

NGHIÊN CỨU VỀ DỊCH THUẬT VĂN HỌC HÀN QUỐC DỰA TRÊN TRÍ TUỆ NHÂN TẠO VÀ KHẢ NĂNG ỨNG DỤNG TRONG GIÁO DỤC: TẬP TRUNG VÀO ĐÁNH GIÁ BẰNG THƯỚC ĐO TỐI ƯU HÓA LIÊN NGÔN NGỮ VÀ PHÂN TÍCH THEO KHUNG TIÊU CHÍ CHẤT LƯỢNG ĐA CHIỀU

Phạm Thị Ngọc^{1*}, Yun Juseok¹

*Tác giả liên hệ, Email: omonisaranghe@gmail.com. ORCID: 0009-0003-4568-7264

Ngày tòa soạn nhận được bài báo: 15/06/2026

Ngày phản biện đánh giá: 22/06/2026

Ngày bài báo được duyệt đăng: 14/09/2026

DOI: 10.59266/houjs.2026.1441

Tóm tắt: Bài viết phân tích chất lượng dịch văn học Hàn Quốc bằng trí tuệ nhân tạo (AI) và khả năng ứng dụng trong giáo dục dịch thuật. Dữ liệu gồm Dongbaekkkot của Kim You-jeong, bản dịch chuyên gia và bản dịch AI; kết quả được đánh giá bằng thước đo tối ưu hóa liên ngôn ngữ đánh giá bản dịch (Crosslingual Optimized Metric for Evaluation of Translation - COMET) trên 39 đoạn và khảo sát 30 giáo viên theo các tiêu chí tham chiếu của khung tiêu chí chất lượng đa chiều (Multidimensional Quality Metrics - MQM). Kết quả cho thấy AI truyền đạt tốt nội dung chính nhưng còn hạn chế về văn phong, cảm xúc, yếu tố văn hóa và hiệu quả văn học, vì vậy phù hợp hơn với vai trò tư liệu đối chiếu trong đào tạo dịch thuật.

Từ khóa: dịch văn học, khung tiêu chí chất lượng đa chiều, thước đo tối ưu hóa liên ngôn ngữ, trí tuệ nhân tạo

I. Đặt vấn đề

Sự phát triển của trí tuệ nhân tạo (Artificial Intelligence, viết tắt là AI) tạo sinh đang mở rộng khả năng ứng dụng AI trong dịch thuật và giáo dục ngoại ngữ. AI có thể tạo bản dịch tương đối tự nhiên và xử lý văn bản theo ngữ cảnh, qua đó hỗ trợ việc sử dụng tác phẩm văn học trong giáo dục tiếng Hàn và đào tạo dịch thuật. Tuy nhiên, dịch văn học không chỉ yêu cầu truyền đạt chính xác nội dung mà còn cần thể hiện phong cách, cảm xúc, quan hệ

nhân vật và yếu tố văn hóa. Vì vậy, chất lượng bản dịch văn học do AI tạo ra cần được xem xét cả về nội dung và các đặc điểm khó lượng hóa của văn bản văn học.

Nghiên cứu này lựa chọn truyện ngắn Dongbaekkkot của Kim You-jeong làm đối tượng phân tích, đối chiếu bản gốc tiếng Hàn, bản dịch tiếng Việt của chuyên gia và bản dịch AI. Nghiên cứu tập trung vào ba mục tiêu: xây dựng quy trình dịch văn bản văn học dài bằng AI; đánh giá chất lượng bản dịch bằng thước

¹ Trường Đại học Hà Nội, Hà Nội, Việt Nam

đo tối ưu hóa liên ngôn ngữ đánh giá bản dịch (Crosslingual Optimized Metric for Evaluation of Translation, viết tắt là COMET) (Rei và cộng sự, 2020, 2023) kết hợp với phân tích dựa trên các tiêu chí tham chiếu của khung tiêu chí chất lượng đa chiều (Multidimensional Quality Metrics, viết tắt là MQM) (Lommel và cộng sự, 2013; Mariana và cộng sự, 2015); và xem xét khả năng sử dụng bản dịch AI trong giáo dục tiếng Hàn và đào tạo dịch thuật. Khoảng trống mà nghiên cứu hướng tới là: các nghiên cứu về dịch máy Hàn-Việt hiện nay chủ yếu khảo sát văn bản thông tin và thường chỉ dùng một loại thước đo, chưa đối chiếu điểm số tự động với đánh giá của người dạy trên một tác phẩm văn học hoàn chỉnh.

II. Cơ sở lý luận và tổng quan nghiên cứu

2.1. Đánh giá dịch máy tự động bằng COMET

Trong lĩnh vực đánh giá dịch máy, các thước đo tự động được sử dụng nhằm xác định chất lượng bản dịch một cách khách quan và thuận tiện. Trước đây, những thước đo dựa trên sự trùng khớp hình thức như thước đo đánh giá song ngữ (Bilingual Evaluation Understudy, viết tắt là BLEU) được sử dụng phổ biến, nhưng các thước đo này thường khó phản ánh đầy đủ sự tương đồng về ý nghĩa và mức độ phù hợp với ngữ cảnh. Để khắc phục hạn chế đó, các thước đo dựa trên mạng nơ-ron đã được phát triển, trong đó COMET là một công cụ tiêu biểu.

