt

Chúng tôi tập trung chủ yếu vào nhận diện các từ gồm hai âm tiết (từ song tiết) trong *Từ điển tiếng Việt* [11] mà không tính đến các từ nhiều hơn hai âm tiết. Điều này xuất phát từ đặc trưng cấu trúc từ vựng của yếu tố Hán Việt: Từ Hán Việt trong tiếng Việt thường xuất hiện dưới dạng từ ghép hai âm tiết như *học sinh*, *nông nghiệp*, *kinh tế*, *văn hoá*..., phản ánh xu hướng phổ biến của từ mượn tiếng Hán được Việt hoá qua nhiều thế kỉ và là đơn vị tiêu biểu thể hiện khả năng tạo từ.

Mặt khác, để xem xét khả năng và mức độ cấu tạo từ của các yếu tố Hán Việt trong tiếng Việt, chúng tôi tiến hành khảo sát cả các từ đơn tiết nhằm làm rõ vai trò và khả năng tồn tại của yếu tố Hán Việt khi đứng một mình.

Những từ gồm ba âm tiết trở lên như *giáo dục học*, *kinh tế* ... được loại bỏ vì một số lí do sau: *i*) chúng là kết quả của quá trình ghép từ mở rộng, kết hợp yếu tố Hán Việt với thành tố khác, dẫn đến sự phức tạp trong nhận diện và phân loại; *ii*) chúng chiếm tỉ lệ nhỏ, không đại diện cho xu hướng vận động và gia nhập chung của yếu tố Hán Việt trong tiếng Việt hàng ngày; *iii*) việc khảo sát chúng sẽ làm tăng khối lượng công việc mà không mang lại giá trị đáng kể cho nghiên cứu này.

Mục đích của thống kê này là để xác định số lượng và tỉ lệ các từ có yếu tố Hán Việt trong tiếng Việt, phân biệt chúng với từ thuần Việt (như *mẹ cha*), từ mượn khác (như *Internet*), hoặc từ tượng thanh trùng lặp không đọc theo âm Hán Việt (như *rì rầm*), từ đó cung cấp dữ liệu định lượng cho việc đánh giá khả năng độc lập tạo từ và mức độ tạo từ của chúng.

Kết quả bước đầu chúng tôi thu được như sau:

Bảng 1. Số lượng từ trong *Từ điển tiếng Việt* (Hoàng Phê)

Kiểu từ	Số lượng	Tỉ lệ
Số từ đơn tiết	8.226	22.1%
Số từ song tiết	25.283	68%
Số từ loại khác	3.666	9.9%
Tổng số	37.175	100%

Theo đó, trong *Từ điển tiếng Việt* [10], có 37.175 từ, trong đó chiếm tỉ lệ cao nhất là các từ song tiết (68%). Trong số đó, số lượng từ có yếu tố Hán Việt tham gia cấu tạo là 15.027 từ song tiết (có 1 hoặc 2 yếu tố cấu tạo là yếu tố Hán Việt) và 905 yếu tố Hán Việt tạo nên từ đơn tiết tiếng Việt. Như vậy, xét một cách tương đối, trong phạm vi nguồn ngữ liệu khảo sát của chúng tôi, với 2.944 yếu tố Hán Việt, người Việt đã Việt hoá và Việt tạo ra khoảng 15.900 từ đơn tiết và song tiết tiếng Việt (tương đương khoảng 42,8% tổng số từ tiếng Việt khảo sát được trong từ điển). Số liệu này cho thấy con số chênh lệch không nhỏ so với nhận định trước đây của các nhà Việt ngữ học vẫn cho rằng có khoảng 60 - 70% đơn vị từ vựng tiếng Việt có nguồn gốc vay mượn từ tiếng Hán. Một trong số các nguyên nhân dẫn tới sự chênh lệch này là do sự vận động, thay đổi của từ vựng tiếng Việt hiện đại so với thời kì trước; mặt khác, không thể không tính đến sai số có thể xảy ra trong quá trình nhận diện, thống kê ngữ liệu khảo sát.

