



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

# Dependency Parsing

Viện CNTT-TT  
Đại học Bách khoa Hà Nội

# Các nội dung chính

1. Tổng quan về Dependency Parsing.
  - a. Định nghĩa
  - b. Ứng dụng
  - c. Tính chất
2. Các phương pháp giải quyết bài toán.
  - a. Transition-based
  - b. Graph-based
  - c. Các cách tiếp cận hiện nay
3. Các kết quả cài đặt.

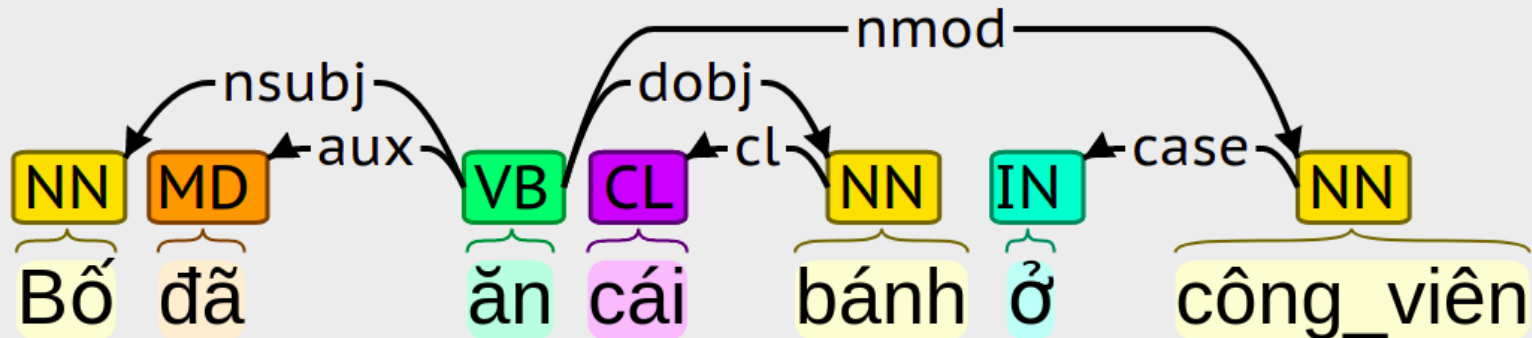
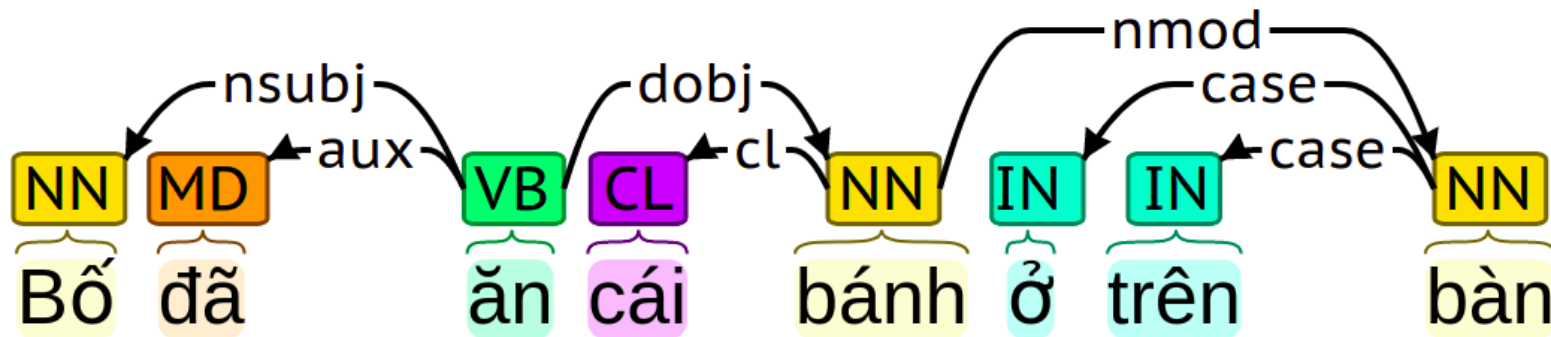
# Tổng quan về Dependency Parsing

