

HƯỚNG DẪN TÁCH CÂU TIẾNG VIỆT

1. Tiền đề cơ sở để tách câu:

Theo sách ngữ pháp tiếng Việt của Ủy ban Khoa học Xã hội (1980): “ *Câu là đơn vị dùng từ hay đúng hơn dùng ngữ mà cấu tạo nên trong quá trình tư duy, thông báo; nó có nghĩa hoàn chỉnh, có cấu tạo ngữ pháp, và có tính chất độc lập*”. Dựa vào quan điểm này về câu ta sẽ xét một đơn vị ngôn ngữ có phải là câu hay không.

2. Mục tiêu:

- Xác định ranh giới rõ ràng và nhất quán giữa các câu tiếng Việt. Phân biệt đơn vị câu với các đơn vị nhỏ hơn câu (từ, ngữ...) và lớn hơn câu (đoạn, văn bản).
- Làm tiêu chí chính để xây dựng nên câu tiếng Việt trong ngữ liệu tiếng Việt.
- Làm cơ sở để gán các nhãn ngôn ngữ cao hơn (tách từ, gán nhãn từ loại, phân tích cú pháp...)
- Làm tiền đề cho các bài toán khác, như: đóng câu trong song ngữ Anh-Việt, Pháp-Việt, dịch tự động Việt-Anh...

3. Phân tích và nhận diện câu:

3.1. Phân tích câu:

Xét về cấu tạo có câu đơn, câu ghép.

3.1.1 Câu đơn:

Một câu đơn cơ bản gồm có một nòng cốt đơn. Nòng cốt đơn gồm có hai phần, phần đề và phần thuyết (theo quan điểm ngữ pháp chức năng) mà quan điểm ngữ pháp truyền thống gọi là chủ ngữ và vị ngữ.

Ví dụ 1:

Bão Lekima cấp 11 / đang hướng vào Nghệ An - Hà Tĩnh.

Mọi chuyện / rồi sẽ qua đi.

Trong cấu tạo câu đơn có thể có những thành phần ngoài nòng cốt như thành phần than gọi, thành phần chuyển tiếp, thành phần chú thích, thành phần tình huống, thành phần khởi ý.

Ví dụ 2:

*Nhiều lúc, tôi cũng muốn gào thét thật to, đập tung, phá vỡ tất cả...
Con người, đó là cái vốn quý nhất.
Chao, đường còn xa lắm!*

Riêng với thành phần than gọi thì ta chỉ xét nó thuộc nòng cốt câu khi nó đứng ở cuối hoặc ở giữa câu.

Ví dụ 3:

Chúng ta đi về đi, bà con ơi!

Khi thành phần than gọi đứng ở đầu câu thì ta xem nó là một câu. Vì vốn dĩ thành phần than gọi đã có tính chất độc lập. Hơn nữa, nó được ngăn cách với nòng cốt câu bằng dấu (!) nên ta xem nó như một câu đặc biệt.

Ví dụ 4:

*Trời! Nó lại quay trở về.
→ Tách thành 2 câu:
Trời!
Nó lại quay trở về.*

Câu đơn đặc biệt là câu mà nòng cốt đơn chỉ có một thành phần.

Ví dụ 5:

*Chỉ còn lại những ngày cuối cùng...
Điều chỉnh lại mình đi!*

3.1.2 Câu ghép:

Về cấu trúc, câu ghép được tạo nên bởi ít nhất hai vế, mỗi vế là một nòng cốt đơn. Câu ghép cũng có thể có những thành phần ngoài nòng cốt như câu đơn. Về cấu trúc câu ghép thì có hai loại câu ghép cơ bản là câu ghép song song (câu ghép đẳng lập) và câu ghép qua lại (câu ghép chính phụ).

Câu ghép song song (câu ghép đẳng lập)

Nếu cấu trúc câu đơn giản, ngắn gọn (gồm 2 vế mà mỗi vế là một nòng cốt đơn) thì ta giữ nguyên cấu trúc câu của ngữ liệu.

Ví dụ 6:

Giọng của cháu đôi lúc đã nghẹn lại trong quá trình phiên dịch cho tổng thống và Chủ tịch nước, cháu đã cố kiềm chế những giọt nước mắt của mình vì quá xúc động.

Nếu cấu trúc câu ghép song song có hơn hai vế và quá phức tạp (gồm nhiều nòng cốt đơn) thì ta có thể tách thành những câu đơn. Bởi vì quan hệ giữa các vế trong câu ghép song song không thật chặt chẽ và tách ra càng đơn giản thì việc xử lý dữ liệu sẽ càng dễ dàng.