3.3. Khả năng độc lập tạo từ của yếu tố Hán Việt

Căn cứ vào khả năng độc lập tạo từ, các yếu tố Hán Việt với tư cách là đơn vị cấu tạo từ tiếng Việt, có thể được chia thành hai nhóm chính: hình vị độc lập (hình vị tự do) và hình vị không độc lập (hình vị phụ thuộc). Chúng tôi tiến hành khảo sát, đối chiếu sự xuất hiện của các yếu tố Hán Việt (thống kê từ nguồn dữ liệu [1]) trong các từ tiếng Việt (thống kê từ nguồn dữ liệu [2]) và thu được kết quả ban đầu như ở bảng dưới đây:

Bảng 2. Khả năng độc lập tạo từ của yếu tố Hán Việt

Yếu tố Hán Việt	Số lượng	Tỉ lệ %
Hình vị độc lập	905	30,74
Hình vị không độc lập	2039	69,26
Tổng	2944	100

Nhóm 1: Các yếu tố Hán Việt có khả năng độc lập tạo từ - Hình vị độc lập

Nhìn chung, số các yếu tố Hán Việt có khả năng độc lập tạo từ (vừa có thể đứng một mình để tạo nên các từ đơn tiết và có thể kết hợp với hình vị khác để tạo nên các từ đa tiết trong tiếng Việt) chỉ chiếm 30,7% tổng số yếu tố Hán Việt. Điều này cũng cho thấy khả năng độc lập tạo từ của các yếu tố Hán Việt thấp hơn so với các yếu tố cấu tạo từ thuần Việt vì các từ đơn tiết do yếu tố Hán Việt tạo ra chỉ chiếm khoảng 11% (905/8.226) trong tổng số từ đơn tiết mà chúng tôi khảo sát được). Ví dụ một số yếu tố:

病 *bệnh* (bệnh tật) trong câu “*Bệnh của cô ấy biến chuyển nặng.*”

花 *hoa* (bông hoa) trong câu “*Hoa nở thơm ngát.*”

京 *kinh* (kinh đô) trong câu “*Các quan đại thần được triệu về kinh ngay trong đêm.*”

陵 *lăng* (lăng tẩm) trong câu “*Bao quanh lăng là những thế núi hùng vĩ.*”

免 *miễn* (miễn trừ) trong câu “*Học sinh đạt điểm cao được miễn làm bài tập bổ sung.*”

Bên cạnh khả năng sử dụng độc lập, các yếu tố thuộc nhóm hình vị độc lập có thể kết hợp với những hình vị khác để tạo từ phức, chẳng hạn như: *bệnh án, bệnh nhân, hoa khô, kinh thành, lăng mộ, miễn chức, miễn trách,...*

Nhóm 2: Yếu tố Hán Việt không có khả năng độc lập tạo từ - Hình vị không độc lập

Yếu tố Hán Việt không có khả năng độc lập tạo từ - có vai trò là các hình vị phụ thuộc trong cấu tạo từ tiếng Việt - chiếm 69,26% trên tổng số yếu tố Hán Việt, phản ánh xu hướng phụ thuộc mạnh mẽ trong cách sử dụng chúng trong tiếng Việt. Phần lớn yếu tố Hán Việt trở thành hình vị phụ thuộc khi được tích hợp vào hệ thống ngôn ngữ của người Việt. Các hình vị phụ thuộc không thể độc lập là một từ, chỉ có thể kết hợp với các yếu tố khác để tạo thành từ có nghĩa. Ví dụ một vài yếu tố như: 犬 *khuyển* (chó) không phải là một từ trong tiếng Việt nhưng xuất hiện trong cách từ *khuyển mã, khuyển ung...*; 勸

khuyến (khuyên giải) + xuất hiện trong các từ *khuyến dụ*, *khuyến khích*, *khuyến học*, *khuyến thiện*, *khuyến nông*...; 山 *son* (núi) xuất hiện trong các từ *son cước*, *son lâm*, *giang son*; 水 *thuỷ* (nước) xuất hiện trong các từ *thuỷ thủ*, *thuỷ sản*, *thuỷ quân lục chiến*, *thuỷ canh*...

Có thể thấy, phần lớn các yếu tố Hán Việt không thể độc lập là một từ. Điều này khẳng định quá trình tiếp nhận có chọn lọc và biến đổi sáng tạo các yếu tố Hán gốc của người Việt. Quá trình tiếp nhận các yếu tố Hán không chỉ dừng lại ở việc sao chép mà còn bao gồm sự cải biên để phù hợp với văn hoá Việt. Trong quá trình lan truyền và hành chức tại cộng đồng, các yếu tố Hán gốc được Việt hoá về mặt âm đọc, ngữ nghĩa, cả phong cách đồng thời được kết hợp với các yếu tố khác để sinh từ mới, mang đặc trưng riêng của người Việt.