- Định nghĩa
- Các ứng dụng của Dependency Parsing
- Các tính chất của cây cú pháp phụ thuộc

# Dependency Parsing là gì

- Phân tích cú pháp phụ thuộc
- Không phân tích chủ ngữ, vị ngữ, các cụm danh từ, cụm động từ,... thay vì đó, phân tích quan hệ phụ thuộc giữa các từ trong câu với nhau.
- Thường liên quan chặt chẽ đến bài toán Gán nhãn từ loại (Part Of Speech Tagging)
- Được bắt đầu quan tâm nhiều từ thập kỷ trước do sự giàu thông tin mà kiểu phân tích này mang lại.

# Dependency Parsing là gì



# Dependency Parsing là gì

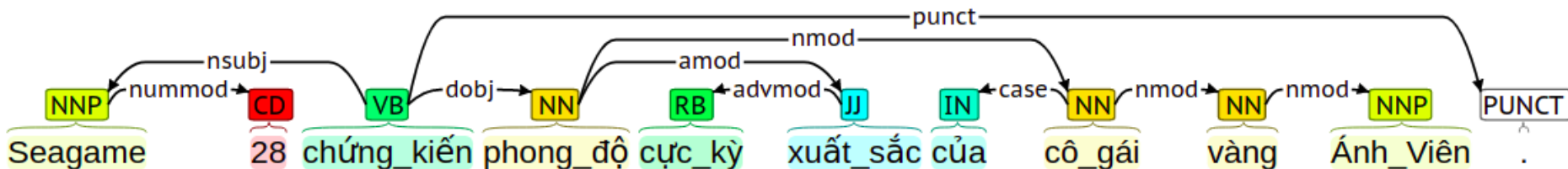
Một quan hệ phụ thuộc thể hiện bằng 1 mũi tên có hướng, trong đó:

- ❖ **head**: đầu không có mũi tên, là từ được bổ nghĩa
- ❖ **dependent**: đầu có mũi tên, là từ bổ nghĩa
- ❖ **label**: quan hệ phụ thuộc giữa 2 từ này.

# Các nhãn phụ thuộc

## ❖ Một số nhãn phụ thuộc:

- nsubj (Nominal subject): chủ ngữ, chủ thể
- dobj (Direct object): tân ngữ trực tiếp
- nmod (Nominal modifier): danh từ bổ nghĩa
- amod (Adjectival modifier): tính từ bổ nghĩa
- nummod (Numeric modifier): số từ bổ nghĩa

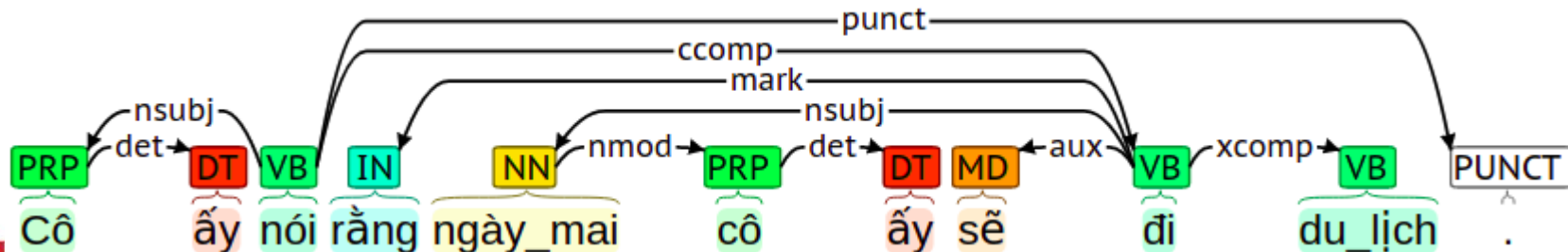


# Các nhãn phụ thuộc

## ❖ Một số nhãn phụ thuộc:

- ccomp (Clausal component): Mệnh đề thành phần
- xcomp (Open clausal component): Mệnh đề thành phần mở rộng
- aux (Auxiliary): phụ từ, trợ động từ

