

# Khái quát các nghiên cứu ngôn ngữ học về nguồn gốc của tiếng Việt

Mark Alves\*

*Khoa Ngoại ngữ và Triết học, Trường Đại học Montgomery County, Hoa Kỳ*

Nhận ngày 22 tháng 10 năm 2008

**Tóm tắt.** Trong khi đa số các nhà nghiên cứu ngôn ngữ học cả ở trong và ngoài nước Việt Nam cho tiếng Việt là một ngôn ngữ Môn-Khmer thuộc họ ngôn ngữ vùng Nam Đảo, thì vẫn có một số ý kiến phản bác và trong công chúng vẫn tồn tại sự nhầm lẫn. Bài viết này xem xét bốn giả thiết về nguồn gốc ngôn ngữ của tiếng Việt, các giả thiết này đặt tiếng Việt vào các nhóm ngôn ngữ khác nhau như: Nam Á, Nam Đảo, Hán hay Tai-Kadai. Trên cơ sở phương pháp luận ngôn ngữ học và các kịch bản có thể xảy ra trong giao tiếp giữa các dân tộc, thì quan điểm phổ biến cho rằng tiếng Việt thuộc hệ ngôn ngữ Nam Á vẫn khả thi nhất.

Lướt qua các từ điển bách khoa toàn thư thông dụng<sup>(1)</sup> thì giới ngôn ngữ học xếp tiếng Việt vào tiểu nhánh Môn-Khmer thuộc họ ngôn ngữ Nam Á (Austroasiatic), do đó gắn tiếng Việt vào nhóm có nguồn gốc chung với hơn 160 ngôn ngữ sử dụng trên toàn lục địa Đông Nam Á<sup>(2)</sup>. Tuy nhiên, có những người phản bác lại quan điểm này. Họ đưa ra cách phân loại khác, và xếp tiếng Việt vào cùng nhóm ngôn ngữ Vùng Nam Đảo, hoặc nhóm ngôn ngữ

thanh điệu Tai-Kadai<sup>(3)</sup>, hay Hán<sup>(4)</sup>. Yếu tố làm cho việc xếp tiếng Việt với nhóm Môn-Khmer khó có thể chứng minh một cách chắc chắn là kho từ vựng rất phong phú của tiếng Việt bao gồm từ có gốc thuộc các hệ ngôn ngữ Nam Á, Hán, Tai-kadai, và ở mức độ ít hơn, là hệ ngôn ngữ Vùng Nam Đảo - TBD. Hơn nữa, loại hình ngôn ngữ học của tiếng Việt, một ngôn ngữ đơn âm và có thanh điệu, làm cho giả thiết về nguồn gốc Nam Á của nó thậm chí còn mờ nhạt hơn, vì các ngôn ngữ Môn-Khmer thường có hai âm tiết và không có thanh điệu. Chỉ có

\* E-mail: jvs@berkeley.edu

<sup>(1)</sup> Ví dụ *Encyclopedia Britannica*, *MSN Encarta* và từ điển bách khoa toàn thư trên mạng internet tại địa chỉ: [www.Wikipedia.org](http://www.Wikipedia.org)

<sup>(2)</sup> Để có thêm số liệu về các ngôn ngữ Môn-Khmer, số người sử dụng các ngôn ngữ này và vị trí địa lý của họ, xem địa chỉ website của Viện Ngôn ngữ học mùa Hè, cơ sở dữ liệu "Ethnologue" ([www.ethnologue.com](http://www.ethnologue.com)) một tập hợp khổng lồ, mặc dù phần nào vẫn chưa hoàn chỉnh, các ngôn ngữ trên thế giới. Để đọc thêm các nghiên cứu ngôn ngữ học về hệ ngôn ngữ Môn-Khmer và các tiểu nhánh, hãy truy cập trang web của Paul Sidwell (Trường Đại học Tổng hợp quốc gia Ô-ster-rây-li-a), [http://www.anu.edu.au/%7Eu9907217/languages/language\\_s.html](http://www.anu.edu.au/%7Eu9907217/languages/language_s.html) (truy cập ngày 24/5/2006).

<sup>(3)</sup> Từ "Tai-Kadai" nói tới ba nhánh của một họ ngôn ngữ hoàn chỉnh, gồm Kam-Tai, Kadai và Htai. Trong bài viết này, từ được sử dụng chung là "Tai", mô tả một tiểu nhóm thuộc tiểu nhóm Kam-Tai. Không nên nhầm lẫn "Tai" với "Thái" (có thêm chữ h), là ngôn ngữ quốc gia của Thái - lan và chỉ là một trong 50 ngôn ngữ Tai.

<sup>(4)</sup> Từ "Hán" trong bài viết này nói tới một nhóm ngôn ngữ có liên quan với nhau - một vài nhóm lớn (ví dụ: tiếng Trung quốc phổ thông, tiếng Yue, Min v.v.) mà mỗi một thứ tiếng có hàng chục thổ ngữ hoặc các biến thể địa phương - chứ không phải để chỉ một biến thể nhất định của tiếng Hán.

việc áp dụng các công cụ tái hiện chuyên dùng trong ngành Ngôn ngữ học lịch sử thì nguồn gốc Môn-Khmer của tiếng Việt mới có thể được minh chứng một cách rõ ràng hơn.

Trong thực tế, tuy đa số các chuyên gia ngôn ngữ ở Việt Nam nhất trí với quan điểm xếp tiếng Việt vào họ Nam Á<sup>(5)</sup>, các bản đồ chính thức của Việt Nam có ghi phân bố địa lý của các nhóm dân tộc thiểu số [1] kèm theo các nhóm ngôn ngữ của họ cho thấy việc xếp loại tiếng Việt vẫn chưa dứt khoát, tuy tiếng Việt và tiếng Mường liên quan với nhau, nhưng ám chỉ rằng hai ngôn ngữ này ít nhiều có nguồn gốc khác hẳn với tất cả các nhóm ngôn ngữ khác ở Việt Nam. Sự không chắc chắn kéo dài này là đáng kể vì tiếng Việt có số lượng người sử dụng lớn nhất (trên 82 triệu) trong tất cả các ngôn ngữ được dùng ở lục địa Đông Nam Á. Do vậy, việc xem xét lại những bằng chứng đã đưa ra để ủng hộ các giả thiết đối ngược nhau là điều cần làm.

Tuy phần lớn những ý tưởng được trình bày trong bài viết này có thể đã được tổng hợp một phần ở nhiều tài liệu khác nhau, nhưng cho đến thời điểm này, chưa có một công trình nghiên cứu nào tập hợp đầy đủ các lý lẽ biện minh cho giả thiết Môn-Khmer và bác bỏ các quan điểm đối chọi khác. Mục đích của bài này là mô tả khái quát và tránh không đi sâu vào chuyên môn (những thuật ngữ cần thiết đều được chú giải) các kết quả nghiên cứu ngôn ngữ lịch sử và những tranh luận về cách phân loại tiếng Việt (chẳng hạn như ngôn ngữ này có chung nguồn gốc với những ngôn ngữ nào) dựa trên nhiều giả thiết khác nhau. Nhìn chung, trên cơ sở áp dụng thận trọng phương pháp nghiên cứu chuẩn mực của ngành ngôn ngữ học lịch sử và xét đến những bối cảnh di trú và tương tác giữa các sắc tộc có khả năng xảy ra thì giả thiết

tiếng Việt bắt nguồn từ nhóm Môn-Khmer vẫn là khả quan nhất. Đúng vậy, bằng chứng ủng hộ cho giả thiết này cũng chính là loại bằng chứng mà người ta thường dựa vào để chứng minh nguồn gốc chung của các ngôn ngữ. Hơn nữa giả thiết này còn cho thấy một hiện tượng thú vị của việc một ngôn ngữ phát triển từ phi thanh điệu sang có thanh điệu và đồng thời rút gọn từ đa âm xuống đơn âm tiết.

## 1. Tóm tắt lịch sử nghiên cứu về nguồn gốc tiếng Việt

Bản tóm tắt đầy đủ nhất các nghiên cứu đầu tiên về nguồn gốc ngôn ngữ học lịch sử của tiếng Việt là từ một bài viết của William Gage[2, 493-524]. Vấn đề phá hệ của nguồn gốc ngôn ngữ học của tiếng Việt bắt nguồn từ đầu thế kỷ 19, khi Giám mục Jean-Louis Taberd lần đầu tiên tuyên bố rằng tiếng Việt là một biến thể của tiếng Hán. Sau đó, vào năm 1856, James Logan gợi ý rằng “tiếng An-nam” là một bộ phận của họ ngôn ngữ Môn-Khmer mà ông gọi là “cấu trúc Môn-An nam”. Chắc chắn quan điểm này mang tính suy đoán vì khối lượng dữ liệu khi đó rất hạn chế[2], và đến đầu thế kỷ 20, quan điểm này đã bị thách thức. Năm 1912, sử gia đồng thời là nhà nghiên cứu ngôn ngữ học nổi tiếng người Pháp là Henri Maspéro đã viết một chuyên khảo trong đó ông khẳng định rằng tiếng Việt thuộc nhóm ngôn ngữ Tai[3]. Mãi tới giữa thế kỷ 20 một nhà ngôn ngữ học người Pháp khác, André Haudricourt xuất bản ba bài [4-6] viết cung cấp căn cứ ngôn ngữ học theo đó tiếng Việt có thể được chứng minh một cách thuyết phục là có nguồn gốc thuộc hệ ngôn ngữ Nam Á cả trên cơ sở dữ liệu về từ vựng và âm vị<sup>(6)</sup>. Từ đó, các nghiên cứu về đề tài này chủ yếu tập trung vào các mối quan hệ giữa tiếng Việt với nhóm Môn-Khmer

<sup>(5)</sup> Thực ra đây là một ý kiến mang tính dự đoán, song có vẻ là ý kiến chung của các nhà ngôn ngữ tại Viện Ngôn ngữ học Hà Nội và hầu hết các nhà nghiên cứu ngôn ngữ ở các trường đại học khác của Việt Nam. Để xem các ý kiến về chủ đề này, hãy truy cập vào trang web của Viện Ngôn ngữ học Hà Nội ([www.ngonngu.net](http://www.ngonngu.net)) trong đó nêu sự chấp nhận về nguồn gốc Nam Á của tiếng Việt.

<sup>(6)</sup> Trong một bài báo của mình viết năm 1954 “Nguồn gốc thanh trong tiếng Việt” Haudricourt cũng đưa ra giả thiết về nguồn gốc của các thanh điệu trong tiếng Việt, một học thuyết có tác động lớn đến công tác nghiên cứu ngôn ngữ học lịch sử về các ngôn ngữ Đông và Đông Nam Á.

Michel Ferlus [7,8], Gage[2], Gerard Diffloth[9,10], Nguyễn Văn Lợi[11] và Nguyễn Tài Căn [12] đã đưa ra một số ấn phẩm quan trọng nhất về vấn đề này, và các công trình của nhiều học giả khác đã cung cấp thêm các bằng chứng hỗ trợ. Tuy nhiên, trong con mắt của một ồ học giả, một số yếu tố phức tạp làm cho vấn đề này vẫn chưa được giải quyết.

