Liệu các máy dịch có thể thay thế cho con người?

NGUYỄN NGỌC VŨ*

Kể từ khi xuất hiện, máy tính đã cách mạng hóa cách thức chúng ta sống và làm việc. Nhờ phát minh ra máy tính mà chúng ta đã đạt được nhiều thành tựu to lớn trong nhiều lĩnh vực và đấy mới chỉ là giai đoạn mở đầu. Trong lĩnh vực học tập, giảng dạy ngoại ngữ cũng như trong ngành ngôn ngữ học, việc dịch thuật bằng máy tính đã được biết đến trong nhiều thập niên qua nhưng câu hỏi liệu máy tính có thể dịch hiệu quả như con người không vẫn còn là một câu hỏi chưa có lời đáp. Bài viết này nhằm mục đích đánh giá khả năng dịch của các máy tính và từ đó giải quết một phần câu hỏi nêu trên.

Trái với suy nghĩ của nhiều người cho rằng việc dịch bằng máy tính là lĩnh vực mới mẻ, dịch bằng máy tính đã có một lịch sử dài (Hutchins 2003). Từ năm 1947, khi các máy tính dùng cho mục đích phi quân sự ra đời, ý tưởng dịch bằng máy tính đã được đưa ra và được hưởng ứng rộng rãi. Năm 1952, hội thảo đầu tiên về dịch bằng máy tính được tổ chức. Hai năm sau đó, hệ thống dịch bằng máy đầu tiên được giới thiệu với công chúng. Lúc đó ai cũng nghĩ rằng kỉ nguyên dịch bằng máy tính mới chỉ bắt đầu, rằng những người làm công tác phiên dịch sẽ thất nghiệp và rằng ai cũng có thể phiên dịch bất kì văn bản nào chỉ bằng một động tác bấm nút. Tuy nhiên, mọi việc lại không như mong đợi. Kết quả thu được từ các dự án dịch bằng máy tính tệ đến nỗi năm 1966 chính phủ Mĩ ngưng tài trợ cho chương trình này. Thay vào đó, người ta tập trung nỗ lực vào việc phát triển các công cụ hỗ trợ người dịch. Các dự án của Liên Xô trong lĩnh vực này vốn được khởi động vào giữa những năm 1950 cũng bị ngưng lại. Dù vậy, các nghiên cứu về việc dịch bằng máy tính vẫn được tiếp tục trong những năm 70. Năm 1976, Canada đã có được bước đột phá khi lắp đặt thành công hệ thống dịch các thông báo về thời tiết (Météo System) và sử dụng rộng rãi trong công chúng. Trong những năm sau đó, Liên Minh Châu Âu cũng đầu tư rất nhiều tiền bạc để phát triển các hệ thống dịch máy cho một số cặp ngôn ngữ ở Tây Âu nhưng kết quả không được như mong đợi. Năm 1981, phần mềm dịch thuật được sử dụng cho các máy tính cá nhân và dần dần việc dịch bằng máy tính trở nên phổ biến hơn. Trong những năm 1980, người ta lại ra sức nghiên cứu về lĩnh vực này. Cũng trong thời gian đó, các công ti Nhật bắt đầu sản xuất các phần mềm và công cụ

* Trường Đại học Sư phạm TP.HCM.

hỗ trợ người dịch. Năm 1990, các trạm dịch trên mạng bắt đầu hoạt động và trong khoảng 8 năm trở lại đây, dịch vụ dịch bằng máy xuất hiện trên mạng Internet.

Với một bề dày lịch sử phát triển và nhiều công sức được bỏ ra như vậy, lẽ ra giờ đây các hệ thống dịch máy có thể làm cho các dịch giả thất nghiệp. Tuy nhiên, theo số liệu của nhóm TTT (Translation-Theory-Technology 2001) thì việc dịch bằng máy cho đến nay chỉ chiếm chưa đến 10% lượng văn bản dịch trên toàn thế giới. Tại sao vậy? Câu trả lời nằm ở tính phức tạp của ngôn ngữ.

