

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



XỬ LÝ NGÔN NGỮ TỰ NHIÊN
XÂY DỰNG TRÌNH PHÂN TÍCH CÚ PHÁP TIẾNG VIỆT

Lớp: CS221.K11

Giáo viên hướng dẫn: ThS. Nguyễn Trọng Chính

Sinh viên thực hiện:

BÙI NGUYỄN MÃO 16520724

NGUYỄN ĐỨC LỢI 16521722

TP. Hồ Chí Minh, ngày 10 tháng 1 năm 2020

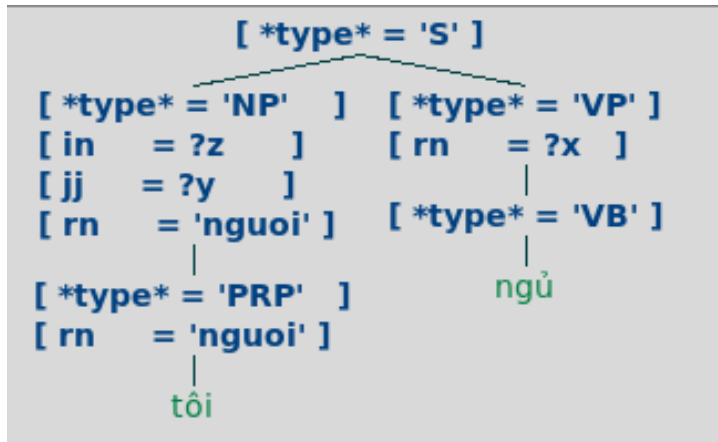
MỤC LỤC

1	<u>PHÂN TÍCH 30 CÂU TIẾNG VIỆT</u>	3
1.1	CHỦ NGỮ (DANH TỪ) + VỊ NGỮ (ĐỘNG TỪ NỘI ĐỘNG):	3
1.2	CHỦ NGỮ (DANH TỪ + TÍNH TỪ) + VỊ NGỮ (ĐỘNG TỪ NGOẠI ĐỘNG + DANH TỪ)	5
1.3	CHỦ NGỮ (DANH TỪ + GIỚI TỪ) + VỊ NGỮ (ĐỘNG TỪ)	7
1.4	CHỦ NGỮ (DANH TỪ) + VỊ NGỮ (ĐỘNG TỪ + GIỚI TỪ)	9
1.5	CHỦ NGỮ (CHỨA LIÊN TỪ) + VỊ NGỮ (TÙY CHỌN ĐỘNG TỪ, TÍNH TỪ, GIỚI TỪ)	12
1.6	CÓ 2 ĐỘNG TỪ NHƯNG KHÔNG CÓ LIÊN TỪ Ở GIỮA	14
2	<u>TỔNG HỢP LUẬT SẢN SINH</u>	17
3	<u>XỬ LÝ RÀNG BUỘC NGỮ NGHĨA</u>	18
3.1	DANH TỪ VÀ ĐỘNG TỪ	18
3.2	DANH TỪ VÀ GIỚI TỪ	19
3.3	DANH TỪ VÀ TÍNH TỪ	20
4	<u>PHÂN TÍCH 20 CÂU TỪ BỘ TỪ VỰNG VÀ ĐÁNH GIÁ</u>	20
4.1	PHÂN TÍCH 20 CÂU	20
4.2	KẾT LUẬN	27

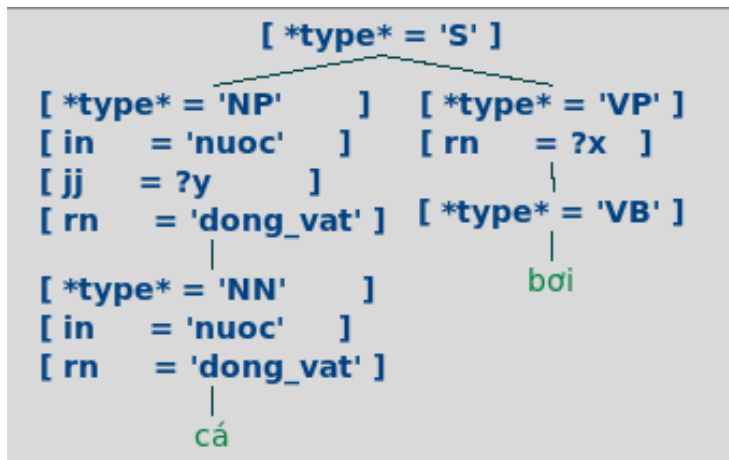
1 Phân tích 30 câu tiếng Việt

1.1 Chủ ngữ (danh từ) + vị ngữ (động từ nội động):

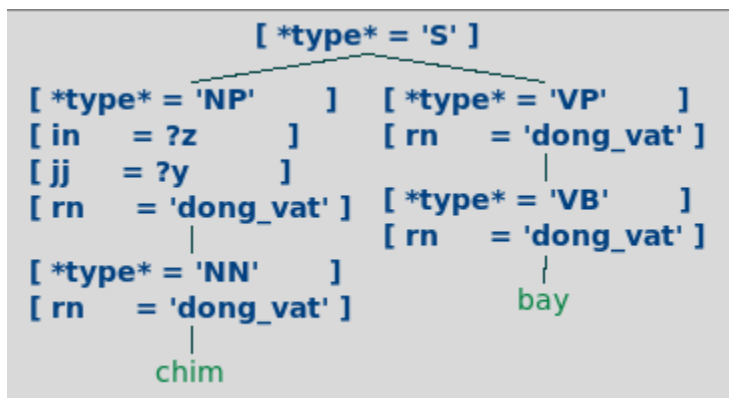
1.1.1 Tôi ngủ.



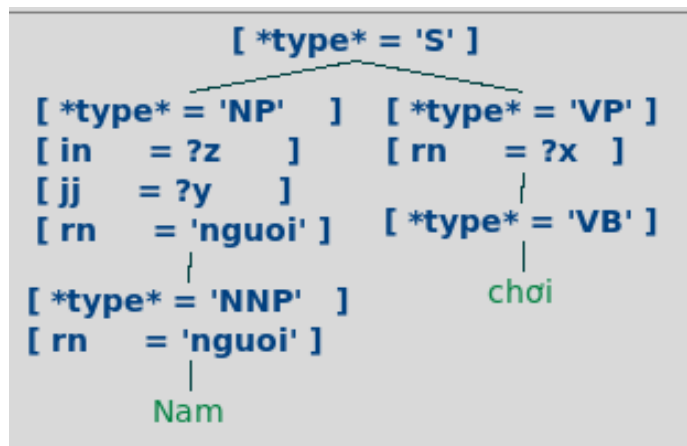
1.1.2 Cá bơi.



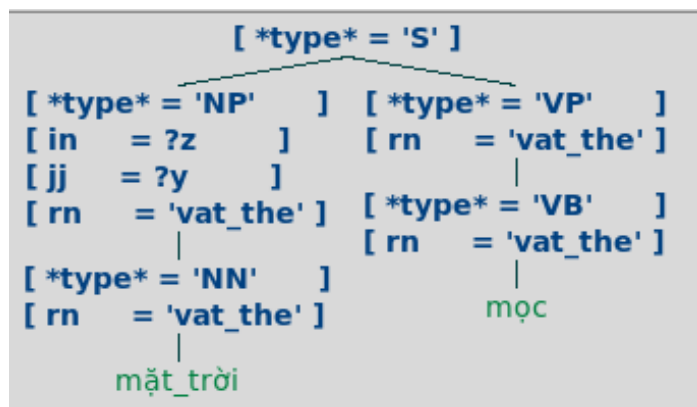
1.1.3 Chim bay.



1.1.4 Nam chơi.



1.1.5 Mặt trời mọc.



❖ Tập luật sản sinh:

$S \rightarrow NP VP$

$NP \rightarrow NN$

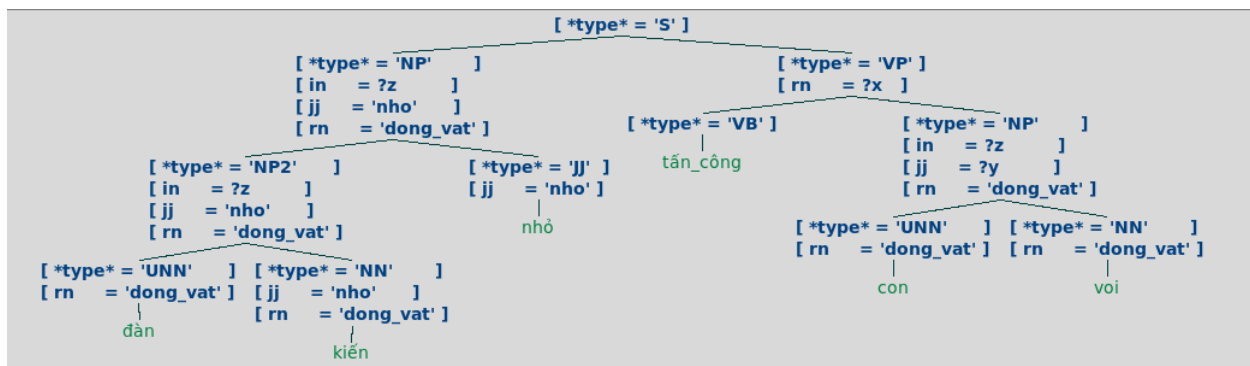
$NP \rightarrow PRP$

$NP \rightarrow NNP$

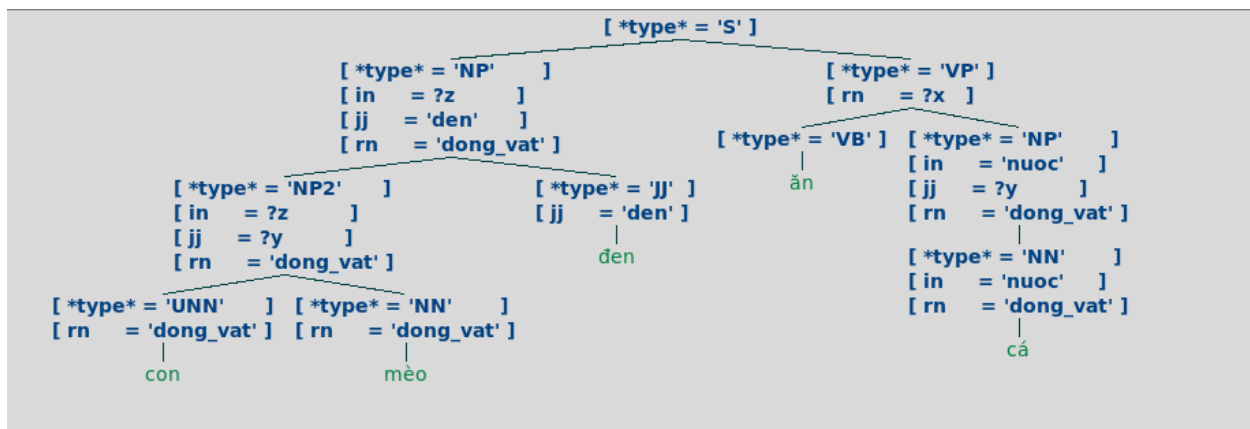
$VP \rightarrow VB$

1.2 Chủ ngữ (danh từ + tính từ) + vị ngữ (động từ ngoại động + danh từ)

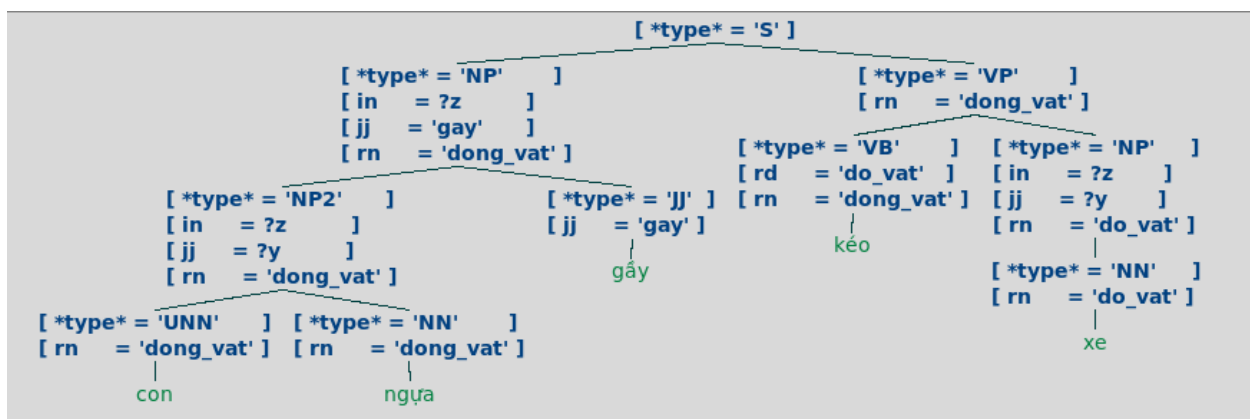
1.2.1 Đàn kiến nhỏ tấn công con voi.