## 1. Công cụ lý thuyết của Ngôn ngữ học lịch sử

Giống như trong khảo cổ học, các phương pháp tiếp cận trong ngành ngôn ngữ học tuy thuộc vào tính thuyết phục của các bằng chứng và các giả thiết hợp lý<sup>(7)</sup>. Suy cho cùng, những suy đoán về quá khứ xa xưa không thể có sự chắc chắn tuyệt đối, nhưng có thể dựa trên các dữ liệu về ngôn ngữ, văn hoá, lịch sử hiện có có thể đưa ra các giả định lô-gic và loại trừ một số khả năng. Sử dụng các thông tin như vậy, người ta có thể trình bày quan điểm về cơ cấu ngôn ngữ học của các ngôn ngữ trong quá khứ (được gọi là “phục nguyên”), tương tác giữa các ngôn ngữ và các nhóm ngôn ngữ (ví dụ sự vay mượn các thành tố từ vựng và cấu trúc ngôn ngữ), và các mối quan hệ phá hệ ngôn ngữ học (ví dụ mối quan hệ với một họ ngôn ngữ nhất định) lại thành tố cơ bản cho phép các nhà nghiên cứu xếp hai ngôn ngữ vào một họ ngôn ngữ là (1) một bộ từ vựng cơ bản chung, và (2) các bộ phận tương xứng lặp lại nhiều lần giữa hai ngôn ngữ (ví dụ một âm trong một ngôn ngữ giống một âm hay tương tự như một âm trong nhiều từ cùng gốc [tức là các từ có chung nguồn gốc] thuộc một ngôn ngữ khác, do đó tạo ra một mô hình tương xứng). Thời gian quá xa xưa có thể làm cho một số vấn đề trở nên đáng nghi ngờ vì đã trải qua nhiều thay đổi, tuy nhiên ngay như

vậy thì mô hình tương xứng về âm vị cũng cần đủ hoàn chỉnh để có thể phục nguyên một ngôn ngữ nguyên thủy, tức là một ngôn ngữ trong quá khứ mà từ đó các ngôn ngữ liên quan được hình thành. Ngoài ra, khi điều kiện cho phép, các chi tiết về lịch sử, khảo cổ, và/hoặc nhân chủng học cần cung cấp các bằng chứng bổ sung và các kịch bản khả thi để giải thích cho việc một ngôn ngữ tách thành nhiều ngôn ngữ.

Thuật ngữ “từ vựng cơ bản” vừa có tính khách quan, vừa rất ấn tượng, song đối với các hệ ngôn ngữ đã hình thành vững chắc như Ấn - Âu, một hệ liên kết các ngôn ngữ có phạm vi địa lý rộng lớn như tiếng Hindi và tiếng Anh, và hệ Hán - Tạng, một nhóm gồm hơn hai trăm ngôn ngữ khác nhau[13,14], bao gồm cả những ngôn ngữ mà tên gọi của nó đã nói nên phần nào bản chất, tiếng Hán và tiếng Tây Tạng - thì thuật ngữ này có thể minh họa dễ dàng. Từ vựng cơ bản gồm các loại từ của một ngôn ngữ ít có khả năng mất đi trong thời gian hàng thế kỷ hay thiên niên kỷ và có nhiều khả năng tồn tại trong thời gian dài hơn các loại từ vựng không phải là cơ bản khác. Những từ có thể được xem là thực sự cơ bản bao gồm các, các bộ phận của cơ thể, các hiện tượng thiên nhiên và động vật thông thường, và các hành vi/hoạt động cơ bản, ngoài các loại ngữ nghĩa thông thường khác và các khía cạnh về sự tồn tại của con người. Đây chính xác là các từ mà chúng không chỉ liên kết các ngôn ngữ Môn-Khme thành một hệ ngôn ngữ, mà còn liên kết chúng với ngôn ngữ tiếng Việt. Trong khi xác định các mối liên kết ngôn ngữ học, các từ có ngữ nghĩa chung chung hơn hữu ích<sup>(8)</sup> hơn các từ có nghĩa cụ thể hơn nhưng lại gắn với các từ thuộc các nhánh nhỏ trong phạm vi các chủng loại rộng hơn. Ví dụ, từ *con chim* [bird], một từ nguyên Môn-Khme nói tới cả một phạm trù ngữ nghĩa, trong khi từ *bồ câu* [pigeon], có thể có gốc từ ngôn ngữ Tai-Kadai, chỉ một loài chim cụ thể, mà nó làm cho từ nguyên Môn-Khme càng hữu

<sup>(7)</sup> Một khác biệt quan trọng giữa ngành khảo cổ tự nhiên và ngôn ngữ học lịch sử là trong khảo cổ, tuổi của các cổ vật được đánh giá bằng niên đại các - bon. Ngôn ngữ học lịch sử suy cho cùng là một ngành không có công cụ chẩn đoán. Công cụ đánh dấu niên đại khách quan gần nhất là các lưu trữ về chữ viết, nhưng chỉ trong trường hợp niên đại của các tài liệu này có thể xác minh được.

<sup>(8)</sup> Từ điển từ nguyên và từ đồng nghĩa Hán - Tây Tạng, phần “Giới thiệu” <http://stedt.berkeley.edu/> (accessed May 24, 2006).

ích hơn. Từ vựng cơ bản có xu hướng đề kháng với các giao lưu ngôn ngữ vừa phải (tức là giao lưu giữa các nhóm dân tộc dẫn đến những thay đổi về ngôn ngữ), một tình huống diễn ra một số vay mượn từ vựng<sup>(9)</sup>. Điều này làm cho chúng loại từ vựng cơ bản trở nên hữu ích hơn cho việc xác định nguồn gốc ngôn ngữ chung của các ngôn ngữ. Cuối cùng, môn thống kê từ vựng học, tức là việc đánh giá mức độ sử dụng chung từ vựng cơ bản giữa các ngôn ngữ, là một công cụ hỗ trợ có thể sử dụng để xác định mức độ liên hệ giữa các ngôn ngữ và đánh giá được phần nào các bằng chứng của nguồn gốc ngôn ngữ. Phương pháp thống kê thẳng thắn này hẳn có thể là nền tảng của các nghiên cứu sau này, nhưng không nhất thiết đưa đến kết quả chắc chắn bởi vì ngôn ngữ vay mượn từ vựng và đôi khi với số lượng lớn.

Bằng chứng về sự cùng nguồn gốc của hai ngôn ngữ liên quan đến việc xác định mô hình âm vị của hai âm đối lập và song song nhau. Hơn nữa, nếu xuất hiện càng nhiều ví dụ về các mô hình như vậy, thì lập luận về tình trạng cùng nguồn gốc càng vững chắc. Ví dụ, tìm được vài ví dụ về các từ có âm răng /d/ trong một ngôn ngữ và âm răng /n/ trong một ngôn ngữ khác sẽ củng cố thêm lập luận rằng hai ngôn ngữ này ít nhiều có liên hệ với nhau, dù có thể nói chắc chắn về các quan hệ phá hệ hay giao tiếp giữa hai ngôn ngữ này hay không. Để xác định cùng nguồn gốc phá hệ, sự tương xứng về âm vị giữa các bộ từ vựng cơ bản trong hai ngôn ngữ cần bao gồm đủ các mục từ vựng để phục nguyên các hệ thống âm hoàn chỉnh và, lý tưởng nhất là một dạng phục nguyên văn hoá xã hội nào đó (tức là sự phục nguyên một hệ thống văn hoá và cách sống chứ không chỉ là hệ thống ngôn ngữ) dựa trên từ vựng cơ bản. Việc này cũng đòi hỏi phải so sánh các giai đoạn trước đó đã được

phục nguyên của các nhóm ngôn ngữ chứ không chỉ là hai ngôn ngữ trong dạng hiện đại của chúng. Có được các từ cùng gốc với sự tương xứng về âm vị trong nhiều ngôn ngữ khác nhau, kể cả các ngôn ngữ xa nhau về địa lý (do đó giảm bớt khả năng vay mượn), có thể củng cố thêm lập luận về cùng nguồn gốc phá hệ. Nếu không có loại bằng chứng này, các lập luận đó sẽ kém thuyết phục hơn.

Trong khi việc xác định kho từ vựng cơ bản và sự tương ứng âm vị đều đặn vẫn là cơ sở để đánh giá cùng nguồn gốc ngôn ngữ, trong trường hợp cụ thể của tiếng Việt, cần có thêm tiêu chuẩn để giảm bớt hơn nữa tính không chắc chắn có thể xảy ra. Một trong các nguy cơ của Sprachbund, hay gọi là vùng ngôn ngữ, như Đông Nam Á, là từ ở các ngôn ngữ khác nhau lại có thể giống nhau về cả âm vị và ngữ nghĩa do sự giống nhau về tượng hình trong các hệ thống âm vị và chức năng ngữ nghĩa, ngữ pháp, chức năng ngôn ngữ thực tiễn nhưng vẫn chưa hình thành các từ nguyên liên hệ với nhau. Hiện tượng “từ cùng gốc giả” hay “trông giống nhau”, các trường hợp giống nhau ngẫu nhiên về âm vị hay ngữ nghĩa là một vấn đề rắc rối nổi bật trong ngôn ngữ học Đông Nam Á vì nhiều ngôn ngữ ở vùng này có số từ đơn âm đáng kể, ngoài những giống nhau về tượng hình khác. Có xu hướng là một số từ nhất định hay loại từ nhất định rất phổ biến trong các ngôn ngữ ở khu vực này, như vùng ngôn ngữ Đông Nam Á, nơi có nhiều hoạt động giao tiếp giữa các nhóm dân tộc đến mức chúng thực sự làm giảm khả năng của chúng ta trong việc thiết lập cùng nguồn gốc phá hệ. Trường hợp này phổ biến trên toàn thế giới đối với trường hợp từ “mẹ”, một từ thường bắt đầu bằng âm /m/, hay từ “bố” thường bắt đầu bằng âm /f/ hay một âm tương tự. Ở Đông và Đông Nam Á, những từ tương tự nhau trong nhiều hệ ngôn ngữ gồm các từ điển đạt “mất” hay các từ chỉ định “này/kia”, và trên cơ sở các dữ liệu hiện có hay về phương pháp luận chưa có thể nói được liệu các trường hợp giống nhau này hay những trường hợp khác là kết quả của những nguồn gốc xa xưa hơn giữa các hệ ngôn ngữ hay đơn thuần chỉ là các giao lưu ngôn ngữ.

<sup>(9)</sup> Cụm từ “more useful - hữu ích hơn” ở đây nên được hiểu với nghĩa tương đối. Bởi vấn đề là sức nặng của bằng chứng, nên rất có thể là một số từ vựng phần nào ít cơ bản hơn lại có thể là có cùng nguồn gốc. Các ý kiến cực đoan về bất kỳ phía nào cũng không hợp lý, nhưng chúng tôi hy vọng rằng việc tích lũy các chi tiết đủ để bảo vệ một quan điểm nào đó.

