

NGÔN NGỮ HỌC ĐỊA LÝ VÀ VIỆC XÂY DỰNG BẢN ĐỒ PHƯƠNG NGỮ*

NGUYỄN TÀI THÁI**

Abstract: This article presents the theoretical foundations and historical development of linguistic geography, emphasizing its role in studying the spatial distribution of linguistic variants and in the development of dialect atlases. Worldwide, many countries have developed linguistic atlases at both national and regional levels, whereas in Vietnam, the development of language/dialect atlases has not received sufficient attention. The article highlights the role of Geographic Information System (GIS) and tools such as ArcGIS, QGIS, and WebGIS in managing, analyzing, and visualizing linguistic data. It concludes that the application of GIS technology is an inevitable direction for modernizing dialect research as well as for developing dialect atlases.

Keywords: *linguistic geography; dialect; linguistic atlas; GIS; ArcGIS; QGIS; WebGIS.*

1. Dẫn nhập

Trước thế kỉ XIX, việc nghiên cứu ngôn ngữ tại các quốc gia châu Âu chủ yếu hướng đến việc xác lập và chuẩn hóa các dạng thức chuẩn của ngôn ngữ quốc gia như tiếng Anh, tiếng Pháp và tiếng Đức. Tuy nhiên, bên cạnh biến thể chuẩn tồn tại trong phạm vi toàn quốc gia, thực tiễn ngôn ngữ cho thấy sự hiện diện của nhiều biến thể khu vực với những khác biệt đáng kể về ngữ âm, từ vựng và cấu trúc. Nhận thức được giá trị tư liệu và ý nghĩa khoa học của các biến thể này, giới nghiên cứu dần mở rộng phạm vi khảo sát sang các vùng phương ngữ khác nhau, coi đây là lĩnh vực cần được xem xét một cách hệ thống.

Trong bối cảnh đó, Johann Andreas Schmeller được xem là một trong những học giả tiên phong khi công bố công trình ngữ pháp về phương ngữ Bavarian vào năm 1821. Công trình này không chỉ đánh dấu bước chuyển từ nghiên cứu ngôn ngữ chuẩn sang nghiên cứu các biến thể ở các khu vực mà còn đặt nền tảng cũng như phương pháp luận cho sự hình thành và phát triển của ngôn ngữ học địa lí trong giai đoạn tiếp theo.

Ngôn ngữ học địa lí có nhiều nội dung và nhiệm vụ nghiên cứu khác nhau. Đó có thể là việc miêu tả các đặc điểm ngôn ngữ (cả bình diện ngữ âm và từ vựng) ở các khu vực địa lí khác nhau để tìm hiểu những đặc trưng ngôn ngữ; xem xét những mối liên quan giữa các hiện tượng ngôn ngữ để làm cơ sở cho việc tìm hiểu những dấu vết lịch sử còn lưu lại trong ngôn ngữ. Bên cạnh đó, nghiên cứu ngôn ngữ địa lí còn hướng đến việc xây dựng các bản đồ ngôn ngữ cũng như các bản đồ phân bố các biến thể ngôn ngữ ở các địa phương. Xét ở khía cạnh này, ngôn ngữ học địa lí tập trung vào nghiên cứu biến thể ngôn ngữ ở bề mặt không gian và có liên quan đến việc trình bày kết quả thông qua các biểu diễn bản đồ. Khi đó, bản đồ được xem là công cụ trực quan nhất trong việc biểu diễn sự phân bố của các biến thể ngôn ngữ.

Dựa trên những cơ sở lí thuyết của ngôn ngữ học địa lí, bài viết sẽ tìm hiểu việc nghiên cứu ngôn ngữ học địa lí ở Việt Nam và trên thế giới cũng như những vấn đề mà ngôn ngữ học địa lí quan tâm. Bên cạnh đó, bài viết cũng giới thiệu một số công cụ có thể ứng dụng vào việc xây dựng bản đồ ngôn ngữ/ phương ngữ. Đây chính là cơ sở giúp các nhà nghiên cứu tiến hành xây dựng bản đồ phương ngữ tiếng Việt trong tương lai.

2. Ngôn ngữ học địa lí và lịch sử hình thành

Ngôn ngữ học địa lí là thuật ngữ được sử dụng cả trong các từ điển về ngôn ngữ học lẫn trong các công trình nghiên cứu. Theo cách hiểu của *Từ điển Ngôn ngữ học và ngữ âm học* (A dictionary of linguistics and phonetics) thì *Ngôn ngữ học địa lí nghiên cứu sự phân bố các ngôn ngữ và phương*

* Bài viết là sản phẩm của Nhiệm vụ nghiên cứu khoa học cấp bộ “Xây dựng bản đồ phân bố các từ chi quan hệ thân tộc trong phương ngữ Bắc”; Mã số: KHXXH/NV2025-06; Đơn vị chủ trì: Viện Hàn lâm Khoa học xã hội Việt Nam.

** Viện Ngôn ngữ học; Email: thaintth@yahoo.com

ngữ theo khu vực của chúng. Ngôn ngữ địa lí hay phương ngữ địa lí cũng có thể thay thế cho cách hiểu ngôn ngữ khu vực hay phương ngữ khu vực. (Crystal David, 2008: 210). Theo đó, việc tìm hiểu sự phân bố ngôn ngữ và phương ngữ ở các khu vực địa lí khác nhau, qua đó có thể lí giải việc phân bố các biến thể ngôn ngữ ở các khu vực nghiên cứu chính là nội dung quan trọng của ngôn ngữ học địa lí. Cũng theo cách hiểu đó, *Từ điển giải thích thuật ngữ Ngôn ngữ học* cho rằng ngôn ngữ học địa lí là: *Bộ phận Ngôn ngữ học nghiên cứu sự phổ biến về lãnh thổ của các hiện tượng ngôn ngữ. Ngôn ngữ học địa lí được tách ra từ phương ngữ học vào cuối thế kỉ XIX. Sự tích góp các cứ liệu về sự tồn tại của những khác biệt về phương ngữ trong các ngôn ngữ khác nhau là cơ sở để xác định sự trùng nhau hay không trùng nhau về ranh giới sự phổ biến của những khác biệt này trong một vùng lãnh thổ phân bố ngôn ngữ nhất định. Ngôn ngữ học địa lí có liên hệ chặt chẽ với Ngôn ngữ học khu vực. Sự xuất hiện và phát triển của Ngôn ngữ học địa lí có liên quan đến việc vẽ bản đồ về những khác biệt phương ngữ của các ngôn ngữ và tạo lập các atlas phương ngữ* (Nguyễn Như Ý chủ biên, 2001: 35). Cách hiểu này cho thấy ngôn ngữ học địa lí gắn liền với phương ngữ học cũng như sự phân bố các đặc điểm ngôn ngữ khác nhau ở các địa phương. Tuy nhiên, đây là hai hướng tiếp cận có mối quan hệ mật thiết nhưng không đồng nhất. Việc ngôn ngữ học địa lí tách ra từ phương ngữ học chính là do sự thay đổi về nội dung, phương pháp cũng như đối tượng nghiên cứu. Nội dung nghiên cứu của phương ngữ học hướng đến việc mô tả và phân tích các đặc điểm ngôn ngữ ở các địa phương như những hệ thống tương đối hoàn chỉnh; “*định vị sự hiện hữu của các biến thể ngôn ngữ trên bản đồ, vạch ra ranh giới các vùng phương ngữ, các vùng đồng ngữ, các đường đồng ngữ*” (Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, 2025: 132). Phương pháp nghiên cứu của phương ngữ học chủ yếu là điều tra thực địa, miêu tả và phân tích đặc điểm ngữ âm, từ vựng, ngữ pháp với đối tượng là các biến thể ở các địa phương. Trong khi đó, ngôn ngữ học địa lí tập trung khảo sát sự phân bố của các hiện tượng ngôn ngữ trong không gian, qua đó làm rõ cách thức các đặc điểm về ngữ âm, từ vựng và ngữ pháp thay đổi theo không gian địa lí; đồng thời cũng chú trọng đến nguồn gốc và quá trình phát triển của chúng dưới tác động của các yếu tố lịch sử theo thời gian. Phương pháp của ngôn ngữ học địa lí chủ yếu là lập bản đồ và phân tích quy luật lan tỏa với đối tượng nghiên cứu là các đặc điểm ngôn ngữ. Như vậy, hai lĩnh vực này có đối tượng và phương pháp nghiên cứu riêng: nếu ngôn ngữ học địa lí nhấn mạnh chiều kích không gian và sự phân bố, thì phương ngữ học lại tập trung vào bản chất và cấu trúc của các biến thể ngôn ngữ.

