

Hướng dẫn gán nhãn từ loại

Nguyễn Phương Thái, Vũ Lương, Nguyễn Thị Minh Huyền, và nhóm dữ liệu

SP 7.3 – VLSP

Nhóm VTB lựa chọn tiêu chí *phân loại từ dựa trên khả năng kết hợp và chức vụ ngữ pháp của từ*. Chẳng hạn danh từ thì thường có chức vụ ngữ pháp là chủ ngữ hoặc bổ ngữ trong câu, thêm vào đó là khả năng kết hợp với số từ (hai, ba) và định từ (mỗi, mọi). Khi gán nhãn từ loại nhóm dữ liệu thường tham khảo từ điển tiếng Việt và các sách ngữ pháp. Tuy nhiên nếu không có một tài liệu hướng dẫn thì vẫn có trường hợp mọi người đưa ra các quyết định khác nhau cho cùng một tình huống. Tài liệu này được xây dựng trong quá trình gán nhãn từ loại cho ngữ liệu thô.

Nội dung

1. Bảng từ loại	2
2. Nhập những động từ - kết từ chính phụ	3
3. Nhập những động từ - trợ từ	3
4. Nhập những động từ - danh từ	3
5. Nhập những động từ - tính từ	3
6. Nhập những động từ - phó từ	3
7. Nhập những danh từ - kết từ chính phụ	4
8. Nhập những tính từ - danh từ	4
9. Nhập những kết từ đẳng lập – phụ từ	4
10. Các định từ ở vị trí -3[NTCẩn77, DQBan05]	4
11. Các cụm danh từ dẫn xuất (hình thái dẫn xuất)	4
12. Các cụm từ “sau khi”, “trong khi”, v.v.	4
13. Các từ “ra”, “vào”, “lên”, “xuống”, v.v.	5
14. Danh sách các từ chức năng không có trong từ điển nhưng lại có trong sách ngữ pháp [DQBan05]	5
15. Danh sách từ (hay tổ hợp từ) chức năng có trong từ điển nhưng lại không được phân loại	5
16. Nhóm từ chỉ vị trí: “trong”, “ngoài”, “trên”, “dưới”, v.v.	6
17. Từ bổ nghĩa cho số từ	6
18. Trợ từ “một cách”	6
Tài liệu tham khảo	6

1. Bảng từ loại

STT	Nhãn	Tên	Ví dụ
1	N	Danh từ	tiếng, nước, thủ đô, nhân dân, đồ đạc, cây cối, chim muông
2	Np	Danh từ riêng	Nguyễn Du, Việt Nam, Hải Phòng, Trường Đại học Bách khoa Hà Nội, Mộc tinh, Hoả tinh, Phật, Đạo Phật
3	Nc	Danh từ chỉ loại	con, cái, đứa, bức
4	Nu	Danh từ đơn vị ¹	mét, cân, giờ, năm, nhóm, hào, xu, đồng
5	V	Động từ	ngủ, ngồi, cười; đọc, viết, đá, đặt; thích, yêu, ghét, giống, muốn
6	A	Tính từ	tốt, xấu, đẹp; cao, thấp, rộng
7	P	Đại từ	tôi, chúng tôi, hắn, nó, y, đại nhân, đại ca, huynh, đệ
8	L	Định từ ²	mỗi, từng, mọi, cái; các, những, mấy
9	M	Số từ	một, mười, mười ba; dăm, vài, mười; nửa, rưỡi
10	R	Phó từ	đã, sẽ, đang, vừa, mới, từng, xong, rồi; rất, hơi, khi, quá
11	E	Giới từ ³ (kết từ chính phụ)	trên, dưới, trong, ngoài; của, trừ, ngoài, khỏi, ở
12	C	Liên từ (kết từ đẳng lập)	và, với, cùng, vì vậy, tuy nhiên, ngược lại
13	I	Thán từ	ôi, chao, a ha
14	T	Trợ từ, tình thái từ (tiểu từ) ⁴	à, a, á, à, ấy, chắc, chẳng, cho, chứ
15	B	Từ tiếng nước ngoài (hay từ vay mượn)	Internet, email, video, chat
16	Y	Từ viết tắt	OPEC, WTO, HIV
17	S	Yếu tố cấu tạo từ	bất, vô, gia, đa
18	X	Các từ không phân loại được	

¹ Mới bổ xung vào tháng 5/2008

² Trong tài liệu hướng dẫn cũ là D

³ Trong từ điển (SP7.2) là O

⁴ Trong từ điển (SP7.2) phân ra hai loại là trợ từ (T) và cảm từ (E)

Khi gán nhãn ngữ liệu, nhãn từ viết tắt sẽ là nhãn kép. Chẳng hạn nếu từ viết tắt là HIV thì nhãn của nó là Ny vì HIV viết đầy đủ thì là danh từ. Tương tự, nhãn từ vay mượn cũng là nhãn kép, ví dụ: email/Nb, chat/Vb

2. Nhập nhằng động từ - kết từ chính phụ

Trong câu sau “vào” là kết từ chính phụ (E) chứ không phải động từ.

