The Chinese University of Hong Kong

Department of Philosophy

Master of Philosophy

"The Formal Properties of Natural Language Syntax"

by

LI, Chi Ho

May 1997

Dedicated to

Eric Cantona,

the forever King of Manchester United

Abstract

The master thesis you are holding originates from my study on the foundation of generative grammar. Syntactic investigations comprise two aspects: the substantive aspect and the formal aspect. The substantive aspect includes the hypothesis of various syntactic rules/principles, the collection of data in support of those rules/principles, and their revisions in the face of counter-examples. Undoubtedly such work is what most linguists have been doing. However, syntactic rules/principles are not as arbitrary as people like them to be, it seems that there are some conditions or formats which are observed by those rules/principles. The discovery of such conditions constitutes what I call the formal aspect of syntactic study, and it is an important area in philosophy of language.

What are these conditions? Historically the most plausible candidates of these conditions have been represented in various grammar formalisms, the context-free grammar (CFG), the context-sensitive grammar (CSG) and some other grammar formalisms with their generative capacities lying somewhere between those of CFG and CSG. The generative capacity of a grammar formalism means the types of sentences/symbol strings generated by this formalism. So, the task of the formal study on syntax, or *mathematical linguistics*, is to figure out the correct formalism which generates and only generates natural language sentences.

Note that there are indeed two senses of the notion "generative capacity": *strong* generative capacity and *weak* generative capacity. In brief, weak generative capacity cares whether a grammar formalism generates all grammatical sentences while strong generative capacity cares whether a formalism generates all grammatical sentences with *correct structural descriptions*. The survey on weak generative capacity has long been humiliated as an uninteresting and unimportant inquiry. "The study of weak generative capacity is of rather marginal linguistic interest. It is important only in those cases where some proposed theory fails even in weak generative capacity." [1] Chomsky suggested that some "fairly elementary theories" like CFG had been proven to be unable to (weakly) generate natural language sentences. Yet the fact is not as simple as he thought. The aim of writing this paper is to evaluate all those arguments against the adequacy of CFG with respect to weak generative capacity; and I will show that almost all these arguments are unsound. However, I will also show that Chomsky is really a linguistic "prophet", because there is a linguistic phenomenon which is beyond the scope of CFG. I will also point out that in a "deeper" sense, the study of weak generative capacity is not only of marginal interest; it is of no theoretical interest at all.

[1] Chomsky (1965) p.60.

Content

The Formal Properties of Natural Language Syntax

1. Introduction

There are many linguistic phenomena that philosophers of language puzzle about. Among them there is the question: why are human beings, endowed only with finite memory and processing resources, able to understand and generate an *infinite* number of sentences, each of which bears a different meaning to others? A very plausible answer is that human beings memorize a finite set of vocabulary and a finite set of syntactic rules, then by combining the lexical items using the syntactic rules, we can generate an infinite number of sentences, just like a logician derives an infinite set of theorems from a finite set of axioms by applying a finite set of inference rules. The process of sentence understanding is done similarly: human beings parse a sentence into constituent phrases and words in accordance with the syntactic rules, then get the meaning of the whole sentence from the meanings of the words by employing those rules again.

To give a simple example showing how syntax makes way for semantics. Assume we have the following semantic rules as a part of our linguistic competence:

(i) the semantic value of '*John*' is an entity with the name 'John';
(ii) the semantic value of '*Mary*' is an entity with the name 'Mary';
(iii) the semantic value of '*loves*' is the relation of loving;
(iv) the semantic value of $[_S$ NP VP$]$ is *true* if and only if

($\exists$ x, p) (the semantic value of NP is x, that of VP is p, and x has p);

(v) the semantic value of $[_{VP}$ V NP$]$ is p if and only if

($\exists$ r, y) (the semantic value of V is r, that of NP is y, and p is the property of being an x such that x bears r to y);

(vi) the semantic value of $[_A$ B$]$ is x if and only if x is the semantic value of B, where A, B stand for any syntactic categories.

Now let us consider the way by which we understand the sentence '*John loves Mary*'. In order to do so, we must first determine its syntactic structure to be

$[_S [_{NP}$ John$] [_{VP} [_V$ loves$] [_{NP}$ Mary$]]]$.

Then according to the 6 semantic rules listed above we can figure out its meaning as:

the entity with the name 'John' bears the relation of loving to the entity with the name Mary.

By the same set of syntactic and semantic rules, the semantic interpretations of sentences '*Mary loves John*', '*John loves John*', '*Mary loves Mary*' can also be fixed. Of course, mere these several rules are insufficient to deal with the various subtleties of any natural language; yet this example illustrates that the modular approach to language is at least possible.

