

NGOẠI NGỮ VỚI BẢN NGỮ

ỨNG DỤNG HÌNH THỨC DỊCH MÁY VÀO SOẠN THẢO VĂN BẢN PHẠM HẰNG NGA*

TÓM TẮT: Ngày nay, cùng với sự phát triển của khoa học công nghệ, ngày càng có nhiều hệ thống dịch tự động được giới thiệu trên Internet và nhận được nhiều sự quan tâm của người sử dụng. Tuy nhiên hiện vẫn có nhiều tranh luận khác nhau về chất lượng và mức độ tin cậy của các kết quả dịch máy. Nghiên cứu tìm hiểu việc ứng dụng hình thức dịch máy vào soạn thảo văn bản giữa cặp ngôn ngữ Việt-Anh qua việc phân tích các lỗi thường gặp khi dịch thuật và quan điểm của người sử dụng với hệ thống dịch máy. Kết quả nghiên cứu cho thấy nhóm lỗi ngữ pháp xuất hiện nhiều nhất trong các văn bản được phân tích. Các nhóm lỗi từ vựng, ngữ nghĩa và diễn ngôn là các nhóm lỗi phổ biến tiếp theo và có một tỉ lệ tương đối cân bằng nhau. Dựa trên phân tích này, người sử dụng và phiên dịch viên có thể tập trung hơn tới các yếu tố đó trong quá trình hiệu đính các văn bản dịch máy thô. Mặt khác, nghiên cứu cũng ghi nhận những phản hồi tích cực của người sử dụng với hình thức dịch máy. Các điểm hạn chế cũng như các lỗi thường gặp đã được xác định qua nghiên cứu sẽ đem lại những định hướng phù hợp hơn để người sử dụng áp dụng trong quá trình dịch máy.

TỪ KHOẢ: dịch thuật; dịch máy; dịch tự động; lỗi dịch thuật; thái độ.

NHẬN BÀI: 27/02/2023.

BIÊN TẬP-CHỈNH SỬA-DUYỆT ĐĂNG: 09/05/2023

1. Đặt vấn đề

Trong kỉ nguyên của công nghệ thông tin ngày nay, sự phát triển của khoa học và công nghệ đã mang lại lợi ích cho nhiều lĩnh vực của cuộc sống. Tiềm năng và ảnh hưởng của công nghệ thông tin trong lĩnh vực ngôn ngữ vì vậy cũng ngày càng tăng cao. Công nghệ thông tin nói chung và các công cụ dịch máy nói riêng đã phần nào hỗ trợ người sử dụng trong quá trình ứng dụng ngoại ngữ. Việc áp dụng các ứng dụng này đem lại sự thích thú, hào hứng cho người sử dụng. Tuy nhiên, bên cạnh đó cũng còn nhiều e ngại về chất lượng của những bản dịch mà các công cụ này đem lại. Các nghiên cứu trước đây về việc áp dụng công cụ dịch tự động vào lĩnh vực ngôn ngữ thường được tiến hành trong bối cảnh trường học. Tuy nhiên, không thể phủ nhận rằng bên cạnh đối tượng người học, còn có một số lượng lớn những người sử dụng công nghệ này là những người đi làm. Vì vậy, nghiên cứu tìm hiểu việc ứng dụng công cụ dịch tự động vào việc soạn thảo văn bản của nhóm đối tượng này. Nghiên cứu cũng hướng tới tìm hiểu tiềm năng của các ứng dụng dịch tự động cũng như tính hữu dụng của chúng để hỗ trợ người dùng trong soạn thảo văn bản.

2. Cơ sở lí thuyết

Những nghiên cứu và phát triển về dịch máy đã được bắt đầu thực hiện từ cách đây gần 100 năm. Ngay từ những năm ba mươi của thế kỉ trước, đã có hai nghiên cứu về quá trình dịch máy được cấp bằng sáng chế Melby (1995). Theo Hutchins và Somers (1992), khái niệm dịch máy trước đây còn được gọi là dịch cơ khí (mechanical translation) hay dịch tự động (automatic translation). Tuy nhiên, hiện nay trong tiếng Anh, thuật ngữ dịch máy được sử dụng phổ biến hơn để chỉ quá trình trong đó một hệ thống được vi tính hoá sản xuất ra bản dịch từ ngôn ngữ này sang ngôn ngữ khác. Quá trình này có thể có hoặc không có sự hỗ trợ của người dịch.

Almahasees (2022) gần đây cũng đưa ra quan điểm tương đồng về khái niệm dịch máy. Theo tác giả này, dịch máy cũng là một quá trình sử dụng máy tính để cung cấp bản dịch từ một ngôn ngữ này sang một ngôn ngữ khác. Tuy nhiên, định nghĩa của Almahasees (2022) cụ thể hoá các công cụ có thể được sử dụng trong quá trình dịch máy, bao gồm các từ điển đơn ngữ/ song ngữ, mạng lưới kho ngữ liệu học và nơ ron và các quy trình thuật toán. Như vậy, dịch máy vẫn có những đặc điểm giống

* Trường Đại học Giao thông Vận tải; Email: ngaph@utc.edu.vn

như hình thức dịch thuật truyền thống. Hai hình thức này đều hướng tới một mục tiêu truyền đạt ý nghĩa từ một ngôn ngữ nguồn sang ngôn ngữ đích. Hình thức dịch máy chỉ khác với phương pháp truyền thống ở việc sử dụng công cụ là máy tính thay vì con người (Almahasees, 2022).

Tuỳ thuộc vào các mức độ khác nhau của tự động hoá trong quá trình dịch thuật, dịch máy có thể được chia thành các nhóm khác nhau. Các nghiên cứu của Lehrberger và Bourbeau (1988) và Hutchins và Somers (1992) cùng chỉ ra ba hình thức dịch máy.

