

Ảnh hưởng của sự thay đổi cách phát âm theo khu vực đối với cách đánh vần của trẻ em và lợi ích tiềm tàng của công cụ kiểm tra chính tả thích ứng với giọng điệu

Emma O'Neill
Trung tâm ADAPT
Đại học Cao đẳng Dublin
emma.l.oneill@ucdconnect.ie

Joe Kenny
Zeeko
Ireland
joe@zeeko.ie

Anthony Ventresque
Trung tâm nghiên cứu
Lero Đại học Cao đẳng Dublin
anthony.ventresque@ucd.ie

Julie Carson-Berndsen
Trung tâm ADAPT
Đại học Cao đẳng Dublin
julie.berndsen@ucd.ie

trường tượng

Một đứa trẻ không quen với cách viết đúng chính tả của một từ thường áp dụng phương pháp “phát âm”: chia từ đó thành các âm cấu thành và sau đó chọn các chữ cái để thể hiện các âm đã xác định. Điều này thường dẫn đến lỗi chính tả rất khác về mặt chính tả so với mục tiêu dự định. Gần đây, những nỗ lực đã được thực hiện để phát triển công cụ kiểm tra chính tả dựa trên ngữ âm nhằm giải quyết tình trạng sai chính tả ở trẻ em. Tuy nhiên, công việc nhỏ đã được thực hiện để điều tra tiềm năng của các công cụ sửa lỗi chính tả kết hợp các biến thể phát âm theo khu vực. Nếu trước tiên một đứa trẻ phải xác định các âm thanh tạo nên một từ thì rõ ràng cách phát âm của chúng sẽ ảnh hưởng đến quá trình này. Chúng tôi điều tra giả thuyết này cùng với tính khả thi và lợi ích tiềm năng của việc điều chỉnh các công cụ sửa lỗi chính tả cho phù hợp với các biến thể ngôn ngữ cụ thể hơn - đặc biệt là tiếng Anh có giọng Ailen. Chúng tôi sử dụng dữ liệu lỗi chính tả từ học sinh trên khắp Ireland để điều chỉnh công cụ kiểm tra chính tả dựa trên phiên âm tiếng Anh hiện có và chúng minh sự cải thiện về hiệu suất. Những kết quả này không chỉ nhắc nhở việc xem xét các dạng ngôn ngữ trong quá trình phát triển công cụ kiểm tra chính tả mà còn góp phần vào tài liệu hiện có về vai trò của giọng vùng miền trong việc đạt được khả năng viết thành thạo.

1. Giới thiệu

Cách đánh vần của trẻ em từ lâu đã được coi là bắt nguồn từ ngữ âm; bị ảnh hưởng bởi sự giống nhau về cách phát âm và giọng nói (Độc, 2018). Lỗi chính tả thường 'sáng tạo' ở chỗ chúng phản ánh khả năng phán đoán về ngữ âm của trẻ và xuất phát từ nỗ lực của trẻ trong việc sử dụng các chữ cái để thể hiện âm thanh trong ngôn ngữ nói của mình. Do đó, những lỗi chính tả này thường là duy nhất và có thể sai lệch nhiều so với nguyên tắc chính tả.

viết đúng chính tả khiến việc tự động sửa những lỗi như vậy trở thành một nhiệm vụ không hề đơn giản.

Ngày nay, trọng tâm phổ biến trong giáo dục đọc viết sớm là việc dạy ngữ âm với các trường học trên toàn thế giới nói tiếng Anh kết hợp phương pháp này vào chương trình giảng dạy của họ (Hội đồng Chương trình giảng dạy và Đánh giá Quốc gia, 2019; Bowers and Bowers, 2017). Các phương pháp đọc và viết dựa trên ngữ âm liên quan đến việc dạy mối quan hệ giữa các chữ cái và âm thanh. Ví dụ: từ 'cat' sẽ được chia thành các chữ cái 'c', 'a', 't' và các âm vị tương ứng /k/, /æ/, /t/ Khi xử lý cách viết của một từ không quen thuộc Sau đó, trẻ em được khuyến khích áp dụng phương pháp “phát âm” bằng cách xác định trình tự ngữ âm của từ và các chữ cái đại diện cho những âm thanh này. Cách tiếp cận này được những người đánh vần có thành tích thấp dựa nhiều vào (Daffern và Critten, 2017).

Bất chấp bản chất ngữ âm của cách đánh vần của trẻ em và việc sử dụng phương pháp giáo dục xóa mù chữ gắn liền với lời nói, rất ít công việc được thực hiện nhằm phát triển các công cụ kiểm tra chính tả có thể giải thích được sự biến đổi phát âm có hệ thống. Phần lớn nghiên cứu hiện tại về sự tương tác giữa biến thể giọng nói và khả năng đọc viết thường tập trung vào trẻ em có phương ngữ tiếng Anh của người Mỹ gốc Phi (AAE) và sự sai lệch về âm vị và hình thái so với cái được gọi là tiếng Anh Mỹ chính thống (MAE) (xem Phần 2.2). Trong đánh giá có hệ thống của họ, Snell và Andrews (2017) chỉ ra rằng chưa có đủ nghiên cứu về mối quan hệ giữa giọng hoặc phương ngữ khu vực và khả năng viết tiếng Anh ở Anh. Từ những công trình được xem xét lại, họ

¹Trong suốt bài viết này, chúng tôi đề cập đến các âm vị sử dụng ký hiệu ARPA-bet phù hợp với Từ điển Phát âm CMU (Weide, 1998) được sử dụng trong mô hình sửa lỗi chính tả.

phỏng đoán rằng không có mối quan hệ đơn giản giữa thành tích đọc viết và nền tảng ngôn ngữ và khẳng định rằng phương ngữ khu vực chỉ có một ảnh hưởng ít nhiều tới việc viết. Tuy nhiên, trong suốt ôn tập, những khó khăn về chính tả không được coi là liên quan đến sự biến đổi cách phát âm mà thay vào đó là kết quả của sự phức tạp giữa chính tả và cách phát âm tiếng Anh ảnh hưởng đến tất cả trẻ em bất kể giọng của họ. Trong khi chúng tôi đồng ý rằng tất cả trẻ em sẽ phải đối mặt với những thách thức trong việc đạt được khả năng đọc viết, chúng tôi tìm cách khám phá xem liệu khu vực sự khác biệt có thể ảnh hưởng đến loại vấn đề mà học sinh gặp phải và các loại lỗi chính tả kết quả là họ sản xuất. Thật vậy, [Terry \(2006\)](#) cho rằng "trong khi tất cả trẻ em phải học cách thương lượng sự không phù hợp giữa lời nói và chữ in để trở thành người đọc và viết giỏi, quá trình này có thể đặc biệt khó khăn đối với những trẻ có ngôn ngữ nói khác biệt đáng kể so với các hình thức viết thông thường". Theo cách tương tự như người ta có thể sử dụng một cách tiếp cận có mục tiêu để giảng dạy một ngoại ngữ dựa trên tiếng mẹ đẻ của người học ngôn ngữ, nó có thể mang lại lợi ích cho trẻ em mắc chứng các biến thể ngôn ngữ khu vực sẽ được dạy đọc và kỹ năng viết với sự đa dạng trong tâm trí và có quyền truy cập vào các công cụ được thiết kế đặc biệt để xử lý biến thể cụ thể của họ.

