



ISO 9001:2008

TRƯỜNG ĐẠI HỌC TRÀ VINH
HỘI ĐỒNG KHOA HỌC

**BÁO CÁO TỔNG KẾT
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CẤP TRƯỜNG**

**CÔNG THÔNG TIN BẢN ĐỒ NGÔN NGỮ DÂN
TỘC HỌC (TRƯỜNG HỢP TỈNH TRÀ VINH)**

Chủ nhiệm đề tài: NGUYỄN THỊ HUỆ

Chức danh: Trưởng khoa

Đơn vị: Khoa Ngôn ngữ - Văn hóa - Nghệ thuật Khmer Nam Bộ

Trà Vinh, ngày tháng năm



ISO 9001:2008

TRƯỜNG ĐẠI HỌC TRÀ VINH
HỘI ĐỒNG KHOA HỌC

**BÁO CÁO TỔNG KẾT
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CẤP TRƯỜNG**

**CÔNG THÔNG TIN BẢN ĐỒ NGÔN NGỮ DÂN
TỘC HỌC (TRƯỜNG HỢP TỈNH TRÀ VINH)**

Xác nhận của cơ quan chủ quản
(Ký, đóng dấu, ghi rõ họ tên)

Chủ nhiệm đề tài
(Ký, ghi rõ họ tên)

Nguyễn Thị Huệ

Trà Vinh, ngày tháng năm

MỤC LỤC

THÔNG TIN CHUNG VỀ ĐỀ TÀI	4
DANH MỤC BẢNG BIỂU	7
DANH MỤC HÌNH ẢNH	8
PHẦN MỞ ĐẦU.....	10
1 Tính cấp thiết của đề tài.....	10
2 Tổng quan nghiên cứu	11
2.1 Tình hình nghiên cứu trong nước.....	11
2.2 Tình hình nghiên cứu ngoài nước	13
3 Mục tiêu.....	19
4 Đối tượng, phạm vi và phương pháp nghiên cứu.....	19
4.1 Đối tượng và địa điểm nghiên cứu.....	19
4.2 Nội dung thực hiện.....	19
4.3 Phương pháp nghiên cứu, kỹ thuật sử dụng.....	19
PHẦN NỘI DUNG	21
CHƯƠNG 1. THU THẬP VÀ SỐ HÓA DỮ LIỆU	21
1.1 Thu thập dữ liệu	21
1.1.1 Giới thiệu về dữ liệu cần thiết cho đề tài	21
1.1.2 Phương pháp thu nhập dữ liệu ngôn ngữ và dân tộc	21
1.1.3 Thu thập mẫu dữ liệu ngôn ngữ và dân tộc	22
1.1.4 Mô phỏng dữ liệu ngôn ngữ.....	25
1.2 Số hóa bản đồ ngôn ngữ dân tộc học	26
1.2.1 Các chỉ số của bản đồ ngôn ngữ	26
1.2.2 Các chỉ số về dân tộc	28
1.2.3 Các chỉ số về ngôn ngữ trên từng dân tộc.....	28
CHƯƠNG 2. PHÁT TRIỂN CÔNG THÔNG TIN BẢN ĐỒ NGÔN NGỮ DÂN TỘC HỌC.....	30
2.1 Cơ sở dữ liệu địa lý, ngôn ngữ và dân tộc	30

2.1.1	Sự cần thiết của cơ sở dữ liệu và chọn hệ quản trị cơ sở dữ liệu	30
2.1.2	Mô tả các khái niệm về dữ liệu	30
2.1.3	Mô hình dữ liệu.....	31
2.1.4	Cơ sở dữ liệu ở mức vật lý.....	32
2.2	Tổng quan về công thông tin bản đồ ngôn ngữ và dân tộc	33
CHƯƠNG 3. ỨNG DỤNG CÔNG THÔNG TIN BẢN ĐỒ NGÔN NGỮ		
DÂN TỘC HỌC TRONG ĐÁNH GIÁ CẢNH HUỐNG NGÔN NGỮ.....		38
3.1	Ngôn ngữ và đời sống cộng đồng	38
3.1.1	Tính thiết yếu của ngôn ngữ trong đời sống.....	38
3.1.2	Bối cảnh ngôn ngữ ở tỉnh Trà Vinh	39
3.2	Khai thác giá trị ứng dụng của công thông tin bản đồ trực tuyến về cảnh huống ngôn ngữ ở tỉnh Trà Vinh	41
3.2.1	Đối với các nhà quản lý chính quyền địa phương	44
3.2.2	Đối với các nhà giáo dục, nhà nghiên cứu.....	45
3.2.3	Đối với người dân	45
PHẦN KẾT LUẬN.....		46
1	Kết quả đề tài và thảo luận	46
1.1	Tổng kết các hoạt động và kết quả.....	46
1.2	Thảo luận.....	48
2	Kiến nghị	49
2.1	Các hoạt động ngắn hạn	49
2.2	Các hoạt động triển khai mở rộng.....	49
TÀI LIỆU THAM KHẢO.....		50
PHỤ LỤC		52
Phụ lục 1 - Phiếu thu thập dữ liệu ngôn ngữ và dân tộc		52
Phụ lục 2 - Thu thập số liệu về dân tộc.....		53
Phụ lục 3 - Vài kết quả từ công thông tin LangMap.....		57

Phụ lục 4 - Từ điển dữ liệu..... 72
Phụ lục 5 - Một số câu lệnh để tạo các dịch vụ trực tuyến (on-the-fly) 74

TÓM TẮT

(Nêu ngắn gọn và rõ ràng về mục đích, nội dung, phương pháp, kết quả và kết luận đã được trình bày trong báo cáo chính, nếu được có thêm phần tiếng Anh).

Ngôn ngữ giữ vai trò thiết yếu cho hoạt động giao tiếp, đồng thời là sợi dây gắn kết trí tuệ của dân tộc, được truyền đạt, kế thừa và phát huy từ đời này sang đời khác. Có thể khẳng định, ngôn ngữ thể hiện bản sắc của dân tộc. Do đó, việc tìm hiểu tình hình sử dụng ngôn ngữ đóng vai trò quan trọng. Đề tài tập trung phân tích dữ liệu thu thập về sự phân bố các ngôn ngữ ở từng khu vực địa phương theo số lượng người chọn ngôn ngữ sử dụng hàng ngày; dữ liệu về năng lực ngôn ngữ (nghe nói/đọc viết) theo từng ngôn ngữ được chọn. Đây là một nghiên cứu trường hợp, khai thác thông tin tại tỉnh Trà Vinh – một tỉnh có đặc điểm nổi bật là đông người Khmer và họ đã sinh sống lâu đời tại nơi đây cùng với người Kinh, người Hoa. Mục đích nghiên cứu là cung cấp một bức tranh cụ thể - công thông tin bản đồ trực tuyến về ngôn ngữ dân tộc của ba ngôn ngữ (Việt, Khmer, Hoa) tại tỉnh Trà Vinh (từ cấp xã đến cấp tỉnh). Việc xác định mối liên hệ theo không gian địa lý và các vấn đề kinh tế - xã hội khác giữa sự lựa chọn ngôn ngữ và mức độ thành thạo ngôn ngữ nhằm cung cấp cơ sở để định hướng các quyết định, chính sách ngôn ngữ, các vấn đề kinh tế, chính trị và xã hội.

DANH MỤC BẢNG BIỂU

Bảng 1. Tổng kết mẫu dữ liệu về ngôn ngữ và dân tộc ở nhóm 8, phường 8, thành phố Trà Vinh	24
Bảng 2. Tổng kết mẫu dữ liệu về ngôn ngữ và dân tộc ở nhóm 1, phường 2, thành phố Trà Vinh	24
Bảng 3. Phiếu thu thập dữ liệu ngôn ngữ và dân tộc ở mức hộ	52
Bảng 4. Số liệu về dân tộc của tỉnh Trà Vinh (<i>Căn cứ theo Quyết định số 572/QĐ-UBND ngày 18/3/2016 của Chủ tịch Ủy ban nhân dân tỉnh Trà Vinh về việc phê duyệt kết quả tổng điều tra, rà soát hộ nghèo, hộ cận nghèo năm 2015 trên địa bàn tỉnh Trà Vinh</i>)	53

DANH MỤC HÌNH ẢNH

Hình 1. Bản đồ tỷ lệ những người có thể nói tiếng Anh ở châu Âu	14
Hình 2. Bản đồ đa dạng ngôn ngữ của các nước trên thế giới.....	15
Hình 3. Bản đồ các phương ngữ của nước Đức.....	16
Hình 4. Bản đồ ngôn ngữ và dân tộc ở miền Nam Việt Nam (năm 1972)	17
Hình 5. Mô tả dữ liệu.....	31
Hình 6. Mô hình thực thể.....	32
Hình 7. Cài đặt cơ sở dữ liệu ở mức vật lý.....	33
Hình 8. Tổng quan về hệ thống.....	34
Hình 9. Các thành phần của hệ thống và công nghệ sử dụng.....	35
Hình 10. Giao diện của trang công khai	36
Hình 11. Giao diện của trang bảo mật	37
Hình 12. Bản đồ về sự đa dạng hóa ngôn ngữ ở tỉnh Trà Vinh.....	41
Hình 13. Bản đồ về sự phân bố tiếng Khmer ở huyện Tiểu Cần.....	42
Hình 14. Bản đồ về sự phân bố tiếng Khmer ở huyện Trà Cú	43
Hình 15. Các bước thực hiện mà đề tài đã hoàn thành	47
Hình 16. Bản đồ dân tộc chiếm đa số trong tỉnh Trà Vinh.....	57
Hình 17. Số liệu về dân tộc của từng huyện trong tỉnh Trà Vinh.....	58
Hình 18. Số liệu về dân tộc của tỉnh Trà Vinh	58
Hình 19. Bản đồ phân bố dân tộc Kinh mức xã của tỉnh Trà Vinh	59
Hình 20. Bản đồ phân bố dân tộc Kinh mức huyện của tỉnh Trà Vinh	59
Hình 21. Bản đồ phân bố dân tộc Khmer mức xã của tỉnh Trà Vinh	60
Hình 22. Bản đồ phân bố dân tộc Khmer mức huyện của tỉnh Trà Vinh	60
Hình 23. Bản đồ phân bố dân tộc Kinh của thành phố Trà Vinh	61
Hình 24. Bản đồ phân bố dân tộc Khmer của thành phố Trà Vinh	61
Hình 25. Bản đồ phân bố dân tộc Kinh của huyện Trà Cú.....	62
Hình 26. Bản đồ phân bố dân tộc Khmer của huyện Trà Cú.....	62

Hình 27. Số liệu về dân tộc của từng xã/phường của thành phố Trà Vinh....	63
Hình 28. Bản đồ phân bố sử dụng tiếng Việt của tỉnh Trà Vinh	63
Hình 29. Bản đồ phân bố sử dụng tiếng Khmer của tỉnh Trà Vinh	64
Hình 30. Bản đồ phân bố sử dụng tiếng Việt với mức thông thạo biết đọc và viết của tỉnh Trà Vinh.....	64
Hình 31. Bản đồ phân bố sử dụng tiếng Khmer với mức thông thạo biết đọc và viết của tỉnh Trà Vinh.....	65
Hình 32. Bản đồ phân bố sử dụng tiếng Việt của thành phố Trà Vinh	65
Hình 33. Bản đồ phân bố sử dụng tiếng Khmer của thành phố Trà Vinh	66
Hình 34. Bản đồ dân tộc chiếm đa số mức xã của tỉnh Sóc Trăng	67
Hình 35. Bản đồ dân tộc chiếm đa số mức huyện của tỉnh Sóc Trăng	67
Hình 36. Bản đồ dân tộc Kinh mức huyện của tỉnh Sóc Trăng	68
Hình 37. Bản đồ dân tộc Khmer mức huyện của tỉnh Sóc Trăng	68
Hình 38. Bản đồ dân tộc Hoa mức huyện của tỉnh Sóc Trăng	69
Hình 39. Bản đồ dân tộc Hoa mức xã của tỉnh Sóc Trăng	69
Hình 40. Số liệu về dân tộc mức tỉnh của tỉnh Sóc Trăng	70
Hình 41. Số liệu về dân tộc mức huyện của tỉnh Sóc Trăng.....	70
Hình 42. Bản đồ đa dạng hóa ngôn ngữ của tỉnh Sóc Trăng	71
Hình 43. Bản đồ phân bố tiếng Khmer của tỉnh Sóc Trăng.....	71

PHẦN MỞ ĐẦU

1 Tính cấp thiết của đề tài

Trà Vinh là tỉnh thuộc khu vực đồng bằng sông Cửu Long, nằm giữa sông Tiền và sông Hậu, có 65 km bờ biển, diện tích tự nhiên 2.341 km², trong đó đất sản xuất nông nghiệp chiếm 81,8%. Dân số chung là 1.015.284 người, trong đó dân tộc Kinh 686.009 người, chiếm tỷ lệ 67,56%; dân tộc Khmer 321.084 người, chiếm tỷ lệ 31,62%; dân tộc Hoa 7.690 người, chiếm tỷ lệ 0,77% còn lại là dân tộc Chăm và một số dân tộc khác 501 người, chiếm tỷ lệ 0,05%. Tỉnh có 61 xã, phường, thị trấn có trên 20% đồng bào dân tộc Khmer. Do đó, cần thực hiện nội dung thứ 14 là Xây dựng bản đồ ngôn ngữ ở các tỉnh có nhiều dân tộc thiểu số, thực hiện từ năm 2017 đến năm 2019. Nội dung này nằm trong lộ trình triển khai thực hiện đề án “tăng cường tiếng Việt cho trẻ mầm non và học sinh tiểu học vùng dân tộc thiểu số giai đoạn 2016-2020, định hướng đến 2025” (Quyết định số 1008/QĐ-TTG ngày 02/6/2016 của Thủ tướng Chính phủ, ban hành kèm theo quyết định số 2805/QĐ-GDDT ngày 15 tháng 8 năm 2016 của Bộ Giáo dục và Đào tạo).

Trong suốt quá trình phát triển cùng các dân tộc Kinh, Hoa tại vùng đất Nam Bộ, cộng đồng Khmer đã giao hoà, gắn kết tạo nên một nền văn hóa, ngôn ngữ phong phú, đa dạng. Ngôn ngữ giữ vai trò thiết yếu cho hoạt động giao tiếp, đồng thời là sợi dây gắn kết trí tuệ của dân tộc, được truyền đạt, kế thừa và phát huy từ đời này sang đời khác. Có thể khẳng định, ngôn ngữ thể hiện bản sắc của dân tộc. Do đó, việc tìm hiểu tình hình sử dụng ngôn ngữ đóng vai trò quan trọng nhằm xác định vai trò, chức năng, giá trị của từng ngôn ngữ đang tồn tại trong cộng đồng.

Cảnh huống ngôn ngữ là một khái niệm được coi là quan trọng bậc nhất của ngôn ngữ học xã hội vì ngôn ngữ học xã hội nghiên cứu về mặt chức năng của ngôn ngữ (Trần Phương Nguyên, 2012). Cảnh huống ngôn ngữ “là toàn bộ các hình thái tồn tại (bao gồm cả phong cách) của một ngôn ngữ, hay toàn bộ các ngôn ngữ trong một mối quan hệ tương hỗ về lãnh thổ, xã hội và sự tương tác về mặt chức năng, trong giới hạn của một khu vực địa lý nhất định hay một thực thể hành chính-chính trị” (Nguyễn Hữu Hoàn; Nguyễn Văn Lợi; Tạ Văn Thông, 2013). Trong đó, các tác giả khẳng định rằng, khi miêu tả cảnh huống ngôn ngữ ở nước ta, cần quan tâm đến một số nội dung, trong đó bao gồm 1) số lượng các dân tộc – ngôn ngữ đang hoạt động hành chức trên địa bàn; 2) số

lượng người sử dụng từng ngôn ngữ và cách phân bố của các đối tượng này (bao gồm cả sự phân hóa xã hội, trình độ học vấn, ...). Như vậy, để xác định cảnh hướng ngôn ngữ tại tỉnh Trà Vinh, đòi hỏi một nghiên cứu chuyên sâu nhưng cụ thể. Diễn đạt cảnh hướng ngôn ngữ tại một địa phương bằng công cụ cổng thông tin địa lý và bản đồ số về ngôn ngữ và dân tộc qua giao diện trực tuyến là một việc làm cần thiết. Nhận thức rõ về bối cảnh ngôn ngữ sẽ giúp phân tích một số nguyên nhân gây nên sự khác biệt trong lựa chọn ngôn ngữ và năng lực sử dụng ngôn ngữ theo từng khu vực địa lý, xác định mối liên hệ theo không gian địa lý và các vấn đề kinh tế-xã hội khác, tạo cơ sở cho việc định hướng các quyết định, chính sách ngôn ngữ, đặc biệt là ngôn ngữ dân tộc.

Đề tài nghiên cứu hướng đến phương pháp và công cụ nghiên cứu về cảnh hướng ngôn ngữ một cách linh hoạt và hiện đại, áp dụng công nghệ nhằm thuận lợi cho quá trình tiếp nhận và cập nhật kịp thời các diễn biến ngôn ngữ. Từ đó, hoạt động quản lý dân cư, triển khai thực hiện các chính sách ngôn ngữ, nghiên cứu tình hình ngôn ngữ đang được sử dụng sẽ được hiệu quả hơn. Ngoài ra, đây còn là một công cụ nghiên cứu ngôn ngữ và dân tộc nhằm góp phần vào hoạch định các chủ trương chính sách ngôn ngữ cho tỉnh nhà và cho quốc gia, khai thác hiệu quả vai trò là Khoa trọng điểm quốc gia.

2 Tổng quan nghiên cứu

2.1 Tình hình nghiên cứu trong nước

Trần Trí Dõi (Trần Trí Dõi, 2001), (Trần Trí Dõi, 1999) với “Nghiên cứu các dân tộc thiểu số ở Việt Nam”, (Khổng Diễn, 1995) với “Dân số và dân số tộc người ở Việt Nam”, (Tạ Văn Thông, 2009)(chủ biên) với “Tìm hiểu ngôn ngữ các dân tộc ở Việt Nam”, ...và một loạt các bài viết của (Nguyễn Hữu Hoàn; Nguyễn Văn Lợi; Tạ Văn Thông, 2013): “Ngôn ngữ, chữ viết các dân tộc thiểu số ở Việt Nam (Những vấn đề chung)”, (Nguyễn Văn Khang, 2014) với “Nhìn lại chính sách ngôn ngữ của Đảng và nhà nước Việt Nam về tiếng Việt và những vấn đề đặt ra đối với tiếng Việt hiện nay”... Nhìn chung các công trình chỉ nghiên cứu từng khía cạnh khác nhau của tình hình sử dụng một ngôn ngữ nào đó hoặc một nhân tố nào đó tại một địa phương trên lãnh thổ Việt Nam.