Vào/E dịp/N nghỉ/V hè/N rất/R đông/A học sinh/N phổ thông/N theo/V học/V
./.

Hay “lên” trong ví dụ sau:

Rmah Rô/Np khoác/V gửi/N lên/E vai/N ./.

3. Nhập nhằng động từ - trợ từ

Bé đang tập đi/V .
trông sạch quá đi/T

4. Nhập nhằng động từ - danh từ

Hiện tượng nhập nhằng này khá phổ biến. Một số trường hợp là:

- Động từ chỉ hành động như “cày”, “cuốc”, “đục”, v.v. chuyển thành danh từ chỉ đồ vật tương ứng “(cái) cày”, “(cái) cuốc”, “(cái) đục”, v.v.

Tôi mượn *cuốc*/N để *cuốc*/V đám đất sau nhà.

- Động từ chỉ hành động như “suy nghĩ”, “đắn đo”, v.v. chuyển thành danh từ chỉ khái niệm hay sự vật trừu tượng “(những) suy nghĩ”, “(những) đắn đo”, v.v.

Vài *suy nghĩ*/N về Toán học Việt Nam
Hôm nay, đọc lại hồi kí Hồi ức và *suy nghĩ*/N của Trần Quang Cơ trên tinh thần “ôn cố tri tân”, bạn đọc Thông Luận có thể nhận ra nhiều điều thú vị.

5. Nhập nhằng động từ - tính từ

Đồng hồ này *chạy*/V rất chính xác .
Hàng bà ấy dạo này bán không *chạy*/A .

6. Nhập nhằng động từ - phó từ

Nhằm “được” là động từ trong trường hợp sau:

Mình/P không/R nuôi/V chúng nó/P *được*/R ./.

Từ này có độ nhập nhằng cao vì nó vừa có thể là động từ, vừa có thể là tính từ, phó từ, và trợ từ. Trong ví dụ trên “được” là phụ từ vì nó biểu thị điều vừa nói đến (động từ “nuôi”) là có khả năng thực hiện.

7. Nhập nhằng danh từ - kết từ chính phụ

Nhà ấy nhiều *của/N* lắm !
sách *của/E* thư viện

8. Nhập nhằng tính từ - danh từ

Từ chỉ tính chất như “khó khăn”, “gian khổ”, v.v. chuyển thành danh từ chỉ sự vật trừu tượng “(những) khó khăn”, “(những) gian khổ”, v.v.

Tuy nhiên, sức chống đỡ trước *khó khăn/N* của Việt Nam đã tốt lên rất nhiều.
Tăng giá sách, *khó khăn/N* sẽ thêm chồng chất

9. Nhập nhằng kết từ đẳng lập – phụ từ

Cháu Ngọc rất bé *mà/C* rất khỏe .
Mẹ đã bảo *mà/T* !

10. Các định từ ở vị trí -3[NTCần77, DQBan05]

Danh sách : “tất cả”, “tất thầy”, “toàn bộ”, “hầu hết”, “phần lớn”
tất thầy/L học sinh trường này, *hầu hết/L* những giáo sư này

11. Các cụm danh từ dẫn xuất⁵ (hình thái dẫn xuất)

Có nhiều cụm danh từ có cấu tạo gồm động từ hoặc tính từ đi sau danh từ chỉ loại. Ví dụ như: “sự nóng hổi”, “vụ hồi lộ”, v.v. Các từ này phải được gán nhãn là động từ hay tính từ.

sự/Nc nóng hổi/A; *vụ/Nc hồi lộ/V*

Trường hợp của một số cụm từ như: “những vất vả”, “những thành công”, v.v. Thì ta gán từ loại danh từ cho “vất vả”, “thành công”, v.v.

12. Các cụm từ “sau khi”, “trong khi”, v.v.

Các cụm từ này gồm hai từ, một là kết từ chính phụ (chẳng hạn “sau”), từ còn lại là danh từ thời gian (chẳng hạn “khi”).

sau/E khi/N; *trong/E khi/N*; *trước/E khi/N*

Ngoài ra còn có: “đến khi”, “tới khi”, “trước lúc”, “trong lúc”, “đến lúc”, “tới lúc”, v.v.; “trong đó” cũng có thể được xếp vào nhóm này (*trong/E đó/P*).

⁵ Trong tiếng Anh hậu tố dẫn xuất làm thay đổi từ loại (*happy/A* → *happiness/N*), trong tiếng Việt thì lại sử dụng từ chức năng.

13. Nhóm từ “ra”, “vào”, “lên”, “xuống”

Danh sách : ra, vào, lên, xuống, đến, tới, sang, qua, lại, về

Các từ này có thể được gán nhãn là động từ, phó từ, và kết từ chính phụ (giới từ) tùy thuộc vào ngữ cảnh:

- Khi không có vị từ là thực từ (như “đi”, “hiểu”, “béo”) đứng trước, các từ này có tư cách động từ:

Tôi/P ra/V sân/N. Tôi/P vào/V lớp học/N.