Machine-aided human translation (MAHT; Dịch bởi con người có trợ giúp của máy): Hình thức này được đánh giá là hình thức gần nhất với phương pháp dịch truyền thống. Người dịch sẽ vẫn chịu trách nhiệm chính để đưa ra bản dịch. Máy móc sẽ có những hỗ trợ giới hạn như các chương trình kiểm tra chính tả, ngữ pháp hoặc văn phong.

Human-aided machine translation (HAMT; Dịch máy có trợ giúp của người): Dựa trên bản dịch thô do máy tính sản xuất, người dịch sẽ hiệu đính lại để có được sản phẩm dịch hoàn chỉnh cuối cùng. Hai nghiên cứu của các nhóm tác giả trên cũng đều thống nhất rằng sự can thiệp của người dịch với quá trình hiệu đính có thể diễn ra tại bất kì giai đoạn nào của quá trình dịch máy, từ trước khi bắt đầu, trong khi dịch và sau khi dịch.

Fully automatic machine translation (FAMT; dịch máy hoàn toàn tự động): Trong hình thức này, con người sẽ không can thiệp vào quá trình chuyển đổi văn bản gốc thành bản dịch thô. Tuy nhiên, việc chỉnh sửa lại bản dịch thô sau đó có thể được tiến hành. Dịch máy tự động hoàn toàn không có nghĩa là loại bỏ việc hiệu đính của con người sau đó Lehrberger và Bourbeau (1988). Theo Nguyễn Ninh Bắc (2014), với hình thức này người sử dụng có thể là những người không chuyên về dịch thuật hoặc không biết ngôn ngữ gốc và ngôn ngữ đích cần dịch, người dùng chỉ cần nhập đoạn văn bản cần dịch vào ứng dụng và ra lệnh dịch. Tự máy tính sẽ đưa ra một bản dịch trọn vẹn.

Các nhà nghiên cứu Hutchins và Somers (1992) một lần nữa xác nhận việc phân định rõ các hình thức dịch máy không dễ dàng, vì vậy, chúng có thể được gọi chung là *Computer-aided translation* (dịch thuật có máy hỗ trợ).

Sự phổ biến và tầm ảnh hưởng lan rộng của dịch máy đã phần nào chứng minh được tính hiệu quả của những công cụ này. Một trong những ưu điểm của nhiều chương trình dịch máy hiện nay đó chính là tốc độ dịch thuật. Đa phần các công cụ dịch thuật tự động có thể cho ra một bản dịch với thời gian ngắn hơn so với người dịch Trujillo (1999). Almahasees (2022) bổ sung thêm một lợi thế của dịch máy đó là tính dễ tiếp cận và chi phí thấp thậm chí là miễn phí. Trong bối cảnh toàn cầu hoá, nhu cầu về dịch thuật ngày càng tăng cao. Tuy nhiên chi phí dịch thuật chuyên nghiệp do các phiên dịch viên cung cấp đôi khi vẫn vượt quá khả năng tài chính của người dùng. Google Translate là một trong những minh hoạ điển hình của ưu điểm này. Đây có thể được coi là một trong các lí do chủ yếu khiến cho các công cụ dịch máy rất phổ biến với đa số người sử dụng. Công Tôn Huyền Tôn Nữ Ý Nhiệm (2016) đã chỉ ra rằng, nhu cầu sử dụng công cụ dịch tự động là một xu hướng tất yếu. Trong lĩnh vực nghiên cứu khoa học, nhờ có các công cụ dịch tự động, các nhà nghiên cứu không còn gặp khó khăn khi tiếp cận những nguồn tài liệu bằng các ngôn ngữ khác nhau. Bên cạnh đó, công cụ dịch tự động còn hỗ trợ việc giải mã ngôn ngữ để nắm bắt được các nội dung của tài liệu. Tác giả cũng nhấn mạnh rằng, đối với đối tượng này, những lỗi ngữ pháp hoặc chất lượng bản dịch cũng không được đòi hỏi quá cao vì nhóm đối tượng này chủ yếu cần nắm được các vấn đề cốt lõi và tinh thần nội dung của văn bản gốc. Điều này cũng có thể được mở rộng ra cho các nhóm đối tượng khác ví dụ như các sinh viên không chuyên về ngoại ngữ và có nhu cầu tiếp cận các tài liệu chuyên ngành. Nhóm đối tượng người dùng là người đi làm cũng có thể tận dụng và phát huy ưu điểm này của công cụ dịch tự động.

Bên cạnh những ưu điểm nổi bật của dịch máy, hình thức dịch thuật này vẫn còn tồn tại một số nhược điểm. Theo Almahasees (2022), chưa có sự tương quan giữa tốc độ dịch thuật và chất lượng. Điều này có thể được hiểu rằng, mặc dù dịch máy có thể đưa ra những bản dịch trong thời gian ngắn hơn rất nhiều so với người dịch, nhưng chất lượng của những bản dịch đó có thể chưa cao. Nhược

điểm tiếp theo của dịch máy mà tác giả Almahasees (2022) bổ sung thêm việc hệ thống dịch máy có thể cung cấp một bản dịch với chất lượng tốt ở lần này nhưng lần khác lại là một bản dịch chưa đạt yêu cầu. Điều đó cho thấy tính ổn định, đặc biệt về chất lượng của các hệ thống dịch máy vẫn còn là một quan ngại với người sử dụng.