Xem thêm: <http://universaldependencies.org/u/dep/>



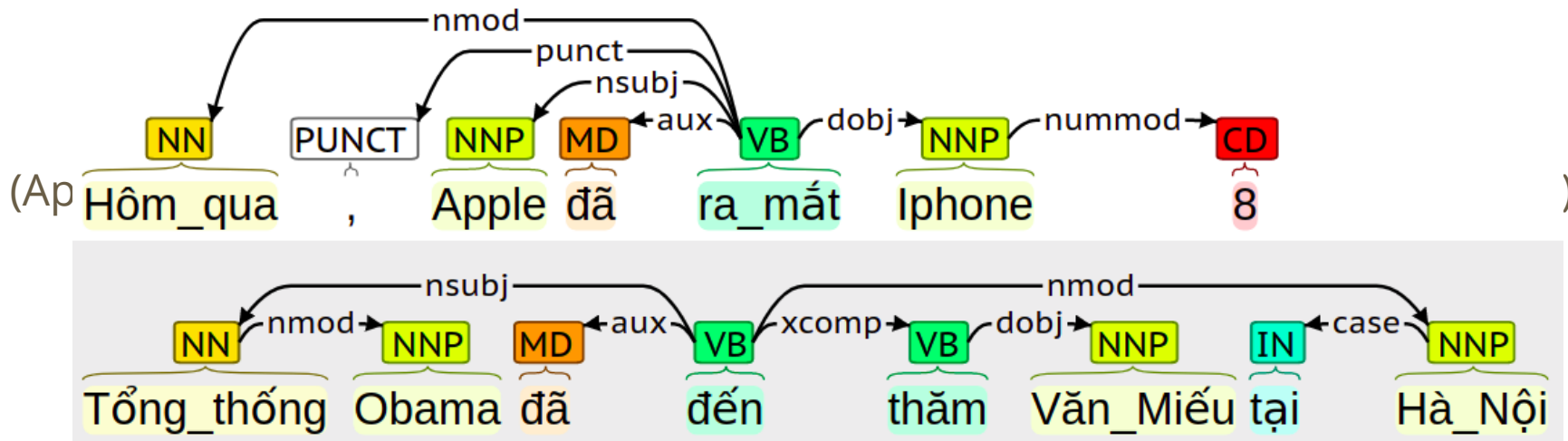


# Tổng quan về Dependency Parsing

- Định nghĩa
- Các ứng dụng của Dependency Parsing
- Các tính chất của cây cú pháp phụ thuộc

# Ứng dụng.

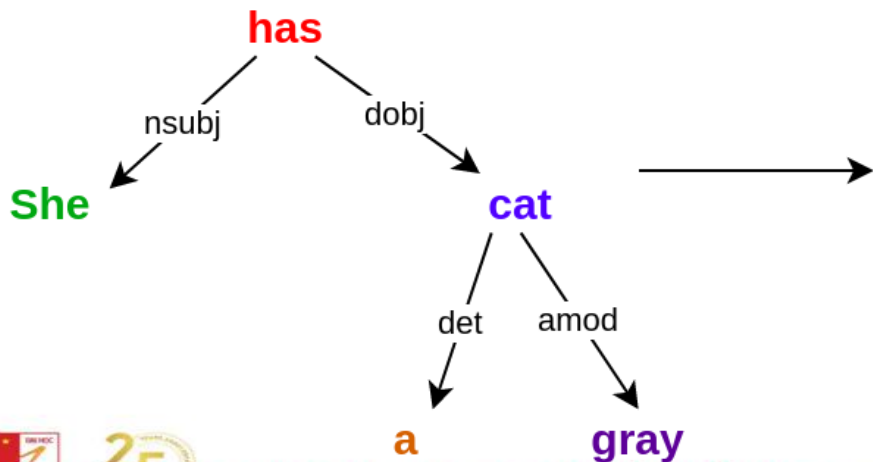
Xây dựng cơ sở tri thức dựa trên Trích rút quan hệ (Relation Extraction)