Bài viết này tìm cách nghiên cứu giả thuyết sự khác biệt trong khu vực trong cách phát âm ảnh hưởng viết sai chính tả trong bài viết của trẻ em. ĐẾN làm điều này, chúng tôi khám phá liệu một giọng có thích nghi không trình kiểm tra chính tả hoạt động tốt hơn hệ thống cơ bản sẽ gợi ý một số điểm chung cơ bản giữa các lỗi chính tả do những người nói cùng một biến thể ngôn ngữ. Khi làm như vậy, chúng tôi cũng điều tra tính khả thi của việc phát triển một hệ thống như vậy và những lợi ích tiềm tàng mà nó mang lại cho các diễn giả của một biến thể khu vực. Công cụ sửa lỗi chính tả dựa trên ngữ âm tiếng Anh hiện có đã được tinh chỉnh sử dụng dữ liệu lỗi chính tả được thu thập từ học sinh trên khắp Ireland. Chúng tôi chứng minh rằng người Ireland này Mô hình điều chỉnh bằng tiếng Anh có dấu (IAE) đạt được kết quả tốt hơn trong bộ thử nghiệm lỗi chính tả IAE của chúng tôi. Hơn nữa, chúng tôi chứng minh rằng những lợi ích này trong hiệu suất không chỉ là kết quả của việc đào tạo bổ sung, bằng cách tinh chỉnh mô hình cơ bản với Dữ liệu sai chính tả của IAE cho thấy hiệu suất đạt được ít hơn. Những kết quả này sẽ gợi ý rằng mối quan hệ giữa sự biến đổi cách phát âm theo vùng và khả năng đọc viết đáng được khám phá thêm.

2 công việc liên quan

2.1 Phương pháp sửa lỗi chính tả

[Kukich \(1992\)](#) cho rằng lỗi chính tả thuộc một trong hai loại; kiểu chữ và nhận thức. Điều đầu tiên xảy ra khi người viết biết chính xác viết đúng chính tả của một từ nhưng lại mắc lỗi khi sản xuất nó. Ví dụ: nhập một địa chỉ khác ký tự hơn dự định do nhấn nhầm một phím liên kề trên bàn phím. Những loại lỗi này chỉ có xu hướng sai lệch so với chính tả mục tiêu bằng một thao tác chỉnh sửa duy nhất (thay thế, xóa, chèn hoặc chuyển vị) ([Damerau, 1964](#)) và do đó có thể dễ dàng sửa chữa bằng các phương pháp khoảng cách chỉnh sửa ký tự được sử dụng bởi những người kiểm tra chính tả truyền thống. Ngược lại, lỗi nhận thức là kết quả của việc thiếu biết cách đánh vần một từ cho đúng. Lỗi chính tả do cố gắng nắm bắt thông tin chuỗi âm thanh của một từ thuộc một tập hợp con của lỗi nhận thức được gắn nhãn là lỗi ngữ âm và chúng thường đi chệch hướng đáng kể so với từ mục tiêu ([Kukich, 1992](#)).

Các mô hình hiệu chỉnh xác suất, như của [Brill và Moore \(2000\)](#) hoặc [Church và Gale \(1991\)](#), đã cải thiện hiệu suất đối với các lỗi nhận thức bằng cách lập mô hình khả năng thực hiện nhiều thao tác chỉnh sửa. Việc sửa lỗi ngữ âm nói riêng đã được giải quyết bằng cách kết hợp phát âm thông tin trái ngược với việc chỉ biểu diễn chính tả. Ví dụ: bằng cách sử dụng chỉnh sửa có trọng số khoảng cách xem xét sự giống nhau về mặt ngữ âm giữa các biểu đồ ([Veronis, 1988](#)), bằng cách gán một mã chữ và số dùng để ghi lại phiên âm các tính năng ([Russell và Odell, 1918](#)), hoặc bằng cách chuyển đổi lỗi chính tả sang âm vị tương ứng của nó trình tự sử dụng các quy tắc chuyển âm thanh từ chữ cái sang âm thanh ([Fisher, 1999; Toutanova và Moore, 2002](#)). Trong cuộc khảo sát gần đây của họ, [Hládek và cộng sự. \(2020\)](#) ghi nhận sự phổ biến ngày càng tăng của kiến trúc bộ mã hóa-giải mã và mạng lưới thần kinh sâu mạng xử lý vấn đề sửa lỗi chính tả là một trong những dịch máy thống kê.

Các công cụ sửa lỗi chính tả dành riêng cho trẻ em cũng đã được phát triển gần đây.

[Downs và cộng sự. \(2020\)](#) đã phát hành Kidspell: một công cụ kiểm tra chính tả ngữ âm, dựa trên quy tắc, hướng đến trẻ em. Của họ hệ thống sử dụng các quy tắc ngữ âm của tiếng Anh để ánh xạ các chữ cái tới các phím nhằm mục đích nắm bắt chính xác biểu đạt ngữ âm. Đề xuất ứng viên để sửa lỗi chính tả được tạo ra bằng cách xác định các từ có ngữ âm phù hợp hoặc tương tự. Tương tự, [S-capade \(O'Neill và cộng sự, 2020\)](#), công cụ kiểm tra chính tả tiếng Anh dành cho trẻ em của chúng tôi, tạo ra ứng cử viên

những chỉnh sửa có âm vị giống hoặc tương tự trình tự dự đoán của từ sai chính tả.

Nó tạo ra một biểu diễn âm vị của lỗi chính tả và tính toán khoảng cách chỉnh sửa có trọng số giữa lỗi chính tả và cách sửa lỗi dựa trên khả năng các âm vị được chèn, xóa hoặc thay thế cho người khác. Bản chất của phương pháp chỉnh sửa khoảng cách của S-capade cho phép điều chỉnh trọng lượng âm vị và chúng tôi sử dụng điều này làm cơ sở để tinh chỉnh dữ liệu của người nói IAE trong công việc này.

2.2 Cách nói và khả năng đọc viết

Sự mua lại

Nghiên cứu ban đầu về các yếu tố ảnh hưởng đến khả năng đọc viết của trẻ em đã phỏng đoán rằng việc nói một loại ngôn ngữ "không chuẩn" hoặc "không chính thống" cản trở hiệu suất đọc và viết với Schwartz (1982) đặt ra thuật ngữ "giao thoa phương ngữ". Gần đây hơn, một giải pháp thay thế (mặc dù không loại trừ lẫn nhau) giải thích cho mối liên hệ giữa các sản phẩm không chính thống và thành tích biết đọc biết viết đã trở nên phổ biến ở văn học. Điều này liên quan đến ý tưởng rằng những đứa trẻ tạo ra nhiều hình thức không chính thống hơn, đặc biệt là trong những bối cảnh mà điều này sẽ không phù hợp (tức là trong lớp học), có khả năng có ít ngôn ngữ hơn nhận thức nói chung. Sự thiếu nhận thức này hoặc tính linh hoạt có khả năng mở rộng sang các khía cạnh khác bao gồm nhận thức về âm vị học được coi là không thể thiếu trong quá trình tiếp thu kiến thức (Terry và Scarborough, 2011). Vì vậy, đó không phải là việc sử dụng các hình thức không chính thống này hoặc cách sử dụng phương ngữ không chuẩn trong nói chung có tác động tiêu cực đến việc đọc và viết mà đúng hơn là tần suất sử dụng cao là dấu hiệu cho thấy điểm yếu tiềm ẩn về ngôn ngữ

cũng ảnh hưởng đến khả năng đọc viết.

