

## XÂY DỰNG HỆ THỐNG DỊCH TỰ ĐỘNG TIẾNG VIỆT

**Đào Hồng Thu**  
(TS, Đại học Bách khoa HN)

### 1. Mở đầu

Hơn nửa thế kỉ đã trôi qua kể từ thời điểm dịch tự động được hình thành như một xu hướng khoa học độc lập<sup>1</sup>. Trong toàn bộ lịch sử tồn tại của mình, dịch tự động đã trải qua nhiều thăng trầm nhất định.<sup>2</sup> Và mặc dù còn tồn tại không ít hạn chế về các khả năng hoạt động của hệ thống dịch tự động, vấn đề về xây dựng các hệ thống dịch bằng máy vẫn đang ngày càng thu hút sự chú ý của các nhà nghiên cứu chúng.<sup>3</sup>

Hệ thống dịch tự động được phát triển liên tục từ các chương trình đơn giản nhất với việc sử dụng nguồn máy tính hạn chế, đến các hệ thống hiện đại với nguồn máy tính dồi dào và hiện đại nhất để trợ giúp cho biên phiên dịch viên. Lượng thông tin khổng lồ trong tiến trình khu vực hóa và toàn cầu hóa ngày nay đòi hỏi tự động hóa việc giải quyết một loạt vấn đề nhất định ở các phạm vi hoạt động xã hội khác nhau. Dịch tự động (dịch bằng máy) hiện nay được coi là mẫu

chốt trong các vấn đề kinh tế và xã hội của các quốc gia thời đại giao tiếp thông tin quốc tế, mặc dù khả năng sử dụng các hệ thống dịch trên cơ sở máy tính còn có những hạn chế nhất định.

### 2. Tính cấp thiết của việc xây dựng các hệ thống dịch tự động từ tiếng nước ngoài ra tiếng Việt và ngược lại

Vấn đề dịch tự động từ tiếng nước ngoài ra tiếng Việt được bắt đầu nghiên cứu vào những năm 60 thế kỉ XX. Theo các tài liệu nước ngoài<sup>4</sup>, dưới sự bảo trợ của không lực Hoa Kỳ, năm 1969 công ty Logos đã được Bernard E. Scott thành lập với mục đích tiếp tục nghiên cứu việc tổ chức hệ thống dịch tự động từ tiếng Anh ra tiếng Việt. Scott bắt đầu chuẩn bị cho việc tổ chức hệ thống dịch tự động này vào mùa xuân năm 1965 tại Viện công nghệ máy tính tại New York, USA. Tháng 6 năm 1970 hệ thống dịch tự động Logos I ra đời với từ điển tự động hóa hỗ trợ chỉ bao gồm hơn 1000 từ tiếng Việt. Việc nghiên cứu hệ thống dịch tự động này chấm dứt vào năm 1973.

Gần như đồng thời, vào đầu những năm 70 thế kỉ XX, một dự án khác về xây dựng hệ thống dịch tự động từ tiếng Anh ra tiếng Việt đã được tiến hành tại Tập đoàn viễn thông Xyzyx, California. Hệ thống này đầu tiên được xây dựng để dịch văn bản Anh - Pháp về vũ trụ học trên máy IBM 360 theo

<sup>1</sup> Dịch tự động được coi như là một xu hướng khoa học độc lập từ đầu những năm 50 thế kỉ XX trên cơ sở đan xen của một số ngành khoa học như toán học, điều khiển học, ngôn ngữ học và lập trình hóa sau khi Warren Weaver cho lưu truyền bản ghi nhớ "Bản dịch" đối với các chuyên gia thuộc các lĩnh vực kiến thức khác nhau. [Hutchins W.J., 1986; Кулагина О. С., 1979].

<sup>2</sup> Hutchins W.J. Machine Translation: past, present, future // Ellis Horwood Series in Computers and their Applications - Wiley, Halsted Press, 1986. - 382 pp.

