



CHƯƠNG TRÌNH DỊCH

Bài 6: Phân tích cú pháp



Nội dung

1. Vai trò của bộ phân tích cú pháp (PTCP)
2. Nhiệm vụ của PTCP
3. Các mục tiêu của PTCP
4. Đầu vào và đầu ra của PTCP
5. Các bước xây dựng bộ PTCP
6. Bài tập và thảo luận

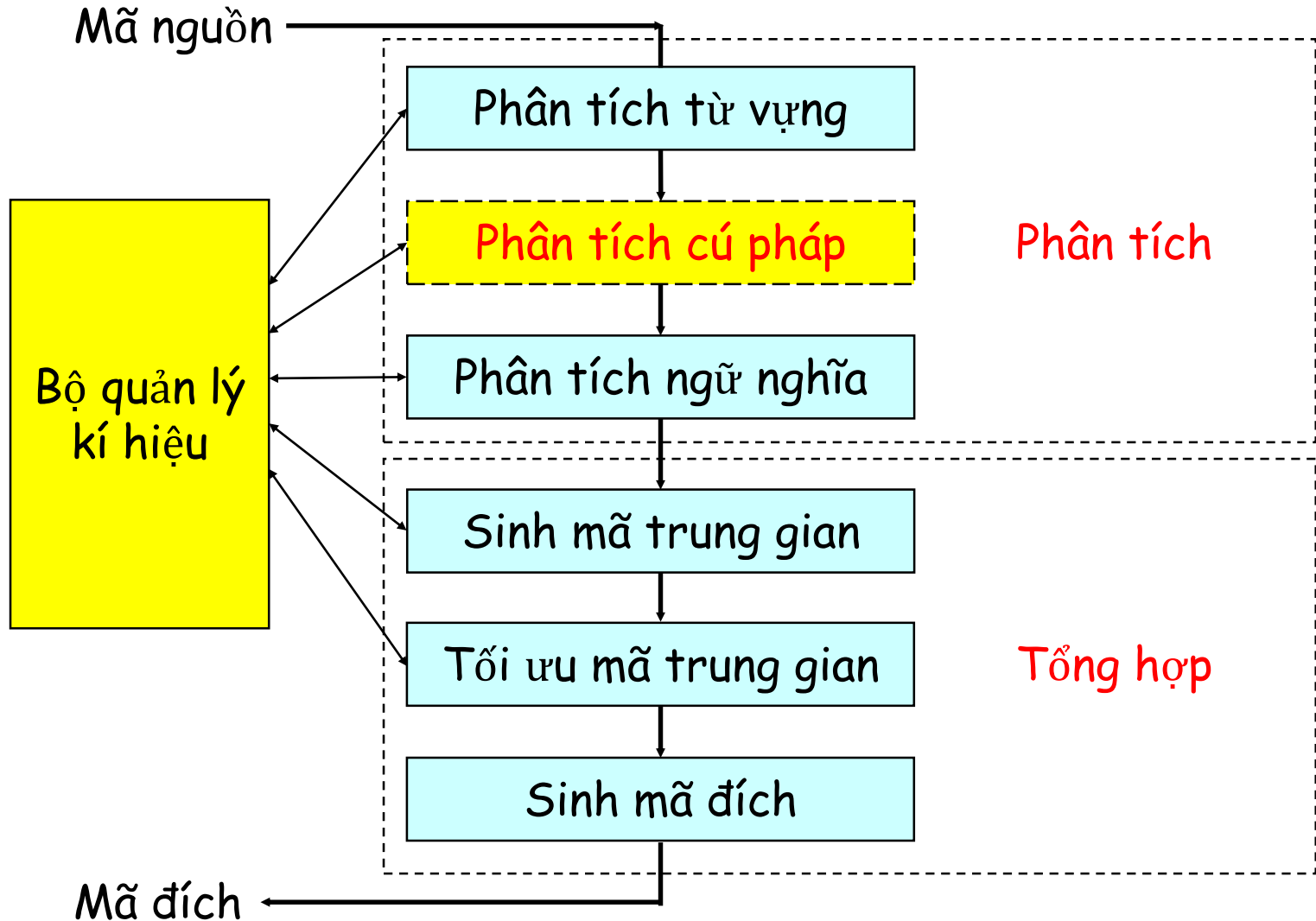


Phần 1

Vai trò của bộ phân tích cú pháp (PTCP)