Tốt nhất là loại trừ những từ có thể có nhiều nguồn gốc ngôn ngữ, hay ít nhất thừa nhận chúng là bằng chứng kém thuyết phục hơn của lập luận về nguồn gốc giữa các họ ngôn ngữ riêng biệt, một quan điểm ít được xem xét đến trong các nghiên cứu lịch sử ngôn ngữ ở Đông Nam Á. Điều cốt yếu là phải xem xét yếu tố địa lý, như trường hợp Bắc Việt Nam, khi kiểm tra những liên quan về từ vựng có thể có với các ngôn ngữ láng giềng - mà không coi trọng nhóm này hơn nhóm kia - bởi vì vùng này là một mắt xích địa lý giữa Trung Quốc, lục địa Đông Nam Á với vùng đảo Đông Nam Á. Cuối cùng, hiện tượng tượng thanh là một vấn đề nữa làm suy yếu lập luận về từ cùng nguồn gốc; ví dụ tiếng Việt có từ *cắt* [cut] cũng như các dạng ngữ âm tương tự của cùng một từ trong các ngôn ngữ Môn-Khme (ví dụ tiếng Khme *kat*), tiếng Tai (ví dụ *tət*), và tiếng Trung Quốc phía nam (ví dụ tiếng Quảng Đông *chit*[15]). Một phương pháp tiếp cận bảo thủ hơn có thể gồm rất nhiều các từ cùng gốc nhưng sau đó loại bỏ hay xử lý cẩn thận tới mức có thể theo các tiêu chí loại bỏ nhất định (xem danh mục dưới đây). Như vậy, những dữ liệu còn lại có thể có nhiều giá trị chứng minh cho nguồn gốc ngôn ngữ hơn. Các tiêu chí để xác định nguồn gốc ngôn ngữ có thể gồm:

1. Số lượng từ vựng cơ bản (toàn bộ)
2. Các mô hình tương xứng về âm vị của từ vựng cơ bản (toàn bộ)
3. Số lượng các ngôn ngữ trong một hệ ngôn ngữ có từ gốc chung (toàn bộ)
4. Các khả năng đa dạng trong các hệ ngôn ngữ khác nhau (loại trừ)
5. Tính chất tượng thanh (loại trừ)

Ngoài phương pháp luận chung được mô tả ở trên, các kịch bản về lịch sử của sự di cư và giao lưu của con người cũng phải được xem xét. Nếu bằng chứng ngôn ngữ cho thấy hai ngôn ngữ có chung bối cảnh lịch sử thì bằng chứng này gợi ra mối quan hệ dân tộc nào? Trước những dữ liệu lịch sử có được thì các giả thiết vững chắc đến đâu, và ta đã biết được gì về các hoàn cảnh ngôn ngữ, xã hội và địa lý? Mặc dù các nghiên cứu như vậy hẳn mang tính

phòng đoán, tính vững chắc của mỗi giả thiết về nguồn gốc ngôn ngữ khi nói về con người phải được xem xét ít nhiều.

### 3. Tiếng Hán<sup>(10)</sup>

Sự giao tiếp về văn hoá của người Việt với người Trung Quốc có từ hơn hai ngàn năm trước đây - một ngàn năm đô hộ của Trung Quốc, tiếp theo là một ngàn năm của thân phận chư hầu phải cống nạp và sự giao lưu văn hoá với các cường quốc phía bắc. Như nói ở trên, giả thiết về nguồn gốc phả hệ của tiếng Việt với tiếng Hán có từ thời thuộc địa và có vẻ vẫn còn rơi rớt trong các giả thiết phổ biến. Quan điểm như vậy chủ yếu dựa vào số lượng rất lớn dễ nhận thấy từ vựng Hán trong tiếng Việt, những nét giống nhau quan trọng về loại hình ngôn ngữ giữa hai ngôn ngữ và, ít nhất tới đầu thế kỷ hai mươi, là việc sử dụng chữ viết Hán trong tiếng Việt (song song với chữ *Nôm* dựa trên cách viết của tiếng Hán).

Tuy nhiên, những khía cạnh này đều có vấn đề. Từ ngữ có thể được vay mượn, và chính là loại từ dùng chung, chứ không chỉ là số lượng, cùng cố các giả thiết về nguồn gốc chung. Tiếp theo, trong khi cả tiếng Hán và tiếng Việt đều có thanh điệu, thì các ngôn ngữ lân cận như Tai-Kadai và Hơ-mông - Miên cũng có. Cuối cùng, chữ viết (một sản phẩm của con người có thể lưu truyền và không phải là một tài năng di truyền) chung cho cả hai ngôn ngữ rõ ràng không phải là một chỉ thị của nguồn gốc ngôn ngữ gốc; chỉ có ngôn ngữ nói được đại diện bằng chữ viết là có thể sử dụng được cho mục đích này. Một điều quan trọng nữa cần phải lưu ý là những giống nhau giữa tiếng Việt và tiếng Hán không nhất thiết là kết quả của ảnh hưởng của tiếng Hán. Một số nét giống nhau có thể là kết quả của khu vực địa lý bất kể hệ ngôn ngữ nào, hay những sự giống nhau cũng có thể là

<sup>(10)</sup> Nói chính xác hơn, tiếng Hán là một nhóm ngôn ngữ. Hai người mà nói những phương ngữ Trung Quốc khác, ví dụ tiếng Bắc kinh và tiếng Quảng đông, không có thể hiểu nhau trừ khi họ có học phương ngữ Trung Quốc khác.

kết quả của những thay đổi tự nhiên nội tại của ngôn ngữ. Ít nhất, nói rằng thực trạng hiện nay của tiếng Việt được xác định chỉ bằng những giao tiếp với tiếng Hán là quá đơn giản hoá sự việc [16, 221-242].

Trong khi việc xem xét những nét giống nhau nổi bật giữa tiếng Việt và tiếng Hán rất lý thú và quan trọng, việc xem xét những nét không giống nhau hay những gì trong tiếng Việt không bắt nguồn từ tiếng Hán cũng quan trọng không kém. Đặc biệt chúng ta phải nhận ra các yếu tố cần thiết để xác lập nguồn gốc ngôn ngữ giữa tiếng Việt và tiếng Hán. Trừ một số ít tài liệu và các trường hợp chuyên biệt, không một đại từ hay số từ nào trong tiếng Việt có nguồn gốc từ tiếng Hán<sup>(11)</sup>. Những từ biểu đạt các hiện tượng thiên nhiên thực sự cơ bản cũng không xuất phát từ tiếng Hán, ví dụ từ *chó* /dog/, *chim* /bird/, *củi* /firewood/, nước /water/, lá /leaf/, rễ /root/ và một số từ khác, mà trong thực tế có nguồn gốc từ hệ Môn-Khme<sup>(12)</sup>. Nói chung, những loại từ vựng cơ bản gắn kết nhiều biến thể khác nhau của tiếng Hán thành một nhóm ngôn ngữ hầu như không có trong tiếng Việt<sup>(13)</sup>. Về ngữ pháp, trật tự danh từ trong cụm từ tiếng Việt đặt danh từ trước từ bổ nghĩa (một đặc tính có cả trong ngôn ngữ Môn-Khme và Tai ở khu vực này) là ngược lại với tiếng Hán từ bổ nghĩa đứng trước danh từ. Quả vậy, loại từ vựng mà tiếng Việt vay mượn từ tiếng Hán phần lớn

chính là loại từ vựng mà các ngôn ngữ thường vay mượn nhau: đó là từ vựng chỉ các vật dụng mang tính văn hoá cụ thể và các biểu đạt về văn hoá (ví dụ các dụng cụ trong gia đình, các vật dụng hay khía cạnh về y tế, về cơ quan quản lý của chính phủ, và nghệ thuật), và không phải là từ vựng cơ bản là các từ tương đối có sức chống lại việc vay mượn.

Thay vì các từ có nguồn gốc chung với tiếng Hán, tồn tại các tầng từ vay mượn từ tiếng Hán. Những nhà nghiên cứu như Vương Ly[17], Haudricourt [6], Đào Duy Anh[18] và các tác giả khác có sử dụng truyền thống âm vị lịch sử của tiếng Hán đã xác định được một tầng cổ xưa của từ vựng Hán từ Triều đại nhà Hán (năm 206 trước Công Nguyên đến 220 sau Công Nguyên), bằng cách xác định các mô hình thay đổi trong các phụ âm đầu và thanh điệu. Những từ như “kind/type”, “paper” và “well” cho thấy mô hình của âm vòm ban đầu và thanh sắc[19] được minh hoạ trong Bảng 1. Những từ như vậy thường được người Việt xem là một phần của chữ *Nôm* truyền thống, chữ viết biểu đạt tiếng Việt nói, chứ không phải là chữ Hán. Điều này trái ngược với một tầng từ vựng vay mượn sau này từ tiếng Hán thuộc triều đại nhà Đường (năm 618 - 907), được gọi là Hán - Việt, trong đó mỗi từ là một chữ viết Hán (*Hán tự*) phát âm kiểu Việt.

Bảng 1. Các từ vay mượn Hán - Việt và Hán - Việt cổ (HV)

Chú giải	Hán - Việt	Hán - Việt cổ	Chữ viết
kind/type	chùng	giống	種 種
paper	chi	giấy	紙 紙
Well	trình	giếng	井 井

Điều quan trọng là phải lưu ý rằng truyền thống ngữ âm học lịch sử Hán phải được tính tới để thiết lập sự chắc chắn về nguồn gốc của các từ cho mượn xa xưa này<sup>(14)</sup>. Nếu không ứng dụng các nghiên cứu ngôn ngữ học lịch sử về

<sup>(11)</sup> Ví dụ là muôn /ten thousand/, từ Hán cổ của chữ Hán-Việt vạn 萬 và y /he-she/, chữ Hán 伊.

<sup>(12)</sup> Những bằng chứng như vậy là kết quả của việc so sánh với nhiều ngôn ngữ Môn-Khme. Trong một số trường hợp, những từ như “fish – cá” và “leaf – lá” thậm chí có quan hệ rộng lớn hơn với các nhóm ngôn ngữ khác trong vùng (“fish – cá” trong ngôn ngữ vùng Nam Đảo, và “leaf – cá” trong cả ngôn ngữ Hán – Tây tạng và ngôn ngữ vùng Nam Đảo), nhưng chỉ những nghiên cứu có tính chất hệ thống áp dụng phương pháp phê phán mạnh mẽ mới có thể xác định được điều này một cách chắc chắn hơn. Hơn nữa, một số lượng ít ỏi từ có thể là cùng gốc sẽ không đủ cơ sở để xác định liệu vấn đề này là sự giao tiếp về ngôn ngữ hay nguồn gốc ngôn ngữ chung. Laurent Sagart thậm chí đưa ra quan điểm về sự quan hệ lâu đời giữa các ngôn ngữ Hán-Tây tạng, Nam Á và Tai-Kadai.

<sup>(13)</sup> Jerry Norman đã đưa ra một danh sách rất hay về các từ nguyên Hán – Tây Tạng; không một từ nào trong danh sách này có vẻ có từ cùng gốc tương ứng trong vốn từ vựng bản địa của tiếng Việt.