Ví dụ 7:

“Mong ước của tôi là: đấu tranh cho đến khi đất nước giành được độc lập và sau đó lập quan hệ ngoại giao và bình thường hóa quan hệ giữa VN và Mỹ, được như vậy thì tôi có thể mỉm cười mà nhắm mắt xuôi tay bất cứ lúc nào cũng thỏa lòng rồi”.

→ Câu trên là một câu ghép đẳng lập gồm nhiều nòng cốt đơn. Ta có thể tách thành:

Mong ước của tôi là: đấu tranh cho đến khi đất nước giành được độc lập và sau đó lập quan hệ ngoại giao và bình thường hóa quan hệ giữa VN và Mỹ.

Được như vậy thì tôi có thể mỉm cười mà nhắm mắt xuôi tay bất cứ lúc nào cũng thỏa lòng rồi.

Ví dụ 8:

Đa số bà con ủng hộ chủ trương xây dựng khu đô thị mới Thủ Thiêm và họ sẵn sàng giao đất để thực hiện dự án, nhưng họ muốn phải được đảm bảo quyền lợi và cuộc sống sau khi di dời.

→ Theo ngữ nghĩa thì câu này có thể tách:

Đa số bà con ủng hộ chủ trương xây dựng khu đô thị mới Thủ Thiêm.

Họ sẵn sàng giao đất để thực hiện dự án, nhưng họ muốn phải được đảm bảo quyền lợi và cuộc sống sau khi di dời.

Tuy nhiên ta nên hạn chế việc tách câu này, đặc biệt là với những câu ghép đẳng lập mà các vế câu được nối với nhau bằng kết từ (và, rồi, hay, còn). Vì việc tách câu này có thể làm cho câu cú gọn gàng nhưng ý nghĩa tự nhiên của ngữ liệu ít nhiều đã bị thay đổi.

Câu ghép qua lại (câu ghép chính phụ)

Câu ghép chính phụ là câu ghép mà các vế trong câu phụ thuộc lẫn nhau, không thể tách ra được.

Có thể nhận biết câu ghép chính phụ qua các cặp từ quan hệ như: nếu...thì, tuy...nhưng, do...mà, ...

Ví dụ 9:

- Dù họ là nhà thầu Nhật Bản nhưng nếu họ vi phạm pháp luật VN thì vẫn xử họ theo qui định của pháp luật VN.

- Và lại, đây là loại tội phạm mới thuộc về lĩnh vực khoa học kỹ thuật, vì vậy ngoài lực lượng điều tra của ngành công an, chúng tôi cần phải phối hợp với các ngành chuyên môn khoa học kỹ thuật khác để tìm ra nguyên nhân.

Giả sử mẹ nắm 60% vốn của công ty con; vậy mẹ phải cử đại diện dự các phiên họp của ĐHCD của công ty con và biểu quyết theo số vốn góp.

3.2. Nhận diện câu:

3.2.1 Nhận diện chung:

Với các kiểu câu bình thường như trên ta có thể nhận biết câu qua dấu câu: dấu chấm (câu tả, câu trần thuật, câu kể), dấu chấm than (câu cảm, câu cầu khiến), dấu chấm hỏi (câu hỏi).

3.2.2 Nhận diện câu trong hội thoại:

Trong hội thoại dấu 2 chấm (:) báo hiệu cho lời nói trực tiếp, và lời nói trực tiếp này nằm trong dấu ngoặc kép (“...”) hoặc bắt đầu sau dấu gạch đầu dòng (-). Trong trường hợp này, ta sẽ tách câu (nhận diện câu qua dấu hai chấm (:)).

Ví dụ 10:

Ông cho biết:

- Căn cứ vào kết quả kiểm tra, khảo sát và những chứng cứ thu thập ban đầu từ các đơn vị nghiệp vụ, tôi nhận thấy đây là một vụ án đặc biệt nghiêm trọng, gây hậu quả lớn về người và của.

Hắn nói: “Mày chạy trước đi.”

→Tách thành hai câu:

Hắn nói:

“Mày chạy trước đi.”

Đối với đoạn hội thoại có vế trích dẫn nằm ở cuối câu thì ta cũng sẽ tách câu. Vì trong lời nói trực tiếp có nhiều câu, khi ta tách chúng ra thành những câu riêng

biệt, vế trích dẫn cuối cùng sẽ gắn với câu cuối cùng làm thành một câu khác có ý nghĩa khác thì câu sẽ trở nên sai. Vì vậy ta sẽ tách vế này ra thành một câu.