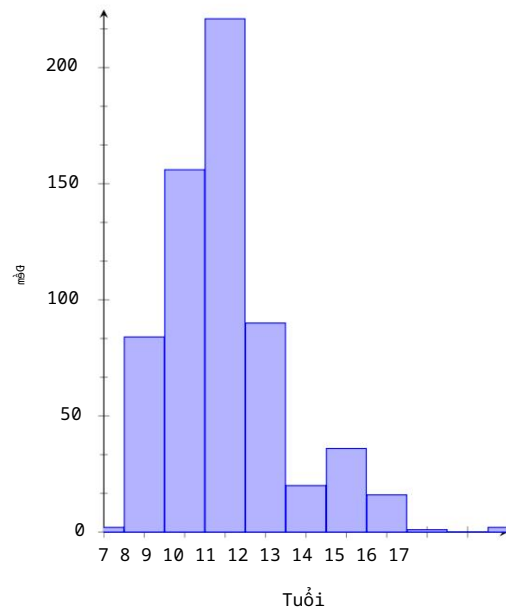
Một số công trình đã nghiên cứu khả năng đọc viết kỹ năng của trẻ nói AAE - đặc biệt là những với những hình thức không chính thống thường xuyên xuất hiện trong lời nói của họ (Charity và cộng sự, 2004; Terry, 2006; Terry và Scarborough, 2011; Terry, 2012; Terry và Connor, 2010; Connor và Craig, 2006). Cái này nghiên cứu xác nhận rằng những đứa trẻ thường xuyên sử dụng biến thể AAE thường gặp nhiều rắc rối hơn với học đọc và viết. Đặc biệt, Terry và Connor (2010) đã chứng minh rằng các từ có đặc điểm nhạy cảm với phương ngữ khác nhau gây ra các vấn đề về chính tả trong vừa gặp khó khăn vừa thường thu hút được đọc giả. Ngoài việc kiểm tra AAE, nghiên cứu hỗ trợ tương tự về ngôn ngữ nói và giáo dục

đã được thực hiện liên quan đến các biến thể ngôn ngữ tiếng Creole của Hà Lan, Ả Rập, Hy Lạp và Carribean (Driessen và Withagen, 1999; Saiegh-Haddad, 2003; Siegel, 1999; Yiakoumetti, 2007). Đây là rõ ràng là một vấn đề không chỉ giới hạn ở giống AAE hoặc thậm chí với nhiều loại tiếng Anh. Như vậy, mặc dù công việc này tập trung vào IAE, việc điều chỉnh các công cụ ngôn ngữ thành các biến thể nói có thể mang lại lợi ích khắp các ngôn ngữ trên thế giới.

3. Phương pháp luận

3.1 Bộ dữ liệu

Tập dữ liệu sử dụng trong thí nghiệm này được biên soạn từ một bộ sưu tập các cuộc khảo sát học sinh trên khắp Ireland được thực hiện bởi Zeeko (Everri và Park, 2018). Một tập hợp con của các khảo sát này bao gồm các câu hỏi mở có dạng "Tại sao bạn nghĩ...?" hoặc "Bạn thích điều gì ở...?" Và người tham gia trả lời ở dạng văn bản tự do. Tổng cộng, các câu trả lời khảo sát từ 628 sinh viên bao gồm những câu trả lời bằng văn bản tự do này và được phân tích để tìm các lỗi chính tả không phải từ. Học sinh có độ tuổi khác nhau từ 7 đến 17 tuổi với độ tuổi trung bình là 10 tuổi. Việc phân bổ đầy đủ độ tuổi của người trả lời có thể được thấy trong Hình 1.



Hình 1: Phân bố độ tuổi của người trả lời khảo sát.

Mặc dù đã nhận được phản hồi từ các địa điểm trên toàn quốc, 50,6% số người được hỏi đến từ trường học ở Quận Dublin. Như vậy, những phân tích sau này ảnh hưởng của cách phát âm theo vùng được chú trọng nhiều đối với các dạng tiếng Anh của Dublin. Thu thập bổ sung và đa dạng hơn về mặt địa lý

dữ liệu được coi là một lĩnh vực công việc trong tương lai.

Các lỗi chính tả không phải từ được trích xuất từ các câu trả lời và người chú thích sẽ đánh giá các mục tiêu từ thực tương ứng của chúng. Trong trường hợp không thể xác định được mục tiêu từ ngữ cảnh, lỗi chính tả sẽ bị xóa. Điều này dẫn đến một kho văn bản gồm 232 cặp lỗi chính tả và các mục tiêu từ thực từ đó được gọi là kho văn bản IAE. Ví dụ về một số lỗi chính tả: cặp mục tiêu có thể được xem trong Bảng 1.

Viết sai chính tả	Target
thực sự là bekos vì	
khó chịu	khác biệt
egicasinol	giáo dục
mishon	Sứ mệnh
tình hình sichweshen	

Bảng 1: Ví dụ về lỗi chính tả và mục tiêu từ thực từ kho dữ liệu IAE.

3.2 Ma trận khoảng cách âm vị của S-capade

Các chỉnh sửa của ứng viên và thứ hạng của họ do công cụ S-capade gợi ý được xác định bằng cách sử dụng thước đo khoảng cách chỉnh sửa có trọng số. Công cụ tạo biểu đồ thành âm vị (CMUSphinx, 2016) giữa chuỗi âm vị được dự đoán của lỗi chính tả và chuỗi âm vị của ứng viên. Theo cách tiếp cận tương tự như của Wagner và Fischer (1974), các chuỗi được căn chỉnh tối ưu để có tổng khoảng cách chỉnh sửa thấp nhất được tính bằng cách tính tổng chi phí của mỗi lần thay thế, xóa và chèn.

Chi phí của các hoạt động này được lấy từ ma trận khoảng cách âm vị mô hình hóa sự giống nhau giữa các âm vị. Vì sự giống nhau có thể được coi là một hàm của khả năng nhầm lẫn (Gallagher và Graff, 2012), nên ma trận khoảng cách này được tạo ra dựa trên khả năng nhầm lẫn của các âm vị cả về mặt âm học và phân phối (Kane và Carson-Berndsen, 2016; O'Neill và Carson-Berndsen, 2019). Nếu hai âm vị có khả năng bị nhầm lẫn thì chúng được coi là rất giống nhau và do đó có điểm khoảng cách thấp và chi phí thay thế thấp.

Việc xóa và thêm vào được coi là sự thay thế của một âm vị bằng chuỗi trống và ngược lại. Các giá trị khoảng cách giữa các âm vị và chuỗi trống được dựa trên tài liệu hiện có về phép kết âm và phép chêm (Collins và Mees, 2013; Fourakis và Port, 1986; Gimson và Ramsaran, 1970; Itô, 1989; Yip, 1987).

