

NGOẠI NGỮ VỚI BẢN NGỮ

**NHỮNG LỖI NGÔN NGỮ CỦA GOOGLE TRANSLATE
KHI DỊCH VĂN BẢN TỪ TIẾNG ANH SANG TIẾNG VIỆT****PHAN THỊ NGỌC LỆ***

TÓM TẮT: Trong những năm gần đây, dịch máy đã trở thành chủ đề xu hướng trong nghiên cứu dịch thuật và giảng dạy tiếng Anh. Trong số các ứng dụng dịch máy, Google Translate với những tính năng thuận tiện và dễ sử dụng cho người dùng đã cho thấy là công cụ dịch máy phổ biến nhất. Tuy nhiên, chất lượng bản dịch của Google Translate vẫn còn gây ra nhiều tranh cãi từ nhiều học giả trên thế giới. Từ thực tiễn đó, nghiên cứu này được thực hiện nhằm tìm ra những lỗi ngôn ngữ cơ bản của Google Translate khi dịch văn bản từ tiếng Anh sang tiếng Việt. Để đạt được mục đích nghiên cứu, tôi sử dụng phương pháp tổng hợp và phân tích lỗi trên các văn bản dịch của Google Translate từ bài tập nhóm của hai lớp biên dịch Anh-Việt cơ bản tôi được phân công trực tiếp giảng dạy tại Đại học Quốc gia Hà Nội. Kết quả cho thấy các nhóm lỗi chính của Google Translate bao gồm lỗi ngữ nghĩa, lỗi từ vựng, và lỗi ngữ pháp. Tôi hi vọng rằng nghiên cứu này sẽ cho giáo viên và người học thấy được những hạn chế của Google Translate, từ đó có những phương pháp giảng dạy và học tập phù hợp trong môn dịch thuật và tiếng Anh.

TỪ KHOA: Dịch máy; Google Translate; lỗi ngôn ngữ; dịch Anh-Việt; sinh viên Việt Nam.

NHẬN BÀI: 01/07/2024.

BIÊN TẬP-CHỈNH SỬA-DUYỆT ĐĂNG: 02/08/2024

1. Dẫn nhập

Trong những năm gần đây, sự tiến bộ nhanh chóng của ngôn ngữ học máy tính đã mang tới cho việc dạy và học ngoại ngữ nhiều kết quả hữu ích. Đặc biệt, việc sử dụng dịch máy trong giao tiếp, viết hoặc dịch thuật tài liệu đã trở nên phổ biến đối với người học và người sử dụng ngôn ngữ. Đã có một số công cụ dịch máy được thiết kế cho đến nay như Microsoft Translator, Amazon Translate, Bing Translator và Yandex Translate. Trong số các công cụ này, Google Translate được coi là công cụ hàng đầu và phổ biến nhất trên toàn thế giới (Indah, 2023), cung cấp cho người dùng hàng tỷ bản dịch mỗi ngày (Shankland, 2013).

Theo Srikandi và Daulay (2022), việc sử dụng Google Translate làm công cụ dịch khá đơn giản vì người dùng chỉ cần gõ văn bản vào trang web, chọn ngôn ngữ đích và nhấn vào nút dịch. Điều này mang lại sự tiện lợi cho người học vì nó cải thiện kỹ năng viết và đóng vai trò như một phương pháp học tập hữu ích. Tuy nhiên, vì công cụ này chỉ dịch theo từng từ nên nó tạo ra những câu không chính xác. Trong nghiên cứu của Srikandi và Daulay (2022), Google Translate thường xuyên sử dụng cấu trúc của Indonesia mà khác với cấu trúc tiếng Anh và điều này có thể dẫn đến một số lỗi về ngữ pháp. Như vậy, sử dụng Google Translate để dịch văn bản lại phát sinh một vấn đề mới bởi vì nó chỉ đơn giản là chuyển ngôn ngữ chứ không biết áp dụng đúng ngữ pháp (Phạm & cộng sự, 2022). Tuy nhiên, theo Tongpoon-Patanasorn (2020), nghiên cứu về dịch máy nói chung và về ứng dụng Google Translate nói riêng nhìn chung vẫn còn ít. Trên thế giới, một số nghiên cứu gần đây đã xem xét chất lượng và các vấn đề của văn bản dịch bằng dịch máy như Borodina và cộng sự (2021), Tsai (2023), Al Farisi và cộng sự (2023). Trong bối cảnh Việt Nam, vẫn còn một khoảng trống nghiên cứu về dịch máy nói chung và về ứng dụng Google Translate nói riêng được sử dụng trong thực tế mặc dù chúng đóng vai trò ngày càng quan trọng trong các lĩnh vực xã hội, kinh doanh và học thuật.