Trong công trình “Ngôn ngữ, chữ viết các dân tộc thiểu số ở Việt Nam” (Nguyễn Hữu Hoàn; Nguyễn Văn Lợi; Tạ Văn Thông, 2013), ở chương II tác giả Nguyễn Hữu Hoàn viết về “cảnh hướng ngôn ngữ các dân tộc thiểu số ở

Việt Nam”. Trong đó, tác giả trình bày các vấn đề về cảnh huống ngôn ngữ nói chung và các vấn đề về cảnh huống ngôn ngữ ở Việt Nam. Nguyễn Văn Hoàn đã giới thiệu một số quan niệm về cảnh huống ngôn ngữ của thế giới và Việt Nam, trong đó quan niệm của Giáo sư Nguyễn Văn Lợi là rõ ràng hơn cả. Có thể nói, khái niệm này đã khái quát được tất cả các vấn đề cần quan tâm cũng như mối quan hệ của các vấn đề của một cảnh huống ngôn ngữ. Về cảnh huống ngôn ngữ ở Việt Nam, Nguyễn Hữu Hoàn đã trình bày cụ thể và sâu về các bình diện như: đặc điểm về số lượng dân tộc, ngôn ngữ; đặc điểm về quan hệ cội nguồn, loại hình của các ngôn ngữ; đặc điểm về sự hình thành và phát triển của các dân tộc, ngôn ngữ; đặc điểm về dân số - tộc người và địa lí - tộc người; trình độ phát triển, vai trò vị thế của các ngôn ngữ; thái độ ngôn ngữ của các cộng đồng dân tộc.

Ở khu vực phía Bắc, tình hình nghiên cứu về cảnh huống ngôn ngữ thu hút được nhiều sự quan tâm lựa chọn của các nhà nghiên cứu:

- Trong công trình “Cảnh huống ngôn ngữ ở Thái Nguyên” (Dương Thị Thanh Hoa, 2010), tác giả nhận diện cảnh huống ngôn ngữ về mặt khái niệm, nhân tố hình thành, phân loại. Bên cạnh đó, đặc điểm cảnh huống ngôn ngữ ở Thái Nguyên được tác giả xem xét theo tiêu chí định lượng (số lượng các ngôn ngữ, số lượng người nói mỗi thứ tiếng, số lượng phạm vi giao tiếp của mỗi ngôn ngữ, số lượng các ngôn ngữ nổi trội về chức năng); và theo tiêu chí định chất (nguồn gốc của các ngôn ngữ, tương quan về chức năng giữa các ngôn ngữ, tình hình sử dụng ngôn ngữ của người Tày, người Nùng, người Hoa).

- Bên cạnh mô tả tình hình sử dụng ngôn ngữ của người Mông, người Tày, người La Chí, tác giả luận văn “Cảnh huống ngôn ngữ ở Hà Giang” (Nguyễn Thị Thanh Huyền, 2010) còn đề cập phân tích những nhân tố nảy sinh hiện tượng đa ngữ, sự pha tạp ngôn ngữ...ở tỉnh Hà Giang.

- Luận án tiến sĩ “Nghiên cứu cảnh huống ngôn ngữ dân tộc Tày ở vùng đông Bắc Việt Nam” của Hà Thị Tuyết Nga (2014) đề xuất phương pháp đánh giá vị thế ngôn ngữ, cụ thể là vị thế của tiếng Tày vùng Đông Bắc trong bối cảnh hội nhập và phát triển ở Việt Nam. Sự chuyển giao ngôn ngữ giữa các thế hệ, sự tiếp nhận của cộng đồng trong các lĩnh vực mới được phân tích. Để xác định thực tế cảnh huống ngôn ngữ, tác giả luận án mô tả các yếu tố dân tộc-nhân khẩu, yếu tố xã hội, yếu tố văn hóa và yếu tố thể chế.

• Nguyễn Thị Thu Dung (Nguyễn Thị Thu Dung, 2015) nhận diện việc xác định năng lực ngôn ngữ theo các mối tương quan. Cụ thể, tác giả vận dụng xét năng lực tiếng Việt trong các mối tương quan như:

- Mối tương quan giữa năng lực tiếng Việt và giới tính.
- Mối tương quan giữa năng lực tiếng Việt và tuổi tác.
- Mối tương quan giữa năng lực tiếng Việt và trình độ học vấn.
- Mối tương quan giữa năng lực tiếng Việt và nghề nghiệp.
- Mối tương quan giữa năng lực tiếng Việt và tình hình kinh tế của gia đình.
- Mối tương quan giữa năng lực tiếng Việt và mức độ thường xuyên của việc đi ra khỏi làng.

Mỗi công trình hướng vào khai thác sâu các bình diện thuộc cảnh huống ngôn ngữ trên địa bàn một tỉnh hay một khu vực cụ thể.

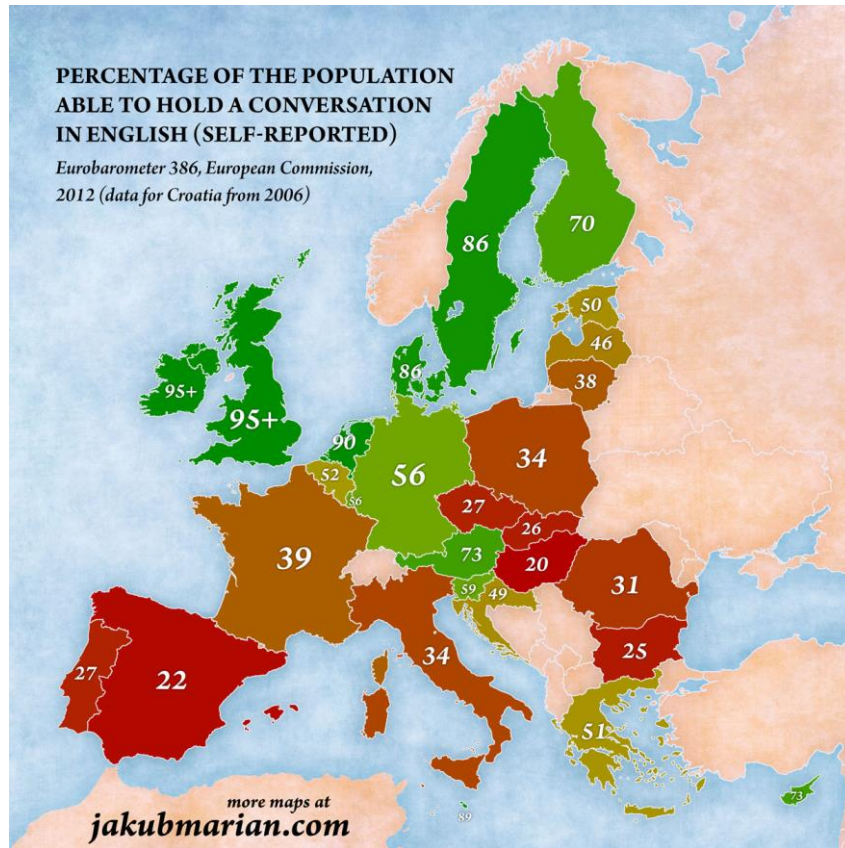
Trong bài *Những con số nói về ngôn ngữ trên thế giới* (VNN, 2015), nhà báo Thu Phương đã liệt kê một loạt các công trình nghiên cứu về số lượng các ngôn ngữ đang được sử dụng trên thế giới, phân chia theo từng châu lục. Châu Á là châu lục đang dẫn đầu về sự đa dạng ngôn ngữ với 2301 các ngôn ngữ khác nhau. Thứ tiếng được nhiều người bản ngữ sử dụng nhất, xếp theo thứ tự, là Trung, tiếng Hindi-Udru, tiếng Anh, tiếng Ả Rập, tiếng Tây Ban Nha... Tính toán theo chỉ số đa dạng ngôn ngữ Greenberg, khi đạt giá trị 100% thì có nghĩa là từng người trong nước sẽ có riêng ngôn ngữ mẹ đẻ. Ở Papua New Guinea, giá trị lên đến 98.8%. Còn ở Mỹ, tỷ lệ là 33%. Tiếng Anh được sử dụng rộng rãi trên 101 quốc gia, tiếng Ả Rập sử dụng ở 60 quốc gia, tiếng Pháp với 51 quốc gia sử dụng... Dự báo những ngôn ngữ có nguy cơ biến mất vào cuối thế kỷ thứ 21 tập trung ở các khu rừng Amazon, vùng châu Phi cận sa mạc Sahara, Châu Đại Dương, Úc và Đông Nam Á.

2.2 Tình hình nghiên cứu ngoài nước

Trên thế giới đã có những công trình công bố liên quan đến việc miêu tả cảnh huống ngôn ngữ bằng bản đồ ngôn ngữ số nhưng công trình về *Language and Space: Language Mapping* (Lameli, A.; Kehrein, R.; Rabanus, S. (Eds), 2010), *Language mapping in the Atlas* (Moseley, 2010).

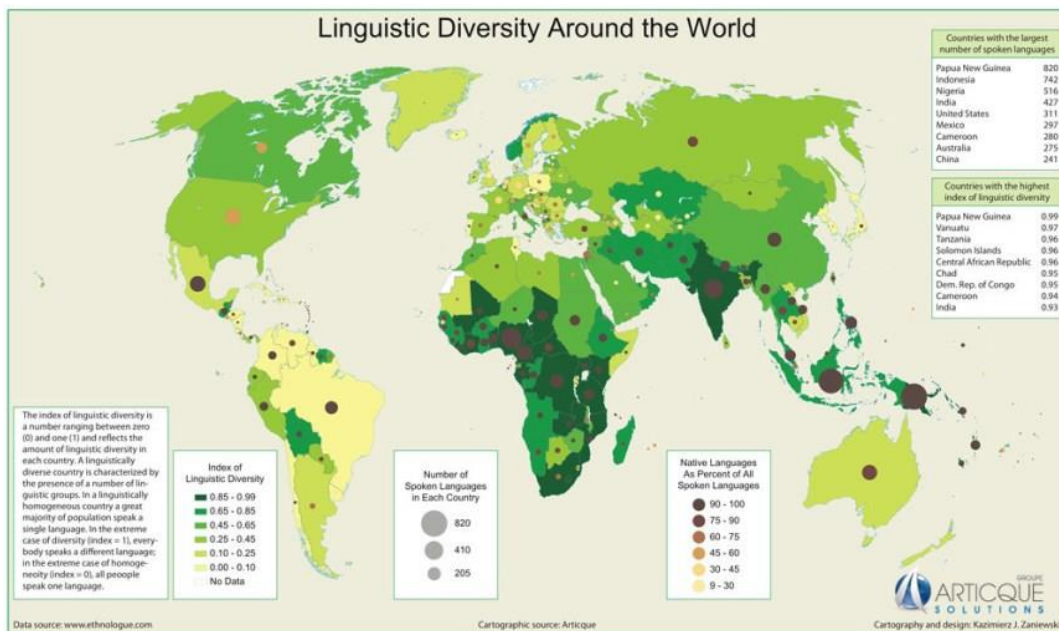
Ví dụ như bản đồ trong Hình 1 (Marian, 2011) sau đây cho thấy tỷ lệ phần trăm của những người có thể nói chuyện bằng tiếng Anh phân theo từng quốc gia trong khối cộng đồng châu Âu. Bản đồ này được hình thành dựa trên dữ liệu

trong công trình "Special Eurobarometer 386" (Eurobarometer, 2012). Các dữ liệu mang tính ước lượng, có nghĩa là người được phỏng vấn tạo thành một mẫu đại diện dân số, làm cơ sở để xác lập khả năng "nói được" các ngôn ngữ khác nhau dưới sự hướng dẫn của một người phỏng vấn. Khả năng thực tế của họ chưa được kiểm nghiệm.



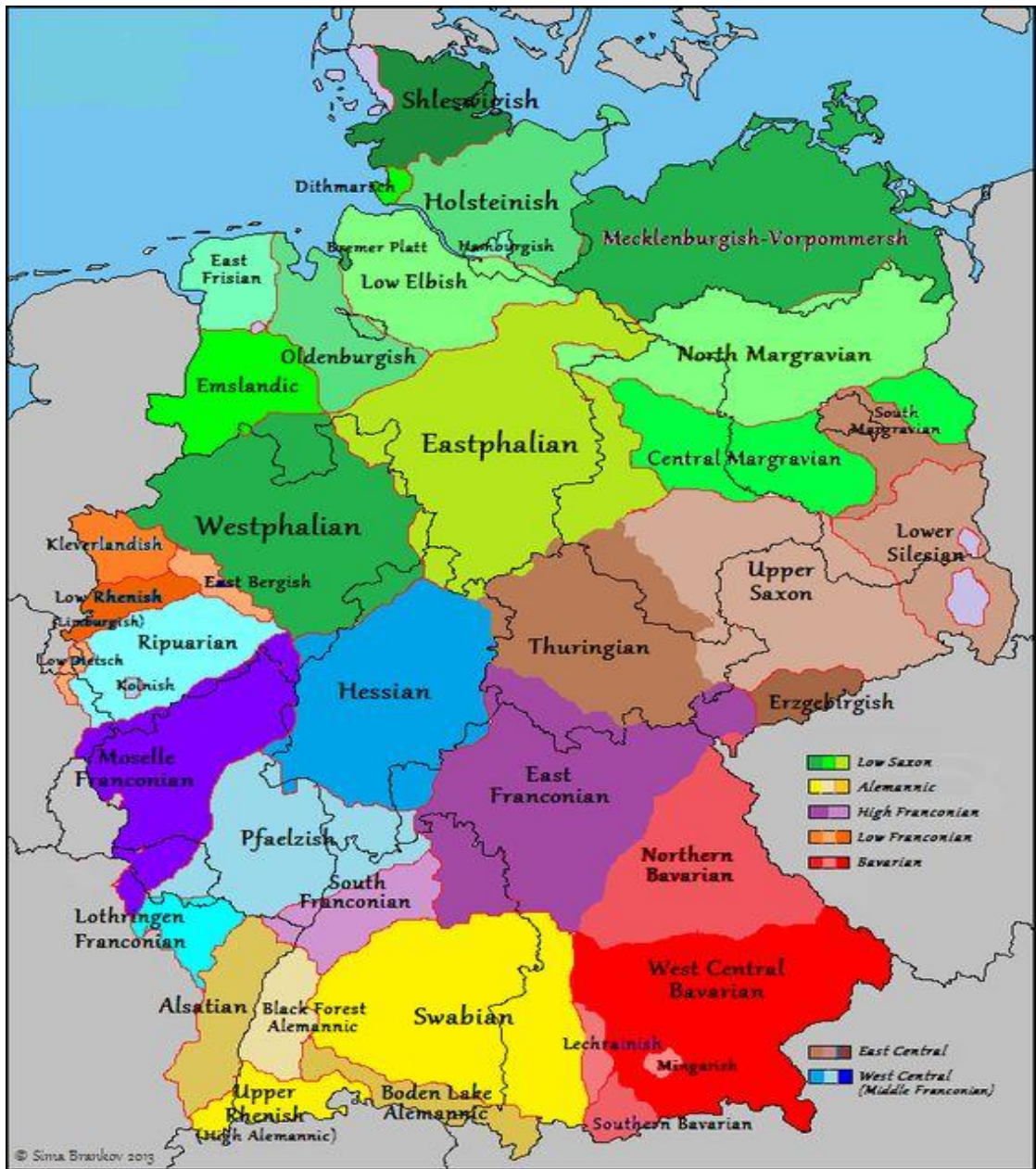
Hình 1. Bản đồ tỷ lệ những người có thể nói tiếng Anh ở châu Âu

Kazimierz Zaniewski (Đại học Wisconsin Oshkosh) tạo ra bản đồ trên Hình 2 (Zaniewski, 2011) miêu tả sự đa dạng ngôn ngữ của các nước trên thế giới bằng cách sử dụng dữ liệu của Ethnologue. Bản đồ này đã cho thấy mức độ đa dạng ngôn ngữ ở quốc gia khác nhau, trong đó số 0 có nghĩa là tất cả mọi người nói cùng một ngôn ngữ và số 1 có nghĩa là tất cả mọi người có ngôn ngữ riêng của họ. Kết quả nghiên cứu cho thấy nước Papua New Guinea có số 0,99, thể hiện mức độ đa dạng ngôn ngữ cao. Trong khi đó, Việt Nam trong khoảng 0.10-0.25, nêu lên mức độ đa dạng ngôn ngữ thấp. Điều này cho thấy tiếng Việt chiếm ưu thế vượt trội so với các ngôn ngữ đang có tại Việt Nam.