COMET là mô hình đánh giá dịch máy được huấn luyện dựa trên dữ liệu đánh giá của con người. Công cụ này sử dụng văn bản nguồn, bản dịch tham chiếu

và bản dịch cần đánh giá để dự đoán chất lượng bản dịch. Rei và cộng sự (2020) cho rằng COMET có khả năng phản ánh đánh giá của con người tốt hơn nhiều thước đo truyền thống vì nó chú ý đến quan hệ ngữ nghĩa giữa các văn bản. Rei và cộng sự (2023) cũng chỉ ra rằng COMET phản ánh khá tốt mức độ truyền đạt ý nghĩa, song kết quả chủ yếu được thể hiện bằng điểm số nên các yếu tố như phong cách, cảm xúc, văn hóa và hiệu quả văn học khó được đánh giá đầy đủ bằng một chỉ số tự động.

Trong nghiên cứu này, COMET được sử dụng để phân tích định lượng chất lượng bản dịch văn học Hàn Quốc do AI tạo ra. Vì vậy, COMET ở đây được sử dụng như một công cụ hỗ trợ đánh giá và để lựa chọn những đoạn cần phân tích bằng phương pháp định tính, chứ không phải là tiêu chuẩn duy nhất để kết luận về chất lượng bản dịch.

2.2. Đánh giá chất lượng dịch dựa trên MQM

Bên cạnh đánh giá tự động bằng COMET, việc đánh giá chất lượng bản dịch cũng cần được thực hiện bằng phương pháp định tính. Điều này đặc biệt quan trọng đối với dịch văn học, vì bản dịch văn học không chỉ cần truyền đạt đúng nội dung mà còn phải thể hiện được phong cách, giọng điệu nhân vật, sắc thái cảm xúc, yếu tố văn hóa và hiệu quả văn học của nguyên tác. Một bản dịch tuy đúng về nghĩa cơ bản vẫn có thể chưa đạt yêu cầu nếu làm mất đi không khí văn học hoặc quan hệ tinh tế giữa các nhân vật.

MQM là khung đánh giá chất lượng dịch được sử dụng để phân loại lỗi dịch theo nhiều phương diện khác nhau. Khung này không chỉ đưa ra điểm số chung mà còn giúp xác định bản dịch có vấn đề ở

phương diện nào, chẳng hạn như độ chính xác của ý nghĩa, độ tự nhiên của cách diễn đạt, phong cách, thuật ngữ, yếu tố văn hóa hoặc mức độ phù hợp với mục đích dịch. Lommel và cộng sự (2013) cho rằng MQM có ưu điểm ở tính linh hoạt và khả năng phân tích lỗi chi tiết. Mariana và cộng sự (2015) cũng chỉ ra rằng MQM có thể hỗ trợ quá trình đánh giá bản dịch một cách hệ thống và đáng tin cậy hơn.

Trong nghiên cứu này, MQM được điều chỉnh cho phù hợp với mục tiêu đánh giá bản dịch văn học Hàn Quốc sang tiếng Việt do AI tạo ra. Các tiêu chí đánh giá được chia thành năm nhóm chính: độ chính xác, độ tự nhiên của tiếng Việt, phong cách, sự phù hợp về văn hóa và hiệu quả văn học. Những tiêu chí này giúp xem xét bản dịch AI có duy trì được giọng kể, lời thoại, quan hệ nhân vật, cảm xúc và tính văn học của nguyên tác hay không.

Điểm khác biệt của nghiên cứu này nằm ở việc kết hợp COMET và MQM để đánh giá bản dịch AI của văn học Hàn Quốc sang tiếng Việt trong mối liên hệ với giáo dục tiếng Hàn và đào tạo dịch thuật. Các nghiên cứu trước đây thường đánh giá dịch máy bằng chỉ số tự động hoặc phân tích lỗi dịch theo một khung riêng lẻ, trong khi nghiên cứu này đồng thời sử dụng điểm số COMET và phân tích định tính dựa trên MQM để xem xét cả mức độ truyền đạt ý nghĩa lẫn các yếu tố văn học khó lượng hóa.

III. Phương pháp nghiên cứu

3.1. Đối tượng và tư liệu nghiên cứu

Nghiên cứu này lựa chọn truyện ngắn Dongbaekkkot của Kim You-jeong làm đối tượng nghiên cứu nhằm phân tích phương pháp dịch văn học bằng AI và chất lượng bản dịch do AI tạo ra. Đây là tác phẩm có dung lượng tương đối ngắn, phù hợp để phân tích toàn văn, đồng thời có sự kết hợp cân bằng giữa lời thoại, lời kể và diễn biến sự kiện.

Tác phẩm cũng có nhiều đặc điểm đáng chú ý như cách diễn đạt mang tính khẩu ngữ, sắc thái địa phương, sự thể hiện cảm xúc và quan hệ giữa các nhân vật thông qua lời thoại. Không khí văn học, yếu tố hài hước và những biểu hiện hàm ẩn trong tác phẩm giúp người nghiên cứu xem xét AI xử lý phong cách, giọng điệu và cảm xúc của văn bản văn học như thế nào.