Nguyễn Văn Bình và Huỳnh Công Pháp (2021) nhận định, mặc dù trong 20 năm phát triển gần đây, lĩnh vực dịch tự động đã đạt được những tiến triển vượt bậc, nhưng đến nay, kết quả của các hệ thống dịch máy vẫn chưa đạt được hiệu quả cao như các bản dịch do con người thực hiện. Tính chính xác của hệ thống dịch máy có thể cao hơn với một số các ngôn ngữ phổ biến như tiếng Anh, tiếng Pháp và trong một số lĩnh vực thông dụng. Tuy nhiên, đối với nhóm ngôn ngữ ít hoặc các lĩnh vực chuyên ngành như y tế, kỹ thuật, pháp lý, công nghệ, chất lượng của hệ thống dịch máy còn thấp, khó có thể áp dụng rộng rãi trong thực tế do các hệ thống dịch máy không đưa ra được các kết quả dịch đúng với các khái niệm chuyên môn, làm cho nhiều văn bản dịch trở nên khó khăn, không có giá trị sử dụng cao. Nguyễn Ninh Bắc (2014) cũng đồng tình với quan điểm này. Tác giả khẳng định hình thức dịch tự động hoàn toàn bằng máy tính có thể được áp dụng tương đối hiệu quả trong các trường hợp khi ngôn ngữ gốc và ngôn ngữ đích có chung ngữ hệ và cách diễn đạt tương đối giống nhau. Bên cạnh đó, nếu văn bản gốc chứa nhiều phép tu từ, nghĩa bóng thì bản dịch công nghệ này chỉ có thể dùng để tham khảo và để hiểu được nội dung sơ bộ. Vì vậy, việc người phiên dịch rà soát và hiệu đính lại các sản phẩm dịch máy là rất cần thiết để các văn bản có thể được sử dụng với mức độ chính xác cao hơn và phù hợp với ngữ cảnh cũng như yêu cầu sử dụng.

Huỳnh Công Pháp và Nguyễn Văn Bình (2017) có nêu trong lĩnh vực dịch máy, hiệu đính có thể được hiểu là quá trình trong đó bản dịch thô của máy tính được chỉnh sửa và hoàn thiện bởi con người. Qua đó, chất lượng bản dịch có thể được cải thiện và đạt được chất lượng tiêu chuẩn. Việc hiệu đính có thể được tiến hành dưới nhiều hình thức. Nếu bản dịch chỉ được chỉnh sửa với mục đích có thể hiểu được nội dung cơ bản và tiết kiệm thời gian nhất có thể, đây là hình thức hiệu đính sơ bộ (light-post-editing). Trong khi đó, hiệu đính tổng quát (full-post editing) sẽ đem lại chất lượng cao hơn khi văn bản cuối cùng sẽ có văn phong, ngữ pháp và ngữ nghĩa thích hợp. Nếu văn bản dịch tự động tốt sẽ tốn ít thời gian để hiệu đính hơn một văn bản dịch tự động chưa tốt và ngược lại. Các tác giả cũng một lần nữa khẳng định việc hiệu đính với các bản dịch máy là cần thiết để hoàn thiện bản dịch. Trong quá trình hiệu đính, từ ngữ, chính tả, tên riêng, thuật ngữ chuyên ngành có thể được rà soát, kiểm tra. Thêm vào đó, quá trình hiệu đính cũng giúp xây dựng kho ngữ liệu để phục vụ đánh giá và nâng cao chất lượng các hệ thống dịch máy.

Theo Almahasees (2022), có nhiều hình thức đã được áp dụng để đánh giá kết quả của các hệ thống dịch máy, phân tích lỗi là một trong những phương pháp đó. Đây được coi như một công cụ quan trọng trong việc tìm ra nguyên nhân cũng như đưa ra biện pháp để cải thiện các công cụ dịch máy. Almahasees (2022) và Coasta et al. (2015) đều đưa ra một số những nhóm lỗi nhất định có thể bắt gặp trong các bản dịch máy bao gồm chính tả, từ vựng, ngữ pháp, ngữ nghĩa và diễn ngôn.

Theo Công Huyền Tôn Nữ Ý Nhiệm (2014), mặc dù vấn đề chất lượng do các hệ thống dịch tự động vẫn luôn gây tranh cãi tuy nhiên các hệ thống này vẫn được ưu chuộng và sử dụng rộng rãi không chỉ với người học ngoại ngữ mà cả với người không chuyên về ngoại ngữ. Vì vậy, các công ty tin học trên thế giới vẫn liên tục nghiên cứu và phát triển để nâng cao chất lượng của các công cụ và ứng dụng này.

3. Phương pháp nghiên cứu

3.1. Câu hỏi nghiên cứu

Nghiên cứu được tiến hành để tìm hiểu việc áp dụng các hệ thống dịch máy trong soạn thảo văn bản với đối tượng người sử dụng là những người đi làm dựa trên các câu hỏi nghiên cứu:

1. Các lỗi thường gặp khi sử dụng công cụ dịch máy để soạn thảo văn bản là gì?
2. Thái độ của người sử dụng đối với công cụ dịch máy trong soạn thảo văn bản như thế nào?

3.2. Phương pháp nghiên cứu và cách thức tiến hành

Nghiên cứu áp dụng phương pháp nghiên cứu tình huống (case study). Nghiên cứu được tiến hành với 45 đối tượng là các chuyên viên thuộc Ban Quản lý đường sắt đô thị Hà Nội (MBR) với 83,7% người tham gia trên 35 tuổi, 16,3% trong độ tuổi từ 20-34. Do đặc thù chuyên môn của cơ quan chuyên về giao thông và kỹ thuật, 88,4% đối tượng tham gia nghiên cứu là nam giới, chỉ có 11,6% đối tượng là nữ giới. Các đối tượng nghiên cứu cũng có trình độ tiếng Anh khác nhau theo Khung tham chiếu ngôn ngữ châu Âu (CEFR). 25,6% người tham gia nghiên cứu có trình độ ngoại ngữ A2. Trình độ ngoại ngữ B1 chiếm một tỉ lệ lớn nhất trong nhóm đối tượng nghiên cứu với 34,9%. Một tỉ lệ tương đối cao các đối tượng nghiên cứu có trình độ từ B2 trở lên (34,9%) và vẫn còn 4,7% người tham gia đánh giá trình độ ngoại ngữ của mình chỉ ở mức độ A1.