Cấu trúc một chương trình dịch





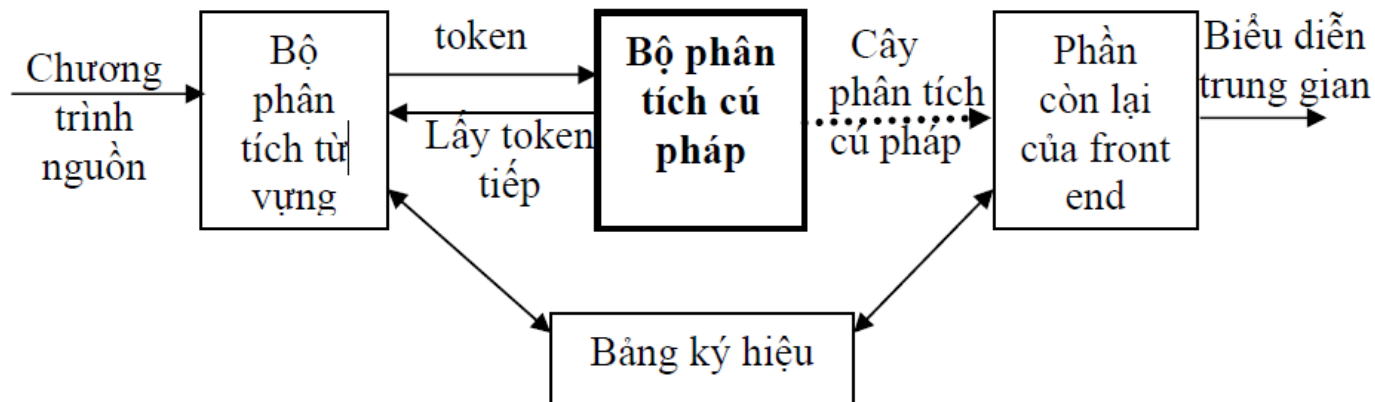
Vai trò của bộ phân tích cú pháp

- Phân tích cú pháp là pha thứ hai của trình dịch
- Bộ PTCP nhận dữ liệu đầu vào là dãy các từ tố (cùng với các thông tin kèm theo), dựa theo các luật văn phạm của ngôn ngữ, xây dựng cây cú pháp (syntax tree) của chuỗi vào
 - PTCP làm việc chặt chẽ với PTTV và thường có thể bắt đầu thực hiện công việc ngay khi PTTV mới có những kết quả ban đầu (không cần đợi PTTV kết thúc)
 - Đối với một số ngôn ngữ đơn giản, thiết kế trình dịch còn đi xa hơn bằng cách ghép PTTV và PTCP thành một module duy nhất (dịch trực tiếp văn bản)



Vai trò của bộ phân tích cú pháp

- PTCP cung cấp dữ liệu cho bộ phân tích ngữ nghĩa
 - Làm việc độc lập với bộ PTNN
 - Chỉ trả về kết quả cho phân tích ngữ nghĩa sau khi đã hoàn thành đầy đủ (hoặc tương đối đầy đủ) việc tạo cây cú pháp
- Ngoài ra PTCP cũng cung cấp dữ liệu về lỗi và gợi ý sửa lỗi cho bộ soạn thảo





Vai trò của bộ phân tích cú pháp

- Trái với bộ PTTV, thường được “đính kèm” nhiều nhiệm vụ khác, thiết kế bộ PTCP thường chỉ nhằm tới nhiệm vụ duy nhất là dựng cây cú pháp
 - Lý do chính là việc xây bộ PTCP hiệu quả khá phức tạp
- Tuy PTCP trả về kết quả cho PTNN, nhưng trong một số thiết kế, chính bộ PTCP sẽ quyết định khi nào thực hiện phân tích ngữ nghĩa dựa trên các điều kiện khởi động (trigger)
 - Chẳng hạn: sau khi kết thúc khai báo một class (trong ngôn ngữ C++) thì có thể tiến hành luôn việc phân tích ngữ nghĩa của class đó



