LING 120, Fall 2017: Language and Computers

Topic: Overview of Natural Language Processing

Instructor: Sowmya Vajjala

Iowa State University, USA

4 October 2017 (Week 7)

Class outline

- Solution to yesterday's problem
- Tasks in NLP: continuation
- Midterm preparation time

BrokEnglish! - solution

- Problem: `http://www.nacloweb.org/resources/problems/2011/E.pdf`
- Solution:`http://nacloweb.org/resources/problems/2011/ES.pdf`

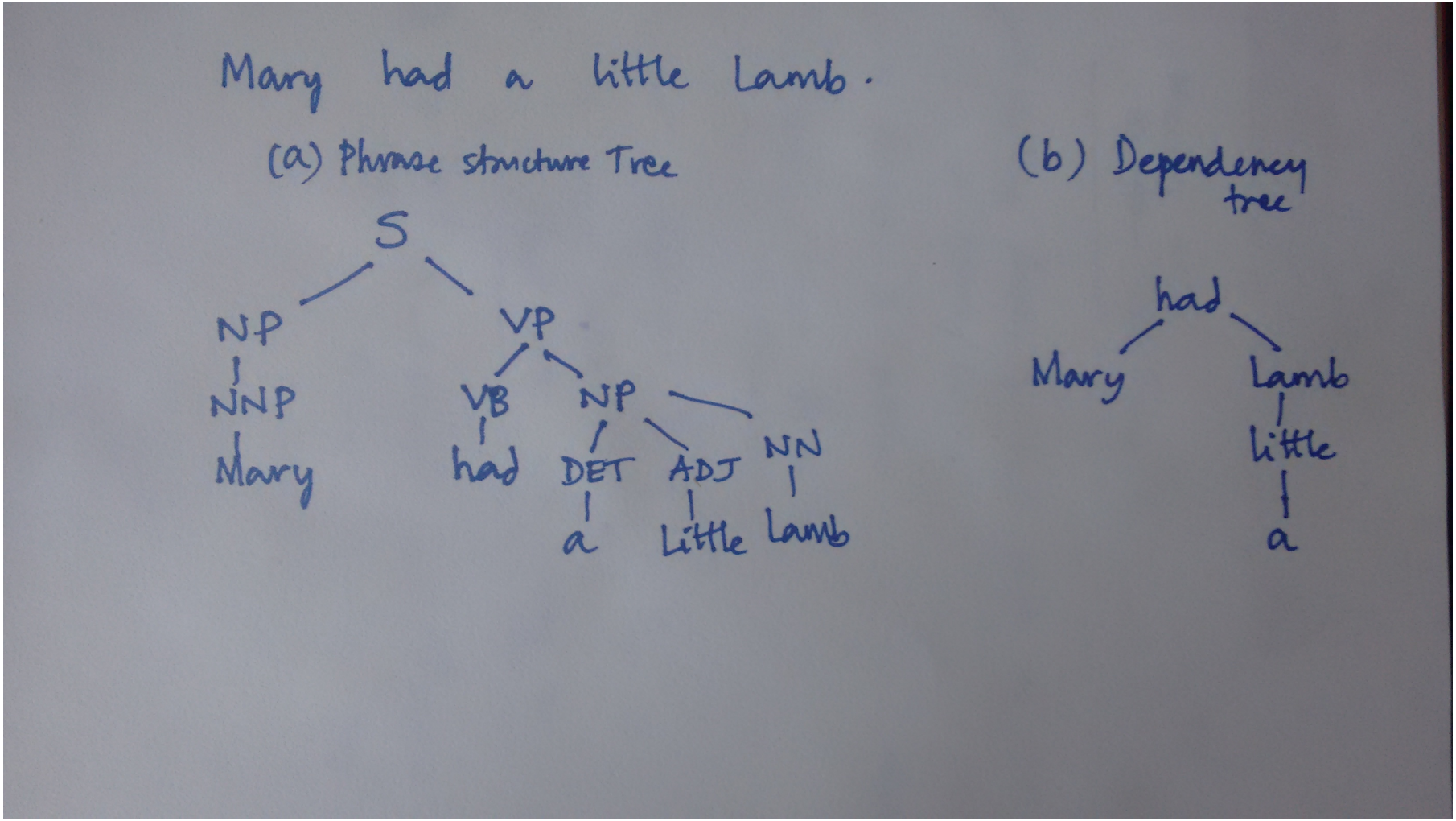
Madly Ambiguous: a web-game@OSU

http://madlyambiguous.osu.edu:1035/

NLP tasks: Parsing

- Task: Construct the syntactic structure of a given sentence.
- Two kinds of trees can be generated in NLP: Phrase structure tree (Constituency tree), Dependency tree
- PST: shows parse structure in terms of Noun Phrases, Verb Phrases, Prep. Phrases etc.
- Dependency Tree: shows relations between words in a sentence in terms of a pre-defined set of relations
- Useful to develop various applications such as question-answering systems (like Siri)
- Important note: POS tagging errors can carry over and affect parser efficiency.

NLP tasks: Parsing



(to be continued)

NLP tasks: Word Sense Disambiguation

- Task: For words that can have multiple meanings, what is the right sense of the word in a given sentence?
- Example: "Let us go inside, it is cold" vs "I have cold and cough"
- Very important for applications such as machine translation, information retrieval
- Good progress for English WSD. One of the active areas of research in the field.

NLP tasks: Named Entity Recognition

- Task: Identify and classify named entities (e.g., person names, organization names, locations etc.,)
- Application: Information extraction from text
- Some NER is domain specific (biomedical NER, financial NER etc)
- Current methods of NER: hand-crafted or automatically compiled lists + statistical machine learning models
- Active area of research for English and other languages.

NLP tasks: Semantic Role Labeling

- SRL is all about doing a "semantic parse" of a sentence. The task here is to identify argument structure of a sentence and thematic roles of different entities.
- Example: (source: `http://www.cs.upc.edu/~srlconll/`)

The following sentence, taken from the PropBank corpus, exemplifies the annotation of semantic roles:

[$_{A0}$ He] [$_{AM\text{-}MOD}$ would] [$_{AM\text{-}NEG}$ n't] [$_{V}$ **accept**] [$_{A1}$ anything of value] from [$_{A2}$ those he was writing about] .

Here, the roles for the predicate **accept** (that is, the *roleset* of the predicate) are defined in the PropBank Frames scheme as:

V: verb
A0: acceptor
A1: thing accepted
A2: accepted-from
A3: attribute
AM-MOD: modal
AM-NEG: negation

Midterm Preparation + Attendance

- Form into your mid-term groups, work on your presentations for next week
- Before you leave, post a 3 sentence summary of your presentation on Canvas, giving the team member names.
- That counts as your attendance for today.

Next class

- Conclusion of NLP Tasks overview
- Quick introduction to machine learning and its relevance for language processing
- Assignment 4 description
- Instructions for mid-terms
- (Probably) time for mid-term preps.