



ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Bài 3. Văn phạm sản sinh

Lý thuyết ngôn ngữ

- Mô hình cho tất cả các ngôn ngữ
- Ngôn ngữ là tập các xâu (sentence, string) trên một bảng chữ nào đó
- Ví dụ về xâu
 - Dãy các bit
 - Số thực
 - Chương trình C
 - Câu tiếng Việt

Vấn đề biểu diễn ngôn ngữ

- Thực chất là biểu diễn cú pháp của ngôn ngữ
- Biểu diễn phải hữu hạn
- Công cụ sản sinh: văn phạm
- Công cụ đoán nhận: ô tômat

Phân cấp Chomsky

Lớp ngôn ngữ	Công cụ sản sinh	Công cụ đoán nhận	Ghi chú
Đệ quy kể được	Văn phạm loại 0 (ngữ cấu)	Máy Turing	Các bài toán tổng quát
Cảm ngữ cảnh	Văn phạm cảm ngữ cảnh	Ôtômat tuyến tính giới nội	Ngôn ngữ tự nhiên
Phi ngữ cảnh	Văn phạm phi ngữ cảnh	Ôtômat đẩy xuống	Ngôn ngữ lập trình, phần chính của ngôn ngữ tự nhiên
Chính quy	Văn phạm chính quy Công cụ biểu diễn: Biểu thức chính quy	Ôtômat hữu hạn	Từ vựng của ngôn ngữ tự nhiên, ngôn ngữ lập trình