Phần 2

Nhiệm vụ của phân tích cú pháp



Nhiệm vụ của phân tích cú pháp

- PTCP đảm nhiệm nhiệm vụ phức tạp nhất của trình dịch, đó là kiểm tra lỗi cú pháp của chuỗi vào (vốn có thể làm sai lệch hoàn toàn ý nghĩa của input)
- Các nhiệm vụ chính (nhất thiết phải có để đảm bảo hoạt động của chương trình dịch):
 - Xây dựng cây cú pháp cho chuỗi vào
 - Thực hiện một số thao tác ngữ nghĩa phục vụ cho việc phân tích tiếp theo
 - Phát hiện các lỗi về văn phạm và lựa chọn phương pháp xử lý phù hợp
 - Xử lý lỗi để tiếp tục thực hiện việc phân tích
 - Đưa ra các gợi ý sửa lỗi cho mã nguồn



Phần 3

Các mục tiêu của phân tích cú pháp



Các mục tiêu của PTCP

- **Chính xác:** đây là mục tiêu quan trọng nhất, kết quả phân tích cần trả về chính xác cây phân tích
- **Tốc độ:** khó xây dựng các bộ PTCP tuyến tính theo độ dài của chuỗi vào (ngoại trừ ngôn ngữ đầu vào có văn phạm quá đơn giản), nhưng bộ PTCP cần hoạt động đủ nhanh (nên cần tuyến tính)
- **Chịu lỗi:** bộ PTCP cần có khả năng chịu lỗi và có chiến lược khắc phục lỗi phù hợp
- **Hiệu quả về bộ nhớ:** bộ PTCP cần sử dụng bộ nhớ một cách hiệu quả (do việc phải lưu trữ toàn bộ cây phân tích cho mã nguồn)