<sup>3</sup> Беляева Л.Н. Теория и практика перевода. - Санкт - Петербург, 2003. - стр.4-7

<sup>4</sup> Hutchins W.J. Machine Translation: past, present, future // Ellis Horwood Series in Computers and their Applications - Wiley, Halsted Press, 1986. - 382 pp. - §12.2. Logos Development Corporation (1969-78)

nguyên tắc hoạt động, về cơ bản, tương tự như của hệ thống Logos và của phiên bản Systran những năm 70 thế kỉ trước.<sup>5</sup>

Có thể nói rằng, cùng với sự phát triển của các thế hệ máy tính, dịch tự động từ tiếng nước ngoài (chủ yếu là tiếng Anh) ra tiếng Việt được bắt đầu nghiên cứu ở Việt Nam từ cuối những năm 80 thế kỉ XX. Trong những năm gần đây, việc sử dụng được ngoại ngữ trong giao lưu quốc tế đóng vai trò ngày càng quan trọng. Nhiều loại từ điển điện tử, các loại sách song ngữ đã và đang được hình thành nhằm phục vụ cho việc dạy và học ngoại ngữ. Một số nhóm gồm các lập trình viên đã có những nghiên cứu về tổ chức hệ thống dịch tự động từ tiếng Anh ra tiếng Việt (ví dụ, EVTrans, nhóm nghiên cứu về tổ chức dịch tự động của Trường Khoa học Tự nhiên - Đại học Quốc gia, Thành phố Hồ Chí Minh).

Trong khi các nhà ngôn ngữ chưa hoàn toàn sẵn sàng cho việc tổ chức phần mềm dịch tự động thì có quan niệm cho rằng nhà ngôn ngữ chỉ cần mô tả các hiện tượng ngôn ngữ, còn thành lập lôgarit dịch là phần việc của lập trình viên. Tuy nhiên, thực tế đã cho thấy rằng quan niệm trên không chính xác. Điều này thể hiện ở chỗ, cho đến nay Việt Nam vẫn chưa có một hệ thống dịch tự động tiếng nước ngoài - Tiếng Việt - tiếng nước ngoài nào có độ chuẩn xác cao như ETAP, PROMT, SILOD v.v.

Dịch tự động là dạng đặc biệt trong hoạt động dịch thuật của biên phiên dịch viên. Đồng thời, dịch tự động là phương tiện đặc biệt trong giao tiếp giữa các ngôn ngữ và

cung cấp thông tin kinh tế, xã hội hiện đại cần thiết không những chỉ đối với các quốc gia trên toàn thế giới, mà còn cần thiết đối với cá nhân mỗi con người trong hoạt động đời sống xã hội của mình.

Là một nước đang phát triển, Việt Nam cần sử dụng nhiều công nghệ tiên tiến của thế giới. Trong số các công nghệ tiên tiến đang được các nước trên thế giới sử dụng nhiều nhất hiện nay có công nghệ dịch tự động. Việc cần thiết sử dụng các hệ thống dịch tự động tiếng Việt bắt nguồn từ hàng loạt các nguyên nhân khách quan. Thứ nhất, ở Việt Nam hiện nay, tăng mạnh nhu cầu đọc nhanh để lấy được nhiều thông tin mới, bổ ích, đặc biệt là các thông tin về khoa học, công nghệ và kĩ thuật mới trong các nước tiên tiến trên thế giới. Thứ hai, trong tất cả các trường học và công sở của Việt Nam ngày càng đòi hỏi sự nắm vững tiếng Anh như một phương tiện cần thiết trong giao tiếp giữa các ngôn ngữ. Thứ ba, nhu cầu giao lưu bằng nhiều ngoại ngữ về chuyên ngành khoa học ngày càng tăng. Thứ tư, sự phát triển của công nghệ máy tính và Internet đã thúc đẩy nhu cầu khai thác thông tin trên nhiều trang web, bằng nhiều ngôn ngữ khác nhau.

Cần thừa nhận rằng dịch tự động không thể thay thế hoàn toàn việc dịch thuật được thực hiện bởi con người. Tuy nhiên, nhờ hệ thống tự động trong chương trình dịch bằng máy, có thể nâng cao đáng kể hiệu quả lao động của người dịch và chất lượng của cán bộ chuyên ngành. Ngày nay, dịch tự động đã phát triển đến mức có thể thực hiện dịch trực tiếp trên mạng. Tuy vậy, trong số các ngôn ngữ có thể dịch tự động trực tuyến, chưa có tiếng Việt.

