Câu 1: Các mức độ xử lý ngôn ngữ tự nhiên?

* Phonology and phonetic : âm vị
* Morphology : hình thái
* Syntax: cú pháp
* Semantics: ngữ nghĩa.
* Pramatics: ngữ dụng
* Discourse: diễn ngôn

Một số lượng nhỏ các mô hình bao gồm: máy trạng thái(state machine), formal rule system, các mô hình logic và các mô hình xác suất được sử dụng để có thể phân tích và thu nhận các đặc tính trên của ngôn ngữ.

Câu 2: Tại sao xử lý ngôn ngữ tự nhiên lại khó:

Khó do sự nhập nhằng của ngôn ngữ. Nhập nhập do: từ loại, cú pháp, ngữ nghĩa. Cái gì cũng nhập nhằng.

Câu 3: Các đặc điểm chính của học bán giám sát(semi-supervised learning)? Khi nào thì dùng nó.

Sử dụng cả dữ liệu có nhãn và không có nhãn. Mục đích là tăng cường độ chính xác so với học có giám sát khi chỉ sử dụng dữ liệu có nhãn. Còn dùng dữ liệu như nào là tùy thuật toán.

Bài toán nào cũng có thể dùng không chỉ phân lớp hay phân cụm. Thường dùng khi dữ liệu có nhãn ít. Không đảm bảo xậy dựng được hệ thống có độ chính xác cao.