- Khi có vị từ là thực từ đứng trước, các từ này làm phó từ:

đi/V ra/R, bước/V xuống/R, đẩy/V xe/N ra/R, kéo/V xe/N lên/R

- Khi có động từ đằng trước⁶ và danh từ (thường là chỉ vị trí) đằng sau hoặc khi ở đầu câu, các từ này làm kết từ chính phụ:

đi/V xuống/E Hải Phòng/Np, đi/V lên/E Kontum/Np, bê/V bàn/N ra/E vườn/N

- Khi động từ đứng trước có mẫu NP-VP thì các từ này là động từ :

bảo/V nó/P xuống/V đây/P

14. Danh sách các từ chức năng không có trong từ điển nhưng lại có trong sách ngữ pháp [DQBan05]

Kết từ chính phụ (giới từ): “cho đến”, “cho tới”, “để mà”, “để cho”, “cùng với”

Kết từ đẳng lập: “cũng như”

15. Danh sách từ (hay tổ hợp từ) chức năng có trong từ điển nhưng lại không được phân loại

Kết từ chính phụ: “đến nỗi”

khỏe/A đến nỗi/E có thể/V dùng/V hai/M tay/N nắm/V hai/M sừng/N trâu/N ghi/V xuống/E đất/N

Phó từ: “đúng ra”⁷, “lẽ ra”, “mới đây”, “vừa rồi”, “vừa qua”; “hình như”, “dường như”; “ít nhất”

đúng ra/R ,/, nó/P phải/V bị/V kỷ luật/V

Các từ này có đặc điểm là có thể xuất hiện ở đầu câu hoặc giữa chủ ngữ và vị ngữ.

Kết từ đẳng lập: “chẳng hạn”; “ngoài ra”; “mặt khác”; “tức”, “tức là”; “nghĩa là”; “ngược lại”; “nói chung”, “nhìn chung”; “nói riêng”; “nói tóm lại”

Mặt khác/C, thông thường các sản phẩm tẩy trang đều chứa thành phần dầu.

⁶ Chú ý là giữa từ ngữ cảnh (động từ, tính từ) và từ đang xét phải có quan hệ ngữ pháp

⁷ [DQBan05] phân loại từ này là trợ từ

Một số từ khác: “do đó”, “do vậy”, “tuy vậy”, “dẫu vậy”

Dẫu vậy/C, cả hai ông và bất kỳ người nào khác đều không thể thống nhất về những phẩm chất cần có, chính xác của một ông chủ Nhà Trắng.

16. Nhóm từ chỉ vị trí: “trong”, “ngoài”, “trên”, “dưới”, v.v.

Có một số trường hợp ta không làm theo hướng dẫn trong từ điển. Nhóm danh từ này thuộc một trong các số ít đó. Xét từ “trên” như một ví dụ.

(1) Danh từ theo từ điển → giới từ:

Máy bay/N lượn/V *trên*/E thành phố/N ./.
Nhà/N anh/P ấy/P ở/V *trên*/E tầng/N năm/M ./.

(2) Danh từ theo từ điển → tính từ

hàng/N ghế/N *trên*/A ./.
đọc/V lại/R mấy/L trang/N *trên*/A ./.

Có hai lý do để ta qui định lại:

- Hành vi của từ “trên” chả khác gì giới từ trong trường hợp (1) và tính từ trong trường hợp (2)
- Qui định lại sẽ giúp việc gán nhãn đơn giản hơn (nhất quán hơn)

17. Từ bổ nghĩa cho số từ

Danh sách từ: “hơn”, “trên”, “dưới”, “gần”, “khoảng”, v.v.

hơn/A 200/M đại biểu/N, trên/A mười/M suất/N học bổng/N, khoảng/A 50/M bảng/Nu

Yêu cầu: Phân loại cho các từ này là tính từ (A).

18. Trợ từ “một cách”

Nước/N xã/V Miwon/Np bỗng nhiên/R trong/A *một cách*/T kỳ lạ/A !/.

Trợ từ này có tác dụng nhấn mạnh. Nó luôn đứng ở trước bộ phận cần được nhấn mạnh [LBiên99].

19. Mẫu “từ .. sang ..”

Đó/P là/V hải_trình/N lớn/A nhất/R từ/E tây/N sang/E đông/N

“từ” và “sang” cần được gán nhãn là E. Đây là các thành phần của cụm giới từ chỉ hướng.

Tài liệu tham khảo

[DQBan05] Diệp Quang Ban. 2005. Ngữ pháp tiếng Việt (2 tập). NXB Giáo dục.

[LBiên99] Lê Biên. 1999. Từ loại tiếng Việt hiện đại. NXB Giáo dục.

[NTCần77] Nguyễn Tài Cẩn. 1977. Ngữ pháp tiếng Việt: tiếng – từ ghép – đoản ngữ. NXB ĐH & THCN

[UBKHXH83] Ủy ban Khoa học Xã hội Việt Nam. 1983. Ngữ pháp tiếng Việt. NXB Khoa học Xã hội.

[Vietlex08] Trung tâm Từ điển học. 2008. Từ điển tiếng Việt. NXB Đà Nẵng.