Nếu một lỗi chính tả có cùng một chuỗi âm vị

vì mục tiêu từ thực của nó thì khoảng cách chỉnh sửa sẽ là 0 và mục tiêu có thể sẽ là ứng viên điều chỉnh được xếp hạng cao nhất. Tuy nhiên, nếu chuỗi âm vị được dự đoán của lỗi chính tả không khớp với bất kỳ từ thực tế nào, thì từ sửa được đề xuất sẽ là từ được cho là giống nhất do

điểm khoảng cách chỉnh sửa thấp (O'Neill và cộng sự, 2020).

Các giá trị trong ma trận khoảng cách không đặc trưng cho một dạng nói mà thay vào đó mô hình hóa sự nhầm lẫn về âm vị chung và quá trình âm vị học của tiếng Anh. Tuy nhiên, cách hiểu khác nhau về âm vị trong một biến thể của tiếng Anh có thể dẫn đến sự khác biệt về mức độ giống nhau giữa các âm vị. Ví dụ, hãy xem xét âm vị /TH/, trong IAE và đặc biệt là trong các giọng Dublin, thường trải qua quá trình cứng cổ và được coi là điểm dừng của phẩy nang chứ không phải là âm sát răng (Hickey, 2004). Điều này có thể khiến trẻ áp dụng cách đánh vần "sound out" để mã hóa âm vị /T/. Thật vậy, trong tập dữ liệu IAE (xem Phần 3.1), chúng ta thấy các ví dụ có thể có về hiệu ứng này như trong *someting (cái gì đó) và *tink (nghĩ). Do đó, việc tinh chỉnh các giá trị trong ma trận khoảng cách bằng cách sử dụng lỗi chính tả do người nói IAE tạo ra sẽ nắm bắt được các hiệu ứng cụ thể của biến thể này và mang lại hiệu suất tốt hơn cho công cụ sửa lỗi chính tả từ những người nói như vậy.

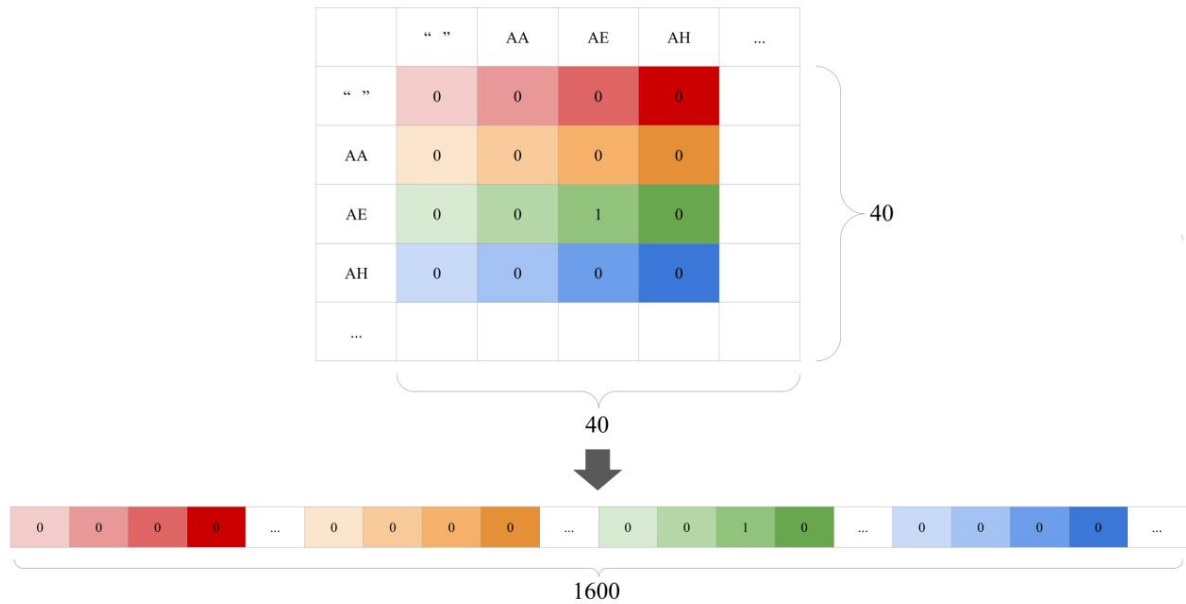
3.3 Tinh chỉnh tiếng Anh có giọng Ailen

Do kích thước hạn chế của tập dữ liệu, phương pháp xác thực chéo k-fold đã được sử dụng để điều chỉnh và đánh giá hiệu suất của công cụ sửa lỗi chính tả được điều chỉnh theo IAE. IAE-corpus được tách ngẫu nhiên thành 10 lần. Mỗi nếp gấp được tổ chức như một bộ thử nghiệm trong khi 9 nếp gấp còn lại được sử dụng để tinh chỉnh ma trận khoảng cách của công cụ S-capade.

Trong trường hợp hiệu suất tốt nhất, khoảng cách chỉnh sửa có trọng số giữa lỗi chính tả và mục tiêu từ thực tương ứng của nó sẽ thấp hơn so với các chỉnh sửa tiềm năng khác. Do đó, trọng số liên quan đến việc thay thế, xóa và chèn được quan sát thấy trong các cặp sai chính tả: đích này sẽ giảm trong khi trọng số được quan sát giữa các cặp sai chính tả: không phải đích sẽ tăng lên.

Để làm điều này, chúng tôi triển khai một mạng lưới thần kinh nông thực hiện hồi quy logistic.

Đối với mỗi lỗi chính tả trong tập huấn luyện, mô hình cơ sở S-capade được sử dụng để tạo ra 10 lỗi sửa có khả năng xảy ra nhất dựa trên khoảng cách chỉnh sửa có trọng số. Trình tự âm vị dự đoán của lỗi chính tả được thu được bằng cách sử dụng biểu đồ



Hình 2: Trực quan hóa việc tạo các vectơ đầu vào thưa thớt từ số lần thay thế.

đến công cụ âm vị (CMUSphinx, 2016) được đào tạo về

Từ điển Phát âm CMU (Weide, 1998).

Các trình tự này sau đó được sắp xếp phù hợp với trình tự của từng ứng cử viên như có thể thấy trong Bảng 2. Trong trường hợp này

ví dụ chúng tôi quan sát sự thay thế sau đây:

- AE : AE
- K : chuỗi rỗng
- SH : CH
- AH : UW
- L : L
- IY : IY

Lưu ý rằng việc không có thao tác chỉnh sửa là được coi là sự thay thế của một âm vị bằng chính nó và việc chèn và xóa được coi là các thay thế liên quan đến chuỗi trống. Mỗi cặp viết sai chính tả: ứng cử viên sau đó được biểu diễn dưới dạng một vectơ thưa có độ dài 1600 với mỗi giá trị biểu thị số lần xuất hiện của tất cả các vectơ có thể sự thay thế âm vị (xem Hình 2 để hình dung quá trình này).

chính tả	Chuỗi âm vị
thực ra	AE K SH AH L IY
đầu đôn	AE CH UW L IY

Bảng 2: So sánh trình tự âm vị giữa 'thực sự' và 'achuly'.

