

NGOẠI NGỮ VỚI BÀN NGŨ

MỘT SỐ VẤN ĐỀ VỀ DỊCH MÁY

NGUYỄN NGỌC HẢI*

TÓM TẮT: Cho dù đã được hiện thực hoá vào giữa thế kỉ 20, nhưng mãi đến những thập kỉ cuối cuối thế kỉ 20 và hai thập kỉ đầu thế kỉ 21 dịch máy (machine translation) mới thực sự phát triển mạnh mẽ và trở nên phổ biến không những trong giới dịch thuật chuyên nghiệp, mà còn phổ biến trong những người dịch không chuyên nghiệp, các nhà nghiên cứu, sinh viên và cả những người có tiếp xúc với người nước ngoài hoặc ngoại ngữ. Trước tình hình đó, người dạy ngoại ngữ không những phải thay đổi quan điểm về dịch máy mà còn phải trau dồi kiến thức về dịch máy, sẵn sàng dạy dịch máy nhằm đào tạo sinh viên chuyên ngữ khi tốt nghiệp vừa giỏi về kĩ năng tiếng, về kĩ năng dịch truyền thống lại vừa giỏi về sử dụng công nghệ đặc biệt là thành thạo dịch máy trong môi trường công việc đầy cạnh tranh trong tương lai. Với mục đích trên, bài viết xin giới thiệu về lịch sử hình thành, phát triển của dịch máy và phương pháp dịch máy cũng như phân tích những điểm mạnh và điểm yếu của dịch máy và một số biện pháp khắc phục.

TỪ KHÓA: dịch máy; quá trình phát triển dịch máy; phương pháp dịch máy; sử dụng dịch máy.

NHẬN BÀI: 16/04/2024.

BIÊN TẬP-CHỈNH SỬA-DUYỆT ĐĂNG: 08/05/2024

1. Đặt vấn đề

Ý tưởng về dịch máy (Machine Translation - MT) được hình thành từ thế kỉ thứ 17, nhưng mãi đến những năm giữa và cuối thế kỉ thứ 20 máy dịch và dịch máy mới trở thành hiện thực. Bất chấp những thăng trầm, dịch máy thực sự phát triển mạnh mẽ, ngày càng phổ biến và đạt nhiều tiến bộ đáng kể vào những năm cuối thế kỉ 20 và hai thập kỉ đầu của thế kỉ 21, nhờ vào sự phát triển như vũ bão của công nghệ điện tử, công nghệ bán dẫn, công nghệ thông tin, Internet và các ngành khoa học khác cũng như sự bùng nổ về nhu cầu dịch thuật do đòi hỏi cấp thiết của toàn cầu hoá về kinh tế và các lĩnh vực khác của xã hội. Việc sử dụng phần mềm dịch thuật đã phổ biến rộng rãi trên phạm vi toàn cầu không những trong giới dịch thuật chuyên nghiệp, sinh viên đại học và đặc biệt là sinh viên chuyên ngữ mà còn phổ biến tới tận những người dân bình thường khi cần tiếp xúc với người nước ngoài, với ngoại ngữ miễn là họ có máy tính để bàn, laptop, tablet hoặc điện thoại thông minh.

Điều đó đặt ra cho những người đào tạo biên-phiên dịch cần phải có sự thay đổi cách nhìn nhận về dịch máy (machine translation-MT), phần mềm hỗ trợ dịch thuật (computer-assisted/aided translation-CAT) và đã đến lúc chúng ta phải dạy cho sinh viên về dịch máy, về những điểm mạnh, điểm yếu của dịch máy cũng như tác hại của sử dụng dịch máy không đúng phương pháp, đúng lúc. Đặc biệt là đối với giáo viên chuyên ngữ: cần đề cập và dạy về dịch máy, những môn học cần đến dịch máy, v.v. để sinh viên chuyên ngữ vừa có kiến thức, khả năng ngôn ngữ vừa phát triển được các kĩ năng ngôn ngữ, kĩ năng dịch (human translation skills) và thành thạo tích hợp dịch máy (MT) và các phần mềm hỗ trợ dịch thuật (CAT), các thiết bị điện tử với đích nhằm tới là họ sẽ trở thành những người biên-phiên dịch 4.0. (Human translator 4.0- Jakup Absolon- 2019) trong một môi trường công việc tương lai đầy cạnh tranh. Đây cũng chính là nội dung và mục đích của bài viết này.

2. Nội dung

2.1. Tóm lược về sự hình thành và phát triển dịch máy

Theo Isabel Lacruz, Riita Jaaskelainen (2018, tr.164), với nỗ lực khôi phục một ngôn ngữ toàn trí (a pansophic language) - một ngôn ngữ toàn diện, bao trùm tất cả các ngôn ngữ khác, có khả năng diễn đạt mọi kiến thức chung và có thể hiểu được tất cả các ngôn ngữ khác, các học giả thế kỉ thứ 17 đã có ý tưởng về máy dịch (Translation machine-TM) và dịch máy (MT). Tuy nhiên, đến tận những năm giữa và cuối thế kỉ 20, ý tưởng này mới trở thành hiện thực với sự xuất hiện các thiết bị dịch cơ

*Trường Đại học Phenikaa; Email: ngochai213d@gmail.com

học (mechanical translating devices) vào những năm đầu của thập kỉ 1930 tại Nga và Pháp.

Mặc dù vậy, các nhà nghiên cứu về dịch máy đều lấy mốc khi Warren Weaver trao đổi thư thông báo (Weaver's memo) vào năm 1949 nhằm thúc đẩy nghiên cứu về dịch máy ở Mỹ và Tây Âu và gợi ý rằng, máy tính điện tử hiện đại có thể được sử dụng trong dịch thuật và rằng có thể vẫn còn nhiều yếu tố tiêu cực trong dịch máy. Thực tế là, chính thư thông báo của Warren Weaver (Weaver's memo) đã khởi đầu cho nghiên cứu và hiện thực hoá dịch máy [Lane Shwarts, 2018, tr.165-166].