Qua những nghiên cứu trong và ngoài nước về dịch máy và ứng dụng Google Translate, có thể thấy chất lượng của bản dịch Google Translate cần phải được cải thiện đáng kể để các văn bản dịch có chất lượng tốt nhất. Trong quá trình giảng dạy ngoại ngữ, giáo viên cần phải nhận thức được những hạn chế về mặt ngôn ngữ khi dịch của Google Translate để đưa ra hướng dẫn phù hợp cho người học. Do đó, tôi nghĩ rằng một nghiên cứu về những lỗi ngôn ngữ cơ bản của Google Translate khi dịch văn bản từ tiếng Anh sang tiếng Việt là thực sự cần thiết để người học nhận thức rõ các vấn đề ngữ pháp có thể phát sinh khi sử dụng Google Translate. Từ đó, người học có thể tránh được những vấn đề này và cải thiện kiến thức ngoại ngữ của họ.

*TS; Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội; Email: phanthingocle@yahoo.com.vn

2. Cơ sở lý thuyết

2.1. Tổng quan về ứng dụng dịch Google Translate

Google Translate là một công cụ dịch máy tự động miễn phí được Google triển khai vào năm 2006. Giống như các công cụ dịch máy khác, nó phụ thuộc vào công nghệ máy tính và phần mềm để dịch các từ, câu, đoạn văn từ ngôn ngữ nguồn sang ngôn ngữ đích ngay lập tức (Boitet & cộng sự, 2009). Theo quan điểm của Groves và cộng sự (2015), Google Translate gần đây đã trở thành công cụ dịch máy phổ biến nhất vì nó chạy trên cơ sở dữ liệu khổng lồ và cung cấp các văn bản được dịch chính xác hơn so với các ứng dụng dịch máy khác. Người dùng có thể truy cập miễn phí và dễ dàng vì nó cung cấp cả giao diện trang web và ứng dụng di động Android và iOS để dịch văn bản sang ngôn ngữ đã chọn. Kể từ khi ra đời, dịch vụ này hỗ trợ 103 ngôn ngữ ở nhiều cấp độ khác nhau, trong đó có tiếng Việt và dịch hơn 100 tỷ từ mỗi ngày (McGuire, 2018).

Hiện tại, Google Translate sử dụng hệ thống dịch máy nơ-ron (Neural Machine Translation) cho hầu hết các cặp ngôn ngữ và hệ thống này cung cấp kết quả chính xác hơn các kết quả khác. Tuy nhiên, chất lượng bản dịch của Google Translate còn phụ thuộc vào nguồn tài liệu được nhập vào hệ thống. Theo đó, Google Translate chỉ có bản dịch chuẩn khi chứa các dữ liệu liên quan đến nội dung dịch được yêu cầu. Theo quan điểm của Kurniasih (2017), sử dụng dịch máy nói chung và Google Translate nói riêng có cả lợi ích và nhược điểm. Tác giả liệt kê một số ưu điểm của Google Translate như: dịch ngôn ngữ nhanh, miễn phí cho tất cả người dùng, có thể sử dụng trên tất cả các trình duyệt Internet, truyền thông toàn cầu. Bên cạnh đó, Google Translate cũng có một số nhược điểm như: không chính xác 100%, dễ dàng sử dụng để gian lận trong giờ học ngoại ngữ (Kurniasih, 2017). Vì vậy, để đánh giá chính xác chất lượng dịch thuật của Google Translate, các học giả cần nghiên cứu chất lượng dịch thuật của Google Translate ở nhiều lĩnh vực khác nhau. Các quan điểm về những lỗi ngôn ngữ của Google Translate trong và ngoài nước được tổng hợp trong phần dưới đây.

2.2. Tổng quan tình hình nghiên cứu về những lỗi ngôn ngữ của Google Translate trong và ngoài nước

Tại Việt Nam, một số tác giả đã đề cập tới các vấn đề khi sử dụng Google Translate. Trong luận án nghiên cứu về những lỗi của Google Translate khi dịch tự động Anh-Việt các hợp đồng kinh tế, Trần Lê Tâm Linh (2017) đứng trên quan điểm ngôn ngữ học để khảo sát các loại lỗi ngôn ngữ, chủ yếu là lỗi chính tả, lỗi từ vựng và lỗi ngữ pháp khi dịch tự động hợp đồng kinh tế Anh-Việt để tìm ra quy luật phát sinh lỗi. Các kết quả của nghiên cứu này làm cơ sở cho ngành tin học cài đặt các thuật toán khắc phục lỗi. Một nghiên cứu khác của Trần Trương Huỳnh Lê và Lê Minh Thanh (2020) cũng tập trung vào các lỗi ngữ pháp của Google Translate, nhưng trên tư liệu tiếng Hán. Trên tổng số 175 câu khảo sát, các tác giả phân tích lỗi ở mức độ cơ bản nhất và đã phát hiện khoảng 100 lỗi. Con số này thể hiện số câu dịch sai chiếm hơn nửa số câu khảo sát. Từ đó cho thấy, công cụ này chưa thể đáp ứng được nhu cầu của người học. Tương tự như vậy, trong nghiên cứu về dịch tự động và kỹ năng viết của sinh viên, tác giả Công (2016) đã phân chia một số nhóm lỗi cơ bản khi sử dụng Google Translate như: lỗi dịch về từ vựng, lỗi dịch về ngữ pháp và cú pháp, lỗi dịch về tham chiếu. Từ đó, tác giả đưa ra một số đề xuất nhằm hạn chế những ảnh hưởng tiêu cực của dịch tự động đối với kỹ năng viết của sinh viên. Có thể thấy chưa nhiều nghiên cứu tại Việt Nam đề cập tới vấn đề chất lượng dịch của các công cụ dịch máy nói chung, và Google Translate nói riêng.