Ví dụ 11:

"CSGT có nhìn thấy cũng chịu chết vì đâu có len vào được mà xử phạt. Nếu bắt dừng xe thì kẹt đường ngay", một CSGT chốt tại đây nói.

→Tách thành ba câu:

CSGT có nhìn thấy cũng chịu chết vì đâu có len vào được mà xử phạt.

Nếu bắt dừng xe thì kẹt đường ngay.

Một CSGT chốt tại đây nói.

Ví dụ 12:

"Điều khác lạ ở VN so với nhiều nước châu Âu là các doanh nghiệp sản xuất có thể tham gia phân phối,, các qui định trong kinh doanh được hiện có không qui định nhiệm vụ cụ thể của từng tổ chức trong dây chuyền phân phối " - ông Andre nhận xét.

→ tách thành 2 câu:

"Điều khác lạ ở VN so với nhiều nước châu Âu là các doanh nghiệp sản xuất có thể tham gia phân phối, các qui định trong kinh doanh được hiện có không qui định nhiệm vụ cụ thể của từng tổ chức trong dây chuyền phân phối".

Ông Andre nhận xét.

3.2.3 Nhận diện câu sau dấu chấm phẩy (;)

Dấu chấm phẩy (;) thường dùng để chỉ ranh giới giữa các vế trong câu ghép song song. Vì vậy ta có thể tách câu giống như câu ghép song song.

Ngoài những tiêu chí nhận diện câu qua câu ghép song song ta có những trường hợp khác sau:

Không nên tách câu khi sau dấu (;) là “thì”, “và”, “nên”

Ví dụ 13:

Giả sử, theo bản điều lệ, HĐQT có sáu thành viên; thì công ty mẹ phải thuyết phục các cổ đông trong ĐHCD bầu bốn người đại diện của họ vào HĐQT. → không tách câu.

Nói một cách khác theo ngôn từ ta thường dùng, cơ quan chủ quản ra lệnh cho công ty con (1) qua số vốn mình nắm và theo quyền biểu quyết đa số tương đối hay tuyệt đối trong ĐHCĐ của công ty con; và (2) có người đại diện của mình nắm đa số thành viên trong HĐQT. → không tách câu

Sau dấu (;) không phải là “thì”, “và”, “nên” thì ta có thể tách câu được. Riêng trường hợp sau “và” không phải là động từ, không phải là sự liệt kê thì cũng có thể tách được.

Ví dụ 14:

Vốn của nó do Nhà nước bỏ vào; nó hoạt động theo chỉ thị của cơ quan chủ quản; và cơ quan này là người nắm vốn duy nhất.

→ Nên tách thành:

Vốn của nó do Nhà nước bỏ vào.

Nó hoạt động theo chỉ thị của cơ quan chủ quản.

Và cơ quan này là người nắm vốn duy nhất.

Sau dấu (;) là cặp từ “nhưng (đề/ nêu/ muốn)...thì” thì cũng có thể tách câu được vì cặp từ này có khả năng tạo thành một câu có đủ ý nghĩa và hoạt động độc lập được

Ví dụ 15:

Đối với chiến lược của tập đoàn, việc nâng cao hiệu quả sử dụng đất là cần thiết; nhưng để cho các công ty con thực hiện thì đại diện của PetroVietnam tại ĐPM phải họp ĐHCĐ hay HĐQT để ra quyết định.

→ Nên tách thành:

Đối với chiến lược của tập đoàn, việc nâng cao hiệu quả sử dụng đất là cần thiết.

Nhưng để cho các công ty con thực hiện thì đại diện của PetroVietnam tại ĐPM phải họp ĐHCĐ hay HĐQT để ra quyết định.

Sau dấu (;) là một cụm từ có đầy đủ chủ vị và có khả năng độc lập thì cũng nên tách câu

Ví dụ 16:

Một nghiên cứu đã chứng minh rằng đối với các nước có trình độ phát triển thấp, mức độ phát triển xã hội là một nhân tố thích ứng với tăng trưởng; ở một trình độ cao hơn, mức độ này dẫn đến thay đổi về phát triển cơ sở hạ tầng và các thể chế kinh tế...

→ Nên tách thành:

Một nghiên cứu đã chứng minh rằng đối với các nước có trình độ phát triển thấp, mức độ phát triển xã hội là một nhân tố thích ứng với tăng trưởng.