3.4. Mức độ tạo từ của yếu tố Hán Việt

Bên cạnh khả năng độc lập tạo từ, mức độ tạo từ của yếu tố Hán Việt cũng thể hiện rõ tính năng sản (productivity) và tính khả biến của tiếng Việt nói chung và yếu tố Hán Việt nói riêng, đồng thời cho thấy khả năng ứng biến linh hoạt của cộng đồng ngôn ngữ trong việc tạo ra và tiếp nhận từ ngữ mới, ngay cả khi chúng chưa từng xuất hiện trước đó. Nhờ hiện tượng này, kho từ vựng của tiếng Việt không ngừng được mở rộng, làm phong phú thêm và càng trở nên thích nghi với những biến động của đời sống xã hội.

Chúng tôi tiến hành khảo sát, đối chiếu hai nguồn dữ liệu như đã trình bày ở trên và bước đầu thống kê được kết quả như sau: Trong tổng số 25.283 từ song tiết tiếng Việt, số từ có chứa yếu tố Hán Việt là 15.027 từ; số lượt yếu tố Hán Việt tham gia tạo từ tổng là 24.841 và chỉ có 2716/2944 yếu tố Hán Việt tham gia cấu tạo từ (xuất hiện trong nguồn ngữ liệu [2]).

Trong tổng số từ có yếu tố Hán Việt, sự xuất hiện của các yếu tố Hán Việt lại được phân hoá cụ thể như bảng dưới đây:

Bảng 4. Số lượng từ có yếu tố Hán Việt trong tiếng Việt

STT	Kiểu từ	Số lượng	Tỉ lệ %
1	Từ có 2 âm tiết đều là yếu tố Hán Việt	11.860	78,92
2	Từ có 1 âm tiết là yếu tố Hán Việt	3.167	21,08
Tổng		15.027	100

Dựa vào 15.027 từ chứa yếu tố Hán Việt, chúng tôi tiến hành tính trung bình số lượt mà mỗi yếu tố Hán Việt tham gia vào cấu tạo từ thông qua công thức dưới đây:

$$\frac{\text{Số từ chứa yếu tố Hán Việt}}{\text{Số yếu tố Hán Việt}} \times 100\% = \frac{15027}{2944} \times 100\% \approx 5$$

Như vậy, tính trung bình, mỗi yếu tố Hán Việt tham gia vào cấu tạo khoảng 5 từ trong tiếng Việt. Căn cứ vào con số trung bình này, chúng tôi phân chia mức độ tạo từ của các yếu tố Hán Việt thành các nhóm sau:

- Nhóm 1 - Yếu tố Hán Việt có mức độ tạo từ cao: gồm các yếu tố Hán Việt tham gia vào cấu tạo trên 5 từ trong tiếng Việt.

- Nhóm 2 - Yếu tố Hán Việt có mức độ tạo từ trung bình: gồm các yếu tố Hán Việt tham gia vào cấu tạo 5 từ trong tiếng Việt.

- Nhóm yếu tố Hán Việt có mức độ tạo từ thấp: nhóm gồm các yếu tố Hán Việt tham gia vào cấu tạo dưới 5 từ trong tiếng Việt.

Kết quả phân loại yếu tố Hán Việt theo các nhóm mức độ tạo từ được cụ thể hoá trong bảng dưới đây:

Bảng 5. Mức độ tạo từ của yếu tố Hán Việt

STT	Nhóm yếu tố Hán Việt	Số lượng	Tỉ lệ
1	Mức độ tạo từ cao	1076	36,55%
2	Mức độ tạo từ trung bình	131	4,45%
3	Mức độ tạo từ thấp	1737	59%
Tổng		2944	100%

Thống kê sơ bộ dựa trên đối chiếu giữa hai cuốn từ điển và trung bình mức độ tạo từ của một yếu tố là 5 từ, chúng tôi nhận thấy, nhóm yếu tố có tính năng sản thấp mỗi yếu tố chỉ sản sinh ít hơn 5 từ gồm 1.737 yếu tố chiếm 59%, ví dụ: yếu tố 棟 *đống* chỉ xuất hiện trong từ *luơng đống* (rường và cột), 哮 *hao* chỉ xuất hiện trong từ *bào hao* (gầm thét/bồn chồn, không yên lòng), yếu tố 婪 *lam* trong *tham lam*, 氧 *duyơng* trong *duyơng khí*... Sự tồn tại của những yếu tố này mang giá trị học thuật và văn hoá nhất định. Chúng góp phần bảo lưu những lớp từ ngữ cổ, gợi nhắc đến những biểu hiện ngôn ngữ của quá khứ, đồng thời cung cấp tư liệu cho việc nghiên cứu lịch sử ngôn ngữ, văn hoá dân tộc và quá trình phát triển từ vựng tiếng Việt.