Rõ ràng rằng trong các điều kiện hiện nay, khi "người dịch chuyên nghiệp cần phải biết vô vàn các thuật ngữ của nhiều chuyên ngành và cần phải thuộc rất nhiều tên gọi chính xác các chủng loại chi tiết, linh kiện,

<sup>5</sup> Дао Хонг Тху. Проблема создания систем машинного перевода с вьетнамского / на вьетнамский язык - Прикладная лингвистика в науке и образовании. Третья международная научная конференция, 16-17 марта 2006, С-Петербург, с.49-54;

dụng cụ, cơ cấu, các chất v.v. khác nhau<sup>6</sup>, thì một phiên dịch viên dù giỏi đến đâu cũng không thể cập nhật hết được một lượng thông tin khổng lồ trong nền kinh tế toàn cầu phát triển. Lúc này, việc sử dụng dịch tự động để trợ giúp cho quá trình dịch thuật là tất yếu và cần thiết.

### 3. Lời cuối

Hiện nay, việc xây dựng hệ thống dịch tự động bằng tiếng Việt là rất cần thiết để đáp ứng nhu cầu to lớn của sự phát triển kinh tế và xã hội mang tính chất toàn cầu.

Trên thực tế, còn chưa có các số liệu về khả năng xây dựng được hệ thống dịch tự động hóa hoàn toàn có thể đảm bảo chất lượng dịch cao mà không đòi hỏi bất cứ sự hiệu đính nào. Vì vậy, khi xây dựng hệ thống dịch tự động cần phải tính đến các yêu cầu thực tế của người sử dụng hệ thống và khả năng mô hình hóa quá trình tư duy lời nói của chương trình hỗ trợ dịch tự động<sup>7</sup>.

### TÀI LIỆU THAM KHẢO

1. Байрамова Л.К. *Вопросы машинного перевода*. Казань, 1973.- 96 с.
2. Бельская И.К. *Язык человека и машина*. М., 1969. - 410 с.
3. Беляева Л.Н. *Теория и практика перевода*. Санкт-Петербург, 2003. - 84с.
4. Зубов А.В., Зубова И.И. *Информационные технологии в лингвистике*. М., 2004. - 204 с.
5. Кулагина О.С. *Машинный перевод: современное состояние*.// В сб.: Семиотика и информатика. Вып. 29. // М., ВИНТИ, 1989.
6. Дорот В., Новиков Ф. *Толковый словарь современной компьютерной лексики*. БХВ - Пет., 2001. - 512 с.

7. Марчук Ю.Н. *Проблемы машинного перевода*. М., 1983. - 233с.

8. Microsoft Press. *Толковый словарь по вычислительной технике*. М., Русская редакция, 1995. - 496 с.

9. Нелюбин Л.Л. *Компьютерная лингвистика и машинный перевод*. М., 1991, 151 стр.

10. Стеблин-Каменский М.И. *Значение машинного перевода для языкознания*. // В сб.: Материалы по машинному переводу. Сб. 1. // Л., Изд. ЛГУ, 1958.

11. Федоров А.В. *Основы общей теории перевода*. М., 1983.

12. Холзнер С.-Perl. *Специальный справочник*. СПб: Питер, 2001. - 496с.

13. Grover, C., Carroll, J. and Briscoe, T. *The Alvey Natural Language Tools Grammar (4-th Release)*. University of Cambridge, 1993.

14. Kay M. *The Proper Place of Men and Machines in Language Translation*. Working paper CSL - 80 - 11. Xerox PARC, 1980

15. *Machine Translation - Theoretical and methodological issues*. Ed. by Sergei Nirenburg. Cambridge, 1987. - 350 с.

16. Hutchins W.J. *Machine Translation: past, present, future* // Ellis Horwood Series in Computers and their Applications - Wiley, Halsted Press, 1986. - 382pp.

17. Дао Хонг Тху. *Проблема создания систем машинного перевода с вьетнамского / на вьетнамский язык* - Прикладная лингвистика в науке и образовании. Третья международная научная конференция, 16-17 марта 2006, С-Петербург, с.49-54;

18. Дао Хонг Тху. *Машинный перевод с вьетнамского языка и на вьетнамский язык (вьетнамско-английский и англо-вьетнамский машинный перевод)* - Труды международной конференции "Корпусная лингвистика - 2006" 10-14 октября 2006 г., С-Петербург, изд. С-Петербургского университета, 2006, с.99-105.

<sup>6</sup> Беляева Л.Н. Теория и практика перевода. Санкт-Петербург, 2003, с.19.

<sup>7</sup> Беляева Л.Н. Лингвистические автоматы в современных информационных технологиях. - Санкт - Петербург, 2001. - стр.50