Trước hết, các ngôn ngữ khác nhau ở rất nhiều điểm. Điểm khác biệt lớn nhất là ở thuộc tính ngữ nghĩa. Giữa các ngôn ngữ, một từ có thể có nhiều cách dịch. Chẳng hạn, từ "know" ở tiếng Anh có thể dịch thành "savoir" hoặc "connaitre" trong tiếng Pháp. Thêm vào đó, đôi lúc người ta không thể tìm ra được từ tương đương của ngôn ngữ này trong ngôn ngữ kia. Đối với những trường hợp này, bản thân những dịch giả chuyên nghiệp cũng gặp rất nhiều khó khăn. Năm ngoái tôi có tham gia dịch một quyển sách có nhan đề "grammatology". Trong từ điển đơn ngữ thì từ này có nghĩa là "art of writing" (nghệ thuật viết). Nhưng nếu dịch là "nghệ thuật viết" hay "viết học" thì lạ tai quá. Trong nhóm dịch lúc đó mới có một thầy giáo nảy ra ý kiến hay. Thầy này tra từ điển Anh-Hoa rồi từ đó suy ra được một từ Hán-Việt là "văn tự học". Điều này có thể thực hiện được vì 70% từ vựng tiếng Việt có nguồn gốc từ tiếng Hoa. Thêm vào đó, việc dịch các thành ngữ, khẩu ngữ hay các câu nói ẩn dụ cũng vô cùng khó khăn. Nhiều dịch giả chuyên nghiệp cũng phải thừa nhận rằng có một số thành ngữ chỉ có thể giải thích chứ không thể dịch được. Ngay cả bản thân con người cũng gặp nhiều rắc rối như vậy thì máy móc làm sao có thể giải quyết được.

Vấn đề từ vựng tuy nhiên mới chỉ là khó khăn nhỏ. Các vấn đề về cấu trúc còn phức tạp hơn nhiều. Các ngôn ngữ khác nhau có cấu trúc câu khác nhau. Đối với một số ngôn ngữ như tiếng Anh, tiếng Đức, tiếng Hoa, cấu trúc câu chính là SVO (subject-verb-object chủ ngữ-động từ-tân ngữ). Trong một số ngôn ngữ như La tinh, Nhật, Ba tư v.v... thì cấu trúc chính là SOV (subject-object-verb chủ ngữ-tân ngữ-động từ). Tiếng Ailen, Ai Cập và Arập cổ thì lại có cấu trúc VSO (verb-subject-object). Một số ngôn ngữ còn không có trật tự rõ ràng trong cấu trúc. Chính vì máy tính chỉ có thể dịch theo kiểu từ sang từ nên nó sẽ gặp rắc rối về cấu trúc và ngữ nghĩa. Chẳng hạn, trong phụ lục 2, câu tiếng Anh "Hundreds of outraged Palestinians quickly crowded around the wreckage, some waving bloodstained clothing lifted from the gutted car" được dịch thành "Les centaines de Palestiniens outragés rapidement serrés autour de l'épave, une partie ondulant bloodstained l'habillement soulevé et de la voiture étripée". Một người Pháp khi đọc đoạn dịch này chắc chắn sẽ rất khó hiểu.

Đoạn phân tích ở trên cho thấy, việc dịch qua lại từ tiếng Anh sang tiếng Pháp gặp nhiều khó khăn dù cả hai đều thuộc họ ngôn ngữ Ấn-Âu. Nếu vậy thì chuyện dịch ngôn ngữ thuộc hai họ khác nhau sẽ ra sao? Từ tiếng Anh sang tiếng Hoa chẳng hạn. Các ngôn ngữ châu Á vốn có nguồn gốc rất khác ngôn ngữ châu Âu (Chriss). Các vấn đề về ngôn ngữ học này gây khó khăn cho cả người dịch chứ đừng nói gì đến máy móc.

Một điểm cần chú ý nữa là công việc dịch thuật không chỉ yêu cầu kiến thức về ngôn ngữ học. Giáo sư Dennett (1995:19) ở trường đại học South Bank London đã đưa ra năm loại tri thức cần thiết cho việc dịch thuật:

1. Tri thức về ngôn ngữ mẹ đẻ
2. Tri thức về ngoại ngữ
3. Tri thức về mối liên hệ giữa ngoại ngữ và ngôn ngữ mẹ đẻ.
4. Tri thức tổng quát về lĩnh vực của tài liệu dịch
5. Tri thức về văn hoá và tập quán ngôn ngữ của cả ngôn ngữ mẹ đẻ và ngoại ngữ.

