

Với ngôn ngữ Tiếng Việt, thuộc loại hình đơn lập, mang đặc điểm là từ Tiếng Việt không biến đổi hình thái, ranh giới từ không được xác định mặc nhiên bằng khoảng trắng. Tiếng Việt có đặc điểm là ý nghĩa ngữ pháp nằm ở ngoài từ, phương thức ngữ pháp chủ yếu là trật tự từ và từ hư. Cho nên có trường hợp một câu có thể có nhiều ngữ nghĩa khác nhau tùy vào cách ta tách từ như thế nào, gây nhập nhằng về ngữ nghĩa của câu. Ví dụ:

Với câu **"Xoài phun thuốc sâu không ăn."** có thể được tách từ như sau, với ý nghĩa hoàn toàn khác nhau:

Xoài / phun thuốc / sâu / không / ăn.

Xoài / phun / thuốc sâu / không / ăn.

Với câu **"Ăn cơm không được uống rượu."**, có thể được tách từ như sau:

Ăn / cơm / không / được / uống / rượu.

Ăn / cơm không / được / uống / rượu.

Với câu **"Mẹ vào ca ba con ngủ với dì."**, có thể được tách từ như sau:

Mẹ / vào / ca ba / con / ngủ / với / dì.

Mẹ / vào ca / ba con / ngủ / với / dì.

Điều đó cho thấy, công việc tách từ trong Tiếng Việt không phải là chuyện dễ dàng, vì nó tạo ra các câu có ngữ nghĩa hoàn toàn khác nhau, gây ảnh hưởng đến chất lượng huấn luyện mô hình học. Chính vì vậy, công việc tách từ là rất quan

trọng đối với xử lý ngôn ngữ Tiếng Việt, nhất là khi giải quyết các bài toán liên quan đến ngữ nghĩa của văn bản.