Nhóm 131 yếu tố có tính năng sản ở mức trung bình tham gia vào cấu tạo 5 từ gồm 131 yếu tố, chiếm 4,45%, chẳng hạn như, yếu tố 帥 *soái* tham gia cấu tạo các từ *soái phủ*, *chủ soái*, *nguyên soái*, *thống soái*, *tướng soái*; yếu tố 疊 *diệp* tạo ra các từ *diệp âm*, *trùng diệp*, *diệp khúc*, *diệp ngữ*, *diệp vận*. Tuy số lượng không lớn nhưng nhóm này đại diện cho những yếu tố vừa có tính phổ biến nhất định, vừa giữ được mức ổn định trong hệ thống từ vựng. Tính năng sản của nhóm này thường gắn liền với các lĩnh vực có tính chuyên môn hoá cao, vốn từ ngữ ổn định, ít bị biến đổi theo thời gian. Dù không năng động như nhóm có mức độ tạo từ cao nhưng nhóm các yếu tố này vẫn cho thấy sức sống và vai trò quan trọng trong ngôn ngữ chuyên ngành, đồng thời khẳng định sự hiện diện của yếu tố Hán Việt trong các hệ thống tri thức đặc thù của người Việt.

Nổi bật nhất là nhóm yếu tố có mức độ tạo từ cao, gồm **1.076 yếu tố**, chiếm **36,55%**. Đây là những yếu tố có khả năng kết hợp mạnh mẽ, thường xuyên tham gia vào việc hình thành từ mới và được sử dụng rộng rãi trong mọi lĩnh vực đời sống. Chẳng hạn như yếu tố 大 *đại* kết hợp với các hình vị là yếu tố Hán Việt để tạo từ có các từ *đại ác*, *đại bác*, *đại bại*, *đại bản doanh*, *đại bộ phận*, *đại ca*, *đại cáo*, *đại chiến*, *đại chúng*, *đại công nghiệp*, *đại cục*, *đại dịch*, *đại dương*, *đại đa số*, *đại đoá*, *đại đoàn kết*, *đại đội*, *đại đồng*, *đại đức*, *đại gia*, *đại gia đình*..., kết hợp với các hình vị không phải yếu tố Hán Việt để tạo từ có các từ *đại bip*, *đại bọm*, *đại ngàn*, *đại trà*, *trống đại*.... Các yếu tố 生 *sinh* xuất hiện trong 122 từ, 情 *ình* trong 109 từ, 大 *đại* trong 105 từ, 學 *học* trong 105 từ .

Kết quả khảo sát này phản ánh sự ảnh hưởng sâu sắc và bền vững của yếu tố Hán Việt trong hệ thống từ vựng tiếng Việt. Mặc dù phần lớn các yếu tố Hán Việt không thể độc lập tạo thành một từ, nhưng chúng vẫn đóng vai trò quan trọng trong việc kết hợp cấu tạo từ mới, góp phần làm phong phú và đa dạng hoá ngôn ngữ Việt. Hơn nữa, tính năng sản của yếu tố Hán Việt còn là biểu hiện cụ thể của sự giao thoa văn hoá giữa Việt Nam và Trung Hoa. Những yếu tố gốc Hán, khi đi vào tiếng Việt, không còn giữ nguyên hình thức và nội hàm ban đầu mà được bản địa hoá, biến đổi, thậm chí kết hợp với từ thuần Việt để tạo nên từ mới. Điều này cho thấy ngôn ngữ Việt không bị đồng hoá, mà trái lại, luôn chủ động tiếp thu, lựa chọn và biến đổi các yếu tố ngoại lai để làm giàu cho kho tàng từ vựng tiếng Việt.

4. Kết luận

Kết quả khảo sát, thống kê chi tiết của chúng tôi đã cho thấy một thực tế quan trọng: mặc dù tỉ lệ yếu tố Hán Việt tham gia cấu tạo từ tiếng Việt so với tổng số yếu tố gốc Hán là không lớn (chỉ khoảng 12%),

nhưng sức sản sinh từ vựng của chúng lại vô cùng mạnh mẽ, đóng vai trò quan trọng trong từ vựng tiếng Việt, đặc biệt ở lớp từ ngữ thuộc lĩnh vực hành chính, khoa học, giáo dục, chính trị, văn hoá...