3.2. Quy trình dịch văn bản văn học bằng AI

Đối với văn bản văn học có dung lượng tương đối dài, việc đưa toàn bộ tác phẩm vào một lần trong môi trường đối thoại thông thường của AI có thể dẫn đến hiện tượng dịch không đầy đủ, tóm tắt nội dung hoặc bỏ sót một số đoạn. Vì vậy, nghiên cứu này sử dụng chương trình Python kết hợp với phương thức gọi giao diện lập trình ứng dụng (Application Programming Interface, viết tắt là API) để chia văn bản thành các phần nhỏ, gửi lần lượt từng phần đến AI và lưu lại kết quả dịch một cách tự động.²

² Giao diện lập trình ứng dụng (Application Programming Interface, viết tắt là API) là phương thức trong đó chương trình gửi văn bản cần xử lý đến mô hình AI và nhận, lưu kết quả phản hồi, thay vì người dùng nhập trực tiếp vào cửa sổ trò chuyện. Hiện nay, nhiều dịch vụ AI tạo sinh cung cấp API cho nhà phát triển và nhà nghiên cứu bên cạnh giao diện hội thoại dành cho người dùng phổ thông. Trong nghiên cứu này, phương thức gọi API được sử dụng để chia tác phẩm văn học dài thành nhiều phần và dịch tuần tự từng phần.

Quy trình dịch được thực hiện theo hai bước. Ở bước thứ nhất, AI được yêu cầu lập kế hoạch dịch dựa trên toàn bộ tác phẩm, bao gồm thông tin về nhân vật, cách xử lý tên riêng, xưng hô, phong cách trần thuật, bối cảnh và tiêu chí chia đoạn. Ở bước thứ hai, từng đoạn văn bản được dịch sang tiếng Việt dựa trên kế hoạch này. Khi dịch từng đoạn, các thông tin ngữ cảnh quan trọng được cung cấp lại nhằm duy trì tính nhất quán về tên nhân vật, giọng điệu, cách xưng hô và không khí chung của tác phẩm. Bản dịch tiếng Việt của chuyên gia dùng làm bản tham chiếu trong nghiên cứu này được lấy từ Lee và Nguyễn (2010).

Bản dịch tiếng Việt do AI tạo ra theo quy trình này sau đó được sử dụng làm đối tượng để phân tích định lượng bằng COMET và đánh giá định tính dựa trên MQM.

3.3. Phân tích định lượng bằng COMET

Để đánh giá định lượng chất lượng bản dịch do AI tạo ra, nghiên cứu này sử dụng công cụ COMET. Bản gốc tiếng Hàn, bản dịch tiếng Việt của chuyên gia và bản dịch tiếng Việt do AI tạo ra được chia thành các đơn vị nội dung tương ứng, rồi sắp xếp sao cho mỗi hàng dữ liệu bao gồm một đoạn văn bản gốc, một đoạn bản dịch tham chiếu và một đoạn bản dịch AI.

Trong nghiên cứu này, mô hình ‘Unbabel/wmt22-comet-da’ được sử dụng theo phương thức đánh giá có bản tham chiếu. Cụ thể, bản dịch tiếng Việt của chuyên gia được dùng làm bản dịch tham chiếu, còn bản dịch tiếng Việt do AI tạo ra là đối tượng đánh giá. Dữ liệu được lưu dưới dạng tệp giá trị phân tách bằng

dấu phẩy (Comma-Separated Values, viết tắt là CSV) và xử lý bằng chương trình Python để tính điểm chất lượng cho từng đoạn cũng như điểm trung bình chung của toàn bộ bản dịch.

Kết quả COMET được sử dụng để xác định xu hướng chất lượng của bản dịch AI ở cấp độ định lượng, đồng thời làm chỉ báo để chọn những đoạn cần tiếp tục phân tích bằng phương pháp định tính dựa trên MQM, chứ không được xem là tiêu chuẩn tuyệt đối về chất lượng bản dịch văn học.

3.4. Đánh giá định tính dựa trên MQM

Bên cạnh phân tích định lượng bằng COMET, nghiên cứu tiến hành khảo sát dựa trên bộ tiêu chí tham chiếu MQM nhằm xem xét cụ thể hơn chất lượng bản dịch AI. Ngữ liệu khảo sát gồm năm đoạn trích, mỗi đoạn tương ứng với một giai đoạn của cốt truyện (mở đầu, phát triển, cao trào, khủng hoảng, kết thúc) và được chọn theo tiêu chí bao quát cả lời kể lẫn lời thoại. Người tham gia đối chiếu nguyên tác tiếng Hàn, bản dịch chuyên gia và bản dịch AI, sau đó đánh giá bản dịch AI theo thang Likert 5 mức với 10 câu hỏi thuộc năm nhóm tiêu chí sau.

Độ chính xác

1) Ý nghĩa của nguyên tác có được truyền đạt chính xác không?

2) Có hiện tượng lược bỏ hoặc làm sai lệch thông tin không?

Độ tự nhiên trong diễn đạt

3) Cấu trúc câu của bản dịch có tự nhiên trong tiếng Việt không?

4) Việc lựa chọn từ ngữ có tự nhiên không?

Phong cách

5) Phong cách trần thuật của nguyên tác có được duy trì không?

6) Cách diễn đạt trong lời thoại có tự nhiên không?

Sự phù hợp về văn hóa

7) Các yếu tố văn hóa có được chuyển tải phù hợp không?