Hơn một thập kỉ sau đó việc nghiên cứu về dịch máy phát triển rất nhanh ở Mỹ, Canada và châu Âu (kể cả Liên Xô cũ) đánh dấu bởi hai hội nghị quốc tế về dịch máy năm 1952 và 1956, tại Massachusetts Institute of Technology (MIT) với một khoản đầu tư của chính phủ Mỹ lên đến 20 triệu đô la (trị giá vào thời điểm đó) vào nghiên cứu dịch máy và các lĩnh vực có liên quan. Đặc biệt, vào tháng 1/1954, tại New York, các nhà nghiên cứu đến từ Công ty IBM và Trường Đại học Georgetown (Mỹ) tiến hành trình diễn máy dịch đầu tiên từ tiếng Nga sang tiếng Anh [Mona Baker, 2021, tr.40].

Tuy nhiên, vào năm 1964, tại Mỹ, Ủy ban Tư vấn Xử lí Ngôn ngữ tự động (Automatic Language Processing Advisory Committee - ALPAC) được thành lập để đánh giá về nghiên cứu dịch máy và vào năm 1966 và đã đưa ra một báo cáo (ALPAC Report) kết luận rằng, tốc độ dịch máy chậm hơn, chất lượng sản phẩm dịch máy thiếu chính xác hơn và giá thành lại cao hơn gấp đôi so các sản phẩm do con người dịch; đặc biệt là dịch máy trong thời gian trước mắt không có triển vọng gì [Mona Baker, 2021:140]. Mặc dù báo cáo này của ALPAC được cho là đã bị chỉ trích rất nhiều vì đã dựa trên các dữ liệu lạc hậu và thiếu nhìn xa trông rộng, nhưng nó giáng một đòn mạnh vào dịch máy. Theo đó, chính phủ Mỹ ngừng các khoản đầu tư và dịch máy trải qua một giai đoạn trầm lắng. Mặc dù vậy, ngay tại Mỹ, một số nhóm vẫn tiếp tục nghiên cứu về dịch máy và nghiên cứu dịch máy vẫn tiếp tục phát triển tại Canada, Tây Âu, Liên Xô (cũ), đặc biệt là tại Nhật Bản. Chính vì vậy, nghiên cứu dịch máy vẫn đạt được một số kết quả đáng kể. Cụ thể là, các máy dịch Systran bắt đầu được sử dụng để dịch từ tiếng Nga sang tiếng Anh cho không quân Hoa Kỳ vào những năm 1970; các máy dịch Meteo dịch các bản dự báo thời tiết vào năm 1976 tại Canada; cũng vào năm 1976, Ủy ban Liên hiệp châu Âu cho lắp đặt các máy dịch Systran để dịch từ tiếng Anh sang tiếng Pháp [Arturo Trujillo -1999, tr.5].

Những năm đầu của thập kỉ 1980 chứng kiến sự bùng nổ trở lại của dịch máy và nghiên cứu về dịch máy. Chẳng hạn, năm 1982, Cộng đồng châu Âu khởi động Dự án Eurotra dựa trên các nghiên cứu đã được Đại học Grenoble và Đại học Saarbrücken thực hiện từ những năm 1960 và 1970. Cũng trong năm 1982, Nhật Bản bắt đầu dự án Mu và năm 1983, Mỹ cũng bắt đầu nghiên cứu các máy dịch dựa trên kiến thức (knowledge - based MT); đồng thời bắt đầu có một số máy dịch, phần mềm dịch thuật đã được bán ra thị trường vào thời gian này.

Vào cuối những năm 1980 và những năm đầu 1990, rất nhiều công ty, đặc biệt là các công ty Nhật Bản bắt đầu bán ra thị trường các phần mềm dịch thuật dùng trên máy tính. Lúc này, phần mềm dịch thuật và công cụ hỗ trợ người dịch (CAT) đã được sử dụng trên máy tính cá nhân và các thiết bị điện tử khác. Trong giai đoạn này, người ta cũng chứng kiến sự xuất hiện của các phần mềm phiên dịch, các phần mềm dịch thuật và phương pháp thống kê trong dịch máy (Statistical machine translation) [Arturo Trujillo, 1999, tr.6].

Những năm cuối thập kỉ 1990 và hai thập kỉ đầu của thế kỉ 21 chứng kiến sự bùng nổ mạnh mẽ trong sử dụng phần mềm dịch thuật do nhu cầu dịch thuật ngày càng cao, giao lưu quốc tế và toàn cầu hoá về kinh tế và các lĩnh vực khác ngày càng quyết liệt hơn; thiết bị điện tử có thể sử dụng phần mềm dịch thuật ngày càng nhỏ hơn về kích cỡ. Nhờ đó, thiết bị công nghệ dịch máy tiện lợi hơn, thông minh hơn và cài đặt được nhiều ứng dụng hơn. Ngoài ra, một số yếu tố khác cũng thúc đẩy sự bùng nổ sử dụng dịch máy, chẳng hạn như: sự phát triển và phổ biến của mạng Internet và các phần mềm dịch thuật trực tuyến, một số phần mềm dịch thuật trực tuyến như Google Translate và Microsoft Translator v.v. hoàn toàn miễn phí, còn một số phần mềm có trả tiền thì lại rất linh hoạt, có

thể trả tiền tính theo lượt sử dụng, theo thời gian sử dụng ngắn, theo thời gian theo tuần, tháng hoặc năm sử dụng. Kết quả là việc tiếp cận và sử dụng các phần mềm dịch thuật trở nên cực kì dễ dàng cho mọi đối tượng từ những người dịch chuyên nghiệp, sinh viên chuyên ngữ, sinh viên đại học, các kĩ sư và kể cả những người không có kiến thức ngôn ngữ học lẫn kĩ thuật máy tính vẫn có thể sử dụng phần mềm dịch.

