

510 Language project: English Sentence Syntax

I was originally doing my language around blackjack, but I had to keep stepping it back and wasn't able to get anything I was happy with. I was talking to my roommate about my struggle and while we were chatting it occurred to me, because my roommate is a linguistics major, that using grammar, i.e. spoken language grammar, is very very similar to the grammar we use in this class, which is by design. So scrapping my blackjack idea I chose English sentence syntax.

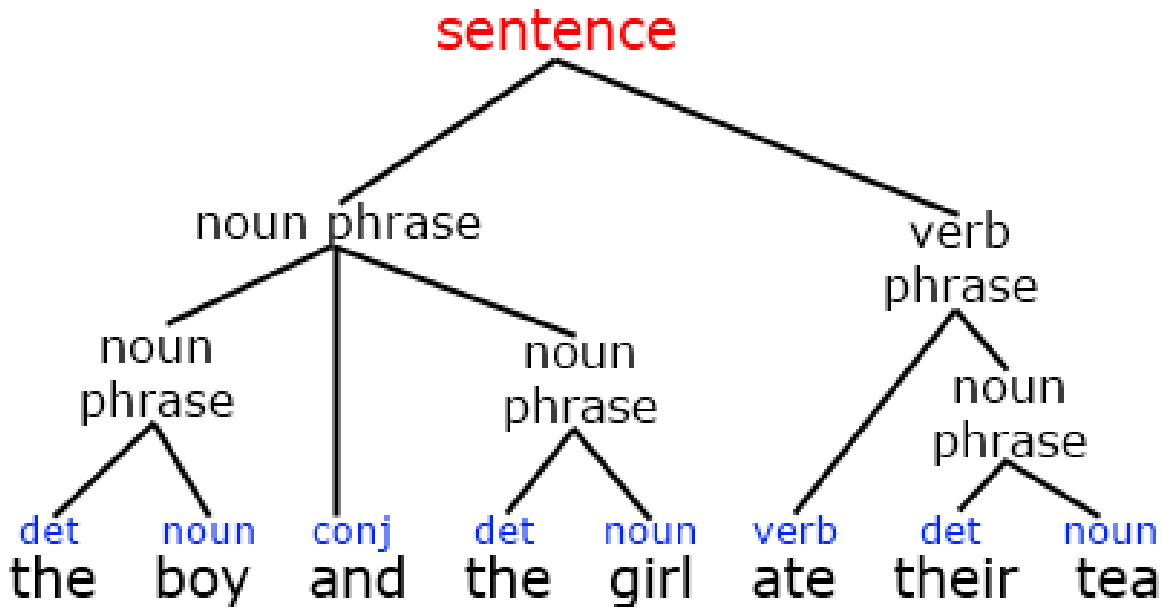


Figure 1: Example syntax tree

Above is what linguists call a syntax tree. While this is the most basic form of one it's the one that is parsible by someone who hasn't spent the last 2 years taking syntax classes. What it does is break down an English sentence into pieces that every language is made up of.

In the tree above you can see that these parts of speech, or in the case of my language our terminals, are determiners, nouns, conjunctions, verbs, and the one that is not listed is preposition and adverb. By finding the rules of how these pieces are organized we can create a grammar that can make any syntactical sentence. These also act like our alphabet, being

$$\Sigma = \{\text{noun, verb, det, adv, adj, prep}\}$$

Rules of Syntax

Sentences are made up of two main pieces, the noun phrase (NP) and the verb phrase (VP), from this we can create the most simple language out of the NP and VP being a noun (N)

and a verb (V), i.e. sam runs. While this is a valid sentence the meaning is not clear, this is what syntax is, it is only focused on what makes a valid sentence.

This translates very easily into a grammar, with Sentence being the starting state and then NP and VP next with their transitions and then adjectives and prepositional phrases having some transitions as well with everything else being a terminal.

$$\begin{aligned} S &\rightarrow NP VP \mid S \text{ conj } S \mid \text{adv } S \\ NP &\rightarrow \text{noun} \mid \text{det noun} \mid \text{det AdjP noun} \mid NP PP \mid \\ &NP \text{ conj } NP \\ VP &\rightarrow \text{verb} \mid \text{verb NP} \mid \text{verb NP PP} \mid \text{verb PP} \mid VP \\ &\text{conj VP} \mid \text{adv VP} \mid VP \text{ adv} \\ \text{AdjP} &\rightarrow \text{adj} \mid \text{adj AdjP} \\ PP &\rightarrow \text{prep NP} \end{aligned}$$

Figure 2: Grammar

This is the grammar I came up with. I originally had actual words for more of the terminal states but thought that the grammar doesn't really cover all English sentences if there is only a limited amount of words. A sample valid string for this would be:

**Det adj adj noun conj noun prep det noun adv verb
conj verb**

This into an actual sentence would be something such as: "The large black dog and cat near the house quickly barked and ran."

PDA

Converting the grammar to a PDA leaves us with a two state machine much like one we saw in class (using LaTeX).