Chúng ta có thể lập trình máy tính để chúng có được ba loại tri thức đầu tiên nhưng để làm cho chúng có thể tổng hợp kiến thức và vốn sống vào ngôn ngữ dịch là điều không thể có được. Có rất nhiều trường hợp người dịch phải hiểu vấn đề mình đang dịch thì mới có thể dịch đúng được (Hutchins 1991). Chẳng hạn, khi dịch cụm từ "pregnant women and children" sang tiếng Pháp, chúng ta hiểu rằng từ "pregnant" (có thai) không thể dùng cho "children". Cho nên bản dịch phải là "des femmes enceintes et des enfants" chứ không phải là "des femmes et des enfants enceintes". Tất cả chúng ta đều biết là trẻ con thì không có thai nhưng làm thế nào máy tính hiểu được vấn đề trên? Tất nhiên người ta có thể lập trình trong từ điển của máy tính rằng từ "pregnant" chỉ dùng với danh từ có thuộc tính ngữ nghĩa là "nữ" (female) và "người lớn" (mature). Nhưng rõ ràng là chúng ta không thể lập trình tất cả mọi thứ cho máy tính. Trong quyển sách "Machine Translation", tác giả Arnol (1994) đã thừa nhận rằng đây là vấn đề nan giải:

"Chúng ta không thể lường hết được loại tri thức về thế giới nào máy tính cần được lập trình. Lượng thông tin ngầm mà người viết và người đọc trao đổi với nhau một cách tự động là rất lớn và khó đoán trước. Mọi nỗ lực giải mã loại thông tin này đều khó mà thành công".

Trong các thập niên vừa qua, các dự án dịch bằng máy tính tiến triển rất chậm và không đạt được mục đích đề ra ban đầu. Đúng là các hệ thống máy dịch đã dễ sử dụng hơn nhưng chất lượng các bản dịch bằng máy vẫn không tiến bộ nhiều (Hutchins 2003). Lí do chính là có rất nhiều vấn đề trong việc xử lí ngôn ngữ cho dù có sử dụng loại thiết bị máy tính nào đi nữa. Tác giả cũng cho rằng đây là bản chất

của ngôn ngữ và nó gắn liền với các khác biệt về văn hoá. Chính vì vậy máy tính sẽ luôn gặp phải khó khăn này và rất khó vượt qua được. Bây giờ khi người ta không còn hào hứng với việc dịch bằng máy nữa thì nhiều máy dịch cũng không còn được sử dụng. Công ti Weidner từng bán được cả ngàn máy dịch ở Nhật bây giờ phải đóng cửa do thua lỗ. Winger – hệ thống máy tính dịch các ngôn ngữ bắc Âu – cũng phải rút lui mặc dù có triển vọng. Gần đây, công ti Sail Labs đã từng sản xuất hệ thống dịch Đức - Anh - Đức đầy ấn tượng cũng gặp khó khăn tài chính. Trên đây chỉ là vài ví dụ trong một danh sách dài các hệ thống máy dịch bị phá sản mà Hutchins đưa ra. Thật ra, chất lượng bản dịch kém là nguyên nhân chính làm cho các máy dịch thất bại. Trong phụ lục 1, chúng tôi có thử cho phần mềm dịch trên mạng dịch thử một đoạn văn bản đơn giản nhưng kết quả cuối cùng là đúng thất vọng. Kết quả dịch trong phụ lục 2 cũng không khá hơn.

Tuy nhiên, chúng ta cũng phải thừa nhận rằng có một số tình huống có thể dùng máy dịch. Khi ta không cần một bản dịch chất lượng cao thì dịch bằng máy là một lựa chọn tốt. Trong nhiều trường hợp, khi ta chỉ cần biết đại ý của văn bản thì máy tính có thể thỏa mãn yêu cầu này trong tích tắc. Đôi lúc người dịch phải dịch những văn bản khoa học kĩ thuật. Loại văn bản này thường dùng đi dùng lại một số cụm từ mà nghĩa chỉ giới hạn trong một lĩnh vực cụ thể. Trong trường hợp này máy dịch là công cụ hữu hiệu giúp người dịch tiết kiệm thời gian và công sức. Dịch bằng máy cũng có hiệu quả cao trong các lĩnh vực mà ngôn ngữ sử dụng một lượng từ hay cấu trúc giới hạn. Hệ thống dự báo thời tiết "Météo" của Canada là một ví dụ điển hình. Do chỉ phải dịch những văn bản có lượng từ, cấu trúc giới hạn và lặp lại nhiều lần, hệ thống nay cho kết quả khả quan khi dịch từ Anh sang Pháp và ngược lại. Các nhà ngôn ngữ học và các lập trình viên máy tính đang chú ý đến việc phát triển các phần mềm dịch chuyên ngành khoa học tự nhiên tương tự. Do đó, trong một số lĩnh vực, dịch bằng máy vẫn có hiệu quả.