Trên thế giới, nhiều nghiên cứu đã được tiến hành về các lỗi ngôn ngữ khi dịch bằng Google Translate. Một trong những nghiên cứu điển hình về phân loại lỗi của dịch máy không thể không nhắc tới tiêu chí đánh giá của Farrús và cộng sự (2010). Tác giả phân ra 5 loại lỗi cơ bản của dịch máy bao gồm: lỗi chính tả (dấu câu, viết hoa, chính tả), lỗi hình thái (liên quan đến các hình thức (tức là biến tố, thường là hậu tố) của động từ, danh từ, và những từ khác (tính từ và trạng từ)), lỗi từ vựng (thừa từ hoặc thiếu từ), lỗi ngữ pháp (liên từ, giới từ, mạo từ, thành phần cú pháp), và lỗi ngữ nghĩa (ý nghĩa của từ). Một số nhà nghiên cứu khác cũng đồng ý rằng bản dịch do Google Translate cung cấp đòi hỏi phải có nhiều chỉnh sửa do có vấn đề trong việc xác định “cấu trúc câu và phong cách viết” (Kirchoff & cộng sự, 2011), và “sử dụng sai từ và bỏ sót từ” (Sheppard, 2011). Groves và Mundt (2015) giải thích rằng, Google Translate thiếu khả năng “thỏa mãn các tiêu chuẩn của một diễn ngôn” khi xử lý bản dịch vượt quá cấp độ câu. Điều này có thể do sự khác biệt giữa

ngôn ngữ nguồn và ngôn ngữ đích (Costa-jussá & cộng sự, 2012). Trong những năm gần đây, theo nghiên cứu của Egamberdievna và Daminovna (2020), sinh viên kỹ thuật dầu khí đã nhận thấy bản dịch bằng Google Translate khó hiểu, đôi khi gây nhầm lẫn cho sinh viên do các lỗi về ngữ pháp, từ vựng, và hình thái cú pháp. Borodina và cộng sự (2021) cũng cho rằng, dịch máy là một quá trình dịch văn bản hoàn toàn tự động. Nếu không có sự hỗ trợ của người dịch, dịch máy hiện tại chỉ có thể đạt được độ chính xác 90%. Trong một cuộc khảo sát thông qua bảng câu hỏi do Tsai (2023) thực hiện, tác giả cho rằng Google Translate cho kết quả chưa hẳn chính xác, nhưng vẫn khá hữu ích cho sinh viên trong việc tìm từ vựng để lên ý tưởng bài viết. Al Farisi và cộng sự (2023) cũng nhấn mạnh rằng các bản dịch của Google Translate cần phải được chỉnh sửa kỹ lưỡng về mặt ngữ pháp để hạn chế số lỗi.

Qua các nghiên cứu trong và ngoài nước, có thể thấy rằng cách phân loại lỗi của dịch máy khá đa dạng tùy thuộc vào từng ngữ cảnh nghiên cứu. Tôi cho rằng với mục đích của nghiên cứu này, cách phân loại lỗi dịch máy của Farrús và cộng sự (2010) là phù hợp nhất vì nhóm tác giả đã giải thích chi tiết các loại lỗi dịch thuật và cách phân loại này cũng dễ dàng áp dụng ở cấp độ từ, cụm từ hoặc câu đơn giản. Do đó, cách phân loại của Farrús và cộng sự (2010) (gồm 5 tiêu chí: lỗi chính tả, lỗi hình thái, lỗi ngữ pháp, lỗi từ vựng, lỗi ngữ nghĩa) được tôi sử dụng như khung lý thuyết để thiết kế các công cụ nghiên cứu được trình bày cụ thể trong phần dưới đây.

3. Cách thức nghiên cứu

Tham gia nghiên cứu này bao gồm 90 sinh viên đến từ hai lớp dịch thuật Anh-Việt cơ bản tại Đại học Quốc gia Hà Nội. Điều kiện tiên quyết để tham gia khóa học là sinh viên cần có trình độ tiếng Anh B2 theo Khung tham chiếu chung châu Âu (tương đương cấp độ 4 theo Khung năng lực ngoại ngữ 6 bậc). Đây là hai lớp tôi được phân công trực tiếp giảng dạy trong học kỳ 2 năm học 2023-2024.

Trong khóa học, sinh viên được yêu cầu thực hiện bài tập biên dịch theo nhóm. Tôi lựa chọn 01 bài báo tiếng Anh (độ dài khoảng 350-400 từ về chủ đề Kinh tế-xã hội) và gửi cho các nhóm vào tuần 5. Sinh viên các nhóm được yêu cầu sử dụng Google Translate để dịch toàn bộ bài báo sang tiếng Việt. Sau đó, các nhóm sẽ xác định và phân tích các vấn đề trong bản dịch của Google Translate và đề xuất bản dịch của nhóm. Vào tuần 7, tôi thu các sản phẩm nhóm, và đánh giá các bản dịch theo tiêu chí cụ thể để tổng kết xem bản dịch của Google Translate có những lỗi gì, nhóm nào có phương án đề xuất tốt hơn, và bình chọn cho sản phẩm dịch có chất lượng tốt nhất sau khi hiệu đính bản dịch của Google Translate.