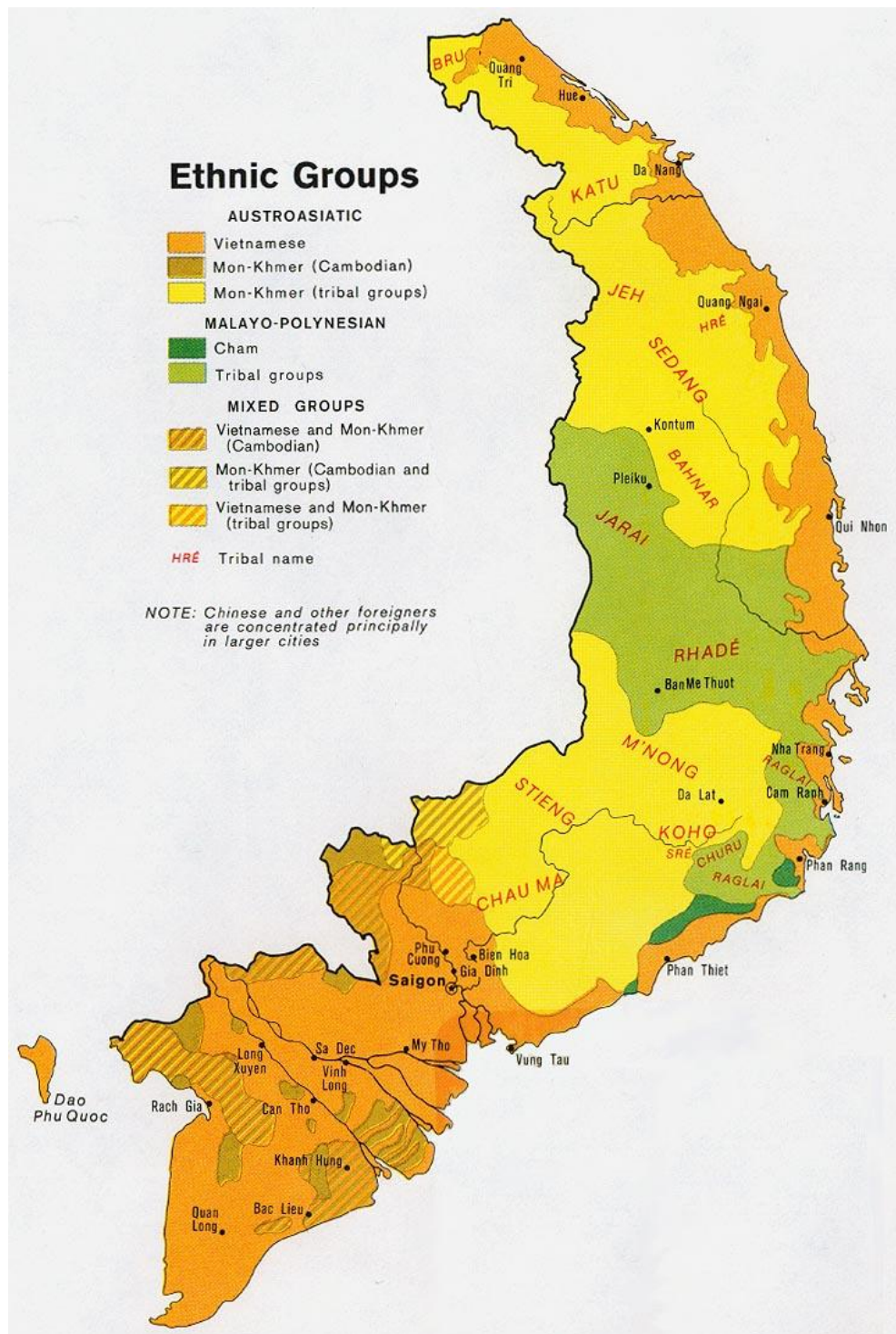
Hình 2. Bản đồ đa dạng ngôn ngữ của các nước trên thế giới

Ngoài ra, bản đồ ngôn ngữ còn được sử dụng để hiển thị sự phân bố các phương ngữ trong một địa phương hay quốc gia. Hình 3 là bản đồ phương ngữ của nước Đức (GermanAccentsDialects). Với 250 phương ngữ, người dân ở nước Đức đôi khi gặp phải trở ngại một khi ngôn từ địa phương không trùng lặp với ngôn từ toàn dân. Phương ngữ thường là ngôn ngữ thứ nhất để giao tiếp nói năng, chứ không phải là ngôn ngữ trong lớp học hoặc khi học đọc, học viết. Ngôn ngữ toàn dân được xem là ngôn ngữ chuẩn, thường được sử dụng trên báo, đài, sách vở. Tuy nhiên, thú vị là học đọc và học viết chữ theo ngôn ngữ toàn dân trong khi đó ngôn ngữ địa phương (phương ngữ) lại được dùng trong quá trình tư duy. Do đặc điểm mức độ phương ngữ nhiều nên bối cảnh ngôn ngữ ở nước Đức trở nên hấp dẫn đối với các nhà nghiên cứu ngôn ngữ, phương ngữ trên toàn thế giới.



Hình 3. Bản đồ các phương ngữ của nước Đức

Đối với Việt Nam, bản đồ ngôn ngữ dân tộc học đã được lập để thể hiện bối cảnh ngôn ngữ ở miền Nam Việt Nam vào năm 1972 (CLPP).



Hình 4. Bản đồ ngôn ngữ và dân tộc ở miền Nam Việt Nam (năm 1972)

Nhìn chung, hầu hết các bản đồ hiện có đều tập trung mô tả trạng thái ngôn ngữ trên diện rộng (quốc gia hoặc khu vực), chi tiết về bối cảnh ngôn ngữ ở mức tỉnh, huyện hay xã chưa thấy được đề cập. Mặt khác, bản đồ về ngôn ngữ ở Việt Nam còn hạn chế về số lượng và chủ yếu khai thác các bản đồ giấy, do đó đã hạn chế khá nhiều đến sự thuận lợi khi cập nhật hoặc sử dụng cho đông đảo mọi đối tượng quan tâm.

Những nghiên cứu về ngôn ngữ góp phần xác định sự lớn mạnh của tộc người trong các quốc gia. Ở Việt Nam, các nhà nghiên cứu tập trung miêu tả trạng thái ngôn ngữ ở từng địa phương. Mặc dù vậy, các dữ liệu nghiên cứu về tình hình sử dụng ngôn ngữ, năng lực ngôn ngữ và thái độ đối với ngôn ngữ của người dân, sự đa dạng ngôn ngữ hay mức độ song ngữ... chưa được nhiều người tiếp cận hoặc khó cho mọi tầng lớp trong xã hội tiếp cận. Để tạo sự thuận lợi cho việc nhận diện trạng thái ngôn ngữ ở từng vùng địa phương, ý tưởng về việc hiển thị bằng bản đồ ngôn ngữ dân tộc tương tác trực tuyến được tác giả nghiên cứu đề tài phát triển. Điều này vừa thể hiện sự tích hợp công nghệ trong nghiên cứu ngôn ngữ, vừa mở ra một hướng tiếp cận mới cho các nhà nghiên cứu ngôn ngữ trong nước.

Để nhận diện cảnh huống ngôn ngữ ở tỉnh Trà Vinh, một số đặc điểm cần chú ý như sau:

- Cảnh huống ngôn ngữ là một khái niệm quan trọng của ngôn ngữ học xã hội và có hiệu quả trực tiếp đối với chính sách ngôn ngữ. Cảnh huống ngôn ngữ của một quốc gia hay một khu vực luôn chiếm một vị trí hết sức quan trọng trong việc xác định địa vị ngôn ngữ. Chức năng xã hội của ngôn ngữ có thể được chính phủ điều tiết bằng hoạt động kế hoạch hóa ngôn ngữ. Việc làm này tập trung vào các nội dung như: lựa chọn một mã ngôn ngữ mới để phục vụ cho một mục đích cụ thể; phân bổ các chức năng khác nhau cho một ngôn ngữ hay nhiều ngôn ngữ; chỉnh đốn cách sử dụng những ngôn ngữ trong một cộng đồng. Trong địa bàn tỉnh Trà Vinh, chức năng xã hội của từng ngôn ngữ như Việt, Khmer cần được chú trọng tìm hiểu. Người dân Khmer có thể thông thạo (nghe/nói) cả hai ngôn ngữ (Việt, Khmer) nhưng lại không thể đọc/viết tiếng mẹ đẻ (tiếng Khmer) hoặc tiếng Việt. Tính chất song ngữ mất cân bằng khá cao này đã dẫn đến sự đa dạng và phức tạp của cảnh huống ngôn ngữ tại nơi đây. Trần Phương Nguyên chú ý hai hiện tượng. Thứ nhất, nếu đối tượng sử dụng ngôn ngữ mặc dù tương đối đồng nhưng sống rải rác phân bố xa nhau thì sẽ làm giảm khả năng hành chức của ngôn ngữ vì thiếu môi trường giao tiếp, khó tổ chức các hoạt động giáo dục bằng ngôn ngữ, khó phát hành các ấn phẩm, như vậy có thể dẫn đến sự chuyển hóa từ việc sử dụng tiếng mẹ đẻ sang sử dụng các ngôn ngữ khác của các cộng đồng xung quanh. Thứ hai, không có sự phụ thuộc trực tiếp giữa sức mạnh về dân số và sức mạnh về chức năng của ngôn ngữ. Đây là những yếu tố cần thiết khi cân nhắc ước tính số lượng ngôn ngữ được từng hộ dân chọn sử dụng.

3 Mục tiêu

Mục tiêu chung: Hình thành công thông tin bản đồ ngôn ngữ dân tộc học trực tuyến ứng dụng trên một địa bàn cụ thể (tỉnh Trà Vinh)

- **Mục tiêu cụ thể 1:** Khảo sát, thu thập dữ liệu về dân tộc, trạng thái các ngôn ngữ đang được sử dụng.
- **Mục tiêu cụ thể 2:** Lập biểu đồ thống kê về sự phân bố dân tộc, ngôn ngữ trên từng địa phương thuộc tỉnh Trà Vinh (từ cấp xã/phường đến cấp huyện/thị/thành phố).
- **Mục tiêu cụ thể 3:** Thiết kế công thông tin bản đồ ngôn ngữ dân tộc học (bản đồ phân bố dân tộc, bản đồ tỷ lệ các ngôn ngữ được sử dụng, bản đồ sự đa dạng ngôn ngữ theo các địa phương trong tỉnh Trà Vinh).
- **Mục tiêu cụ thể 4:** Phân tích ứng dụng của công thông tin bản đồ ngôn ngữ dân tộc học trong việc đánh giá cảnh huống ngôn ngữ tại một địa phương.

4 Đối tượng, phạm vi và phương pháp nghiên cứu

4.1 Đối tượng và địa điểm nghiên cứu

Nghiên cứu tập trung vào tỉnh Trà Vinh, phân bố dân cư (theo dân tộc, theo địa phương – cấp xã). Nội dung chủ yếu mô tả sự phân bố dân tộc, ngôn ngữ hành chức theo mức đơn vị hành chính cấp xã.

4.2 Nội dung thực hiện

Nội dung 1. Thu thập dữ liệu và mô phỏng dữ liệu

Nội dung 2. Nghiên cứu các chỉ số về ngôn ngữ và dân tộc học

Nội dung 3. Lập công thông tin bản đồ về ngôn ngữ dân tộc học

Nội dung 4. Phân tích ứng dụng của công thông tin bản đồ ngôn ngữ học trực tuyến trong đánh giá cảnh huống ngôn ngữ

4.3 Phương pháp nghiên cứu, kỹ thuật sử dụng

Để thực hiện đề tài, tác giả áp dụng các phương pháp nghiên cứu sau đây:

- Phương pháp nghiên cứu tư liệu: trên cơ sở kế thừa các công trình đã thực hiện trong và ngoài nước về các vấn đề có liên quan đến đề tài như ngôn ngữ học như ngôn ngữ học xã hội, cảnh huống ngôn ngữ, thái độ đối với ngôn ngữ, chính sách ngôn ngữ, sự phân bố ngôn ngữ; các công trình về đặc điểm văn hóa,

xã hội vùng đồng bằng sông Cửu Long; và những công trình về việc thiết lập bản đồ hiển thị đặc điểm ngôn ngữ.

- Phương pháp điều tra xã hội học (khảo sát thực tế bằng kỹ thuật phỏng vấn, trò chuyện, quan sát) nhằm thu thập dữ liệu về dân tộc; số lượng người sử dụng ngôn ngữ, năng lực sử dụng ngôn ngữ cụ thể trong từng hộ dân. Mẫu chọn khảo sát là 10% tổng số hộ trong 2 phường (phường 8 và phường 2). Phường 8 đại diện cho khu vực có mật độ dân cư Khmer đông và sống liền kề (nông thôn); phường 2 đại diện cho khu vực có mật độ dân cư Khmer thưa thớt, rải rác (thành thị).

- Phương pháp mô phỏng dữ liệu: việc thu thập dữ liệu ngôn ngữ trên toàn tỉnh đòi hỏi kinh phí rất cao, vì vậy việc mô phỏng dữ liệu ngôn ngữ tạm thời được sử dụng. Phương pháp nội suy được sử dụng để mô phỏng dữ liệu ngôn ngữ.

- Kỹ thuật làm bản đồ trực tuyến bằng các dịch vụ theo chuẩn quốc tế Open Geospatial Consortium (OGC): các chuẩn Web Map Service (WMS) và Web Feature Service (WFS) được sử dụng để tạo ra các dịch vụ trực tuyến làm ra các bản đồ đa dạng hóa ngôn ngữ, bản đồ về dân tộc và phân bố dân tộc, bản đồ về phân bố ngôn ngữ chi tiết. Các dịch vụ này không những sẽ được sử dụng trong cổng thông tin bản đồ ngôn ngữ dân tộc học này mà còn có thể cung cấp để các trang Web của các sở ban ngành có thể gọi để hiển thị trên đó (giống như gọi bản đồ của Google vào để hiển trên các trang Web của công ty, trường học). Ngoài ra kỹ thuật tạo màu cho bản đồ theo chuẩn Styled Layer Descriptor (SLD) cũng được sử dụng.

PHẦN NỘI DUNG

CHƯƠNG 1. THU THẬP VÀ SỐ HÓA DỮ LIỆU

1.1 Thu thập dữ liệu

1.1.1 Giới thiệu về dữ liệu cần thiết cho đề tài

Các loại dữ liệu cần thiết để lập bản đồ ngôn ngữ và dân tộc:

- Dữ liệu địa lý để vẽ đường biên của địa phương các cấp
- Dữ liệu ngôn ngữ và dân tộc

Dữ liệu địa lý có thể được thu thập từ bản đồ địa chính của tỉnh. Ngoài ra có một số bản đồ địa chính của cả nước được sưu tập từ nguồn dữ liệu mở để thể hiện đường biên giữa các tỉnh của cả nước.

Dữ liệu ngôn ngữ và dân tộc cần được thu thập ở từng hộ hay từng cá nhân trong từng địa phương cấp thấp nhất có thể (ví dụ cấp xã/phường). Khi có dữ liệu ở mức thấp nhất, công thông tin sẽ tự động tổng hợp dữ liệu cho các cấp còn lại.

Do việc thu thập dữ liệu từng hộ hay cá nhân trên toàn tỉnh đòi hỏi chi phí rất cao nên phạm vi về thu thập dữ liệu được thay đổi bằng cách thu thập dữ liệu có sẵn về dân tộc (Bảng 4 là số liệu về dân tộc do Ban Dân tộc tỉnh Trà Vinh cung cấp) và một số mẫu dữ liệu để tiến hành mô phỏng các vùng chưa có dữ liệu về ngôn ngữ.

Tiếp theo trong chương này là các thảo luận về thu thập dữ liệu về ngôn ngữ và dân tộc (Mục 1.1.2), Việc tiến hành thu thập mẫu dữ liệu ngôn ngữ và dân tộc (Mục 1.1.3), và việc mô phỏng dữ liệu về ngôn ngữ (Mục 1.1.4).

1.1.2 Phương pháp thu nhập dữ liệu ngôn ngữ và dân tộc

Thông thường, để đánh giá tương đối chính xác số liệu thống kê về vấn đề nào đó trong một tỉnh thì nên tiến hành điều tra trong toàn tỉnh, chẳng hạn như điều tra dân số. Các tính toán để lập bản đồ về ngôn ngữ và dân tộc cho các quốc gia trên thế giới thường bắt đầu bằng việc thu thập dữ liệu ngẫu nhiên trên từng cá thể và chỉ xem xét tiếng mẹ đẻ của từng cá thể. Trong thực tế, việc tiến hành điều tra từng cá thể sẽ mất nhiều thời gian và công sức. Bên cạnh đó, sự lựa chọn sử dụng ngôn ngữ trong đời sống là xuất phát từ nhu cầu và ý nguyện của cá nhân nên dữ liệu về số lượng người sử dụng ngôn ngữ luôn thay đổi. Một số

người Khmer đã chọn hẩu tiếng Việt cho con cái của họ khi vừa được sinh ra, mặc dù vẫn giữ nguyên thành phần dân tộc là Khmer. Vì vậy một giải pháp khác đề nghị ở đây để thu thập dữ liệu về dân tộc và ngôn ngữ là thu thập dữ liệu ở mức hộ. Cơ sở để tác giả đề nghị khảo sát mức hộ dân là dựa vào cách thức mà Cục Khảo sát Cộng đồng Hoa Kỳ (American Community Survey) thực hiện để tính toán số lượng người sử dụng ngôn ngữ hiện nay tại Mỹ (Ryan, 2013). Chính vì vậy, việc đề xuất phiếu thu thập dữ liệu, xác định số hộ, chọn lựa các hộ để khảo sát là việc làm quan trọng, cung cấp tư liệu thực tế để phân tích bối cảnh ngôn ngữ và lập các biểu bảng thống kê.

Trong tương lai, để triển khai cập nhật số liệu về ngôn ngữ sử dụng trên toàn tỉnh, thông tin về ngôn ngữ và dân tộc nên được kết hợp cùng với đợt điều tra dân số. Các giải pháp thu thập dữ liệu về ngôn ngữ và dân tộc như sau:

- Giải pháp dài hạn: thu thập dữ liệu cùng với đợt điều tra dân số để giảm chi phí, đồng thời dữ liệu ở các đợt điều tra khác nhau sẽ được sử dụng để xác định sự biến động về ngôn ngữ và dân tộc trong từng địa phương các cấp.

- Giải pháp ngắn hạn trong phạm vi đề tài: tiến hành thu thập mẫu dữ liệu và sử dụng mẫu này để ước tính các số liệu về ngôn ngữ và dân tộc. Ngoài ra dữ liệu mẫu này còn được sử dụng để kiểm chứng và đánh giá độ tin cậy của các ước lượng được đưa ra.

1.1.3 Thu thập mẫu dữ liệu ngôn ngữ và dân tộc

1.1.3.1 Cách thức thực hiện

a) Địa điểm

Việc khảo sát được tiến hành tại hai phường thuộc thành phố Trà Vinh, phường 2 và phường 8.

Đặc điểm của phường 2: nằm tại trung tâm thành phố Trà Vinh, có 22,6% hộ dân tộc thiểu số.

Đặc điểm của phường 8: thuộc khu vực ngoại ô thành phố, có tỷ lệ người Khmer chiếm trên 70%.

Mục đích của việc chọn hai địa bàn khảo sát này là tìm ra sự lựa chọn ngôn ngữ sử dụng hàng ngày của người dân ở hai khu vực: thành thị và ven ô (hay có thể tạm gọi là nông thôn); và nơi có đông đồng bào dân tộc thiểu số và nơi có ít

đồng bào dân tộc thiểu số. Thông thường, ngôn ngữ của nhóm đông sẽ chiếm ưu thế.