8) Cách xưng hô và biểu hiện quan hệ xã hội có tự nhiên không?

Hiệu quả văn học

9) Không khí và cảm xúc của tác phẩm có được truyền tải không?

10) Bản dịch có tự nhiên khi đọc như một tác phẩm văn học không?

Kết quả khảo sát theo 10 câu hỏi trên được sử dụng để bổ sung và giải thích cho kết quả phân tích định lượng bằng COMET.

IV. Kết quả và thảo luận

4.1 Kết quả phân tích định lượng bằng COMET

Kết quả phân tích bằng COMET được thực hiện trên 39 đoạn văn bản đã được đối chiếu giữa nguyên tác tiếng Hàn, bản dịch tiếng Việt của chuyên gia và bản dịch tiếng Việt do AI tạo ra. Tất cả các đoạn đều được sử dụng trong quá trình đánh giá. Kết quả cho thấy điểm trung bình toàn hệ thống đạt 0,7911 (độ lệch chuẩn = 0,0330; trung vị = 0,7929; khoảng dao động từ 0,7027 đến 0,8530); 87,2% số đoạn đạt từ 0,75 điểm trở lên và 43,6% đạt từ 0,80 điểm trở lên. Điểm số này cho thấy bản dịch AI nhìn chung phản ánh tương đối ổn định nội dung của nguyên tác tiếng Hàn và có mức độ tương đồng nhất định với bản dịch tham chiếu của chuyên gia.

Bảng 1. Kết quả đánh giá COMET trên 39 đoạn

MODEL: Unbabel/wmt22-comet-da MODE: reference-based SYSTEM_SCORE: 0.791132 SEGMENT_COUNT: 39 USED_ROWS: 39 SKIPPED_ROWS: 0 SRC_COL: src_text REF_COL: ref_text MT_COL: mt_text	
[SEGMENT SCORES]	csv_row=21, score=0.792136
csv_row=2, score=0.853016	csv_row=22, score=0.798715
csv_row=3, score=0.783071	csv_row=23, score=0.710074
csv_row=4, score=0.800577	csv_row=24, score=0.825800
csv_row=5, score=0.796507	csv_row=25, score=0.805730
csv_row=6, score=0.742187	csv_row=26, score=0.776015
csv_row=7, score=0.819652	csv_row=27, score=0.792932
csv_row=8, score=0.809433	csv_row=28, score=0.761060
csv_row=9, score=0.766153	csv_row=29, score=0.774295
csv_row=10, score=0.841209	csv_row=30, score=0.774188
csv_row=11, score=0.806345	csv_row=31, score=0.790592

csv_row=12, score=0.810095	csv_row=32, score=0.745406
csv_row=13, score=0.805798	csv_row=33, score=0.788542
csv_row=14, score=0.800783	csv_row=34, score=0.783062
csv_row=15, score=0.836485	csv_row=35, score=0.811911
csv_row=16, score=0.841697	csv_row=36, score=0.739864
csv_row=17, score=0.828123	csv_row=37, score=0.775673
csv_row=18, score=0.772710	csv_row=38, score=0.810835
csv_row=19, score=0.791506	csv_row=39, score=0.801190
csv_row=20, score=0.788032	csv_row=40, score=0.702749.

Xét theo từng đoạn, điểm COMET cao nhất đạt 0,8530 và thấp nhất đạt 0,7027, cho thấy chất lượng bản dịch AI không hoàn toàn đồng đều giữa các đoạn. AI truyền đạt khá tốt ở những đoạn nội dung rõ ràng, cấu trúc câu đơn giản, nhưng dễ bộc lộ hạn chế ở những đoạn giàu hình ảnh văn học và sắc thái cảm xúc phức tạp.

Có thể thấy rõ điều này qua đoạn có điểm COMET thấp nhất. Trong nguyên tác tiếng Hàn, cụm “꽃 밑을 살금살금 기어서” không chỉ miêu tả hành động bò mà còn gợi một chuyển động nhẹ, kín đáo trong không gian tự nhiên. Bản dịch AI chuyển thành “bò dưới những bông hoa”, có liên hệ về nghĩa nhưng chưa tự nhiên trong tiếng Việt và không làm rõ được không gian, động tác của nhân vật. Trong khi đó, bản dịch tham chiếu của

chuyên gia không dịch hoàn toàn sát từng từ mà tái cấu trúc thành “luồn qua kẽ đá”, qua đó làm nổi bật đường di chuyển và trạng thái lẩn tránh của nhân vật.

Tương tự, cụm “바위를 끼고 엉금엉금 기어서” trong nguyên tác gợi cảm giác nhân vật bám sát vào tảng đá và bò một cách chậm chạp, vụng về. Bản dịch AI dùng “bò quanh những tảng đá”, nhưng từ “quanh” chỉ diễn tả vị trí xung quanh nên chưa thể hiện rõ cảm giác nép sát và che giấu trong nguyên tác.

Nhìn chung, kết quả COMET cho thấy AI có khả năng tạo ra bản dịch văn học ở mức có thể chấp nhận được về mặt truyền đạt nội dung cơ bản. Tuy nhiên, kết quả định lượng này cần được kết hợp với phân tích định tính dựa trên MQM để đánh giá toàn diện hơn đặc điểm của bản dịch AI.