Trong nghiên cứu này để đánh giá chất lượng bản dịch của mỗi nhóm và tổng kết được các lỗi ngôn ngữ của Google Translate, tôi sử dụng phương pháp phân tích lỗi để xác định các loại lỗi chính trong bản dịch của Google Translate. Sau đó dựa vào khung phân loại lỗi dịch máy của Farrús và cộng sự (2010), tôi phân tích và xếp các lỗi vào 5 nhóm chính gồm: lỗi chính tả, lỗi hình thái, lỗi ngữ pháp, lỗi từ vựng, lỗi ngữ nghĩa với những ví dụ minh họa cụ thể. Kết quả thu được về các nhóm lỗi ngôn ngữ của Google Translate khi dịch văn bản từ tiếng Anh sang tiếng Việt được trình bày cụ thể trong phần dưới đây.

4. Kết quả nghiên cứu và thảo luận

Để tìm hiểu về chất lượng dịch của Google Translate và phân loại các lỗi ngôn ngữ chính, tôi đã tổng hợp kết quả phân tích dịch Google Translate của các nhóm phân công trong lớp, sử dụng phương pháp phân tích lỗi để kiểm tra lại các bản dịch của từng nhóm. Bộ dữ liệu gồm 230 câu văn bản tiếng Việt do Google Translate dịch và 380 lỗi đã được phát hiện nhờ những dữ liệu này. Dựa theo khung phân loại lỗi ngôn ngữ của Farrús và cộng sự (2010), các loại lỗi chính được trình bày trong Bảng 1 dưới đây.

Bảng 1. Các loại lỗi ngôn ngữ của Google Translate

TT	Các loại lỗi		Tần suất	Tỉ lệ %
1	Lỗi ngữ nghĩa	Thiếu ngữ cảnh	53	13,9
		Câu không rõ nghĩa	21	5,5
		Hiểu sai về văn bản nguồn	25	6,6
		Dịch nguyên văn	19	5
		Dịch thiếu tự nhiên	34	8,9

2	Lỗi từ vựng	Lựa chọn từ không phù hợp	67	17,6
		Thiếu từ	64	16,8
		Thừa từ	15	3,9
3	Lỗi ngữ pháp	Sai liên từ	18	4,7
		Sai trật tự từ	58	15,3
4	Lỗi chính tả	Lỗi viết hoa	6	1,6
5	Lỗi hình thái	Hậu tố của từ	0	0%

4.1. Lỗi ngữ nghĩa

Từ kết quả phân tích và tổng hợp các bản dịch của Google Translate, lỗi ngữ nghĩa là nhóm lỗi có tần suất mắc phải cao nhất (chiếm 39,9%), trong đó những câu bị thiếu ngữ cảnh hoặc dịch thiếu tự nhiên chiếm số phần trăm cao nhất trong nhóm. Những ví dụ cụ thể của nhóm lỗi này được minh họa trong bảng 2 dưới đây.

Bảng 2. Ví dụ về lỗi ngữ nghĩa

TT	Loại lỗi	Văn bản nguồn	Văn bản đích (Google Translate dịch)	Đề xuất dịch
1	Thiếu ngữ cảnh	<i>Find it on every happening Hanoi corner, sometimes paired with food,...</i>	<i>Tìm thấy nó trên mọi ngõ ngách Hà Nội đang diễn ra, đôi khi được kết hợp với đồ ăn...</i>	<i>Các quán bia có mặt trên mọi góc ngách của Hà Nội, đôi khi người ta uống bia kèm chút đồ ăn/ đồ nhắm/ nhậu...</i>
2	Câu không rõ nghĩa	<i>The flavor depends on the rate at which the beer is being sold and how much the seller has decided to water it down that day.</i>	<i>Hương vị phụ thuộc vào tốc độ bán bia và lượng nước của người bán đã quyết định giảm nước vào ngày hôm đó.</i>	<i>Hương vị của bia phụ thuộc vào tốc độ tiêu thụ và công thức pha chế.</i>
3	Hiểu sai về văn bản nguồn	<i>Leap-of-faith traffic</i>	<i>Lưu lượng truy cập nhảy vọt</i>	<i>Giao thông theo kiểu tin nhau là chính/ Qua đường bằng “niềm tin”</i>
4	Dịch nguyên văn	<i>In the next five years, there should be no change in curricula and textbooks.</i>	<i>Trong 5 năm tới không thay đổi chương trình và sách giáo khoa.</i>	<i>Trong 5 năm tới, tạm giữ ổn định chương trình và sách giáo khoa.</i>
5	Dịch thiếu tự nhiên	<i>Vietnamese education: urgent solutions are in bad need</i>	<i>Nền giáo dục Việt Nam: đang rất cần những giải pháp cấp bách</i>	<i>Mấy giải pháp cấp bách về giáo dục ở Việt Nam</i>