Ở một trình độ cao hơn, mức độ này dẫn đến thay đổi về phát triển cơ sở hạ tầng và các thể chế kinh tế...

Ví dụ 17:

Theo đó, chủ xe khách 63L-5796 Võ Hồng Xuân bị phạt 2,1 triệu đồng; tài xế Đặng Hữu Thành (con bà Xuân) bị phạt 2,6 triệu đồng; tài xế xe khách 63L-5691 Lê Ngọc Trân bị phạt 2,1 triệu đồng; tài xế xe khách 63L-5634 Nguyễn Văn Thủy bị phạt 430.000 đồng.

→ Nên tách thành:

Theo đó, chủ xe khách 63L-5796 Võ Hồng Xuân bị phạt 2,1 triệu đồng.

Tài xế Đặng Hữu Thành (con bà Xuân) bị phạt 2,6 triệu đồng.

Tài xế xe khách 63L-5691 Lê Ngọc Trân bị phạt 2,1 triệu đồng.

Tài xế xe khách 63L-5634 Nguyễn Văn Thủy bị phạt 430.000 đồng.

3.2.4 Nhận diện câu sau dấu ngang (-):

Dấu ngang dùng để chỉ ranh giới của thành phần chú thích, đặt trước những lời đối thoại, liệt kê.

Đối với câu có dấu ngang dùng để chỉ thành phần chú thích thì ta không nên tách câu.

Ví dụ 18:

Cơn sốt vé trong năm nay không còn nghi ngờ gì nữa phải thuộc về ngôi sao nhạc nhẹ mới 14 tuổi Miley Cyrus, diễn viên ngôi sao của bộ phim truyền hình Hannah Montana trên Disney Channel - bộ phim nói về cuộc sống thú vị của một cô nàng vừa là sinh viên vừa là ngôi sao nhạc nhẹ.

TTO - Sau một thời gian chạy thử nghiệm, Công ty VinaGame sẽ chính thức giới thiệu Zing MP3 - công cụ tìm kiếm âm nhạc trực tuyến đầu tiên tại Việt Nam vào đầu tháng tới.

Trên đây là những trường hợp thông thường và một số trường hợp đặc biệt mà công việc tách câu thường gặp phải (đặc biệt là đối với ngữ liệu lấy từ báo chí).

3.2.5 Thực tế nhận diện câu và một số vấn đề lưu ý khác:

- Nhận diện câu trong văn bản thơ:

Khi trích dẫn thơ xuất hiện dấu / chúng ta phải tách câu.

Ví dụ 19:

“Tôi muốn tắt nắng đi / Cho màu đừng nhạt mất / Tôi muốn buộc gió lại / Cho hương đừng bay đi”

Chúng ta phải tách thành:

“Tôi muốn tắt nắng đi

*Cho màu đừng nhạt mắt
Tôi muốn buộc gió lại
Cho hương đừng bay đi”*

- Nhận diện câu qua dấu hai chấm, ngay sau đó có đánh số:

Ví dụ 20:

Người ta tổng kết có năm nguyên nhân bỏ học: (1) kinh tế gia đình khó khăn; (2) cha mẹ không quan tâm; (3) quản lý của nhà trường kém, chưa tập trung bồi dưỡng HS yếu; (4) phối hợp giữa nhà trường và gia đình chưa chặt chẽ, thường xuyên; (5) HS thiếu chuyên cần, học lực kém.

Tách thành:

Người ta tổng kết có năm nguyên nhân bỏ học:

- (1) kinh tế gia đình khó khăn;*
- (2) cha mẹ không quan tâm;*
- (3) quản lý của nhà trường kém, chưa tập trung bồi dưỡng HS yếu;*
- (4) phối hợp giữa nhà trường và gia đình chưa chặt chẽ, thường xuyên;*
- (5) HS thiếu chuyên cần, học lực kém.*

Như vậy, gặp trường hợp hai chấm (số 1, 2, 3...) chúng ta cần tách câu. Nếu không có dấu hai chấm, chỉ có (số 1,2,3...) thì chúng ta không tách.

Ví dụ 21:

Nói một cách khác theo ngôn từ ta thường dùng, cơ quan chủ quản ra lệnh cho công ty con (1) qua số vốn mình nắm và theo quyền biểu quyết đa số tương đối hay tuyệt đối trong ĐHCD của công ty con; và (2) có người đại diện của mình nắm đa số thành viên trong HĐQT.

Trường hợp này không tách.

*Tp.HCM ngày 15 tháng 1 năm 2008
TM Nhóm biên soạn VCL*

Đinh Điền