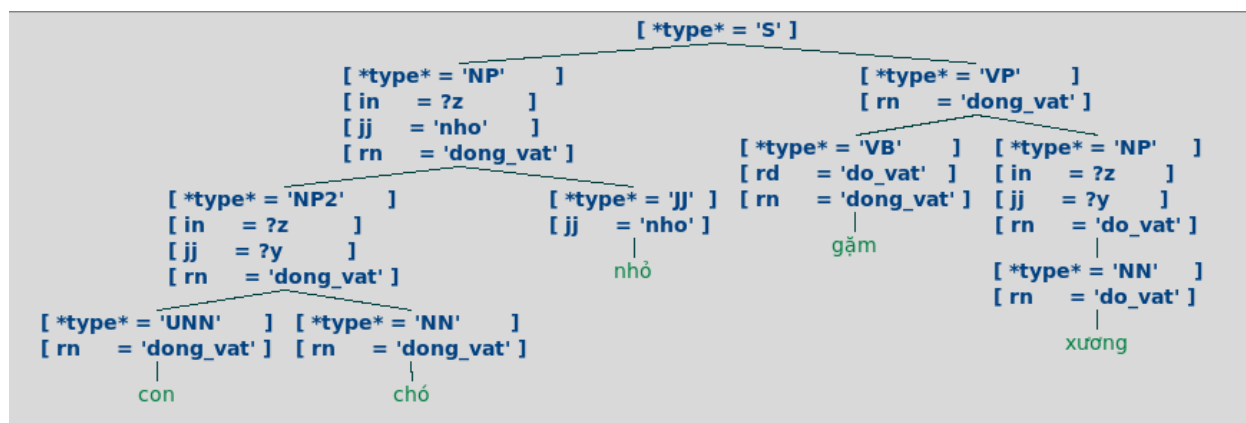
1.2.2 Con mèo đen ăn cá.



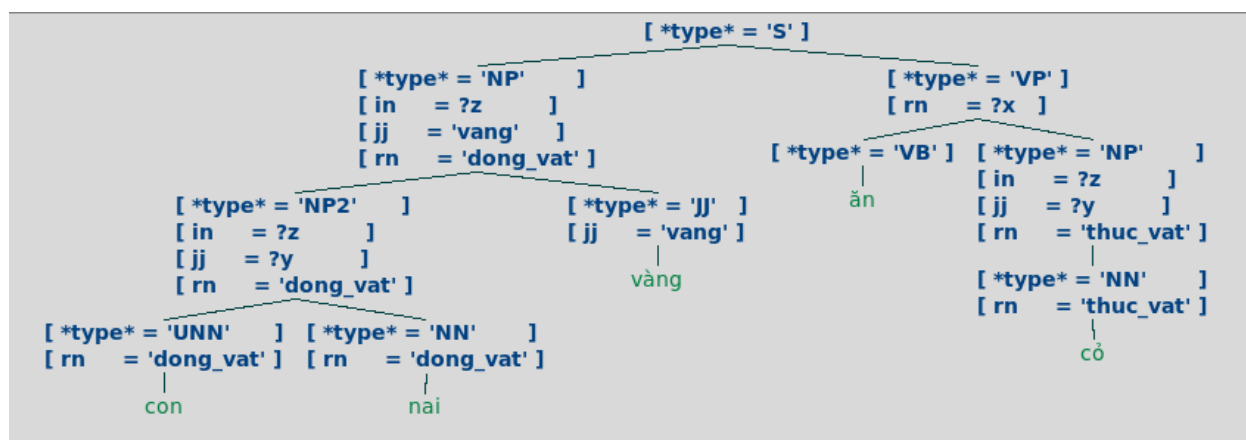
1.2.3 Con ngựa gầy kéo xe.



1.2.4 Con chó nhỏ gặm xương.



1.2.5 Con nai vàng ăn cỏ.



❖ Tập luật sản sinh:

$S \rightarrow NP VP$

$NP \rightarrow NN$

$NP \rightarrow PRP$

$NP \rightarrow NNP$

$NP \rightarrow UNN NN$

$NP \rightarrow NP1 JJ$

$NP1 \rightarrow NN$

$NP1 \rightarrow PRP$

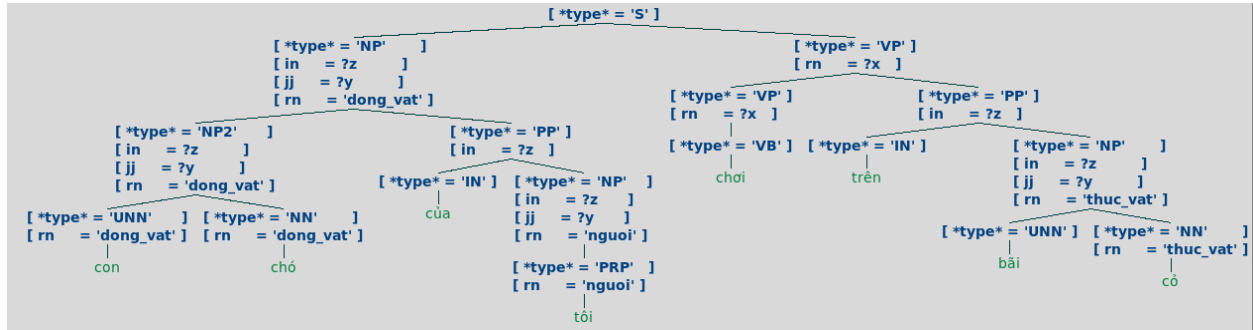
NP1 → NNP

NP1 → UNN NN

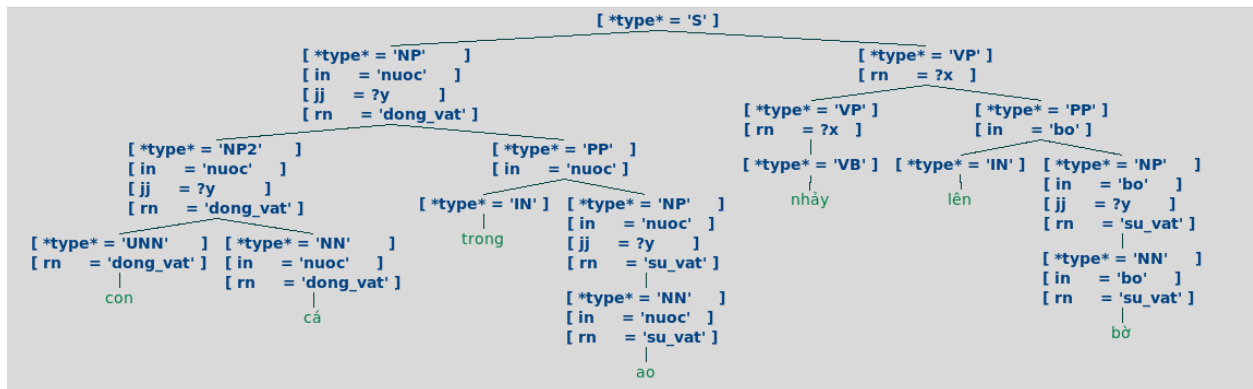
VP → VB NP

1.3 Chủ ngữ (danh từ + giới từ) + vị ngữ (động từ)

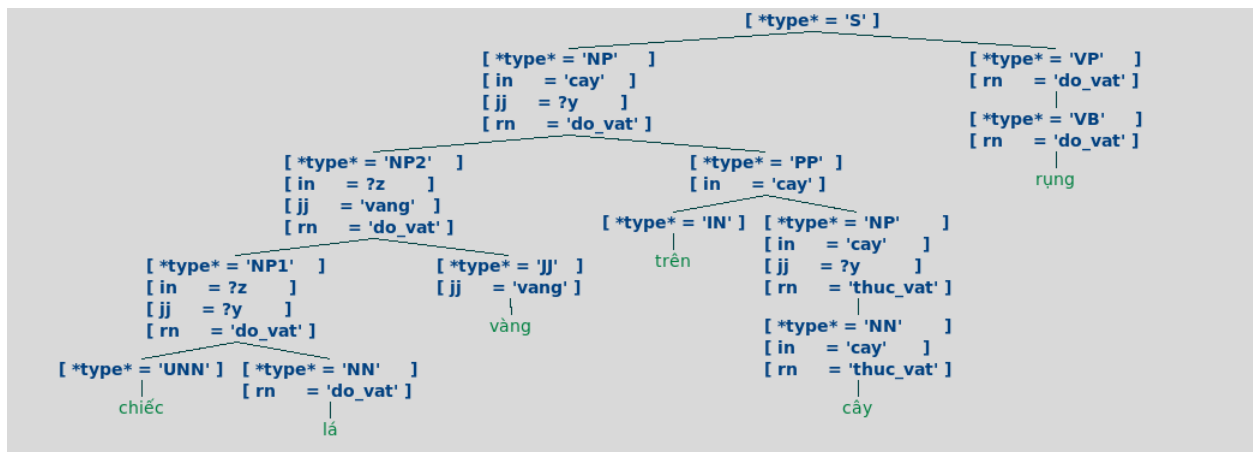
1.3.1 Con chó của tôi chơi đùa trên bãi cỏ



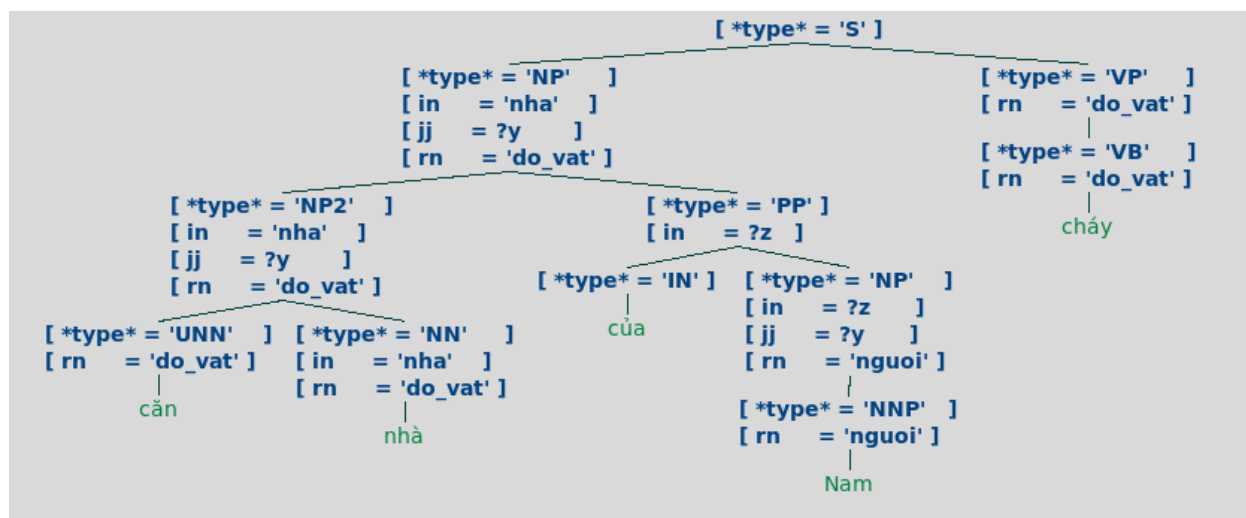
1.3.2 Chú cá trong ao nhảy lên bờ.



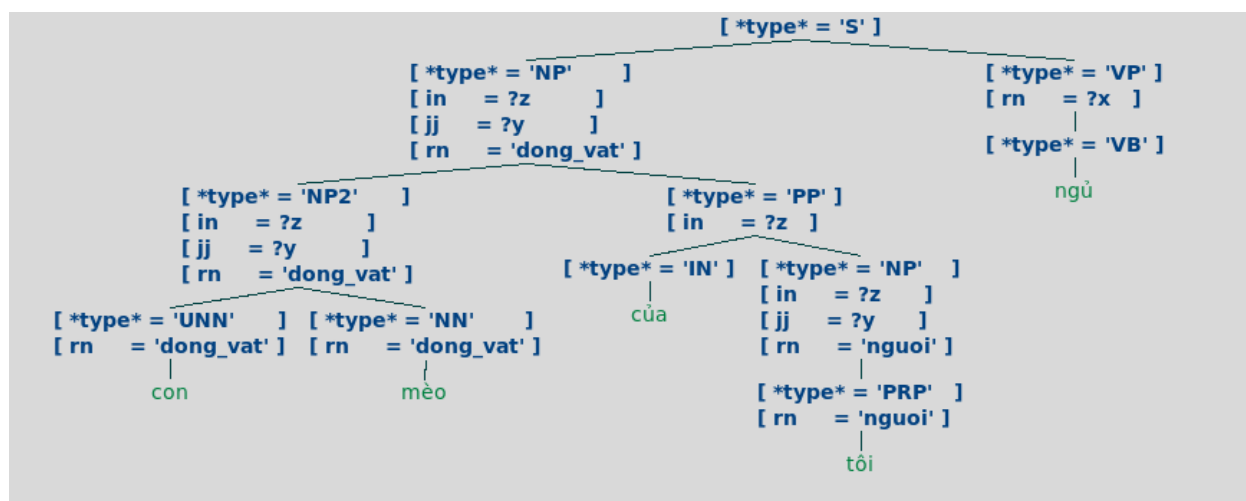
1.3.3 Chiếc lá vàng trên cây rụng.