Qua các ví dụ trên, có thể thấy rằng do Google Translate không hiểu ngữ cảnh của văn bản, nên dịch từ không chính xác, khiến cho câu văn không rõ nghĩa và thiếu tự nhiên. Theo như Zhou và cộng sự (2019), thông tin sai lệch là hành động thay thế một từ hoặc cụm từ nhưng theo cách sai dẫn đến hiểu lầm. Việc sử dụng ngôn ngữ phù hợp với tình huống có ảnh hưởng trực tiếp đến mức độ tự nhiên của ngôn ngữ. Các ví dụ cho thấy Google Translate không "nhạy cảm" với ngữ cảnh, mặc dù có ngân hàng bộ nhớ khổng lồ về từ vựng, ngữ nghĩa và cú pháp. Cùng một cụm từ hoặc thuật ngữ sẽ có ý nghĩa riêng biệt tùy thuộc vào ngữ cảnh sử dụng nó. Vấn đề này càng trở nên thách thức hơn trong tiếng Việt, nơi mà nhiều thuật ngữ khác nhau được sử dụng để giao tiếp trong các ngữ cảnh khác nhau. Google Translate thường dịch thì của câu không chính xác hoặc thiếu chủ ngữ trong câu khiến cho nghĩa của câu không rõ ràng.

Trên thế giới, các nhà nghiên cứu cũng đã có những thử nghiệm tương tự nhưng trên những cặp ngôn ngữ khác. Ví dụ, Handschuh (2013) đã tiến hành một nghiên cứu về bản dịch tiếng Đức-Anh bằng cách sử dụng bốn bản dịch máy trực tuyến khác nhau (ví dụ: Google Translate, SYSTRAN, Bing và Babylon) và phát hiện ra rằng vẫn còn những lỗi được tìm thấy trong bản dịch văn bản đích do dịch máy tạo ra. Cụ thể là ý nghĩa ban đầu của câu thường bị thay đổi khi dịch các văn bản dài hơn. Những kết quả nghiên cứu này cho thấy lỗi về ngữ nghĩa là một vấn đề khá lớn khi sử dụng Google Translate để dịch văn bản. Tuy nhiên, theo Vilar và cộng sự (2006), loại lỗi phổ biến nhất của Google Translate là nhóm lỗi về từ vựng gồm có lỗi “chọn từ không phù hợp”, tiếp đến là lỗi “thiếu từ”, lỗi “trật tự từ” và sau đó mới là lỗi “thiếu ngữ cảnh”. Theo ý kiến của tôi, lỗi về từ vựng cũng là một nhóm lỗi chiếm tần suất cao (xem Bảng 1). Nhóm lỗi này được trình bày trong phần tiếp theo.

4.2. Lỗi từ vựng

Nhóm lỗi về từ vựng đứng thứ hai với tần suất mắc lỗi chiếm tổng 146 (38,3%) mặc dù chỉ có 3 loại lỗi xuất hiện trong các văn bản dịch. Đáng chú ý, loại lỗi lựa chọn từ không phù hợp là loại lỗi có tần suất mắc phải phổ biến nhất trong tất cả các loại lỗi (17,6%). Một số ví dụ cụ thể của nhóm lỗi này được minh họa trong Bảng 3 dưới đây.

Bảng 3. Ví dụ về lỗi từ vựng

TT	Loại lỗi	Văn bản nguồn	Văn bản đích (Google Translate dịch)	Đề xuất dịch
1	Lựa chọn từ không phù hợp	They may wander at school all the daytime simply for learning by heart exams' sample answers.	Họ có thể lang thang ở trường cả ngày chỉ để học thuộc lòng các câu trả lời mẫu của các bài kiểm tra.	Họ có khi lên lớp từ sáng sớm đến tối mịt, môn nào cũng học đi học lại các bài mẫu cho kỳ kiểm tra.
2	Thiếu từ	In conclusion, it is desperately needed to notice that.....	Tóm lại, cần thấy rằng.....	Tóm lại, chúng ta cần phải nhận thức rằng.....
3	Thừa từ	Inadequately high financial cost for textbooks comparing to their actual effectiveness.	Chi phí tài chính cho sách giáo khoa cao chưa tương xứng so với hiệu quả thực tế của chúng.	Kinh phí cho sách giáo khoa cao quá mức so với hiệu quả thực tế.