Như vậy, sự phát triển nhanh chóng các ngành khoa học đặc biệt là Internet cũng như với nỗ lực nghiên cứu của các nhà khoa học và sự đầu tư của các chính phủ, các tổ chức và cá nhân đã giúp cho dịch máy, tuy có lúc trầm lắng, phát triển không ngừng và ngày càng phổ biến trên thế giới.

2.2. Tóm lược quá trình phát triển của Phương pháp dịch máy

Theo W. John Hutchins và Harold L. Somers (1992), *dịch máy* (machine translation) là thuật ngữ được sử dụng một cách truyền thống và tiêu chuẩn đề cập đến máy dịch được máy tính hoá về các chương trình máy tính có khả năng dịch nhiều thể loại văn bản từ một ngôn ngữ tự nhiên này sang một ngôn ngữ tự nhiên khác, có thể có hoặc không có sự trợ giúp của người dịch [tr.26]. Như vậy, thuật ngữ *dịch máy* và thuật ngữ *phần mềm dịch thuật* có thể dùng thay thế cho nhau. (Trong bài viết này, chúng tôi dùng thuật ngữ *dịch máy* bao hàm cho cả hai thuật ngữ này)...

2.2.1. Dịch máy dựa trên các nguyên tắc (Rule-based machine translation)

Có hai mô hình máy dịch mà con người đầu tiên sáng chế ra: 1/Thiết bị dịch của Georges Artsrouni, Pháp, được cấp bản quyền vào năm 1933, hoạt động dựa trên hệ thống tra cứu và truy cập cơ học với các tấm băng giấy; 2/Thiết bị dịch thứ hai của Troyanskii, người Nga, đã hoạt động theo một quy trình toàn diện theo ba giai đoạn: Giai đoạn 1: phân tích văn bản nguồn qua việc tìm ra từ căn của các từ, xác định chức năng cú pháp của chúng; giai đoạn 2: chuyển dịch các từ căn và chức năng ngữ pháp từ ngôn ngữ nguồn sang ngôn ngữ đích; và giai đoạn 3: tổng hợp và tạo ra văn bản ngôn ngữ đích. Ba giai đoạn như vậy thật sự tương tự như các giai đoạn của các máy dịch sau này [Alan Melby, 1995, tr.17].

2.2.2. Phương pháp trực tiếp (Direct approach)

Ban đầu, các máy dịch còn được gọi là *hệ thống dịch máy trực tiếp* (Direct transfer system) hoạt động theo phương pháp còn rất giản đơn là tra cứu và thay thế bằng từ vựng của ngôn ngữ đích giống như phương pháp dịch từ-đối-từ [Lynne Bowker, Jairo Buitrago Ciro, 2019, tr.54]. Ví dụ: cụm từ *chiếc ô tô màu đỏ* được máy dịch là *the car red* (mà không phải là *the red car*). Từ quan điểm ngôn ngữ học, điều bị mất đi khi dịch theo phương pháp này là máy dịch trực tiếp bỏ qua khâu phân tích cấu trúc bên trong của văn bản nguồn.

2.2.3. Phương pháp gián tiếp (Indirect approach)

Phương pháp này có hai biến thể chính là phương pháp chuyển dịch (Transfer method) và ngôn ngữ trung gian (Interligua) [John Hutchins, Harold Somers, 1992, tr.72].

Phương pháp gián tiếp (Indirect approach) gồm ba khâu là phân tích, chuyển dịch và tổng hợp. Nhiệm vụ của khâu đầu tiên là phân tích văn bản nguồn, phân tích các câu nhằm giải thích một cách trừu tượng về văn bản đó, phân tích để xác định rõ từ loại cũng như mối quan hệ của các từ trong câu (danh từ, động từ, tính từ, v.v.). Qua đó, máy dịch sẽ tạo ra bài giải thích trừu tượng (abstract representation). Từ đó, máy dịch sẽ tạo ra văn bản dịch từ bài giải thích trừu tượng trên, mà không đơn thuần chuyển ngữ từ dịch chuỗi các từ không theo cấu trúc của ngôn ngữ nguồn sang ngôn ngữ đích. Đây chính là phương pháp dịch từ văn bản nguồn sang văn bản đích qua ngôn ngữ trung gian (Interligua) [John Hutchins, Harold Somers, 1992, tr.73].

Dịch máy hoạt động dựa trên các nguyên tắc (rule-based machine translation; RBMT) với ba khâu phân tích, chuyển dịch và phân tích là nguyên lí hoạt động chủ yếu của máy dịch trong khoảng 50 năm kể từ sau chiến tranh thế giới lần thứ hai cho đến thập kỉ cuối của thế kỉ 20 [Lynne Bowker, Jairo Buitrago Ciro, 2019, tr.56]. Tuy nhiên, một trong những hạn chế lớn nhất của phương pháp hoạt động dựa trên các nguyên tắc này là phải tạo ra một chương trình phân tích cho từng ngôn ngữ cũng như chương trình tổng hợp cho từng cặp ngôn ngữ. Kết quả là phải viết rất nhiều chương trình phân

tích và tổng hợp cho các ngôn ngữ khác nhau. Theo Lynne Bowker và Jairo Buitrago Ciro (2019, nếu như các nước Liên minh châu Âu có 24 ngôn ngữ chính thức, thì tổ chức này phải phát triển 552 chương trình chuyên dịch giữa các ngôn ngữ, mà riêng mỗi chương trình lại có hơn 10 nghìn các nguyên tắc của các ngôn ngữ chính thức của các quốc gia trong tổ chức. Đây thực sự là một công việc khổng lồ nếu chúng ta suy ra là phải phát triển khổng lồ các chương trình phân tích, chuyên dịch và tổng hợp cho các ngôn ngữ khác trên thế giới.