Ở mỗi phường, tỷ lệ số hộ được chọn là 10% trong tổng hộ dân tại mỗi phường (dựa theo tỷ lệ mà Cục Khảo sát Hoa Kỳ - US Census Bureau sử dụng). Cụ thể như sau: tổng số có 926 hộ ở phường 2, nên chọn 100 hộ để khảo sát; phường 8 có 2159 hộ, nên chọn 300 hộ để khảo sát. Các hộ đều nằm liền kề, trong cùng một địa bàn. Đặc biệt, ở phường 8, việc khảo sát thực hiện ở khóm 8, nơi có đông người Khmer sinh sống.

b) Phương pháp tiến hành

Phiếu thu thập dữ liệu về ngôn ngữ và dân tộc ở mức hộ được đề xuất như trong Bảng 3 của phần Phụ lục 1. Mỗi hộ thuộc một dân tộc và mỗi hộ có thể sử dụng nhiều ngôn ngữ như tiếng Việt, tiếng Khmer với mức độ thông thạo khác nhau như chỉ biết nghe/nói (A), biết nghe/nói và đọc/viết (C). Ngoài ra, họ tên, giới tính và số người trong hộ cũng cần được thu thập để phục vụ cho các nghiên cứu về xã hội học dân tộc của Khoa Ngôn ngữ - Văn hóa - Nghệ thuật Khmer Nam bộ nói riêng và trường Đại học Trà Vinh sau này.

Mỗi hộ được tiếp cận trực tiếp và cung cấp thông tin để điều tra viên điền vào phiếu thu thập dữ liệu. Khi có từ 50% số thành viên trong hộ sử dụng cùng một ngôn ngữ (ví dụ như tiếng Việt) trong hộ thì hộ đó được xem là hộ biết nghe nói tiếng Việt. Tương tự như vậy khi xác định năng lực ngôn ngữ. Từ 50% thành viên trong hộ viết và đọc được tiếng Khmer thì hộ đó được xem là biết đọc viết chữ Khmer.

1.1.3.2 Kết quả thu được

Tại phường 8, thành phố Trà Vinh, nhóm điều tra đã đến gặp lần lượt 294 hộ, trong đó hộ người Khmer có 255 hộ, 37 hộ người Kinh và chỉ có 2 hộ người Hoa. Đặc điểm của các hộ này là sống liền kề, chung trong khóm (khóm 8). Mức độ sử dụng tiếng Việt của người Khmer thấp hơn so với tiếng Khmer (669 người nói được tiếng Khmer, trong khi chỉ có 175 người nói được tiếng Việt). Bên cạnh đó, số lượng người biết chữ Khmer cũng khá cao (365 người). Người Kinh cũng biết tiếng Khmer để thuận lợi trong giao tiếp. Có 7 người Kinh biết chữ Khmer. 17 người Kinh sử dụng tiếng Khmer để giao tiếp thông thường hàng ngày.

Bảng 1. Tổng kết mẫu dữ liệu về ngôn ngữ và dân tộc ở khóm 8, phường 8, thành phố Trà Vinh

Dân tộc	Số hộ thu thập	Số người	Tiếng Việt		Tiếng Khmer		Tiếng Hoa	
			chỉ biết nghe/nói	đọc/viết/ nghe/nói	chỉ biết nghe/nói	đọc/viết/ nghe/nói	chỉ biết nghe/nói	đọc/viết/ nghe/nói
Kinh	37	156	15	141	17	7	0	0
Khmer	255	1057	175	584	669	365	0	0
Hoa	2	15		15	0	0	0	0
Tổng số	294	1228	190	740	686	372	0	0

Dân tộc	Số hộ thu thập	Số người	Tiếng Việt		Tiếng Khmer		Tiếng Hoa	
			A	C	A	C	A	C
Kinh	37	156	10%	90%	11%	4%	0%	0%
Khmer	255	1057	17%	55%	63%	35%	0%	0%
Hoa	2	15	0%	100%	0%	0%	0%	0%

Tại phường 2, thành phố Trà Vinh, khu vực được chọn khảo sát là khu dân cư, sống tập trung trong cùng khu vực, gần chợ và trung tâm thành phố. Số lượng người Khmer ở phường 2 khá ít. Nhóm khảo sát đã thực hiện tại 5 tổ thuộc khóm 1, phường 2. Tổng số hộ khảo sát là 134 hộ, trong đó có 3 hộ người Khmer và 2 hộ người Hoa. Hầu hết người dân ở đây (kể cả Kinh, Khmer, Hoa) đều sử dụng tiếng Việt. Khả năng biết đọc và viết chữ Việt của người dân ở đây rất cao, 10/13 người Khmer biết đọc và viết tiếng Việt, trong khi đó, không có người nào đọc viết được chữ Khmer.

Bảng 2. Tổng kết mẫu dữ liệu về ngôn ngữ và dân tộc ở khóm 1, phường 2, thành phố Trà Vinh

Dân tộc	Số hộ thu thập	Số người	Tiếng Việt		Tiếng Khmer		Tiếng Hoa	
			chỉ biết nghe/nói	đọc/viết/ nghe/nói	chỉ biết nghe/nói	đọc/viết/ nghe/nói	chỉ biết nghe/nói	đọc/viết/ nghe/nói
Kinh	129	575	49	526	0	0	0	0
Khmer	3	13	3	10	2	0	0	0
Hoa	2	7	3	4	0	0	0	0
Tổng số	134	595	55	540	2	0	0	0

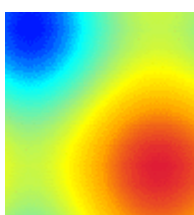
Dân tộc	Số hộ thu thập	Số người	Tiếng Việt		Tiếng Khmer		Tiếng Hoa	
			A	C	A	C	A	C
Kinh	129	575	9%	91%	0%	0%	0%	0%
Khmer	3	13	23%	77%	15%	0%	0%	0%
Hoa	2	7	43%	57%	0%	0%	0%	0%

1.1.4 Mô phỏng dữ liệu ngôn ngữ

1.1.4.1 Phương pháp

Việc mô phỏng dữ liệu được tiến hành sau khi có mẫu số liệu về ngôn ngữ trên từng dân tộc và số liệu về dân tộc trên toàn tỉnh. Phương pháp đa nội suy (multivariate interpolation) được áp dụng để mô phỏng dữ liệu về ngôn ngữ với các yếu tố đầu vào đa chiều bao gồm dữ liệu về dân tộc, mật độ trên vùng địa lý và các tỷ lệ ngôn ngữ trên từng dân tộc được thu thập ở Phụ Lục 2 của chương này và yếu tố địa lý.

Hình dưới đây được trích dẫn để diễn tả một phần ý nghĩa của nội suy đa chiều.



Ví dụ màu đỏ là khu vực đông người Kinh và tỷ lệ sử dụng tiếng Việt rất cao, vùng màu xanh là khu vực đông người Khmer và tỷ lệ sử dụng tiếng Khmer cao, đồng thời còn có tỷ lệ các ngôn ngữ của từng dân tộc trong 2 khu vực này. Các khu vực khác thì đã biết tỷ lệ dân tộc (thu thập từ ban dân tộc tỉnh) nhưng chưa được biết tỷ lệ ngôn ngữ. Tỷ lệ sử dụng các loại ngôn ngữ của các vùng chưa biết sẽ được nội suy từ số liệu dân tộc và tỷ lệ ngôn ngữ của khu vực xanh đậm và khu vực đỏ đậm đã biết.

1.1.4.2 Độ tin cậy của dữ liệu mô phỏng

So sánh tỷ lệ các ngôn ngữ trên từng dân tộc của mẫu dữ liệu thu thập ở một khu vực của phường 8 và một khu vực của phường 2 thì thấy rằng các tỷ lệ này tương đối khớp với tỷ lệ của dữ liệu mô phỏng cho cả phường ở các phường này. Để mô phỏng chính xác sự phân bố dân tộc và ngôn ngữ thì cần nghiên cứu chuyên sâu riêng biệt trong lĩnh vực Toán ứng dụng và địa lý ngôn ngữ, và cần sự đóng góp của các nhà nghiên cứu trong lĩnh vực này. Trong phạm vi đề tài về ngôn ngữ, việc đánh giá độ tin cậy chỉ dừng lại ở mức so sánh số liệu từ mẫu dữ liệu thu thập được và số liệu từ kết quả mô phỏng. Tuy nhiên, việc thu thập dữ liệu về ngôn ngữ và dân tộc nên được tiến hành trên toàn Tỉnh cùng đợt với điều tra dân số để có dữ liệu chính xác hơn.

1.2 Số hóa bản đồ ngôn ngữ dân tộc học

1.2.1 Các chỉ số của bản đồ ngôn ngữ

1.2.1.1 Chỉ số đa dạng hóa ngôn ngữ (Linguistic Diversity Index)

Chỉ số đa dạng hóa ngôn ngữ được Joseph H. Greenberg giới thiệu vào năm 1956 trên Journal Language (Vol. 32) (Greenberg, 1956). Từ đó chỉ số đa dạng hóa được áp dụng rộng rãi. UNESCO thành lập bản đồ trực tuyến và biểu bảng thống kê về sự đa dạng hóa ngôn ngữ của tất cả các nước trên thế giới (Simons, Gary F. and Charles D. Fennig (eds.), 2017).

Công thức tính toán chỉ số đa dạng hóa ngôn ngữ khá đơn giản:

$$1 - \left(\frac{L_1}{P}\right)^2 - \left(\frac{L_2}{P}\right)^2 - \dots - \left(\frac{L_n}{P}\right)^2 \quad \text{hoặc} \quad 1 - \frac{L_1^2 + L_2^2 + \dots + L_n^2}{P^2}$$

Trong đó P là tổng số dân trong địa phương đó. n là số loại ngôn ngữ sử dụng trong địa phương. $L_1, L_2, \dots, L_i, \dots, L_n$ là số người sử dụng ngôn ngữ i .

1.2.1.2 Chỉ số tính toán số lượng hộ sử dụng ngôn ngữ

Thông thường bản đồ phân bố ngôn ngữ được tính trên số lượng người sử dụng tiếng mẹ đẻ (chỉ một loại tiếng mẹ đẻ cho mỗi người). Tuy nhiên, ở Trà Vinh có nhiều trường hợp mỗi người sử dụng tiếng Khmer và cả tiếng Việt như tiếng mẹ đẻ. Ngoài ra còn phải tính đến mức độ sử dụng trong việc thiết lập bản đồ ngôn ngữ. Vì vậy có nhiều cách số hóa khác nhau để lập bản đồ ngôn ngữ. Một ví dụ cụ thể dưới đây nhằm để thảo luận chi tiết hơn về vấn đề này:

Một địa phương có 10 hộ với các chi tiết về ngôn ngữ như sau:

- 7 hộ dân tộc Kinh trong đó có 5 hộ chỉ sử dụng tiếng Việt, 2 hộ vừa sử dụng tiếng Việt vừa sử dụng tiếng Khmer.
- 3 hộ dân tộc Khmer trong đó có 1 hộ chỉ sử dụng tiếng Khmer còn 2 hộ sử dụng tiếng Khmer và tiếng Việt.

7 hộ dân tộc Kinh	5 hộ chỉ sử dụng tiếng Việt
	2 hộ vừa sử dụng tiếng Việt vừa sử dụng tiếng Khmer
3 hộ dân tộc Khmer	1 hộ chỉ sử dụng tiếng Khmer
	2 hộ sử dụng tiếng Khmer và tiếng Việt

Như vậy tổng số hộ là 10, nhưng số lượt hộ sử dụng ngôn ngữ là 14 (bao gồm 9 hộ sử dụng tiếng Việt và 5 hộ sử dụng tiếng Khmer).

Ngôn ngữ chiếm đa số của địa phương là ngôn ngữ có số hộ sử dụng nhiều nhất. Từ đó, bản đồ ngôn ngữ chiếm đa số là sự thể hiện của ngôn ngữ chiếm đa số của từng địa phương.

Tuy nhiên, bản đồ sự phân bố (phần trăm) của một loại ngôn ngữ có khác đi vì lý do mỗi hộ có thể sử dụng nhiều ngôn ngữ.

Địa phương có số hộ là P , có n ngôn ngữ được sử dụng với số lượng lần lượt là $L_1, L_2, \dots, L_i, \dots, L_n$. $L_1 + L_2 + \dots + L_n$ có thể lớn hơn P .

- Cách tính 1: Phần trăm số hộ sử dụng ngôn ngữ L_i trong địa phương là số hộ sử dụng ngôn ngữ L_i chia tổng số hộ P .

Trong ví dụ trên thì phân bố số hộ sử dụng tiếng Việt trong địa phương là $9/10 \cdot 100$ (%); phân bố số hộ sử dụng tiếng Khmer trong địa phương là $5/10 \cdot 100$ (%)

- Cách tính 2: Phần trăm số hộ sử dụng ngôn ngữ L_i trong địa phương là số hộ sử dụng ngôn ngữ L_i chia tổng số lượt hộ sử dụng ngôn ngữ ($L_1 + L_2 + \dots + L_n$).

Trong ví dụ trên thì phân bố số hộ sử dụng tiếng Việt trong địa phương là $9/14 \cdot 100$ (%); phân bố số hộ sử dụng tiếng Khmer trong địa phương là $5/14 \cdot 100$ (%).

a) Chỉ số mức độ sử dụng ngôn ngữ trong địa phương

Chỉ số mức độ sử dụng ngôn ngữ trong địa phương là chỉ số được đề cập trong cách tính 1 của ví dụ trên.

Địa phương có số hộ là P , có n ngôn ngữ được sử dụng với số lượng lần lượt là $L_1, L_2, \dots, L_i, \dots, L_n$. $L_1 + L_2 + \dots + L_n$ có thể lớn hơn P .

Chỉ số mức độ sử dụng ngôn ngữ L_i trong địa phương là $\frac{L_i}{P}$

b) Tỷ lệ các ngôn ngữ được sử dụng trong địa phương

Tỷ lệ các ngôn ngữ được sử dụng trong địa phương là chỉ số được đề cập trong cách tính 2 của ví dụ trên.

Địa phương có số hộ là P , có n ngôn ngữ được sử dụng với số lượng lần lượt là $L_1, L_2, \dots, L_i, \dots, L_n$. $L_1 + L_2 + \dots + L_n$ có thể lớn hơn P .

Tỷ lệ các ngôn ngữ L_i được sử dụng trong địa phương là $\frac{L_i}{L_1 + L_2 + \dots + L_n}$

1.2.1.3 Ngôn ngữ chiếm đa số của địa phương

Ngôn ngữ chiếm đa số của địa phương cũng khá đơn giản, là ngôn ngữ có số hộ sử dụng nhiều nhất. Từ đó, bản đồ ngôn ngữ chiếm đa số là sự thể hiện của ngôn ngữ chiếm đa số của từng địa phương.

1.2.2 Các chỉ số về dân tộc

Địa phương có số hộ là P, có m dân tộc với số lượng lần lượt là E_1, E_2, \dots, E_m .
 $E_1 + E_2 + \dots + E_m = P$.

1.2.2.1 Dân tộc chiếm đa số trong địa phương

Dân tộc chiếm đa số trong địa phương là dân tộc có số hộ nhiều nhất trong địa phương.

1.2.2.2 Tỷ lệ phân bố của dân tộc

Tỷ lệ phân bố của dân tộc E_i trong địa phương là $\frac{E_i}{P}$

1.2.3 Các chỉ số về ngôn ngữ trên từng dân tộc

Chỉ số ngôn ngữ trên từng dân tộc là tỷ lệ phân bố của mỗi loại ngôn ngữ trên từng dân tộc.

Chẳng hạn một địa phương có số hộ là P, có m dân tộc với số lượng lần lượt là E_1, E_2, \dots, E_m , và có n ngôn ngữ được sử dụng $L_1, L_2, \dots, L_i, \dots, L_n$ cùng với số lượng sử dụng của mỗi loại ngôn ngữ trong từng dân tộc được ký hiệu là $L_i E_j$ ($L_i E_j$ là ký hiệu của một con số để nói về số lượng sử dụng ngôn ngữ i của dân tộc j). Tỷ lệ phân bố của loại ngôn ngữ i trên dân tộc j là $L_i E_j / E_j$.

Như vậy, nếu tỉnh Trà Vinh có 3 dân tộc Kinh, Khmer và Hoa và các ngôn ngữ sử dụng trong gia đình là Việt, Khmer và Hoa thì sẽ có 9 chỉ số về ngôn ngữ trên từng dân tộc lần lượt là:

- Tỷ lệ sử dụng tiếng Việt của người Kinh
- Tỷ lệ sử dụng tiếng Việt của người Khmer
- Tỷ lệ sử dụng tiếng Việt của người Hoa
- Tỷ lệ sử dụng tiếng Khmer của người Kinh
- Tỷ lệ sử dụng tiếng Khmer của người Khmer
- Tỷ lệ sử dụng tiếng Khmer của người Hoa

- Tỷ lệ sử dụng tiếng Hoa của người Kinh
- Tỷ lệ sử dụng tiếng Hoa của người Khmer
- Tỷ lệ sử dụng tiếng Hoa của người Hoa

Nếu xét thêm mức độ thông thạo của từng ngôn ngữ trên mỗi dân tộc thì sẽ có 18 chỉ số về ngôn ngữ trên từng dân tộc với mức độ nghe/nói (A) hay nghe/nói/đọc/viết (C).

CHƯƠNG 2. PHÁT TRIỂN CÔNG THÔNG TIN BẢN ĐỒ NGÔN NGỮ DÂN TỘC HỌC

2.1 Cơ sở dữ liệu địa lý, ngôn ngữ và dân tộc

2.1.1 Sự cần thiết của cơ sở dữ liệu và chọn hệ quản trị cơ sở dữ liệu

Dữ liệu về ngôn ngữ và dân tộc được thu thập theo từng hộ trong từng địa phương. Mỗi địa phương có đường ranh giới với tọa độ địa lý để vẽ lên đường ranh giới. Để lưu trữ thông tin ranh giới của từng địa phương và tạo thuận lợi cho việc lập bản đồ ngôn ngữ và dân tộc, hệ quản trị cơ sở dữ liệu về không gian sẽ rất cần thiết.