Ở khía cạnh nghiên cứu, nhiều tác giả cả trong và ngoài nước đều quan tâm đến cách hiểu ngôn ngữ học địa lí. Các nhà ngôn ngữ học châu Âu cho rằng, ngôn ngữ học địa lí là một nhánh của ngôn ngữ học liên quan đến ngôn ngữ của các khu vực và được hiểu bằng các thuật ngữ khác nhau như địa lí phương ngữ (dialect geography), địa lí ngôn ngữ (language geography), ngôn ngữ học địa lí (linguistics geography) (*Dẫn theo* John M. Kirk, Steward Sanderson, 1985: 124). Về cơ bản có thể thấy các nhà nghiên cứu đều chia sẻ quan niệm cho rằng ngôn ngữ học địa lí chính là việc nghiên cứu về sự đa dạng của ngôn ngữ ở các khu vực khác nhau; nó giúp chỉ ra các vị trí địa lí và sự phân bố của ngôn ngữ ở các khu vực đó; với sự đa dạng của ngôn ngữ và văn hóa. Theo đó, sự phân bố ngôn ngữ ở các khu vực địa lí khác nhau chính là đối tượng nghiên cứu của ngôn ngữ học địa lí.

Nghiên cứu ngôn ngữ học địa lí có lịch sử từ khá sớm. Georg Wenker được xem là người đã đặt nền móng cho hướng nghiên cứu ngôn ngữ học địa lí, đặc biệt là việc vẽ bản đồ phương ngữ. Tiếp theo nghiên cứu của G. Wenker, các nhà ngôn ngữ học của Nga (Viện Hàn lâm Khoa học Nga, 1903), Anh (Kolb E., 1966; Kolfb, E, B. Glauser, W. Elmer, R. Stamm, 1979; Orton, H. and S. Sanderson and Widdowson, 1978) và Hoa Kỳ (Allen, H. B., 1973; William L., S. Ash S., Boberg C., 2006)... cũng tiến hành việc khảo sát ngôn ngữ ở các địa phương và thể hiện chúng trên bản đồ. Gần đây một số quốc gia Châu Á như Nhật Bản, Trung Quốc và Hàn Quốc cũng đã xây dựng các tập bản đồ ngôn ngữ của quốc gia cũng như các địa phương.

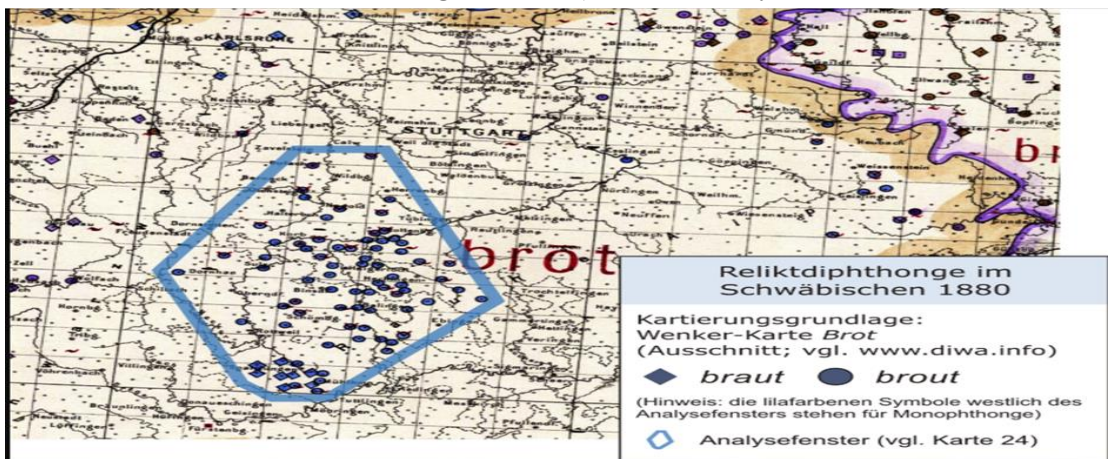
Ở Việt Nam, hướng nghiên cứu ngôn ngữ học địa lí gắn với việc tìm hiểu sự phân bố không gian của các hiện tượng ngôn ngữ chưa được chú ý nhiều. Hoàng Thị Châu trong công trình *Tiếng Việt trên các miền đất nước* mà sau này đổi tên thành công trình *Phương ngữ học* thì bên cạnh việc miêu

tả một cách khái quát nhất tiếng địa phương ở cả ba miền Bắc - Trung - Nam cũng chú ý đến việc thể hiện sự phân bố các đặc điểm ngôn ngữ ở các địa phương qua việc định vị sự xuất hiện của một số biến thể ngôn ngữ ở các khu vực địa lí khác nhau. Cũng theo hướng đó, một số nghiên cứu của các tác giả như Võ Xuân Trang (1997), Nguyễn Văn Nguyên (2003), Phạm Hữu Viện (2004), Nguyễn Tài Thái (2015, 2018, 2023), Lê Quốc Hưng (2018), Nguyễn Thị Thu Hà (2019), Trịnh Cẩm Lan (2023a, 2023b)... bên cạnh việc miêu tả các đặc điểm ngữ âm, từ vựng ở các phương ngữ, thổ ngữ đã chú ý nhiều hơn đến sự phân bố về mặt địa lí của các biến thể ngôn ngữ. Các nghiên cứu theo hướng này đã áp dụng các phương pháp của ngôn ngữ học địa lí trong việc nghiên cứu phương ngữ, đặc biệt là đã ứng dụng phần mềm ArcGIS để xây dựng các bản đồ. Đây là sự kết hợp mang tính mới và có giá trị trong nghiên cứu ngôn ngữ học nói chung và phương ngữ học nói riêng.