Do tính chất trọng điểm của dự án mà MRB đang tiến hành, dự án có sự tham gia của các đối tác nước ngoài như tổng thầu, tư vấn, nhà tài trợ. Do đó việc trao đổi văn bản được tiến hành theo hai hình thức tiếng Việt và tiếng Anh.

Nghiên cứu được tiến hành theo quy trình như sau. Đầu tiên, các đối tượng soạn thảo các văn bản bằng tiếng Việt. Sau đó, họ sử dụng các công cụ dịch thuật tự động để tạo ra một bản dịch thô bằng tiếng Anh. Bản dịch thô sẽ được phiên dịch viên rà soát và chỉnh sửa để có được một phiên bản cuối cùng hoàn thiện nhất để sử dụng liên hệ với các đối tác nước ngoài.

Các bản dịch thô cũng được phân tích để tìm hiểu tần suất của các lỗi dịch thuật.

Thêm vào đó, một bảng câu hỏi điều tra cũng đã được thực hiện để xác nhận các kết quả phân tích lỗi dịch thuật do các hệ thống máy đồng thời cũng để tìm hiểu cảm nhận của người sử dụng đối với các hệ thống dịch máy.

4. Kết quả và thảo luận

4.1. Tổng hợp và phân tích các lỗi trong các văn bản dịch máy

Kết quả phân tích các cặp văn bản dịch Việt-Anh với tổng số lượng trung bình 150-200 từ/văn bản, trong số đó có 272 lỗi thuộc các nhóm khác nhau được phát hiện. Tỉ lệ ghi nhận trung bình 13,6 lỗi trên một văn bản.

Dựa trên nghiên cứu của Almahasees (2022) và Coasta et al. (2015), các loại lỗi trong các văn bản dịch có thể được phân loại chi tiết như sau:

Bảng 1. Kết quả phân tích lỗi sai dịch máy

STT	Nhóm lỗi/ tỉ lệ	Chi tiết
1	Chính tả 9,92%	Lỗi Ngày tháng 77,78%
		Lỗi chính tả 22,22%
2	Từ vựng 19,10%	Lỗi thiếu từ 55,77%
		Lỗi thêm từ 23,06%
		Lỗi không dịch 21,17%
3	Lỗi ngữ pháp 33,36%	Lỗi trật tự từ 9,89%
		Lỗi về sự hoà hợp 10,90%
		Lỗi động từ 41,76%
		Lỗi từ loại 24,17%
		Các lỗi khác 13,28%
4	Ngữ nghĩa 18,52%	Lỗi chọn chưa phù hợp 86,27%
		Lỗi Kết hợp từ 13,73%
5	Diễn ngôn 19,10%	Lỗi văn phong 45,09%
		Lỗi ngữ cảnh 54,91%

Theo kết quả phân tích ở trên, các lỗi ngữ pháp có tần suất xuất hiện cao nhất với 33,36%. Như vậy có thể suy luận rằng, công cụ dịch máy mắc các lỗi về ngữ pháp nhiều hơn các lỗi khác khi dịch

văn bản từ tiếng Việt sang tiếng Anh. Trong các nghiên cứu trước đó của Almahasees (2022) và Coasta et al. (2015), nhóm lỗi ngữ pháp của hệ thống dịch máy chiếm tỉ lệ ít hơn so với nhóm lỗi ngữ nghĩa và từ vựng, tuy nhiên tại nghiên cứu này, nhóm lỗi ngữ pháp lại nổi trội hơn. Điều này có thể được lí giải do sự khác biệt về mặt cấu trúc ngữ pháp; đồng thời, kho ngữ liệu cho dịch máy tiếng Việt được đánh giá còn khá khiêm tốn so với những nhóm ngôn ngữ khác Nguyễn Văn Bình và Huỳnh Công Pháp (2021). Ba nhóm lỗi từ vựng, ngữ nghĩa và diễn ngôn cùng chiếm tỉ lệ khá cân bằng là 19,10%, 18,52% và 19,10% trong tổng số các lỗi ghi nhận được. Như vậy, bên cạnh các lỗi ngữ pháp, các lỗi về từ vựng và ngữ nghĩa cũng khá phổ biến. Kết quả khi phân tích lỗi dịch máy giữa hai ngôn ngữ tiếng Việt và tiếng Anh này cũng tương đồng với một số những nghiên cứu trước đây khi phân tích lỗi của các cặp ngôn ngữ khác. Cụ thể, nghiên cứu của Coasta et al. (2015) cũng ghi nhận kết quả tương đồng giữa nhóm lỗi từ vựng và ngữ nghĩa khi áp dụng dịch máy giữa cặp ngôn ngữ Tiếng Bồ Đào Nha và tiếng Anh. Như vậy, trong quá trình hiệu đính sau dịch máy, người dịch cần phải lưu ý đến các khía cạnh này để có được bản dịch đạt yêu cầu.

Phân tích lỗi cũng ghi nhận các kết quả giống với nghiên cứu của Coasta et al. (2015) và Almahasees (2020) trong đó so với các nhóm lỗi nêu trên, các lỗi về chính tả ít nghiêm trọng nhất với 9,92%. Các ứng dụng dịch máy cũng giống như công cụ trong Microsoft Word đều có thể hỗ trợ khá tốt việc kiểm tra chính tả của văn bản. Tuy nhiên việc vẫn còn tồn tại những lỗi chính tả trong các văn bản dịch máy cho thấy bản thân người sử dụng đôi khi vẫn còn chủ quan và bất cẩn khi đánh máy.