1.3.4 Căn nhà của Nam cháy



1.3.5 Con mèo của tôi ngủ



❖ Tập luật sản sinh

$S \rightarrow NP VP$

$NP \rightarrow NN$

$NP \rightarrow PRP$

$NP \rightarrow NNP$

$NP \rightarrow UNN NN$

$NP \rightarrow NP2 PP$

NP2 → NP1 JJ

NP2 → NN

NP2 → PRP

NP2 → NNP

NP2 → UNN NN

NP1 → NN

NP1 → PRP

NP1 → NNP

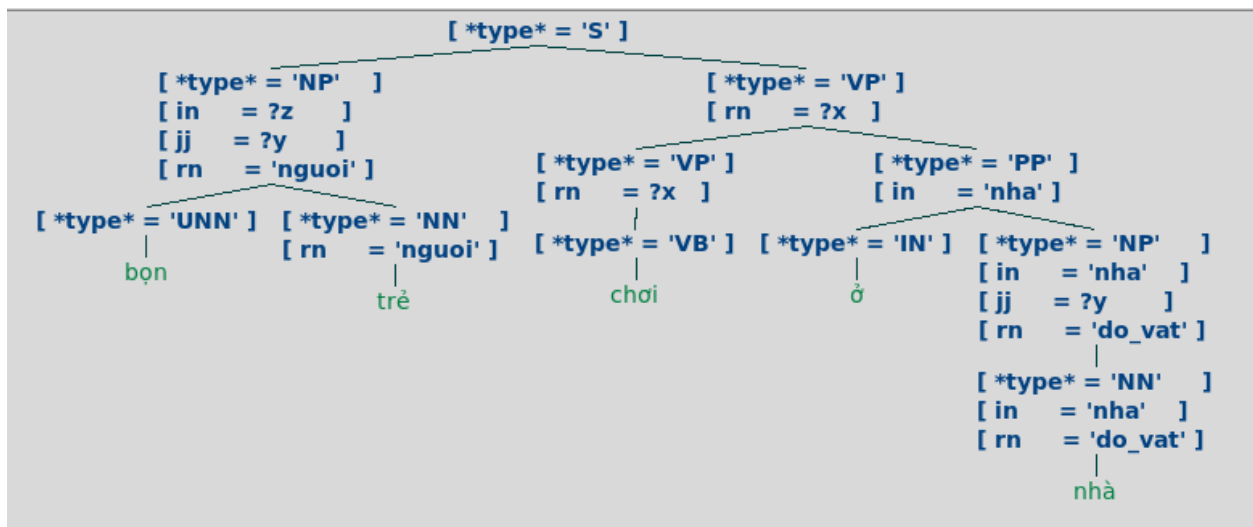
NP1 → UNN NN

PP → IN NP

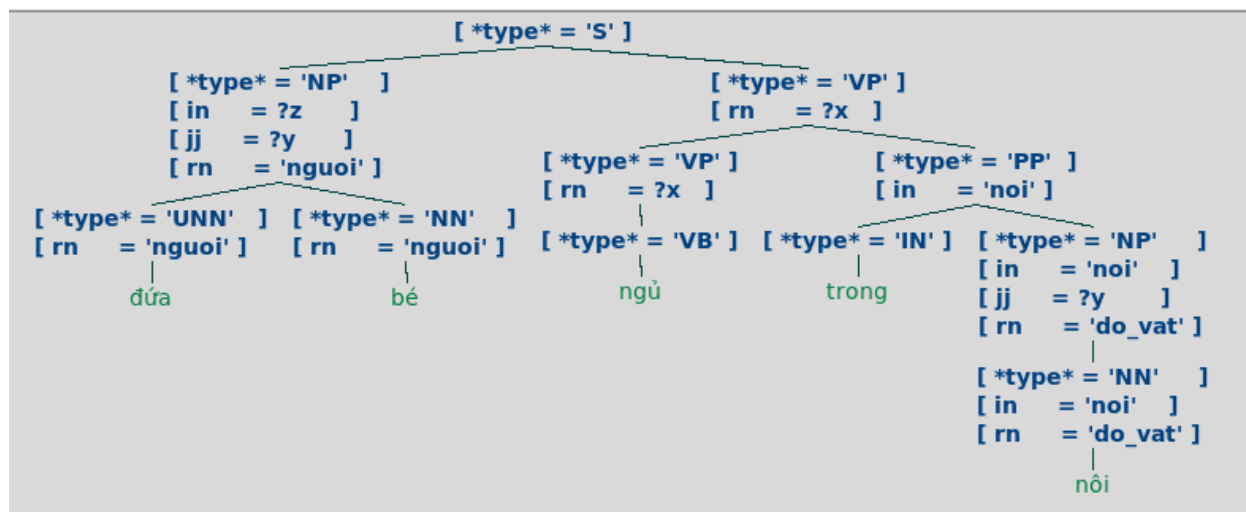
VP → VB NP

1.4 Chủ ngữ (danh từ) + vị ngữ (động từ + giới từ)

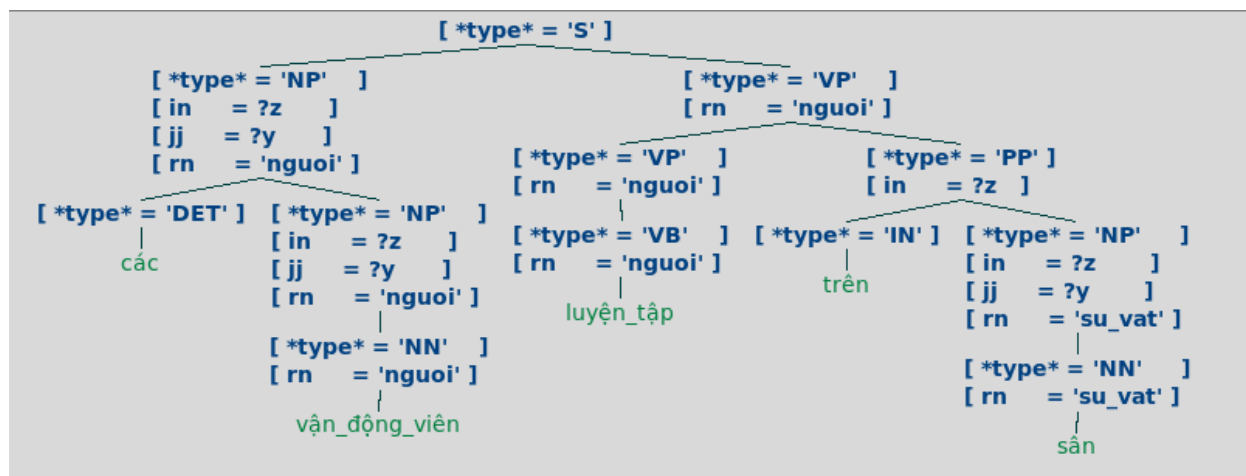
1.4.1 Bọn trẻ chơi ở nhà



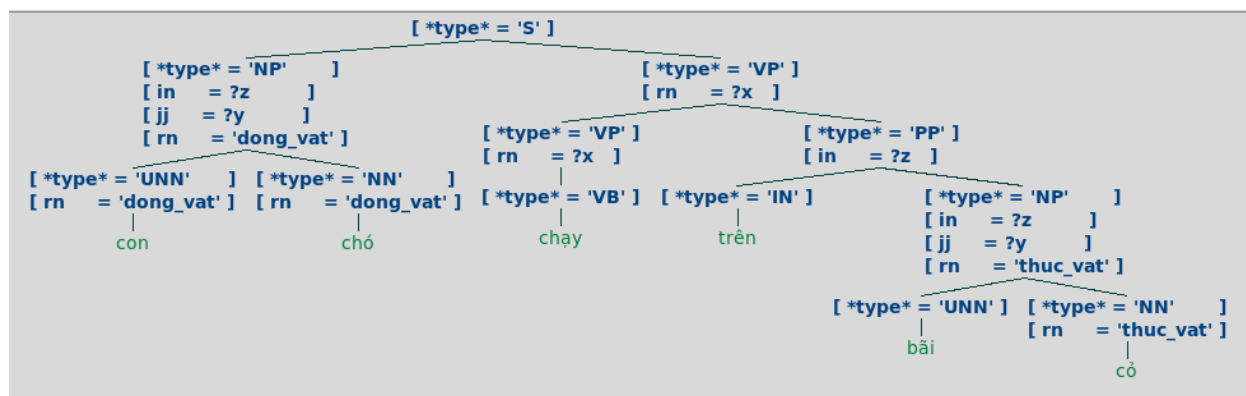
1.4.2 Đưa bé ngủ trong nôi



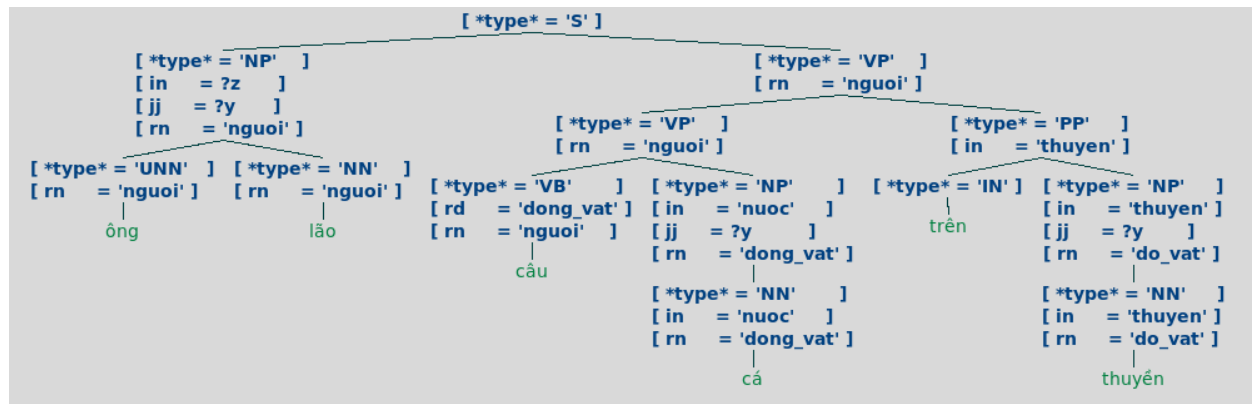
1.4.3 Các vận động viên luyện tập trên sân