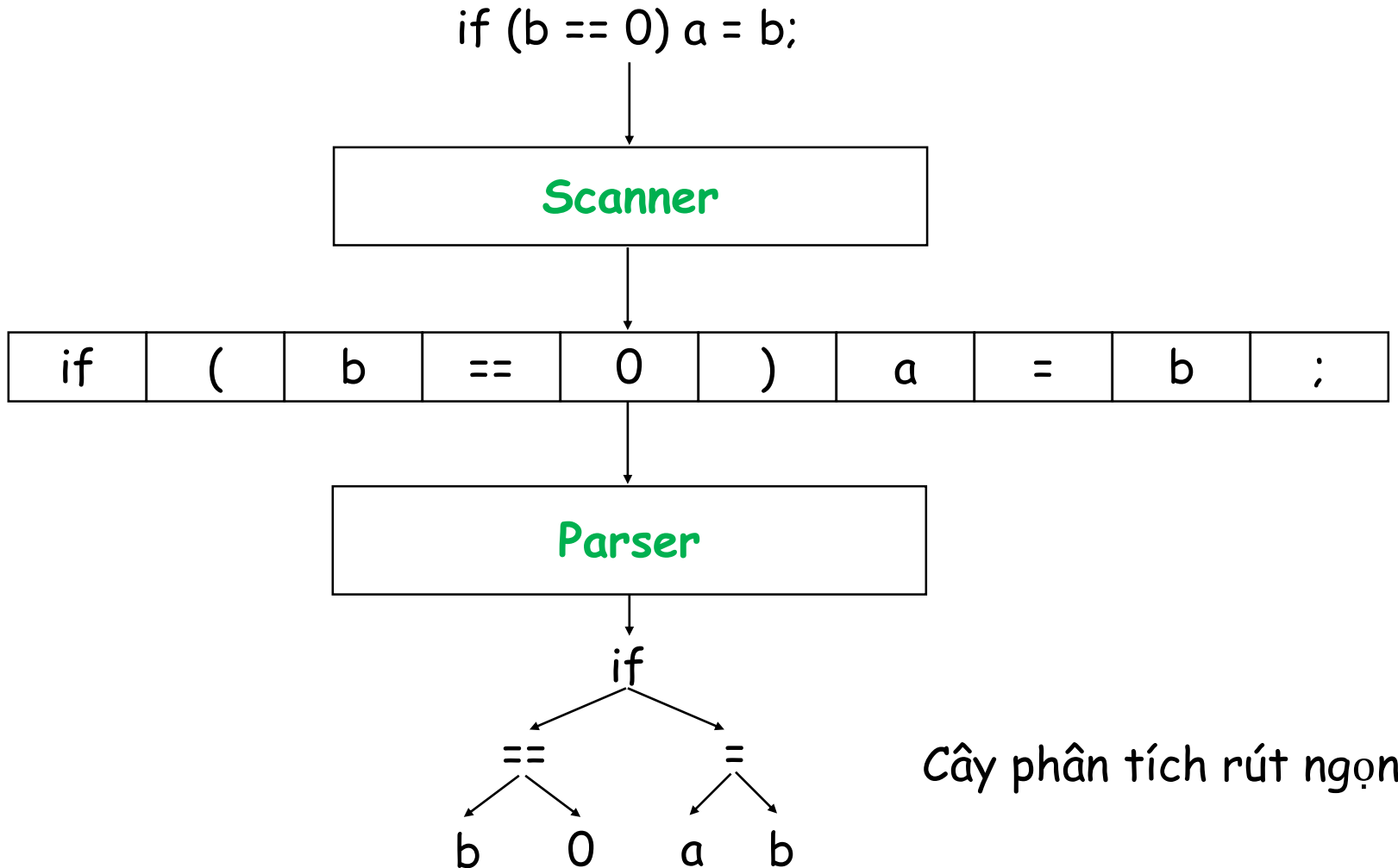


Phần 4

Đầu vào và đầu ra của phân tích cú pháp



Ví dụ về phân tích cú pháp



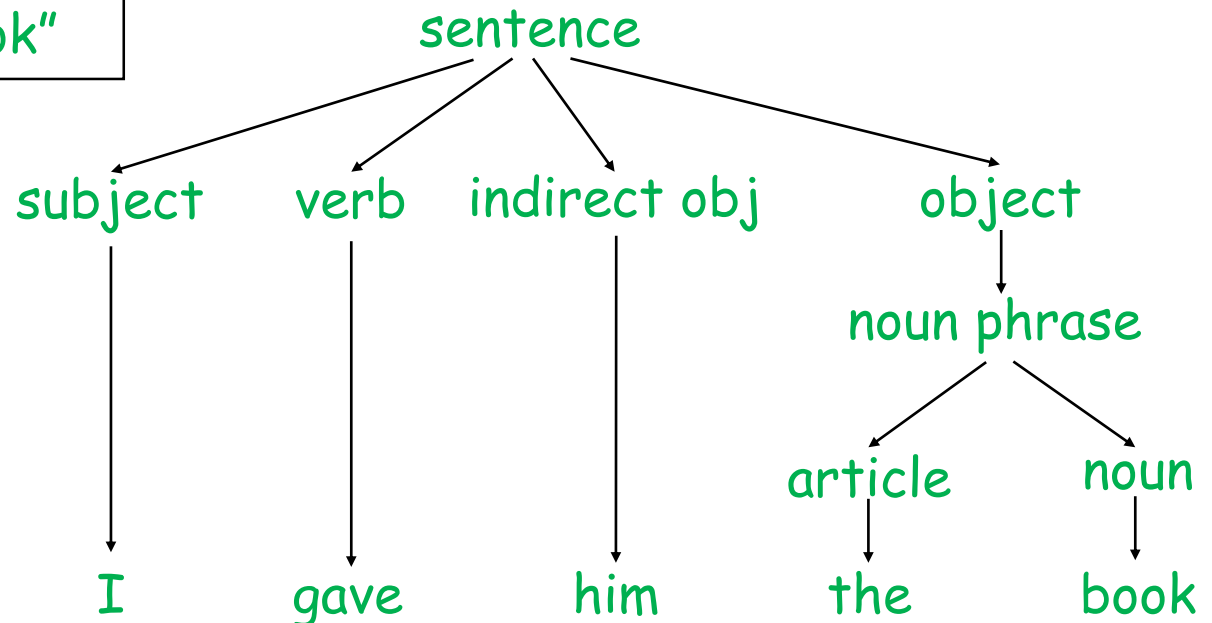


Ví dụ về phân tích cú pháp

PTCP có thể áp dụng với ngôn ngữ tự nhiên, nhưng:

- Ngôn ngữ tự nhiên có luật ngữ pháp phức tạp
- Ngôn ngữ tự nhiên có yếu tố ngữ cảnh