<sup>(14)</sup> Công trình ngôn ngữ học lịch sử hoàn chỉnh nhất về Hán - Việt, với ý kiến thảo luận đầy đủ về truyền thống âm vị lịch sử của tiếng Hán.

các “từ điển gieo vần” tiếng Hán<sup>(15)</sup> ngược trở lại thời gian khoảng một ngàn năm trăm năm và các nghiên cứu liên quan về các biến thể hiện đại của tiếng Hán cũng như cách đọc chữ Hán của người Nhật và người Triều tiên, thì không thể có được lập luận xác đáng về nguồn gốc ngôn ngữ với tiếng Hán. Trong một số trường hợp, thông qua việc vay mượn lẽ tẻ từ ngữ của một ngôn ngữ khác, các từ riêng biệt có thể thay đổi hình thái âm vị của chúng một cách thường xuyên, song hiện tượng các mô hình tượng hình chung và tự nhiên lặp đi lặp lại nhiều lần của những thay đổi về âm thì phải được xác định để chứng minh cho một trường hợp khả thi nào đó.

Hiện nay, với sự ứng dụng phương pháp luận ngôn ngữ học lịch sử một cách lô-gic, bất kỳ một lập luận nào cho rằng tiếng Việt có nguồn gốc từ tiếng Hán và vì thế có quan hệ với hệ ngôn ngữ lớn hơn là Hán - Tây tạng, là không có cơ sở.

#### 4. Hệ ngôn ngữ Tai-Kadai

Tuyên bố hùng hồn nhất về nguồn gốc Tai của tiếng Việt do Maspéro đưa ra trong một bài viết dài 120 trang đăng trên tạp chí *L'École Francoise d'Extrême - Orient*<sup>(16)</sup>. Trong bài viết đó, Maspéro chỉ ra khoảng một trăm từ mà ông coi là có cùng nguồn gốc chung giữa tiếng Việt và tiếng Tai-Kadai. Mặc dù ông cũng thừa nhận các từ cùng nguồn gốc Môn-Khme trong tiếng Việt, nhưng ông cho rằng dựa trên số từ nguyên Tai rõ ràng và các đặc tính âm vị chung (như thanh điệu và cơ cấu âm tiết đơn âm là phổ biến), tiếng Việt thuộc hệ ngôn ngữ Tai. Tuy

nhiên, Haudricourt[5] chỉ ra rằng một số lượng đáng kể hình thái mà người ta cho là ngôn ngữ Tai trong thực tế có nguồn gốc từ tiếng Hán. Hơn nữa, có thể thấy rằng dữ liệu của Maspéro bộc lộ nhiều trường hợp là bằng chứng kém thuyết phục về từ vựng và âm vị, như các trường hợp từ tượng thanh (ví dụ *mèo* [cat]), hay các hình thái được thấy trong các ngôn ngữ ở khắp vùng Đông Nam Á, như đã được nêu trong phần nói về các công cụ lý thuyết của ngôn ngữ học lịch sử. Một khả năng phức tạp khác là một số tiếng Tai trong lãnh thổ Việt nam lại vay mượn từ tiếng Việt, một khả năng thường không được các học giả trong ngành văn học về ngôn ngữ học lịch sử Đông Nam Á thảo luận tới. Trong tình hình như vậy, các dữ liệu về tiếng Tai vay mượn từ tiếng Việt có thể tạo ra một cảm tưởng sai lầm về từ tiếng Tai mượn từ tiếng Việt. Cuối cùng, sự thiếu vắng vốn từ vựng cơ bản chung, như đại từ, số từ, từ chỉ các bộ phận của cơ thể, hay các danh từ và động từ cơ bản khác khá có sức thuyết phục trong việc không xếp tiếng Việt vào hệ ngôn ngữ Tai-Kadai.

Điều đó không có nghĩa là tiếng Việt không giao tiếp với các ngôn ngữ Tai. Một danh mục từ mới được Nguyễn Tài Căn[12] tổng hợp. Ông đã sưu tầm được gần hai chục từ có thể được tiếng Việt vay mượn từ tiếng Tai. Việc so sánh các từ này với các phục nguyên từ Tai nguyên thủy<sup>(17)</sup> cho thấy rằng đúng là có một số hình thái có thể là các ứng cử viên nặng ký của các từ vay mượn. Phương pháp tiếp cận tốt nhất có thể là so sánh các hình thái được phục nguyên với những từ này của tiếng Việt với các hình thái tiếng Tai nguyên thủy được phục nguyên. Nếu không, các mô hình về sự biến đổi âm không thể đánh giá chính xác được, và vì thế khó tránh các giả thiết sai lầm. Một công trình như vậy đang chờ đợi sự phục nguyên hoàn chỉnh nhóm ngôn ngữ Vietic nguyên thủy (tức là nhóm ngôn ngữ bao gồm tiếng Việt,

<sup>(15)</sup> Các từ điển gieo vần cho thấy các chủng loại phát âm chữ Hán thuộc triều đại Đường và Tống. Sự nhận biết chính xác về các âm được trình bày có thể được suy ra chỉ bằng việc so sánh những chủng loại này với các biến thể hiện đại của tiếng Hán, một đề tài nghiên cứu ngôn ngữ học lịch sử từ cuối những năm 1800.

<sup>(16)</sup> Maspéro, “Études sur la Phonétique Historique de la Langue Annamite: Les Initiales.” Một công trình nghiên cứu ít được trích dẫn hơn chứa đựng điều mà tác giả coi là các từ cùng gốc giữa tiếng Việt và một vài tiếng Tai khác.

<sup>(17)</sup> Những từ vay mượn từ ngôn ngữ Tai trong tiếng Việt do Nguyễn Tài Căn đưa ra được so sánh với “Proto - Tai - O-Matric” trực tuyến, một cơ sở dữ liệu các từ proto-Tai được tái hiện.

tiếng Mường và khoảng hai chục ngôn ngữ Chứt)<sup>(18)</sup>. Rất nhiều từ vay mượn từ tiếng Tai sang tiếng Việt liên quan đến nông nghiệp (ví dụ: vịt [duck] và đực [male animal]), một tình huống gợi ý về một mối giao lưu ngôn ngữ nhưng không phải là mối giao lưu có ảnh hưởng đáng kể tới tiếng Việt.

Sự giống nhau về hệ thống phân loại giữa tiếng Việt và tiếng Tai, nhất là về thanh điệu, đã được các học giả theo trường phái này cho là biểu hiện của cùng nguồn gốc, hay có thể là kết quả của những giao tiếp lâu dài mà tiếng Việt đã phát triển một số đặc tính của tiếng Tai, nhưng khu vực địa lý của những ngôn ngữ có thanh điệu và biệt lập/đơn âm ở khắp Trung Quốc và lục địa Đông Nam Á rộng lớn tới mức không thể chỉ sử dụng sự có mặt của thanh điệu (mà không dùng các yếu tố khác, nhất là sự tương thích về âm vị) là căn cứ chứng minh nguồn gốc ngôn ngữ. Quả vậy, ngay cả một số ngôn ngữ Môn-Khme và ngôn ngữ Chăm ở vùng Nam Đảo cũng có xu hướng đơn âm và phát triển thanh điệu<sup>(19)</sup>. Việc sử dụng đặc tính hệ thống phân loại theo vùng này để củng cố các lập luận về nguồn gốc ngôn ngữ tốt nhất cũng rất rủi ro. Dựa trên cơ sở của sự gần gũi về địa lý giữa tiếng Việt và nhóm tiếng Tai và trên cơ sở một số bằng chứng về từ vựng nói trên, có thể nói một cách tương đối chắc chắn rằng hai nghìn năm trước tổ tiên người Việt đã giao lưu với người Tai ở khu vực Nam Trung Quốc và Bắc Việt Nam hiện nay. Phạm Đức Dương đã thảo luận về điều có vẻ là cơ sở ngôn ngữ học – các bằng chứng về từ vựng nói trên – để chứng minh cho sự vay mượn của nền văn hoá lúa nước giữa người Việt và người Tai[20]. Bằng chứng về gien sinh học cũng gợi ra những liên kết chặt chẽ có thể có giữa các nhóm người Việt với người Tai, song các công trình nghiên cứu khác lại chỉ ra các quan hệ giữa các nhóm người Môn-Khme với người Tai-Kadai[21]. Tuy nhiên, nhìn chung, dựa trên các dữ liệu hiện tại và các phương pháp nghiên cứu, nhiều

nhất thì mối quan hệ Việt - Tai chỉ có thể được xem là một trong các giao lưu ngôn ngữ vừa phải.

## 5. Các nguồn gốc Thái Bình Dương: Nhóm Nam Đảo và tiếng Nhật

Cũng tồn tại một nhóm học thuyết liên quan và trùng lặp khác, tất cả đều gợi ý về nguồn gốc Thái Bình Dương của tiếng Việt, kể cả những lập luận về nguồn gốc chung với họ ngôn ngữ vùng Nam Đảo và tiếng Nhật. Keith Taylor, trong phần mở đầu cuốn sách *Sự ra đời của Việt Nam* của mình, đã viết nhận xét dưới đây, dựa trên cơ sở những phân tích của ông về truyền thống văn hoá và dân gian: "Những truyền thống huyền thoại về Lạc Long Quân và nguồn gốc của các Vua Hùng cho thấy một nền văn hoá hướng về biển để thích nghi với môi trường lục địa. Nền văn minh đã xuất hiện cùng với một anh hùng văn hoá tới từ vùng biển đã đánh bại quyền lực lục địa bằng cách bắt giữ vợ của kẻ thù của mình và biến người phụ nữ đó thành mẹ của những đứa con thừa tự của mình. Nguồn gốc huyền thoại của các Vua Hùng phản ánh một nền văn hoá biển dựa trên sự lớn dần về chính trị từ các ảnh hưởng của lục địa"[22].

Những người khác đã đưa ra bằng chứng về từ vựng để hỗ trợ cho một khái niệm như vậy. Những công trình nghiên cứu ban đầu để chứng minh cho mối quan hệ giữa tiếng Việt với ngôn ngữ vùng Nam Đảo là của Nobuhiro Matsumoto, Bình Nguyên Lộc và Nguyễn Ngọc Bích[23-25]. Matsumoto là người đi xa nhất, gợi ra mối quan hệ giữa tiếng Việt và cả tiếng Nhật lẫn tiếng Malai-Pô-lê-ni-zian. Ông đã đưa ra hơn bảy mươi từ gốc có khả năng liên quan giữa tiếng Việt và tiếng Nhật, nhưng chỉ có 9 hình thái để bảo vệ cho quan điểm về mối liên kết giữa tiếng Việt và tiếng vùng Nam Đảo<sup>(20)</sup>. Bằng chứng này bao gồm một số từ vựng có vẻ

<sup>(18)</sup> Như đã được thảo luận bởi Nguyễn Tài Căn trong *Giáo trình lịch sử ngữ âm tiếng Việt*

<sup>(19)</sup> Để xem các ý kiến thảo luận tương tự.

