

THAM LUẬN

KẾT QUẢ TRIỂN KHAI THỬ NGHIỆM VÀ ĐỊNH HƯỚNG PHÁT TRIỂN, NHÂN RỘNG SẢN PHẨM ĐỀ TÀI CHO CÁC TIẾNG DÂN TỘC KHÁC TRÊN CẢ NƯỚC

ThS. Trần Duy Linh

*Giám đốc Trung tâm Công nghệ thông tin và Truyền thông
Quảng Ngãi, Chủ nhiệm Đề tài*

I. SỰ CẦN THIẾT VÀ ĐẶT VẤN ĐỀ NGHIÊN CỨU

1. Xuất phát điểm, sự cần thiết thực hiện Đề tài

Hiện nay, trên địa bàn tỉnh Quảng Ngãi, dân số vùng đồng bào dân tộc thiểu số và miền núi khoảng 230.705 người (với 64.223 hộ). Trong đó, đồng bào dân tộc thiểu số là 187.072 người (với 49.818 hộ), gồm: dân tộc Hrê là 133.103 người, sinh sống chủ yếu ở các huyện Ba Tơ, Minh Long và Sơn Hà; dân tộc Co là 33.227 người, sinh sống chủ yếu ở huyện Trà Bồng; dân tộc Xê đăng (Ca dong) là 19.690 người, sinh sống chủ yếu ở huyện Sơn Tây và các dân tộc thiểu số khác là 1.052 người¹.

Trong những năm qua, đồng bào dân tộc thiểu số trong tỉnh cũng đã phát huy truyền thống anh hùng, tinh thần vượt khó, hưởng ứng nhiệt tình các phong trào thi đua yêu nước, tích cực tham gia ngày càng sâu rộng vào các chương trình, chính sách phát triển kinh tế, đảm bảo an sinh xã hội. Tuy nhiên, vẫn có khoảng cách lớn giữa người Kinh và đồng bào dân tộc thiểu số, trong đó khoảng cách giao tiếp là vấn đề cần khắc phục đầu tiên.

Trong những năm qua, với sự quan tâm của các cấp ngành có liên quan, hàng năm, Sở Nội vụ phối hợp với Trường Đại học Phạm Văn Đồng có tổ chức các lớp giảng dạy tiếng đồng bào dân tộc thiểu số (cụ thể là tiếng Hrê và tiếng Co) cho đối tượng cán bộ công chức, viên chức, giáo viên. Tuy nhiên, chưa có Cơ sở dữ liệu tiếng đồng bào dân tộc thiểu số nào được biên soạn để phục vụ việc tra cứu ngôn ngữ quá trình học mà chỉ có các tài liệu giáo trình giảng dạy, tài liệu nghiên cứu của các nhóm tác giả. Tài liệu mới nhất và chuẩn nhất được cấp có thẩm quyền công nhận của tiếng Hrê và tiếng Co được sử dụng hiện nay là:

- Tài liệu Đào tạo, bồi dưỡng tiếng Hrê (dùng cho cán bộ công chức tại miền núi và công tác dân tộc tỉnh Quảng Ngãi) của UBND tỉnh Quảng Ngãi (ban hành tạm thời kèm theo Quyết định số 893/QĐ-UBND ngày 27/4/2007 của UBND tỉnh Quảng Ngãi). Ban biên soạn: Chủ biên: Đỗ Anh; cộng tác viên chính:

¹ Báo cáo số 1295/BC-BDT ngày 21/11/2023 của Ban Dân tộc tỉnh Quảng Ngãi về tình hình triển khai thực hiện Chiến lược công tác dân tộc giai đoạn 2021-2030, tầm nhìn đến năm 2045 trên địa bàn tỉnh Quảng Ngãi

Trương Hồng Hoà, Đinh Cơi, Nga Ri Vê.