- Ngôn ngữ tự nhiên có nhiều lỗi (trong thực tế sử dụng)

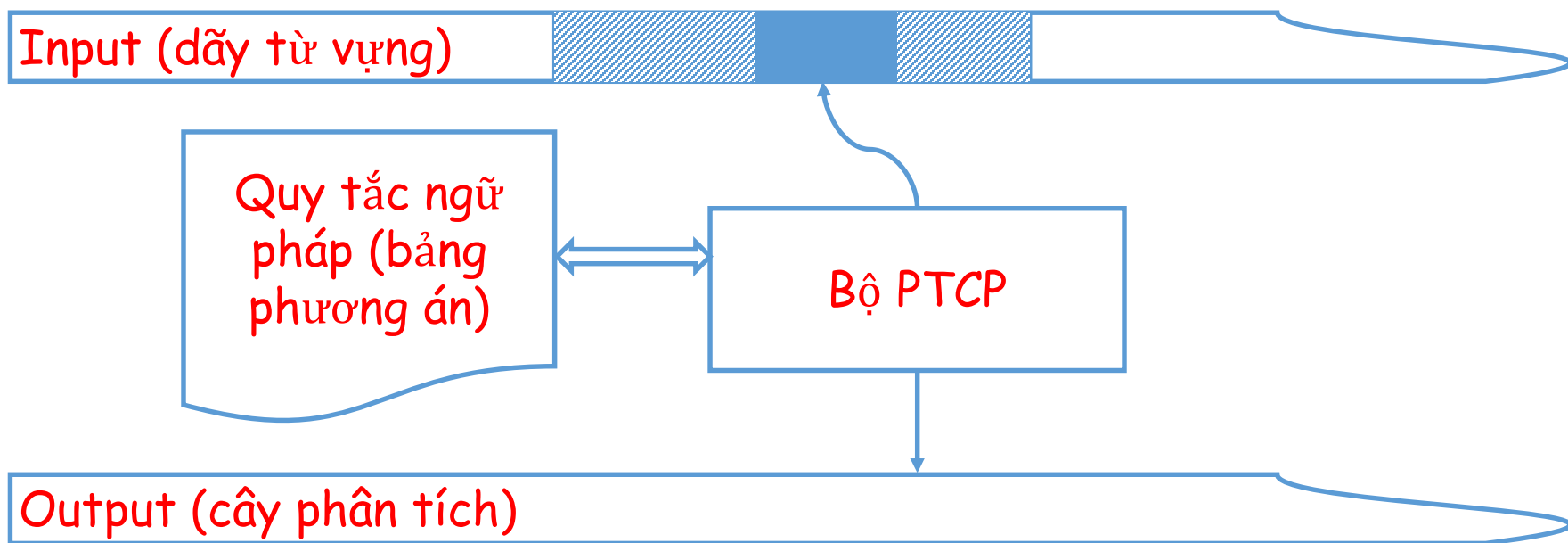
"I gave him the book"





Đầu vào của bộ PTCP

- Đầu vào của bộ PTCP là dãy các từ vựng đã được xác định chi tiết về từ loại
- Bộ PTCP thường cần quan sát nhiều hơn 1 kí hiệu đầu vào để ra quyết định dựng cây phân tích





