Ling 473 Assignment 1

Daniel Campos dacampos@uw.edu

07/26/2018

1 Why Computational Linguistics

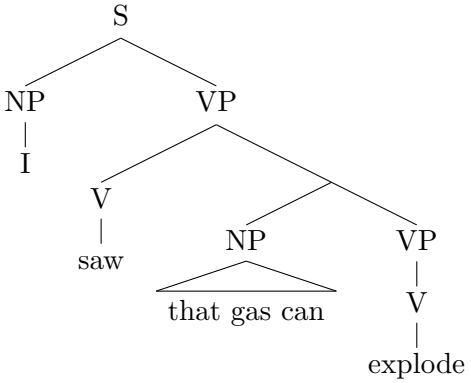
I became interested in computational linguistics in the spring of 2014 when I took a break from college to work. Having just had my mind blown by my first Machine Learning class I stumbled my way into the research team at Basis Technology working on Multi-lingual Entity Resolution. Getting to build an incredible product that mplemented linguistics fundamentals fascinated me. When I was back at school, I took some introductory NLP and linguistics courses which further peaked my interest. In Fall 2015 I started working at Microsoft and soon joined a team working on software translation where I implemented and built various Machine Translation systems. That project was challenging and forced me to dive into the literature to understand how MT works. After that project was implemented I joined the Core Relevance team at Bing working on Machine Reading Comprehension (MSMARCO is a data set I work on)and search metrics. My current work and the literature I have encountered so far has me interested in Question Answering and Natural Language Generation but I also am really interested by morphology and sociolingustics. Honestly I haven't read about a sub field in computational linguistics that hasn't interested me.

2 Phrase Structures

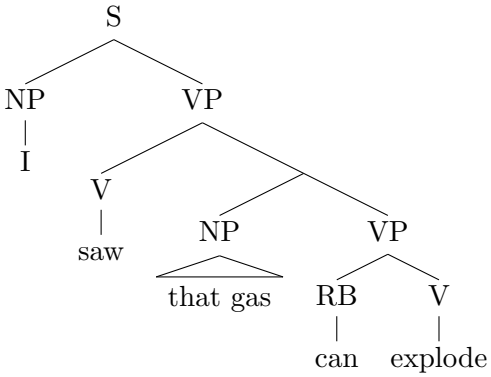
Consider the following sentence: I saw that gas can explode.

2.1 How many phrase structure trees can you find for this sentence?

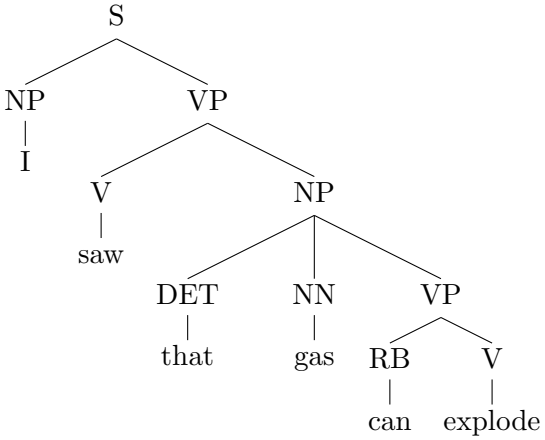
I see 3 different phrase structures that revolve around treating the sub phrase 'that gas can'. FYI i really struggled to get the qtree library to compile so my trees are a little shoddy.



This phrase structure is referring to 'that gas can' as a specific entity meaning this phrase structure read as someone saw a specific gas can explode.



This next phase structure takes the other meaning of can(as in possible) but keeps that gas as a NN which can be best understood as saying a specific 'gas' is able to explode.



This final structure separates 'that gas can' into 3 separate concepts. That ceases to be a DET and become a qualifier. This whole sentence can be understood as someone saw that gas(as a universal substance) is able to explode.

2.2 Write using Penn Treebank notation.

First Structure

(S (NP I) (VP saw (NP that gas can)(VP explode)(. .)))

Second Structure

S (NP I)(VP saw (SBAR (S (NP that gas) (VP (RB can) (VP explode)(. .))))))
Third Structure
((S (NP I)(VP saw (SBAR that(S (NP gas)(VP can (VP explode)(. .))))))

3 Word Creation

How many six-letter "words" can be formed from the alphabet a – z ? A "word" for this question must have at least one vowel a e i o u , and may not contain all vowels. Show your work and explain your answer.

First let c(i) stand for combinations with i amount vowels. Since we don't have to worry about repeating vowels we can just treat letter as two classes, vowels and consonants. Using this information we create our equation

$$combinations = total - c(0) - c(6) \tag{1}$$

$$Combinations = 26^6 - (\binom{6}{0}(21^6)(5^0)) - (\binom{6}{6}(21^0)(5^6)) \tag{2}$$

$$Combinations = 308915776 - 85766121 - 15625 = 223134030 \tag{3}$$

4 Character Arrangement

How many ways can the characters in the following tuple be arranged? 葡萄葡萄橙橙苹梨蕉 This is just like our MISSISIPI Problem. 葡 occurs 4 times, 橙 occurs 2 times and each of 苹梨蕉 occurs once.

$$Arrangements = (9!)/(3!2!1!1!1!) = 362880/12 = 30240 \tag{4}$$

5 Document Processing System

Consider a document processing system which performs pairwise comparisons and a corpus containing 19 documents as follows:

Topic	Count
Conference Proceedings	7
Journal Articles	9
Workshop Abstracts	3

5.1 How many pairwise comparisons are possible between documents on the same topic?

let C(x) represent the comparisons possible

$$C(ConferenceProceedings) = \binom{7}{2} = 21 comparisons \tag{5}$$

$$C(JournalArticles) = \binom{9}{2} = 36 comparisons \tag{6}$$

$$C(WorkshopAbstracts) = \binom{3}{2} = 3comparisons \tag{7}$$

5.2 How many pairwise comparisons are possible between documents on different topics?

Let J represent Journal articles, W represent Workshop abstracts, and C represent Conference proceedings)

$$Comparisons = CP = C(J\&W) + C(J\&C) + C(W\&C) \tag{8}$$

$$CP = \binom{9}{1} * \binom{3}{1} + \binom{9}{1} * \binom{7}{1} + \binom{3}{1} * \binom{7}{1} \tag{9}$$

$$CP = 9 * 3 + 9 * 7 + 3 * 7 = 27 + 63 + 21 = 111 \tag{10}$$

6 Extra Credit Unordered sets with repetition

Since we are dealing with unordered sets with repetition we can simplify the initial problem from producing the number of unordered sets to how many combinations of k items can be formed from N items where repetition is allowed. In other words since we are going to choose k items from n n becomes N+K-1.

$$C(N, K) = (N + K - 1)!/(K!(N - 1)!) \tag{11}$$