Ngoài ra, các máy dịch còn có những lợi thế quan trọng giúp nó tồn tại trong thời gian dài. Lợi thế đầu tiên là máy tính dịch nhanh hơn con người hàng ngàn lần. Trong thời đại bùng nổ thông tin như hiện nay, lượng văn bản cần dịch là rất lớn và không phải khi nào con người cũng dịch hết được. Máy tính chỉ cần một phút để dịch lượng văn bản mà con người phải mất cả tuần lễ mới hoàn thành xong. Các văn bản trong phụ lục 1 và 2 được dịch trong vòng chưa đến 1 giây. Quả thật tốc độ là ưu thế lớn nhất của việc dịch bằng máy. Ngoài tốc độ dịch cao, các máy dịch còn giúp người ta tiết kiệm được nhiều chi phí. Theo Heuberger, một chuyên gia về dịch bằng máy, người ta có thể tiết kiệm được từ 5% đến 50% chi phí khi dùng máy dịch. Một lí do khác nữa khiến người ta cần các máy dịch là trong nhiều trường hợp văn bản dịch phải dùng một số thuật ngữ thống nhất từ đầu đến cuối. Người dịch thường có xu hướng dùng từ khác nhau để diễn đạt một khái niệm mà điều này lại kiêng kị trong

các văn bản khoa học kĩ thuật. Do sử dụng bộ từ điển và cơ sở dữ liệu được lập trình sẵn nên máy tính có thể đảm bảo tính nhất quán đến mức tối đa. Như vậy, tốc độ cao, chi phí thấp và tính nhất quán cao là những lợi thế lớn mà người ta có được khi dịch bằng máy.

Nói tóm lại, do bản chất phức tạp của ngôn ngữ, các máy tính khó có thể dịch một cách hiệu quả mặc dù các nhà khoa học đã bỏ rất nhiều công sức cho công việc này. Từ khi ra đời đến nay, các máy và phần mềm dịch đạt được rất ít tiến bộ về chất lượng. Tuy nhiên, các máy dịch lại rất công hiệu khi ta không cần chất lượng dịch cao hoặc khi có sự giới hạn về từ vựng và cấu trúc trong văn bản. Dịch bằng máy còn có những ưu điểm khác như tốc độ cao, chi phí thấp và nhất quán. Có lẽ chỉ khi nào máy tính có thể tư duy như con người thì chúng mới có thể dịch một cách hiệu quả được. Nhưng trong tương lai gần thì điều này là không thể. Ngày mà con người có thể rút máy dịch bỏ túi ra kề vào miệng người bản xứ và nhận lại lời dịch vẫn còn rất xa vời (Fromkin 1999). Trong khoảng thời gian ấy, các nhà phiên dịch và biên dịch vẫn đóng một vai trò rất quan trọng.

TÀI LIỆU THAM KHẢO

[1]. Arnol, D., Balkan, L., Humphreys, L.R., Meijer, S. & Sadler, L., (1994), *Machine Translation – An Introductory Guide*. NCC Blackwell, Oxford.

[2]. Chriss, R., (n.d.), *Asian Languages in Translation*. http://www.translationdirectory.com/article39.htm [Accessed 15th April 2004].

[3]. Danett, G., (1995), *Translation Memory: Concept, products, impact and prospects*. South Bank University, London.

[4]. Fromkin, V., Blair, D., and Collins, P. (1999), *An Introduction to Language*. Harcourt, Sydney.

[5]. Heuberger, A., (2001), *Machine Translation vs. Translation Memory*. Foreign Exchange Translation. http://www.multilingualwebmaster.com/library/mt-vs-tm.html [Accessed 10th April 2004].

[6]. Hutchins, W.J., (1986), *Machine translation: past, present, future*. Chichester, Ellis Horwood.

[7]. Hutchins, W.J., (1991), "Why computers do not translate better" in *Translating and the Computer 13: the theory and the practice of machine translation - a marriage of convenience?* Papers presented at a conference jointly sponsored by Aslib, The Association for Information Management; Aslib Technical Translation Group; Institute of Translation and Interpreting, 18-19 November 1991, CBI Conference Centre, London WC1. http://ourworld.compuserve.com/homepages/WJHutchins/Aslib13.pdf [Accessed 10th April 2004].

[8]. Hutchins, W.J., (2003), *Machine translation and computer-based translation tools:* what's available and how it's used. http://ourworld.compuserve.com/homepages/WJHutchins/Valladolid-2003.pdf [Accessed 10th April 2004].

[9]. *Machine Translation.* Linguistic Data Consortium, University of Pelsyvania. http://www.ldc.upenn.edu/sb/cis530/materials/CL20.pdf [Accessed 12th April 2004].

[10]. Translation research group, (2001), *Machine Transaltion.* TTT (Translation-Theory-Technology) organization. http://www.ttt.org/theory/mt4me/mtwhat.html [Accessed 15th April 2004].