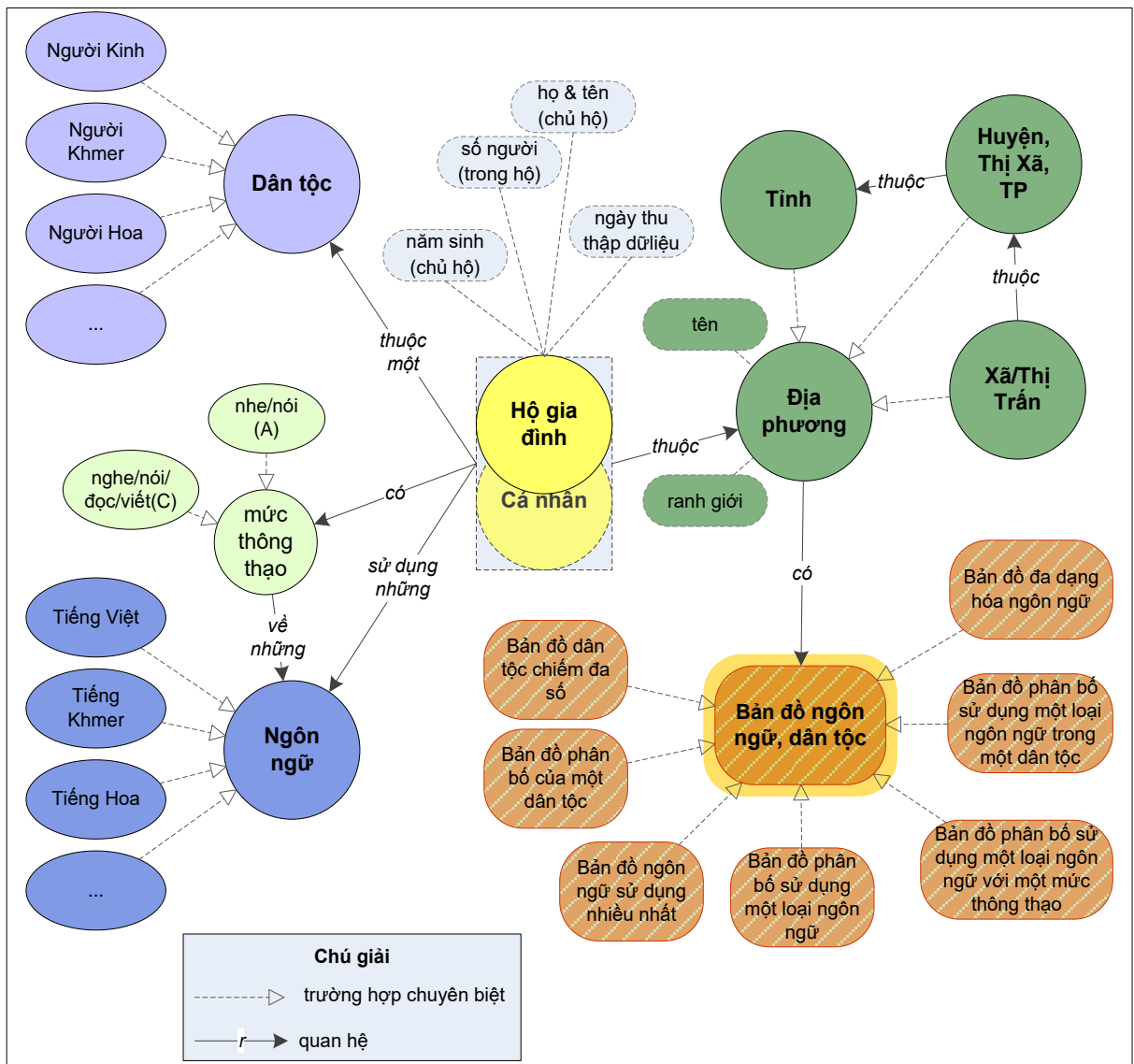
Có các hệ quản trị cơ sở dữ liệu có thể hỗ trợ truy xuất dữ liệu không gian là:

- Oracle, with spatial extension (cần bản quyền)
- Microsoft SQL Server (cần bản quyền)
- IBM – DB2 (cần bản quyền)
- PostgreSQL/POSTGIS (mã nguồn mở)
- MySQL (mã nguồn mở)

Trong phạm vi đề tài nghiên cứu, các hệ quản trị cơ sở dữ liệu cần bản quyền sẽ không được xem xét vì vấn đề kinh phí. PostgreSQL/PostGIS và MySQL đều là phần mềm miễn phí và được sử dụng rộng rãi. PostgreSQL/PostGIS có nhiều chức năng truy cập về không gian hơn và là chuẩn mã nguồn mở cho việc quản lý thông tin địa lý và không gian. Ngoài ra, lý do PostgreSQL/PostGIS được chọn lựa để lưu trữ dữ liệu địa lý, ngôn ngữ và dân tộc là vì PostgreSQL/PostGIS cung cấp cách quản lý dữ liệu nhiều chiều về không gian và thời gian (temporal/spatial), điều này rất thuận lợi cho việc phân tích biến động ngôn ngữ và dân tộc theo thời gian sau này.

2.1.2 Mô tả các khái niệm về dữ liệu

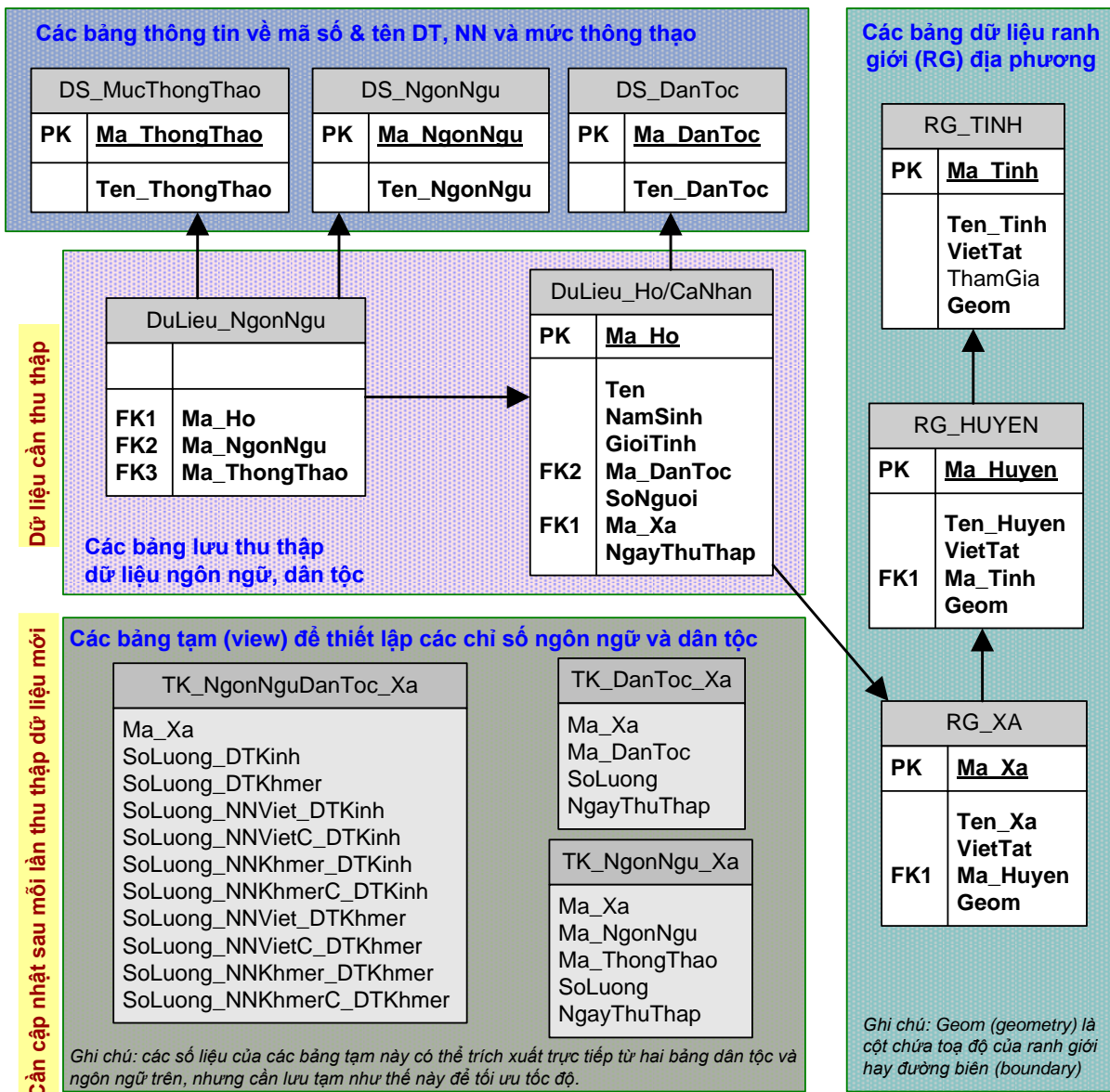
Để tạo cơ sở dữ liệu lưu trữ thông tin đòi hỏi sự phát triển của một mô hình dữ liệu để phản ánh mối quan hệ thời gian, không gian (địa phương), ngôn ngữ và dân tộc. Bước đầu tiên là xây dựng một mô hình khái niệm đại diện cho một mối quan hệ trực quan giữa các yếu tố ngôn ngữ, dân tộc và địa phương như trong Hình 5 dưới đây:



Hình 5. Mô tả dữ liệu

2.1.3 Mô hình dữ liệu

Bước tiếp theo trong việc thực hiện một cơ sở dữ liệu là xây dựng một mô hình dữ liệu. Mô hình dữ liệu sẽ gồm các thực thể và các mối quan hệ giữa các thực thể và các thuộc tính, các kết nối giữa các bảng như trong Hình 6.

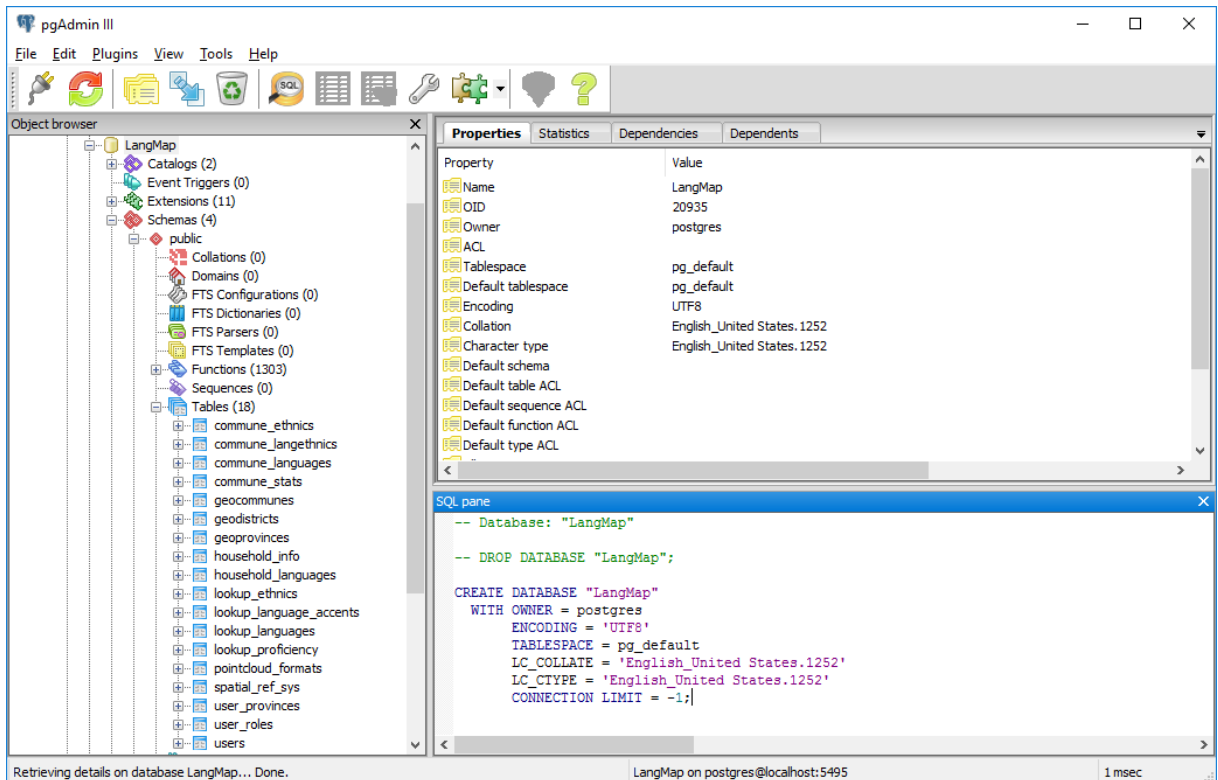


Hình 6. Mô hình thực thể

Ngoài ra còn các bảng chứa thông tin người dùng và vai trò của người dùng (admin có thể xem tất cả các tỉnh có dữ liệu của trang bảo mật, người dùng thường chỉ thấy dữ liệu của tỉnh liên quan trong trang bảo mật). Vì lý do đề tài nghiên cứu về ngôn ngữ và dân tộc nên các mô hình dữ liệu khác không được đề cập chi tiết.

2.1.4 Cơ sở dữ liệu ở mức vật lý

Bước cuối của việc thiết kế và cài đặt cơ sở dữ liệu là cài đặt ở mức vật lý. Các bảng và các mối quan hệ cơ sở dữ liệu được cài đặt trong PostgreSQL phiên bản 9.5 với phần mở rộng PostGIS phiên bản 2.2 và được cài trên hệ điều hành Windows, Windows Server hay Unix.



Hình 7. Cài đặt cơ sở dữ liệu ở mức vật lý

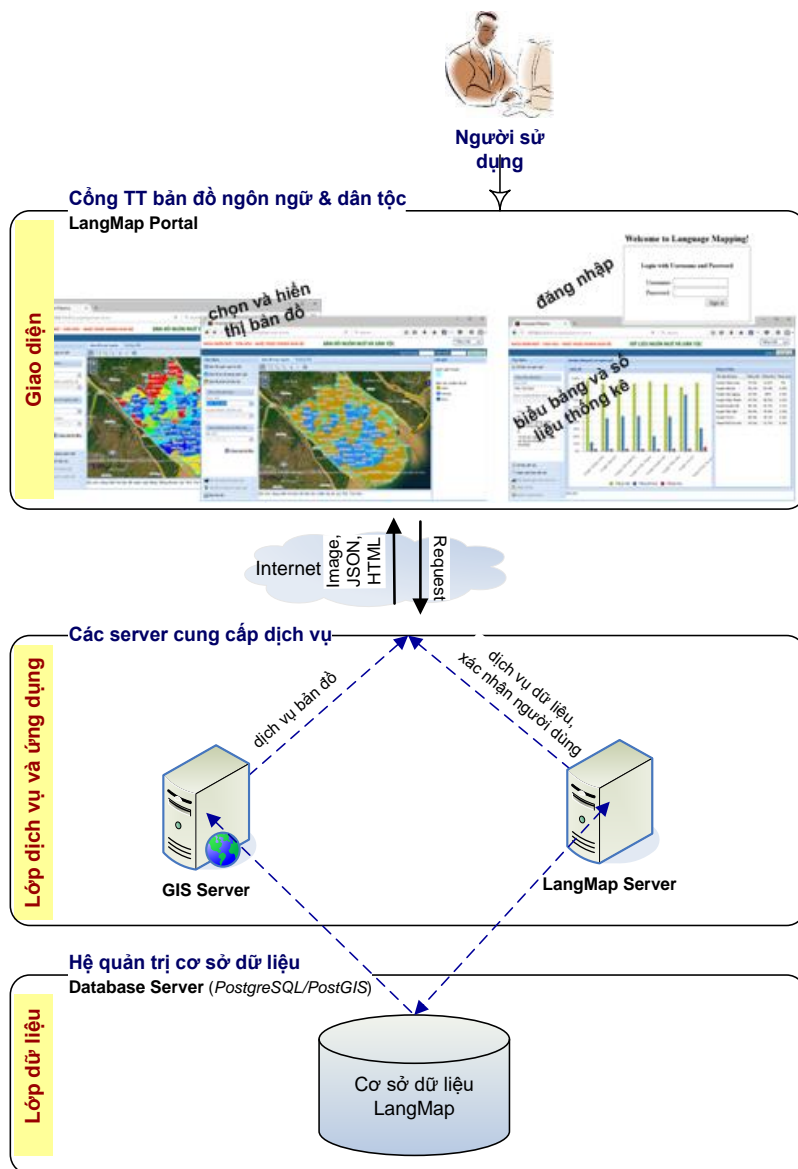
Lưu ý rằng các tên bảng và các thuộc tính được viết bằng tiếng Anh khi cài đặt ở mức vật lý để thuận tiện trong việc trao đổi. Có thể tra cứu trong từ điển về dữ liệu, vui lòng xem thêm ở Phụ lục 4.

2.2 Tổng quan về cổng thông tin bản đồ ngôn ngữ và dân tộc

Cổng thông tin bản đồ ngôn ngữ và dân tộc được thiết kế theo cấu trúc hướng dịch vụ SOA (Service-Oriented Architecture). Hệ thống sẽ bao gồm 3 lớp chính: lớp dữ liệu, lớp dịch vụ và ứng dụng, và lớp giao diện như trong Hình 8.

- **Lớp dữ liệu:** là lớp ở mức thấp nhất sẽ là cơ sở dữ liệu LangMap, chính là nơi lưu trữ toàn bộ dữ liệu không gian địa lý (đường ranh giới địa phương), dữ liệu về ngôn ngữ và dân tộc, và cả thông tin về người dùng để truy cập vào trang bảo mật xem các thông tin quan trọng cho việc hoạch định chiến lược về ngôn ngữ học dân tộc.