2.2.4. Phương pháp dựa trên cơ sở khối liệu (Corpus-based approach)

Vào cuối những năm 1990 và đầu những năm của thế kỉ 21, thế giới chứng kiến sự xuất hiện mô hình dịch máy mới: *Phương pháp trên cơ sở khối liệu* (Corpus-based approach) hoặc dựa trên dữ liệu (Data-driven approach) đã mở ra kỉ nguyên mới cho dịch máy và nghiên cứu dịch máy. Phương pháp này không dựa trên các nguyên tắc ngôn ngữ học mà dựa trên cơ sở dữ liệu lớn các ví dụ trong các văn bản song ngữ cùng đề tài đã được người dịch chuyên nghiệp dịch trước đó. Các cặp văn bản được căn chỉnh, so sánh sao cho từng câu hoặc đoạn câu trong văn bản ngôn ngữ nguồn kết nối với câu hoặc đoạn câu tương ứng trong văn bản ngôn ngữ đích. Máy dịch có thể tham khảo nguồn tài nguyên này để xác định từ và đoạn câu cụ thể nào đã được dịch từ trước và sau đó sử dụng chúng để đưa ra một bản dịch cho văn bản mới phải dịch.

Phương pháp dựa trên cơ sở khối liệu gồm các văn bản song ngữ cùng đề tài chia thành hai cách sau:

Cách 1: Dịch dựa vào các ví dụ (Example-based machine translation). Theo đó, khi phải dịch một câu mới, máy tính sử dụng thao tác tham khảo các bài viết song ngữ cùng đề tài để tìm ra một câu, hoặc ít nhất là một đoạn câu giống y hệt, phù hợp một cách hoàn hảo (perfect match), phù hợp hoàn toàn (full match) nhưng chi khác nhau về số liệu, ngày tháng và địa danh hoặc gần phù hợp (fuzzy match) rồi đưa ra câu dịch.

Cách 2. Dịch máy thống kê (Statistical machine translation). Cách này tương tự như cách dựa trên các ví dụ từ các bài viết cùng đề tài, nhưng nó dựa trên mô hình thống kê nhằm nghiên cứu các câu, cụm từ, các từ (chủ yếu là cụm từ) và quan hệ của chúng giữa văn bản nguồn và văn bản đích. Đồng thời máy dịch sẽ tính toán xác suất từ số lượng lớn các bài viết bằng cả hai cặp ngôn ngữ có cùng đề tài sau đó máy đưa ra bản dịch thông qua việc tối đa hoá khả năng xảy ra của các câu trong ngôn ngữ đích từ số lượng câu nhất định của văn bản nguồn [Lynne Bowker et al. 2019, tr.59].

Tuy ý tưởng thì thật hấp dẫn, nhưng cả hai phương pháp này đều có những khó khăn riêng. Cả hai phương pháp này rất có hiệu quả nếu các văn bản cùng chủ đề để máy nghiên cứu phải rất giống nhau. Ngoài ra, để có bản dịch tốt, máy dịch theo hai phương pháp này cần một lượng lớn các bài viết cùng chủ đề chứa hàng triệu câu thí dụ hoặc từ cùng đề tài. Mặc dù, ngày nay Internet và các chương trình dịch miễn phí hoặc trả tiền trên Internet cung cấp rất nhiều các bài viết song ngữ cùng đề tài, nhưng số lượng các bài viết này phải là rất lớn, đặc biệt là các bài viết như vậy bằng các thứ tiếng ít được sử dụng thì lại rất hiếm [Lynne Bowker, Jairo Buitrago Ciro, 2019, tr.59].

2.2.5. Dịch máy nơron (Neural Machine Translation)

Vào những năm 2010, các nhà nghiên cứu và phát triển dịch máy quan tâm rất nhiều đến phương pháp dịch máy khác, gọi là Dịch máy nơron (Neural Machine Translation): dịch máy hoạt động mô phỏng các bước làm việc của nơron thần kinh và đây là ứng dụng cụ thể của Nghiên cứu sâu (Deep Learning) - một nhánh nhỏ của Trí tuệ nhân tạo (Artificial Intelligence-AI) trong lĩnh vực xử lí ngôn ngữ tự nhiên. Phương pháp dịch máy này có mục đích phát triển hệ thống máy tính có khả năng thực hiện các nhiệm vụ đòi hỏi trí thông minh của con người, chẳng hạn, tìm hiểu ngôn ngữ tự nhiên, nghiên cứu các ví dụ, suy luận, giải quyết vấn đề, dựa vào cơ sở dữ liệu mà đưa ra quyết định dịch từ ngôn ngữ nguồn sang ngôn ngữ đích. Dịch máy nơron cho thấy đây là tiến bộ quan trọng trong việc cải thiện chất lượng dịch máy so với các phương pháp dịch máy trước đây. Cụ thể, nếu các cặp văn bản song ngữ cùng một đề tài được đưa vào mạng làm việc nơron của máy dịch, các thuật toán nghiên cứu sâu (deep-learning algorithms) của mạng sẽ tự nghiên cứu các văn bản này (kể cả văn bản

nguồn và văn bản đích), tự tìm ra các nguyên tắc, các mẫu câu, các cụm từ, các từ và các ngữ cảnh sử dụng chúng trong các câu, đoạn văn và cả văn bản đã được đưa vào mạng. Khi được yêu cầu dịch một văn bản, mạng làm việc nơron của chương trình dịch sẽ truy cập cơ sở dữ liệu của máy và đưa ra gợi ý các đoạn câu, các câu của văn bản đích dựa trên ngữ liệu mà mạng làm việc nơron đã nghiên cứu từ trước.

Có thể nói, dịch máy nơron có kết quả tốt nhất ở các lĩnh vực nhất định khi mạng làm việc nơron của máy dịch có nhiều văn bản để nghiên cứu từ trước. Trong khi đó, dịch máy theo phương pháp này, tương tự như dịch máy theo phương pháp thống kê lại gặp rất nhiều khó khăn trong những lĩnh vực mà không có hoặc không có nhiều các bài viết song ngữ cùng đề tài, cũng như các từ hiếm gặp hoặc các câu dài [Lynne Bowker, Jairo Buitrago Ciro, 2019, tr.60].