3. Việc xây dựng bản đồ ngôn ngữ trên thế giới và ở Việt Nam

Việc xây dựng bản đồ ngôn ngữ/ phương ngữ đã được nhiều quốc gia trên thế giới tiến hành. Dựa trên những dữ liệu, tư liệu khảo sát theo cách làm của phương ngữ học địa lí kết hợp với các kĩ thuật, phương pháp xây dựng bản đồ ngôn ngữ cũng như sự trợ giúp của các phần mềm máy tính thì các Atlas ngôn ngữ (tập bản đồ ngôn ngữ) đã được xây dựng. Từ đó, có thể có các bản đồ ngôn ngữ quốc gia, bản đồ phân bố các ngôn ngữ dân tộc thiểu số, các bản đồ về phương ngữ, thậm chí bản đồ các thổ ngữ.

Các quốc gia Châu Âu (gồm cả Nga) đã hướng đến việc xây dựng bản đồ ngôn ngữ từ khá sớm. Trong suốt những năm từ 1877 đến 1887, Georg Wenker đã khảo sát một nguồn tư liệu phong phú các biến thể/ cách nói địa phương của các câu tiếng Đức chuẩn với số lượng cộng tác viên là các giáo viên tham gia trả lời lên đến 45.000 người. Mỗi bảng câu hỏi bao gồm hơn 40 câu hỏi khác nhau từ mức độ đơn giản đến hơi khó với nhiều dạng thức khác nhau để từ đó có thể thấy được các biến thể địa phương. Với số lượng tư liệu đồ sộ như vậy nhưng đáng tiếc ngay thời điểm đó chỉ có 2 bản đồ được vẽ bằng phương thức thủ công được xuất bản, đó là *Tập Bản đồ ngôn ngữ miền Bắc và miền Trung Đức (Linguistic Atlas of Northern and Central Germany)*. Sau đó, G. Wenker tiếp tục công việc thu thập dữ liệu nhưng mãi đến năm 1926 thì tập bản đồ tiếng Đức (Deutscher Sprachatlas) mới được xuất bản. Mặc dù vẫn còn nhiều vấn đề cần xem xét cũng như bổ sung nhưng có thể nói, hướng đi tiên phong của G. Wenker đã đặt nền móng cho hướng nghiên cứu ngôn ngữ học địa lí, đặc biệt là việc vẽ bản đồ phương ngữ (dẫn theo Chambers, J.K. và Peter Trudgill, 1980:60). Hình dưới đây thể hiện một bản đồ ở Đức.



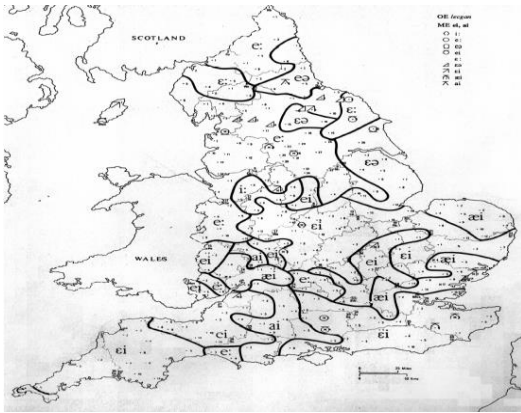
Hình 1. Các nguyên âm đôi còn sót lại ở vùng Swabia (nước Đức) năm 1880

(Nguồn: https://www.uni-marburg.de/de/fb09/dsa/personen/professor-innen-1/juergen-erich-schmidt/publikationen-1/sprachdynamik/medien_karten/karte23.jpg).

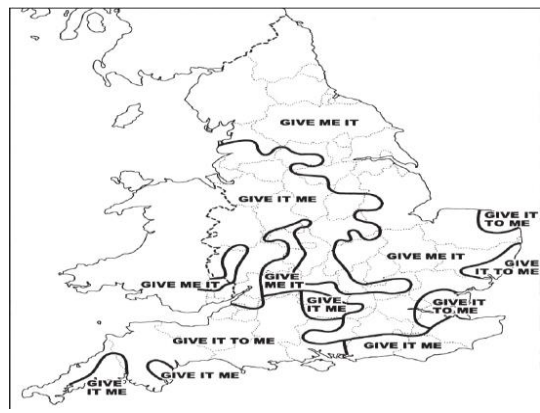
Ngay từ những năm đầu của thế kỉ XX, các nhà ngôn ngữ học thuộc Học viện Khoa học Hoàng gia Nga đã tiến hành xây dựng *Tập bản đồ ngôn ngữ tiếng Nga (Linguistics maps of Russia)* và xuất bản năm 1914. Tập bản đồ này không chỉ cho thấy các biến thể ngôn ngữ của tiếng Nga mà còn có cả

các ngôn ngữ dân tộc thiểu số. Tiếp đó vào những năm 1940 dự án xây dựng *Tập bản đồ phương ngữ tiếng Nga (Dialectological Atlas of the Russian Language)* cũng được tiến hành nhằm ghi lại các biến thể phương ngữ trong tiếng Nga; được xuất bản theo từng giai đoạn với tổng cộng hơn 400 bản đồ các hiện tượng phương ngữ.

Ở Anh, việc xây dựng bản đồ ngôn ngữ có truyền thống lâu đời và mang tính nền tảng trong phương ngữ học địa lí. Công trình quan trọng nhất là *Survey of English Dialects (SED)* do Harold Orton khởi xướng (1950-1961), tiến hành khảo sát quy mô lớn trên hàng trăm điểm ở khắp nước Anh nhằm ghi nhận tiếng Anh nông thôn truyền thống. Trên cơ sở dữ liệu này, các nhà nghiên cứu đã xây dựng *The Linguistic Atlas of England*, trong đó mỗi bản đồ phản ánh một hiện tượng ngôn ngữ cụ thể như bản đồ âm vị, bản đồ từ vựng, hình thái học hoặc một số đặc điểm ngữ pháp. Đây được xem là tập bản đồ bao quát tất cả các đặc điểm ngôn ngữ của tiếng Anh. Tập bản đồ có giá trị lớn nhờ dữ liệu thực địa công phu và việc sử dụng đường đồng ngữ (isogloss) để thể hiện sự phân bố phương ngữ một cách rõ ràng. Hình dưới đây thể hiện bản đồ biến thể ngữ âm và ngữ pháp của tiếng Anh.