1.4.4 Con chó chạy trên bãi cỏ



1.4.5 Ông lão câu cá trên thuyền



❖ Tập luật sản sinh

$S \rightarrow NP VP$

$NP \rightarrow NN$

$NP \rightarrow PRP$

$NP \rightarrow NNP$

$NP \rightarrow UNN NN$

$NP \rightarrow DET NP$

$VP \rightarrow VB$

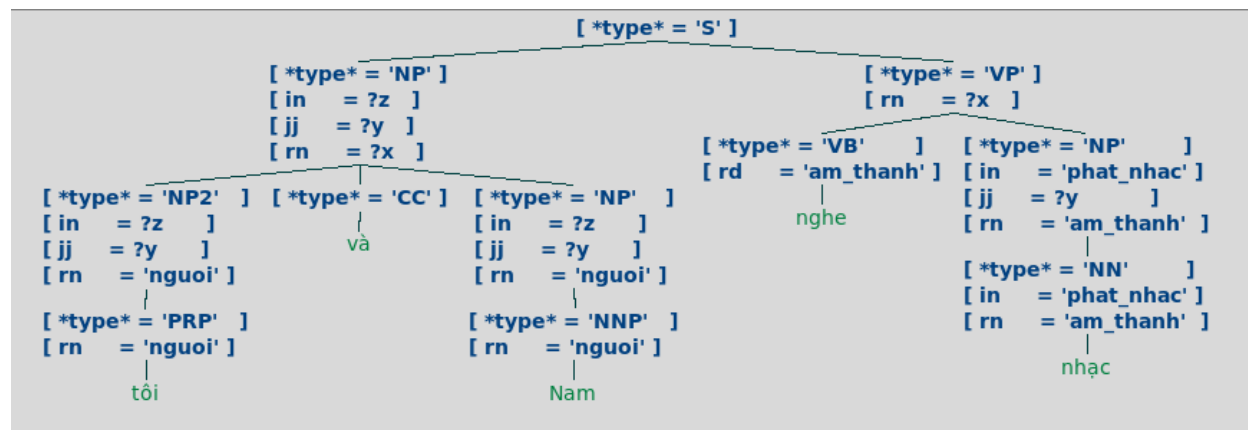
$VP \rightarrow VB NP$

$VP \rightarrow VP PP$

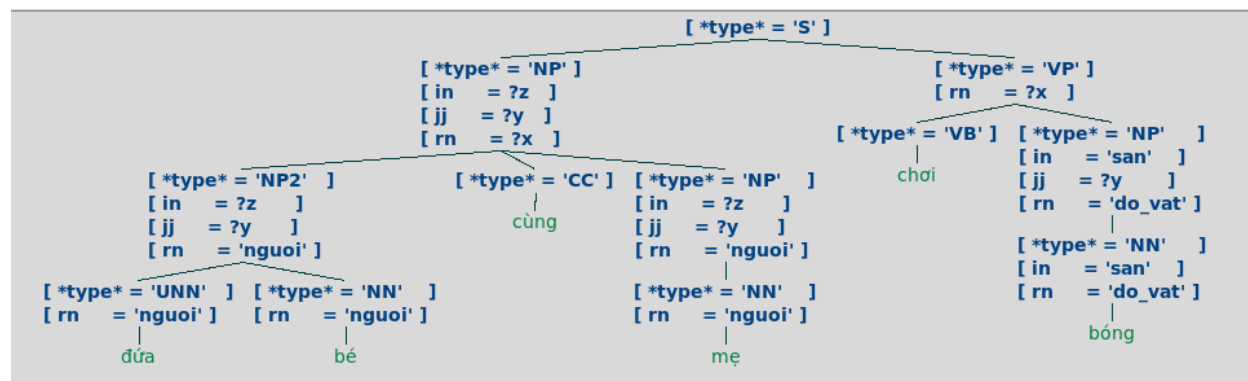
$PP \rightarrow IN NP$

1.5 Chủ ngữ (chứa liên từ) + vị ngữ (tùy chọn động từ, tính từ, giới từ)

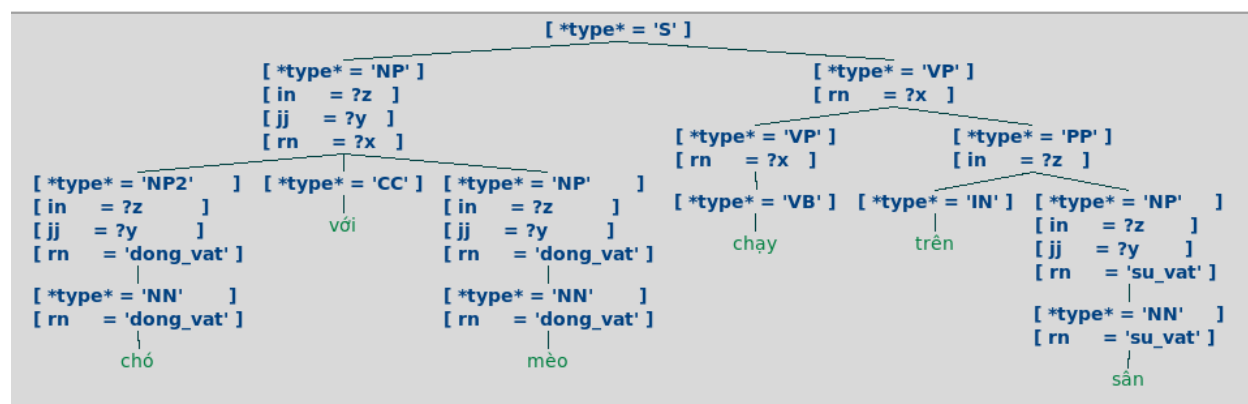
1.5.1 Tôi và Nam nghe nhạc



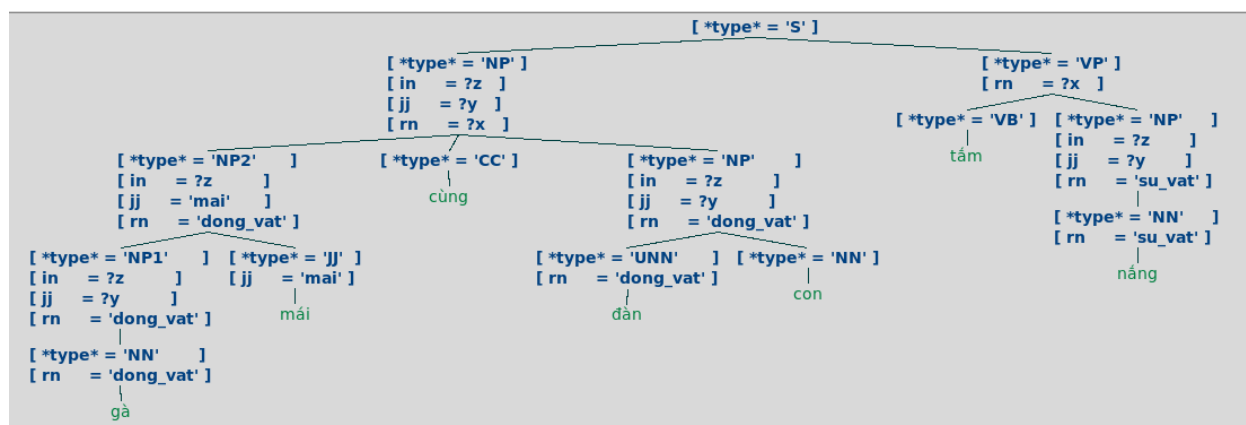
1.5.2 Đứa bé cùng mẹ chơi bóng



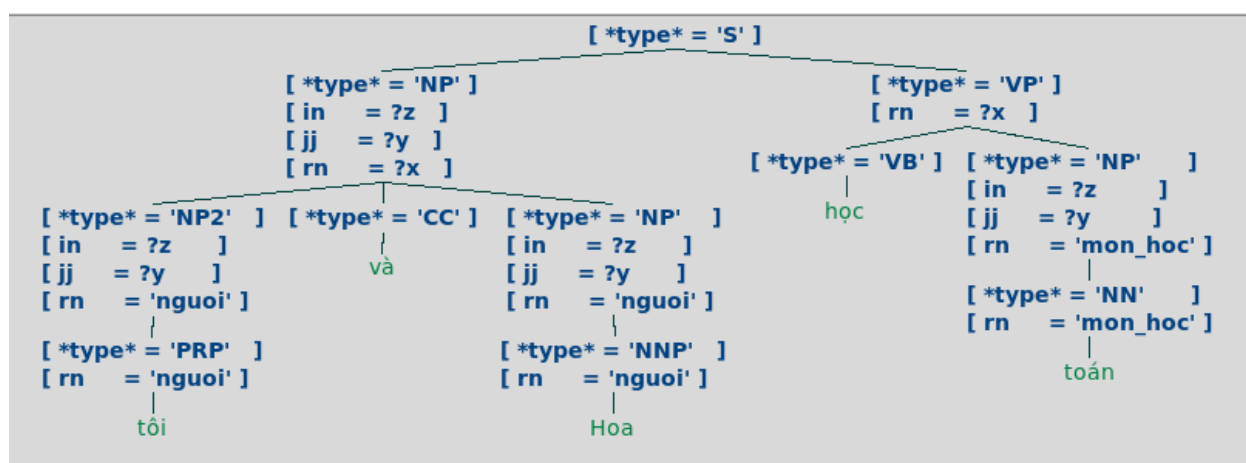
1.5.3 Chó với mèo chạy trên sân



1.5.4 Gà mái cùng đàn con tắm nắng



1.5.5 Tôi và Hoa học toán



❖ Tập luật sản sinh

$S \rightarrow NP VP$

$NP \rightarrow NN$

$NP \rightarrow PRP$

$NP \rightarrow NNP$

$NP \rightarrow UNN NN$

$NP \rightarrow DET NP$

$PP \rightarrow IN NP$

$NP \rightarrow NP2 CC NP$

NP2 → NP1 JJ

NP2 → NN

NP2 → PRP

NP2 → NNP

NP2 → UNN NN

NP1 → NN

NP1 → PRP

NP1 → NNP

NP1 → UNN NN

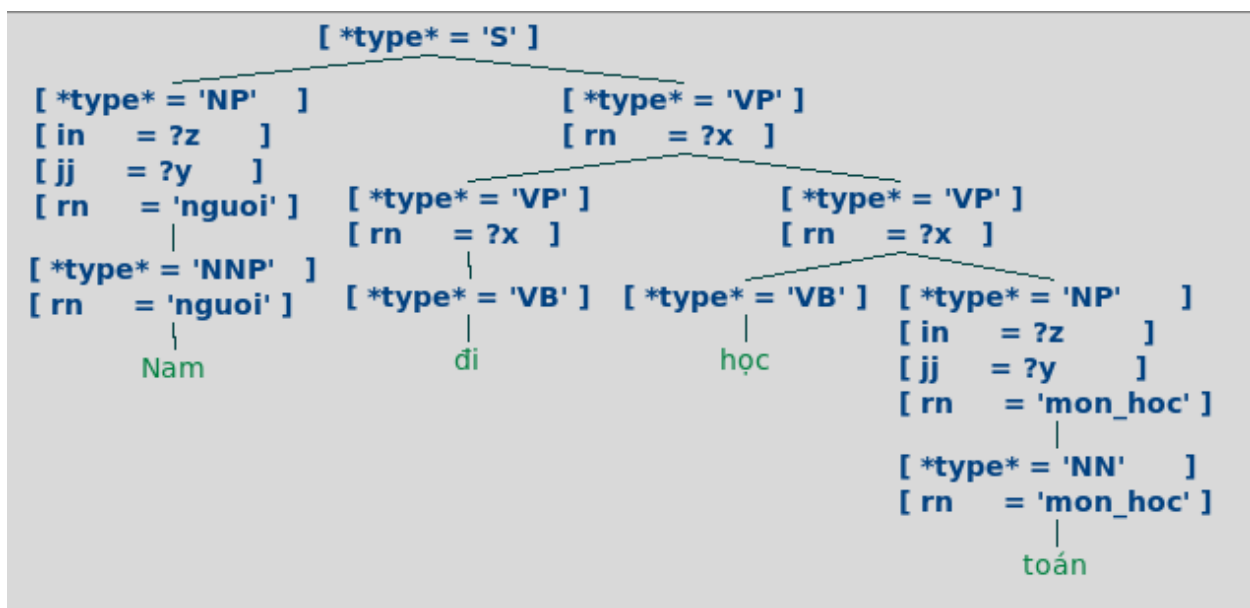
VP → VB

VP → VB NP

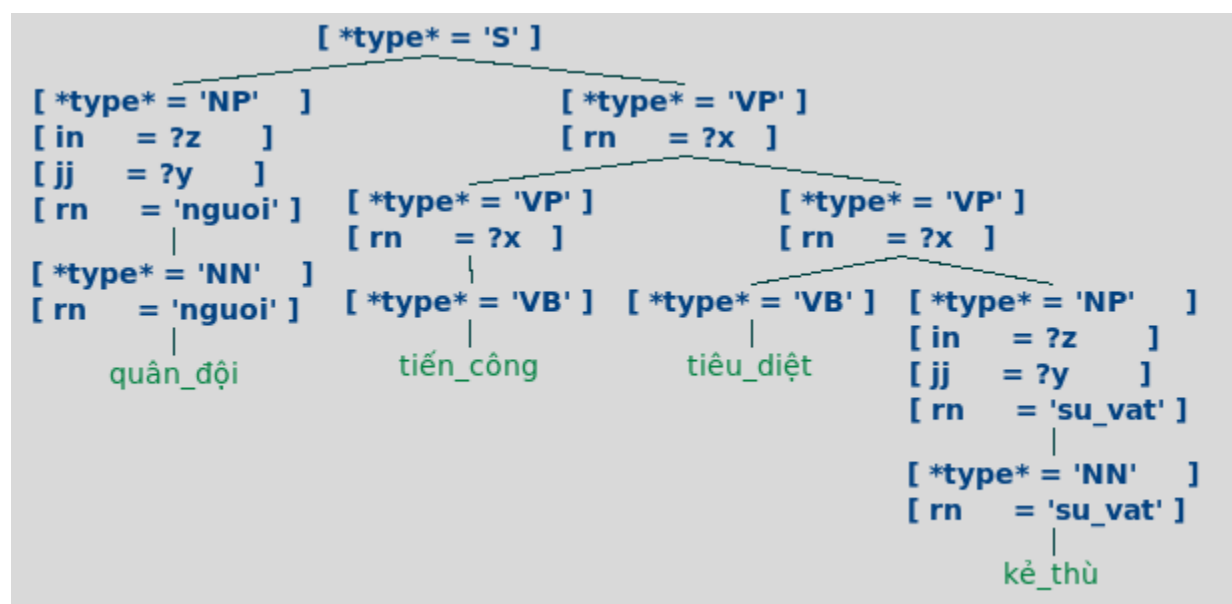
VP → VP PP

1.6 Có 2 động từ nhưng không có liên từ ở giữa

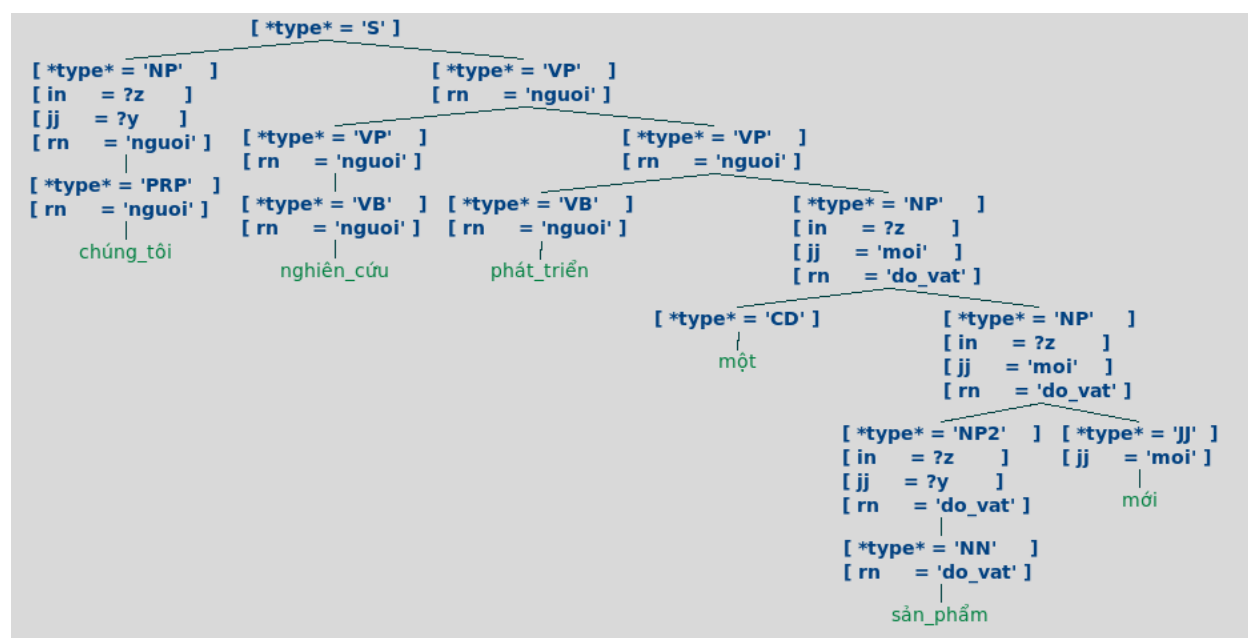
1.6.1 Nam đi học



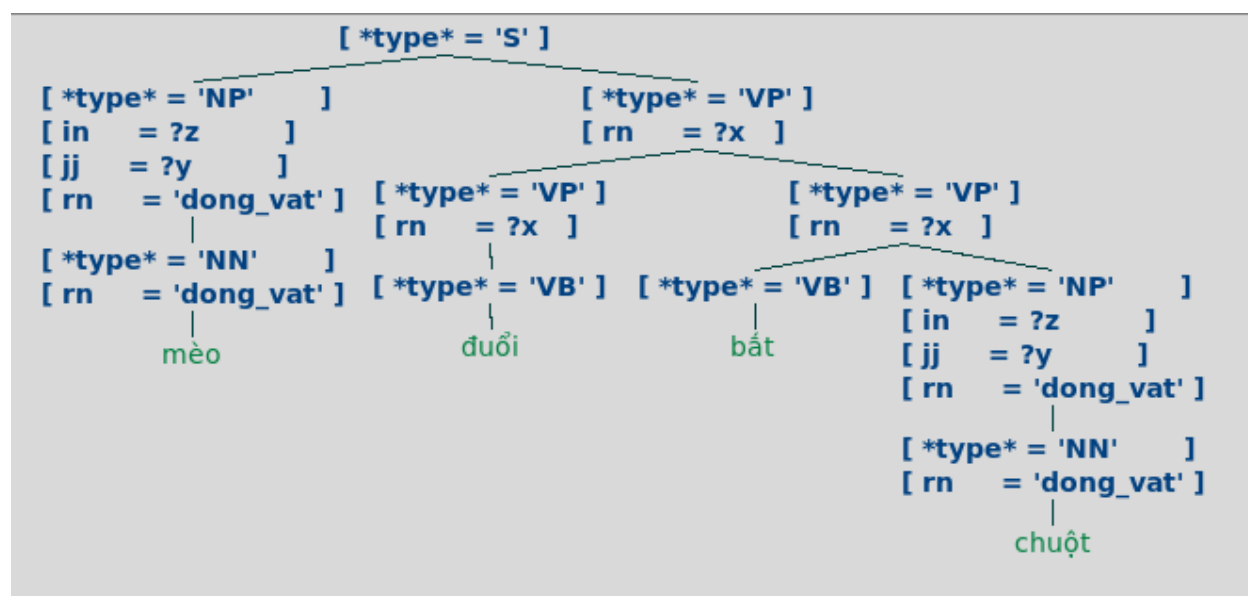
1.6.2 Quân đội tiến công tiêu diệt kẻ thù



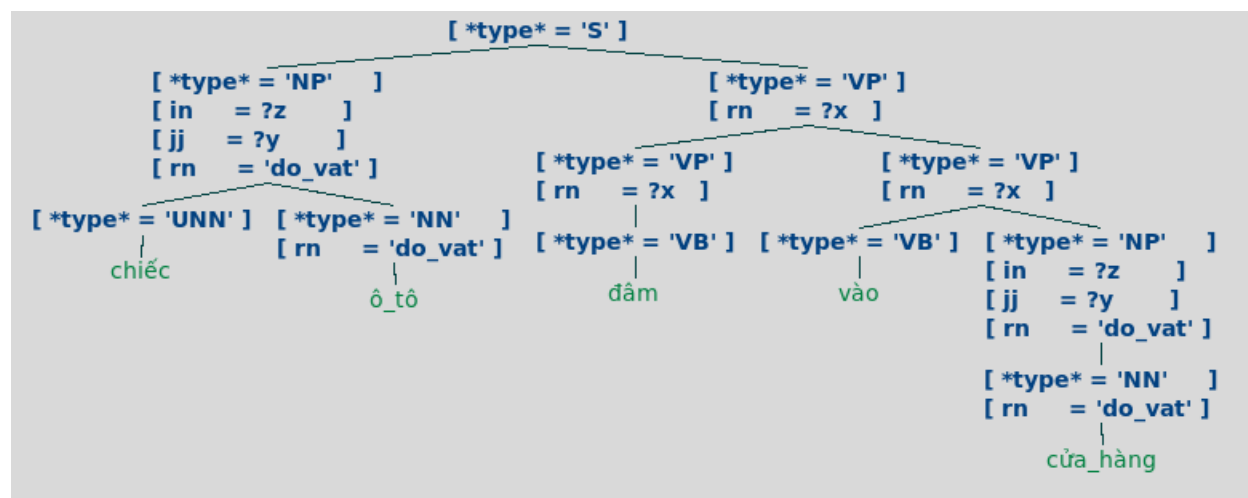
1.6.3 Chúng tôi nghiên cứu phát triển một sản phẩm mới



1.6.4 Mèo đuổi bắt chuột



1.6.5 Chiếc ô tô đâm vào cửa hàng



❖ Tập luật sản sinh

$S \rightarrow NP VP$

$NP \rightarrow NN$

$NP \rightarrow PRP$

$NP \rightarrow NNP$

$NP \rightarrow UNN NN$

$NP \rightarrow DET\ NP$

$VP \rightarrow VB$

$VP \rightarrow VB\ NP$

$VP \rightarrow VP\ PP$

$VP \rightarrow VP\ VP$

$PP \rightarrow IN\ NP$

2 Tổng hợp luật sản sinh

Từ 30 câu trên, ta rút ra được tập các luật sản sinh như sau:

$S \rightarrow NP\ VP$

$NP \rightarrow NN$

$NP \rightarrow NNP$

$NP \rightarrow PRP$

$NP \rightarrow UNN\ NN$

$NP \rightarrow DET\ NP$

$NP \rightarrow CD\ NP$

$NP \rightarrow NP_2\ CC\ NP$

$NP \rightarrow NP_2\ PP$

$NP \rightarrow NP_2\ JJ$

$NP_2 \rightarrow NP_1\ PP$

$NP_2 \rightarrow NP_1\ JJ$

$NP_2 \rightarrow NN$

$NP_2 \rightarrow NNP$

$NP_2 \rightarrow PRP$