Mỗi tập huấn luyện có từ 208-209 lỗi chính tả và mỗi lỗi chính tả có 10 ứng cử viên

đỉnh chính. Nếu chưa có trong danh sách ứng viên, mục tiêu từ thực sẽ được thêm vào. Một số danh sách ứng cử viên chứa nhiều phiên bản của mục tiêu như là kết quả của cách phát âm khác nhau trong từ điển. Như vậy, đối với mỗi tập huấn luyện, chúng ta có khoảng 250-266 trường hợp viết sai chính tả:target cặp và giữa 1856-1883 trường hợp tiêu cực

lỗi chính tả: các cặp không phải mục tiêu.

Trong một quy trình tương tự như trong Hình 2, ma trận khoảng cách âm vị gốc được sử dụng bởi S-capade được chuyển thành vectơ trọng số có độ dài 1600. Mạng nơ-ron một lớp được sử dụng để thực hiện

hồi quy logistic trong nỗ lực phân loại các trường hợp huấn luyện thành lỗi chính tả thực sự: cặp mục tiêu (nhân 0) hoặc lỗi chính tả phủ định: các cặp không phải mục tiêu (1 nhân). Điều này cho phép chúng ta điều chỉnh trọng số bằng cách sử dụng lan truyền ngược để giảm thiểu tổn thất. Bằng cách nhân mỗi vectơ mẫu huấn luyện theo vectơ trọng số chúng tôi thu được điểm khoảng cách chỉnh sửa có trọng số được sử dụng bởi S-capade để xếp hạng ứng viên. Một thành kiến được thêm vào và một phép biến đổi sigmoid sau đó được áp dụng để xác định xác suất một trường hợp đào tạo là âm lỗi chính tả: cặp không phải mục tiêu (xem Công thức 1 trong đó W là vectơ trọng số, Xi là vectơ huấn luyện i, và b là độ lệch).

$$\hat{y}^i = \sigma(W \cdot X_i + b) \quad (1)$$

	Thu hồi MRR@1	Thu hồi@3	Thu hồi@5	Thu hồi@10	
S-capade Mẫu cơ sở Mẫu	0,623	0,543	0,694	0,746	0,789
tiếng Anh có giọng Ailen 0.693 Holbrook	0,629	0,746	0,746	0,781	0,828
Mẫu tiếng Anh Anh 0.673	0,599	0,728	0,728	0,784	0,823

Bảng 3: Xếp hạng đối ứng trung bình và Recall@K trung bình cho ba mô hình.

Sau đó, vectơ trọng số được điều chỉnh bằng cách sử dụng lan truyền ngược và mất entropy chéo (xem Công thức 2 trong đó N là số trường hợp huấn luyện, y là nhãn đúng và \hat{y} là xác suất của kết quả âm sai chính tả: cặp không phải mục tiêu).

$$L = \frac{1}{N} \sum_{n=1}^N [-y \log \hat{y} - (1-y) \log (1-\hat{y})] \quad (2)$$

Sau đó, vectơ trọng số điều chỉnh được định hình lại thành ma trận khoảng cách 40×40 để sử dụng với công cụ sửa lỗi chính tả S-capade.

3.4 Tinh chỉnh không IAE

Để xác định liệu hiệu suất được cải thiện của mô hình tinh chỉnh IAE có phải là kết quả hay không thích ứng với các loại lỗi được tạo ra dưới ảnh hưởng của một biến thể khu vực chữ không chỉ sản phẩm của việc tăng cường đào tạo, chúng tôi đã tìm cách kiểm tra một mô hình được điều chỉnh bằng cách sử dụng dữ liệu sai chính tả trẻ em với một biến thể tiếng Anh khác. Vì điều này nhằm mục đích chúng tôi chọn tập dữ liệu Holbrook (Mitton, 1985) - một tập hợp các lỗi chính tả được trích từ các bài viết của học sinh người Anh (Holbrook, 1964). Để so sánh với kho dữ liệu IAE-, 232 lỗi chính tả được lấy mẫu ngẫu nhiên từ kho văn bản Holbrook nhằm mục đích tinh chỉnh. Phương pháp được thảo luận trước đây trong Mục 3.3 được áp dụng tương tự ngoại trừ việc xác thực chéo k-Fold. Thay vào đó tất cả 232

lỗi chính tả được cung cấp dưới dạng một tập huấn luyện duy nhất. Kết quả điều chỉnh trọng số hồi quy logistic sau đó được chuyển tới bộ kiểm tra chính tả S-capade để được sử dụng làm chỉ phí vận hành chỉnh sửa và Holbrook mô hình tinh chỉnh đã được thử nghiệm trên IAE-corpus của lỗi chính tả. Chúng tôi kỳ vọng mô hình này sẽ cải thiện đáng kể hoạt động tốt hơn mô hình cơ sở vì nó được tinh chỉnh trên dữ liệu lỗi chính tả của trẻ nhưng không thực hiện được như cũng như mô hình do IAE điều chỉnh vì nó không được điều chỉnh đến các tính năng cụ thể của IAE (mặc dù người Anh Tiếng Anh và IAE sẽ chia sẻ một số tính năng trong chung).

4 kết quả

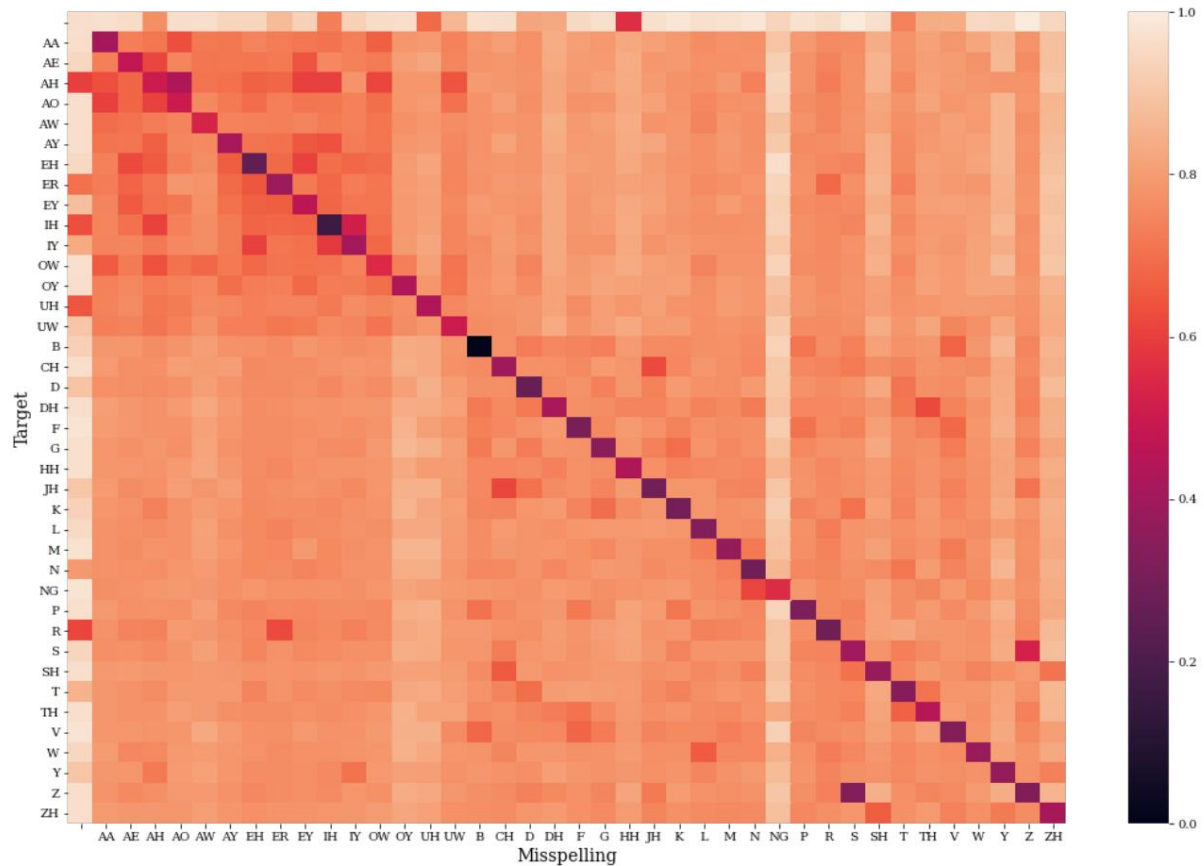
Hiệu quả của công cụ sửa lỗi chính tả nằm ở khả năng đề xuất ứng viên phù hợp sửa chữa và xếp hạng mục tiêu thực sự cao trong số họ. Do đó, chúng tôi xác định tính hiệu quả của các mô hình dựa trên việc liệu chính tả mục tiêu xảy ra ở K ứng cử viên hàng đầu (Recall@K) và ở Xếp hạng đối ứng trung bình (MRR). Vì chỉ có một mục liên quan trong danh sách của các ứng cử viên, cụ thể là từ mục tiêu dự định, Recall@K là 1 nếu mục tiêu nằm trong K ứng cử viên hàng đầu và 0 nếu không. MRR (xem Công thức 3) là thước đo mức độ thường được xếp hạng của mục tiêu trong danh sách các ứng cử viên có giá trị gần bằng 1 cho biết thứ hạng mục tiêu trung bình cao hơn.