Cột thứ ba trong bảng thể hiện số phần trăm chi tiết của từng loại lỗi cụ thể trong từng nhóm lỗi lớn tương ứng.

a) Nhóm lỗi chính tả

Chiếm tỉ lệ lớn hơn trong nhóm lỗi chính tả chính là việc sử dụng ngày tháng. Người dùng vẫn có xu hướng diễn đạt ngày tháng trong các văn bản tiếng Anh theo định dạng tiếng Việt (ví dụ sử dụng 30/9/2022 thay vì 30 September 2022). Cách viết này tuy không ảnh hưởng đến ý nghĩa và việc hiểu nội dung, tuy nhiên, với những văn bản chính thức gửi tới các đối tác quốc tế cũng như các bên tài trợ vốn của MRB, việc các cụm từ ngày tháng được diễn đạt đúng văn phong sẽ phù hợp hơn. Các chi tiết nhỏ này cũng góp phần thể hiện tính chuyên nghiệp của cơ quan đối với các đối tác nước ngoài.

b) Nhóm lỗi từ vựng

Kết quả tổng hợp cho thấy trong nhóm từ vựng, lỗi thiếu từ có tần xuất mắc phải cao nhất trong quá trình dịch máy với 57,77%.

Ví dụ:

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Tiến độ thương thảo	Negotiation	Negotiation progress
Họp giao ban	Meeting	Progress meeting

Tuy nhiên phân tích cho thấy các từ dịch thiếu không quá nghiêm trọng và không ảnh hưởng nhiều đến việc truyền tải ý nghĩa và nội dung.

Nhóm lỗi không dịch với 21,17% trong nghiên cứu xảy ra với các từ viết tắt (HGU, PIC). Trong bối cảnh thông thường, nhóm từ không được dịch đặc biệt là các từ viết tắt cho các thuật ngữ chuyên môn sẽ gây khó khăn cho người sử dụng. Tuy nhiên, trong ngữ cảnh này, người soạn thảo và người nhận văn bản đều nắm được bối cảnh của công việc cũng như các từ viết tắt đều được diễn giải trong tài liệu hợp đồng dự án, vì vậy tương tự như nhóm lỗi thiếu từ, việc không dịch các từ viết tắt không gây nhiều khó khăn cho người nhận bản dịch. Trong nhóm này, lỗi thêm từ phổ biến thứ 2, chiếm 23,06%.

Ví dụ:

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Gói thầu	Bidding package	Package

c) Nhóm lỗi ngữ pháp

Các lỗi về ngữ pháp phổ biến nhất trong các bản dịch máy và cũng có nhiều nhóm đa dạng nhất, trong đó lỗi về động từ chiếm tỉ lệ lớn nhất với 41,76%. Do cách sử dụng động từ trong tiếng Việt đa phần đơn giản hơn so với cách chia thì của động từ trong tiếng Anh, nên tỉ lệ các lỗi về chia thì của động từ xuất hiện tương đối nhiều trong kết quả phân tích. Lỗi hình thái của động từ cũng thường xuất hiện trong nhóm lỗi này.

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Đến nay một số mốc độ tình báo cáo đã quá hạn	Today, there are some milestones to be expired	To date, there are several submission milestones which have been expired.
Đảm bảo các báo cáo của ABC được đệ trình trước...	...ensure that ABC's reports are submitted before...	...ensure that ABC's reports will be submitted before....

Bên cạnh đó, trong tiếng Việt, có hiện tượng các từ thuộc các nhóm từ loại khác nhau đôi khi có cách viết chính tả giống nhau (Bồ Thị Lý, 2019). Đây là một nguyên nhân gây ra khá nhiều khó khăn cho các hệ thống dịch máy khi chuyển đổi văn bản từ tiếng Việt sang tiếng Anh. Điều này cũng được chứng minh trong phân tích. Tỉ lệ lỗi này nhóm đứng vị trí thứ 2 trong nhóm lỗi ngữ pháp với 24,17%.

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Thống nhất (động từ) họp	Agreement (danh từ) to held a meeting	Agree to hold a meeting...
Nhiều lần nhắc nhở (động từ)	Several reminders (danh từ)	Have reminded several times

Sự khác biệt trong ngữ pháp tiếng Anh và tiếng Việt về sự hoà hợp số nhiều cũng là nguyên nhân phổ biến thứ hai khiến cho nhóm lỗi ngữ pháp chiếm một tỉ lệ lớn nhất. Lỗi về sự hoà hợp chiếm 10,90%. Các lỗi này trong nghiên cứu chủ yếu xuất hiện khi không có sự thống nhất giữa diễn giải số nhiều trong hai phiên bản tiếng Việt và tiếng Anh, ví dụ như:

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Các cập nhật	Update	Updates
Tư vấn ABC	ABC consultants	ABC consultant

d) Nhóm lỗi ngữ nghĩa

Nhóm lỗi về ngữ nghĩa trong nghiên cứu có thể được chia thành hai nhóm với tỉ lệ khá chênh lệch, lần lượt là 86,27% với lỗi lựa chọn từ sai hoặc chưa phù hợp và 13,73% lỗi về kết hợp từ (collocations).

