

2015 LSA Summer Institute: Computational Psycholinguistics

Homework #2

Klinton Bicknell (Northwestern) and Roger Levy (UCSD)

Friday 10 July 2015

Due: Tuesday 14 July 2015

1. Japanese is a strictly verb-final language with null pronouns and considerable word-order freedom. Let's consider the following "toy" fragment probabilistic context-free grammar (PCFG) of Japanese. In this fragment, case annotations such as NP[nom] simply distinguish among atomic categories – so NP[nom] and NP[acc] are simply two different categories in the grammar.

0.4	$S \rightarrow \text{NP}[\text{nom}] \text{NP}[\text{acc}] V$	1.0	$\text{NP}[\text{nom}] \rightarrow \text{NP } ga$
0.3	$S \rightarrow \text{NP}[\text{nom}] V$	1.0	$\text{NP}[\text{acc}] \rightarrow \text{NP } o$
0.1	$S \rightarrow \text{NP}[\text{acc}] \text{NP}[\text{nom}] V$	0.5	$\text{NP} \rightarrow \textit{Tanaka-san}$ ('Mr. Tanaka')
0.2	$S \rightarrow \text{NP}[\text{acc}] V$	0.5	$\text{NP} \rightarrow \textit{Ota-san}$ ('Mr. Ota')
		0.5	$V \rightarrow \textit{deta}$ ('went out')
		0.5	$V \rightarrow \textit{yonda}$ ('called')

- Given the input prefix *Tanaka-san ga*, what is the probability that the next element that the comprehender encounters in the input will be the clause-final verb?
- Compute the probability that the next element will be the clause-final verb for the input prefixes *Tanaka-san ga Ota-san o*, *Ota-san o*, and *Ota-san o Tanaka-san ga* as well.
- Let us call the first NP in a sentence NP_1 and the second NP in the sentence, if one appears, NP_2 . In each of the nominative-initial and accusative-initial examples, how many **bits of surprisal** that would be associated with encountering the clause-final verb immediately after encountering NP_1 are reduced by encountering NP_2 first? Remember that surprisal is negative log probability, or log inverse-probability:

$$\begin{aligned} \text{Surprisal of } x \text{ in context } C &= -\log_2 P(x|C) \\ &= \log_2 \frac{1}{P(x|C)} \end{aligned}$$

so the bits of surprisal reduced by NP₂ would be

$$\log_2 \frac{1}{P(x|NP_1)} - \log_2 \frac{1}{P(x|NP_1 NP_2)}$$

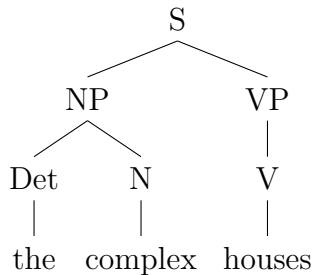
- Could you imagine a reasonable PCFG in which adding a pre-verbal constituent would ever *increase* the amount of surprisal associated with the final verb? Justify your answer.

2. (a) Consider the following PCFG:

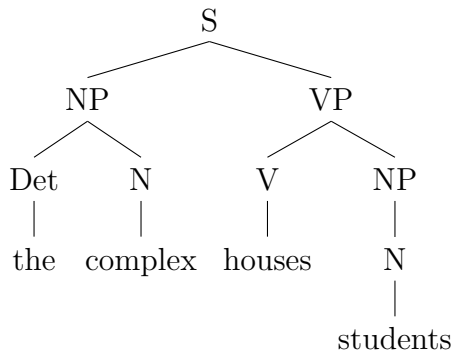
1.0 S → NP VP	1.0 Det → the
0.8 NP → Det N	0.5 N → children
0.2 NP → NP PP	0.5 N → parents
1.0 PP → P NP	1.0 P → of
1.0 VP → V	1.0 V → exist

For a sentence that starts with the word *the*, what is the probability that the second word is *children*?

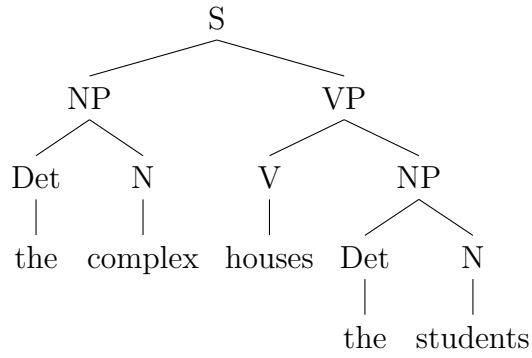
- (b) Now recall that an **incremental tree** in the sense of Jurafsky (1996), as we covered in class on Thursday 10 July, is a completely connected tree that exhaustively specifies the syntactic content connecting all the input encountered thus far to the root of the tree, but does not commit to any syntactic content exclusively covering input that has not yet been encountered. For example, the following incremental tree:



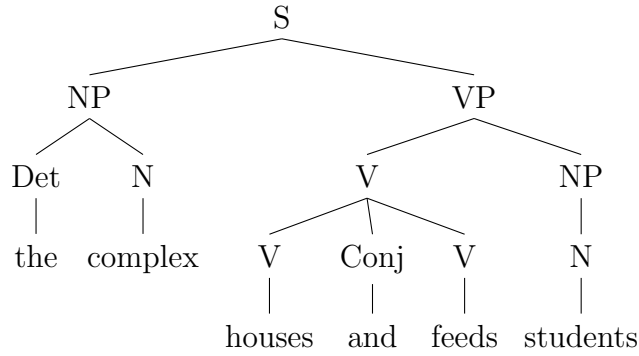
would be consistent with complete-tree elaborations (based on further input) such as:



and



but not complete-tree continuations such as



The probability of an incremental tree is the sum of the probabilities of all the complete trees consistent with it.

Question: How many incremental trees, in the sense of Jurafsky (1996), are consistent with the partial input *the ...*? with the partial input *the children ...*?