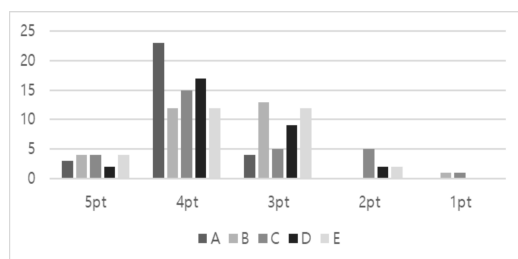
Bảng 2. Ví dụ đoạn có điểm COMET thấp nhất

Hạng mục	Nội dung
Nguyên văn	점순이가 겁을 잔뜩 집어먹고 꽃 밑을 살금살금 기어서 산 아래로 내려간 다음 나는 바위를 끼고 엉금엉금 기어서 산 위로 치빠지 않을 수 없었다.
Bản dịch của chuyên gia	Jeom Sun hoảng hốt nhẹ nhàng luồn qua kẽ đá đi xuống, còn tôi lát sau cũng trườn qua kẽ đá chuồn thẳng lên trên núi.
Bản dịch của AI	Jeom-sun sợ hãi tột độ, lén lút bò dưới những bông hoa rồi đi xuống chân núi. Sau đó, tôi cũng không thể không lồm cồm bò quanh những tảng đá rồi trèo vội lên đỉnh núi.

4.2 Kết quả đánh giá định tính dựa trên MQM

Tổng số người tham gia là 30. Về kinh nghiệm giảng dạy tiếng Hàn, 16,7% có trên 11 năm, 16,7% trên 6 năm, 33,3% trên 3 năm và 33,3% trên 1 năm. Về trình độ, 63,3% có bằng thạc sĩ, 23,3% cử nhân và 13,3% tiến sĩ trở lên. Về thời gian sinh sống tại Hàn Quốc, 30% trên 5 năm, 16,7% trên 3 năm, 23,3% trên 1 năm và 30% dưới 1 năm.

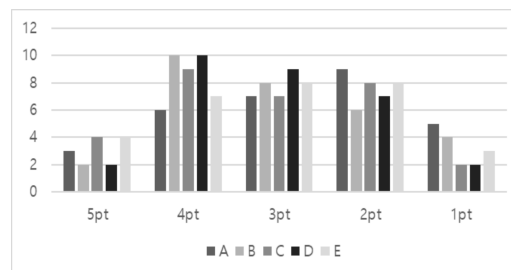
Khảo sát dựa trên MQM gồm 10 câu hỏi như đã trình bày ở mục 3.4. Người tham gia được cung cấp toàn bộ nguyên tác, bản dịch của chuyên gia và bản dịch AI, đồng thời đánh giá năm đoạn trích thuộc phần mở đầu, phát triển, cao trào, khủng hoảng và kết thúc. Hệ số Cronbach's α lần lượt là 0,921, 0,943, 0,942, 0,955 và 0,947, cho thấy độ nhất quán nội tại cao ở tất cả các phần. Trong các biểu đồ dưới đây, năm phần này được ký hiệu lần lượt là A, B, C, D và E.



Hình 1. Độ chính xác về ý nghĩa của nguyên tác

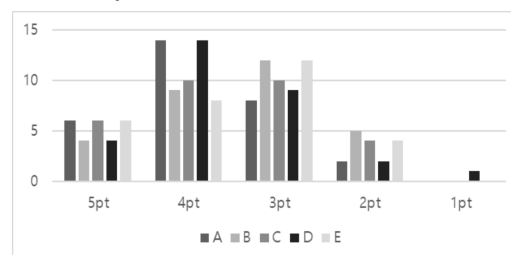
Các câu trả lời tập trung chủ yếu ở mức 4, cho thấy bản dịch AI nhìn chung truyền đạt khá chính xác ý nghĩa của nguyên tác. Phần mở đầu nhận đánh giá cao và ổn định nhất, trong khi ở phần phát

triển, cao trào và kết thúc, tỷ lệ mức 3 tăng và xuất hiện một số mức 1-2. Điều này cho thấy độ chính xác có xu hướng giảm hoặc phân tán hơn ở những đoạn có nhiều hàm ý, cảm xúc và chuyển biến quan trọng của cốt truyện.



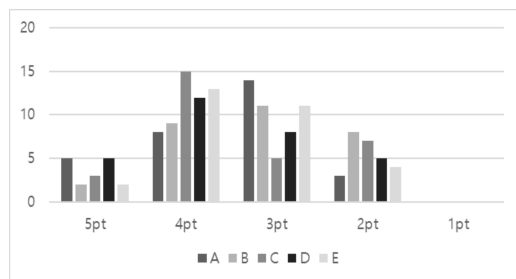
Hình 2. Tình trạng lược bỏ hoặc sai lệch thông tin

Kết quả phân tán hơn so với tiêu chí độ chính xác về ý nghĩa. Người tham gia nhìn chung cho rằng AI truyền đạt được nội dung chính, nhưng vẫn có khả năng lược bỏ hoặc làm sai lệch một số chi tiết. Đặc biệt, ở phần mở đầu, tỷ lệ mức 1-2 tương đối cao, cho thấy thông tin ban đầu về nhân vật, bối cảnh và tình huống chưa được truyền đạt hoàn toàn ổn định.