- Tài liệu “BAI HOC APOK KOOL (BÀI HỌC TIẾNG CO)” do nhóm biên soạn: Nguyễn Minh Trí, Tạ Văn Thông, Hồ Văn Năm, Tạ Quang Tùng thực hiện. Tài liệu này là một công trình thuộc Đề tài KHCN cấp tỉnh: “Nghiên cứu, biên soạn tài liệu dạy – học tiếng Co cho cán bộ, công chức (người Kinh) công tác tại huyện Trà Bồng và huyện Tây Trà, tỉnh Quảng Ngãi” do chủ nhiệm đề tài ThS. Nguyễn Minh Trí thực hiện năm 2016. Tài liệu “Bài học tiếng Co” hướng đến đối tượng sử dụng là cán bộ công chức người Kinh công tác tại huyện Trà Bồng, với mục đích cung cấp phương tiện cho người học biết được cách viết cách đọc chữ Co, bước đầu nắm những đặc điểm cơ bản trong lời ăn tiếng nói của người Co để từ đó có điều kiện tìm hiểu và nâng cao các kỹ năng sử dụng ngôn ngữ này trong những hoàn cảnh đa dạng của đời sống. Ngày 27/8/2019, UBND tỉnh Quảng Ngãi có công văn số 4796/UBND-NC về việc cho phép Trường Đại học Phạm Văn Đồng sử dụng kết quả đề tài nghiên cứu khoa học cấp tỉnh này để xây dựng chương trình dạy - học và thực hiện bồi dưỡng tiếng Co cho cán bộ, công chức, viên chức huyện Trà Bồng.

Như vậy, hiện nay chỉ mới có tài liệu phục vụ học tập là giáo trình ở dạng giấy do những người có kinh nghiệm biên soạn. Tài liệu dạng giấy bị giới hạn trong phạm vi các lớp bồi dưỡng tiếng Hrê, Co và những người nghiên cứu được tiếp cận bộ tài liệu. Đồng thời, để học được tiếng cần có người hướng dẫn cách đọc, phát âm chứ không thể tự học, tự nghiên cứu. Đây là một số hạn chế của phương pháp học truyền thống mà việc ứng dụng công nghệ số sẽ giải quyết tốt được. Và các tài liệu ở dạng giấy đã được UBND tỉnh đồng ý cho sử dụng trong quá trình dạy học nêu trên sẽ là nguồn dữ liệu đầu vào quan trọng để đưa vào Cơ sở dữ liệu điện tử.

Nhóm thực hiện đã tham khảo, nghiên cứu những công trình Cơ sở dữ liệu điện tử ngôn ngữ trên thế giới, trong nước và trong tỉnh và nhận thấy tại thời điểm năm 2022, trong nước đã có một số công trình nghiên cứu về xây dựng nguồn dữ liệu song ngữ tiếng dân tộc thiểu số ở Việt Nam. Cụ thể, trong nghiên cứu “Xây dựng kho dữ liệu song ngữ Việt - Cơ tu phục vụ tra cứu văn hóa dân tộc Cơ tu”², tác giả đã xây dựng kho dữ liệu song ngữ Việt – Cơ tu nhằm phục vụ tra cứu văn hóa dân tộc Cơ tu. Nghiên cứu này chỉ mới dừng lại ở việc số hóa dữ liệu từ điển Việt – Cơ tu, chưa trang bị cho người dùng công cụ để tra cứu từ điển. Trong nghiên cứu “Xây dựng kho dữ liệu từ vựng song ngữ Việt – Bhnong”³, tác giả đã tìm hiểu về cơ sở công nghệ để xây dựng ứng dụng từ điển điện tử: công nghệ dotNET, mô hình ADO.NET, hệ quản trị cơ sở dữ liệu SQL Server và tìm hiểu về cơ sở dữ liệu từ vựng đa ngữ, các công cụ trợ giúp xây dựng CSDL, các kỹ thuật cập nhật dữ liệu, qua đó tác giả đã đi sâu tìm hiểu về XML và đã chọn XML để xây dựng CSDL kho ngữ vựng. Tại nghiên cứu “Xây

² Luận văn “Xây dựng kho dữ liệu song ngữ Việt - Cơ tu phục vụ tra cứu văn hóa dân tộc Cơ tu”, 2013, Đỗ Gia Trinh

³ Luận văn “Xây dựng kho dữ liệu từ vựng song ngữ Việt – Bhnong”, 2011, Nguyễn Văn Toàn

dựng môi trường xử lý tiếng Ê Đê ứng dụng trong dạy và học tiếng Ê Đê”⁴, tác giả đã xây dựng được kho ngữ vựng Việt-Ê Đê được xây dựng dựa trên mô hình hợp nhất nguồn dữ liệu từ điển giấy Việt-Ê Đê và Ê Đê-Việt. Định hướng phát triển kho ngữ vựng thông qua trang web quản lý kho ngữ vựng để các chuyên gia ngôn ngữ Ê Đê cùng tham gia đóng góp, cập nhật dữ liệu vào kho ngữ vựng, góp phần nâng cao chất lượng kho ngữ vựng Việt-Ê Đê. Tuy nhiên, trong tỉnh Quảng Ngãi chưa có bộ Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số tiếng Hrê, Co nào để phục vụ học tập, tra cứu tiếng Hrê, Co trên địa bàn tỉnh Quảng Ngãi.