Văn phạm xuất phát từ ngôn ngữ tự nhiên

<câu>::=<chủ ngữ> <vị ngữ>

<chủ ngữ>::=<danh ngữ>

<danh ngữ>::=<danh từ> <tính từ>

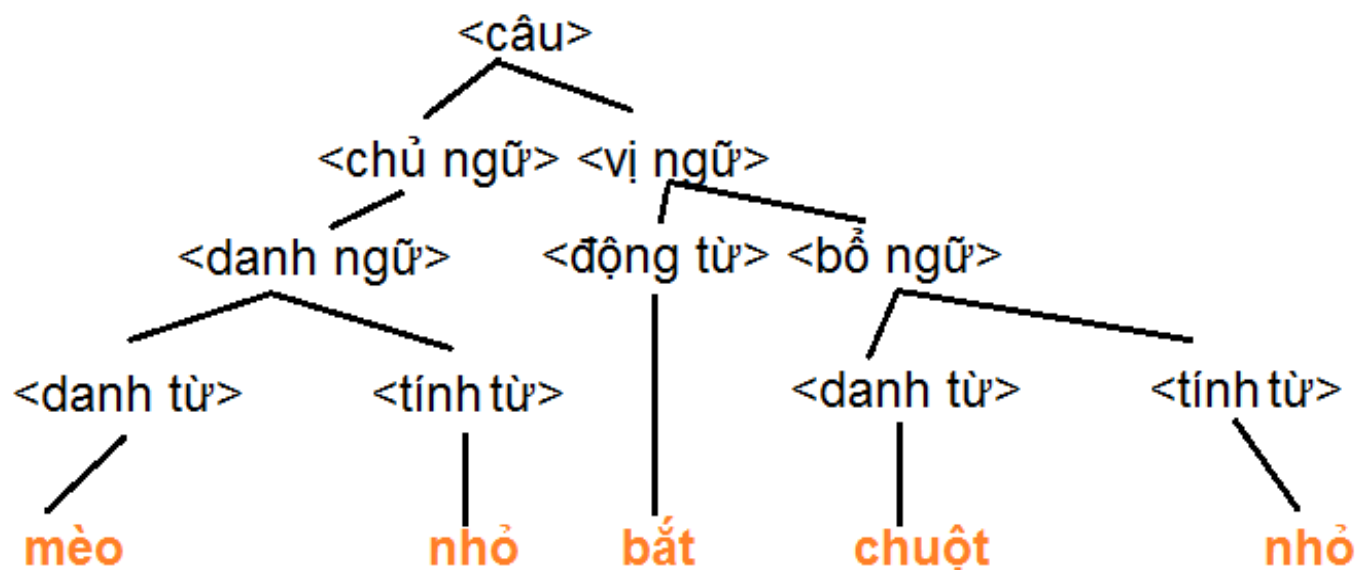
<vị ngữ>::=<động từ> <bổ ngữ>

<bổ ngữ>::=<danh từ> <tính từ>

<động từ> ::= **bắt**

<danh từ>::= **mèo** | **chuột**

<tính từ>::= **nhỏ**



Văn phạm sản sinh các số thực

$\langle \text{Số} \rangle ::= -\langle \text{Số thập phân} \rangle \mid \langle \text{số thập phân} \rangle$

$\langle \text{Số thập phân} \rangle ::= \langle \text{Dãy chữ số} \rangle \mid$

$\langle \text{Dãy chữ số} \rangle . \langle \text{Dãy chữ số} \rangle$

$\langle \text{Dãy chữ số} \rangle ::= \langle \text{Chữ số} \rangle \mid \langle \text{Chữ số} \rangle \langle \text{Dãy chữ số} \rangle$

$\langle \text{Chữ số} \rangle ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

Làm thế nào để sản sinh ra các xâu ?

Văn phạm phi ngữ cảnh có thể dùng để sản sinh ra các xâu thuộc ngôn ngữ như sau:

$X = \text{Ký hiệu đầu}$

While còn ký hiệu không kết thúc Y trong X ***do***

Áp dụng một trong các sản xuất của, văn phạm chẳng hạn $Y \rightarrow w$

Khi X chỉ chứa ký hiệu kết thúc, nó là xâu được sản sinh bởi văn phạm.

Ví dụ

$$S \rightarrow -A \mid A$$
$$A \rightarrow B.B \mid B$$
$$B \rightarrow BC \mid C$$
$$C \rightarrow 0 \mid 1 \mid 2 \mid \dots \mid 9$$

Quá trình sản sinh xâu -3.14

Quá trình thay thế

S

-A

-B.B

-B.BC

-C.BC

-C.CC

-3.CC

-3.1C

-3.14

Sản xuất được sử dụng

$S \rightarrow -A$

$A \rightarrow B.B$

$B \rightarrow BC$

$B \rightarrow C$

$B \rightarrow C$

$C \rightarrow 3$

$C \rightarrow 1$

$C \rightarrow 4$

Quá trình suy dẫn (derivation)

Suy dẫn (Derivations)

- Mỗi lần thực hiện việc thay thế là một bước suy dẫn.
- Nếu mỗi dạng câu có nhiều ký hiệu không kết thúc để thay thế có thể thay thế bất cứ ký hiệu không kết thúc nào

Suy dẫn trái và suy dẫn phải

- Nếu giải thuật phân tích cú pháp chọn ký hiệu không kết thúc cực trái hay cực phải để thay thế, kết quả của nó là suy dẫn trái hoặc suy dẫn phải

Ví dụ suy dẫn trái:

$$\begin{aligned} S &\Rightarrow -A \Rightarrow -B.B \Rightarrow -C.B \Rightarrow -3.B \Rightarrow -3.BC \Rightarrow -3.CC \\ &\Rightarrow -3.1C \Rightarrow -3.14 \end{aligned}$$