Như vậy, dịch máy theo phương pháp dựa trên các ví dụ (Example-based approach), máy dịch sẽ chọn các ví dụ từ các văn bản song ngữ cùng đề tài bằng cách so sánh các ví dụ; dịch máy theo phương pháp thống kê (Statistical machine translation) là lựa chọn các ví dụ là các câu, các cụm từ và các từ của các văn bản song ngữ cùng đề tài theo các thuật toán thống kê. Còn dịch máy nơron (Neural machine translation) lựa chọn các ví dụ, các câu, đoạn câu, cụm từ và từ phù hợp với các ngữ cảnh sử dụng chúng trong các văn bản mà máy đã nghiên cứu từ trước. Mỗi một phương pháp mới xuất hiện đã khắc phục một phần hạn chế của phương pháp trước. Chính vì thế, một số nhà phát triển dịch máy sau này đã kết hợp dịch máy thống kê và dịch máy nơron với nhau. Chẳng hạn: năm 2006 Google Translate hoạt động theo phương pháp thống kê, đến năm 2016 Google Translate lại lai hai phương pháp thống kê và phương pháp dịch máy nơron.

2.3. Hạn chế của dịch máy và một số cách khắc phục

Có thể thấy, dịch máy nơron (Neural Machine Translation) cho thấy sự phát triển đầy hứa hẹn trong lĩnh vực dịch máy: vừa cho sản phẩm dịch nhanh, năng suất cao mà lại có khả năng học qua các ví dụ, đưa ra quyết định qua bản dịch, vừa có khả năng chỉnh sửa để lần dịch sau chất lượng tốt hơn. Tuy nhiên, cũng như dịch máy dựa trên các phương pháp khác, dịch máy nơron có những điểm mạnh và những điểm yếu. Chẳng hạn, máy dịch nơron vẫn phải vật lộn với ngôn ngữ tự nhiên vốn dĩ rất mập mờ và không rõ ràng; đặc biệt dịch máy nơron gặp khó khăn khi văn bản ngôn ngữ nguồn có nhiều câu phức có cấu trúc phức tạp và có nhiều từ mới mà trong cơ sở dữ liệu của máy không có [Thierry Poibeau, 2017, tr.121-122]. Chẳng hạn:

Các từ ngữ gây khó khăn cho quá trình dịch máy là: từ đồng âm khác nghĩa (homophone), từ cùng cách viết khác nghĩa (homonym), từ đa nghĩa (polysemy), từ khác nhau về từ loại. ví dụ: *bare* (trần trụi)- *bear* (con gấu); *plain* (đơn giản, đồng bằng, v.v...) - *plane* (máy bay); *bank* (ngân hàng; bờ sông); *polish* (có thể là động từ, có thể là tính từ hoặc danh từ với các nghĩa khác nhau hoàn toàn)

Ở cấp độ cấu trúc cũng có rất nhiều điểm không rõ ràng và có thể giải thích theo nhiều cách khác nhau gây khó khăn cho dịch máy. Ví dụ:

Lucia told John that she had delivered a baby over the phone (Lucia gọi điện thoại cho John và thông báo rằng cô ta đã sinh con) nhưng máy có thể dịch là: Lucia nói với John rằng cô ta đã sinh con bên trên điện thoại.

John watched the bird with a pair of binoculars, có thể hiểu theo hai cách là: John quan sát con chim bằng ống nhòm hoặc John quan sát con chim có ống nhòm.

There is a pile of inflammable trash next to your car. You should get rid of it: trong câu này không rõ đại từ *it* hồi chi (anaphoric reference) từ nào: *a pile of inflammable trash* hay là *your car*.

Cách sắp xếp các bộ ngữ đôi khi cũng tạo ra các cụm từ gây khó khăn cho dịch máy, ví dụ: *a liquid oxygen tank* (bê chứa oxy lỏng); máy dịch có thể dịch là “bê chứa lỏng chứa ôxy”.

Ngoài ra, những khó khăn mà dịch máy gặp phải là: đầu đề các bài báo bằng tiếng Anh, tục ngữ, thành ngữ, nghĩa giao tiếp, tính đa nghĩa của từ, nghĩa gắn với ngữ cảnh, ngữ điệu trong khẩu ngữ, v.v. [Lynne Bowker, 2019, tr.63]. Vì thế, ngay trong thư thông báo của Warren Weaver (Warren Weaver's memo) cũng chỉ rõ là máy tính điện tử hiện đại có thể được sử dụng trong dịch thuật và

rằng có thể vẫn còn nhiều yếu tố tiêu cực trong dịch máy [Lane Shwarts, 2018, tr.165-166] và ông cũng định hướng cho các nhà nghiên cứu dịch máy khắc phục những khó khăn trên.

Có thể nói, các phương pháp dịch máy sau đều cố gắng khắc phục những hạn chế của phương pháp dịch trước, đã có hiệu quả thiết thực, nhưng vẫn chưa đạt chất lượng của người dịch. Theo Thierry Poibeau, máy tính có thể đánh bại nhà vô địch cờ tướng thế giới vào năm 1997 và nhà vô địch trò chơi Go vào năm 2016, nhưng cho đến tận bây giờ chưa có máy tính nào có thể dịch ngôn ngữ tự nhiên này sang ngôn ngữ tự nhiên khác như người dịch [2017, tr.124].

Chính vì thế, các nhà nghiên cứu và phát triển dịch máy đều cho rằng, để nâng cao chất lượng bản dịch, dịch máy cần có sự tương tác giữa người dịch và máy dịch. Tương tác giữa người dịch và dịch máy gồm có hai hình thức: Người dịch tương tác trước với dịch máy “chỉnh sửa trước” (pre-editing) và người dịch tương tác sau với dịch máy “chỉnh sửa” (post-editing) có thể giúp cho chúng ta phát huy được điểm mạnh của dịch máy (nhanh và năng suất cao hơn) đồng thời nâng cao chất lượng bản dịch máy.