<sup>(20)</sup> Khối lượng lớn các từ cùng gốc Nhật - Việt này được một người sử dụng tiếng Nhật đưa ra có thể là một dấu hiệu của khả năng sáng tạo để các nhà ngôn ngữ học nhận ra những mô hình trong các ngôn ngữ mà họ hiểu biết sâu sắc nhất, chứ không phải là dấu hiệu của các kiến thức có

là cơ bản nhưng không có các hình thái cơ bản hệ thống nào, và không có số đếm, đại từ, từ mô tả các bộ phận cơ thể hay các từ vựng thực sự là cơ bản khác. Một vấn đề khác rất dễ nhận thấy đối với danh mục các từ này là thiếu những tương ứng âm vị, nhất là về khía cạnh thanh điệu và không có sự giải thích hợp lý nào được đưa ra.

Xem xét các dữ liệu từ vựng ở Bảng số 2, lấy từ tài liệu của tác giả Bình Nguyên Lộc[24], “Cắm” là một từ mô tả bộ phận của cơ thể, nhưng về mặt ngữ nghĩa không phải là một bộ phận cơ bản của cơ thể, như từ “mặt” hay “cánh tay”, và cũng thiếu các từ mô tả các bộ phận cơ thể khác. Từ tiếng Việt *bạn* [friend] là một từ vay mượn tiêu biểu giữa tiếng Hán và tiếng Việt (từ [companion] trong tiếng Hán). Các dạng tương tự về ngữ âm của từ “mặt ong” được thấy trong nhiều ngôn ngữ ở khu vực, kể cả tiếng Môn-Khme và tiếng Hán, và không thể dùng để chứng minh cho lập luận về nguồn gốc lịch sử chung hay thậm chí là vay mượn. “Đào” rõ ràng là một từ mượn từ tiếng Chăm vì nó là một trong số ít từ có hai âm tiết trong tiếng Việt (tức là hai âm tiết không có nghĩa riêng lẻ và đó cũng không phải là một từ láy âm). “Lá” rất có

thể là từ Môn-Khme; đây có thể là hình thái Chăm, song có vẻ dễ gây nhầm lẫn<sup>(21)</sup>. Việc loại trừ các ví dụ này làm cho các từ “cây” và “núi” có vẻ là những trường hợp tương tự tình cờ<sup>(22)</sup>. Một vấn đề khác là các hạng mục từ vựng trong tiếng Ma-lay và tiếng Chăm không nhất quán, trong khi xem xét sự gần nhau tương đối về ngôn ngữ học giữa hai ngôn ngữ này (cả hai ngôn ngữ đều là những phần gắn bó chặt chẽ của tiểu nhánh Ma-lai của ngôn ngữ vùng Nam Đảo), cần phải có thêm nhiều trường hợp nữa. Tiếp theo, như trong trường hợp từ “cây” và “đào”, khi xem xét về chiều sâu đáng kể của thời gian (hơn hai ngàn năm), những tương ứng về âm thực sự ít rõ ràng hơn, mặc dù vẫn tồn tại các hình mẫu có thể giải thích được và có hệ thống về sự biến đổi ngữ âm. Vì vậy, trong trường hợp nhiều nhất thì bằng chứng này cũng đáng nghi ngờ, và những gì chúng ta có được rất có thể là một tập hợp các từ cùng gốc già. Cuối cùng, khi xem xét cùng với nhau, những điều này không phải là một tập hợp các từ nguyên tiêu biểu của vùng Nam Đảo, có thể có sức thuyết phục để chứng minh cho lập luận về nguồn gốc ngôn ngữ.

Bảng 2. Các hình thái tiếng Việt - Nhật và Ngôn ngữ Nam Á được đề xuất (Matsumoto) <sup>(21)</sup>

Tiếng Anh	Tiếng Việt	Tiếng Malai - Polynesian	Tiếng Nhật	Nguồn
chin	cắm	dagu, anka', angko'	ago, agi	không rõ
friend	bạn, bạn	têman	tomona-fu	Hán
honey	mật	kemet, kemut	kimo	Đông Nam Á
island	củ lao	(Chăm) kalau, kulau, pulau	ikuri	Chăm
leaf	lá	(Chăm) hala	*pa > ha	Môn-Khme
mountain	non	Mênum, bênum, bênom	mine	Không rõ
tree	cây	(Ma-lai) kayu	ko, ke	không rõ

ý nghĩa. Khả năng tìm ra các từ có âm tương tự chắc chắn được củng cố khi số các âm vị có thể tìm thấy ít hơn, như trong trường hợp tiếng Nhật, một ngôn ngữ chỉ có năm nguyên âm, ngược với tiếng Việt có tới mười tám nguyên âm, tùy thuộc vào sự biến thể giữa các vùng.

<sup>(21)</sup> Về giả thiết mà nói thì từ Chăm có thể là từ vay mượn từ Môn-Khme, khi người Chăm có giao tiếp sâu rộng với người Môn-Khme ở vùng Trung bộ Việt Nam trong vài thế kỷ qua.

<sup>(22)</sup> Một khía cạnh nữa cần xem xét là âm tiết nào trong một từ hai âm tiết bị mất đi. Ví dụ trong các ngôn ngữ Môn-Khme và Chăm, trọng âm trong các từ hai âm tiết rơi vào âm tiết thứ hai, do đó làm cho âm tiết đầu chữ không phải là âm tiết cuối cùng bị mất đi. Giả sử tiếng Việt theo mô hình này, như bằng chứng ngôn ngữ học lịch sử chỉ ra, thì những từ như *cây* sẽ không giữ được âm tiết đầu tiên.

Theo một phương pháp tiếp cận khác, theo giả thiết về nhóm ngôn ngữ liên họ ở vùng Đông Nam Á, Thái Bình Dương và Tiểu lục địa Ấn Độ, thì tiếng Việt có thể được xem là gần với hệ ngôn ngữ vùng Nam Đảo[26]. Bằng chứng rõ ràng nhất về sự gần bó giữa họ ngôn ngữ vùng Nam Đảo với hệ Nam Á là hình thái ngôn ngữ[27] và bằng chứng về từ vựng thì cực kỳ hiếm[28]. Tiếng Việt chỉ cho thấy các mô hình về hình thái và cấu trúc từ chung, như ta thấy trong các ngôn ngữ thuộc tiểu nhánh Vietic rất gần gũi. Các mô hình này chỉ ra những dấu vết còn lại của tiếp đầu ngữ /pa-/ của động từ, một tiếp đầu ngữ phổ biến trong nhiều ngôn ngữ vùng Nam Đảo và Môn-Khme. Bằng chứng về từ vựng cũng rất hiếm hoi, chỉ có vài khả năng là các từ “chó”, “mắt”, “cá” và vài từ khác nữa. Ngay dù giả thiết này đứng vững được, thì quan hệ trực tiếp nhất vẫn là giữ tiếng Việt với tiếng Môn-Khme, chứ không phải là với ngôn ngữ vùng Nam Đảo.

Khi xem xét vị trí địa lý của ngôn ngữ Việt, thì quan điểm cho rằng tiếng Việt có nguồn gốc từ vùng đảo Thái Bình Dương có thể đúng. Bờ biển phía Bắc Việt Nam có thể được những người du hành từ vùng Nam Đảo tiếp cận, và theo các học thuyết hiện tại[29] thì những người này đi từ Nam Trung Quốc sang Đài Loan và tiến xuống phía Nam, tới tận Phi-lip-pin và quần đảo In-đô-nê-xia. Rõ ràng người Chăm là tổ tiên của làn sóng di cư vĩ đại này. Họ đã tới bờ biển phía nam Việt Nam khoảng hai ngàn năm trước và người Việt là láng giềng liền kề của họ ở phía Bắc Đế chế Chăm-pa cho đến khi Đế chế này rơi vào tay người Việt hồi thế kỷ 15. Có thể đã có giao tiếp về ngôn ngữ giữa người Việt và người Chăm, nhưng người Việt đã có mối quan hệ chính trị không thân thiện với những người láng giềng của mình ở phía Nam. Theo các dữ liệu về ngôn ngữ học, sự giao tiếp giữa họ rất ít; chỉ có vài ví dụ về các từ vay mượn từ tiếng Chăm hay ngôn ngữ vùng Nam Đảo được đưa ra một cách chắc

chắn<sup>(23)</sup>. Sự thiếu vắng một cơ sở chắc chắn về từ vựng cơ bản và các mô hình về sự tương ứng âm vị làm cho lập luận rằng tiếng Việt có nguồn gốc từ họ ngôn ngữ vùng Nam Đảo trở nên rất bấp bênh.

## 6. Ngôn ngữ vùng Nam Á và ngôn ngữ Môn-Khme

Hơn 150 ngôn ngữ thuộc Liên họ ngôn ngữ vùng Nam Á được trải rộng khắp vùng lục địa Đông Nam Á, từ Việt Nam tới tỉnh Vân Nam Trung Quốc, xuống tới vùng núi của bán đảo Ma-lai-xia, và thậm chí sang cả đông Ấn Độ, mà vùng này là lãnh địa của tiểu nhánh Munda của liên họ ngôn ngữ vùng Nam Á. Bằng chứng về từ vựng và âm vị cho thấy một cách chắc chắn rằng có một mối quan hệ giữa tiếng Việt với liên họ ngôn ngữ Nam Á. Một số sưu tập ban đầu về khía cạnh này của thuật ngữ tiếng Việt gồm của Maspéro (mặc dù ông ta lập luận nguồn gốc tiếng Việt là tiếng Tai), Gordon Luce, David Thomas với Robert Headley, và Franklin Huffman[3,30-33]. Sử dụng phương pháp luận thống kê từ vựng, Thomas và Headley cho thấy rằng tỷ lệ từ vựng cơ bản chung giữa tiếng Việt và tiếng Môn-Khme là 25 phần trăm, tương tự với tỷ lệ giữa các nhánh của ngôn ngữ Môn-Khme (khoảng 25 đến 35 phần trăm). Các công trình nghiên cứu cung cấp bằng chứng về sự tương xứng về âm vị gồm cả tác phẩm của Haudricourt, Ferlus, Gage, Diffloth và Nguyễn Tài Căn[2,5,7,9,12]. Trong một cuốn sách dành toàn bộ cho đề tài này, Ferlus và Nguyễn Tài Căn chỉ ra rằng con đường phát triển âm vị của tiếng Việt đã trải qua hơn hai nghìn năm từ gốc Môn-Khme của nó. Từ các nguồn khác nhau này, ba trăm từ nguyên thuộc hệ Nam Á có thể được xác định trong tiếng Việt, kể từ thời tiền - liên họ Nam Á. Cũng có thể xác định được vốn từ vựng thuộc các mức độ thấp hơn và các tiểu nhánh

<sup>(23)</sup> Đã thảo luận ba chục từ gốc vùng Nam Đảo trong tiếng Việt, song phần lớn các từ này phải bị loại bỏ vì nhiều trong số từ đó bắt nguồn từ nhiều nhóm ngôn ngữ hoặc hầu hết số từ đó đều có vấn đề về âm vị khi tái hiện.