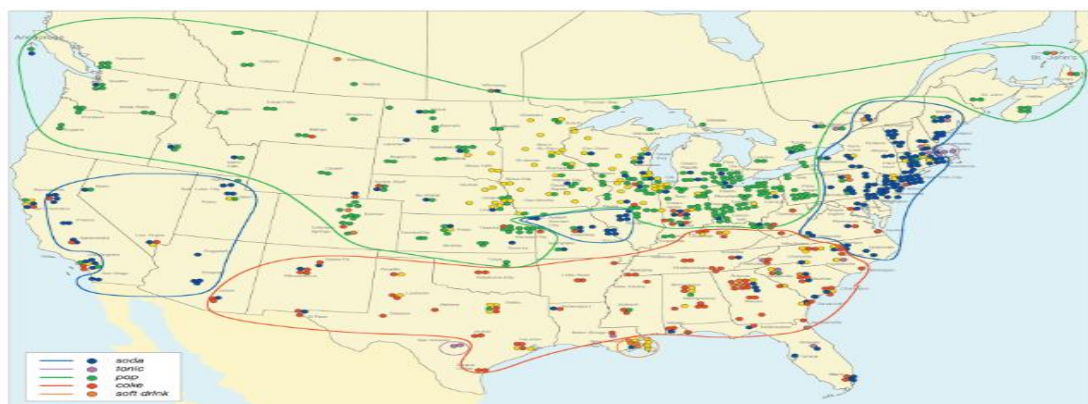
Hình 2: Bản đồ biến thể của nguyên âm trong từ “lay” (đặt/đế) (Nguồn: Kolfb, E, B. Glauser, W. Elmer, R. Stamm, 1979)



Hình 3: Bản đồ biến thể của cách nói “Give me it” (đưa nó cho tôi) (Nguồn: Kolfb, E, B. Glauser, W. Elmer, R. Stamm, 1979)

Các quốc gia ở Châu Mỹ cũng đã có những nghiên cứu về xây dựng bản đồ ngôn ngữ. Công trình *Linguistic Atlas of the United States and Canada (Bản đồ ngôn ngữ học Hoa Kỳ và Canada)* là dự án lớn từ những năm 30 của thế kỷ XX. Đây là tập bản đồ dựa trên tư liệu khảo sát phương ngữ ở diện rộng tại Hoa Kỳ và Canada. Công trình này được xem là đặt nền móng cho nghiên cứu ngôn ngữ địa lí ở Bắc Mỹ. Sau đó không lâu, *Linguistic Atlas of New England* với hơn 700 bản đồ (do Hans Kurath và các cộng sự) đã được xây dựng dựa trên tư liệu khảo sát ở các tiểu bang New Hampshire, Massachusetts, Connecticut, Maine, Vermont, Rhode Island, khu vực Long Island của New York. Ngoài ra, khu vực phía nam New Brunswick của Canada cũng được đưa vào nghiên cứu. Tập bản đồ này nổi bật với phương pháp phỏng vấn chi tiết và kí hiệu hóa dữ liệu trên bản đồ. (<http://findingaids.library.umass.edu/ead/mums330>).

Dựa trên những công nghệ hiện đại, *The Atlas of North American English* là công trình quan trọng của William Labov, Sharon Ash và Charles Boberg, xuất bản năm 2006 bởi Mouton de Gruyter. Khác với các atlas truyền thống, công trình không chỉ mô tả phân bố phương ngữ mà còn tập trung vào các quá trình biến đổi âm thanh đang diễn ra, từ đó xác định lại các vùng phương ngữ lớn như miền Bắc, miền Nam hay Canada. Đây là một công trình mang tính bước ngoặt, chuyển nghiên cứu bản đồ ngôn ngữ từ mô hình tĩnh, truyền thống sang mô hình động kết hợp với xã hội và công nghệ. Kết quả của nghiên cứu này là có 139 bản đồ màu được xây dựng. (<https://books.google.com.vn/books?id=qa4-dFqi6iMC&printsec=frontcover#v=onepage&q&f=true>). Hình 4 dưới đây thể hiện một biến thể của từ “carbonated beverage” (đồ uống có ga) ở các khu vực khảo sát.

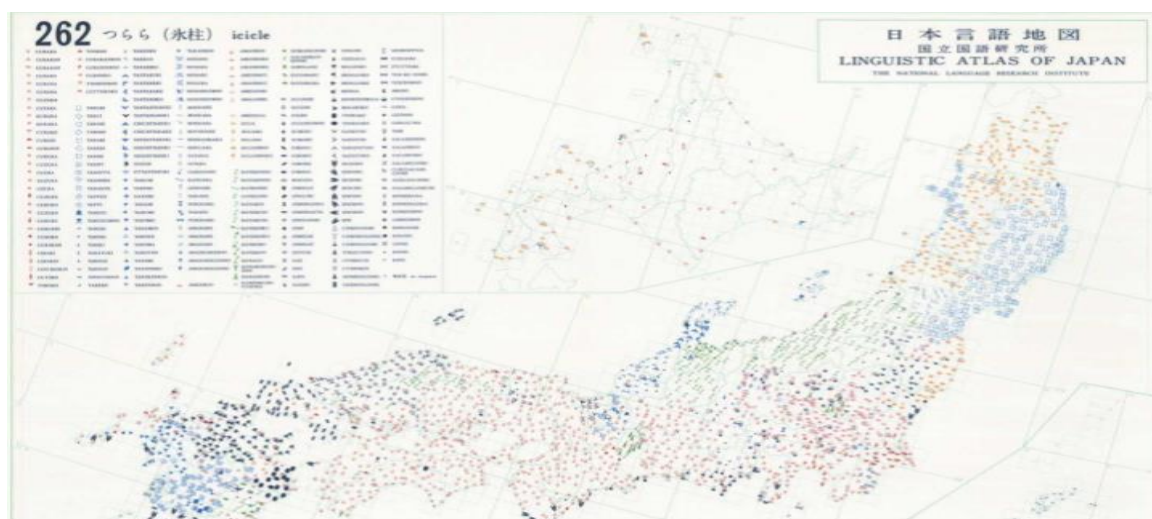


Map 21.1. Geographic distribution of terms for 'carbonated beverage'

Hình 4. Bản đồ phân bố từ “carbonated beverage” (đồ uống có ga)
(Nguồn: Kolfb, E, B. Glauser, W. Elmer, R. Stamm, 1979)

Mexico là quốc gia thuộc Mỹ Latinh cũng tiến hành xây dựng bản đồ ngôn ngữ có tên *Atlas lingüístico de México* (Bản đồ ngôn ngữ Mexico) do nhóm tác giả Juan M. Lope Blanch và các đồng nghiệp. Tập bản đồ bắt đầu được xuất bản vào năm 1990 và kéo dài đến khoảng năm 2000; gồm 958 bản đồ các loại. Tập bản đồ này được xây dựng dựa trên phương pháp xây dựng bản đồ truyền thống của ngôn ngữ học. (<https://www.jstor.org/stable/40300379?seq=11>)