Lỗi từ vựng xảy ra khi Google Translate lựa chọn từ sai hoặc dùng thừa từ hoặc thiếu từ. Theo Putri (2019), lỗi thiếu từ xảy ra khi thiếu một từ đúng ra phải có mặt để đảm bảo câu đúng ngữ pháp. Bên cạnh đó, lỗi thừa từ được định nghĩa là sự tồn tại của một từ không nên có trong câu hoặc cụm từ (Chou, 2014). Qua các ví dụ trong Bảng 3, bản dịch của Google Translate với việc sử dụng từ ngữ sai phần nào đó đã ảnh hưởng đến ý nghĩa của văn bản. Mặt khác, việc dịch thiếu hoặc thừa từ là do Google Translate không tìm được từ tương đương như trong văn bản nguồn. Điều này xảy ra khi Google Translate không thể dịch được từ liên quan đến văn hóa và hệ thống ngôn ngữ của ngôn ngữ đích, từ đó ảnh hưởng đến chất lượng của bản dịch. Trong nghiên cứu của Putri và Ardi (2015), các tác giả cũng cho rằng lỗi từ vựng bao gồm dùng từ không chính xác và dịch thiếu từ là những lỗi phổ biến nhất của Google Translate. Khi phân tích các văn bản dịch từ Google Translate, tôi thấy rằng hầu hết các câu dịch đều bị thiếu chủ ngữ khi câu tiếng Anh chứa những chủ ngữ giả như “it” hoặc “there” (xem Bảng 3). Khác với kết quả này, Putri và Ardi (2015) lại cho rằng những từ Google Translate hay dịch thiếu nhất là những từ có chức năng lấp đầy (filler words) như giới từ, mạo từ hoặc động từ liên kết. Tôi cho rằng mục đích của những từ lấp đầy thường thể hiện mối quan hệ ngữ pháp hơn là ý nghĩa từ vựng của câu. Vì thế, một số lỗi về ngữ pháp của Google Translate được trình bày trong phần tiếp theo.

4.3. Lỗi ngữ pháp

Theo kết quả phân tích lỗi từ các văn bản dịch của Google Translate, có thể thấy rằng lỗi ngữ pháp chiếm không quá lớn trong tổng số các loại lỗi (20%) với hai loại lỗi chính là dịch sai liên từ kết nối và sai trật tự từ. Một số ví dụ cụ thể của nhóm lỗi này được minh họa trong Bảng 4 dưới đây.

Bảng 4. Ví dụ về lỗi ngữ pháp

TT	Loại lỗi	Văn bản nguồn	Văn bản đích (Google Translate dịch)	Đề xuất dịch
1	Sai liên từ	Hopefully, the following proposals would be paid sufficient attention to by functional agencies. Thus, tougher solutions would be made to give the national education a brand new profile.	Mong rằng những đề xuất sau đây sẽ được các cơ quan chức năng quan tâm đúng mức. Do đó, các giải pháp cứng rắn hơn sẽ được thực hiện để mang lại cho nền giáo dục quốc dân một diện mạo hoàn toàn mới.	Mong rằng những đề xuất sau đây sẽ nhận được sự quan tâm đúng đắn từ phía các cơ quan chức năng, từ đó đưa ra những giải pháp hiệu quả hơn để mang lại một diện mạo hoàn toàn mới cho nền giáo dục Việt Nam.
2	Sai trật tự từ	That the above proposals are well taken does not necessarily require increased investment.	Việc các đề xuất trên được thực hiện tốt không nhất thiết phải tăng đầu tư.	Để thực hiện tốt các đề xuất trên, không nhất thiết phải tăng vốn đầu tư.

Các ví dụ trên cho thấy các câu dịch bởi Google Translate đều thiếu tính tự nhiên do bản dịch sử dụng sai các liên từ kết nối các vế câu hoặc sai trật tự từ trong câu khiến câu bị sai ngữ pháp khi dịch sang tiếng Việt. Lỗi sắp xếp sai trật tự từ xảy ra khi một từ hoặc một nhóm từ được đặt không chính xác trong cụm từ hoặc câu (Kembarén & cộng sự, 2023). Tôi cho rằng một trong những lý do chính khiến Google Translate nói riêng và dịch máy nói chung chưa có khả năng đưa ra một bản dịch chính xác nhất là vì dịch máy chỉ có khả năng phiên dịch ngôn ngữ dựa trên các thuật toán và quy trình được lập trình sẵn. Các thuật toán và quy trình này thường được thiết kế xoay quanh các quy tắc giao tiếp chính thức. Điều này khiến cho bản dịch từ dịch máy thường khô khan và kém “tự nhiên”. Trong nghiên cứu về chất lượng bản dịch từ tiếng Anh sang tiếng Indonesia, Putri và Ardi (2015) cho rằng lỗi trật tự từ xảy ra do sự khác biệt lớn về cấu trúc giữa tiếng Anh và tiếng Indonesia. Điều này cũng tương tự như sự khác biệt lớn giữa cấu trúc tiếng Anh và tiếng Việt. Trong nghiên cứu về chất lượng dịch tiêu đề phim tiếng Anh sang tiếng Việt, nhóm tác giả Trần Thị Thanh Huyền và Lê Nhân Thành (2021) cũng đưa ra một ví dụ về lỗi ngữ pháp của Google Translate khi dịch tiêu đề “People on Sunday” thành “Những người vào Chủ nhật”. Bản dịch này theo các tác giả tuy hay nhưng nghe khá lạ tai với người Việt Nam bởi thứ tự cú pháp sai. Do đó, các tác giả đề xuất bản dịch phù hợp là “Chủ nhật của mọi người”.

Trường hợp này là một vấn đề nghiêm trọng mà người học ngôn ngữ sẽ phải đối mặt khi dịch các ngôn ngữ có cấu trúc khác nhau. Vì vậy, điều quan trọng là người học ngôn ngữ phải hiểu được ý chính của ngôn ngữ nguồn, sau đó điều chỉnh cho phù hợp với cấu trúc của ngôn ngữ đích để không làm thay đổi thông điệp truyền tải trong bản dịch.