(ví dụ Môn-Khme, Đông Môn-Khme v.v) do đó có thể đưa ra một kịch bản ngôn ngữ học lịch sử hoàn chỉnh hơn về quá trình di cư dần dần của người Môn-Khme ở khắp vùng Đông Nam Á[34].

Cơ sở từ vựng học liên kết tiếng Việt với tiếng Môn-Khme gồm vốn từ vựng cơ bản chung chung về mặt ngữ nghĩa, và điều quan trọng là đó chính là vốn từ vựng liên kết giữa các ngôn ngữ khác với các tiểu nhánh của các ngôn ngữ Môn-Khme. Danh sách gồm một trăm từ mà Huffman đưa ra năm 1977 cho thấy vài chục<sup>(24)</sup> từ có thể là từ cùng gốc Việt/Khme với nhiều tương ứng về âm với hàng chục ngôn ngữ Môn-Khme khác. Điều mà bằng chứng này cho thấy là (1) vốn từ cơ bản, (2) sự tương ứng về âm vị, và (3) sự lan toả của vốn từ vựng này trong nội tại các ngôn ngữ cũng như giữa các tiểu nhánh của ngôn ngữ Môn-Khme. Điểm thứ ba làm cho các dữ liệu này trở nên có ý nghĩa

về mặt phương pháp luận. Ví dụ về các vấn đề này được trình bày trong Bảng 3 (chữ viết từ các ấn phẩm gốc được giữ nguyên), gồm các tiếng Rục, một thổ ngữ nói Vietic ở tỉnh Quảng Bình[11]; tiếng Pa-cô ở tỉnh Thừa Thiên và Quảng Trị[35]; tiếng Mông ở Thái Lan và Miến Điện[30]; và tiếng Rieng cũng ở Miến Điện[30]. Những từ không có trong dữ liệu được đánh dấu bằng ký hiệu "x". Điều quan trọng là phải xem xét thực tế rằng các ngôn ngữ Môn-Khme khác bắt nguồn từ bốn tiểu nhánh khác nhau của Môn-Khme và bao phủ phạm vi địa lý rộng lớn, song tất cả đều có các từ nguyên Môn-Khme rất phổ biến này, một thực tế củng cố thêm việc gộp tiếng Việt vào họ Môn-Khme và làm suy yếu khả năng của một tình huống vay mượn nhiều về ngôn ngữ.

Bảng 3. Bảng chứng so sánh tiếng Việt và tiếng Môn-Khme

Tiếng Anh	Tiếng Việt	Tiếng Rục	Tiếng Pa-co	Tiếng Môn	Tiếng Rieng
1 shoot	bắn	pinh <sup>1</sup>	pe:p	pun	pa:n <sup>1</sup>
2 four	bốn	pon <sup>1</sup>	pu:p	pun	kə:pun <sup>1</sup>
3 three	ba	pa <sup>1</sup>	pe:	pi	kə:wa:n <sup>1</sup>
4 fish	cá	aka <sup>1</sup>	x	ka	ka:ʔ <sup>1</sup>
5 leg	chân	chun <sup>2</sup>	ju:ŋ	ju:ŋ	tja:ŋ <sup>1</sup>
6 grandchild	cháu	chu <sup>1</sup>	ʔa:caw	cau	x
7 bird	chim	ichim <sup>1</sup>	ʔa:cə:ʔ	gacem <sup>1</sup>	s'um <sup>1</sup>
8 dog	chó	achó <sup>1</sup>	ʔa:cə:	x	s'o:ʔ <sup>1</sup>
9 npe	chín	chin <sup>1</sup>	cə:n	cin	s'um <sup>1</sup>
10 child	con	kon <sup>1</sup>	ʔa:kə:n	kon	kuan <sup>1</sup>
11 -- -- --	đào	tao <sup>1</sup>	ta:p	tən	tao <sup>1</sup>
12 wind	gió	kaio <sup>1</sup>	x	kyó	ku:ʔ <sup>1</sup>
13 two	hai	hai <sup>1</sup>	ba:r	fa	kə:ʔa:r <sup>1</sup>
14 leaf	lá	ola <sup>1</sup>	ʔa:la:	sla	la:ʔ <sup>1</sup>
15 one	một	móc <sup>*</sup>	mə:ʔ	moa	x
16 new	mới	hoi <sup>1</sup>	təm.me:	tu:m	tə:nə:mə:ʔ <sup>1</sup>
17 you (singular)	này	ni <sup>1</sup>	ma:	beh	mi:ʔ <sup>1</sup>
18 nose	mũi	muih <sup>1</sup>	moh	muh	x
19 year	năm	nam <sup>1</sup>	x	cnam	x
20 year	ngáp	tap <sup>1</sup>	ʔa:ʔap	k'a-op	hap
21 sun/day	ngày	x	ʔi:ŋaj	ŋoi	s'ŋə:ʔi:ʔ <sup>1</sup>
22 water	nước	da:k <sup>1</sup>	da:ʔ	đak	x
23 root	củ	lech <sup>2</sup>	tja:ʔ	ruih	ria:s <sup>1</sup>
24 make	đón	pu:pjuh <sup>1</sup>	ku:sə:p	x	hai:n <sup>1</sup>
25 thunder	sấm	kru:m <sup>*</sup>	kru:m	x	tə:nə:m <sup>1</sup>
26 deep	sâu	djom <sup>1</sup>	tru:	sjuih	tjə:ru:ʔ
27 hand	tay	si <sup>1</sup>	ʔa:ti:	tai	ti:ʔ <sup>1</sup>
28 hair	tóc	usak <sup>1</sup>	sək	sok	huk <sup>1</sup>
29 betel	trầu	thu <sup>2</sup>	x	jablu	plu <sup>1</sup>
30 hundred	trăm	kləm <sup>1</sup>	ku.lam	klam	x
31 fruit	trái	puh <sup>1</sup>	ku.laj	x	plə:ʔ <sup>1</sup>
32 thatch	tranh	mleñh <sup>1</sup>	plə:ŋ	x	plə:ŋ <sup>1</sup>
33 far	ngái	chongai <sup>1</sup>	jo:ŋ	ŋoi	s'ŋə:ʔi:ʔ <sup>1</sup>

Ghi chú: - "x" chỉ các từ bị thiếu trong nguồn dữ liệu.

- I. Từ ngái được nói trong các thổ ngữ Việt ở vùng núi tỉnh Nghệ An, Bắc Trung bộ Việt Nam.

<sup>(24)</sup> Theo một phương pháp tiếp cận khác, theo giả thiết về nhóm ngôn ngữ liên họ ở vùng Đông Nam Á, Thái Bình Dương và Tiểu lục địa Ấn Độ, thì tiếng Việt có thể được xem là gắn với hệ ngôn ngữ vùng Nam Đảo[26].

Nghiên cứu danh mục các từ cùng gốc này cho thấy các mô hình tiêu biểu của sự tương xứng về âm vị giữa tiếng Việt và các tiếng Môn-Khme khác. Ferlus mô tả bằng chứng về tình trạng các cụm phụ âm Môn-Khme giảm từ hai ba âm xuống các âm quặt lưỡi trong tiếng Việt (ví dụ âm “tr”, “s” và “f” trong tiếng Việt)[7]. Một số loại âm điệu trong tiếng Việt

tương ứng với một số âm tiết cuối cùng trong tiếng Môn-Khme, như Haudricourt đã nêu trong một bài trình bày của mình tại hội thảo bảo vệ giả thiết của mình về nguồn gốc của thanh điệu trong tiếng Việt[5]. Ví dụ về các lập luận này được trình bày ở Bảng 4, gồm các thể loại tương ứng minh họa cho các trường hợp như vậy trong Bảng 3<sup>(25)</sup>.

Bảng 4. Tương ứng âm vị giữa tiếng Việt và tiếng Môn-Khme

Tiếng Việt	Tiếng Môn-Khme	Các ví dụ trong Bảng 3
Các âm quặt lưỡi (“tr” và “r”)	Các nhóm phụ âm và tiền âm tiết	#23-26, #30-33
Các âm tắc (“b” và “d”)	Các âm tắc (/p/ và /t/)	#1-4, 12
Thanh sắc và nặng	Các âm tắc cuối âm tiết (/p/, /t/, /c/, /k/ và /ʔ/)	#21, #23, #29
Thanh hỏi và ngã	Các âm xát cuối âm tiết (/h/ và /s/)	#19, #24

Chất lượng của vốn từ vựng cơ bản Môn-Khme trong tiếng Việt cũng cao. Ví dụ, ngược với từ có thể là vay mượn từ tiếng Tai gả [chicken], hay công [peacock] từ tiếng Môn-Khme, chúng ta sẽ thấy từ ít tiêu biểu nhất về mặt ngữ nghĩa, và phổ thông hơn là từ chim [bird]. Trong số các từ có gốc Môn-Khme trong tiếng Việt, chúng ta cũng sẽ thấy các từ chỉ con vật thực sự cơ bản như từ chó [dog], cá [fish], và rắn [snake], đối ngược với những chủng loại cụ thể hơn của các động vật này. Về các bộ

phận của cơ thể, chúng ta thấy các từ chân, [leg/foot], tay [arm/hand], tai [ear] và mũi [nose], là những từ có liên hệ trong toàn bộ hệ Môn-Khme và thậm chí vươn xa tới tận các ngôn ngữ Munda của miền Đông Ấn Độ. Tuy có ít ví dụ hơn, nhưng có các động từ Môn-Khme cơ bản, như từ ngồi [to sit], lắng [to listen], và chết [to die]. David Thomas đã chỉ ra rằng tất cả số đếm từ một đến mười trong tiếng Việt đều thuộc họ Môn-Khme hay một tiểu nhóm của Môn-Khme<sup>(26)</sup>. Mặc dù hệ thống đại từ trong tiếng Việt phần lớn được thay thế bằng một hệ thống các từ có điều kiện xã hội hầu hết bắt nguồn từ các từ có quan hệ gia đình, các từ chỉ sự thân mật/thông tục như mày [you] và nó [he/she/it] là các từ thuộc vùng Nam Đảo[36], chắc chắn chịu tác động của quá trình tái cấu trúc toàn bộ các từ về địa chỉ và sự tra cứu trong tiếng Việt[37].