Cùng với các quốc gia ở Châu Âu, Châu Mỹ thì nhiều quốc gia Châu Á cũng có những đóng góp quan trọng trong việc xây dựng bản đồ ngôn ngữ. Nhật Bản được xem là quốc gia có định hướng xây dựng bản đồ khá sớm. Vào những năm đầu của thế kỉ XX, Nhật Bản đã tiến hành xây dựng các bản đồ phương ngữ như *Phonetic Dialect Atlas* (Bản đồ ngữ âm phương ngữ) năm 1905 với 29 bản và năm 1906 xây dựng *Grammatical Dialect Atlas* (Bản đồ ngữ pháp phương ngữ) với 37 bản (Dẫn theo Takuichiro Onishi, tr.138). Tiếp đó, trong giai đoạn từ 1966 đến 1974 các nhà ngôn ngữ học nước này cũng đã xây dựng *Tập bản đồ ngôn ngữ Nhật Bản* (*Linguistics Atlas of Japan*) với một số lượng đồ sộ lên tới 300 bản đồ, chủ yếu là các mục từ vựng. Mặc dù có tên *Linguistics Atlas of Japan* nhưng trên thực tế, các tư liệu phục vụ phân tích được hoàn thiện nhờ dữ liệu phân bố phương ngữ trên toàn Nhật Bản.

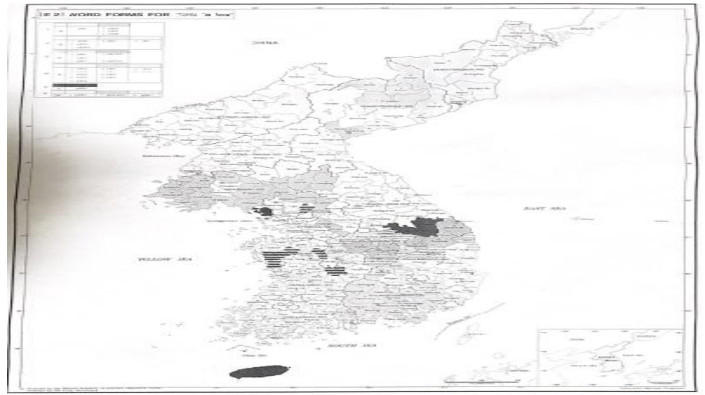


Hình 5. Bản đồ của từ “icicle” (cột băng/trụ băng) trong bản đồ ngôn ngữ Nhật Bản.
(Nguồn: National Language research Institute Japan (1966-1974))

Không dừng lại ở đó, các nhà ngôn ngữ học ở Viện Ngôn ngữ học Quốc gia Nhật Bản tiếp tục thực hiện việc xây dựng *Tập bản đồ ngữ pháp các phương ngữ tiếng Nhật (Grammar Atlas of Japanese Dialects)* với số lượng 350 bản đồ được chia thành 6 tập trong giai đoạn 1986 đến 2006 (Theo National Language research Institute Japan, tr.143). Có thể thấy, các bản đồ của Nhật Bản đã đánh dấu bước phát triển quan trọng khi không chỉ khảo sát từ vựng mà còn phân tích các biến thể ngữ pháp như cách chia động từ, trợ từ và cấu trúc câu giữa các vùng. Nhờ đó, hệ thống bản đồ ngôn ngữ Nhật Bản phản ánh đầy đủ cả ba cấp độ: từ vựng, ngữ âm và ngữ pháp. Mặt khác, ở giai đoạn này, với sự phát triển mạnh mẽ của hệ thống thông tin địa lý (GIS) nên các phương pháp và kỹ thuật hiện đại đã được áp dụng trong việc xây dựng bản đồ; đặc biệt là sự hỗ trợ của phần mềm ArcGIS.

Ở Trung Quốc, việc xây dựng bản đồ ngôn ngữ mặc dù được triển khai tương đối muộn nhưng lại có quy mô lớn và tính hệ thống cao. Tập bản đồ ngôn ngữ Trung Quốc (*Language Atlas of China*) được biên soạn trong giai đoạn 1987-1989 gồm 36 bản đồ, bao quát cả hệ thống phương ngữ Hán và các ngôn ngữ của các dân tộc thiểu số; phản ánh bức tranh đa dạng ngôn ngữ trên toàn lãnh thổ. Các bản đồ trong tập chủ yếu thể hiện sự phân bố không gian của các đặc điểm ngữ âm, từ vựng và phân vùng phương ngữ lớn như Quan thoại, Ngô, Mân... đồng thời có các bản đồ riêng cho khu vực ngôn ngữ dân tộc thiểu số. Phương pháp thể hiện chủ yếu dựa trên kí hiệu và màu sắc để phân vùng, giúp nhận diện rõ ranh giới và khu vực chuyển tiếp giữa các phương ngữ. Có thể thấy *Language Atlas of China* đã đánh dấu bước tiến quan trọng trong việc hệ thống hóa và trực quan hóa sự đa dạng phương ngữ và ngôn ngữ tại Trung Quốc trên quy mô quốc gia.

Viện Hàn lâm Khoa học Hàn Quốc đã tiến hành xây dựng *Tập bản đồ ngôn ngữ Hàn Quốc (Language Atlas of Korea)* vào những năm đầu của thập kỉ 90. Tập bản đồ này gồm 9 bản đồ đơn lẻ thể hiện sự phân bố các đặc điểm về ngôn điệu, dạng thức từ, sự kết hợp... của các từ cụ thể trong tiếng Hàn như *mar (từ)*, *yəu (con cáo)*, *tít (nghe)*, *kop (đẹp)*... Mặc dù là bản đồ ngôn ngữ quốc gia nhưng có thể thấy tính chất phương ngữ được thể hiện khá rõ qua tập bản đồ này. Hình 6 dưới đây thể hiện sự phân bố các biến thể của từ “fox” (con cáo) ở các địa bàn khảo sát.



Hình 6. Bản đồ phân bố dạng thức từ “a fox” (con cáo) trong *Language Atlas of Korea*. (Nguồn: National Academy of Sciences Republic of Korea (1987-1990))

Trong những năm gần đây, Viện Nghiên cứu Ngôn ngữ và Văn hóa Châu Á và Châu Phi (ILCAA) đã tổ chức nhiều dự án nghiên cứu liên quan đến việc phân bố các đơn vị từ vựng cơ bản (như “sun”, “milk”, “rice”, “ion”...) trong các ngôn ngữ thuộc nhiều ngữ hệ khác nhau ở Châu Á (Theo: *Study in Asian Geolinguistics I-VIII*). Cho đến nay đã có 8 ấn phẩm (xuất bản từ 2015) là kết quả của các nghiên cứu này với tên gọi *Study in Asian Geolinguistics* được đánh số từ I-VIII. Nhiều ngôn ngữ đã được nghiên



Map.1 Rice in Tungusic and Uralic languages

Hình 7. Bản đồ từ “Rice” (com/gạo) trong tiếng Tungusic và Uralic (Nguồn: *Study in Asian Geolinguistics II*, tr.6)

cứu kèm theo hàng loạt bản đồ về sự phân bố các đơn vị từ vựng ở các điểm khảo sát; cho thấy sự phong phú của các biến thể ngôn ngữ ở khu vực Châu Á. Bên cạnh đó, các *Hội thảo quốc tế về ngôn ngữ học địa lí Châu Á (International Conference on Asian Geolinguistics)* cũng đã xuất bản những cuốn kỉ yếu với nhiều bài nghiên cứu và bản đồ ngôn ngữ ở các khu vực Châu Á (Trịnh Cẩm Lan, Trần Thị Hồng Hạnh, 2025: 135). Đặc biệt, công nghệ GIS (chủ yếu với phần mềm ArcGIS) đã được sử dụng một cách rất hiệu quả khi xây dựng bản đồ.