- **Lớp dịch vụ và ứng dụng:** là lớp có các máy cung cấp các dịch vụ để vẽ bản đồ và thống kê dữ liệu. Các dịch vụ sẽ được cung cấp bởi 2 server là GIS Server và LangMap server:



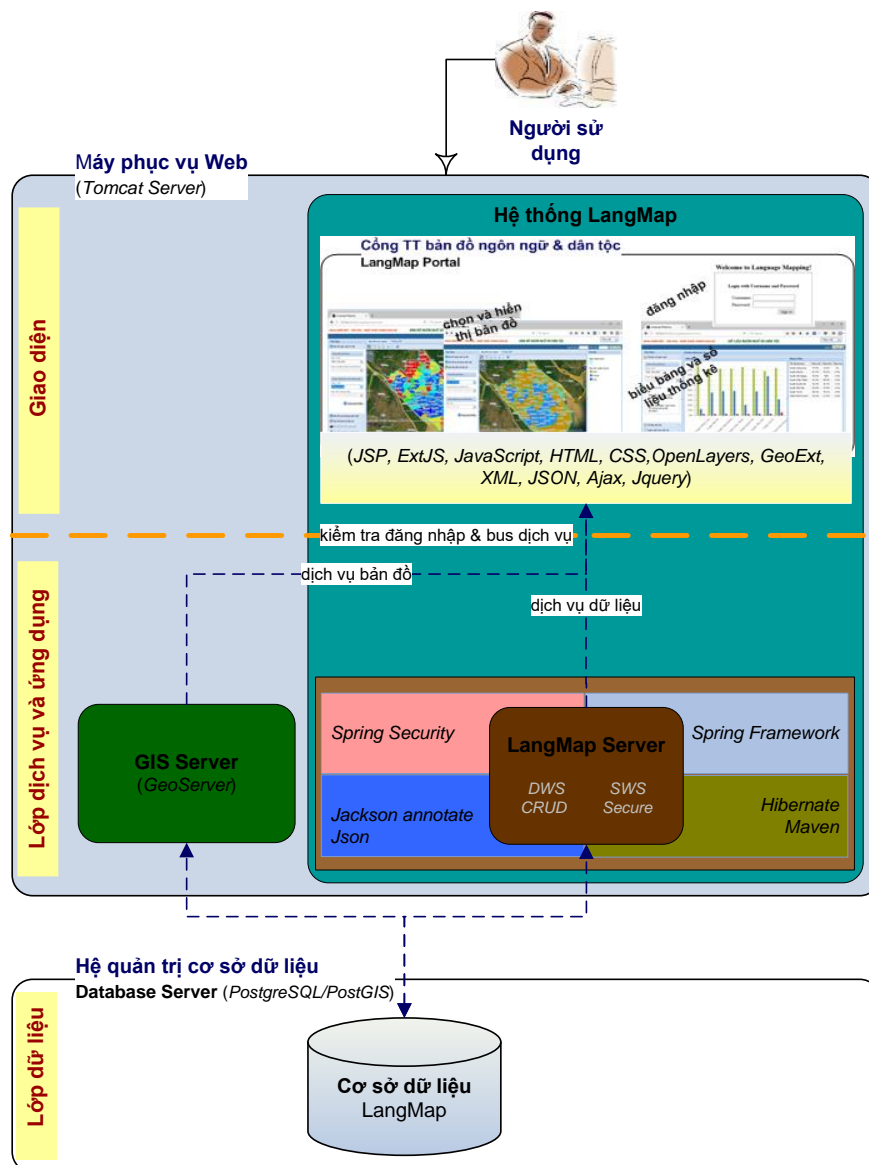
Hình 8. Tổng quan về hệ thống

- GIS Server sẽ cung cấp các dịch vụ vẽ bản đồ trực tuyến (on-the-fly rendering) theo chuẩn OGC (Open Geospatial Consortium). Các câu lệnh truy xuất dữ liệu để tạo bản đồ trực tuyến sẽ được cài đặt trực tiếp trên GIS Server kèm với các tham số về địa lý, ngôn ngữ, dân tộc và các chỉ số tính toán. Một vài câu lệnh quan trọng sẽ được liệt kê trong Phụ lục 5 - Một số câu lệnh để tạo các dịch vụ trực tuyến (on-the-fly)
- LangMap server là server được viết bởi nhóm nghiên cứu để cung cấp các dịch vụ về dữ liệu không thuộc về không gian (non-spatial). Các dịch vụ này cung cấp các danh sách ngôn ngữ dân tộc theo cảnh hướng của địa phương, các số liệu tính toán của các chỉ số về ngôn ngữ và dân tộc đề cập ở Mục 1.2 của Chương 1. Dữ liệu từ các dịch vụ này sẽ dưới dạng XML và JSON để

tiện cho việc làm các danh sách chọn lựa và biểu đồ trong giao diện của cổng thông tin.

Các công nghệ dùng để phát triển hệ thống được mô tả trong Hình 9:

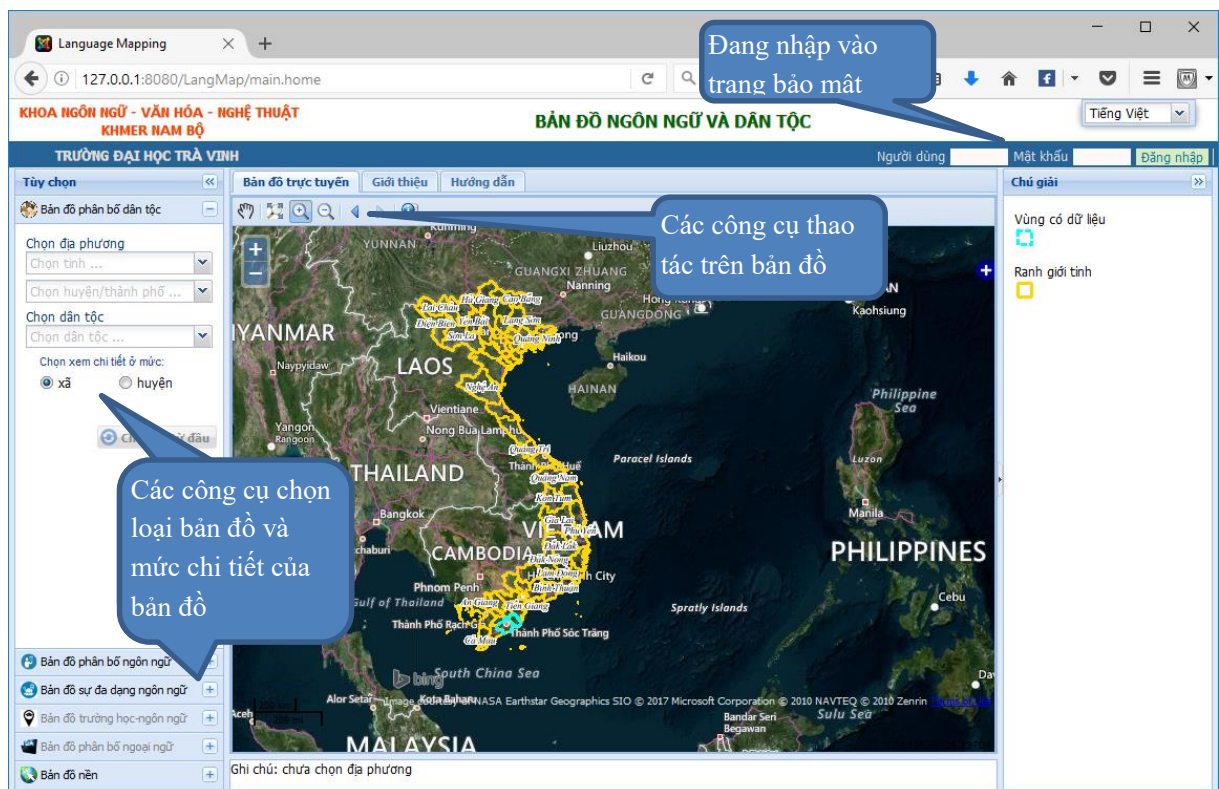
- Hệ quản trị cơ sở dữ liệu: PostgreSQL 9.5 /PostGIS 2.2 64 bit
- GIS Server là phần mềm mã nguồn mở GeoServer (hoặc MapServer tùy theo kinh nghiệm của người phát triển)
- LangMap Server do nhóm phát triển dựa trên các nền tảng của công nghệ Java là Spring Security, Spring MVC và công nghệ Hibernate để kết nối với cơ sở dữ liệu.



Hình 9. Các thành phần của hệ thống và công nghệ sử dụng

- Công nghệ dùng phát triển giao diện là các công nghệ phát triển Web hiện đại như ExtJS, JQuery, OpenLayers, GeoExt,...
- Apache Tomcat phiên bản 8.5 cài đặt để làm máy chủ phục vụ Web. GIS Server và LangMap Server được triển khai trong Apache Tomcat.

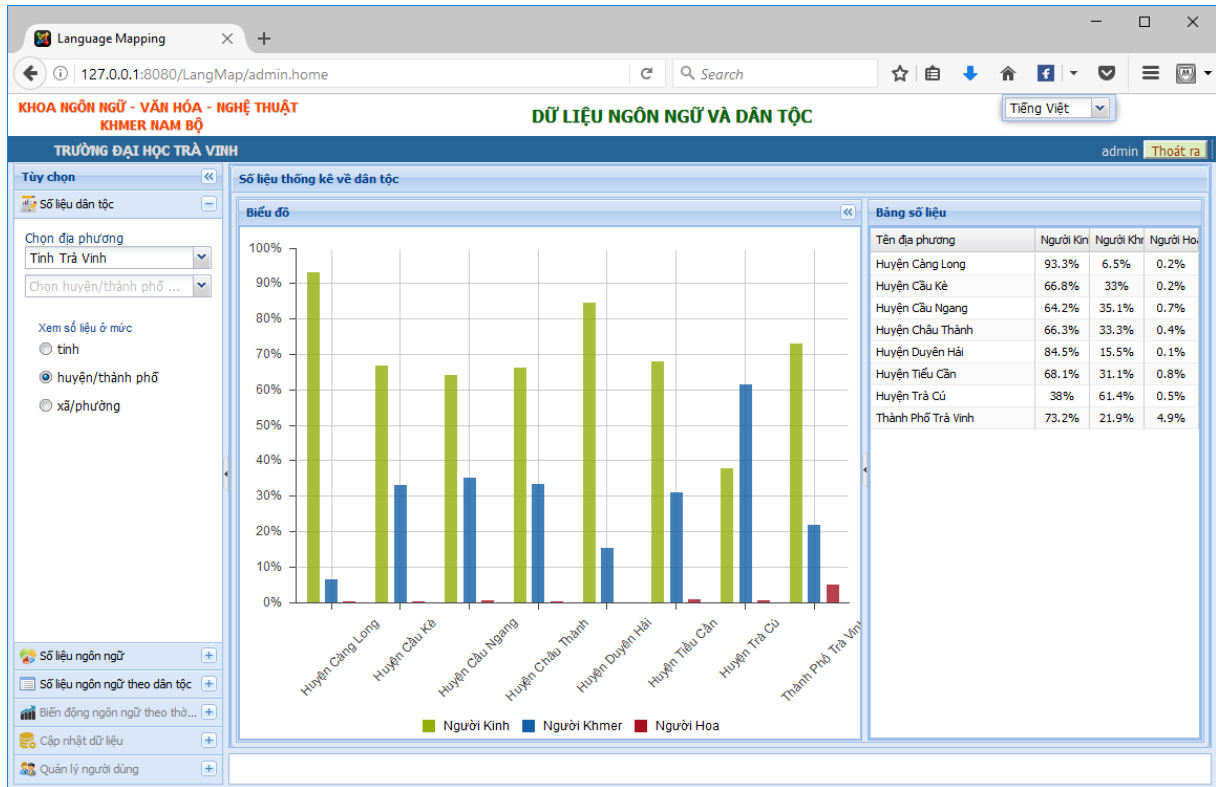
Giao diện trang công khai (Hình 10) cho phép người sử dụng chọn lựa và xem bản đồ ngôn ngữ, bản đồ dân tộc, bản đồ đa dạng hóa ngôn ngữ và có thể chọn bản đồ nền như là OpenStreetMap, Google Streets, Google Hybrid, Google Physical, Bing Map Aerial, Bing Map Road. Với mỗi loại bản đồ trên mà người dùng chọn lựa, thì người dùng có thể lựa chọn chi tiết hơn về địa phương (tỉnh hoặc huyện/thị hoặc phường/xã) và chọn ngôn ngữ để xem bản đồ về ngôn ngữ ở địa phương đó. Giao diện này cũng sẽ có chỗ để nhà quản lý (Admin) đăng nhập vào trang bảo mật được đề cập dưới đây. Các ứng dụng khác như tạo bản đồ phân bố ngoại ngữ, bản đồ quan hệ giữa trường học và ngôn ngữ giảng dạy sẽ được thực hiện trong một đề tài khác, nhưng sẽ thêm vào hệ thống LangMap này.



Hình 10. Giao diện của trang công khai

Trên giao diện này, bên trái là các tùy chọn để người sử dụng chọn xem các loại bản đồ và bản đồ sẽ được hiện ở giữa, bên phải là chú giải cho bản đồ.

Giao diện trang bảo mật (Hình 11) sẽ cho phép người dùng xem thông tin về ngôn ngữ và dân tộc được dạng biểu bảng và biểu đồ. Các số liệu chi tiết về ngôn ngữ và dân tộc sẽ giúp ích cho việc hoạch định chiến lược về ngôn ngữ và cách chính sách về dân tộc. Các kết quả về bản đồ ngôn ngữ và dân tộc cũng nhưng kết quả trong trang bảo mật của tỉnh Trà Vinh sẽ được đề cập trong phần Phụ Lục 3.



Hình 11. Giao diện của trang bảo mật

Các chức năng bản đồ biến động ngôn ngữ và dân tộc theo thời gian (cần có dữ liệu về ngôn ngữ và dân tộc ở nhiều năm khác nhau), chức năng nhập dữ liệu và chức năng quản lý người dùng sẽ được phát triển trong các đề tài tiếp theo.

CHƯƠNG 3. ỨNG DỤNG CÔNG THÔNG TIN BẢN ĐỒ NGÔN NGỮ DÂN TỘC HỌC TRONG ĐÁNH GIÁ CẢNH HUỐNG NGÔN NGỮ

Vai trò của ngôn ngữ trong đời sống hàng ngày, đặc biệt là trong bối cảnh kinh tế, chính trị, xã hội của tỉnh Trà Vinh cần thiết được xác định. Bằng việc nhận diện bối cảnh ngôn ngữ dân tộc học đang diễn ra, các nhà quản lý (nhà nước và tư nhân), các tổ chức phi chính phủ sẽ có cơ sở để phân bổ nguồn ngân sách, theo dõi diễn biến của cư dân, có kế hoạch ưu tiên cho một số địa phương, ghi nhận thông tin về tình hình đang diễn ra ở từng vùng...

3.1 Ngôn ngữ và đời sống cộng đồng

3.1.1 Tính thiết yếu của ngôn ngữ trong đời sống

Ngôn ngữ thể hiện ý thức xã hội. Mỗi hệ thống ngôn ngữ phản ánh bản sắc của cộng đồng nói ngôn ngữ đó: (phong tục, tập quán, thói quen, của cả một cộng đồng).

Ngôn ngữ tồn tại phát triển gắn liền với sự tồn tại và phát triển của xã hội. Xã hội ngày càng đa dạng, phức tạp, phong phú hơn thì ngôn ngữ càng ngày càng phong phú, đa dạng hơn để kịp thời phản ánh sự tiến bộ của xã hội.

Ngôn ngữ là một trong những động lực bảo đảm sự tồn tại và phát triển của xã hội loài người. Bởi vì nhờ ngôn ngữ mà những tư tưởng, tình cảm, nhận thức của con người có thể truyền từ người này đến người khác, từ thế hệ này đến thế hệ khác để con người đồng tâm hiệp lực chinh phục tự nhiên, cải tạo xã hội làm cho đời sống xã hội ngày càng đi lên. Khoa học kỹ thuật đòi hỏi phải nâng cao chất lượng của hoạt động giao tiếp để đưa kiến thức khoa học đang không ngừng tăng lên vào các lĩnh vực của đời sống, của xã hội.

Không nên đồng nhất ngôn ngữ với các đặc trưng chủng tộc hay dân tộc. Có nghĩa là người Việt thì nói tiếng Việt. Thực ra, ranh giới dân tộc và ranh giới ngôn ngữ không trùng nhau. Sống cộng cư giữa các dân tộc hoặc chủng tộc, chắc chắn sẽ dẫn đến sự xuất hiện người song ngữ, hoặc người Việt nhưng tiếng Việt lại không là tiếng mẹ đẻ (trường hợp kết hôn đa văn hóa, cha người Việt, mẹ người Khmer, v.v...).

Nguyên nhân chủ yếu quyết định phương hướng và cách thức phát triển của ngôn ngữ phải là các nhân tố: điều kiện về kinh tế, chính trị, văn hóa và các điều

kiện xã hội khác. Chẳng hạn sản xuất phát triển, giai cấp xuất hiện, nhà nước hình thành có nhu cầu giao dịch theo qui thức hành chính. Vì thế chữ viết xuất hiện kéo theo sự hình thành ngôn ngữ viết và tác động làm ngôn ngữ phát triển. Ngoài ra các nhân tố như hình thức tổ chức cộng đồng, dân số, trình độ văn hóa, thể chế nhà nước, môi trường sống, truyền thống văn hóa, tốc độ phát triển kinh tế, xã hội, mối quan hệ và tương quan về các mặt với các dân tộc xung quanh... càng ảnh hưởng không ít đến sự phát triển của ngôn ngữ. Ngôn ngữ phát triển theo quy luật khách quan, chủ yếu do sự tác động của các nhân tố khách quan. Tuy nhiên nhân tố chủ quan của con người cũng góp phần vào sự phát triển của nó. Chính sách ngôn ngữ có những ảnh hưởng đáng kể đến sự phát triển của ngôn ngữ.

3.1.2 Bối cảnh ngôn ngữ ở tỉnh Trà Vinh

Là một tỉnh có đông đồng bào dân tộc Khmer, nên sự tiếp xúc giữa hai ngôn ngữ Việt, Khmer đã diễn ra lâu đời. Điều này đã mang đến một số đặc điểm khác đặc trưng của tiếng Việt ở vùng đồng bằng sông Cửu Long nói chung và Trà Vinh nói riêng.

Số lượng người Khmer biết nói tiếng Việt khá lớn đến 90%. Hầu hết người Khmer đều biết nói được tiếng Khmer. Thậm chí các gia đình buôn bán ngay tại thành phố Trà Vinh, mặc dù thời gian tham gia học tập và sinh hoạt hoàn toàn gắn kết với người Việt nhưng khi vào những dịp lễ, đám tiệc cùng với bà con thân thuộc họ sử dụng tiếng Khmer để giao tiếp. Vấn đề ngại nói tiếng Khmer xảy ra khi nói chuyện trước đám đông (phát biểu hội họp...) hoặc ở nơi có sự xuất hiện của người Kinh. Đối với các tình huống thường ngày, người Khmer cảm thấy “thuận tiện” hơn khi sử dụng tiếng Khmer để nói chuyện với nhau.

Các phương tiện nghe nhìn (sách báo, tạp chí, chương trình phát thanh, truyền hình) bằng tiếng Khmer khá đa dạng: Ấn phẩm “Tin ảnh dân tộc và miền núi” của Thông Tấn xã Việt Nam được phát hành có phiên bản tiếng Khmer, nhật báo tiếng Khmer trang của hai tỉnh Trà Vinh và Sóc Trăng, “báo Trà Vinh” và “báo Sóc Trăng, tạp chí “Vappa-tho Khmer” (Văn hoá Khmer); chương trình phát thanh tiếng Khmer của cơ quan thường trú Đài Tiếng Nói Việt Nam, các đài phát thanh – truyền hình của các tỉnh trong vùng đồng bằng Sông Cửu Long; chương trình truyền hình biểu diễn văn nghệ như ca, múa, nhạc kịch bằng tiếng Khmer... Nhìn chung, các chương trình truyền thông đại chúng bằng tiếng Khmer khá đa dạng và phong phú nội dung lẫn hình thức thực hiện, thu hút sự

quan tâm của khán, thính giả người Khmer. Tuy nhiên, khi đi sâu vào chi tiết nhất là đối với những từ ngữ vay mượn từ tiếng Pali, Sanskrit, hoặc những từ ngữ dịch từ các thuật ngữ mới của tiếng Việt phổ thông, khả năng tiếp nhận thông tin của các tầng lớp nhân dân có khác nhau. Đa số đồng bào Khmer chỉ sử dụng tiếng Khmer với chức năng khẩu ngữ giao tiếp hàng ngày nên các thuật ngữ chính trị, kinh tế-xã hội bằng tiếng Khmer rất xa lạ với họ. Tỷ lệ mù chữ Khmer trong đồng bào còn khá cao nên nhiều người không thể đọc báo chữ Khmer. Đối với giới trí thức am hiểu tiếng Khmer còn nhiều ý kiến tranh luận chưa thật sự thống nhất cho cách dùng từ ngữ, hay cách phiên dịch, biên dịch từ tiếng Việt ra tiếng Khmer trên báo đài.