Mặt khác, với xu hướng chuyển đổi số hiện nay, chúng ta có thể hoàn toàn tận dụng và ứng dụng công nghệ để hỗ trợ việc học và tự học, qua đó tiết kiệm được rất nhiều chi phí, công sức, tạo điều kiện cho rất nhiều người có nhu cầu tự học, tự nghiên cứu về tiếng đồng bào dân tộc thiểu số của tỉnh, mà phổ biến nhất trên địa bàn tỉnh hiện nay là tiếng Hrê và tiếng Co. Đặc biệt, các công nghệ của cuộc cách mạng khoa học công nghệ lần thứ Tư đang mang lại nhiều lợi ích to lớn cho mọi mặt của đời sống xã hội, trong đó công nghệ IOT – internet kết nối vạn vật và AI – trí tuệ nhân tạo nếu được ứng dụng vào lĩnh vực giáo dục – ngôn ngữ sẽ mang lại lợi ích hết sức thiết thực cho xã hội, góp phần đẩy mạnh việc học và tự học, xây dựng xã hội học tập trên nền tảng số.

Vì vậy, việc nghiên cứu, xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co là hết sức cần thiết nhằm thu hẹp khoảng cách giao tiếp, ngôn ngữ, giúp cho người đồng bào dân tộc thiểu số học hỏi, nâng cao kiến thức, hòa nhập với sự phát triển chung của tỉnh, của đất nước; đồng thời cũng giúp cho các cá nhân người Kinh đang làm việc với người đồng bào có thể hiểu và giao tiếp thuận lợi hơn, qua đó truyền đạt các kiến thức, các chủ trương, chính sách của chính quyền đến người đồng bào, tiếp thu kinh nghiệm và các vấn đề của người đồng bào được sâu sắc hơn. Bên cạnh đó, việc xây dựng CSDL điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co cũng là một hình thức số hoá tiếng đồng bào dân tộc thiểu số của tỉnh Quảng Ngãi. Việc thu thập, lưu trữ để xây dựng kho dữ liệu ngữ vựng dạng số của tiếng Hrê, Co không chỉ giúp cho người bản ngữ có ý thức bảo tồn, phát triển tiếng mẹ đẻ, mà còn hình thành một kho tài nguyên thông tin đầy đủ, chân thực, giúp cho các thế hệ nghiên cứu hiện tại và sau này có tư liệu chính xác. Đồng thời, góp phần gìn giữ cho các thế hệ sau không chỉ ngôn ngữ mà còn những giá trị văn hoá truyền thống đặc sắc của người đồng bào dân tộc tỉnh Quảng Ngãi.

Đề tài đã được phê duyệt tại Quyết định số 2109/QĐ-UBND ngày 23/12/2021 của Chủ tịch UBND tỉnh Quảng Ngãi về việc phê duyệt danh mục nhiệm vụ khoa học và công nghệ cấp tỉnh triển khai thực hiện năm 2022:

- Tên đề tài: **Xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co**

⁴ Luận án Tiến sỹ kỹ thuật “Xây dựng môi trường xử lý tiếng Ê Đê ứng dụng trong dạy và học tiếng Ê Đê”, 2017, Hoàng Thị Mỹ Lê

- Cấp quản lý: **Cấp tỉnh**
- Thời gian thực hiện: **18 tháng** (Từ tháng 02/2023 đến tháng 08/2024)
- Tổ chức chủ trì đề tài: **Trung tâm Công nghệ Thông tin và Truyền thông Quảng Ngãi**
- Tổ chức phối hợp chính thực hiện đề tài: **Ban Dân tộc tỉnh Quảng Ngãi**
- Chủ nhiệm đề tài: **ThS. Trần Duy Linh – ThS. Trần Văn Mẫn**

2. Đặt vấn đề nghiên cứu và kết quả cần đạt được

Đề tài **Xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co** thực hiện và hoàn thành bốn nhiệm vụ lớn đặt ra, gồm:

Nhiệm vụ thứ nhất: Xây dựng hệ thống dữ liệu - kho ngữ vựng Việt – Hrê, Việt – Co: Trên cơ sở nguồn dữ liệu tiếng Hrê và tiếng Co tại 02 bộ tài liệu đã được UBND phê duyệt: “Tài liệu Đào tạo, bồi dưỡng tiếng Hrê” (dùng cho cán bộ công chức tại miền núi và công tác dân tộc tỉnh Quảng Ngãi) của UBND tỉnh Quảng Ngãi (ban hành tạm thời kèm theo Quyết định số 893/QĐ-UBND ngày 27/4/2007 của UBND tỉnh Quảng Ngãi) và Tài liệu “BAI HOC APOK KOOL (BÀI HỌC TIẾNG CO)” (được UBND tỉnh cho phép tại công văn số 4796/UBND-NC ngày 27/8/2019), Đề tài đã tiếp cận và nghiên cứu toàn bộ vốn từ Việt, Hrê và Co cơ bản thường xuyên sử dụng trong cuộc sống trong 02 bộ tài liệu. Từ đó, thực hiện việc số hóa đồng nhất giữa các từ tương ứng để thành lập một kho dữ liệu Việt – Hrê, Việt – Co. Cơ sở dữ liệu gồm hơn 3.000 từ gồm dạng viết (text) và dạng âm thanh:

- Kho ngữ vựng song ngữ Việt – Hrê và ngược lại;
- Kho ngữ vựng song ngữ Việt – Co và ngược lại;

Nhiệm vụ thứ hai: Xây dựng công cụ hỗ trợ gõ tiếng dân tộc (Hrê, Co) trên giao diện phần mềm: Trong phạm vi kết quả của đề tài bao gồm: Cơ sở dữ liệu điện tử Việt – Hrê và ngược lại (Hrê – Việt); Cơ sở dữ liệu điện tử Việt – Co và ngược lại (Co – Việt). Đối với tiếng Việt đã có các công cụ hỗ trợ gõ tiếng Việt tích hợp font Unicode nên việc sử dụng để tra từ trên giao diện phần mềm dễ dàng. Đối với tiếng Hrê và Co, hiện chưa có công cụ hỗ trợ gõ có tích hợp font Unicode, vì vậy để tra cứu từ Hrê – Việt, Co – Việt phải nghiên cứu và xây dựng công cụ cho phép gõ tiếng Hrê, Co bằng font Unicode trực tiếp trên giao diện phần mềm. Qua quá trình nghiên cứu, thực hiện, nhóm thực hiện đề tài đã hoàn thành xây dựng công cụ hỗ trợ gõ tiếng Hrê và Co với các font chữ Unicode tích hợp vào phần mềm để sử dụng trong tra cứu từ Hrê – Việt, Co – Việt.

Nhiệm vụ thứ ba: Xây dựng bộ phần mềm Cơ sở dữ liệu điện tử Việt – Hrê, Việt – Co: Nhờ vào kho dữ liệu đã xây dựng, đề tài sẽ phát triển bộ phần mềm Cơ sở dữ liệu điện tử có thể sử dụng được trên các thiết bị điện tử như: máy tính, điện thoại thông minh... nhằm mục đích tra ngữ nghĩa giữa các ngôn ngữ Việt – Hrê, Việt – Co. Bên cạnh đó, phần mềm còn trang bị thêm một kênh

tường tác để người dùng có thể đóng góp thêm một số từ chưa có trong cơ sở dữ liệu.

Bộ phần mềm Cơ sở dữ liệu điện tử Việt – Hrê, Việt – Co đã được xây dựng hoàn thành, gồm phiên bản chạy trên web và ứng dụng (app) chạy trên thiết bị di động.

Nhiệm vụ thứ tư: Ứng dụng trí tuệ nhân tạo (AI) để xây dựng hệ thống nhận diện giọng nói tiếng dân tộc và tích hợp vào hệ thống CSDL điện tử: Hiện nay, trí tuệ nhân tạo (AI) đã có nhiều ứng dụng rộng rãi trong nhiều lĩnh vực, trong đó có lĩnh vực từ điển điện tử với công nghệ nhận diện giọng nói (SpeechToText). Tuy nhiên, việc nhận diện giọng nói là tiếng đồng bào dân tộc thiểu số Hrê, Co thì chưa có trước đây nên đây là một nghiên cứu mới, vừa mang lại hiệu quả thiết thực với sản phẩm của Đề tài và sẽ nền tảng cho các nghiên cứu khác liên quan đến ứng dụng công nghệ vào lĩnh vực ngôn ngữ. Nhóm thực hiện đã nghiên cứu thành công mô hình huấn luyện dữ liệu nhận diện tiếng Hrê, Co và đã tích hợp thành công kết quả vào các sản phẩm phần mềm (phiên bản web và phiên bản app), cho phép tra cứu bằng âm thanh, giọng nói.