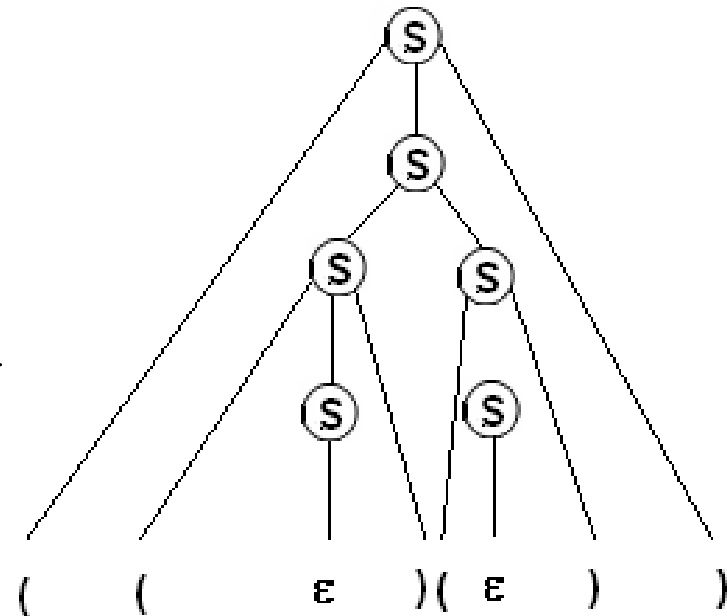
Ví dụ suy dẫn phải:

$$\begin{aligned} S &\Rightarrow -A \Rightarrow -B.B \Rightarrow -B.BC \Rightarrow -B.B4 \Rightarrow -B.C4 \Rightarrow -B.14 \\ &\Rightarrow -C.14 \Rightarrow -3.14 \end{aligned}$$

Cây suy dẫn (Cây phân tích cú pháp)

Cây suy dẫn có những đặc điểm sau

- 1) Mỗi nút của cây có nhãn là ký hiệu kết thúc, ký hiệu không kết thúc hoặc ϵ (xâu rỗng)
- 2) Nhãn của nút gốc là S (ký hiệu đầu)
- 3) Nút trong có nhãn là ký hiệu không kết thúc (nút lá có nhãn là ký hiệu kết thúc hoặc ϵ)
- 4) Nút A có các nút con từ trái qua phải là X_1, X_2, \dots, X_k thì có một sản xuất dạng $A \rightarrow X_1 X_2 \dots X_k$
- 5) Nút lá có thể có nhãn ϵ chỉ khi tồn tại sản xuất $A \rightarrow \epsilon$ và nút cha của nút lá chỉ có một nút con duy nhất



Ví dụ: Cây phân tích cú pháp của văn phạm

G: $S \rightarrow SS \mid (S) \mid \epsilon$ $w = (())$

Văn phạm nhập nhằng

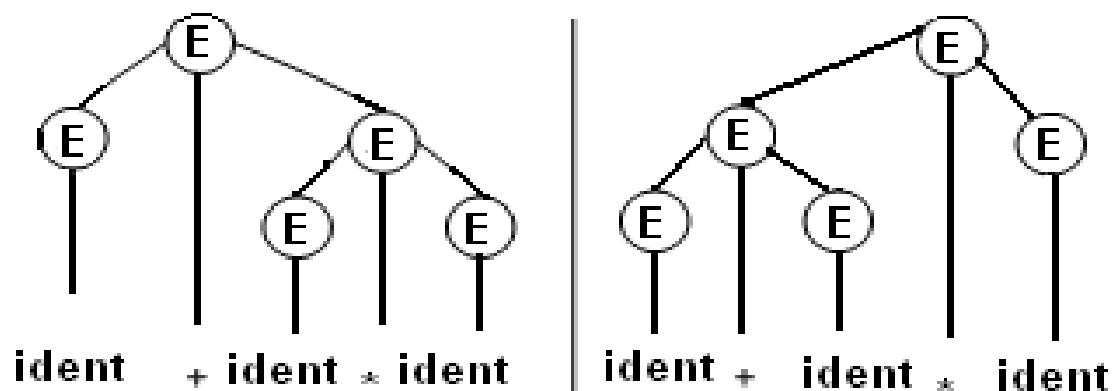
Văn phạm


$E \rightarrow E + E$

$E \rightarrow E * E$

$E \rightarrow (E)$

$E \rightarrow \text{TK_IDENT}$



Cho phép đưa ra hai suy dẫn khác nhau cho
xâu $\text{TK_IDENT} + \text{TK_IDENT} * \text{TK_IDENT}$
 (chẳng hạn $x + y * z$)

