### DỊCH MÁY (MACHINE TRANSLATON)

### Dịch máy hay dịch tự động (machine translation) là một ứng dụng trên máy tính được áp dụng để chuyển tự động một văn bản từ ngôn ngữ này sang một ngôn ngữ khác. Ngày nay, nhu cầu của một hệ thống dịch tự động đang trở nên vô cùng bức thiết khi số lượng văn bản xuất hiện trên môi trường mạng toàn cầu gia tăng một cách khủng khiếp. Sự tồn tại của một hệ dịch tự động hiệu quả sẽ giúp tiết kiệm một khoảng chi phí không nhỏ cho các tổ chức, và cá nhân. Chất lượng của bản dịch phụ thuộc nhiều vào phương pháp dịch thuật. Trong bài này tôi sẽ giới thiệu 3 phương pháp dịch tự động: *dịch máy dựa trên luật, dịch máy dựa trên thống kê và dịch máy dựa trên ví dụ.*

### 1. Dịch máy dựa trên luật

Đó là việc áp dụng các tri thức ngôn ngữ của các cặp ngôn ngữ nguồn-đích do các nhà ngôn ngữ học xây dựng để từ đó mở ra một hướng tiếp cận phổ biến trong một thời gian dài: tiếp cận dựa trên luật (rule-based machine translation). Cách tiếp cận này đòi hỏi rất nhiều công sức để xây dựng nên cơ sở tri thức cần thiết cho một hệ thống dịch máy. Hơn nữa sẽ có những lỗi do mâu thuẫn giữa các luật hoặc do tập luật không bao quát sẽ dẫn đến hiệu quả không cao. Khó khăn lớn nhất trong việc áp dụng cách tiếp cận này đối với những ngôn ngữ mới là những yêu cầu về phân tích từ pháp, cú pháp,...cũng như những yêu cầu nghiên cứu về việc chuyển đổi từ vựng, cấu trúc giữa cặp ngôn ngữ. Vì vậy, xây dựng một hệ dịch trên ngôn ngữ mới (với ngôn ngữ tiếng Việt chẳng hạn) sẽ gặp rất nhiều khó khăn.

### 2. Dịch máy dựa trên thống kê

Dịch máy dựa trên thống kê là hướng tiếp cận hoàn toàn dựa trên dữ liệu nên nó có tính độc lập ngôn ngữ. Xây dựng một hệ thống dịch dựa trên hướng tiếp cận này chỉ cần tối thiểu một ngữ liệu song ngữ giữa cặp ngôn ngữ của hệ thống. Những tham số thống kê thu được từ việc huấn luyện trên ngữ liệu song ngữ sẽ được sử dụng cho việc dịch ở lần sau. Do vậy, không cần phải có một đội ngũ các chuyên gia ngôn ngữ học để hỗ trợ, kiểm tra các tri thức ngôn ngữ. Với yêu cầu tối thiểu của một hệ thống dịch máy, chi phí để xây dựng sẽ giảm được rất nhiều cũng như thời gian xây dựng một hệ thống dịch máy cho một cặp ngôn ngữ mới cũng sẽ giảm đi đáng kể. Và điểm đặc biệt hơn cả nếu ngữ liệu song ngữ được bổ sung càng nhiều sẽ càng nâng cao chất lượng cho hệ dịch.

### 3. Dịch máy dựa trên ví dụ

Hiện tại, dịch máy thống kê đang ngày càng khẳng định sức mạnh của nó do ngày càng có nhiều kho ngữ liệu lớn cho việc thống kê, nhưng nó không cho được kết quả như ý muốn trong một số trường hợp như các câu chuyên ngành.

Các tiếp cận theo dịch máy dựa trên ví dụ rất đơn giản, không đòi hỏi phải có sự phân tích ngôn ngữ học, cú pháp, ngữ nghĩa vì mọi câu dịch đều dựa vào việc “so khớp” mẫu. Việc “so khớp” mẫu dựa hoàn toàn vào kho ngữ liệu song ngữ  để xác định mẫu nào gần đúng nhất và xuất ra thành phần dịch tương ứng của mẫu đó. Vậy thì việc xây dựng kho ngữ liệu song ngữ như thế nào để hệ dịch máy dựa trên ví dụ có thể dịch ra kết quả khả quan hơn là điều đáng quan tâm.

Tuy nhiên, hiện nay chất lượng của các bản dịch tự động chưa thật tốt , các bản dịch chỉ cho biết đại ý và nó hoàn toàn có thể dịch sai một phần hoặc toàn bộ nội dung cốt lõi của văn bản. Vì vậy, hầu hết các sản phẩm dịch tự động đều chỉ mang tính tham khảo.