$$MRR = \frac{1}{|Q|} \sum_{t=1}^{|Q|} \frac{1}{\text{thứ hạng}} \quad (3)$$

Mỗi ma trận khoảng cách được điều chỉnh bởi IAE được cung cấp cho công cụ sửa lỗi chính tả S-capade (thay thế ma trận mô hình cơ sở ban đầu) sau đó được chạy trên những lỗi chính tả ở ô kiểm tra tương ứng. ĐẾN đánh giá hiệu suất của trình kiểm tra chính tả được tinh chỉnh bằng cách sử dụng dữ liệu IAE, chúng tôi báo cáo giá trị trung bình trên tất cả các lần thử nghiệm cho từng biện pháp đánh giá của chúng tôi và so sánh với kết quả thu được bằng cách sử dụng đường cơ sở Mô hình S-capade và trên Holbrook được tinh chỉnh người mẫu. Kết quả được thể hiện ở Bảng 3 với hiệu suất tốt nhất trên ba mô hình được chỉ định in đậm.

5 Thảo luận

Từ kết quả trong Bảng 3, chúng ta có thể thấy rằng mô hình điều chỉnh IAE vượt trội hơn cả mô hình cơ sở và mô hình điều chỉnh Holbrook trong hầu hết các số liệu đánh giá (với kết quả có thể so sánh được cho Nhớ lại@5). Độ chính xác, (Recall@1), tăng thêm 8,6% tuyệt đối so với mô hình cơ sở và Recall@10 tăng 3,9% tuyệt đối. Những sự gia tăng này sẽ có nghĩa là hệ thống có nhiều khả năng đề xuất từ mục tiêu dự định là gợi ý hàng đầu và, trong một môi trường kiểm tra chính tả tương tác, còn hơn thế nữa có khả năng cung cấp cho người dùng một danh sách ứng cử viên



Hình 3: Ma trận khoảng cách âm vị kết quả từ quá trình tinh chỉnh IAE.

các chỉnh sửa có chứa mục tiêu. Do đó, trải nghiệm của trẻ với biến thể IAE sẽ được cải thiện thông qua việc điều chỉnh mô hình. Hiệu suất lợi ích của mô hình được điều chỉnh bởi IAE là bằng chứng sơ bộ cho thấy tồn tại một số điểm chung hoặc điểm chung. Đặc điểm của các loại lỗi chính tả do trẻ em nói tiếng IAE sản xuất. Những tính năng này không phải tất cả đều được chia sẻ bởi lỗi chính tả của người Anh trong Holbrook corpus không đạt được mức tăng hiệu suất cao, đặc biệt là về MRR và độ chính xác.

Về lý thuyết, ma trận khoảng cách tổng hợp sau Tinh chỉnh IAE sẽ nắm bắt được điểm chung này và các giá trị phải liên quan đến các khía cạnh của IAE hệ thống âm vị học và sự hiện thực hóa âm vị học. Hình 3 cho thấy ma trận khoảng cách được điều chỉnh này, tính trung bình trên mười nếp gấp, được hiển thị dưới dạng bản đồ nhiệt trong đó màu tối hơn biểu thị khoảng cách thấp hơn và do đó mức độ tương đồng cao hơn giữa các âm vị. Tính đến kích thước giới hạn của kho dữ liệu IAE và thực tế là không phải tất cả các tính năng của IAE sẽ biểu hiện trong lỗi chính tả, ma trận khoảng cách sẽ xuất hiện để thể hiện ảnh hưởng của âm vị học IAE. Nhiều các giá trị quan sát được có thể được coi là những đặc điểm chung của tiếng Anh. Ví dụ: âm vị /NG/ trong

một mục tiêu có khả năng được thay thế cho /N/ (như trong viết sai chính tả *bein (being)) hoặc mức độ cao của sự biến đổi giữa các nguyên âm. Tuy nhiên, một số giá trị đặc trưng cho IAE và tiếng Anh Dublin nói riêng (Hickey, 2004). Ví dụ, rất có thể âm vị được thay thế cho /TH/ trong mục tiêu là /T/. Điều này có thể là kết quả của sự cứng cố của ma sát răng đến điểm dừng phẩy nang trong nhắc nhở IAE cách viết như *someting (cái gì đó) và *tink (nghĩ). /AY/ việc bắt đầu tăng cũng phổ biến và có thể giải thích lỗi chính tả *niss (đẹp) và âm vị mục tiêu /AY/ có nhiều khả năng thay thế nhất càng có nhiều nguyên âm trung tâm /AH/, /EY/ và /IH/.

Hiệu suất so sánh của IAE được điều chỉnh công cụ sửa lỗi chính tả và phân tích định tính của ma trận khoảng cách thu được sẽ gợi ý rằng sự khác biệt theo vùng trong cách phát âm có ảnh hưởng nào đó đến việc tạo ra lỗi chính tả trong bài viết của trẻ em. Cách tiếp cận được thực hiện để tinh chỉnh mô hình hiện tại, mặc dù đơn giản, đã được chứng minh là hiệu quả trong việc điều chỉnh công cụ hiệu chỉnh cho phù hợp với IAE. Cái này điểm báo tốt cho nghiên cứu và phát triển trong tương lai trong lĩnh vực này ở quy mô lớn hơn. Sự thích nghi của các công cụ hỗ trợ tiếp thu kiến thức và các công cụ giáo dục để

một biến thể ngôn ngữ cụ thể có thể cung cấp hỗ trợ có mục tiêu hơn cho những trẻ em có thể phải đối mặt với trở ngại trong việc học tập của các em. Hy vọng rằng công việc này sẽ thúc đẩy nghiên cứu sâu rộng hơn về sự tương tác giữa ngôn ngữ nói và khả năng đọc viết và làm thế nào điều này có thể được kết hợp vào sự phát triển các công cụ ngôn ngữ để đảm bảo tất cả trẻ em có thể được hưởng lợi bất kể biến thể ngôn ngữ.