Chỉnh sửa trước (pre-editing: người dịch đọc hoặc máy dịch tự quét và kiểm tra văn bản nguồn nhằm phát hiện ra những khó khăn mà dịch máy có thể gặp phải (như đề cập ở phần trên) và cố gắng chỉnh sửa chúng trước khi dịch máy. Cụ thể hơn, chỉnh sửa trước là trình bày lại văn bản nguồn bằng cách sử dụng ngôn ngữ có điều chỉnh - a controlled language [W.J. Hutchins và H.L. Somers, 1992, tr.171]. Đối với các thể hệ máy dịch sau này như máy dịch thống kê (Statistical machine translation), máy dịch neuron (Neural machine translation) có thể bỏ qua khâu chỉnh sửa trước vì chúng hoạt động dựa trên các ví dụ từ các bài viết song ngữ cùng đề tài.

Chỉnh sửa sau (post-editing): chỉnh sửa sản phẩm của dịch máy để đạt được sản phẩm dịch cuối cùng có chất lượng chấp nhận được. Chỉnh sửa sau bao gồm chau chuốt bản dịch máy thô, sửa các lỗi, hiệu đính lại toàn bộ hoặc thậm chí dịch lại toàn bộ hoặc một phần bản dịch.

Mức độ chỉnh sửa còn tùy thuộc vào mục đích sử dụng sản phẩm dịch máy, nếu mục đích sử dụng chỉ là tham khảo thông tin, thì sản phẩm chỉ cần chỉnh sửa một cách tối thiểu, nhưng nếu sản phẩm dịch được sử dụng để xuất bản và phổ biến rộng rãi, thì việc chỉnh sửa phải cẩn thận hơn nhiều [W.J. Hutchins và H.L. Somers, 1992, tr.172].

2.4. Việc sử dụng dịch máy

Theo Sakdanh, Gabriela, Mona Baker (2009, tr.185), các hình thức máy dịch phổ biến đang đang sử dụng hiện nay là các máy dịch cho mục đích riêng/ đặc biệt (special-purpose machine translation system); máy dịch nói/ phiên dịch (speech machine translation system) và dịch máy trực tuyến (online machine translation system).

Dịch máy đa năng (dịch nhiều ngôn ngữ và nhiều lĩnh vực) gặp nhiều khó khăn khi phải dịch tất cả các văn bản với nhiều chủ đề khác nhau một cách đáng tin cậy được nên không thể thiếu khâu chỉnh sửa trước (pre-editing) và chỉnh sửa sau (post-editing). Còn máy dịch có mục đích riêng/ đặc biệt (special purpose machine translation systems) có hiệu quả khi dịch văn bản có từ vựng và cấu trúc ngôn ngữ theo một khuôn khổ nhất định (formulaic language) hoặc các văn bản chuyên về kỹ thuật, như chỉ số về các sản phẩm, hướng dẫn bảo dưỡng, các bản tin và văn bản pháp lý của chính phủ, v.v. ví dụ điển hình về thành công của máy dịch với mục đích đặc biệt là máy dịch METEO khi dịch các bản tin về dự báo thời tiết cho Trung tâm thời tiết Canada. Meteo được đặc biệt phát triển để dịch các bản tin thời tiết từ tiếng Anh sang tiếng Pháp (ở các tỉnh của Canada) và từ tiếng Pháp sang tiếng Anh (ở tỉnh Quebec). Meteo bắt đầu hoạt động vào 1981 và kết thúc hoạt động vào năm 2001; mỗi năm Meteo dịch 20 triệu từ từ tiếng Anh sang tiếng Pháp và 10 triệu từ từ tiếng Pháp sang tiếng Anh, với tốc độ 10 trang tin thời tiết trong một giây, và khoảng 45.000 từ một ngày [Mona Baker, 2001, tr.158].

Máy dịch lời nói/ phiên dịch (Speech machine translation) được hiện thực hoá từ những năm 1980. Máy dịch lời nói/ phiên dịch tổng hợp kết quả của module nhận dạng lời nói (speech-to-text), module dịch máy (machine translation technology) và module tạo lời nói (text-to-speech) [Sakdanh,

Gabriel, Mona Baker, 2009, tr.186]. Ví dụ: Google translate, Microsoft translator, Amazon translate, ITranslate Voice và một số người dùng phải trả phí như: IBM Watson language translator. Cũng như dịch máy, Dịch máy lời nói/ phiên dịch có chất lượng không cao (theo đúng thời gian thực) và muốn có chất lượng tốt hơn vẫn cần có sự tương tác giữa máy dịch và người dịch. Trong trường hợp này người dịch vẫn đưa ra quyết định cuối cùng và quá trình dịch sẽ cần nhiều thời gian hơn.

Với sự phát triển mạnh mẽ của công nghệ thông tin, công nghệ bán dẫn, dịch máy hay các ứng dụng dịch máy trực tuyến thực sự bùng nổ từ giữa những năm 1990. Ví dụ: CompuServe mời chào dịch vụ dịch trực tuyến thư điện tử (e-mail); SYSTRAN giới thiệu ứng dụng dịch trực tuyến các văn bản, các trang web trên nền tảng dịch vụ AltaVista Babel Fish [Sakdanha, Gabriel, Mona Baker, 2009, tr.186]. Dịch vụ AltaVista Babel Fish cho phép người sử dụng dịch các văn bản hoặc nội dung các trang web từ ngôn ngữ này sang ngôn ngữ khác. Hiện nay, hầu hết các cổng Internet bao gồm cả Google, Yahoo, đều có các dịch vụ dịch máy trực tuyến miễn phí.

