Nirdesh Bhandari
CS: 380 Theory of Computation
<u>Christopher Culy: The Complexity of the Vocabulary of Bambara</u>

A simpler way to understand context free grammar is to think about the core of vocabulary construction. When we create a word in English we take a root (lets use the example verb: button) and then we can either add suffixes or prefixes (lets add 'un' to create: unbutton) and then depending on the tense it could we can add a suffix (lets add 'ed' to create: unbuttoned) to create a vocabulary word. In English we have a finite set of rules governing what can be created out of roots and how many/what suffixes or prefixes can be added. We can say that the word generation from vocabulary in English can be confined to these specific sets of rules. This quality of being able to use these rules to create correct words is what we mean when we say that the generative capacity of vocabulary of a language is context free.

It is based on this (assumption) that vocabulary can be correctly generated by these rules that we form sentences structures (or syntax). We look at how the components of a sentence-the noun, verb, subject, participles- must be within a defined order or sets of order for the sentence to be a correct sentence belonging to the English language.

 This paper by Christopher Culy looks at Bambara, a language spoken by the people of Mali, and proves that the vocabulary of a language is a language in itself and cannot be taken as context free. It gives the readers examples of word formations within Bambara that show us that Bambara, unlike English, indeed has infinitely many possible vocabulary words because certain structures/arrangement of words follow their own rules in vocabulary creation rather than the conventional rules for adjacent lexical items.

As a native Nepali speaker, I completely agree with the Culy in his interpretation that the vocabulary of a language need not be context free. What makes Nepali uniquely non-context free, in comparison to Bambara, is its sentence structure as well as its vocabulary. The noun, pronouns, adjectives and the verbs depend inextricably with each other and must be in par with the gender, number, tense, mood of the situation as well as the status of the person you are speaking to or the subject[3]. This gives rise to numerous sentence structures, so much so complex that we have a completely different set of morphemes attached to verbs when speaking about about the royal family. Furthermore, the fact that the Nepali alphabet is composed of 36 consonants and 26 vowels (or vowel diacritics)- with numerous rules governing word construction- only adds to the complexity[2].

My claim that Nepali vocabulary, like Bambara is also non-context free, is based not only in the fact that word construction later in the sentence depends on a variety of factors expressed as the sentence is formed, but, also by the existence and placement of interjectory words[1] that sometimes completely change the construction of all words that follow or precede it. These are words with no

meanings in themselves but rather change the tone of the sentence of brings emphasis to certain things being discussed. A lot of times there is no rule governing syntactical the placement of these words other than tonal correctness (and their relation with other tonal words linked specifically to certain nouns and/or verbs). Thus, Nepali vocabulary too, like Bambara, cannot be boiled down to context free grammar.

This implication that the vocabulary of a language can in fact be infinite rolls over to the syntax creating/checking capabilities to determine the correctness of a language. In case of Bambara Culy showed us how certain rules can work on a recursive pattern and therefore change the rules for syntaxes for a language (which we previously believed could be boiled down to context be free language). But if this were to be a problem then modern text editors that use the rules of vocabulary and grammar to check the correctness should not exists for these languages. But they do!

So, how then have modern text editors gotten around this problem of infinite vocabulary to create all sorts of complex linguistic tools like correction checkers and Google-Translate for these languages? Gerald Gazdar[4], from the University of Sussex believes that recursive patterns Culy mentioned are not a problem because one can simply modify the CFL parser (think of this as a device that creates/analyzes the syntax by knowing all the rules) to include a simple new rule that looks at whether the first half of the syntax is the same as the second half or not, and if found creates and replaces a new word (in accordance to those 'peculiar' rules).

Indeed, this has been done in the case of many languages. One particular example I would highlight is about the use of a Predictive Parser on Bangala[5] (a language very similar to Hindi and Nepali) to recognize grammar. Another solution that modern linguists have come up is using computers to create Probabilistic Context Free Grammar (PCFG) for a language. This has been done for Hindi[6] works by looking at previous data to narrow down to a probable parser. Computers have in a way found a shortcut around infinite vocabulary to some extent.

However, the tools and methods I talked about earlier do not contradict Culy's findings but rather implement a practical application based on the knowledge that a vocabulary can be infinite. Once I meet Culy, I would definitely ask him a question as to what he thinks about these applications that computers use.

In conclusion, I believe Christopher Culy's work was insightful in helping the readers think about language in a more technical way. I think he changed the way linguists had thought of language (particular English), and by using the example of Bambara, forced them to change their flow of logic when it comes to analyzing the correctness of a natural language. Even though we now have computer models that can(almost) accurately predict correctness based on previous data, Culy's work in essence forced the then intellectual(English speaking) linguists to look at language from a different view.

SOURCES:

1. "Interjectory Words." *Nepalgo*. N.p., n.d. Web. 4 Nov. 2016.

   http://nepalgo.tumblr.com/post/104076063965/interjectory-words

2. "Nepali Alphabet, Pronunciation and Language." N.p., n.d. Web. 4 Nov. 2016.

   http://www.omniglot.com/writing/nepali.htm

3. "Nepali Grammar." *Wikipedia* 23 Oct. 2016. *Wikipedia*. Web. 4 Nov. 2016.

   https://en.wikipedia.org/wiki/Nepali_grammar

   Side note : Nepal is a small Himalayan country with a population only 27million people but a staggering 123 unique languages spoken among different ethnic groups in it. '

4. Gazdar, Gerald. "Computationally Relevant Properties of Natural Languages and Their Grammars." N.p., 25

   Feb. 1985. Web. 4 Nov. 2016.   (pg 6 prg 2)

https://pdfs.semanticscholar.org/45c8/7de29ade64da33326ccc428ec2f2a27d48da.pdf

   SIDE notes : http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog2302_2/full
   A paper about how the human brain processes these recursive patterns in a probabilistic manner. (noun o noun o noun o noun o noun o noun)

5. K. M. Azharul Hasan, Al-Mahmud, Amit Mondal, Amit Saha. "RECOGNIZING BANGLA GRAMMAR

   USING PREDICTIVE PARSER." N.p., Dec. 2011. Web. 4 Nov. 2016.

https://arxiv.org/pdf/1201.2010.pdf

6. Harsh Verma. "Probabilistic Context Free Grammar For Hindi." N.p., n.d. Web. 4 Nov. 2016.

http://www.cse.iitk.ac.in/users/cs365/2008/2008-Reports/Verma-final.pdf