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Gần nhất	The last	The latest
Thống nhất	Unify	To reach agreement
Thuế thu nhập cá nhân	Personal input	Personal Income tax

Đây có thể coi là một trong những nhóm lỗi quan trọng nhất vì nhóm lỗi này có ảnh hưởng lớn đến việc truyền tải ý nghĩa giữa văn bản nguồn và văn bản đích. Việc sử dụng từ với ý nghĩa sai trong quá trình dịch sẽ có thể dẫn tới kết quả dịch không chính xác, gây hiểu nhầm cho người đọc. Phân tích cho thấy, người sử dụng có thể mắc lỗi do việc sử dụng sai giữa các từ hay gây nhầm lẫn trong tiếng Anh (effect-affect, last-latest) hoặc do ảnh hưởng của yếu tố liên ngôn ngữ. Mặc dù đã có sự hỗ trợ của máy tính trong quá trình dịch thuật, nhưng do ảnh hưởng của tiếng mẹ đẻ, người sử dụng vẫn có thể lựa chọn những từ ngữ chưa phù hợp. Vì vậy, quá trình hiệu đính do một phiên dịch chuyên nghiệp càng chứng tỏ được sự cần thiết trong ngữ cảnh này bởi họ sẽ xác định được rõ các từ loại và cấu trúc phù hợp để sử dụng.

Về phương diện ngữ nghĩa, nhóm lỗi nhầm lẫn về ý nghĩa giữa các từ đồng âm được liệt kê trong các nghiên cứu trước đây của Coasta et al. (2015) và Almahasees (2020) không xuất hiện trong phân tích tại nghiên cứu này. Điều này chứng minh hệ thống dịch tự động đã có thể phân biệt khá tốt các từ đồng nghĩa khi dịch từ tiếng Việt sang tiếng Anh. Có thể suy ra rằng tính năng này ngày càng được cải thiện so với thời gian trước đây.

e) Nhóm lỗi diễn ngôn

Nhóm lỗi về diễn ngôn có thể được phân chia thành hai nhóm lỗi nổi bật bao gồm lỗi về văn phong chiếm 45,09% và lỗi về ngữ cảnh với 54,91%.

Lỗi về văn phong có thể không phải là những lỗi sai về mặt ngữ pháp, từ vựng hoặc ý nghĩa nhưng ngôn ngữ sử dụng trong các trường hợp này chưa phù hợp với yêu cầu học thuật của văn bản. Với vai trò là các công văn chính thức của MRB gửi tới các bên liên quan, các văn bản này cũng sẽ đóng vai trò pháp lý cho những công việc của MRB vì vậy mức độ trang trọng của ngôn ngữ cần phải đáp ứng được yêu cầu. Tuy nhiên, một số cách diễn đạt trong các văn bản được phân tích vẫn được đánh giá là chưa phù hợp, do đó, các cách diễn đạt này đã được biên dịch viên chỉnh sửa lại cho phù hợp hơn.

Ví dụ:

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Hàng tuần	Weekly	On weekly basis
Sau khi thi công xong...	After the construction is done...	After finishing the construction...

Ngoài ra yếu tố văn phong còn có thể được xét đến cả tính cô đọng hay tính lặp của ngôn ngữ.

Ví dụ:

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Gia hạn và giảm giá trị bảo lãnh tạm ứng	Extend the guarantee for advanced payment and to reduce the advanced payment	Extend and reduce the advanced payment
Tiến hành các thủ tục...	Carry out procedures...	Implement procedures....

Nguyên nhân của yếu tố văn phong có thể được xem xét bởi thực tế những đối tượng nghiên cứu đa phần có nền tảng đào tạo thiên về kỹ thuật thay vì ngôn ngữ và họ cũng làm việc trong môi trường kỹ thuật nhiều hơn. Vì vậy cũng có thể hiểu được việc họ chưa thành thạo những cách diễn đạt đúng văn phong. Một lần nữa, cần có phiên dịch chuyên nghiệp hiệu đính lại văn bản để hỗ trợ người sử dụng cách thức diễn ngôn.

Các đối tượng nghiên cứu khi được khảo sát, cũng bày tỏ quan ngại với yếu tố văn phong diễn đạt trong các bản dịch máy. 27,9% người sử dụng có đánh giá cao về văn phong của kết quả dịch máy. 46,5% chỉ cho mức điểm trung bình và 25,7% thậm chí chỉ đánh giá dưới trung bình cho yếu tố này của dịch máy.

Yếu tố ngữ cảnh dẫn tới 54,91% số lỗi trong nhóm diễn ngôn. Một lần nữa, yếu tố lỗi này không gây ra nhầm lẫn nghiêm trọng về mặt truyền đạt ý nghĩa của kết quả dịch máy nhưng lại được đánh giá không phù hợp với bối cảnh mà văn bản được sử dụng. Almahasees (2020) chỉ ra rằng người dịch có thể phân biệt được các ngữ cảnh khác nhau khi dịch ví dụ ngữ cảnh gia đình, tình thân. Trong khi đó, các hệ thống dịch máy vẫn còn khá hạn chế trong vấn đề này. Khía cạnh này một lần nữa được chứng minh trong nghiên cứu với một loạt các minh chứng như sau:

<i>Tiếng Việt</i>	<i>Bản dịch máy tiếng Anh</i>	<i>Bản tiếng Anh được người dịch hiệu đính</i>
Hoàn trả	Return	Reinstate
Rào chắn thi công	Construction barrier	Construction fence
Chủ đầu tư	Investor	Employer

Bản thân người sử dụng cũng ý thức được rằng ngữ cảnh là một yếu tố sẽ khó được diễn tả hoàn chỉnh trong các kết quả dịch máy. Điều này được thể hiện qua kết khảo sát với chỉ 32,2% người khảo sát đánh giá cao yếu tố ngữ cảnh do hệ thống dịch máy truyền tải, 37,2% giữ quan điểm trung lập và có tới một phần ba còn lại đánh giá thấp tiêu chí này.