NP2 → UNN NN

NP2 → DET NP

NP1 → NN

NP1 → NNP

NP1 → PRP

NP1 → UNN NN

NP1 → DET NP

VP → VB

VP → VP VP

VP → VB NP

VP → VP PP

PP → IN NP

3 Xử lý ràng buộc ngữ nghĩa

3.1 Danh từ và động từ

PRP[rn=nguoi] -> "tôi"

PRP[rn=nguoi] -> "chúng_tôi"

NNP[rn=nguoi] -> "Nam"

NNP[rn=nguoi] -> "Hoa"

UNN[rn=dong_vat] -> "con"

UNN[rn=dong_vat] -> "đàn"

UNN[rn=do_vat] -> "căn"

UNN[rn=nguoi] -> "ông"

UNN[rn=nguoi] -> "đứa"

NN[rn=mon_hoc] -> "toán"

NN[rn=dong_vat] -> "cá"

NN[rn=nguoi] -> "trẻ"

NN[rn=nguoi] -> "bé"

NN[rn=nguoi] -> "lão"

NN[rn=nguoi] -> "mẹ"

NN[rn=nguoi] -> "vận_động_viên"

NN[rn=dong_vat] -> "chim"

NN[rn=dong_vat] -> "voi"
 NN[rn=dong_vat] -> "gà"
 NN[rn=vat_the] -> "mặt_trời"
 NN[rn=thuc_vat] -> "cỏ"
 NN[rn=thuc_vat, in=cay] -> "cây"
 NN[rn=do_vat] -> "xe"
 NN[rn=do_vat] -> "xương"
 NN[rn=do_vat] -> "bóng"
 NN[rn=am_thanh] -> "nhạc"
 NN[rn=do_vat] -> "thuyền"
 NN[rn=do_vat] -> "nôi"
 NN[rn=su_vat] -> "ao"
 NN[rn=su_vat] -> "bờ"
 NN[rn=su_vat] -> "năng"
 NN[rn=su_vat] -> "sân"
 NN[rn=do_vat] -> "nhà"
 NN[rn=do_vat] -> "lá"
 NN[rn=do_vat] -> "sản_phẩm"
 NN[rn=do_vat] -> "ô_tô"
 NN[rn=do_vat] -> "cửa_hàng"
 NN[rn=dong_vat] -> "mèo"
 NN[rn=dong_vat] -> "chuột"
 NN[rn=dong_vat] -> "ngựa"
 NN[rn=dong_vat] -> "kiến"
 NN[rn=dong_vat] -> "chó"
 NN[rn=dong_vat] -> "nai"
 NN[rn=nguoì] -> "quân_đội"
 NN[rn=su_vat] -> "kẻ_thù"
 VB[rn=nguoì] -> "luyện_tập"
 VB[rn=dong_vat] -> "bay"
 VB[rn=vat_the] -> "mọc"
 VB[rn=do_vat] -> "rụng"
 VB[rn=do_vat] -> "cháy"
 VB[rn=nguoì] -> "nghiên_cứu"
 VB[rn=nguoì] -> "phát_triển"
 VB[rn=do_vat] -> "kéo"
 VB[rn=dong_vat, rd=do_vat] -> "gặm"
 VB[rn=nguoì, rd=dong_vat] -> "câu"
 VB[rd=am_thanh] -> "nghe"

3.2 Danh từ và giới từ

NN[rn=dong_vat, in=nuoc] -> "cá"
 NN[rn=thuc_vat, in=cay] -> "cây"
 NN[rn=do_vat, in="san"] -> "bóng"
 NN[rn=am_thanh, in=phat_nhac] -> "nhạc"

NN[rn=do_vat, in=thuyen] -> "thuyền"
 NN[rn=do_vat, in=noi] -> "nôi"
 NN[rn=su_vat, in=nuoc] -> "ao"
 NN[rn=su_vat, in=bo] -> "bờ"

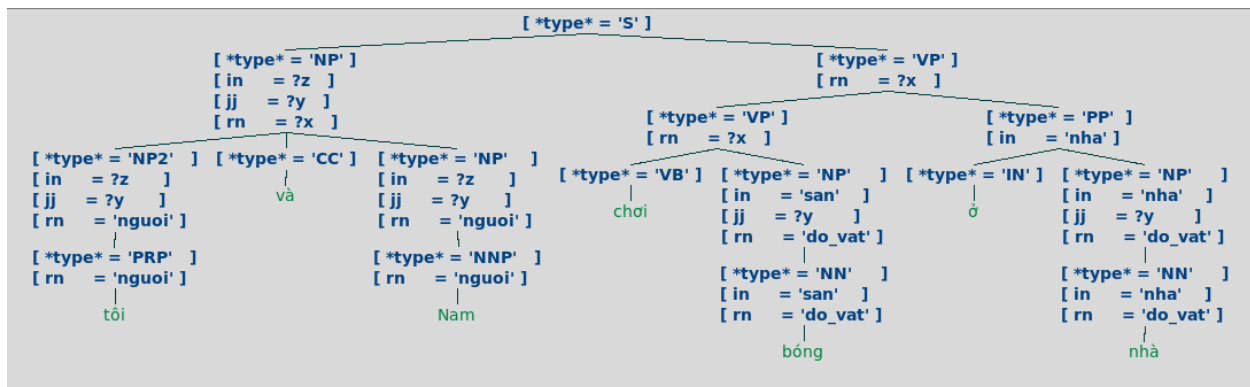
3.3 Danh từ và tính từ

NN[rn=dong_vat, jj=nho] -> "kiến"
 JJ[jj=mai] -> "mái"
 JJ[jj=den] -> "đen"
 JJ[jj=gay] -> "gầy"
 JJ[jj=nho] -> "nhỏ"
 JJ[jj=vang] -> "vàng"
 JJ[jj=moi] -> "mới"

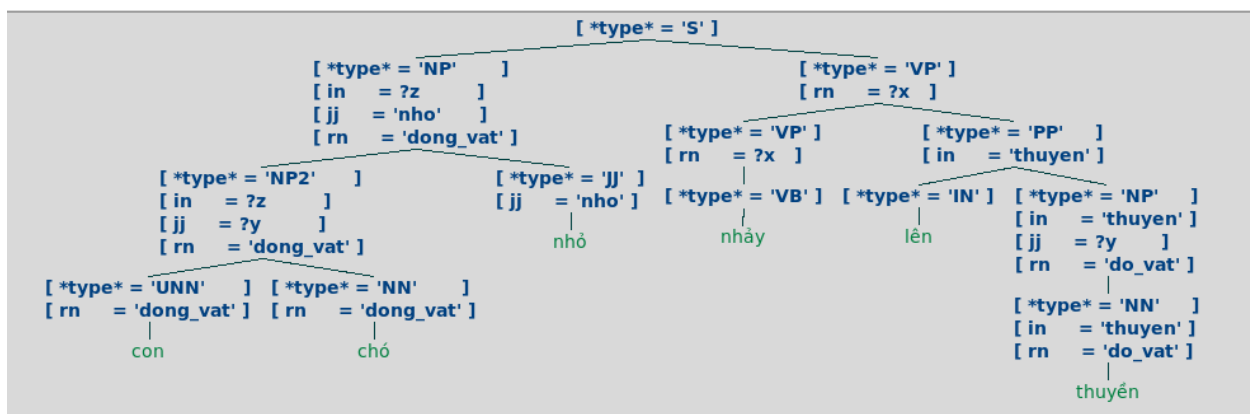
4 Phân tích 20 câu từ bộ từ vựng và đánh giá