Như vậy, qua quá trình đề xuất nội dung và gần 18 tháng nghiên cứu, thực hiện, nhóm thực hiện Đề tài đã hoàn thành bốn nhiệm vụ lớn đã đặt ra, với các sản phẩm đạt yêu cầu đã đặt ra, gồm:

- Sản phẩm kho ngữ vựng Việt – Hrê, Việt – Co với **trên 3.000 từ**, có phát âm từ giọng nam và giọng nữ;
- Sản phẩm **công cụ hỗ trợ gõ tiếng dân tộc (Hrê, Co)**;
- Sản phẩm **bộ phần mềm Cơ sở dữ liệu điện tử Việt – Hrê, Việt – Co**, với hai phiên bản là bản web “**Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co**”, được cài đặt trên hệ thống máy chủ do đơn vị chủ trì vận hành, hoạt động trên môi trường mạng tại tên miền **<https://csdlhreco.nuian.vn>** và **app Hrê – Co – Việt** cho thiết bị di động trên kho ứng dụng App Store và Google Play cho hai dòng máy sử dụng hệ điều hành iOS và Android.
- Sản phẩm Ứng dụng trí tuệ nhân tạo (AI) để xây dựng hệ thống nhận diện giọng nói tiếng dân tộc và tích hợp vào hệ thống CSDL điện tử: đã được nghiên cứu thành công và tích hợp vào cả hai phiên bản web và app, cho phép tra cứu từ bằng âm thanh tiếng Hrê, tiếng Co.

Ngoài ra, còn có các báo cáo khoa học và công nghệ theo từng chuyên đề nghiên cứu, các bài báo khoa học công nghệ được đăng tải trên các tạp chí uy tín của ngành, phim tư liệu để quảng bá sản phẩm.

II. KẾT QUẢ TRIỂN KHAI THỬ NGHIỆM VÀ ĐỊNH HƯỚNG PHÁT TRIỂN, NHÂN RỘNG SẢN PHẨM

1. Triển khai thử nghiệm

Để triển khai thử nghiệm và phát huy hiệu quả của sản phẩm Đề tài đưa sản phẩm đi vào sử dụng trong thực tế, nhóm nghiên cứu đã thực hiện:

Một là, hướng dẫn khai thác Cơ sở dữ liệu điện tử Hrê, Co thông qua các lớp tập huấn tại 5 huyện miền núi của tỉnh: đơn vị chủ trì thực hiện Đề tài đã phối hợp với Ban dân tộc tỉnh tổ chức 05 lớp tập huấn hướng dẫn quản lý và hướng dẫn khai thác CSDL điện tử cho các đối tượng sử dụng, gồm: cán bộ công chức tại phòng Dân tộc các huyện, thành phố, cán bộ thôn bản, cán bộ Đoàn thanh niên, phụ nữ, hội nông dân xã, huyện miền núi; giáo viên và học sinh miền núi của tỉnh tại các huyện: Ba Tơ, Minh Long, Trà Bông, Sơn Hà, Sơn Tây; 01 lớp hướng dẫn quản trị, vận hành hệ thống cho Ban Dân tộc tỉnh và phòng Dân tộc các huyện, thành phố.

Các lớp tập huấn đã có gần 300 học viên được giới thiệu về sản phẩm (gồm phiên bản web và app), được hướng dẫn các thao tác để khai thác, sử dụng các tính năng của CSDL điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co, biết cách gửi đóng góp về CSDL điện tử. Qua các lớp tập huấn, nhóm thực hiện Đề tài cũng được tiếp xúc, trao đổi trực tiếp với đa dạng các đối tượng sử dụng, qua đó nắm bắt thêm những ý kiến, góp ý từ các học viên để hoàn thiện thêm sản phẩm, giúp sản phẩm ngày càng dễ dùng hơn, hiệu quả hơn với đông đảo người dân. Những lớp học này không chỉ giúp các đối tượng tham gia nắm vững cách sử dụng sản phẩm của Đề tài mà còn đóng vai trò quan trọng trong việc chia sẻ, lan truyền ứng dụng này đến cộng đồng.