4.4. Lỗi chính tả và lỗi hình thái

Đây là hai nhóm lỗi chiếm vị trí thấp nhất trong Bảng 1. Trong các bản dịch của Google Translate, lỗi viết hoa chỉ chiếm 1,6% và không có câu nào mắc lỗi về hình thái. Ví dụ về lỗi viết hoa được minh họa trong Bảng 5 dưới đây.

Bảng 5. Ví dụ về lỗi chính tả

TT	Loại lỗi	Văn bản nguồn	Văn bản đích (Google Translate dịch)	Đề xuất dịch
1	Lỗi viết hoa	authorize the local	ủy quyền cho sở giáo	ủy quyền cho Sở Giáo

		<i>Department of Education and Training to organize senior secondary school leaving examinations.....</i>	<i>dục và đào tạo địa phương tổ chức thi tốt nghiệp trung học phổ thông</i>	<i>dục và Đào tạo địa phương tổ chức kỳ thi tốt nghiệp trung học phổ thông</i>
--	--	---	---	--

Trong nghiên cứu của Trần Thị Thanh Huyền và Lê Nhân Thành (2021), các tác giả cũng cho rằng lỗi chính tả là một nhóm lỗi chiếm tần suất rất nhỏ khi dịch văn bản bằng Google Translate. Hai tác giả có đưa ra một ví dụ khác của nhóm lỗi này là cụm “Jurassic Park” được Google Translate dịch thành “công viên kỷ Jura”. Trong ví dụ này, chữ c trong “công” và k trong “kỷ” không tuân theo quy tắc viết hoa trong tiếng Việt.

Về nhóm lỗi hình thái, tôi đồng tình với lập luận của Nguyễn Phú Hoàng Như (2011), đó là có sự khác biệt lớn về mặt hình thái giữa tiếng Anh và tiếng Việt. Trong tiếng Anh, hậu tố được sử dụng để thay đổi hình thức của một từ, thường là danh từ, động từ và tính từ. Ngược lại, tiếng Việt không có hậu tố. Nhờ đó, Google Translate và người dịch thường không mắc lỗi hình thái trong bản dịch từ tiếng Anh sang tiếng Việt.

5. Kết luận

Nghiên cứu này tổng hợp và phân tích lỗi ngôn ngữ của các bản dịch từ tiếng Anh sang tiếng Việt của Google Translate được thực hiện bởi các nhóm sinh viên từ hai lớp dịch thuật Anh-Việt cơ bản tôi được phân công trực tiếp giảng dạy trong học kỳ 2 năm học 2023-2024. Dựa theo khung phân loại các nhóm lỗi của Farrús và cộng sự (2010), tôi chia các loại lỗi thu được vào 5 nhóm chính: lỗi ngữ nghĩa, lỗi từ vựng, lỗi ngữ pháp, lỗi chính tả, và lỗi hình thái. Trong các nhóm lỗi trên, nghiên cứu chỉ ra rằng nhóm lỗi về ngữ nghĩa (gồm các lỗi về thiếu ngữ cảnh, câu dịch không rõ nghĩa, hiểu sai về văn bản nguồn, dịch nguyên văn, và dịch thiếu tự nhiên) và nhóm lỗi về từ vựng (lựa chọn từ không phù hợp, dịch thiếu từ, dịch thừa từ) có tần suất mắc lỗi cao nhất từ các bản dịch của Google Translate. Các kết quả nghiên cứu cũng được tôi thảo luận và so sánh với các kết quả nghiên cứu của các học giả trong và ngoài nước để có sự đánh giá từ các góc nhìn và quan điểm khác nhau khi nghiên cứu về chất lượng bản dịch của Google Translate.

Từ kết quả nghiên cứu nêu trên, có thể thấy rằng nhìn chung, dịch máy hiện tại chỉ có thể đạt được độ chính xác 90%. Chất lượng bản dịch của Google Translate cần được cải thiện đáng kể để tránh các lỗi về ngữ nghĩa, từ vựng và ngữ pháp. Vì vậy, giáo viên phải nhận thức được những giới hạn đó và đưa ra những hướng dẫn phù hợp cho người học. Về phía người học, bên cạnh việc mở rộng vốn từ vựng, tôi cũng đề xuất người học nên cải thiện kiến thức về ngữ pháp để khi có lỗi xảy ra với bản dịch máy, họ có thể ngay lập tức nhận ra và có thể sửa nó bằng cấu trúc, từ hoặc cụm từ thích hợp. Qua nghiên cứu này, tôi hi vọng có thể khiến cho người học nhận ra rằng những bản dịch từ Google Translate vẫn cần chỉnh sửa sau đó. Do vậy, người học không nên quá phụ thuộc vào các ứng dụng dịch thuật, khiến việc xử lý ngôn ngữ trở nên kém nhạy bén và thiếu biểu cảm.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Công Huyền Tôn Nữ Ý Nhiệm (2016), Dịch tự động và kỹ năng viết của sinh viên. *Kỷ yếu Hội thảo Quốc gia “Nghiên cứu Liên ngành về Ngôn ngữ và Giảng dạy Ngôn ngữ lần thứ hai”*, Trường Đại học Ngoại ngữ, Đại học Huế, 93-105.
2. Nguyễn Phú Quỳnh Như (2015), *A quick analysis of some typical gaps in phonology and morphology between English and Vietnamese that lead to Vietnamese students' common errors* (Phân tích nhanh một số khác biệt điển hình về âm vị, hình thái giữa tiếng Anh và tiếng Việt dẫn đến những lỗi thường gặp của sinh viên Việt Nam). <http://nnkt.ueh.edu.vn/wp-content/uploads/2015/06/21.pdf>.
3. Trần Lê Tâm Linh (2017), *Những lỗi ngôn ngữ của phần mềm Google Translate khi dịch tự động Anh - Việt các hợp đồng kinh tế*. Luận án Tiến sĩ Ngữ văn. Trường Đại học Khoa học Xã hội và Nhân văn, Hồ Chí Minh.
4. Trần Trương Huỳnh Lê và Lê Minh Thanh (2020), *Phân tích lỗi ngữ pháp của công cụ google dịch khi dịch tự động từ Hán sang Việt qua khảo sát 100 câu tiếng Hán thông dụng*. Hội thảo Đổi mới trong dạy và học ngôn ngữ, 226-236.