4.1 Phân tích 20 câu

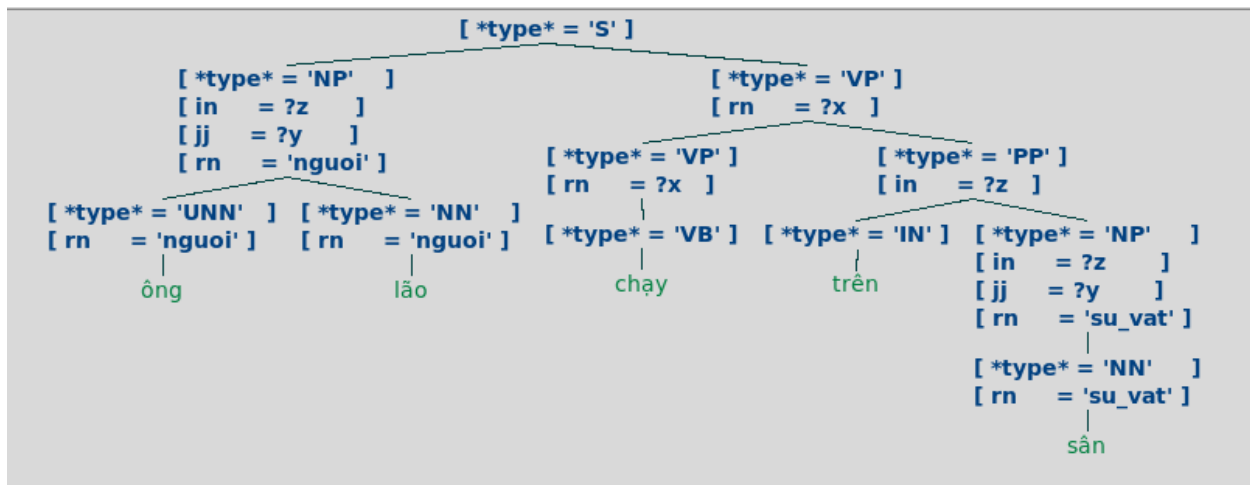
4.1.1 Tôi và Nam chơi bóng ở nhà



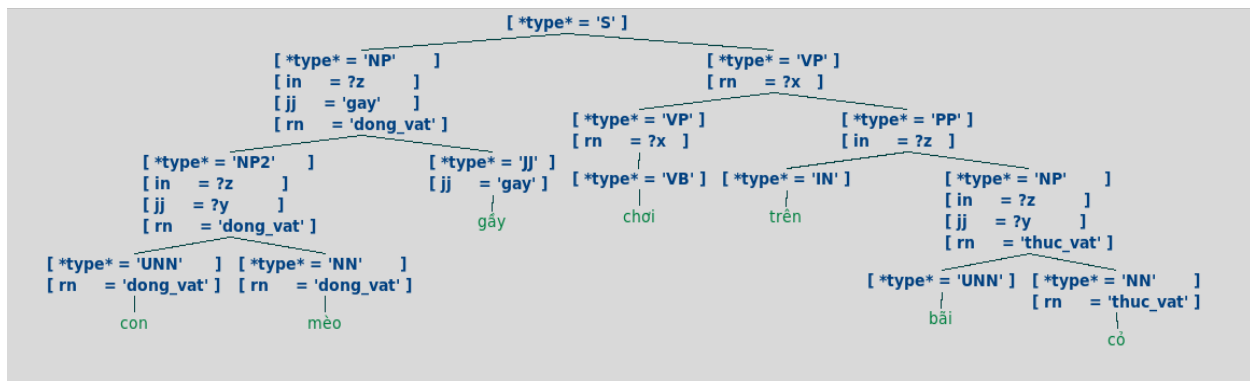
4.1.2 Con chó nhỏ nhảy lên thuyền



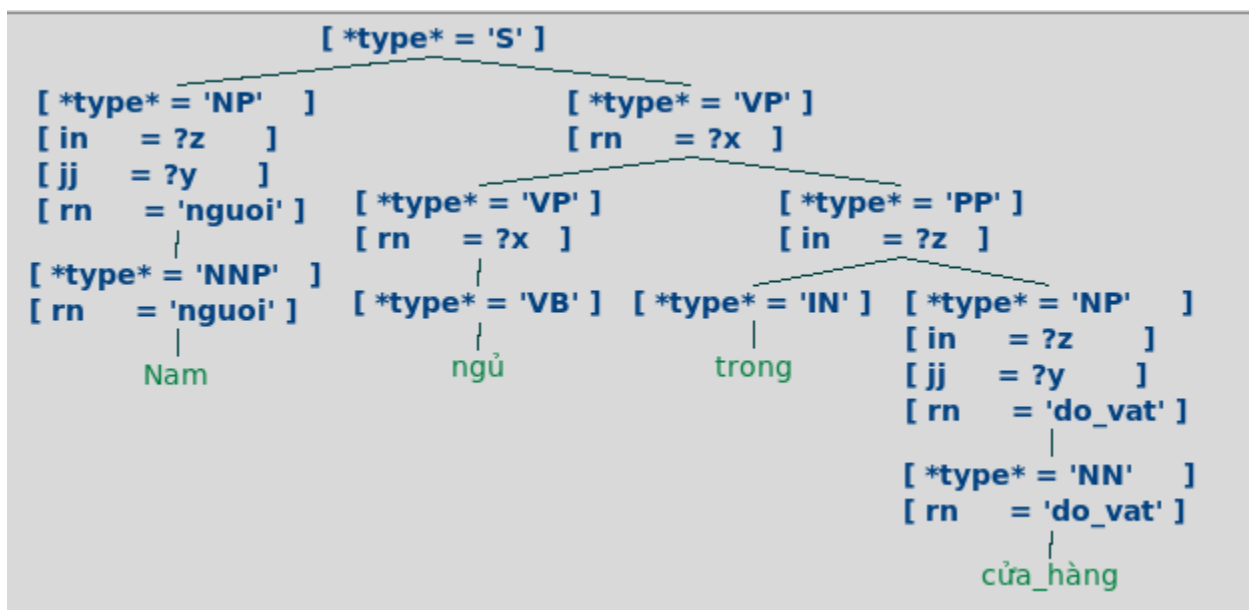
4.1.3 Ông lão chạy trên sân



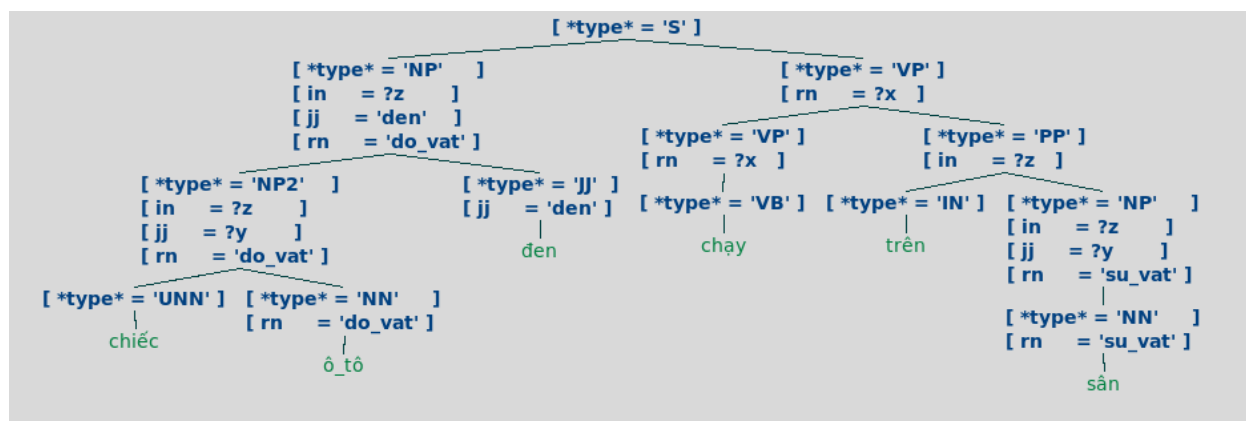
4.1.4 Con mèo gầy chơi trên bãi cỏ



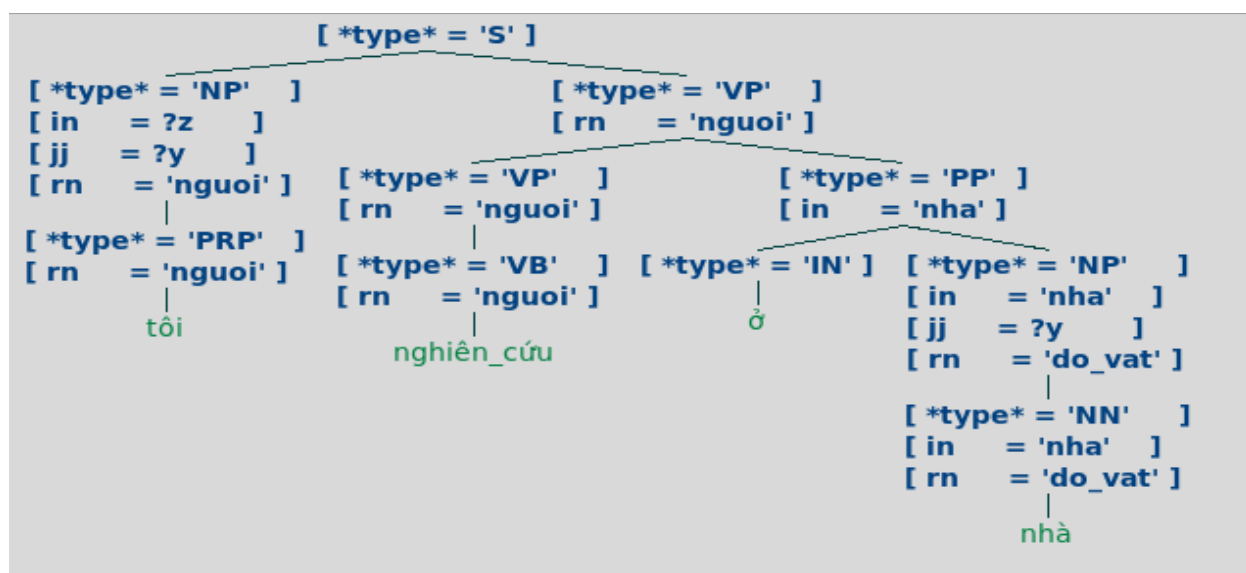
4.1.5 Nam ngủ trong cửa hàng



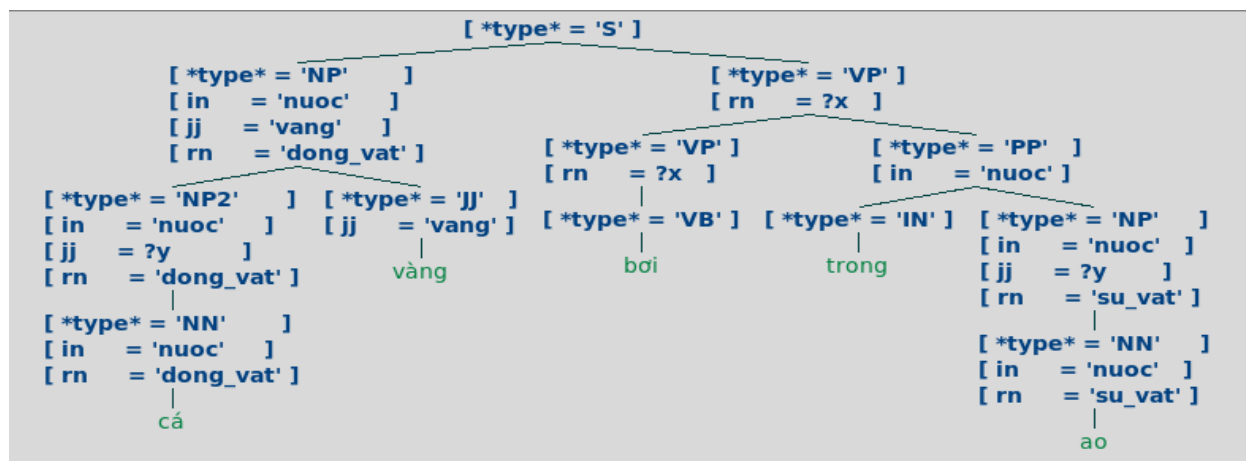
4.1.6 Chiếc ô tô đen chạy trên sân



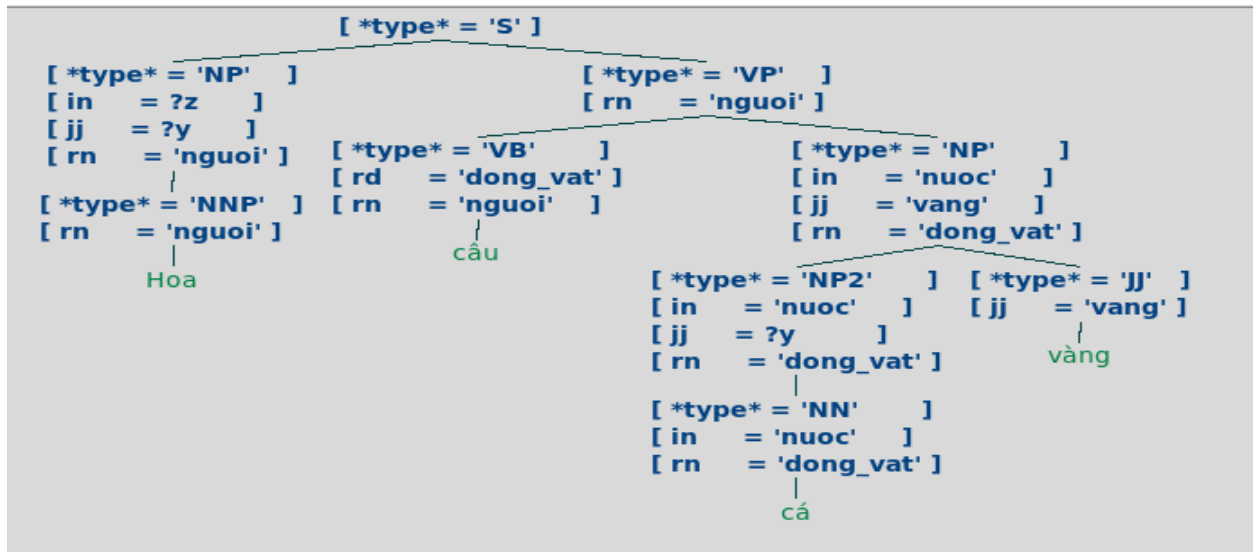
4.1.7 Tôi nghiên cứu ở nhà



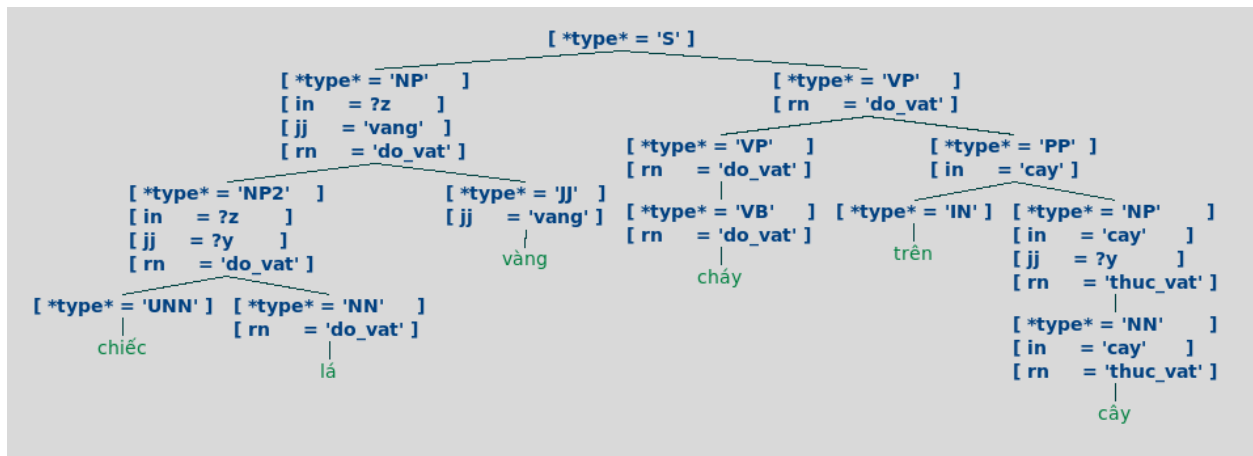
4.1.8 Cá vàng bơi trong ao



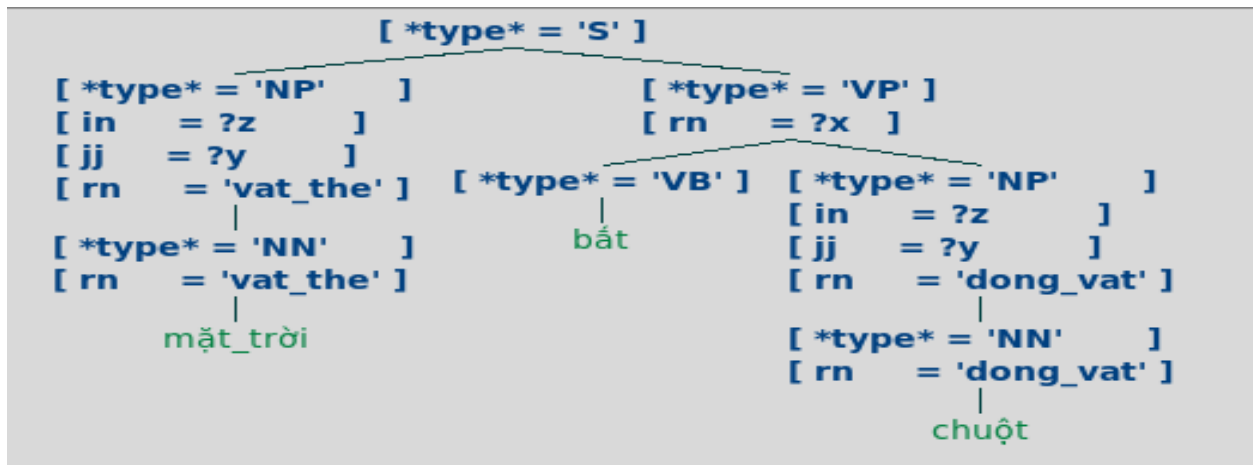
4.1.9 Hoa câu cá vàng



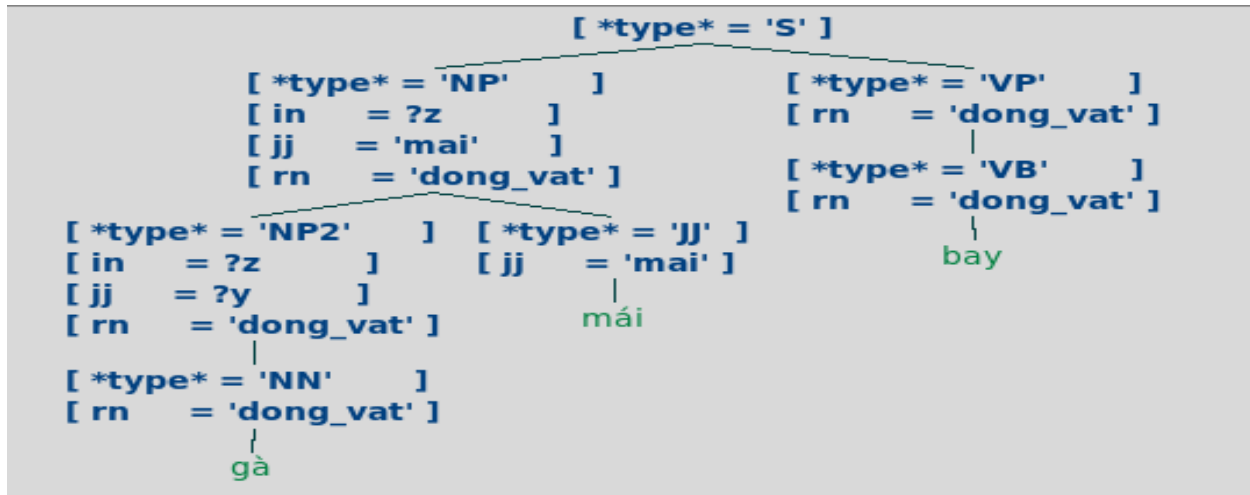
4.1.10 Chiếc lá vàng cháy trên cây



