Phrase Structure Analysis of Sentences

Introduction to NLP, ACS 2014, Assignment 3 © Ted Briscoe (`ejb@cl.cam.ac.uk`) GS18
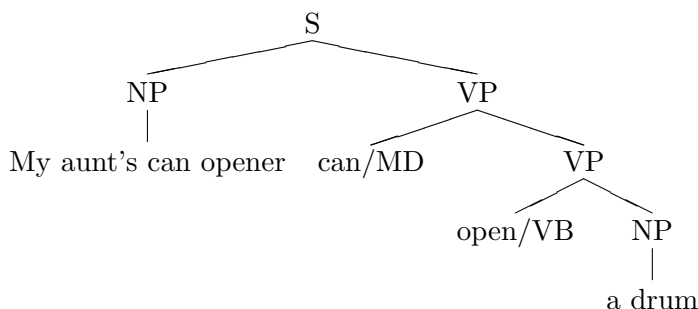
1   Task

Choose 8 sentences from the set below and draw a phrase-structure tree (PST) using non-terminal labels (VP, AP etc). Base your non-terminal labels on those used in the handout (see also Jurafsky and Martin ch12, p422f, p428f and other references on handout). You can invent your own labels for constituents motivated by distributional analysis as necessary.

You can simplify the analysis of NPs e.g: NP(the man who I like). For instance the PST analysis of the first two NPs in:

My aunt's can opener can open a drum

should look something like this:

Write up / draw your answers and hand them in before the following Friday session. Include BRIEF notes on any difficulties or issues you had with specific cases. It is more important to understand and be able to explain your reasoning than to get every constituent right. Be prepared to discuss the difficult cases during the session. Please feel free to work on the task in groups, but the final selection of sentences and their analyses should be your own. I'd recommend trying some new sentences in this assignment, especially if you are finding the analyses easy. Hand in a copy to Student Admin by the deadline, but also bring one to the lectures.

2   Sentences

(1) English also has many words of more or less unique function, including interjections (oh, ah), negatives (no, not), politeness markers (please, thank you), and the existential 'there' (there are horses but not unicorns) among others.

(2) Making these decisions requires sophisticated knowledge of syntax; tagging manuals (Santorini, 1990) give various heuristics that can help human coders make these decisions and that can also provide useful features for automatic taggers.

(3) The Penn Treebank tagset was culled from the original 87-tag tagset for the Brown Corpus. For example the original Brown and C5 tagsets include a separate tag for each of the different forms of the verbs *do* (e.g. C5 tag VDD for *did* and VDG tag for *doing*), *be* and *have.*

(4) The slightly simplified version of the Viterbi algorithm that we present takes as input a single HMM and a sequence of observed words $O = (o_1, o_2, ...o_T)$ and returns the most probable state/tag sequence $Q = (q_1, q_2, q_T)$ together with its probability.

(5) Thus the EM-trained "pure HMM" tagger is probably best suited to cases where no training data is available, for example, when tagging languages for which no data was previously hand-tagged.

(6) Coming home from very lonely places, all of us go a little mad: whether from great personal success, or just an all-night drive, we are the sole survivors of a world no one else has ever seen.

(7) Skill without imagination is craftsmanship and gives us many useful objects such as wickerwork picnic baskets. Imagination without skill gives us modern art.

(8) An MoD spokesman said: "Surveys of Astute have now been completed and she will proceed to Faslane under her own power. She is being escorted by tugs and HMS Shoreham."

(9) But far fewer people fully understand how the Media Lab operates, fits into MIT, and encourages such a creative environment; about half of the anniversary celebration's program focused on simply defining what the Media Lab is.

(10) Instead of constantly worrying about funding, the faculty and students can focus on their project, with the exception of sponsors' weeks, when they have to convince companies to start or continue their support.

(11) The doctors are warning that the NHS cannot make the 20bn of savings by 2014 that ministers expect, while simultaneously undertaking a huge reorganisation that will see England's 152 primary care trusts (PCTs) abolished and consortiums of GPs assume responsibility for the commissioning of services for patients.