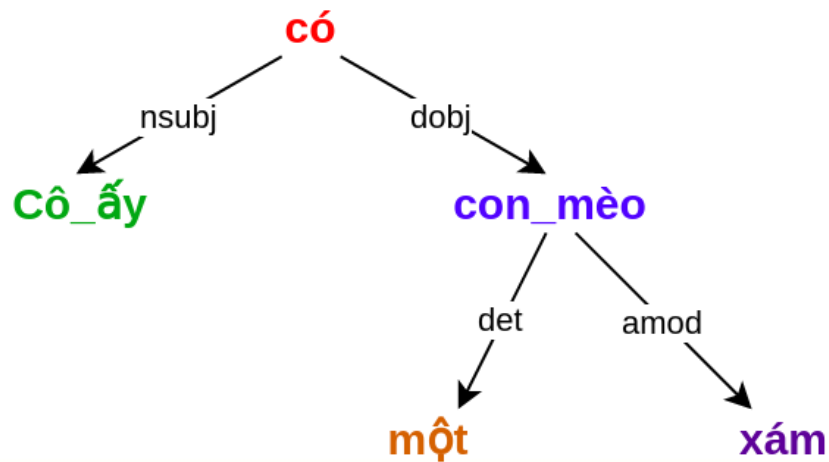
# Ứng dụng.

Dịch máy (Machine Translation)

She has a gray cat



Cô\_ấy có một con\_mèo xám



# Tổng quan về Dependency Parsing

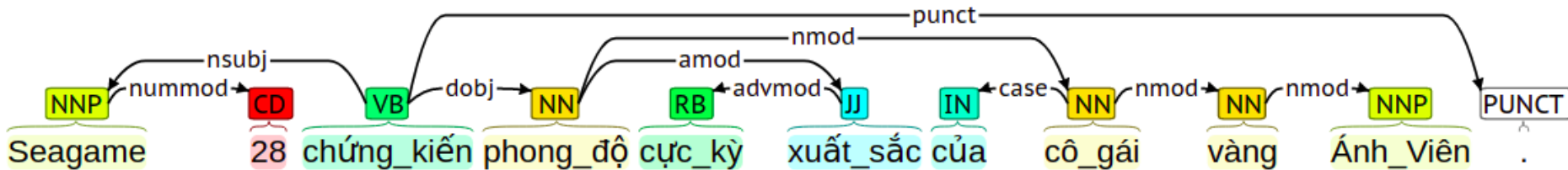
- Định nghĩa
- Các ứng dụng của Dependency Parsing
- **Các tính chất của cây cú pháp phụ thuộc**

# Các tính chất của cây cú pháp phụ thuộc

- ❖ Xét cây cú pháp là 1 đồ thị với các từ là các đỉnh (node), các quan hệ là các cạnh (arc)
- ❖ Đồ thị cú pháp phụ thuộc này có 4 tính chất:
  - Weakly Connected (Kết nối yếu)
  - Acyclic (Không có chu trình)
  - Single head (1 từ chỉ có duy nhất 1 head)
  - Projective

# Các tính chất của cây cú pháp phụ thuộc

- Weakly Connected:
  - Với mọi node  $i$ , luôn tồn tại 1 node  $j$  sao cho có 1 cạnh nối  $i \rightarrow j$  hoặc  $j \rightarrow i$
- Acyclic:
  - Nếu tồn tại cạnh  $i \rightarrow j$ , thì không thể tồn tại 1 đường đi  $j \rightarrow^* i$
- Single head:
  - Nếu có cạnh  $i \rightarrow j$ , thì sẽ không có cạnh  $k \rightarrow j$ , với  $k \neq i$