Ở các gia đình Khmer, tình hình giao tiếp song ngữ thường xảy ra, tập trung nhiều ở các gia đình trí thức. Khả năng song ngữ của các thành viên trong gia đình (đặc biệt khi đối với những gia đình có nhiều tiếp xúc với bên ngoài, hay cha mẹ có khả năng song ngữ tốt và con cái được đi học) rất cao. Ở những gia đình công chức hay giáo viên, hầu hết các phụ huynh đều tập trung rèn luyện tiếng Việt (nghe, nói, đọc, viết) cho con trẻ. Điều này đã tạo ra một môi trường song ngữ khá mạnh tại Trà Vinh. Nguyên do chủ yếu là vì tiếng Việt có thể được sử dụng trong những tình huống liên quan đến việc học tập của con cái hay công việc có liên quan đến chính quyền, đoàn thể. Tùy theo thói quen, những gia đình Khmer chủ yếu sử dụng tiếng Khmer trong gia đình, nhưng điều này có thể thay đổi khi trong cuộc hội thoại có mặt của người ngoài gia đình là người Việt. Tùy vào đặc điểm của người này, các thành viên trong gia đình có thể sử dụng dạng song ngữ hay sẽ luân chuyển sử dụng tiếng Việt, tiếng Khmer hay ngược lại. Hiện tượng song ngữ ở cộng đồng người Khmer ở Trà Vinh diễn ra hoàn toàn theo tính chất tự nhiên, có chiều hướng hòa nhập với tiếng Việt nhiều hơn. Người Khmer khẳng định “Khmer là ngôn ngữ gốc” của họ. Và họ cũng khẳng định sự quan trọng của cả hai ngôn ngữ (Việt, Khmer) trong đời sống hàng ngày.

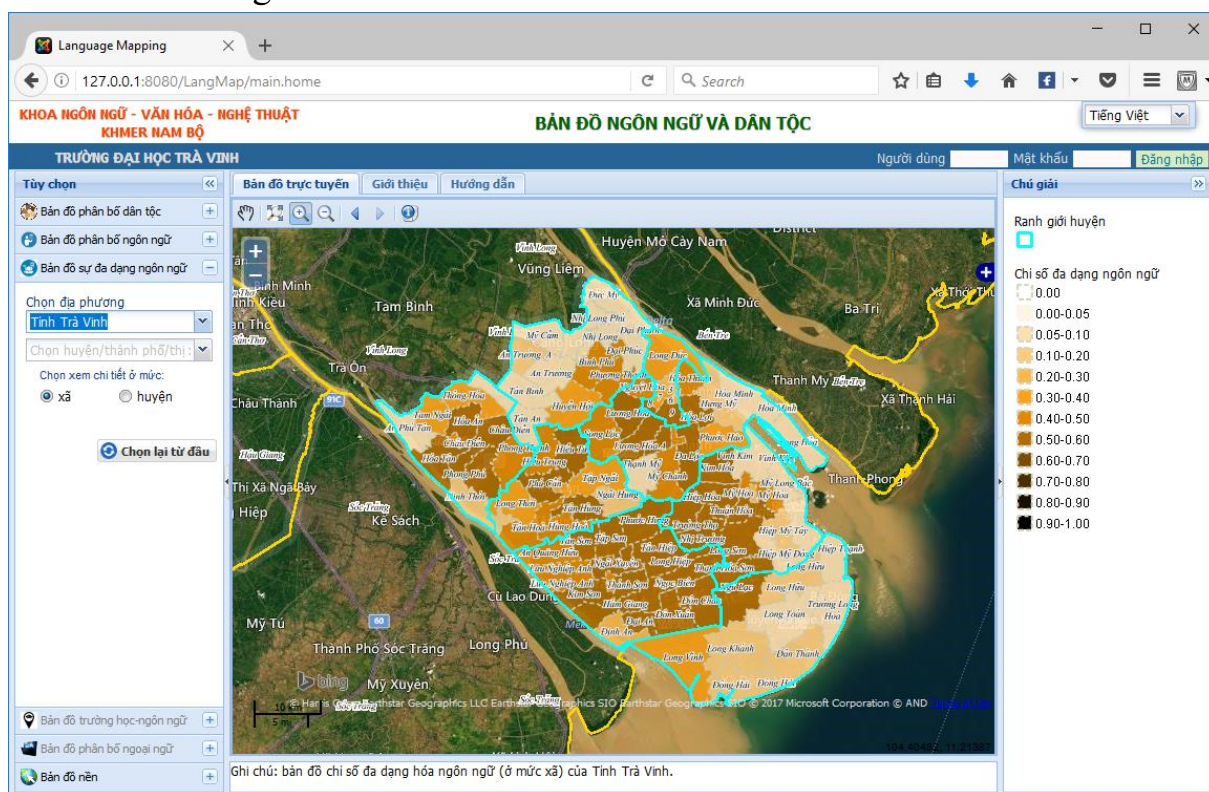
Tiếng Việt chiếm ưu thế khi sử dụng trong các bối cảnh. Hầu như người Khmer không sử dụng tiếng Khmer để giao tiếp với người Kinh (chỉ ngoại trừ trường hợp người Kinh chủ động thể hiện là có thể giao tiếp bằng tiếng Khmer). Vấn đề ở đây là nhằm mục đích thông hiểu. Khi được hỏi tại sao họ cần lựa chọn ngôn ngữ đối với người giao tiếp, thì mọi người đều trả lời rằng “Sợ nói tiếng Khmer người ta không hiểu”. Họ chỉ “ưu tiên” tiếng Khmer khi giao tiếp với người Khmer, còn đối với người Việt họ mặc nhiên là phải sử dụng tiếng

Việt. Kết hợp pha trộn hai ngôn ngữ trong giao tiếp thường xuyên xảy ra. Sự thuận lợi của việc pha trộn ngôn ngữ chính là do sự tương đồng về đặc điểm cấu trúc của hai ngôn ngữ Việt, Khmer nên các từ thay thế xuất hiện trong các pha trộn ngôn ngữ của người dân Khmer Trà Vinh đều thoả mãn về từ loại thay thế và vị trí, trật tự trong câu, hay cụm từ.

3.2 Khai thác giá trị ứng dụng của công thông tin bản đồ trực tuyến về cảnh huống ngôn ngữ ở tỉnh Trà Vinh

Bản đồ trực tuyến về cảnh huống ngôn ngữ sẽ là công cụ trực quan để xác định:

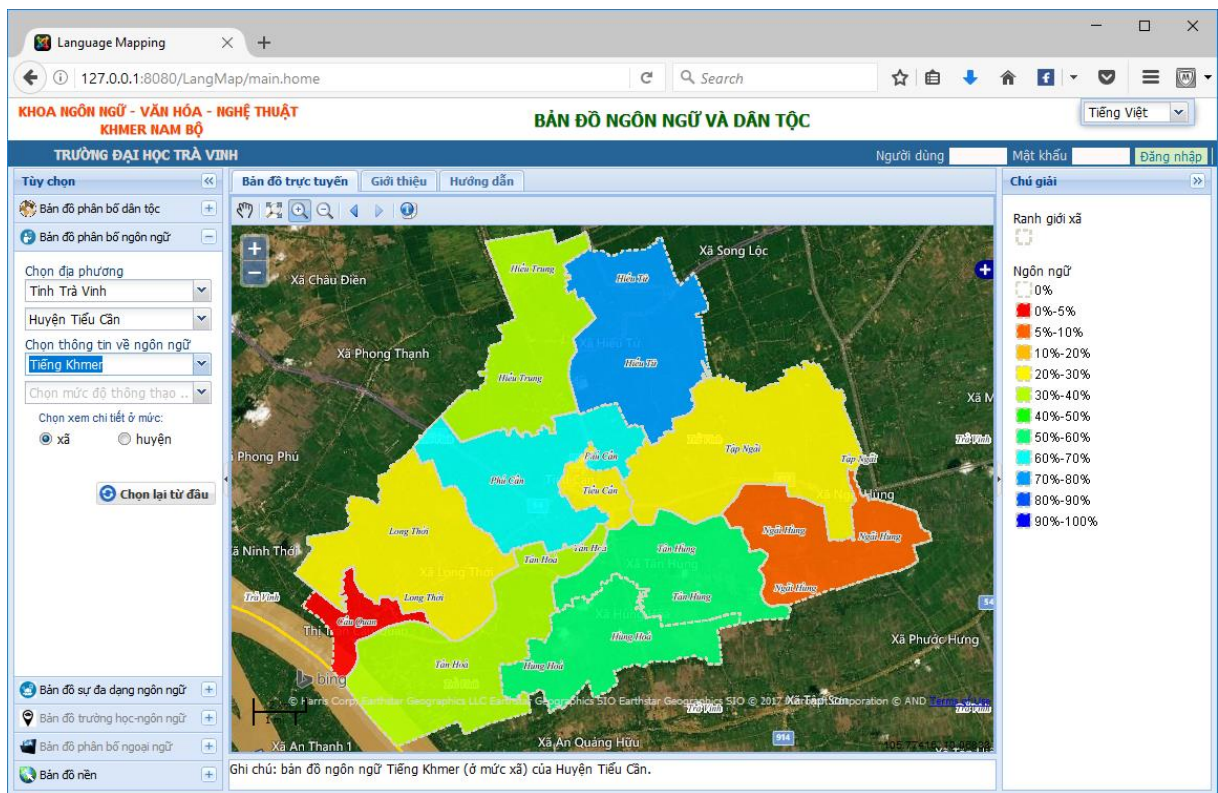
- Sự phân bố của ngôn ngữ đang hành chức, hay có thể nói là mức độ đa dạng ngôn ngữ ở từng khu vực. Các màu hiển thị trên bản đồ càng đậm, chứng tỏ mức độ tiếp xúc ngôn ngữ ở những khu vực này rất cao, và ngược lại, thấp dần khi màu sáng dần.



Hình 12. Bản đồ về sự đa dạng hóa ngôn ngữ ở tỉnh Trà Vinh

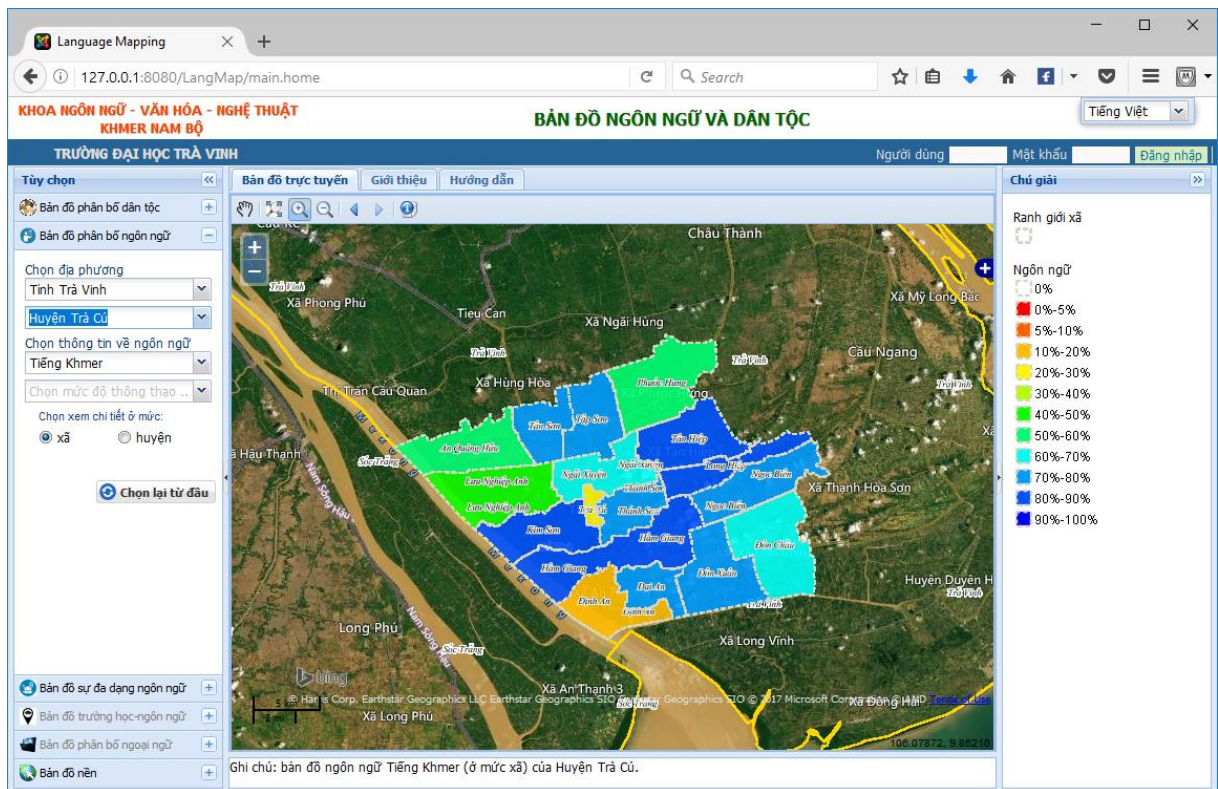
Bản đồ hiển thị mức độ đa dạng hóa các ngôn ngữ đang tồn tại ở tỉnh Trà Vinh. Ở các khu vực ven sông, ven biển, tỷ lệ đa dạng rất thấp (0.00-0.05%), cụ thể như một số xã thuộc huyện Càng Long, Duyên Hải và thành phố Trà Vinh. Mức độ tiếp xúc ngôn ngữ diễn ra mạnh ở các xã như Hòa Ân, Lương Hòa, Nhị Trường... Bản đồ hiển thị vùng đa dạng hóa ngôn ngữ cao sẽ giúp các nhà nghiên

cứu ngôn ngữ xác định vùng trọng tâm cho việc tìm hiểu pidgin và creole. Bên cạnh đó, một khi tiếp xúc ngôn ngữ diễn ra mạnh thì sẽ dẫn đến hiện tượng như trộn mã, chuyển mã hoặc thậm chí vay mượn từ vựng. Về phát âm cũng thay đổi ảnh hưởng lẫn nhau. Có thể, tiếng Việt Trà Vinh ở những vùng này sẽ mang nhiều đặc điểm của ngôn ngữ Khmer (về ngữ âm, chất giọng, từ vựng, cấu trúc, lối diễn giải...); tương tự, tiếng Khmer Trà Vinh cũng vậy. Sẽ càng thêm thú vị, khi thu thập tư liệu ngôn ngữ Khmer ở các vùng này và đối chiếu với đặc điểm ngôn ngữ Khmer ở các vùng ít tiếp xúc như Tân An, Huyện Hội, Phú Tân, Vĩnh Thới... Để cụ thể hơn những vùng có người sử dụng tiếng Khmer nhiều nhất, chúng ta xem Hình 14. Đây là bản đồ về sự lựa chọn sử dụng tiếng Khmer của người dân ở huyện Tiểu Cần, tỉnh Trà Vinh hiện nay.



Hình 13. Bản đồ về sự phân bố tiếng Khmer ở huyện Tiểu Cần

Ngoài ra, để tìm hiểu trạng thái ngôn ngữ Khmer ở huyện Trà Cú (xem Hình 15), bản đồ hiển thị toàn huyện với các màu sắc biểu thị cho mức độ lựa chọn sử dụng tiếng Khmer.



Hình 14. Bản đồ về sự phân bố tiếng Khmer ở huyện Trà Cú

Do ngôn ngữ có tính chất xã hội nên các nhà quản lý nhà nước, hoạch định chính sách ngôn ngữ cần có sự đánh giá về trạng thái ngôn ngữ trên diện rộng (địa lý) và theo quá trình diễn biến (thời gian). Để đánh giá trạng thái ngôn ngữ, về mức độ hành chức của từng ngôn ngữ thì không thể chỉ nhìn trong phạm vi một huyện hay một xã. Từ đây đặt ra yêu cầu cho một tầm nhìn trên diện rộng: cả tỉnh, cả khu vực và quốc gia để có đảm bảo độ tin cậy và bao quát của dữ liệu khi hoạch định chính sách ngôn ngữ. Bản đồ ngôn ngữ dân tộc trực tuyến là một công cụ tốt, đáp ứng một cách hiệu quả yêu cầu này.

Khi vận dụng bản đồ trực tuyến để xem xét năng lực ngôn ngữ, chẳng hạn như khả năng đọc và viết chữ Khmer, ngoài việc hiển thị theo từng địa phương, việc cập nhật dễ dàng được thực hiện. Đáp ứng các hoạt động bảo tồn ngôn ngữ, chữ viết dân tộc, nhiều chương trình học tập và bồi dưỡng tiếng Khmer đã mở ra cho người dân học tập (kể cả người Kinh). Có thể nói, bản đồ trực tuyến sẽ linh hoạt cập nhật theo số lượng người theo học chữ Khmer nhằm hiển thị nhanh nhất và chính xác nhất tỷ lệ biết chữ. Điều này các công cụ hiện tại như bản đồ giấy, hồ sơ lưu về số liệu thống kê chưa làm được.

Bản đồ trực tuyến này sẽ là nguồn tư liệu cung cấp công khai cho mọi người, đáp ứng các quan tâm khác nhau của tất cả các đối tượng trong và ngoài nước.

3.2.1 Đối với các nhà quản lý chính quyền địa phương

Giữ vai trò thiết yếu trong mọi mặt của đời sống xã hội, cảnh huống ngôn ngữ cần được chú trọng và quan tâm nhằm tăng cường hiệu quả quản lý nhà nước, xã hội. Bản đồ ngôn ngữ dân tộc sẽ hỗ trợ các nhà lãnh đạo địa phương trong việc xác định:

- Các ngôn ngữ nào đang được người dân chọn lựa sử dụng trong từng địa phương (đến cấp xã/phường)? Việc lựa chọn ngôn ngữ như vậy có ổn định hay liên tục thay đổi? Nếu liên tục thay đổi thì vì sao? Xu hướng người Khmer đang chọn sử dụng ngôn ngữ nào? Có cần phát triển các lớp học, hoạt động để đảm bảo việc duy trì bản sắc văn hóa, ngôn ngữ dân tộc? Cần thực hiện mạnh, ưu tiên ở vùng nào?