5. Trần Thị Thanh Huyền và Lê Nhân Thành (2021), "Chất lượng dịch tiêu đề film tiếng Anh sang tiếng Việt của Google Translate". Tạp chí *Khoa học - Trường Đại học Quy Nhơn*, 15(4), 69-75. <https://doi.org/10.52111/qnjs.2021.15407>.

Tiếng Anh

6. Al Farisi, M. Z., Tantowi, Y. A., & Nurmala, M. (2023), Language, Culture, and Translation. *Prosiding Pertemuan Ilmiah Internasional Bahasa Arab*, 14(1), 1093–1102.
7. Borodina, M., Golubeva, T. I., Korotaeva, I. E., Shumakova, S. Y., Bessonova, T. V., & Zharov, A. N. (2021), Impact of the Google Translate machine translation system on the quality of training student translators. *Technology. Special Issue on Current Trends in Management and Information Technology*, 18, 68-78. <https://doi.org/10.14704/WEB/V18SI05/WEB18214>
8. Farrús, M., Costa-Jussà, M. R., Mariño, J. B., & Fonollosa, J. A. R. (2010), Linguistic-based evaluation criteria to identify statistical machine translation errors, In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*, Saint Raphaël, France. European Association for Machine Translation. <https://aclanthology.org/2010.eamt-1.12>
9. Groves, M., & Mundt, K. (2015), Friend or foe? Google translate in language for academic purposes. *English for Specific Purposes*, 37, 112-121. <https://doi.org/10.1016/j.esp.2014.09.001>
10. Indah, S. (2023). *An analysis of accuracy level of Google Translate in English-Indonesian translation*. Uin Raden Intan Lampung.
11. Kurniasih, I. (2017), *An analysis of grammatical errors of using Google Translate from Indonesia to English in writing undergraduate thesis abstract among the students' English Department of Iain Metro in the academic year 2016/2017*. An undergraduate thesis. Institute Islamic Studies of Metro.
12. McGuire, N. (2018), *How accurate is Google Translate in 2018?* ARGO Translation.
13. Pham, A., Nguyen, Y., Tran, L., Huynh, K., Le, N., & Huynh, P. (2022), University students' perceptions on the use of Google Translate: Problems and solutions. *International Journal of Emerging Technologies in Learning (IJET)*, 17(4), 79–94. <https://doi.org/10.3991/ijet.v17i04.28179>.
14. Tongpoon-Patanasorn, A., & Griffith, K. (2020), Google Translate and Translation Quality: A Case of Translating Academic Abstracts from Thai to English. *PASSA*, 60(1), 134-163. <http://dx.doi.org/10.58837/CHULA.PASAA.60.1.5>.
15. Tsai, S. C. (2023), Learning with mobile augmented reality-and automatic speech recognition-based materials for English listening and speaking skills: Effectiveness and perceptions of non-English major English as a foreign language students. *Journal of Educational Computing Research*, 61(2), 444–465. <https://doi.org/10.1177/07356331221111203>.

Linguistic errors on English-Vietnamese text translation using Google translate

Abstract: In recent years, machine translation has gained popularity in translation research and English teaching. With its user-friendly features and convenient interface, Google Translate has become the most widely used machine translation application. The accuracy of Google Translate's translations, however, continues to be a source of debate among academics worldwide. From that perspective, this research was conducted to find out the linguistic errors of Google Translate when translating texts from English to Vietnamese. To achieve the research purpose, I used the method of synthesizing and analyzing errors on Google Translate's translated texts from group assignments of two basic English-Vietnamese translation classes I was assigned to teach directly at Vietnam National University. The findings indicate that grammatical, lexical, and semantic problems are the three primary categories of errors in Google Translate. By highlighting Google Translate's shortcomings, the study is hoped to help educators and students choose more effective teaching and learning strategies for translation and English subjects.

Key words: Machine translation; Google Translate; language errors; English-Vietnamese translation; Vietnamese students.