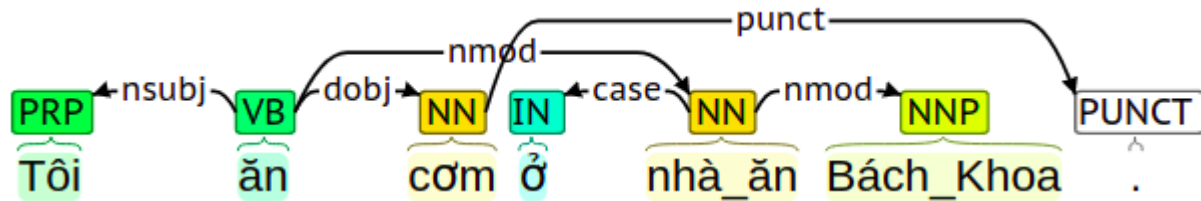
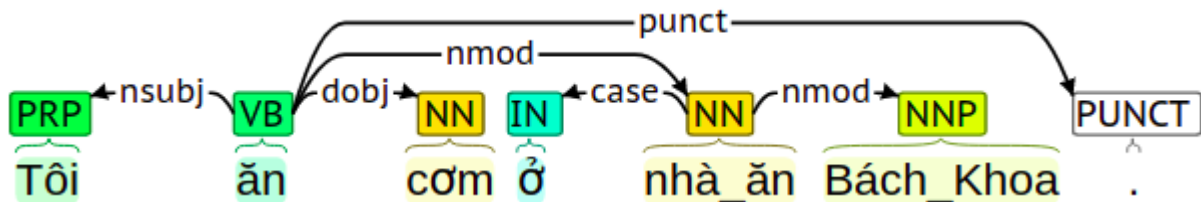


# Các tính chất của cây cú pháp phụ thuộc

- Projective: (tính chất này không bắt buộc)
  - Nếu tồn tại cạnh  $i \rightarrow j$ , thì với mọi  $k$  nằm giữa  $i$  và  $j$ , luôn có đường đi  $i \rightarrow^* k$
  - Một cách trực quan, không có cạnh chéo nhau khi vẽ cây cú pháp tuần tự theo câu

Projective

Non-Projective



# Các nội dung chính

1. Tổng quan về Dependency Parsing.
  - a. Định nghĩa
  - b. Ứng dụng
  - c. Tính chất
2. **Các phương pháp giải quyết bài toán.**
  - a. Transition-based
  - b. Graph-based
  - c. Các cách tiếp cận hiện nay
3. Các kết quả cài đặt.



# Các phương pháp giải quyết.

- **Transition-based**
  - Thuật toán Nivre
- Graph-based
- Các cách tiếp cận hiện nay
  - End to end learning
  - Joint learning

# Transition-based

- Ý tưởng cơ bản của nó là dựa trên các Transition (SHIFT, REDUCE, LEFT-ARC, RIGHT-ARC)
- Khi đọc câu từ trái sang phải, mô hình học máy của hệ thống sẽ quyết định thực hiện transition nào, dãy các transition này giúp xác định được quan hệ phụ thuộc giữa các từ trong câu
- Việc quyết định transition nào sẽ do 1 mô hình học máy (classifier) đảm nhận, cần huấn luyện mô hình này

# Transition-based

- Thuật toán phân tích cú pháp: Nivre, Covington, ...
- Phương pháp huấn luyện Classifier: SVM, Neural network, ...

Ở phần này, chúng ta tìm hiểu về thuật toán Nivre

# Thuật toán Nivre

- Là thuật toán phân tích cú pháp cho phương pháp Transition-based.
- Được Joakim Nivre đề xuất từ những năm 2003-2004
- Là thuật toán phân tích phổ biến nhất hiện tại trong hướng Transition-based, và cũng có nhiều biến thể khác nhau.
- Độ phức tạp  $O(n)$