6 hạn chế và công việc trong tương lai

Công việc này nhằm mục đích điều tra sơ bộ đưa ra ý tưởng rằng sự biến đổi cách phát âm theo vùng biểu hiện ở lỗi chính tả của trẻ em và điều đó một mô hình sửa lỗi chính tả phù hợp vừa khả thi vừa mang lại lợi ích cho những người nói các biến thể đó.

Mặc dù các kết quả ban đầu ủng hộ giả thuyết ban đầu nhưng điều quan trọng cần lưu ý là những hạn chế của cách tiếp cận được trình bày ở đây. Tập dữ liệu IAE được sử dụng cho điều chỉnh và thử nghiệm bị giới hạn về quy mô và phạm vi. Nó là một mẫu lỗi chính tả tương đối nhỏ và thiên về những người nói IAE từ Dublin. Các tập hợp một tập dữ liệu lớn hơn mang tính đại diện hơn cho toàn bộ Ireland là một khu vực được quy hoạch công việc trong tương lai đang chờ thiết kế thử nghiệm và phê duyệt về mặt đạo đức. Phương pháp tiếp cận khoảng cách chỉnh sửa có trọng số được S-capade sử dụng chỉ chiếm ảnh xạ âm vị một-một và không phụ thuộc vào ngữ cảnh. Vì vậy, những đặc điểm có khả năng ảnh hưởng của IAE có thể không được nắm bắt đầy đủ. Ví dụ, trong tiếng Anh Dublin, các nguyên âm ngắn thường là kéo dài khi chúng xuất hiện trước /R/ (Hickey, 2004). Tính năng phụ thuộc vào ngữ cảnh này có thể được mã hóa bằng lỗi chính tả nhưng mô hình hiện tại có thể không có cách nào để phân biệt môi trường _/R/ cụ thể. Nếu điều này tỏ ra là một thiếu sót đáng kể thì cần phải áp dụng một phương pháp sửa lỗi chính tả phức tạp hơn. Một biểu đồ để

Công cụ âm vị được đào tạo trên từ điển phát âm CMU được sử dụng để dự đoán các chuỗi âm vị của lỗi chính tả. Điều này không phải lúc nào cũng chính xác đại diện cho chuỗi âm vị dự định của trẻ. Có thể cải thiện dự đoán bằng cách đào tạo công cụ về dữ liệu sai chính tả.

7. Kết luận

Chúng tôi đã trình bày bằng chứng sơ bộ về ảnh hưởng của sự biến đổi cách phát âm theo vùng đối với lỗi chính tả của trẻ em. Bằng cách tinh chỉnh trình kiểm tra chính tả ngữ âm tiếng Anh theo lỗi chính tả do IAE tạo ra loạ chúng tôi chứng minh hiệu suất được cải thiện trên dữ liệu thử nghiệm tương tự. Phân tích định tính của tái

mô hình sulting cho thấy nó có thể học được một số đặc điểm ngữ âm của biến thể ngôn ngữ. Công việc trong tương lai sẽ tìm cách khám phá sâu hơn mối quan hệ này thông qua việc mở rộng tính năng sửa lỗi chính tả của IAE. tập dữ liệu vì kích thước hạn chế và thiên về một quận cụ thể được xác định là một hạn chế.

Cân nhắc về đạo đức

Zeeko được sự đồng ý để thu thập và phân tích các câu trả lời khảo sát cho mục đích nghiên cứu. Các Tập dữ liệu được sử dụng trong bài viết này có nhiều lỗi chính tả từ những phản hồi này được trích xuất với sự cho phép rõ ràng từ Zeeko. Nó được ẩn danh và chứa không có thông tin nhận dạng cá nhân. Chúng tôi xin giới thiệu công việc ban đầu hướng tới một ứng dụng kiểm tra chính tả nhằm mục đích mang lại lợi ích cho trẻ em có cách phát âm khác nhau theo vùng miền, những trẻ trước đây có thể gặp khó khăn khi sử dụng chính tả ngữ âm hiện có các công cụ sửa lỗi. Bài viết này cũng lập luận rằng nghiên cứu sâu hơn khám phá mối quan hệ giữa giọng địa phương và khả năng đọc viết là cần thiết trong một nỗ lực để kỳ thị sự biến đổi ngôn ngữ như vậy và cung cấp hỗ trợ có mục tiêu cho việc học tập của trẻ em.

Sự nhìn nhận

Trung tâm Công nghệ Nội dung Số ADAPT (www.adaptcentre.ie) được tài trợ theo Chương trình Trung tâm Nghiên cứu SFI (Grant 13/RC/2106_P2). Lero - Trung tâm Nghiên cứu Phần mềm của SFI (www.lero.ie) được tài trợ theo Chương trình Trung tâm Nghiên cứu SFI (Grant 13/RC/2094_P2). Chúng tôi xin cảm ơn Zeeko nhân viên điều hành các cuộc hội thảo và khảo sát về phúc lợi trực tuyến - và tất cả các trường học đã đã gắn bó với Zeeko trong nhiều năm.

Người giới thiệu

Jeffrey S Bowers và Peter N Bowers. 2017. Ngoài ra ngữ âm: Trường hợp dạy trẻ logic của hệ thống chính tả tiếng Anh. Nhà tâm lý học giáo dục, 52(2):124-141.

Eric Brill và Robert C Moore. 2000. Một cải tiến mô hình lỗi để sửa lỗi chính tả kênh ồn ào. TRONG ACL, trang 286-293.

Tổ chức từ thiện Anne H, Hollis S Scarborough và Dar-ion M Griffin. 2004. Làm quen với tiếng Anh học đường ở trẻ em người Mỹ gốc Phi và mối liên hệ của nó đến thành tích đọc sớm. Sự phát triển của trẻ nhỏ, 75(5):1340-1356.