Ở Việt Nam, ngay từ cuối thập niên 60 của thế kỉ XX, việc xây dựng bản đồ ngôn ngữ dân tộc thiểu số đã được tiến hành. Kết quả là đã có *Tập bản đồ về các tiếng địa phương Tây - Nùng* do Trường Đại học Tổng hợp Hà Nội xây dựng. Tiếp đó, *Bản đồ ngôn ngữ dân tộc* được xây dựng năm 1972 để thể hiện bối cảnh ngôn ngữ ở miền Nam Việt Nam. Việc xây dựng những bản đồ này ít nhiều mang tính thử công, đơn giản nhưng đã đánh dấu một bước phát triển mới trong ngôn ngữ học địa lí. Năm 2009, *Atlas địa lí Việt Nam* được xuất bản trong đó có bản đồ Dân tộc với sự phân bố các ngữ hệ và nhóm ngôn ngữ ở Việt Nam. Đây là một bản đồ được xây dựng rất công phu, phản ánh được sự phân bố của các nhóm ngôn ngữ ở Việt Nam. Đáng tiếc, cho đến nay Bản đồ phương ngữ tiếng Việt vẫn chưa được xây dựng.

4. GIS và các công cụ ứng dụng xây dựng bản đồ

Hệ thống thông tin địa lí (Geographic Information System - GIS) là nhóm phần mềm giữ vai trò nòng cốt trong việc xây dựng bản đồ phương ngữ, cho phép tích hợp, quản lí, phân tích và trực quan hóa dữ liệu ngôn ngữ trong mối quan hệ với không gian địa lí. Việc ứng dụng GIS trong nghiên cứu phương ngữ học góp phần chuyển hóa dữ liệu ngôn ngữ từ dạng bảng, văn bản hoặc số liệu khảo sát sang dạng bản đồ, qua đó làm nổi bật quy luật phân bố không gian của các biến thể ngôn ngữ.

Trong nghiên cứu phương ngữ học, phần mềm GIS thực hiện ba chức năng cơ bản. Thứ nhất, GIS cho phép quản lí và tổ chức dữ liệu không gian, bao gồm các điểm khảo sát, ranh giới hành chính, vùng địa lí và các lớp thông tin nền. Thứ hai, GIS hỗ trợ phân tích không gian, giúp xác định mối quan hệ giữa hiện tượng ngôn ngữ và các yếu tố địa lí như khoảng cách, vùng tiếp giáp, trục giao thông hoặc khu vực dân cư. Thứ ba, GIS đảm nhiệm chức năng trực quan hóa kết quả nghiên cứu, thể hiện sự phân bố, mức độ phổ biến và ranh giới của các hiện tượng phương ngữ thông qua các loại bản đồ chuyên đề (dẫn theo Trần Trọng Đức, 2011: 8-15).

Trong bối cảnh chuyển đổi số và sự phát triển mạnh mẽ của công nghệ địa không gian, các phần mềm xây dựng bản đồ đã trở thành công cụ nền tảng trong quản lí tài nguyên, quy hoạch lãnh thổ, phát triển đô thị, giao thông vận tải, môi trường và nhiều lĩnh vực kinh tế - xã hội khác. Các hệ thống này không chỉ hỗ trợ thiết kế bản đồ mà còn tích hợp chức năng thu thập, lưu trữ, xử lí, phân tích và trực quan hóa dữ liệu không gian ở nhiều cấp độ.

Trước hết, ArcGIS là một trong những hệ sinh thái GIS thương mại toàn diện và có ảnh hưởng lớn trên phạm vi toàn cầu, do Esri phát triển; được sử dụng rộng rãi trong nhiều lĩnh vực khoa học, trong đó có ngôn ngữ học xã hội và phương ngữ học. ArcGIS cho phép xây dựng cơ sở dữ liệu không gian phức tạp, trong đó mỗi điểm khảo sát phương ngữ có thể được gắn với nhiều thuộc tính ngôn ngữ như từ vựng, đặc trưng ngữ âm, ngữ pháp hoặc tần suất sử dụng.

Trong nghiên cứu bản đồ phương ngữ, ArcGIS hỗ trợ tạo các bản đồ phân bố hiện tượng ngôn ngữ, bản đồ vùng (areal maps), bản đồ mật độ và bản đồ phân cụm phương ngữ. Ngoài ra, các công cụ phân tích không gian của ArcGIS còn cho phép xác định mức độ tương đồng hoặc khác biệt giữa các điểm khảo sát, từ đó hỗ trợ việc xác lập ranh giới phương ngữ (Nguồn ESRI, “ArcGIS Pro: The World's Leading GIS Software”, Online).

Cùng với ArcGIS thì QGIS là phần mềm GIS mã nguồn mở tiêu biểu, được xem là một nền tảng xây dựng và phân tích bản đồ có tính ứng dụng cao và phổ biến rộng rãi. QGIS cho phép người dùng tự do khai thác, chỉnh sửa và mở rộng chức năng, nhờ đó hình thành một cộng đồng phát triển lớn mạnh và liên tục cập nhật. Về mặt chức năng, QGIS cung cấp bộ công cụ toàn diện cho biên tập, số hóa, quản lý thuộc tính, thiết kế ký hiệu bản đồ và xây dựng bố cục in ấn đạt chuẩn bản đồ học. Với lợi thế chi phí thấp và cộng đồng phát triển tích cực, QGIS trở thành lựa chọn phổ biến trong đào tạo học thuật, nghiên cứu khoa học và triển khai các dự án quy mô vừa và nhỏ.

Bên cạnh các phần mềm GIS cài đặt cục bộ trên máy tính, xu hướng phát triển hiện nay đang chuyển dịch mạnh sang mô hình WebGIS (Geographic Information System trên nền tảng web). WebGIS là hệ thống thông tin địa lý được triển khai thông qua môi trường trình duyệt, cho phép người dùng truy cập, khai thác, cập nhật và chia sẻ dữ liệu không gian mà không cần cài đặt phần mềm chuyên dụng trên thiết bị đầu cuối. Về mặt kiến trúc, WebGIS thường được xây dựng theo mô hình ba tầng (client - server - database), trong đó tầng máy chủ đảm nhiệm chức năng lưu trữ, quản lý và xử lý dữ liệu không gian; tầng cơ sở dữ liệu chịu trách nhiệm quản trị dữ liệu; còn tầng giao diện người dùng cung cấp khả năng hiển thị và tương tác trực tiếp trên nền web.