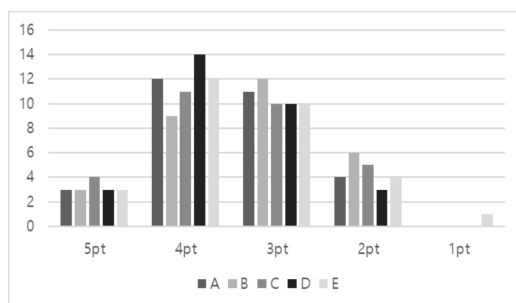
Hình 3. Mức độ phù hợp của cấu trúc câu tiếng Việt

Kết quả nhìn chung tích cực. Ở tất cả các phần, mức 4 và 5 chiếm tỷ lệ cao, trong khi mức 1 hầu như không xuất hiện. Điều này cho thấy bản dịch AI có cấu trúc câu tiếng Việt tương đối tự nhiên và ổn định.



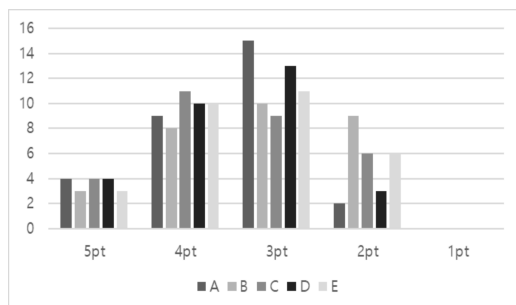
Hình 4. Tính tự nhiên trong lựa chọn từ ngữ

Kết quả chủ yếu phân bố từ mức trung bình trở lên và không có mức 1 ở tất cả các phần. Tuy nhiên, mức 3 và 4 chiếm tỷ lệ lớn, cho thấy cách lựa chọn từ ngữ nhìn chung dễ hiểu và phù hợp về nghĩa, nhưng chưa phải lúc nào cũng đạt được sự tự nhiên và tinh tế của văn bản văn học.



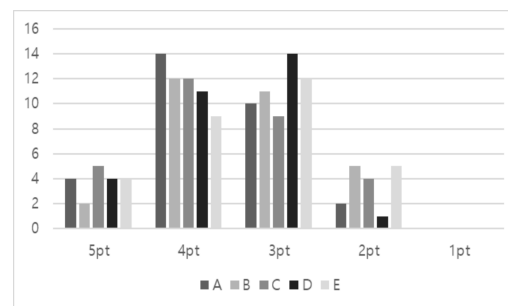
Hình 5. Khả năng duy trì phong cách trần thuật của nguyên tác

Các câu trả lời tập trung chủ yếu ở mức 3 và 4. Kết quả cho thấy AI duy trì được dòng trần thuật cơ bản của nguyên tác, nhưng còn hạn chế trong việc tái hiện đầy đủ đặc trưng văn phong và sắc thái trần thuật của tác phẩm.



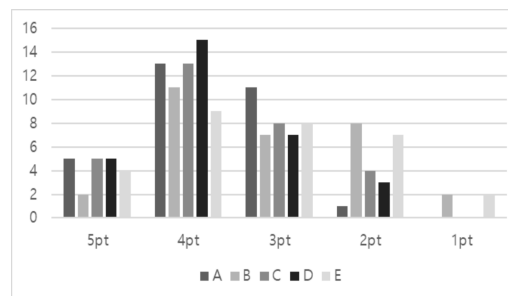
Hình 6. Tính tự nhiên của lời thoại

Kết quả chủ yếu tập trung ở mức 3 và 4. AI có thể truyền đạt nội dung hội thoại ở mức dễ hiểu, nhưng chưa tái hiện đầy đủ tính sinh động của khẩu ngữ cũng như những khác biệt tinh tế về giọng điệu giữa các nhân vật.



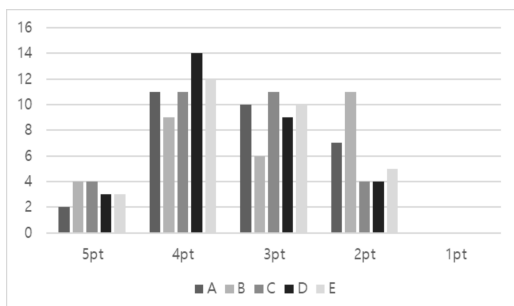
Hình 7. Mức độ phù hợp của các yếu tố văn hóa

Các câu trả lời chủ yếu ở mức 3 và 4, cho thấy các yếu tố văn hóa được truyền đạt ở mức chấp nhận được. Phần mở đầu và cao trào được đánh giá cao hơn, phản ánh sự tái hiện tương đối ổn định về bối cảnh văn hóa.



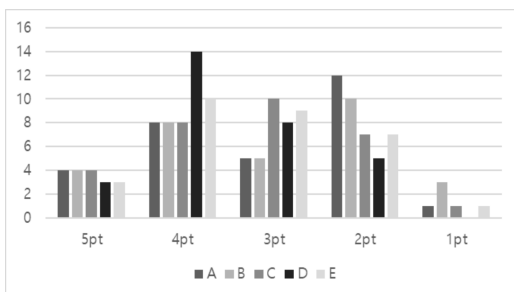
Hình 8. Tính tự nhiên của cách xưng hô và biểu hiện quan hệ xã hội

Các câu trả lời chủ yếu tập trung ở mức 4, cho thấy AI thể hiện quan hệ nhân vật tương đối rõ. Phần mở đầu, cao trào và khủng hoảng được đánh giá cao hơn, phản ánh việc tái hiện khá ổn định quan hệ và thái độ trong lời thoại.