Đầu ra của bộ PTCP

- Đầu ra của bộ PTCP là đầu vào của bộ PTNN, thường thì chỉ có thể hiểu đúng ngữ nghĩa khi đã xác định đầy đủ cấu trúc của câu, vì thế bộ PTCP thường trả về cây phân tích đầy đủ cho bộ PTNN
- Cây phân tích thường có các thành phần sau:
 - Cấu trúc cây (có nhiều lựa chọn kiểu cấu trúc dữ liệu)
 - Cấu trúc một nút cây:
 - Kí hiệu ở nút hiện tại
 - Từ vựng liên quan (nếu là nút lá)
 - Danh sách các nút con
 - Thông tin bổ sung, phục vụ cho việc phân tích tiếp theo



Phần 5

Các bước xây dựng bộ phân tích cú pháp



Các bước xây dựng bộ PTCP

- Mô tả các luật văn phạm của ngôn ngữ nguồn
 - Các mô tả này ban đầu có thể ở dạng ngôn ngữ tự nhiên
 - Đặc tả ý nghĩa các kí hiệu không kết thúc (non-terminal)
 - Chuyển thành các luật văn phạm ở dạng chặt chẽ
- Phân tích bộ văn phạm để lựa chọn phương pháp phân tích cú pháp phù hợp nhất
 - Văn phạm có nhập nhằng hay không?
 - Văn phạm có đệ quy trái hay không?
 - Văn phạm có sự mơ hồ hay không?
 - Văn phạm có độ phức tạp ở mức độ nào?



Các bước xây dựng bộ PTCP

- Lựa chọn phương pháp phân tích cú pháp phù hợp
 - Xây dựng bộ PTCP trực tiếp (dành cho các ngôn ngữ có độ phức tạp đơn giản)
 - Xây dựng bộ PTCP 2 bước
 - Dựa trên văn phạm đầu vào, xây dựng automat đoán nhận
 - Sử dụng automat để xử lý dãy từ tổ từ PTTV
 - Xây dựng bộ PTCP vạn năng: trường hợp văn phạm quá phức tạp, có thể sử dụng các phương pháp phân tích vạn năng để xây dựng bộ PTCP
- Lựa chọn cách xử lý trong tình huống lỗi cú pháp, sinh các gợi ý sửa lỗi và các tình huống cần phải tổ hợp ngữ nghĩa



Phần 6

Bài tập và thảo luận



Thảo luận

1. *Có thể sử dụng biểu thức chính quy và automata hữu hạn để thực hiện việc mô tả ngôn ngữ và dựng cây phân tích được hay không?*
 - *Chỉ ra một vấn đề đơn giản nhất thường gặp trong các ngôn ngữ lập trình và không thể giải quyết bằng automata hữu hạn*
2. *Lựa chọn cấu trúc dữ liệu phù hợp cho:*
 - *Kết quả trả về của bộ PTCP (cây phân tích)*
 - *Lưu trữ các luật văn phạm*
 - *Cấu trúc của nút lá*
3. *Thiết kế prototype cho máy PTCP*
 - *Viết mã minh họa hoạt động của prototype trên*



Bài tập

- 1. Cho biểu thức số học gồm các số nguyên, phép toán và các cặp ngoặc. Ngôn ngữ L được xây dựng bằng cách xóa các thành phần của biểu thức chỉ giữ lại các ngoặc (đóng ngoặc và mở ngoặc)*
 - *Hãy xây dựng văn phạm mô tả ngôn ngữ L*
 - *Ngôn ngữ L có những đặc điểm gì?*
 - **Có bao nhiêu chuỗi w có độ dài 20 thuộc L?*
- 2. Viết bộ luật văn phạm câu tiếng Việt mô tả đồ vật*
 - *Vẽ cây cú pháp của câu “quyển vở đỏ màu xanh nhạt”*
 - *Vẽ cây cú pháp của câu “quyển sách to đỏ màu xanh”*