Về phương diện công nghệ, WebGIS có thể được phát triển dựa trên các nền tảng như ArcGIS Online, Mapbox, hoặc các thư viện JavaScript chuyên dụng như Leaflet và OpenLayers. Những công cụ này hỗ trợ hiển thị dữ liệu bản đồ trên một trang web được xây dựng. Bên cạnh chức năng trực quan hóa, hệ thống WebGIS còn cho phép triển khai các thao tác tương tác nâng cao như tìm kiếm, truy vấn không gian, phân tích chồng lớp và cập nhật dữ liệu theo thời gian thực, qua đó đáp ứng yêu cầu quản lý và ra quyết định dựa trên thông tin địa lý trong môi trường số (dẫn theo Vũ Xuân Định và cộng sự, 2024: 50-52).

Có thể khẳng định rằng phần mềm GIS là công cụ không thể thiếu trong việc xây dựng bản đồ phương ngữ hiện đại. Việc ứng dụng GIS giúp nghiên cứu phương ngữ vượt ra khỏi cách tiếp cận mô tả truyền thống, chuyển sang hướng tiếp cận định lượng và không gian hóa dữ liệu ngôn ngữ. Tùy theo mục tiêu nghiên cứu, quy mô dữ liệu và điều kiện triển khai, người nghiên cứu có thể lựa chọn kết hợp nhiều công cụ GIS khác nhau nhằm đạt được hiệu quả cao nhất trong việc xác lập và biểu diễn không gian phương ngữ.

Ứng dụng những tiến bộ của công nghệ thông tin, một số nghiên cứu của các tác giả Việt Nam cũng như tác giả nước ngoài nghiên cứu về tiếng Việt đã sử dụng công nghệ GIS (chủ yếu là phần mềm ArcGIS) để tiến hành xây dựng bản đồ ngôn ngữ/ phương ngữ. Shumizu Massaki (2014), trong nghiên cứu *The distribution of final palatals in Vietnamese dialects* xây dựng một số bản đồ phân bố các âm cuối trong các phương ngữ tiếng Việt qua các từ cụ thể như “sound” (âm thanh), “eat” (ăn), “canal” (kênh), “news” (tin), “name” (tên)... Mika Kondo (2014) trong nghiên cứu *The Geolinguistic Distribution of the Vietnamese Words for ‘to Squat Down’* xây dựng một bản đồ phân bố các biến thể tiếng Việt của từ trong tiếng Anh là ‘to Squat Down’ (ngồi xổm/ ngồi chò hõ/ ngồi chồm hõm/ ngồi chồm chồm...). Ở Việt Nam có thể kể đến các nghiên cứu như *Biến thể ngữ âm đánh dấu ở huyện Quốc Oai và không gian hành chức của chúng trên bản đồ* của Lê Quốc Hưng (2018); nghiên cứu *Bản đồ thổ ngữ các huyện Giao Thủy, Hải Hậu, Nghĩa Hưng (Tỉnh Nam Định)* của Nguyễn Thị Thu Hà (2019), nghiên cứu *Characteristics and distribution of the variants of maternal kinship terms in Sơn Tây dialect (Hanoi)* của Trịnh Cẩm Lan (2023) hay nghiên cứu *Khảo sát và lập bản đồ các biến thể ngữ âm và từ vựng hiện hữu tại các huyện Quốc Oai, Thạch Thất, Ba Vì* cũng của tác giả Trịnh Cẩm Lan, (2023). Các nghiên cứu này đã áp dụng phương pháp xây dựng bản đồ dựa trên phần mềm ArcGIS bản online. Kết quả là đã có những bản đồ về sự phân bố các biến thể ở các điểm khảo sát. Tuy nhiên, mục

đích chính của các nghiên cứu này là định vị các biến thể ở các điểm điều tra nên các kĩ thuật xây dựng bản đồ cũng như thông tin của bản đồ như tên bản đồ, tỉ lệ, thước tỉ lệ, lưới chiếu, kim chỉ hướng... không được thể hiện. Tiếp đó, nghiên cứu *Xây dựng bản đồ một số hiện tượng ngữ âm huyện Thạch Thất và Quốc Oai (Hà Nội)* (2021) và nghiên cứu *Xây dựng bản đồ biến thể ngữ âm tiếng Thanh Hóa dựa trên phần mềm ArcGIS* (2024) của Nguyễn Tài Thái đã ứng dụng phần mềm ArcGIS bản Desktop 8.0 để xây dựng bản đồ phân bố/ định vị các biến thể ngữ âm qua các thổ ngữ. Có thể nói, các bản đồ được xây dựng từ những nghiên cứu này đã thể hiện được khá nhiều thông tin; bao gồm cả thông tin của bản đồ ngôn ngữ lẫn thông tin của bản đồ hành chính. Đây có thể xem là những hướng đi phù hợp của ngôn ngữ học địa lí; đồng thời kết hợp được những tiến bộ của khoa học công nghệ hiện đại.

5. Kết luận

Ngôn ngữ học địa lí đã có những đóng góp quan trọng và mang tính nền tảng đối với nghiên cứu phương ngữ. Bằng việc đặt các hiện tượng ngôn ngữ trong mối quan hệ với không gian địa lí, hướng nghiên cứu này không chỉ dừng lại ở việc miêu tả các đặc điểm ngữ âm, từ vựng hay ngữ pháp của từng phương ngữ, thổ ngữ, mà nó còn làm rõ quy luật phân bố, mức độ lan tỏa và ranh giới của các biến thể. Nhờ đó, phương ngữ học được mở rộng từ bình diện miêu tả sang bình diện không gian - lịch sử, góp phần nhận diện cấu trúc vùng phương ngữ và các quá trình tiếp xúc, giao thoa cũng như biến đổi ngôn ngữ.