Hình 9. Khả năng truyền tải không khí và cảm xúc của tác phẩm

Kết quả chủ yếu tập trung ở mức 3 và 4, đồng thời xuất hiện khá nhiều câu trả lời ở mức 2. Điều này cho thấy AI vẫn còn hạn chế trong việc truyền tải đầy đủ không khí và cảm xúc của tác phẩm văn học.



Hình 10. Mức độ phù hợp với tư cách là bản dịch văn học

Các câu trả lời chủ yếu phân bố từ mức 2 đến 4. Điều này cho thấy bản dịch AI không bị đánh giá là hoàn toàn không phù hợp, nhưng cũng chưa đạt mức hoàn thiện cần thiết của một bản dịch văn học. Phần khủng hoảng và kết thúc được đánh giá tương đối cao, trong khi phần mở đầu và phát triển có tỷ lệ mức 1-2 cao hơn, cho thấy hạn chế trong việc hình thành không khí, thể hiện cá tính văn phong và mật độ cảm xúc ở phần đầu tác phẩm.

Có thể minh họa bằng đoạn dẫn ở Bảng 2. Hai từ tượng hình “살금살금” và “엇금엇금” được bản dịch AI chuyển thành “lén lút bò” và “lồm cồm bò quanh”,

chỉ giữ nét nghĩa chung của hành động, trong khi bản dịch chuyên gia tái hiện được cả đường di chuyển lẫn sắc thái hài hước. Đây là kiểu khác biệt khiến các tiêu chí về phong cách (câu 5-6) và hiệu quả văn học (câu 9-10) nhận điểm thấp hơn rõ rệt so với các tiêu chí về độ chính xác.

Nhìn chung, kết quả MQM cho thấy bản dịch AI có ưu điểm trong việc truyền đạt nội dung chính và tạo ra văn bản tiếng Việt tương đối dễ hiểu, nhưng còn hạn chế ở những yếu tố cốt lõi của dịch văn học như phong cách, cảm xúc, quan hệ nhân vật, yếu tố văn hóa và hiệu quả văn học. Các ý kiến tự do cũng cho thấy người tham gia đánh giá cao tốc độ xử lý và tính tiện lợi của AI, nhưng cho rằng bản dịch AI nên được sử dụng như tài liệu đối chiếu hoặc bản dịch sơ bộ, chứ không nên xem là bản dịch mẫu hoàn chỉnh.

4.3 Khả năng ứng dụng trong giáo dục dịch thuật tiếng Hàn

Kết quả phân tích cho thấy bản dịch AI có thể truyền đạt tương đối ổn định nội dung chính của nguyên tác, nhưng vẫn còn hạn chế trong việc tái hiện phong cách văn học, sắc thái cảm xúc, quan hệ nhân vật và yếu tố văn hóa. Vì vậy, trong giáo dục dịch thuật tiếng Hàn, bản dịch AI cần được xem như một tài liệu đối chiếu giúp người học phân tích, đánh giá và chỉnh sửa bản dịch một cách phê phán, chứ không phải một bản dịch mẫu hoàn chỉnh.

Cụ thể, trong lớp học dịch thuật dành cho người học chuyên ngành tiếng Hàn, giáo viên có thể cung cấp đồng thời nguyên tác tiếng Hàn, bản dịch tiếng Việt của chuyên gia và bản dịch tiếng Việt do

AI tạo ra. Người học đối chiếu ba loại văn bản này để nhận diện sự khác biệt về nghĩa, cách lựa chọn từ ngữ, giọng điệu, lời thoại, cách xưng hô và hiệu quả văn học. Hoạt động này giúp người học hiểu rằng dịch thuật là quá trình diễn giải văn bản và điều chỉnh bản dịch phù hợp với ngữ cảnh văn hóa, chứ không chỉ là chuyển đổi ngôn ngữ.

Bản dịch AI cũng có thể được sử dụng trong hoạt động sửa lỗi và cải thiện bản dịch. Người học dựa trên các tiêu chí như độ chính xác, độ tự nhiên, phong cách, yếu tố văn hóa và hiệu quả văn học để đánh giá bản dịch AI, sau đó đề xuất phương án chỉnh sửa. Quá trình này giúp người học phát triển năng lực đánh giá bản dịch, năng lực phát hiện lỗi và năng lực biên tập văn bản.

Ngoài ra, kết quả nghiên cứu cũng cho thấy giáo dục dịch thuật trong thời đại AI cần đưa năng lực sử dụng AI vào nội dung đào tạo. Người học không chỉ cần biết sử dụng công cụ AI để tạo bản dịch, mà còn cần biết thiết kế yêu cầu dịch, cung cấp thông tin ngữ cảnh, kiểm tra kết quả, phát hiện vấn đề và chỉnh sửa bản dịch.

