

THÁCH THỨC VÀ TRIỂN VỌNG CỦA VIỆC ỨNG DỤNG CÔNG NGHỆ THÔNG TIN (CNTT) VÀO CÔNG TÁC BIÊN DỊCH

ThS. Nguyễn Ngọc Vũ

Trường ĐHSP Tp. Hồ Chí Minh

I. GIỚI THIỆU

Xu hướng đẩy mạnh ứng dụng CNTT vào các hoạt động nghiên cứu và giảng dạy đang có chiều hướng ngày càng mạnh mẽ ở Việt Nam trong những năm gần đây và tất nhiên công tác biên phiên dịch cũng không nằm ngoài xu hướng ấy. Tuy nhiên trong việc nghiên cứu dịch tự động và ứng dụng CNTT vào dịch thuật đã có lịch sử hơn nửa thế kỷ ở các nước phát triển. Mặc dù đã đạt được những thành tựu nhất định song kết quả của các công trình nghiên cứu vẫn còn khiêm tốn ở nước ngoài vẫn chưa đạt được kết quả mong muốn. Bài viết này sẽ giới thiệu sơ lược lịch sử nghiên cứu ứng dụng CNTT vào dịch thuật, phân tích những khó khăn và triển vọng của lĩnh vực này với mục đích đánh giá khả năng thực sự của các công cụ dịch tự động hiện nay.

II. NỘI DUNG

Trước hết, việc dịch thuật bằng các chương trình tự động không phải là một lĩnh vực mới mẻ như nhiều người tưởng. Từ năm 1974 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 (Hutchins 2003). Năm 1952, hội thảo đầu tiên về dịch bằng máy tính được tổ chức. Hai năm sau đó, hệ thống dịch bằng máy tính 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 biên dịch sẽ thất nghiệp và rằng ai cũng có thể biên dịch bất cứ 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 tẻ đến nỗi năm 1966 chính phủ Mỹ ngưng tài trợ cho dự án 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. Trong thời gian này, các công ty Nhật bắt đầu sản xuất các phần mềm và công cụ 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 tính 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 “connaître” 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. Cách đây không lâu, tác giả 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 nghe không phù hợp. Trong nhóm dịch lúc đó mới có một thầy giáo này 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 nhà 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 Latinh, Nhật, Ba tư v.v... thì cấu trúc chính là SOV (subject-object-verb chủ ngữ-tân ngữ- động từ). Tiếng Ai-len, Ai cập và Nhật cổ thì lại có cấu trúc VSO (verb-subject-object). Một số ngôn ngữ còn lộn xộn có trật tự rõ ràng trong cấu trúc. Chính vì máy tính chỉ có thể dịch eo kiểu từ sang từ nên nó sẽ gặp rắc rối về cấu trúc và ngữ nghĩa. Chẳng hạn, khi tác giả thử dùng dịch vụ dịch tự động trên Internet ([tp://translation.langenberg.com](http://translation.langenberg.com)) để dịch câu tiếng Anh **"Like most offices, my office is a place where I can concentrate on my work and feel comfortable at the same time"** sang tiếng Pháp rồi dịch ngược lại sang tiếng Anh thì được kết quả như sau **"Like the majority of the offices, my office is a place where I then to concentrate me on my work and to tell me at the same time comfortable"**. Rõ ràng là có sự khác biệt lớn giữa câu nguồn và câu dịch mặc dù tiếng Anh và tiếng Pháp có khá nhiều từ tương đồng. Nếu vậy thì chuyện dịch ngôn ngữ thuộc hai họ khác nhau ra sao? Từ tiếng Anh sang tiếng Hoa chẳng hạn. Các ngôn ngữ Châu Á có nguồn gốc rất khác ngôn ngữ Châu Âu (Bowker, 2002). Các vấn đề 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 Bank London đã đưa ra 5 loại tri thức cần 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 3 loại tri thức đầu 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 woment 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 để máy tính hiểu được vấn đề trên? Tất nhiên người ta có thể lập trình trong từ điển của máy tính rằng là 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 tiên tiến 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 lý ngôn ngữ mà máy móc không thể xử lý được 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 ty Weidner từng bán được cả ngàn máy dịch ở Nhật Bản bây giờ cũng đã 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 ty 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 Huttchins đưa ra. Thật ra chất lượng bản dịch kém là nguyên nhân chính làm cho các máy dịch thất bại. Trong một số lần khảo sát gần đây, chúng tôi có thử cho phần mềm dịch trên mạng dịch thử một đoạn văn đơn giản nhưng kết quả cuối cùng lại khá thất vọng. Kết quả dịch những văn bản phức tạp hơn thì lại càng không ra gì.