- Các dự án đầu tư (nhà máy, xí nghiệp...) bên cạnh vấn đề địa lý, thổ nhưỡng, giao thông thì đang tập trung ở những vùng sử dụng ngôn ngữ nào?

- Ngôn ngữ được sử dụng như thế nào ở các vùng có tỷ lệ bỏ học cao hoặc trình độ dân trí thấp?

- Tại các địa bàn sử dụng nhiều tiếng Khmer thì đội ngũ cán bộ xã phường có am hiểu về tiếng Khmer để giao dịch, tiếp cận (tiếp xúc người dân, tuyên truyền chính sách, chuyên giao kỹ thuật nuôi trồng)?

- Trong tương lai, tiếng Khmer sẽ lan rộng hay thu hẹp ở các vùng nào? Mức độ dịch chuyển các khu vực có sự pha trộn ngôn ngữ mạnh dự kiến sẽ diễn ra ra sao?

Bên cạnh việc cung cấp dữ liệu cho các nhà quản lý, bản đồ trực tuyến về cảnh huống ngôn ngữ còn là công cụ để giới thiệu về bối cảnh xã hội đang diễn ra ở tỉnh Trà Vinh cho các đối tượng: đầu tư kinh doanh, tham quan du lịch, tìm hiểu về ngôn ngữ Khmer, tìm hiểu về sự đa dạng ngôn ngữ... Một mặt, giúp tăng thêm hiểu biết của người dân địa phương về chính quê hương mình, mặt khác, thu hút sự chú ý của các nhà nghiên cứu hay đầu tư ở một vùng nào đó tại tỉnh Trà Vinh. Công thông tin về cảnh huống ngôn ngữ ở Trà Vinh sẽ làm cầu nối giữa các địa phương trong tỉnh, trong khu vực, quốc gia và với các nước láng giềng, đặc biệt nước bạn Campuchia, nơi tiếng Khmer là ngôn ngữ chính thức. Nhiều hoạt động kinh doanh sản xuất và dịch vụ đã và đang được tăng cường, thúc đẩy hơn mối quan hệ hữu nghị giữa Việt Nam và Campuchia.

3.2.2 Đối với các nhà giáo dục, nhà nghiên cứu

Có nguồn tư liệu được cập nhật và dễ dàng tiếp cận (web) để theo dõi diễn biến sử dụng ngôn ngữ tại một địa phương trong nước. Từ đây, mở ra triển vọng hợp tác nghiên cứu ngôn ngữ, đặc điểm của ngôn ngữ (Việt, Khmer...) tại Trà Vinh – một trong những nơi qui tụ các đặc điểm đặc trưng của ngôn ngữ Khmer Nam Bộ.

3.2.3 Đối với người dân

Bản đồ ngôn ngữ dân tộc sẽ mở rộng hiểu biết của người dân về các ngôn ngữ đang diễn ra trong cộng đồng mình sinh sống, hoặc những vùng họ quan tâm trong tỉnh Trà Vinh.

Nhận thức về vai trò của ngôn ngữ, về thực tế của ngôn ngữ đang diễn ra, người dân sẽ có những đóng góp thiết thực vào sự giữ gìn, phát triển ngôn ngữ, năng lực ngôn ngữ. Họ là những nhân tố quan trọng và quyết định cho hoạt động bảo tồn và phát huy ngôn ngữ dân tộc.

PHẦN KẾT LUẬN

1 Kết quả đề tài và thảo luận

1.1 Tổng kết các hoạt động và kết quả

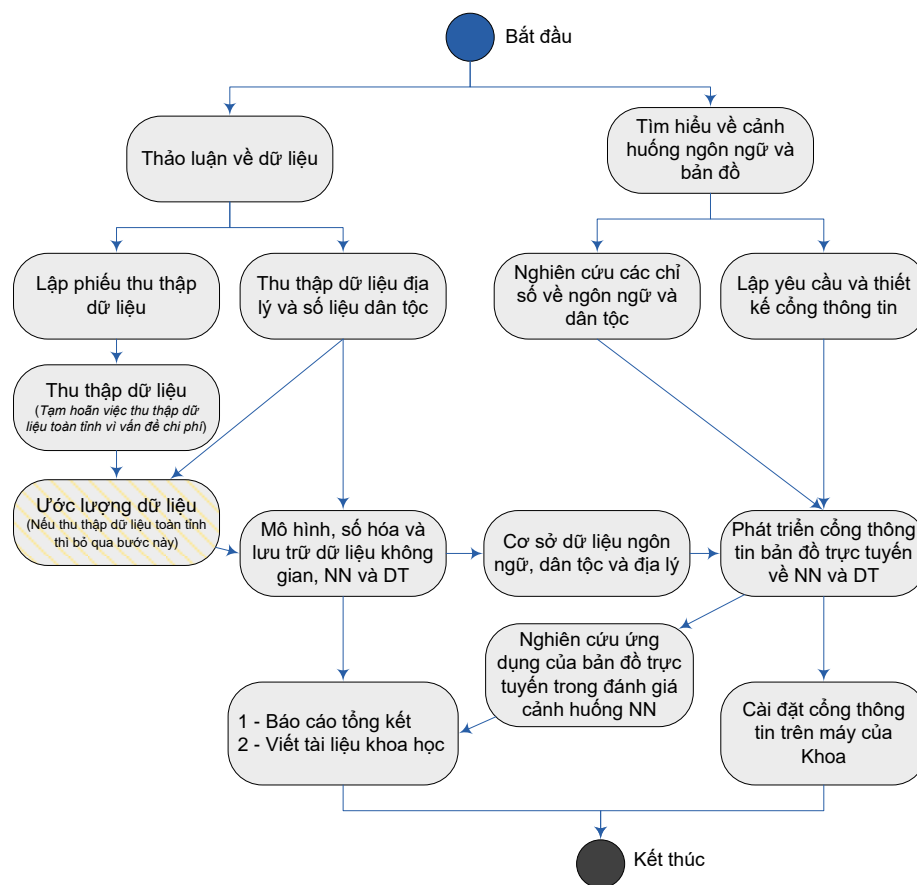
Nhóm nghiên cứu đã hoàn thành các hoạt động đã được xác định trong phạm vi của đề tài nghiên cứu này, bao gồm các hoạt động chính:

- Thu thập dữ liệu
- Nghiên cứu các chỉ số về ngôn ngữ và dân tộc

- Lập cơ sở dữ liệu về ngôn ngữ, dân tộc và địa lý, cũng như là phát triển cổng thông tin bản đồ để nghiên cứu cảnh huống ngôn ngữ của tỉnh Trà Vinh. Hệ thống này được phát triển trên cơ sở sẵn sàng để mở rộng sang các tỉnh khác và cho cả nước. Nhóm nghiên cứu cũng tiến hành thu thập thông tin dân tộc của tỉnh Sóc Trăng và thử nghiệm thành công chức năng mở rộng sang các tỉnh khác.

- Ứng dụng của bản đồ trực tuyến về ngôn ngữ dân tộc học vào nghiên cứu cảnh huống ngôn ngữ.

Các hoạt động chi tiết, bao gồm cả 4 hoạt động chính vừa nêu trên đây, mà đề tài đã hoàn thành được trình bày trong sơ đồ sau đây:



Hình 15. Các bước thực hiện mà đề tài đã hoàn thành

Kết quả đạt được của đề tài để chuyển giao:

TT	Tên sản phẩm	Dạng sản phẩm	Chú thích
1	Phiếu thu thập dữ liệu ngôn ngữ và dân tộc	Bảng Excel	Nên được sử dụng để thu thập dữ liệu ngôn ngữ và dân tộc cùng với đợt điều tra dân số của tỉnh để cập nhật thông tin theo thời gian hàng năm.
2	Các bảng số liệu về dân tộc và mẫu dữ liệu về ngôn ngữ	Bảng Excel	- Số liệu về dân tộc của tỉnh Trà Vinh và Sóc Trăng - Mẫu dữ liệu thu thập ở một số khu vực trong phường 8 và 2.
3	Cơ sở dữ liệu và công thông tin	Hệ thống Web động cùng với CSDL không gian	Đã cài đặt trên máy của Khoa, sẵn sàng để cài trên máy chủ của trường
4	Các tài liệu của đề tài	Các tài liệu dạng PDF	- Đề xuất đề tài - Báo cáo thuyết minh - Báo cáo tổng kết - Bài báo khoa học (đang hoàn thiện)

1.2 Thảo luận

Có một số vấn đề cần được lưu ý và xem xét đối với mức độ thực hiện các hoạt động của đề tài:

- Khó khăn lớn nhất của đề tài chính là thu thập dữ liệu trong thực tế. Vì ngôn ngữ và năng lực ngôn ngữ cần được khảo sát trên từng cá thể trong xã hội. Việc chọn mẫu khảo sát chủ yếu là tìm cơ sở để ước lượng hoặc gọi ra các cách tính chỉ số. Nói chung, thống kê ngôn ngữ học cho phép có độ sai số nhất định. Để khắc phục, cụ thể là áp dụng cho đề tài này, tác giả đề nghị kết hợp thu thập thông tin về tình hình sử dụng ngôn ngữ trong các đợt tổng điều tra dân số hàng năm của tỉnh.

- Ngoài ra, có một số điểm thảo luận đã được xem xét:

(1) Thực hiện đối chiếu mức độ sai lệch giữa phân tính toán chỉ số mang tính ước lượng và dữ liệu thu thập thực tế. Điều này đã giúp khẳng định tính phù hợp của các cách tính và tăng độ tin cậy cho dữ liệu số hóa bản đồ ngôn ngữ toàn tỉnh.

(2) Xác định tính chất phù hợp khi sử dụng bản đồ trực tuyến để hiển thị bối cảnh ngôn ngữ. Cụ thể, các đặc điểm ngôn ngữ sẽ được hiển thị bằng bản đồ trực tuyến như sau:

- Bản đồ ngôn ngữ theo dân tộc
- Bản đồ về năng lực ngôn ngữ theo dân tộc
- Bản đồ ngôn ngữ theo địa phương
- Bản đồ về năng lực ngôn ngữ trong các địa phương
- Bản đồ về sự phân bố các ngôn ngữ (cấp tỉnh, cấp huyện/thị, cấp xã/phường)
- Bản đồ về tỷ lệ sử dụng các ngôn ngữ trong cùng một địa phương

(3) Hướng đến mở rộng ứng dụng bản đồ để sử dụng cho các nghiên cứu về ngôn ngữ học xã hội như: bản đồ ngôn ngữ - trường học; bản đồ phương ngữ; bản đồ về năng lực tiếng Anh; bản đồ về năng lực tiếng Khmer; bản đồ ngôn ngữ theo giới tính; bản đồ ngôn ngữ theo độ tuổi... Ngoài ra, có thể ứng dụng để lập bản đồ văn hóa – du lịch; bản đồ chăn nuôi; bản đồ trồng trọt canh tác....

2 Kiến nghị

2.1 Các hoạt động ngắn hạn

- Dữ liệu của bản đồ cần được cập nhật để đáp ứng hiệu quả sử dụng, do đó, cần có kế hoạch thu thập dữ liệu ở từng hộ (có thể kết hợp với hoạt động điều tra hộ dân hàng năm của tỉnh).
- Nghiên cứu này chỉ mới tập trung mô tả các vùng sử dụng ngôn ngữ theo dân tộc, theo mức độ thông thạo chứ chưa tìm hiểu về số lượng người song ngữ; mục đích sử dụng cho từng ngôn ngữ; mức độ thường xuyên sử dụng ngôn ngữ...
- Cổng thông tin bản đồ sẽ được cài đặt tạm thời trên máy cá nhân ở Khoa, nên được cài trên máy chủ của trường đại học Trà Vinh để người dùng có thể truy xuất qua internet ở bất kỳ nơi đâu. Đề xuất địa chỉ của cổng thông tin là <http://lca.tvu.edu.vn/LangMap/main.home>
- Chức năng quản lý người dùng của trang bảo mật và chương trình nhập dữ liệu vào hệ thống sẽ được phát triển ở đề tài tiếp theo.

2.2 Các hoạt động triển khai mở rộng

- Hướng đến mở rộng ứng dụng bản đồ để sử dụng cho các nghiên cứu về ngôn ngữ học xã hội như: bản đồ ngôn ngữ - trường học; bản đồ phương ngữ; bản đồ về năng lực tiếng Anh; bản đồ về năng lực tiếng Khmer; bản đồ ngôn ngữ theo giới tính; bản đồ ngôn ngữ theo độ tuổi... Ngoài ra, có thể ứng dụng để lập bản đồ về những lĩnh vực nghiên cứu khác của Khoa như là về văn hóa.
- Nên tiến hành thu thập dữ liệu định kỳ, chẳng hạn như thu thập dữ liệu ngôn ngữ cùng với đợt điều tra dân số nhằm nghiên cứu sự biến động của ngôn ngữ theo thời gian. Lợi ích của việc tìm hiểu biến động ngôn ngữ theo thời gian là nhằm tìm hiểu sự biến mất của ngôn ngữ của dân tộc thiểu số, tìm hiểu xem mức độ thông thạo như đọc/viết có được cải thiện theo thời gian hay không để từ đó các nhà quản lý có thể đưa ra chính sách tốt về ngôn ngữ.
- Cổng thông tin bản đồ ngôn ngữ và dân tộc đã có sẵn chức năng mở rộng sang các tỉnh khác và cả nước, tuy nhiên cần thông báo rộng rãi dưới hình thức hội thảo hay giới thiệu đến các tỉnh để cùng chung tay mở rộng sử dụng hệ thống. Một khi các tỉnh thu thập và cung cấp thông tin về ngôn ngữ và dân tộc thì Khoa sẽ phục vụ nạp dữ liệu vào hệ thống và sẵn sàng để các tỉnh có thể sử dụng.

TÀI LIỆU THAM KHẢO

- Dương Thị Thanh Hoa. (2010). Cảnh hưởng ngôn ngữ ở Thái Nguyên. *Luận văn thạc sỹ*. Trường đại học Thái Nguyên.
- Nguyễn Thị Thanh Huyền. (2010). Cảnh hưởng ngôn ngữ ở Hà Giang. *Luận văn Thạc sỹ*. Trường đại học Thái Nguyên.
- CLPP. (không ngày tháng). Đã truy lục 04 21, 2017, từ Consortium for Language Policy and Planning: <http://ccat.sas.upenn.edu/plc/clpp/>
- Eurobarometer. (2012). *Special Eurobarometer 386: Europeans and their Languages*. Đã truy lục Jan 30, 2017, từ European Union Open Data Portal: https://data.europa.eu/euodp/en/data/dataset/S1049_77_1_EBS386
- GermanAccentsDialects. (không ngày tháng). Được truy lục từ The Germany: <https://www.thegermanz.com/german-accents-dialects-german-germans-dont-understand/>
- Greenberg, J. H. (1956). The Measurement of Linguistic Diversity. *Linguistic Society of America*, 32(1), 109-115.
- Hà Thị Tuyết Nga. (2014). Nghiên cứu cảnh hưởng ngôn ngữ dân tộc Tày ở vùng đông Bắc Việt Nam. *Luận án Tiến sỹ*. Đại học quốc gia Hà Nội.
- Khổng Diên. (1995). *Dân số và dân số tộc người ở Việt*. Khoa học xã hội.
- Lameli, A.; Kehrein, R.; Rabanus, S. (Eds). (2010). *Language and Space : Language Mapping: An International Handbook of Linguistic Variation*. (Tập Volume 2). Handbook of linguistics and communication.
- Marian, J. (2011). *Map of the percentage of people speaking English in the EU by country*. Đã truy lục Jan 30, 2017, từ <https://jakubmarian.com/map-of-the-percentage-of-people-speaking-english-in-the-eu-by-country/>
- Moseley, C. (2010). *Atlas of the World's Languages in Danger*. (3rd Edition). Paris: UNESCO Publishing. Đã truy lục Jan 30, 2017, từ UNESCO: <http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/linguistic-diversity-and-multilingualism-on-internet/atlas-of-languages-in-danger/language-mapping/>

- Nguyễn Hữu Hoàn; Nguyễn Văn Lợi; Tạ Văn Thông. (2013). *Ngôn ngữ, chữ viết các dân tộc thiểu số ở Việt Nam (những vấn đề chung)*. NXB từ điển Bách khoa.
- Nguyễn Thị Thu Dung. (2015). Nghiên cứu trạng thái đa ngữ tại huyện Mường Chà-tỉnh Điện Biên. *Luận án Tiến sỹ*. Trường đại học Quốc gia Hà Nội.
- Nguyễn Văn Khang. (2014). Nhìn lại chính sách ngôn ngữ của Đảng và nhà nước Việt Nam về tiếng Việt và những vấn đề đặt ra đối với tiếng Việt hiện nay. *Ngôn ngữ và Đời sống*.
- Ryan, C. (2013). Language Use in the United States: 2011. *U.S. Census Bureau*.
- Simons, Gary F. and Charles D. Fennig (eds.). (2017). *Ethnologue: Languages of the World, Twentieth edition*. Đã truy lục Jan 30, 2017, từ Ethnologue: <https://www.ethnologue.com/statistics/country>
- Tạ Văn Thông. (2009). *Tìm hiểu ngôn ngữ các dân tộc ở Việt Nam*. Khoa học xã hội.
- Tổng cục thống kê. (2016). *Kết quả chủ yếu điều tra biến động dân số và kế hoạch hóa gia đình thời điểm 1/4/2015*. Hà Nội: NXB Thống Kê.
- Trần Phương Nguyên. (2012). Những nhân tố ảnh hưởng đến cảnh huống ngôn ngữ của người Chăm ở Nam Bộ. *Khoa học Xã hội số 11 (171)*, 41.
- Trần Trí Dõi. (1999). *Nghiên cứu ngôn ngữ các dân tộc thiểu số Việt Nam*. Đại học Quốc gia Hà Nội.
- Trần Trí Dõi. (2001). *Ngôn ngữ và sự phát triển văn hoá xã hội*. Văn hoá Thông tin, Hà Nội.
- VNN. (2015). *Những con số về ngôn ngữ trên thế giới*. Được truy lục từ vietnamnet.vn: <http://vietnamnet.vn/vn/giao-duc/nhung-con-so-noi-ve-ngon-ngu-tren-the-gioi-234032.html>
- Zaniewski, K. (2011). *The linguistic diversity of the world in one map*. Đã truy lục Jan 30, 2017, từ <https://matadornetwork.com/read/linguistic-diversity-world-one-map/>