Có một hiện tượng thú vị đã xảy ra trong nghiên cứu khi các đối tượng nghiên cứu đặt niềm tin tương đối lớn vào phiên bản dịch máy và không kiểm tra lại một số thuật ngữ và tên gọi quen thuộc gắn liền với cơ quan của họ. Họ vẫn giữ nguyên phiên bản tên riêng chưa đúng do dịch máy sản xuất khi chuyển văn bản để phiên dịch viên hiệu đính. Ví dụ như Hanoi Urban Railway Management Board thay vì Hanoi Metropolitan Railway Management Board (Ban quản lý đường sắt đô thị Hà Nội); Hanoi pilot urban railway project of Hanoi city thay vì Hanoi pilot light metro line project (Dự án tuyến đường sắt đô thị thí điểm thành phố Hà Nội). Điều này cũng được chứng minh trong khảo sát. 90% người sử dụng đánh giá khá cao tính chính xác của hệ thống dịch tự động.

4.2. Thái độ của người sử dụng đối với công cụ dịch máy

Kết quả khảo sát cho thấy rõ ràng hơn quan điểm và thái độ của người sử dụng đối với các hệ thống dịch máy. Đa số người sử dụng (hơn 60%) khẳng định họ thường xuyên, rất thường xuyên sử dụng công cụ dịch tự động khi soạn thảo văn bản. Điều này chứng minh mức độ phổ biến của các công cụ này đối với nhóm đối tượng người đi làm. Công cụ dịch tự động Google Translate phổ biến nhất. 100% đối tượng nghiên cứu biết và sử dụng công cụ này. Cambridge Dictionary và Microsoft Translator là hai ứng dụng đứng vị trí tiếp theo về mức độ phổ biến, với tỉ lệ phần trăm lần lượt là 25,6% và 20,9% người dùng. Hai ứng dụng Viki-translation và Yandex Translate còn khá mới mẻ với người sử dụng, chỉ 3 người dùng biết tới các ứng dụng này.

Các đối tượng khảo sát đồng ý các ứng dụng dịch tự động hữu ích với việc soạn thảo văn bản bằng tiếng Anh dựa trên một số ưu điểm của các công cụ dịch máy. Ưu điểm về tốc độ dịch thuật được nhiều người sử dụng đánh giá cao với 72% đồng ý và rất đồng ý. Đa số đối tượng khảo sát (79%) đánh giá ứng dụng dịch thuật dễ sử dụng. Trái lại 11,1% không đồng ý và hoàn toàn không đồng ý với ý kiến trên, 9% giữ quan điểm trung lập. Chỉ có 11,1% người dùng gặp khó khăn khi sử dụng các ứng dụng này. Một tỉ lệ tương đối cao đối tượng nghiên cứu (70%) còn đánh giá cao ưu điểm khi các ứng dụng dịch tự động hỗ trợ đa dạng các ngôn ngữ khác nhau.

Bên cạnh những ưu điểm, đối tượng nghiên cứu còn bày tỏ thái độ và quan điểm của họ đối với nhược điểm của các hệ thống dịch máy. Đối tượng nghiên cứu khá cẩn trọng với sự chính xác của kết quả dịch thuật ngữ chuyên ngành trên hệ thống dịch máy. 67% người sử dụng bày tỏ quan ngại với

chất lượng dịch thuật ngữ chuyên ngành, chỉ có 33% tin tưởng vào kết quả dịch máy các thuật ngữ chuyên ngành. Hơn 50% đối tượng nghiên cứu đồng tình với các nhược điểm khác của dịch tự động như vẫn còn giới hạn về số lượng kí tự cho mỗi lần dịch và độ chính xác của dịch máy không nhất quán giữa các ngôn ngữ khác nhau. Hi vọng trong tương lai, các hệ thống dịch máy có thể tiếp tục được nâng cấp để dịch được những văn bản dài hơn và có tính ổn định cao hơn cho các ngôn ngữ không phổ biến ví dụ như tiếng Việt. Bên cạnh đó, các yếu tố như ngữ pháp, từ, các thuật ngữ chuyên ngành cũng chỉ nhận được mức đánh giá trung bình từ phía đối tượng nghiên cứu. Hơn 50% người sử dụng cũng e ngại ảnh hưởng tiêu cực của công cụ dịch tự động với việc phát triển kĩ năng viết của họ. Việc phụ thuộc quá mức vào các công cụ sẵn có này sẽ làm giảm tính chủ động trong việc sử dụng ngôn ngữ và dần dần ảnh hưởng đến năng lực ngôn ngữ của người sử dụng. Như vậy, có thể nhận thấy, người sử dụng cũng có những đánh giá và nhận định nhất định về ưu điểm cũng như những điểm hạn chế của các hệ thống dịch tự động, vì vậy, 72% người dùng khẳng định họ thường xuyên và luôn luôn thực hiện việc kiểm tra và hiệu đính bản dịch tự động thô. 28% còn lại thỉnh thoảng và hiếm khi tiến hành việc rà soát. Không có đối tượng nào không thực hiện các bước hiệu đính sau khi nhận được bản dịch thô.

Mặc dù vẫn còn một số hạn chế, những bản dịch tự động vẫn được người sử dụng đánh giá cao. Và 100% khẳng định sẽ vẫn tiếp tục sử dụng hình thức này trong tương lai để soạn thảo văn bản và tra cứu.

5. Kết luận

Có thể kết luận rằng, các đối tượng nghiên cứu thường xuyên sử dụng các công cụ dịch máy vào quá trình soạn thảo văn bản để hỗ trợ họ chuyển đổi văn bản tiếng Việt sang tiếng Anh phục vụ cho công việc với đối tác nước ngoài. Kết quả phân tích cho thấy các nhóm lỗi ngữ pháp, từ vựng, ngữ nghĩa và diễn ngôn xuất hiện phổ biến trong các bản dịch máy Việt-Anh. Vì vậy, trong quá trình hiệu đính, người dịch cần lưu ý nhiều hơn đến các yếu tố này để nâng cao độ chính xác của bản dịch.