- Nhà thờ Kenneth W và William A Gale. 1991. [Chăm điểm xác suất cho việc sửa lỗi chính tả](#). Thống kê và Máy tính, 1(2):93-103.
- CMUSphinx. 2016. [Dựa trên công cụ chuyển đổi biểu đồ thành âm vị về việc học theo trình tự](#).
- Beverley Collins và Inger M Mees. 2013. Thực tế Ngữ âm học và âm vị học: Sách tham khảo dành cho sinh viên. Routledge.
- Carol McDonald Connor và Holly K Craig. 2006. Ngôn ngữ của trẻ mẫu giáo người Mỹ gốc Phi đang nổi lên kỹ năng đọc viết và sử dụng tiếng Anh của người Mỹ gốc Phi: Một mối quan hệ phức tạp.
- Tessa Daffern và Sarah Critten. 2019. Sinh viên và quan điểm của giáo viên về chính tả. Tạp chí Úc về Ngôn ngữ và Đọc viết, 42(1):40-57.
- Fred J. Damerou. 1964. [Một kỹ thuật phát hiện và sửa lỗi chính tả trên máy tính](#). Cộng đồng. ACM, 7(3):171-176.
- Brody Downs, Oghenemaro Anuyah, Aprajita Shukla, Jerry Alan Thất bại, Sole Pera, Katherine Wright, và Casey Kennington. 2020. [Kidspell: Công cụ kiểm tra chính tả ngữ âm, dựa trên quy tắc, hướng đến trẻ em](#). Trong Biên bản của Hội nghị Đánh giá và Tài nguyên Ngôn ngữ lần thứ 12, trang 6937-6946.
- Geert Driessen và Virgie Withagen. 1999. Ngôn ngữ giống và thành tựu giáo dục của người bản địa học sinh tiểu học. Văn hóa Ngôn ngữ và Chương trình giảng dạy, 12(1):1-22.
- Công viên Marina Everri và Kirsty. 2018. Trẻ em trực tuyến hành vi ở các trường tiểu học và trung học ở Ireland. Báo cáo kỹ thuật, Zeeko, Nova UCD.
- William M Fisher. 1999. [Thống kê tin nhắn tới điện thoại hoạt động bằng cách sử dụng ngram và quy tắc](#). Trong ICASSP, tập 2, trang 649-652.
- Marios Fourakis và Càng Robert. 1986. Dừng epenthe-sis trong tiếng Anh. Tạp chí Ngữ âm học, 14(2):197-221.
- Gillian Gallagher và Peter Graff. 2012. Vai trò của sự giống nhau về âm vị học. Lingua, 2(122):107-111.
- Alfred Charles Gimson và Susan Ramsaran. 1970. Một giới thiệu về cách phát âm tiếng Anh, tập 4. Edward Arnold London.
- Raymond Hickey. 2004. Âm vị học của tiếng Anh Ailen. Sổ tay các loại tiếng Anh, 1:68-97.
- Daniel Hládek, Ján Staš và Matúš Pleva. 2020. Khảo sát sửa lỗi chính tả tự động. Thiết bị điện tử, 9(10):1670.
- David Holbrook. 1964. Tiếng Anh cho người bị từ chối: Đào tạo xóa mù chữ ở bậc trung học cơ sở trường học.
- Junko Ito. 1989. Một lý thuyết điệu điệu về sự kết hợp. Ngôn ngữ tự nhiên & Lý thuyết ngôn ngữ học, 7(2):217-259.
- Mark Kane và Julie Carson-Berndsen. 2016. Tăng cường sự nhầm lẫn của điện thoại dựa trên dữ liệu bằng cách sử dụng các thiết bị bị hạn chế sự công nhận. Trong INTERSPEECH, trang 3693-3697.
- Karen Kukich. 1992. [Kỹ thuật tự động hóa sửa chữ trong văn bản](#). Khảo sát máy tính Acm (CSUR), 24(4):377-439.
- Roger Mitton. 1985. Một bộ sưu tập các lỗi chính tả tiếng Anh có thể đọc được trên máy tính. Tâm lý học thần kinh nhận thức, 2(3):275-279.
- Hội đồng quốc gia về chương trình giảng dạy và đánh giá. 2019. Chương trình giảng dạy ngôn ngữ tiểu học. Chính phủ Ireland, Dublin, Ireland.
- Emma O'Neill và Julie Carson-Berndsen. 2019. [Ảnh hưởng của sự phân bố âm vị đến sự tương đồng về nhận thức trong tiếng Anh](#). INTERSPEECH, trang 1941-1945.
- Emma O'Neill, Robert Young, Elsa Thiaville, Muire-ann MacCarthy, Julie Carson-Berndsen và An-thony Ventresque. 2020. [S-capade: Sửa lỗi chính tả nhằm vào các lỗi sai lệch đặc biệt](#). Trong Hội nghị quốc tế về ngôn ngữ thống kê và Xử lý giọng nói, trang 85-96. Mùa xuân.
- Charles Đọc. 2018. Trẻ em đánh vần sáng tạo. Routledge.
- Robert C. Russell và Margaret King Odell. 1918. Soundex. Bảng sáng chế Hoa Kỳ 1.261.167.
- Elinor Saiegh-Haddad. 2003. Khoảng cách ngôn ngữ và việc đọc lần đầu: Trường hợp của diglos-sia trong tiếng Ả Rập. Tâm lý học ứng dụng, 24(3):431.
- Judith I Schwartz. 1982. Sự can thiệp của phương ngữ vào việc đạt được trình độ đọc viết. Tạp chí Đọc, 25(5):440-446.
- Jeff Siegel. 1999. Tiếng Creole và các phương ngữ thiểu số trong giáo dục: Tổng quan. Tạp chí đa ngôn ngữ và Phát triển đa văn hóa, 20(6):508-531.
- Julia Snell và Richard Andrew. 2017. [Ở mức độ nào Phương ngữ và giọng địa phương có ảnh hưởng đến sự phát triển kỹ năng đọc và viết không?](#) Cambridge Tạp chí Giáo dục, 47(3):297-313.
- Nicole Patton Terry. 2006. Quan hệ giữa phương ngữ các kỹ năng biến thể, ngữ pháp và chính tả sớm. Đọc và Viết, 19(9):907-931.
- Nicole Patton Terry. 2012. Xem xét các mối quan hệ giữa sự đa dạng của phương ngữ và kỹ năng đọc viết mới nổi. Rối loạn giao tiếp hàng quý, 33(2):67-77.
- Nicole Patton Terry và Carol Connor. 2010. Châu Phi Tiếng Anh và chính tả Mỹ: Cách thứ hai học sinh đánh vần các đặc điểm nhạy cảm với phương ngữ của từ? Khuyết tật học tập hàng quý, 33(3):199-210.
- Nicole Patton Terry và Hollis S Scarborough. 2011. Giả thuyết âm vị học như một khuôn khổ có giá trị để nghiên cứu mối quan hệ của biến thể phương ngữ với kỹ năng đọc sớm.

Kristina Toutanova và Robert C Moore. 2002. [Mô hình phát âm để cải thiện việc sửa lỗi chính tả](#) . Trong Kỷ yếu của Đại hội thường niên lần thứ 40 trên Hiệp hội Ngôn ngữ học tính toán, trang 144-151.

Jean Veronis. 1988. [Sửa lỗi âm thanh bằng máy tính](#). Máy tính và nhân văn, 22(1):43-56.

Robert A Wagner và Michael J Fischer. 1974. Vấn đề sửa chuỗi thành chuỗi. JACM, 21(1):168-173.

Robert L Weide. 1998. [Từ điển phát âm của CMU-nary](#).

Androula Yiakoumetti. 2007. Lựa chọn ngôn ngữ lớp học trong cộng đồng hai phương ngữ: bao gồm hoặc để loại trừ phương ngữ? Tạp chí giáo dục Cambridge, 37(1):51-66.

Moira Yip. 1987. Nguyên âm tiếng Anh. Tự nhiên Ngôn ngữ & Lý thuyết ngôn ngữ, trang 463-484.