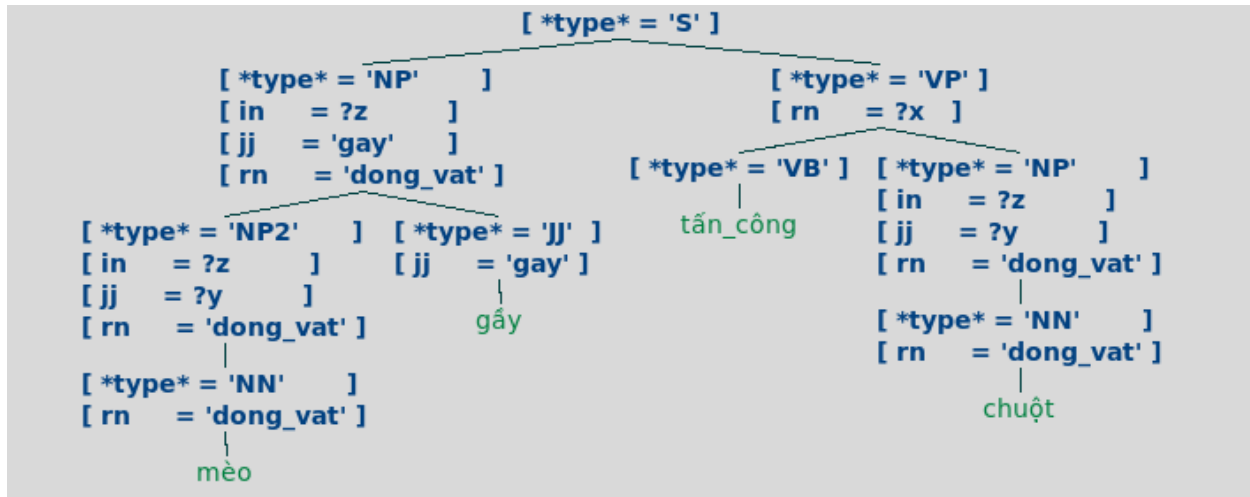
4.1.11 Mặt trời bắt chuột



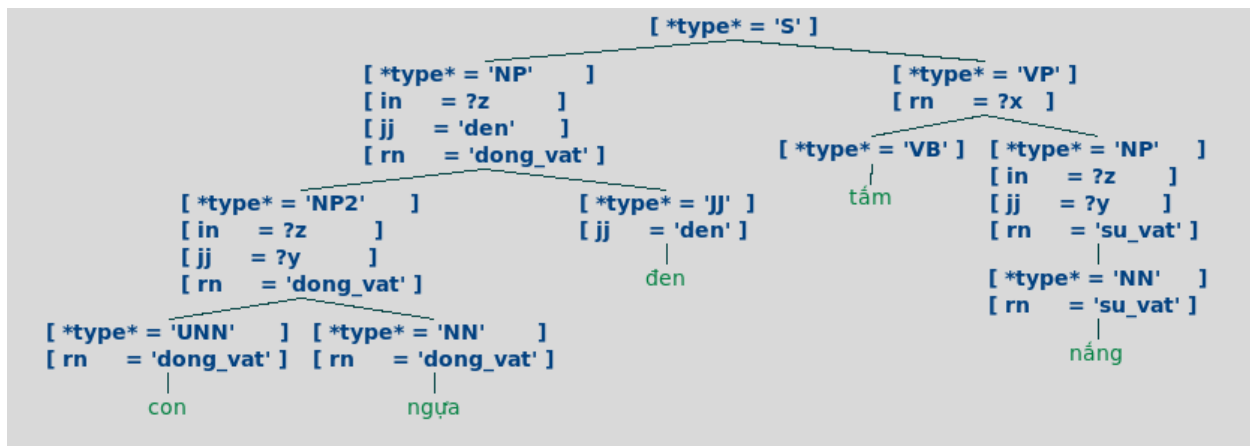
4.1.12 Gà mái bay



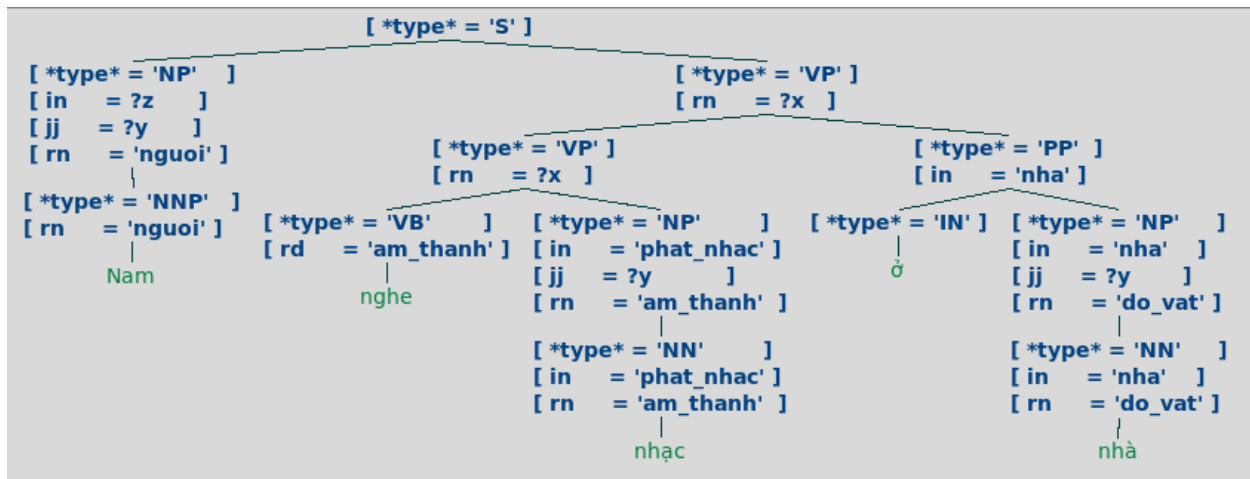
4.1.13 Mèo gây tấn công chuột



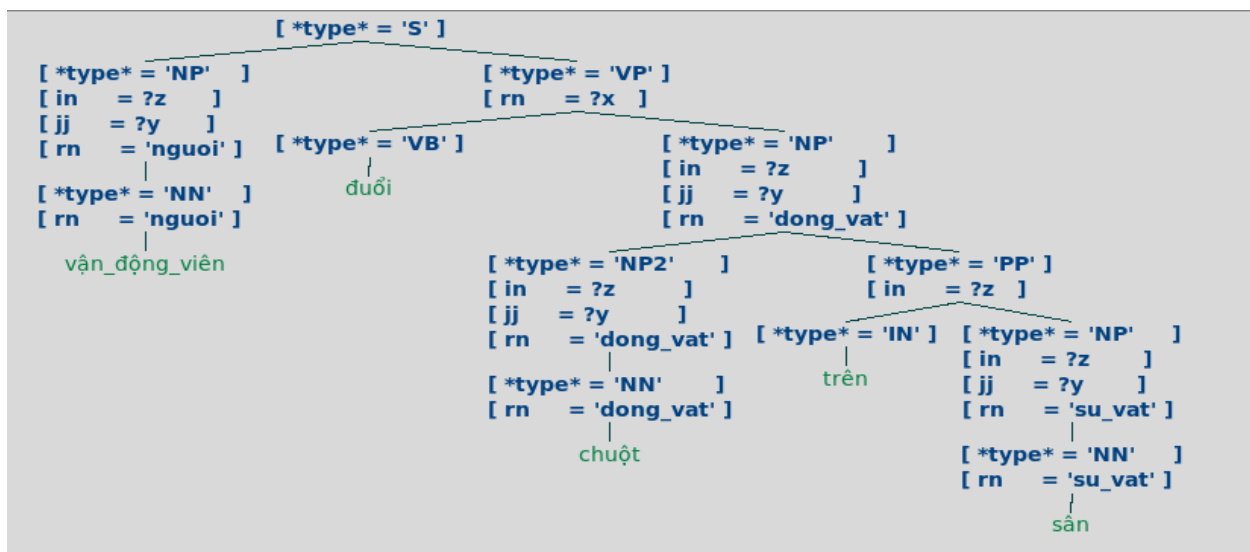
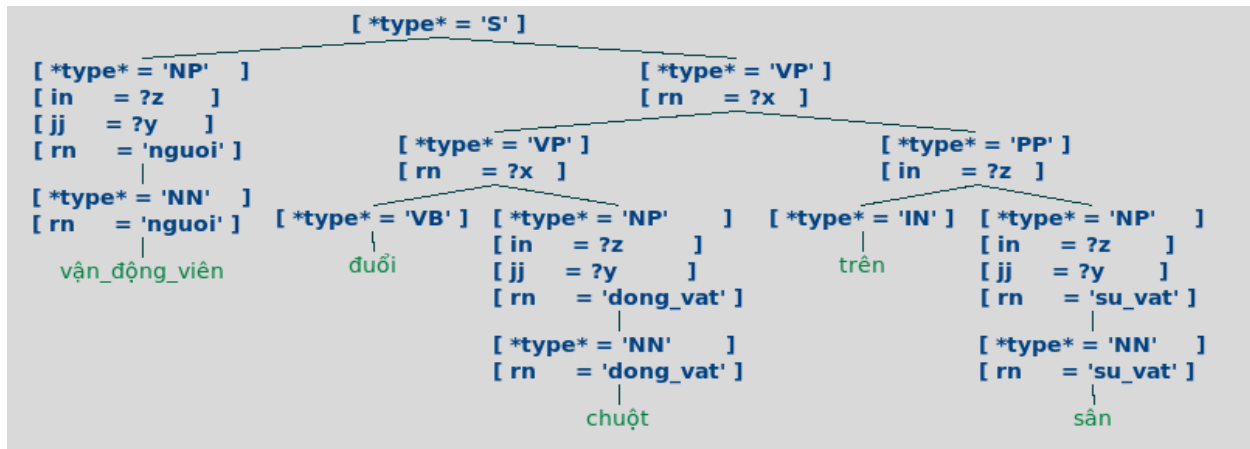
4.1.14 Con ngựa đen tăm nể



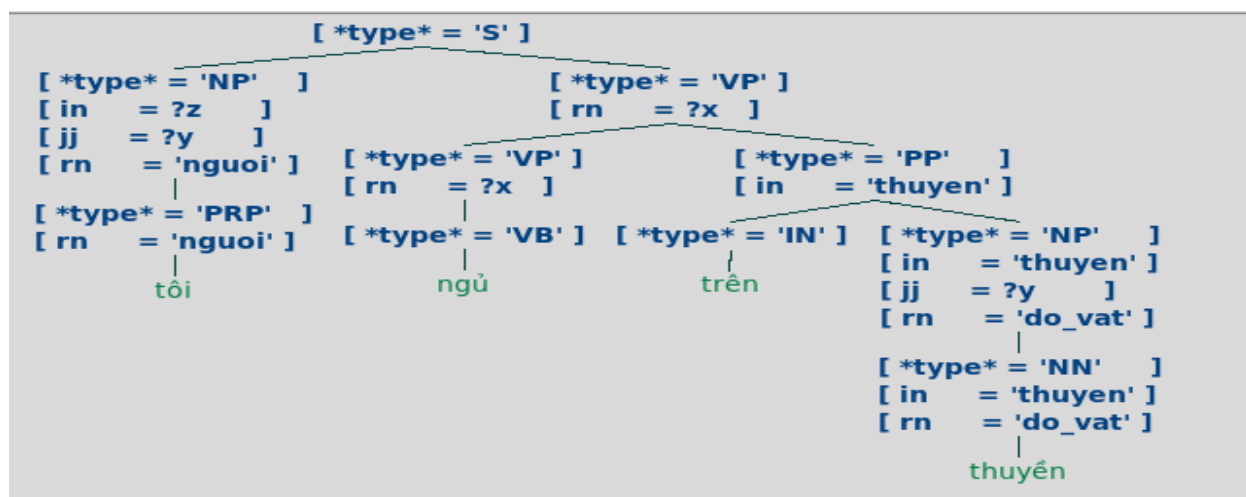
4.1.15 Nam nghe nhạc ở nhà



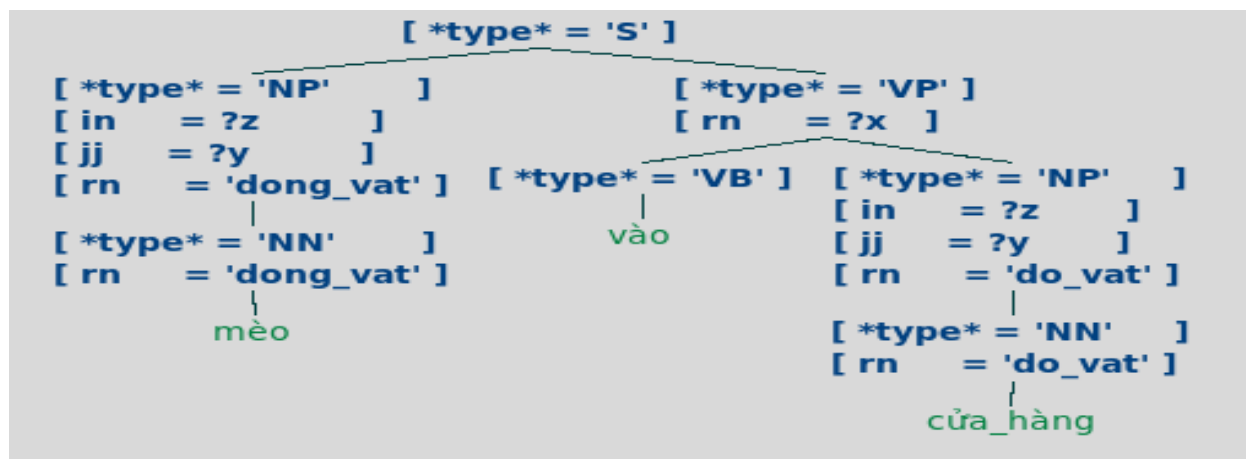
4.1.16 Vận động viên đuổi chuột trên sân



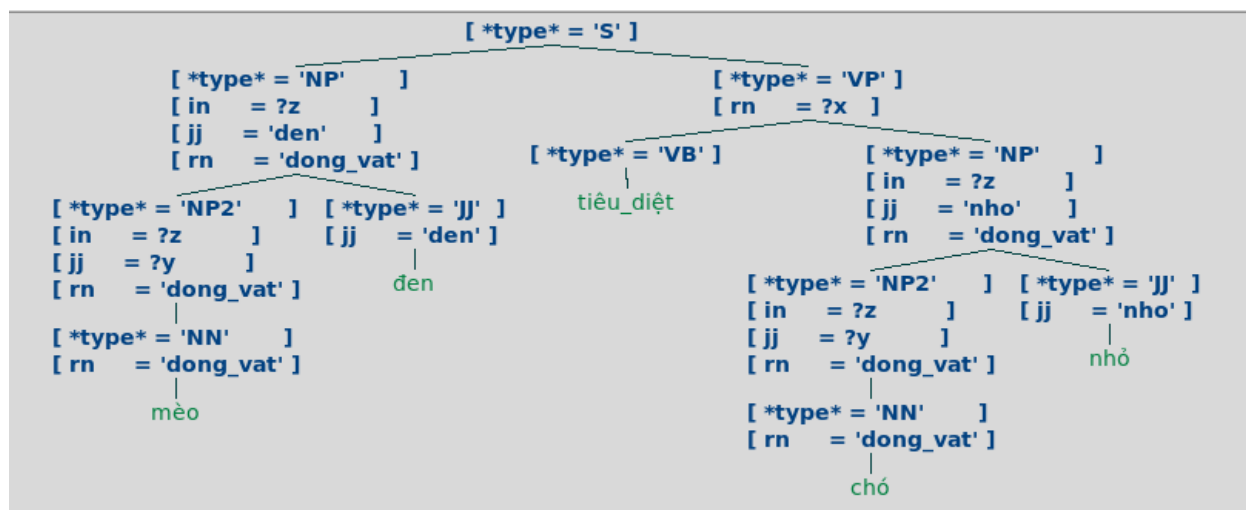
4.1.17 Tôi ngủ trên thuyền



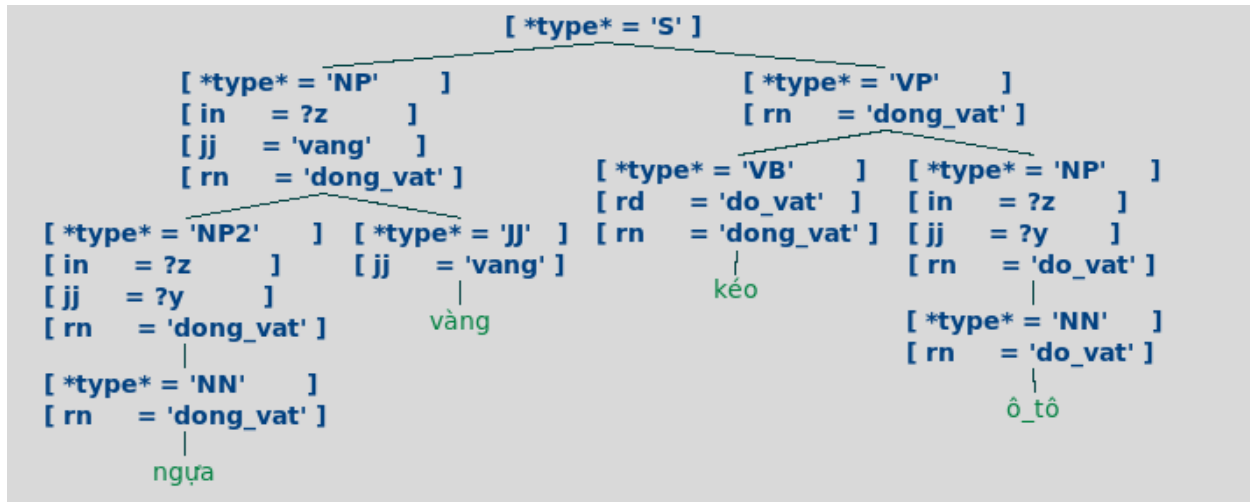
4.1.18 Mèo vào cửa hàng



4.1.19 Mèo đen tiêu diệt chó nhỏ



4.1.20 Ngựa vàng kéo ô tô



4.2 Kết luận

Sau khi thực hiện phân tích cú pháp các câu mới được tạo ra trên chương trình, chúng ta có thể rút ra được các kết luận:

- Chương trình phân tích cú pháp hoạt động tốt và đưa ra kết quả phù hợp với các câu trên.
- Các ràng buộc ngữ nghĩa trong chương trình hiện tại chưa có nhiều nhưng vẫn cho ra kết quả tốt chỉ ra rằng số lượng từ vựng trong bộ dữ liệu ban đầu còn hạn chế, bản chất ngữ nghĩa của các từ xa rời nhau, ít có sự đan xen nghĩa của các câu.
- Các câu mới tạo ra bản chất có cấu trúc khá gần với những câu đặt ra ban đầu nên chương trình phân tích cú pháp dễ dàng đưa ra kết quả tốt.
- Ở câu *vận động viên đuổi chuột trên sân*, chương trình cho ra 2 cách phân tích cú pháp câu tùy thuộc vào ràng buộc ngữ nghĩa danh từ và giới từ. Tuy nhiên, nếu không quy định rõ ràng buộc ngữ nghĩa như ở đây thì chúng ta vẫn thấy cả 2 cách đều phù hợp.