Đáng chú ý, đây là lần đầu tiên các số liệu cụ thể này được thống kê và công bố dựa trên một khối ngữ liệu lớn, có kiểm chứng, thay vì chỉ dừng ở ước đoán hoặc nhận định định tính. Kết quả này có thể góp phần định hướng lại cách nhìn nhận về vai trò của yếu tố Hán Việt nói riêng và từ ngữ vay mượn gốc Hán nói chung trong tiếng Việt. Trước đây, nhiều tài liệu ngôn ngữ học thường dẫn lại con số 60-70% từ vựng tiếng Việt có nguồn gốc Hán (nhưng hầu hết đều chưa phân tách rạch ròi giữa các lớp yếu tố Hán khác nhau, chưa thống kê chi tiết số lượng yếu tố Hán Việt cũng như chưa làm rõ khả năng cấu tạo từ và mức độ hiện diện thực tế của chúng trong hệ thống từ vựng). Bằng việc lượng hoá cụ thể số lượng yếu tố Hán Việt và các kiểu từ chúng tạo ra, nghiên cứu này cung cấp một cơ sở dữ liệu đáng tin cậy để xem xét lại các nhận định truyền thống, đồng thời khẳng định rằng yếu tố Hán Việt tuy là lớp từ vay mượn, nhưng đã được Việt hoá sâu sắc và trở thành một phần hữu cơ của vốn từ tiếng Việt.

Về phương diện phương pháp nghiên cứu, nghiên cứu này chứng minh rằng việc tiến hành thống kê, phân loại dựa trên khối dữ liệu lớn và kiểm chứng chéo là hoàn toàn khả thi, đồng thời mang lại những phát hiện có giá trị học thuật lâu dài. Dữ liệu thu được có thể được sử dụng để xây dựng các từ điển chuyên biệt, các công cụ xử lí ngôn ngữ tự nhiên, cũng như phục vụ cho các nghiên cứu so sánh, đối chiếu trong Việt ngữ học và ngôn ngữ học lịch sử. Từ góc nhìn rộng hơn, đây là minh chứng cho vai trò của khảo sát định lượng trong việc cung cấp nền tảng vững chắc cho các phân tích định tính, góp phần nâng cao độ chính xác và chiều sâu của các công trình nghiên cứu về tiếng Việt.

PHỤ LỤC

1. Từ song tiết có yếu tố Hán Việt trong tiếng Việt: <https://online.fliphtml5.com/aryly/xvjp/>
2. Yếu tố Hán Việt có khả năng độc lập tạo từ tiếng Việt: <https://online.flippingbook.com/view/191005019/6/>

TÀI LIỆU THAM KHẢO

1. Đào Duy Anh. *Hán Việt từ điển*. NXB Văn hoá Thông tin. Hà Nội. 2005.
2. Dương Tuấn Anh, Lê Thị Mai Anh, Nguyễn Thị Phương Anh, Nguyễn Thị Hồng Ngọc, Vi Thị Như Quỳnh. *Tính năng sản của yếu tố Hán Việt*. Tạp chí Văn học và Tuổi trẻ, số tháng 11. 2024.
3. Nguyễn Tài Cẩn. *Nguồn gốc và quá trình hình thành cách đọc Hán Việt*. NXB Đại học Quốc gia Hà Nội. Hà Nội. 2004.
4. Nguyễn Thị Thanh Chung. *Định hướng khảo luận tâm nguyên hệ thống yếu tố Hán Việt bằng phương pháp nghiên cứu lịch sử - so sánh*, Tạp chí Khoa học Công nghệ Việt Nam, số 10. 2019.
5. Nguyễn Thị Thanh Chung. *A discussion on the technique of etymological analysis and its applicability in tracing the origin of Sino-Vietnamese elements*, HNUE Journal of Science; Social Sciences, Volume 64, Issue 11, pp. 71-80; 2019.
6. Nguyễn Thị Thanh Chung. *Khảo luận tâm nguyên từ ghép Hán Việt có cơ chế ngữ âm tương đồng với từ láy*. Tạp chí Hán Nôm, số 2, tr. 3-20. 2021.
7. Hoàng Văn Hành. *Từ điển yếu tố Hán Việt thông dụng*. NXB Khoa học xã hội. 1991.
8. Nguyễn Văn Khang. *Đặc điểm về khả năng tạo từ của các đơn vị đơn tiết Hán Việt*. Tạp chí Khoa học ngoại ngữ, số 52 (tháng 9/2017), tr.15-24. 2017.
9. Lê Đình Khản. *Từ vựng gốc Hán trong tiếng Việt*. NXB Đại học Quốc gia TP. Hồ Chí Minh. 2002.
10. Phan Ngọc. *Bản sắc văn hóa Việt Nam*, NXB Văn hóa thông tin, 1998.
11. Hoàng Phê (chủ biên). *Từ điển tiếng Việt*. Viện Ngôn ngữ học. NXB Hồng Đức. 2023.