Do yêu cầu dịch trực tuyến ngày một tăng và ngày càng có nhiều ứng dụng dịch máy trên mạng Internet cũng như sự phát triển, sự phổ biến của mạng toàn cầu này đã tạo ra sự bùng nổ mạnh mẽ của dịch máy vào những thập kỉ đầu của thế kỉ 21. Các nhà phát triển phần mềm dịch thuật-dịch máy (machine translation software) và các phần mềm/ công cụ hỗ trợ dịch thuật (Computer aided/assisted translation- CAT) cài đặt trên máy tính đã tích hợp vào các sản phẩm của mình chức năng dịch nội dung các trang web và thư điện tử. Lynne Bowker, Jairo Buitrago (2019), cho rằng người sử dụng dịch máy có thể là những người như sau:

Những người dịch chuyên nghiệp tự do hoặc làm cho một trung tâm dịch thuật và có thể một trung tâm dịch thuật trực tuyến nào đó. Họ có kiến thức, kĩ năng ngôn ngữ, kiến thức về văn hoá và kĩ năng dịch; họ tận dụng ưu thế của dịch máy là tốc độ nhanh và chi phí rẻ vừa đảm bảo được chất lượng sản phẩm dịch máy thông qua khâu chỉnh sửa sau (post-editing) do chính họ có thể thực hiện.

Những người đã được đào tạo dịch thuật, họ là nhân viên của một công ty có sử dụng ngoại ngữ, họ tuy cũng có tiếp xúc với người nước ngoài và dịch tài liệu từ ngoại ngữ sang tiếng mẹ đẻ hoặc ngược lại nhưng các công việc không liên quan đến ngoại ngữ lại rất nhiều chiếm đa số thời gian làm việc và năng lực của họ. Do công ty thay đổi phương thức kinh doanh, thay đổi địa bàn hoạt động, các nhân viên lúc này có quá nhiều việc phải làm: vừa phải làm các công việc văn phòng như thường lệ, vừa phải dịch nội dung các trang web của các công ty đối tác tiềm năng trên mạng Internet để nghiên cứu thị trường, vừa phải đi phiên dịch cho lãnh đạo tiếp xúc với khách hàng nước ngoài và vừa phải dịch các sách hướng dẫn sử dụng cho một số lượng lớn các sản phẩm của công ty sang tiếng của nước sở tại để tiếp thị sản phẩm. Trước sức ép công việc khủng khiếp đó, các nhân viên phải sử dụng dịch máy để dịch nội dung của số lượng lớn các trang web và chỉ chỉnh sửa nội dung một phần hoặc toàn bộ nội dung của trang để trình lãnh đạo. Khi làm phiên dịch, các nhân viên sử dụng máy tính bảng, điện thoại thông minh có cài đặt các phần mềm hỗ trợ dịch thuật để công tác dịch thuật chính xác và dễ dàng hơn. Tuy nhiên, khi dịch các quảng cáo, sách hướng dẫn sử dụng hoặc bảo trì các sản phẩm của công ty, các nhân viên phải sử dụng dịch máy và phải chỉnh sửa sau (post-editing) thật cẩn thận cho phù hợp với ngôn ngữ, văn hoá, phong tục tập quán của nước sở tại.

Những người có tiếp xúc với ngoại ngữ ở mọi trường hợp, mọi trình độ. Họ sử dụng ngay bản dịch “thô” (sản phẩm của dịch máy) mà không cần phải chỉnh sửa với chất lượng không cao lắm [W. John Hutchins, Harol L. Somers, 1992, tr.157].

Đó là những nhà nghiên cứu, sinh viên các ngành khoa học khác và một lượng lớn sinh viên quốc tế sang các nước khác học các khoá học bằng ngôn ngữ mà không phải tiếng mẹ đẻ/ bản ngữ của mình. Họ cần đọc, dịch tài liệu cần thiết cho nghiên cứu khoa học, học tập và viết luận văn tốt nghiệp. Các tài liệu, các bài báo trên các tạp chí khoa học như vậy rất nhiều, có bài viết về cái mà họ cần và cũng có nhiều bài viết chỉ có một số phần liên quan đến cái mà họ cần. Họ sử dụng dịch máy để tìm kiếm tài liệu, họ không đòi hỏi quá cao về chất lượng bản dịch tổng thể, họ đọc lướt bản dịch tìm ra các phần, các đoạn cụ thể và chỉ tham khảo người dịch chuyên nghiệp chỉnh sửa một số đoạn

trong tài liệu cần cho nghiên cứu của họ. Tuy nhiên, trước khi in và phổ biến nghiên cứu của mình bằng một ngoại ngữ, họ có thể chỉnh sửa trước (pre-editing) tài liệu hoặc nhờ người dịch chuyên nghiệp chỉnh sửa sau (post-editing) bản dịch máy để có bản dịch cuối cùng có chất lượng cao [Lynne Bowker, 2019, tr.2].

Nhân viên hoặc lãnh đạo cơ quan hoặc một công ty có nhiều giao dịch với đối tác nước ngoài. Họ thường xuyên nhận nhiều thư điện tử, một số thư điện tử có nội dung quan trọng và có thư điện tử khác thì không. Họ sẽ sử dụng dịch máy để dịch và xử lý hết các thư điện tử trên, thư nào không quan trọng thì chỉ xem nội dung và trả lời người gửi, còn thư nào có nội dung quan trọng thì phải cẩn thận chỉnh sửa bản dịch để báo cáo cấp trên.

Khách du lịch, người bán hàng và cả những người dân bình thường: họ biết ít, thậm chí không hề biết ngoại ngữ nhưng họ có thể dùng máy tính bảng hoặc điện thoại thông minh để nói chuyện, giao tiếp với người nước ngoài. Giao tiếp giữa hai người, hai nhóm người như vậy thường là những câu giao tiếp có từ vựng thông thường và ngữ pháp đơn giản và với các chủ đề hàng ngày. Chất lượng dịch của dịch máy lúc này rất thấp, nhưng mục đích giao tiếp là hai bên hiểu nhau là được.