Hai là, tuyên truyền, quảng bá sản phẩm trên các phương tiện truyền thông, nền tảng số: Đơn vị chủ trì chủ động trong việc quảng bá và tuyên truyền sản phẩm Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co trên các phương tiện truyền thông, nền tảng số như Zalo, Facebook, các trang thông tin điện tử và các nền tảng số khác. Nhờ vào đó, thu hút được sự quan tâm của đông đảo cộng đồng mạng và các nhóm quan tâm đến bảo tồn và phát triển ngôn ngữ dân tộc.

Bà là, giới thiệu đến Tổ Công nghệ số cộng đồng: Trên địa bàn tỉnh Quảng Ngãi, hiện đã thành lập tổ công nghệ số cộng đồng đến cấp thôn, với tổng cộng hơn **1.140 Tổ Công nghệ số cộng đồng tại 173 xã, phường, thị trấn và 954 thôn, tổ dân phố trong tỉnh với hơn 7.500 thành viên tham gia.** Thành viên Tổ Công nghệ số cộng đồng là những hạt nhân tích cực, có sứ mệnh đưa công nghệ số đến mọi ngõ ngách của cuộc sống, bằng cách tiếp cận với các công nghệ số, ứng dụng số hữu ích và triển khai, hướng dẫn lại cho người dân tại thôn, bản, tổ dân phố mình nắm bắt, sử dụng.

Nắm bắt được vai trò, lợi thế và khả năng lan tỏa mạng lưới này tại địa phương, nhóm thực hiện Đề tài đã giới thiệu sản phẩm Cơ sở dữ liệu Hrê Co đến với mạng lưới Tổ Công nghệ số cộng đồng, xem nơi đây là phương thức quan trọng trong việc phổ biến sản phẩm đến các cộng đồng người dân trên toàn tỉnh. Hình thức triển khai giới thiệu sản phẩm đến mạng lưới Tổ Công nghệ số cộng đồng sẽ là thiết kế tờ rơi dạng giấy và poster dạng điện tử để phát và gửi đến thành viên của tổ. Tờ rơi, poster cần thể hiện được thông tin giới thiệu của sản

phẩm, cách truy cập để cài đặt, khai thác, sử dụng. Các nội dung trên tờ rơi, poster được thể hiện ngắn gọn, dễ hiểu, dễ làm để phục vụ đối tượng người sử dụng là người dân.

Bốn là, kết hợp giới thiệu, triển khai trong các lớp bồi dưỡng tiếng Hrê, Co: Hằng năm, Sở Nội vụ phối hợp Trường Đại học Phạm Văn Đồng tổ chức các lớp bồi dưỡng tiếng đồng bào dân tộc thiểu số Hrê, Co cho các đối tượng cán bộ công chức, viên chức, giáo viên làm việc tại các huyện miền núi.

Trong các thành viên thực hiện chính của Đề tài, có 03 thành viên hiện đang công tác tại trường Đại học Phạm Văn Đồng, trong đó có 01 thành viên có chuyên môn về ngôn ngữ; đồng thời có 02 thành viên là chuyên gia về ngôn ngữ Hrê và Co, hiện cũng đang làm báo cáo viên giảng dạy tại các lớp bồi dưỡng nêu trên. Thông qua các đầu mối là thành viên thực hiện của Đề tài, sản phẩm sẽ dễ dàng kết nối vào công tác dạy và học tại trường, giới thiệu cho học viên khai thác trong các lớp bồi dưỡng tiếng Hrê, Co. Thông qua đó, vừa giúp công tác đào tạo tập huấn của Trường được sinh động, thuận tiện, hiệu quả hơn, đồng thời góp phần lan tỏa, nâng cao tính dụng của sản phẩm Đề tài.

Năm là, kết hợp giới thiệu, triển khai trong các lớp tập huấn công tác dân tộc: Đơn vị chủ trì thực hiện Đề tài sẽ phối hợp với Ban Dân tộc tỉnh, phòng dân tộc các huyện có người đồng bào dân tộc thiểu số sinh sống để đặt banner liên kết đến sản phẩm Đề tài trên các trang thông tin điện tử của Ban Dân tộc tỉnh. Đồng thời, đơn vị phối hợp thực hiện Đề tài là Ban Dân tộc tỉnh đã và sẽ kết hợp giới thiệu và triển khai sản phẩm cơ sở dữ liệu Hrê Co trong các lớp tập huấn công tác dân tộc hằng năm do Ban Dân tộc tỉnh hoặc phòng dân tộc các huyện tổ chức, giúp mở rộng phạm vi ứng dụng của Đề tài đến các địa phương trong thời gian sau khi Đề tài kết thúc. Thông qua đó, góp phần tăng cường sự phát triển và bảo tồn ngôn ngữ và văn hóa của người dân tộc Hrê, Co.