Tuy nhiên chúng ta cũng phải thừa nhận rằng có một số tình huống có thể dùng các 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 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 kỹ 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 (01) phút để dịch lượng văn bản mà con người mất cả tuần lễ mới dịch xong. Các văn bản mà chúng tôi thử dịch trên các công cụ dịch trên mạng được dịch trong vòng chưa đến 01 giây. Có thể nói tốc độ dịch chính là ưu thế lớn nhất của máy tính. Ngoài tốc độ dịch cao, các máy dịch còn giúp người ta tiết kiệm được nhiều. Theo Heuberger, một chuyên gia về dịch bằng máy, người ta có thể tiết kiệm từ 5-50% chi phí khi dùng máy dịch. Một lý 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 kỵ trong các văn bản khoa học kỹ 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 (Newton, 1998). 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.

III. KẾT LUẬN

Nói tóm lại, do bản chất phức tạp của ngôn ngữ, việc sử dụng máy tính để dịch tự động các văn bản là chuyện rất khó khăn ngay cả đối với các ngôn ngữ họ hàng với nhau, ít nhất là trong giai đoạn hiện nay. Chính vì thế các nỗ lực dịch tự động từ tiếng Việt sang các ngôn ngữ khác họ như tiếng Anh hoặc tiếng Pháp dù rất đáng quý nhưng chắc chắn sẽ còn phải mất rất nhiều thời gian mới có thể cho kết quả chấp nhận được. Tuy nhiên khi chúng ta không cần chất lượng cao hoặc khi có sự giới hạn về từ vựng và cấu trúc trong văn bản thì dịch tự động cũng có thể là một lựa chọn để xem xét. Có lúc chỉ khi nào máy tính có thể tư duy như con người thì chúng mới có thể dịch một cách hiệu quả được. Nói theo cách của Fromkin (1999) thì ngày mà con người có thể rút máy dịch bỏ túi ra kê vào miệng của người bản xứ và nhận lại lời dịch vẫn còn rất xa vời.

Tài liệu tham khảo

- Arnol, D., Balkan, L., Humphreys, L.R., Meijer, S.& Sadler, L., 1994. *Machine Translation – An Introductory Guide*. NCC Blackwell, Oxford
- Bowker, Lynne, 2002. *Computer-aided translation technology: a practical introduction*. University of Ottawa Press, Ottawa.
- Bond, Francis, 2005. *Translating the untranslatable: a solution to the problem of generating English determiners*. Stanford, CA.
- Dorr, B., Jean. 1997. *Machine Translation: a view from the lexicon*. CUP, Cambridge.
- Dannet, G. 1995. *Translation Memory: Concept, products, impact and prospects*. South Bank University, London.
- Fromkin, V., Blair, D., and Collins, P. 1999. *An Introduction to Language*. Harcourt, Sydney.
- Heuberger, A., 2001. *Machine Translation vs. Translation Memory*. Foreign Exchange Translation. <http://www.multilingualwebmaster.com/library/mt-vs-tm.html> [Truy cập 10th April 2006]
- 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> [Truy cập 10th April 2006]
- Hutchins, W.J., 1986. *Machine translation: past, present, future*. Chichester, Ellis Horwood.
- Newton, John, 1998. *Computers in translation: a practical appraisal*. Routledge, London.
- Somers, Harold, 2003. *Computers and translation: a translator's guide*. John Benjamins Publishing Company, Amsterdam