Khách tham quan viện bảo tàng, phòng trưng bày nghệ thuật, thực khách tại một nhà hàng, khách mua hoặc tham quan một hội chợ sách, cửa hàng sách ở nước ngoài; họ chỉ cần đeo Google glass hoặc sử dụng điện thoại thông minh có ứng dụng dịch máy là họ có thể biết được nội dung thông tin về vật trưng bày trong viện bảo tàng, nội dung chính và giá của các cuốn sách, tên và thậm chí là cả giá của các món ăn bằng tiếng mẹ đẻ của mình.

3. Kết luận

Với sự cố gắng khôi phục một ngôn ngữ toàn trí (a pansophic language), dịch máy đã được hình thành trên ý tưởng vào thế kỉ 17 và được hiện thực hoá vào giữa thế kỉ 20, nhưng chính sự phát triển của các ngành khoa học xã hội, công nghệ, Internet, quá trình toàn cầu hoá (globalization) và địa phương hoá (localization) đã tạo ra một sức ép lớn thúc đẩy hoạt động dịch thuật: phải nhanh hơn, tốt hơn, rẻ hơn, tiện lợi hơn và một lần nữa chính sức ép này lại giúp cho dịch máy tiến hoá, phát triển nhanh chóng kể cả về công nghệ lẫn phương pháp dịch máy. Các thập kỉ cuối thế kỉ 20 và hai thập kỉ đầu thế kỉ 21 chứng kiến sự bùng nổ về sự đa dạng và tính phổ biến của dịch máy trong giới dịch thuật chuyên nghiệp, người dịch không chuyên nghiệp, cho các đối tượng có liên quan đến người nước ngoài và ngoại ngữ trong đó có sinh viên đại học. Các yếu tố trên bắt buộc người dịch và giáo viên chúng ta phải thay đổi cả về quan điểm, nếp nghĩ về dịch máy mà còn phải song song giúp sinh viên chuyên ngữ phát triển các kĩ năng ngôn ngữ, kĩ năng dịch và tích hợp giảng dạy và hướng dẫn sinh viên nắm vững kiến thức và thực hành dịch máy với mục đích là sinh viên chuyên ngữ khi tốt nghiệp vừa phải thành thạo các kĩ năng ngôn ngữ, kĩ năng dịch và sử dụng thành thạo dịch máy để sau khi tốt nghiệp dễ dàng hoà nhập vào môi trường làm việc đầy cạnh tranh trước mắt cũng như tương lai sau này.

TÀI LIỆU THAM KHẢO

1. Isabel Lacruz, Riita Jaaskelainen (Eds.) (2018), *Innovation and Expansion in Translation Process research*: John Benjamins Publishing Company.
2. Arturo Trujillo (1999), *Translation Engines - Techniques for Machine Translation*: Springer- Verlag London.
3. Gloria Corpas Pastor, Isabel Duran- Munoz (2018), *Trends in E-Tools and Resources for Translators and Interpreters*, Brill Rodopi.
4. Alan K. Melby, Terry Warner (1995), *The Possibility of Language - A discussion of the nature of language with implications for human and machine translation*, John Benjamins Publishing Company.
5. Cay Dollerup -Anette Lindergaard (1994), *Teaching Translation and Interpreting 2- Insights, aims and visions*, John Benjamins Publishing Company.

6. Lynne Bowker (2002), *Computer-Aided Translation Technology_ A Practical Introduction*, University of Ottawa Press.
7. Joss Morkens, Sheila Castilho, Federico Gaspari, Stephen Doherty (2018), *Translation Quality Assessment*, Springer International Publishing.
8. Michael Cronin (2013), *Translation in Digital Age*, Routledge.
9. Alan Duff (1989), *Translation (Resource Books for Teachers)*, Oxford University Press.
10. Basil Hatim, Jeremy Munday (2005), *Translation - An Advanced Resource Book*, Routledge.
11. Geoffrey Samuelsson- Brown (2010), *A Practical Guide for Translator*, Multilingual Matters.
12. Jakub Absolon (2019), *Human Translator 4.0*, ASAP - Translation.
13. Lynne Bowker, Michael Cronin, Doroty Kenny, Jennifer Pearson (1998), *Unity in Diversity - Recent Trends in Translation Studies*, Routledge.
14. Lynne Bowker, Jairo Buitrago (2019), *Machine Translation and Global Research- Towards improved Machine Translation literacy in the Scholar community*, Emerald Publishing Ltd.
15. Thierry Poibeau (2017), *Machine Translation*, MIT Press.
16. Martha Catalina del Angel Castillo (2019), *Translation Strategies for English learner*, Editorial Digital .
17. Mona Baker (2001), *Routledge Encyclopedia of Translation studies*, Roulledge.
18. Saldanha, Gabriela- Mona Baker (2009), *Routledge encyclopedia of translation studies*, Routledge.
19. Minako O'Hagan (editor) (2020), *Routledge Handbook of Translation technology*, Routledge.
20. W. John Hutchisons, Harold L. Somers (1992), *An Introduction to Machine Translation*, Academic Press.

Issues in machine translation basic

Abstract: In spite of being realized in Mid-twentieth century, until the late decades of the 20th century and early decades of the 21st century, machine translation has really developed and become widely popular among professional, non-professional translators, researchers, scientists, students as well as among people whose daily life activities relate to foreigners and foreign languages. In response to the situation, the writer recommends that it was high time for university foreign language teachers to change the mindset about machine translation as well as to improve the knowledge of machine translation and be ready to teach machine translation to students in order to train foreign language-majored students to be the graduates with good language skills, human translation skills and with sufficient knowledge and skills in using machine translation in a highly competitive working environment in the future. Therefore, the article will briefly introduce the history and the developemnet of machine translation, the evolution of machine translation approaches. The articles will also analyze the strengths as well as the limitations the weaknesses of machine translation, especially several measures to overcome such limitations and weaknesses.

Key words: Machine translation; history and evolution of machine translation; machine translation approach; use of machine translation.