Những tương ứng về âm vị của vốn từ vựng cơ bản nói trên là khá nhiều. Gage đã liệt kê ra hàng chục trường hợp riêng lẻ về sự tương ứng âm vị giữa các từ trong tiếng Việt và tiếng Môn-Khme<sup>(27)</sup>. Các trường hợp này rơi vào một

<sup>(25)</sup> Vẫn còn một số nghi ngờ về các dữ liệu. Tại sao trong tiếng Việt lại có những từ với thanh sắc mà không có phụ âm tắc cuối cùng (như từ bốn [four] và lá [leaf]), như Haudricourt đã nêu trong tài liệu “Nguồn gốc thanh điệu trong tiếng Việt”? Tại sao một số âm tắc đầu từ trong ngôn ngữ Môn-Khme cũng tắc trong tiếng Việt, nhưng một số âm khác lại là âm mũi (như từ nước [water] và nó [it])? Yếu tố nào đã dẫn tới việc giảm các cụm âm thành các phụ âm quặt lưỡi (ví dụ âm “r”, “s” và “tr” trong tiếng Việt), trái ngược với các khả năng khác, nhất là vì các âm quặt lưỡi hiếm thấy trong các ngôn ngữ lân cận như Môn-Khme, Tai và Chăm? Trong cuốn “Tiếng cổ hòng của ngôn ngữ Nam Á cổ”, Diffloth đã đưa ra các dữ liệu mà nó có thể giải thích cho câu hỏi đầu tiên và cũng có thể không. Câu hỏi thứ hai đòi hỏi phải có thêm nghiên cứu. Câu hỏi thứ ba chưa được giải thích một cách đầy đủ, mặc dù tình trạng vay mượn nhiều và sau đó là sự giao tiếp với tiếng Hán, mà nó đã được tái hiện với các âm quặt lưỡi (chỉ có tiếng Trung ở phía bắc lục địa mới còn những âm quặt lưỡi này) có thể là một yếu tố.

<sup>(26)</sup> Đây là một dạng cổ hơn, ít được sử dụng hơn là dạng tiêu chuẩn nghe

<sup>(27)</sup> Quá trình tái cấu trúc chủ yếu do các giao tiếp với tiếng Hán, từ đó nhiều từ tương tự được vay mượn, mặc dù

số chúng loại mà nó chính là loại hỗ trợ cần thiết của bằng chứng về các mối quan hệ chứ không chỉ là các trường hợp giống nhau ngẫu nhiên. Những tương ứng về âm vị này cung cấp không chỉ bằng chứng để liên kết tiếng Việt với tiếng Môn-Khme mà còn là một phương tiện để phục nguyên tiếng Việt ở các giai đoạn cách đây vài thế kỷ và sớm hơn nữa[2]. Một số tiếp đầu ngữ sơ khai của tiếng Môn-Khme cũng có thể được phục nguyên trong tiếng Việt[7,12].

Những dữ liệu ngôn ngữ học về những quan hệ ngôn ngữ rất gần nhau song cực kỳ thận trọng của tiếng Việt, các ngôn ngữ vùng cao Chứt[38] (như tiếng Rục trong Bảng 3) ở vùng cao nguyên Bắc Trung bộ Việt Nam và vùng biên giới Việt Lào<sup>(28)</sup> đưa ra bằng chứng mà nó làm thay đổi các giả thiết trước đó về các giai đoạn và thời điểm của một số phát triển âm vị trong tiếng Việt. Thực tế rằng tiếng Việt có các nhóm phụ âm là không thể tranh cãi được vì một số thổ ngữ đã có các nhóm phụ âm đến tận thế kỷ hai mươi, và chúng thể hiện rất rõ ràng trong tài liệu viết từ thế kỷ mười bảy (như trong từ điển năm 1651 của de Rhodes)<sup>(29)</sup>. Không chứng minh được khi nào các tiền âm tiết chấm dứt tồn tại trong tiếng Việt, mặc dù sự tồn tại của nó trong các ngôn ngữ Chứt rất gần gũi với tiếng Việt cho thấy rằng việc này đã xảy ra cách đây vài thế kỷ hoặc sớm hơn nữa. Sự hình thành cấu trúc hai âm tiết trong tiếng Việt chỉ có khả năng diễn ra trong kho từ vựng Hán - Việt vào triều đại nhà Minh (1368 - 1644), trong đó hai chữ viết, và như vậy là hai âm tiết, được sử dụng để biểu đạt một số từ nhất định trong tiếng Việt thời kỳ đó, và điều này có nhiều khả năng là cách thức đại diện cho các nhóm phụ âm[3]. Khi xem xét các giai đoạn xa

những từ này phát triển một số chức năng không thấy trong tiếng Hán.

<sup>(28)</sup> Từ "Chứt" là một thổ ngữ được sử dụng trong *Giáo trình về lịch sử ngữ âm tiếng Việt* của Nguyễn Tài Cẩn, và vì vậy toàn bộ nhóm ngôn ngữ Vietic bao gồm các tiểu nhánh ngôn ngữ Việt-Mường và Chứt. Tôi dùng từ "Vietic thứ yếu" để nói tới nhóm ngôn ngữ đó.

<sup>(29)</sup> Ví dụ xem dữ liệu về tiếng Rục để thấy các ý kiến thảo luận tại sao một dữ liệu như vậy lại không đưa ra các mối quan hệ.

xưa của âm vị tiếng Việt, sự giống nhau hiện tại giữa tiếng Việt với tiếng Tai hay tiếng Hán được xem xét theo một cách mới. Giai đoạn hiện đại của cấu trúc âm vị tiếng Việt chỉ là điểm kết thúc của một tiến trình đơn âm hoá kéo dài mà nguyên nhân của nó có thể là sự giảm bớt nội tại tiếp theo các xu hướng chữ viết cũng như sự hội nhập dần dần của hàng trăm từ đơn âm tiếng Hán trong hơn hai ngàn năm.

Trong chúng loại này có hai thông tin bổ sung có thể làm nổi bật nguồn gốc Môn-Khme của tiếng Việt. Thứ nhất, tiếng Việt có bằng chứng cả về các nhóm đầu từ và các tiếp đầu ngữ hay trung tố, mà một số tương ứng với các hình thể ngôn ngữ Môn-Khme[39]. Những bằng chứng như vậy gần như bị thất thoát trong quá trình đơn âm hoá. Thứ hai, các mô hình của các từ láy, trong đó một âm tiết đơn nhất được lặp lại một phần nhưng thường với một âm thay đổi, là một khía cạnh gợi ra sự giống nhau về chữ viết với các ngôn ngữ Môn-Khme. Đúng ra, trường hợp này xảy ra với các ngôn ngữ Môn-Khme được sử dụng để nói bên trong hay sát với lãnh thổ tiếng Việt nhiều hơn là các vùng xa về địa lý, và điều đó gợi ra một loại ảnh hưởng vùng và không nhất thiết là nguồn gốc ngôn ngữ[7]. Có lẽ hiện tượng láy nhiều từ có âm vị tách biệt, một đặc điểm có trong tiếng Việt và tiếng Môn-Khme Pa-cô, một đặc tính ngôn ngữ học khác thường và nổi bật, là có sức thuyết phục hơn<sup>(30)</sup>.

## 7. Tóm tắt và so sánh các giả thiết

Tới thời điểm này, chúng ta có thể so sánh các giả thiết khác nhau về sức thuyết phục của các lập luận ngôn ngữ học và mức độ khả thi của các kịch bản nhân chủng học - lịch sử. Bảng 5 đưa ra sự đánh giá của tác giả về chất lượng của các bằng chứng ngôn ngữ học đối với các nhóm ngôn ngữ nguồn khác nhau. Chỉ có ngôn ngữ Môn-Khme được chứng minh là có chung kho từ vựng nòng cốt cơ bản với tiếng

<sup>(30)</sup> Để xem các ý kiến thảo luận về các mô hình này.

Việt, với lượng tương ứng âm vị bất qui tắc đáng kể, trong khi với các nhóm ngôn ngữ khác thì không có hiện tượng như vậy. Để bác bỏ vốn từ vựng cơ bản Môn-Khme trong tiếng Việt, cần phải loại bỏ phương pháp luận so sánh cơ bản, mà phương pháp này, dù không hoàn hảo, nhưng rất quan trọng cho việc sưu tầm thêm các bằng chứng về ngữ nghĩa. Về những

vấn đề liên quan đến âm vị và chữ viết nói chung, giữa tất cả các nhóm ngôn ngữ đều có sự trùng lặp, và có lẽ đây là nguồn gốc lớn nhất của tình trạng nhầm lẫn trong quá trình nghiên cứu này. Có lẽ tốt hơn là coi các đặc điểm ngôn ngữ này ở vùng ngôn ngữ Đông Nam Á là kết quả của quá trình giao lưu ngôn ngữ lâu dài chứ không phải là có chung nguồn gốc ngôn ngữ.

Bảng 5. Chất lượng các bằng chứng về mối quan hệ ngôn ngữ học giữa tiếng Việt với các họ ngôn ngữ khác

	Môn-Khme	Hán	Tai-Kadai	Vùng đảo Thái Bình Dương
Từ vựng	Vững chắc: lượng từ vựng dùng chung đáng kể.	Yếu: Số từ vựng nòng cốt ít, hiện tượng vay mượn rất rõ	Yếu: ít từ vựng cơ bản dùng chung.	Rất yếu: không có từ vựng nòng cốt cơ bản chung
Âm vị	Vững chắc: nhiều tương đồng, nhưng vẫn còn một số điểm chưa chắc chắn	Một số: Có sự trùng lặp đáng kể về các hệ thống âm vị	Yếu: Có hệ thống tương tự nhưng không có các mô hình âm tương ứng giữa các vốn từ vựng cơ bản	Yếu: Không xác định được sự tương đồng âm vị nào
Hệ thống	Một số: có bằng chứng về các mô hình hình thái, phụ tố và âm láy, bỏ ngữ đứng sau	Yếu: Có nhiều khác biệt về hệ thống	Một số: Hệ thống chung ở Đông Nam Á	Yếu: Hệ thống chung ở Đông Nam Á

Về các kịch bản giao lưu giữa các nhóm dân tộc, để hai giả thiết về nguồn gốc Tai-Kadai hay nhóm Nam Đảo có thể đứng vững, có thể có một thời gian trong đó tổ tiên người Việt hiện đại đã ở trong một vị thế phụ thuộc về mặt xã hội. Điều đó có nghĩa có thể có trường hợp một nhóm người nói tiếng Hán, Tai hay ngôn ngữ vùng Nam Đảo chịu ảnh hưởng thông qua tiếp xúc xã hội với ngôn ngữ Môn-Khme hay một nhóm ngôn ngữ của hệ này nhiều đến mức những người sử dụng ngôn ngữ này vay mượn hầu hết vốn từ vựng cơ bản của họ trước khi chịu ảnh hưởng của tiếng Hán. Tất nhiên, không thể hoàn toàn bác bỏ lập luận rằng ba nghìn năm trước đây, tiếng Việt là tiếng Tai, tiếng Nam Đảo hay tiếng Hán mà sau nay phát triển các đặc điểm ngôn ngữ của tiếng Môn-Khme do những giao tiếp mạnh mẽ về ngôn ngữ, song trường hợp này có thể xảy ra với bất kỳ ngôn ngữ nào trong khu vực. Không có thêm bằng chứng nào nữa, thì các giả định như vậy rất mang tính phỏng đoán. Trong khi các giả

thiết khác không thể bị bác bỏ hoàn toàn, thì bằng chứng tích cực mạnh mẽ nhất vẫn là tiếng Việt có nguồn gốc từ vùng Nam Đảo. Trong thực tế, nghiên cứu ở Đông Nam Á cho thấy có sự tương đồng đáng kể giữa người bản địa Việt và Ma-lai-xia (người Orang Asli), mà ngôn ngữ của họ là Austro-Asiatic - một thực tế rất tương đồng với giả thiết nêu trên<sup>(31)</sup>.