Quan điểm của người sử dụng cũng được đề cập trong nghiên cứu này. Người sử dụng cũng ghi nhận những ưu điểm nổi bật của các công cụ dịch tự động như tốc độ, dễ sử dụng và có thể sử dụng với nhiều cặp ngôn ngữ khác nhau.

Tuy nhiên, đối tượng khảo sát cũng còn khá quan ngại về tính chính xác trong dịch máy của các yếu tố về ngữ pháp và các thuật ngữ ngành. Vì các đối tượng nghiên cứu có nền tảng giáo dục và đào tạo không chuyên về ngôn ngữ, vì vậy họ cũng cần được định hướng về những ưu điểm và nhược điểm của các hệ thống dịch máy để áp dụng phù hợp trong quá trình sử dụng (Nga & Chen, 2020). Nguyễn Ninh Bắc (2014) cũng nhấn mạnh, các công cụ hỗ trợ dịch thuật không nên được xem như hình thức thay thế dịch thuật thông thường hay vai trò của phiên dịch. Phần mềm dịch thuật hỗ trợ tăng tốc độ dịch và tính thống nhất của thuật ngữ trong một số loại văn bản nhất định. Đây cũng là một khía cạnh khác được ghi nhận trong bài báo này với hơn một nửa đối tượng nghiên cứu đồng tình rằng việc lạm dụng các công cụ dịch máy không tốt cho kĩ năng viết của họ. Do đó, người sử dụng không nên lạm dụng hay ỷ lại vào các công cụ dịch máy. Người dùng nên chủ động trau dồi kĩ năng dịch và kĩ năng viết để áp dụng vào việc soạn thảo văn bản cũng như phục vụ cho công việc. Nghiên cứu cũng góp phần củng cố thêm khả năng ứng dụng công nghệ thông tin cũng như trí tuệ nhân tạo vào lĩnh vực ngoại ngữ.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Bò Thị Lý (2019), "Phân tích các lỗi sai từ vựng trong bài viết của sinh viên hệ vừa học vừa làm, Trường Đại học Ngoại ngữ, ĐHQG Hà Nội". Tạp chí *Ngôn ngữ & Đời sống*, số 4(284), tr.58-64.
2. Công Huyền Tôn Nữ Ý Nhiệm (2016), *Dịch tự động và kỹ năng viết của Sinh viên*, trong Kì yếu Hội thảo Quốc gia "Nghiên cứu Liên ngành về Ngôn ngữ và Giảng dạy Ngôn ngữ lần hai", Huế, ngày 29-30 tháng 12 năm 2016, Nxb ĐH Huế, Huế, tr.93-105.

3. Huỳnh Công Pháp & Nguyễn Văn Bình (2017), "Cải tiến chất lượng dịch máy kết hợp giải pháp xây dựng kho ngữ liệu phục vụ đánh giá chất lượng hệ thống dịch tự động tiếng Việt". Tạp chí *Khoa học và công nghệ* (Đại học Đà Nẵng), số 05(114), 46-51.
4. Nguyễn Ninh Bắc (2014), *Tổng quan về ứng dụng công nghệ thông tin trong công việc biên dịch*. Nghiên cứu nước ngoài (Trường ĐH Ngoại ngữ, ĐHQG Hà Nội), 30(4), tr.55-64.
5. Nguyễn Văn Bình & Huỳnh Công Pháp (2021), "Đánh giá vai trò của kho ngữ liệu đối với chất lượng dịch tự động tiếng Việt". Tạp chí *Khoa học và công nghệ*, (Đại học Đà Nẵng), 19(1), tr.47-51.

Tiếng Anh

6. Almahasees, Z. (2022), *Analysing English-Arabic Machine Translation: Google Translate, Microsoft Translator and Sakhr*. New York: Routledge.
7. Coasta, A., Ling, W., Luis, T., Correia, R., & Coheur, L. (2015), A linguistically motivated taxonomy for machine translation error analysis. *Machine Translation*, 29(2), 127-161.
8. Huchin, W. J., & Somers, H. L. (1992), *An introduction to machine translation*. London: Academic Press Limited.
9. Lehrberger, J. & Bourbeau, L. (1988), *Machine translation: Linguistic characteristics of MT systems and general methodology of evaluation*. Canada: John Benjamins Publishing Company.
10. Melby, A. K. (1995), *The possibility of language: a discussion of the nature of language, with implications for human and machine translation*. Philadelphia: John Benjamins Publishing Company.
11. Trujillo, A. (1999), *Translation engines: Techniques for machine translation*. New York: Springer.
12. Phan, T. T. N. & Chen, H. C. (2020), Vietnamese engineering students' perceptions of the use of Google Translate tool. Tạp chí *Khoa học và công nghệ Trường Đại học Công nghiệp Thành phố Hồ Chí Minh*, 48(6), 156-165.

Application of machine translation in drafting documents

Abstract: Nowadays, along with the development of science and technology, many automatic translation systems which have received great attention from the users have been introduced. However, there are still different debates on the quality and reliability of machine translation. The research explores the application of machine translation into drafting documents between Vietnamese and English based on the analysis of the popular translation errors and the attitudes of the users toward machine translation systems. The research analysis results reveal that grammatical errors appear the most in the analyzed documents. The categories of lexis, semantics and discourse are also popular errors with approximately similar rates. Based on such analysis, the users and human translators can pay more attention to these factors during the editing process of the raw machine translation. On the other hand, the study acknowledges the positive feedbacks from the users to machine translation. The limitations as well as the common errors identified in the study may provide suitable directions for the users to apply during the process of machine translation.

Key words: translation; machine translation; automatic translation; translation errors; attitudes.