Như vậy, giá trị giáo dục của bản dịch AI không nằm ở việc thay thế người học hoặc người dịch, mà nằm ở khả năng tạo ra một đối tượng để phân tích và thảo luận, khuyến khích người học tham gia vào quá trình so sánh, đánh giá và chỉnh sửa.

Nghiên cứu còn một số hạn chế. Tư liệu chỉ gồm một truyện ngắn với 39 đoạn đối chiếu, năm đoạn trích khảo sát và 30 người đánh giá, nên kết quả chưa đủ để

khái quát cho dịch văn học Hàn-Việt nói chung. Phân tích dừng ở thống kê mô tả, chưa sử dụng kiểm định suy luận để so sánh giữa các phần của tác phẩm. Ngoài ra, nghiên cứu chỉ sử dụng một hệ thống AI và một mô hình COMET duy nhất.

V. Kết luận và kiến nghị

Nghiên cứu xem xét chất lượng bản dịch văn học Hàn Quốc do AI tạo ra và khả năng ứng dụng bản dịch này trong đào tạo dịch thuật, trên tư liệu truyện ngắn Dongbaekkkot của Kim You-jeong với ba văn bản đối chiếu: nguyên tác, bản dịch chuyên gia và bản dịch AI. Kết quả COMET (trung bình = 0,7911; độ lệch chuẩn = 0,0330) cùng khảo sát 30 giáo viên theo tiêu chí MQM đều cho thấy bản dịch AI truyền đạt ổn định nội dung chính, tạo được văn bản tiếng Việt rõ ràng và có cấu trúc câu tương đối tự nhiên.

Tuy nhiên, AI chưa tái hiện đầy đủ phong cách trần thuật, sắc thái cảm xúc, quan hệ nhân vật, yếu tố văn hóa và hiệu quả văn học của nguyên tác, đặc biệt ở những đoạn giàu từ tượng hình và hàm ý. Vì vậy, bản dịch AI khó được xem là bản dịch văn học hoàn chỉnh nếu thiếu khâu kiểm tra và chỉnh sửa của con người; giá trị của nó nằm ở chỗ tạo ra một đối tượng để người học phân tích, đánh giá và biên tập.

Từ kết quả trên, nghiên cứu đưa ra ba kiến nghị. Thứ nhất, với giảng viên dịch thuật tiếng Hàn: sử dụng bản dịch AI như tư liệu đối chiếu song song với nguyên tác và bản dịch chuyên gia thay vì như bản dịch mẫu, và tổ chức hoạt động phát hiện lỗi và đề xuất phương án chỉnh

sửa theo năm nhóm tiêu chí MQM đã nêu. Thứ hai, với chương trình đào tạo: bổ sung nội dung về năng lực sử dụng AI, gồm kỹ năng thiết kế yêu cầu dịch, cung cấp thông tin ngữ cảnh và thẩm định kết quả. Thứ ba, với các nghiên cứu tiếp theo: mở rộng sang nhiều tác phẩm và thể loại, tăng số người đánh giá và áp dụng kiểm định thống kê suy luận nhằm kiểm chứng những khác biệt được ghi nhận trong nghiên cứu này.

Tài liệu tham khảo:

- Lee, J. S., và Nguyễn, T. K. D. (2010). Tuyển tập truyện ngắn Kim You-jeong. Nhà xuất bản Hội Nhà văn.
- Lommel, A. R., Burchardt, A., & Uszkoreit, H. (2013). Multidimensional quality metrics: A flexible system for assessing translation quality. In *Proceedings of Translating and the Computer 35*. Aslib.
- Mariana, V., Cox, T. L., & Melby, A. (2015). The Multidimensional Quality Metrics (MQM) framework: A new framework for translation quality assessment. *The Journal of Specialised Translation*, 23, 137-161.
- Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2685-2702). Association for Computational Linguistics.
- Rei, R., Guerreiro, N. M., Treviso, M., Coheur, L., Lavie, A., & Martins, A. F. T. (2023). The inside story: Towards better understanding of machine translation neural evaluation metrics. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (pp. 1089-1105). Association for Computational Linguistics.

A STUDY ON KOREAN LITERARY TRANSLATION BASED ON AI AND ITS EDUCATIONAL APPLICABILITY: FOCUSING ON CROSSLINGUAL OPTIMIZED METRIC FOR EVALUATION OF TRANSLATION AND ANALYSIS USING MULTIDIMENSIONAL QUALITY METRICS- INFORMED CRITERIA

Pham Thi Ngoc¹, Yun Juseok¹

Abstract: *This study examines the quality of Korean literary translation generated by artificial intelligence (AI) and its applicability to translation education. The data consist of Kim You-jeong's Dongbaekkkot, an expert Vietnamese translation, and an AI-generated translation. Translation quality was evaluated using the Crosslingual Optimized Metric for Evaluation of Translation (COMET) across 39 segments and a survey of 30 Korean language teachers based on criteria informed by the Multidimensional Quality Metrics (MQM) framework. The results show that AI conveys the main content effectively but remains limited in style, emotion, cultural elements, and literary effect. Therefore, AI translation is better suited as comparative material for translation analysis and revision.*

Keywords: *artificial intelligence, crosslingual optimized metric for evaluation of translation, literary translation, multidimensional quality metrics*

¹ Hanoi University, Hanoi, Vietnam