Qua việc xem xét các thể loại từ vựng và nguồn gốc nhóm ngôn ngữ của chúng, một bức tranh lịch sử đơn giản và rõ ràng xuất hiện. Một nhóm người nói một loại tiếng Môn-Khme, những người rất có thể đã sống du canh du cư, như nhiều nhóm Môn-Khme khác, và họ đã phát triển một nền văn minh tinh vi hơn, tiếp xúc với những người nói tiếng Tai ở vùng là phía Bắc của Việt Nam hiện nay. Nhóm Vietic lấy khái niệm từ ngôn ngữ Tai và kết hợp các từ với các hoạt động lúa nước và chăn nuôi, mà các hoạt động này có thể đã cho phép nền văn minh Việt buổi sơ khai phát triển thêm. Tuy

<sup>(31)</sup> Tài liệu trình bày tại Hội nghị hàng năm lần thứ 9.

nhiên, không có vẻ có bằng chứng về sự giao tiếp sâu rộng với nhóm ngôn ngữ Tai, nhất là việc làm thay đổi cấu trúc ngôn ngữ của tiếng Việt, mà chỉ là một quá trình chung của sự đồng nhất ngôn ngữ trong đó các ngôn ngữ thuộc nhiều nguồn gốc trong vùng có thể trở nên giống nhau về loại hình. Những nét tương tự như vậy về ngôn ngữ học giữa các ngôn ngữ có vẻ là kết quả của hàng nghìn năm giao tiếp giữa các nhóm ngôn ngữ khác nhau. Tiếp theo là sự tiếp xúc giữ tiếng Việt và tiếng Hán. Nhất là trong vài thế kỷ qua, thuật ngữ tiếng Việt đã bão hoà từ vựng Hán- Việt và dẫn đến những mô hình hình thành âm vị, hình thái học và từ ngữ tương tự như tiếng Trung Quốc hiện đại<sup>(32)</sup>. Về cơ bản, tiếng Việt là một ngôn ngữ Môn-Khmer với vài từ vay mượn từ tiếng Tai và một tầng ảnh hưởng sâu sắc từ vựng và ít nhiều cả cấu trúc tiếng Hán - một ý kiến đơn giản nhưng hợp lý trên cơ sở những dữ liệu và các công cụ phân tích hiện có.

### Tài liệu tham khảo

- [1] Dang Nghiem Van, Chu Thai Son, Luu Hung, *The Ethnic Minorities in Vietnam*, Foreign Languages Publishing House, Hà Nội, 1984.
- [2] William W. Gage, "Vietnamese in Mon-Khmer Perspective," *Southeast Asian Linguistics Presented to André-G. Haudricourt*, eds. S. Ratankul, D. Thomas, S. Premisirat, Mahidol University, Bangkok, 1985.
- [3] Henri Maspéro, "Études sur la Phonétique Historique de la Langue Annamite: Les Initiales," *Bulletin de l'École Française d'Extrême-Orient* 12 (1912) 1-127.
- [4] André G. Haudricourt, "La Place du Vietnamien dans les Langues Austroasiatiques," *Bulletin de la Société de Paris* 49, 1 (1953) 122-128.
- [5] André G. Haudricourt, "De l'Origine des Tons en Vietnamien," *Journal Asiatique* 242 (1954) 69-82.
- [6] André G. Haudricourt, "Comment Reconstruire le Chinois Archaïque," *Word* 10, 2-3 (1955) 351-364.
- [7] Michel Ferlus, "Vietnamien et Proto-Viet-Muong," *Asia Sud Est un Monde Insulindien* 6, 4 (1975) 21-55.
- [8] Michel Ferlus, Sự biến hóa của các âm tắc giữa (obstruents mediales) trong tiếng Việt, *Ngôn Ngữ Học* 2 (1981) 1-21.
- [9] Gérard Diffloth, "Proto-Austroasiatic Creaky Voice," *Mon-Khmer Studies* 15 (1989) 139-154.
- [10] Gérard Diffloth, "Vietnamese as a Mon-Khmer Language," *Papers from the First Annual Meeting of the Southeast Asian Linguistics Society*, eds. Martha Ratliff, Eric Schiller (Tempe: Arizona State University, 1990), 125-139.
- [11] Nguyễn Văn Lợi, *Tiếng Rục*, NXB Khoa Học Xã Hội, Hà Nội, 1993.
- [12] Nguyễn Tài Cẩn, *Giáo trình lịch sử ngữ âm tiếng Việt*, NXB Giáo Dục, Hà Nội, 1995.
- [13] Hans Heinrich Hock, *Principles of Historical Linguistics*, Mouton de Gruyter, New York, 1991.
- [14] Terry Crowley, *An Introduction to Historical Linguistics*, Oxford University Press, New York, 1992.
- [15] Sarah Grey Thomason and Terrence Kaufman, *Language Contact, Creolization, and Genetic Linguistics*, University of California Press, Berkeley and Los Angeles, 1988.
- [16] Mark J. Alves, "What's So Chinese about Vietnamese?" *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*, ed. Graham W. Thurgood (Tempe: Arizona State University, 2001).
- [17] Wang Li, "Hanyueyu yanjiu [Research on Sino-Vietnamese]," *Lingnan Xuebao* 9, 1 (1948) 1-96.
- [18] Đào Duy Anh, *Chữ Nôm: Nguồn gốc, cấu tạo, diễn biến* [Chu Nom: Origins, Formation, and Transformations], NXB Khoa Học Xã Hội, Hà Nội, 1979.
- [19] Mei Tsu-Lin, "Tones and Prosody in Middle Chinese and the Origin of the Rising Tone," *Harvard Journal of Asiatic Studies* 30 (1970) 86-110, specifically 95-96.
- [20] Phạm Đức Dương, Cội Nguồn Mô hình Văn Hoá-Xã hội Lúa nước của người Việt qua cứ liệu ngôn ngữ, *Nghiên Cứu Lịch Sử* 5 (1982) 43-52.

<sup>(32)</sup> Tuy nhiên, tôi lập luận rằng (a) nhiều cấu trúc ngôn ngữ của tiếng Việt hiện đại có thể do các quá trình ngôn ngữ tự nhiên chứ không phải do giao tiếp với tiếng Hán, (b) tiếng Hán đóng một vai trò tương đối thứ yếu trong các khía cạnh phi từ vựng của tiếng Việt, và (c) nhiều thay đổi mà nó làm cho tiếng Việt giống với tiếng Hán về mặt cấu trúc diễn ra trong vài thế kỷ qua chứ không phải trong thời kỳ chuyển từ triều đại Hán sang triều đại nhà Đường.

- [21] Luigi L. Cavalli-Sforza, *The History and Geography of Human Genes*, NJ: Princeton University Press, Princeton, 1994, 225, 234.
- [22] Keith W. Taylor, *Birth of Vietnam*, University of California Press, Berkeley and Los Angeles, 1983, 1.
- [23] Nobuhiro Matsumoto, *Le Japonais et les langues Austroasiatiques: Etude de vocabulaire compare*, P. Geuthner, Paris, 1928);
- [24] Bình Nguyễn Lộc, *Nguồn gốc Mã Lai của dân tộc Việt Nam*, (Bách Bộc, Sài Gòn, 1971; repr., CA: Xuân Thu, Los Alamitos, 1987).
- [25] Nguyễn Ngọc Bích, "Tiếng Việt, Tiếng Nhật, và họ Mã Lai ĐaĐảo," *Tuyển Tập Ngôn Ngữ và Văn Học Việt-Nam* 2, Dòng Viện, San Jose, 1994, 437-480.
- [26] Pater Wilhelm Schmidt, *Die Mon-Khmer-Völker*, Druck und Verlag von Friedrich Vieweg und Sohn, Braunschweig, 1906.
- [27] Lawrence A. Reid, Morphological Evidence for Austric, *Oceanic Linguistics* 33, 2 (1994) 323-344.
- [28] Gérard Diffloth, "The Lexical Evidence for Austric, So Far," *Oceanic Linguistics* 33, 2 (1994) 309-321.
- [29] Robert Blust, "Beyond the Austronesian Homeland: The Austric Hypothesis and Its Implications for Archaeology," *Prehistoric Settlement of the Pacific*, ed. W. Goodenough, Philadelphia, Transactions of the American Philosophical Society 86, 5 (1996): 117-140).
- [30] Gordon H. Luce, "Danaw, a Dying Austroasiatic Language," *Lingua* 14 (1965) 98-129,
- [31] David D. Thomas, Robert K. Headley, Jr., "More on Mon-Khmer Subgroupings," *Lingua* 25 (1970) 398-418.
- [32] Franklin E. Huffman, "The Relevance of Lexicostatistics to Mon-Khmer Languages," *Lingua* 43 (1976) 171-198;
- [33] Franklin Huffman, "An Examination of Lexical Correspondences between Vietnamese and Some Other Austroasiatic Languages," *Lingua* 43 (1977) 171-198.
- [34] Mark Alves, "The Vieto-Katuic Hypothesis: Lexical Evidence," *SEALS XV: Papers from the 15th Meeting of the Southeast Asian Linguistics Society*, ed. Paul Sidwell, The Australian National University, Research School of Pacific and Asian Studies, Pacific Linguistics Publishers, Canberra, 2005, 169-176.
- [35] Nguyễn Văn Lợi, Đoàn Văn Phúc, and Phan Xuân Thành, *Sách học tiếng Pakoh-Taoh*, UBND Tỉnh Bình Trị Thiên, Canberra, 1984.
- [36] David Thomas, "South Bahnaric and Other Mon-Khmer Numeral Systems," *Linguistics* 174 (1976) 65-80.
- [37] H.J. Pinnow, "Personal Pronouns in the Austroasiatic Languages: A Historical Study," *Lingua* 14 (1965) 3-42.
- [38] Michel Ferlus, "L'infixe instrumental -m- en Khamou et sa trace en Vietnamien," *Cahiers de Linguistique, Asie Orientale* 2 (1977) 51-55.
- [39] Jeremy H.C.S. Davidson, "A New Version of the Chinese-Vietnamese Vocabulary of the Ming Dynasty-1," *Bulletin of the School of Oriental and African Studies* 38, 2 (1975): 296-315 and no. 3 (1975): 586-608.

## Linguistic research on the origins of the Vietnamese language: an overview

Mark Alves

*Department of Reading, ESL, Foreign Language and Philosophy, Montgomery College,  
51 Mannakee Street Rockville, MD 20850, USA*

While the majority of linguistic researchers both inside and outside Vietnam consider the Vietnamese language a Mon-Khmer, Austroasiatic language, some counterarguments and general confusion continue to exist in the public domain. This article looks at four hypotheses regarding the linguistic origins of Vietnamese, hypotheses that place Vietnamese variously within the Austroasiatic, Austronesian, Chinese, or Tai-Kadai language families. Based on linguistic methodology and plausible scenarios of interethnic contact, the commonly held position - Vietnamese is an Austroasiatic language - remains the most tenable.