Văn phạm là nhập nhằng

Khử nhập nhằng (disambiguation)

$E \rightarrow E + T$

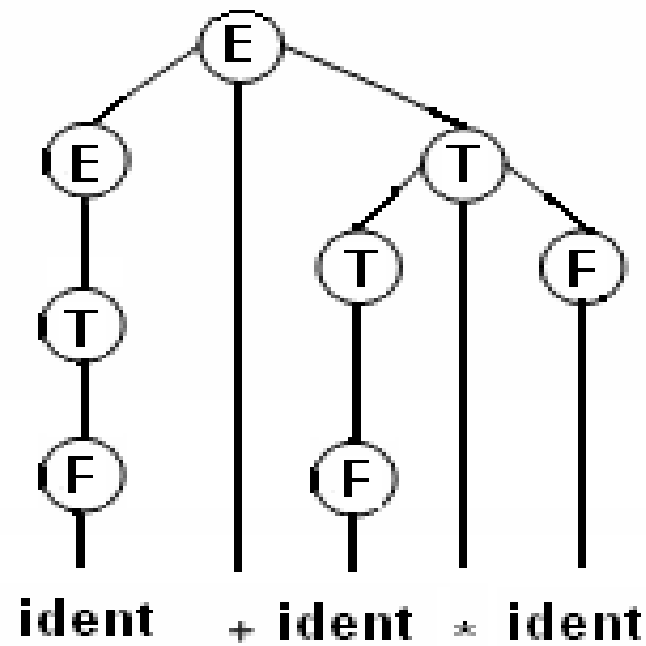
$E \rightarrow T$

$T \rightarrow T * F$

$T \rightarrow F$

$F \rightarrow (E)$

$F \rightarrow \text{TK_IDENT}$



E: Expression, T: Term, F: Factor

(Bằng cách thêm các ký hiệu không kết thúc và các sản xuất để đảm bảo thứ tự ưu tiên)

Đệ quy

- Một sản xuất là đệ quy nếu $X \Rightarrow^* \omega_1 X \omega_2$
- Có thể dùng để biểu diễn các quá trình lặp hay cấu trúc lồng nhau

Đệ quy trực tiếp $X \Rightarrow \omega_1 X \omega_2$

Đệ quy trái $X \Rightarrow b \mid \textcolor{red}{X}a.$

$X \Rightarrow Xa \Rightarrow Xaa \Rightarrow Xaaa \Rightarrow baaaaa \dots$

Đệ quy phải $X \Rightarrow b \mid a \textcolor{red}{X}.$

$X \Rightarrow aX \Rightarrow aax \Rightarrow aaax \Rightarrow \dots aaaaaab$

Đệ quy giữa $X \Rightarrow b \mid (\textcolor{red}{X}).$

$X \Rightarrow (X) \Rightarrow ((X)) \Rightarrow (((X))) \Rightarrow (((... (b) ...)))$

Đệ quy gián tiếp $X \Rightarrow^* \omega_1 X \omega_2$

Khử đệ quy trái

$$E \rightarrow E + T \mid T$$
$$T \rightarrow T * F \mid F$$
$$F \rightarrow (E) \mid \text{TK_IDENT}$$

Khử đệ quy trái bằng cách thêm ký hiệu không kết thúc và sản xuất mới

$$E \rightarrow T E'$$
$$E' \rightarrow + T E' \mid \varepsilon$$
$$T \rightarrow F T'$$
$$T' \rightarrow * F T' \mid \varepsilon$$
$$F \rightarrow (E) \mid \text{TK_IDENT}$$