Sáu là, giới thiệu trong các Hội thảo chuyên đề về ứng dụng công nghệ thông tin, công nghệ trí tuệ nhân tạo AI trong xử lý ngôn ngữ hoặc chuyên đề về nghiên cứu, bảo tồn ngôn ngữ tiếng đồng bào dân tộc thiểu số: Trong phạm vi của Đề tài, có tổ chức 01 Hội thảo cấp tỉnh để giới thiệu về kết quả của Đề tài. Hội thảo mời đại biểu từ Ban Dân tộc một số tỉnh lân cận và Viện Ngôn ngữ học để tăng cường giới thiệu kết quả của Đề tài đến phạm vi bên ngoài tỉnh. Bên cạnh đó, có thể nhận báo cáo tham luận tại các Hội thảo chuyên đề về một số chủ đề mà Đề tài đã nghiên cứu thành công như: ứng dụng công nghệ trí tuệ nhân tạo AI trong nhận diện giọng nói tiếng đồng bào dân tộc thiểu số Hrê, Co.

Thông qua việc triển khai các mô hình ứng dụng sản phẩm, đến nay, web “**Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co**” sự quan tâm, khai thác sử dụng của hơn **32.000 lượt** truy cập, và **app Hre – Co – Việt** đã có hơn **400 lượt tải** và cài đặt ứng dụng.

Phần lớn sau khi triển khai thử nghiệm, thông qua trực tiếp (lớp tập huấn), gián tiếp (qua nền tảng số), đồng đảo người sử dụng cho biết sản phẩm rất dễ sử dụng và hữu ích, giúp việc học tiếng Hrê, Co trở nên thuận tiện và dễ dàng hơn bao giờ hết. Đồng thời, người sử dụng cũng góp ý thêm để hoàn thiện sản phẩm

như: có thể bổ sung các từ ghép từ các từ đơn, bổ sung chức năng cho tra cứu từ theo lĩnh vực để phục vụ tốt hơn nữa nhu cầu sử dụng của người dùng. Ngoài ra đối với tiếng Hrê, có thể bổ sung thêm ghi chú vùng (Ba Tơ, Minh Long, Sơn Hà, Sơn Tây) sử dụng đối với một số từ có sự khác biệt giữa các vùng để giúp người tra cứu hiểu hơn về tính đa dạng của ngôn ngữ tiếng đồng bào dân tộc thiểu số (có một số khác biệt giữa tiếng Hrê “*đường nước*” và “*đường rừng*”).

2. Định hướng phát triển, nhân rộng kết quả Đề tài

Trong phạm vi thực hiện Đề tài, nhóm thực hiện đã thu thập và đưa vào cơ sở dữ liệu trên **3.000** từ tiếng Hrê và tiếng Co. Để tiếp tục phát huy, tiếp tục thu thập, làm giàu dữ liệu tiếng Hrê, Co thông qua kênh tiếp nhận các đóng góp cho cơ sở dữ liệu từ nguồn người sử dụng. Tính năng “Đóng góp cho CSDL” đã được xây dựng trên cả hai phiên bản web và app của phần mềm cơ sở dữ liệu điện tử, cho phép người sử dụng có thể dễ dàng nhập và gửi các đóng góp là chữ viết, hình ảnh, âm thanh về hệ thống. Từ nguồn dữ liệu thu thập phong phú và rộng rãi trong cộng đồng, sẽ là cơ sở để nhóm thực hiện có thể biên tập, chuẩn hóa, bổ sung, làm giàu cho cơ sở dữ liệu.

Những nghiên cứu thành công về kỹ thuật, công nghệ của Đề tài như ứng dụng công nghệ AI nhận diện giọng nói tiếng Hrê, Co; phương pháp để tích hợp bộ mã Unicode để xây dựng bộ gõ tiếng đồng bào dân tộc thiểu số Hrê, Co; ứng dụng thuật toán khoảng cách Levenshtein để tìm và gợi ý từ gần đúng khi tra cứu,... cũng gợi mở cho các nghiên cứu khoa học tiếp theo về ứng dụng công nghệ AI nhận diện giọng nói vào lĩnh vực ngôn ngữ, ví dụ: Có thể ứng dụng **Generative AI – AI tạo sinh** để phát triển thêm tính năng nhận diện câu, dịch câu, dịch đoạn văn.