Việc xây dựng bản đồ ngôn ngữ, phương ngữ là một trong những kết quả tiêu biểu của ngôn ngữ học địa lí. Bản đồ hóa dữ liệu ngôn ngữ không chỉ nâng cao tính trực quan và hệ thống của kết quả nghiên cứu mà còn tạo điều kiện cho việc phân tích định lượng cũng như xác lập ranh giới phương ngữ một cách khoa học. Đặc biệt, trong bối cảnh chuyển đổi số và sự phát triển của công nghệ GIS và hàng loạt công cụ xây dựng bản đồ như hiện nay thì việc xây dựng bản đồ phương ngữ không còn là lựa chọn bổ trợ mà đã trở thành yêu cầu quan trọng và cần thiết, bảo đảm tính hiện đại, chuẩn hóa và khả năng liên thông dữ liệu trong nghiên cứu. Chính vì thế, việc đẩy mạnh ứng dụng công nghệ để xây dựng bản đồ trong nghiên cứu phương ngữ tiếng Việt vừa là hướng tiếp cận phù hợp với xu thế quốc tế, vừa là điều kiện thiết yếu để nâng cao chất lượng và chiều sâu của ngôn ngữ học địa lí trong giai đoạn hiện nay.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Hoàng Thị Châu (2004). *Phương ngữ học tiếng Việt*. NXB Đại học Quốc gia Hà Nội.
2. Vũ Xuân Định và cộng sự (2024). *Ứng dụng WebGIS trong thành lập bản đồ trực tuyến hiện trạng đốt rác thải ngoài trời quận Hoàng Mai, Thành phố Hà Nội*. Tạp chí Khoa học Đo đạc và Bản đồ, số 60 (6/2024).
3. Trần Trọng Đức (2011). *GIS căn bản*. NXB Đại học Quốc gia TP Hồ Chí Minh.
4. Lê Thị Hương Giang (2014). *Cụm thổ âm Ninh Giang - Vĩnh Bảo - Quỳnh Phụ*. Luận văn Thạc sĩ, Trường ĐHSP Hà Nội.
5. Nguyễn Thị Thu Hà (2019). *Bản đồ các thổ ngữ huyện Giao Thủy, Hải Hậu, Nghĩa Hưng (tỉnh Nam Định)*. Luận văn Thạc sĩ, Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia Hà Nội.
6. Lê Quốc Hưng (2018). *Biến thể ngữ âm đánh dấu ở huyện Quốc Oai và không gian hành chức của chúng trên bản đồ*. Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia Hà Nội.
7. Trịnh Cẩm Lan (2023a). *Characteristics and distribution of the variants of maternal kinship terms in Son Tay dialect (Hanoi)*, in *Proceedings of the fifth International Conference on Asian Geolinguistics*. Geolinguistic Society of Japan.
8. Trịnh Cẩm Lan (2023b). *Khảo sát và lập bản đồ các biến thể ngữ âm và từ vựng hiện hữu tại các huyện Quốc Oai, Thạch Thất, Ba Vì*. Đề tài KHCN cấp Đại học Quốc gia Hà Nội.

9. Trịnh Cẩm Lan, Trần Thị Hồng Hạnh (2025). *Ngôn ngữ học địa lí, phương ngữ học và bản đồ phương ngữ: khái niệm, lịch sử và những vấn đề đang đặt ra*. Tạp chí Khoa học và Công nghệ Trường Đại học Công nghiệp Hà Nội. Tập 61, Số 2.
10. Nguyễn Văn Nguyên (2003). *Miêu tả đặc trưng ngữ âm phương ngữ Nghệ Tĩnh*. Luận án Tiến sĩ ngữ văn, Vinh.
11. Trần Thị Phụng (2014). *Giáo trình Hệ thống thông tin địa lí*. NXB Đại học Huế.
12. Nguyễn Tài Thái (2018). *Biến thể ngữ âm một số thổ ngữ ven biển Bắc bộ*, Đề tài nghiên cứu khoa học cấp Bộ, Viện Hàn lâm KHXH Việt Nam.
13. Nguyễn Tài Thái (2021). *Xây dựng bản đồ một số hiện tượng ngữ âm huyện Thạch Thất và Quốc Oai (Hà Nội)*. Viện Ngôn ngữ học. Đề tài nghiên cứu KH cấp cơ sở.
14. Nguyễn Tài Thái (2024). *Xây dựng bản đồ biến thể ngữ âm tiếng Thanh Hóa dựa trên phần mềm ArcGIS*. Đề tài nghiên cứu khoa học cấp Bộ, Viện Hàn lâm Khoa học Xã hội Việt Nam.
15. Võ Xuân Trang (1997). *Phương ngữ Bình Trị Thiên*. NXB Khoa học Xã hội, Hà Nội.
16. Phạm Hữu Viện (2014). *Bản đồ các thổ ngữ tiếng Nghi Lộc - Tỉnh Nghệ An*. Luận văn Thạc sĩ, Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia Hà Nội.
17. Nguyễn Như Ý (chủ biên), Hà Quang Năng, Đỗ Việt Hùng, Đặng Ngọc Lễ (2001). *Từ điển giải thích thuật ngữ Ngôn ngữ học*. NXB Giáo dục.

Tiếng Anh

18. Allen, H. B. (1973). *The linguistic Atlas of the upper Midwest*. University of Minnesota Press, Minneapolis.
19. *Association of Asian Geolinguistics* (2015-2017). Study in Asian Geolinguistics I-VIII.
20. *Australian Academy of the Humanities and Chinese Academy of Social Sciences* (1989, 2012). *Language Atlas of China*.
21. Chambers J. K. and Peter Trudgill (1998). *Dialectology*, Cambridge Textbooks in Linguistics.
22. Charles Boberg, John Nerbonne and Dominic Watt (2018). *The handbook of dialectology*. Wiley Blackwell.
23. Crystal David (2008). *A Dictionary of linguistics and phonetics*. Wiley Blackwell.
24. Dhungel, B.R. (2009). *Linguistic geography: An introduction*. Sampreshan. Kirtipur: Department of Nepali Education, FOE, TU.
25. Francis, W. N., Review of H. Orton and N. Wright (1978). *A word geography of England in American speech*.
26. Georg Wenker (1981). *Linguistic Atlas of Northern and Central Germany*. Strasbourg, London.
27. Harold Orton, Stewart Sanderson and John Widdowson (1978). *The Linguistic Atlas of England*. Croom Helm Ltd, 2-10 St John's Road, London SW11.
28. John M. Kirk, Stewart Sanderson, and J. D. A. Widdowson (1985). *Studies in linguistic geography: The dialects of English in Britain and Ireland*. London: Croom Helm.
29. Kolff, E. (1966). *Phonological Atlas of the Northern region*. Frank, Bern.
30. Kolff, E, B. Glauser, W. Elmer, R. Stamm (1979). *Atlas of English sound*. Frank, Bern.
31. Mika Kondo (2014). *The geolinguistic distribution of the Vietnamese words for 'to Squat Down'*, trong *The second International Conference on Asian Geolinguistics*. Bangkok, Thailand.
32. National Academy of Sciences Republic of Korea (1987-1990). *Language Atlas of Korea*.
33. National Language research Institute Japan (1966-1974). *Linguistics Atlas of Japan*.
34. National Language research Institute Japan (1989-2006). *Grammar Atlas of Japanese Dialects*.
35. Philipp Stoeckle (2014). *Linguistics Geography and GIS*, Conference: Summer School "Tools and Techniques in Geolinguistics", University of Zurich. https://www.researchgate.net/publication/270891234_Linguistic_Geography_and_GIS (download ngày 20/11/2023).
36. Richards, Platt and Weber (1985). *Longman Dictionary of Applied Linguistics*, Longman Group UK Limited.
37. Shimizu M. (2014). *The distribution of final palatals in Vietnamese dialects*, trong *The 2nd International Conference on Asian Geolinguistics*. Bangkok, Thailand.
38. Takuichiro Onishi (2010). *Mapping Japanese dialects*. Dialectologia. Special issue, I, 137-146.
39. William Labov, Sharon Ash and Charles Boberg (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Mouton de Gruyter, Berlin.