Bên cạnh đó, chúng tôi sẽ tiếp tục nghiên cứu, phát triển, nhân rộng kết quả đã đạt được như: ứng dụng trí tuệ nhân tạo trong Mô hình nhận dạng lời nói tiếng đồng bào dân tộc thiểu số, số hóa, xây dựng CSDL đối với các tiếng đồng bào dân tộc khác trên cả nước.

Như vậy, khả năng kế thừa, phát huy, nhân rộng kết quả của Đề tài “Xây dựng cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co” là rất rộng, hứa hẹn mang lại nhiều giá trị trên nhiều lĩnh vực cả ứng dụng lẫn nghiên cứu, cả ngôn ngữ lẫn công nghệ.

3. Đánh giá hiệu quả mang lại của Đề tài:

Sản phẩm của Đề tài “Xây dựng cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co” triển khai ứng dụng vào thực tế dự kiến sẽ mang lại nhiều hiệu quả trên nhiều lĩnh vực:

Hiệu quả về kinh tế - xã hội: Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co sẽ là công cụ quan trọng phục vụ việc học tập, tra cứu tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co một cách nhanh

chống, hiệu quả và tiết kiệm, phục vụ đông đảo các đối tượng có nhu cầu học và tìm hiểu tiếng đồng bào dân tộc thiểu số như: cán bộ công chức, viên chức, giáo viên, sinh viên, học sinh, người dân,... góp phần thiết thực giảm bất đồng ngôn ngữ, xóa bỏ dần khoảng cách giữa người Kinh và người H'rê, Co, tích cực phục vụ cho công tác tuyên truyền các chủ trương, triển khai các chính sách đến người đồng bào dân tộc thiểu số trên địa bàn tỉnh.

Hiệu quả về văn hóa: Đề tài góp phần kịp thời bảo tồn nguồn dữ liệu tiếng đồng bào dân tộc thiểu số tiếng H'rê, Co tại tỉnh Quảng Ngãi, tránh mai một bằng cách số hóa dữ liệu cả dạng văn bản và âm thanh.

Việc thu thập, lưu trữ để xây dựng kho dữ liệu ngữ vựng dạng số của tiếng H'rê, Co không chỉ giúp cho người bản ngữ có ý thức bảo tồn, phát triển tiếng mẹ đẻ, mà còn hình thành một kho tài nguyên thông tin đầy đủ, chân thực, giúp cho các thế hệ nghiên cứu hiện tại và sau này có tư liệu chính xác. Đồng thời, góp phần gìn giữ cho các thế hệ sau không chỉ ngôn ngữ mà còn những giá trị văn hoá truyền thống đặc sắc của người đồng bào dân tộc tỉnh Quảng Ngãi.

Hiệu quả về khoa học công nghệ: Đề tài “Xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – H'rê, Việt – Co” đã tạo ra một cơ sở dữ liệu số đầu tiên về ngôn ngữ các đồng bào dân tộc thiểu số ở tỉnh Quảng Ngãi, đồng thời là công cụ phục vụ đắc lực cho việc nghiên cứu, tra cứu, học tập,... góp phần thay đổi phương pháp học tập bằng cách ứng dụng công nghệ thông tin cũng như sử dụng các hệ thống thông tin điện tử.

Thông qua quá trình nghiên cứu, thực hiện các nội dung công việc của Đề tài, các thành viên tham gia thực hiện Đề tài đã được tiếp cận với những công nghệ mới, chuyên sâu về công nghệ thông tin và trí tuệ nhân tạo. Đây sẽ là nguồn nhân lực quan trọng của tỉnh hỗ trợ đắc lực cho sự phát triển khoa học công nghệ trong quá trình chuyển đổi số đang diễn ra hết sức mạnh mẽ hiện nay trên địa bàn tỉnh Quảng Ngãi.

4. Kết luận

Tóm lại, việc nghiên cứu, xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – H'rê, Việt – Co là cần thiết và sẽ mang lại nhiều lợi ích, giá trị to lớn cho cộng đồng. Đây sẽ là nơi lưu giữ ngôn ngữ truyền thống của các dân tộc thiểu số trên địa bàn tỉnh, đồng thời qua đó cũng sẽ lưu giữ các giá trị văn hóa truyền thống, hình ảnh, phong tục, tập quán và văn hoá của họ, góp phần bảo tồn và phát huy các giá trị truyền thống độc đáo của mỗi dân tộc, cùng hòa vào dòng chảy phát triển chung của tỉnh ta nói riêng và đất nước nói chung.