Adopting this distinction between the semantic and syntactic aspects of language understanding and generation, we are in a position to conceive *grammar* as a formal device manipulating the words as meaningless formal symbols. Mathematically speaking, a grammar of a language is a quadruple $<V_T, V_N, S, R>$, where

(i) V_T is the *terminal alphabet*; it is the set of the words that appear in all the sentences of that language;

(ii) V_N is the *non-terminal alphabet*; it is the set of the symbols that do not appear in the sentences but in the *derivations* of those sentences;

(iii) S is the start symbol; all the derivations of sentences are initiated with this symbol;

(iv) R is the set of all syntactic rules; each of which rewrites a non-terminal symbol into something else.

The derivation of a sentence is a sequence of strings $x_1, x_2, \cdots, x_n$ ($n \geq 1$) such that $x_1 = S$ and for each x_i ($2 \leq i \leq n$), x_i is obtained from x_{i-1} by applying one rule in R. A (grammatical) sentence is the last string in a derivation, and the set of all sentences is called the language generated by this grammar.

The goal of syntactic investigations is to find out the adequate grammars of all natural languages. Here we should notice that there are two senses of an adequate grammar. A *weakly adequate grammar* need merely generate all and only the grammatical sentences, whereas a *strongly adequate grammar* is a weakly adequate grammar which also generates the correct *structural descriptions* of those grammatical sentences. A structural description of a sentence is its constituent structure as assigned by a grammar. For example, if grammars G_1 and G_2 both generate the sentence *John loves Mary* but different phrase structures: $[_S [_{NP}$ John$]$ $[_{VP}$ loves Mary$]$ $]$ by G_1 and $[_S [_{XP}$ John loves$]$ $[_{NP}$ Mary$]$ $]$ by G_2, then we say both G_1 and G_2 are weakly adequate but only G_1 is strongly adequate. As the preface states, this thesis will focus only on the weak adequacy.

Of course, the characterization of the (weakly adequate) grammar(s) of each natural language is subject to linguists' empirical investigations. Yet we must notice that there is some universal constraint for all natural language grammars. Noam Chomsky introduced four kinds of grammar formalisms, each of which is distinguished from others with respect to the format of, and hence the constraint on, syntactic rules. They are

1) Unrestricted rewriting system or type-0 grammar:

the left side of each syntactic rule must contain at least 1 non-terminal symbol

2) Context-sensitive grammar or type-1 grammar: each rule is of the form $\alpha A\beta \rightarrow \alpha\psi\beta$, where $A \varepsilon V_N$, α, β, $\psi \varepsilon (V_T \cup V_N)^*$ and ψ is not an empty string.[1]

3) Context-free grammar or type-2 grammar:

each rule is of the form $A \rightarrow \psi$, where $A \varepsilon V_N$, and $\psi \varepsilon (V_T \cup V_N)^*$.

4) Regular grammar, right linear grammar or type-3 grammar:

each rule is of the form $A \rightarrow x B$ or $A \rightarrow x$, where A, $B \varepsilon V_N$, and $x \varepsilon V_T$.

These four grammar formalisms, constituting the so called *Chomsky hierarchy*, exhibit an increasing restriction on the syntactic rules. Type-0 grammar allows any kind of rule. Type-1 grammar is a special case of type-0 grammar with the restriction that the right hand side of each rule must contain no less symbols than the left hand side.[2] Rules in type-2 grammar are special cases of type-1 grammar rules with α, β being empty. Type-3 grammar rules are also special cases of type-2 grammar rules with ψ being x B or x. Correspondingly, the languages generated by these formalisms exhibit decreasing sizes, since the more restrictive a formalism is, the lesser types of rules are permitted; that means lesser types of sentences are generated.

Hence there is a problem both for philosophers and linguists: **which type of grammar formalism is most suitable for providing (weakly) adequate grammars for all natural languages?** That is to ask, **to what extent should natural language syntactic rules be restricted?** The unrestricted rewriting system is obviously implausible to be a candidate, as it permits any kind of syntactic rules that rewrites a non-terminal symbol, and so it permits grammars generating all ungrammatical sentences as well as those generating grammatical ones. Moreover, the languages generated by unrestricted rewriting systems are recursively enumerable sets and thus, given a sentence, an unrestricted rewriting system may take infinite time to determine its grammaticality.[3] It is also shown[4] that the regular grammar is so restrictive that some kinds of grammatical sentences cannot be generated by it.

Thus we are left with two choices: the context-free grammar (CFG) and the context-sensitive grammar (CSG). CFG is much more preferable because of its simplicity and familiarity. It is simpler than CSG since in each step of derivation it just rewrites a symbol, while CSG must have scanned the neighbouring symbols before

---

[1] * ' is called the Kleene Star operation. The Kleene Star, A*, of a set A is the set formed by concatenating the members of A any number of times in any order.

[2] C.f. Hopcroft & Ullman (1979) pp. 223-4.

[3] C.f. Partee (1990) chapter 19.

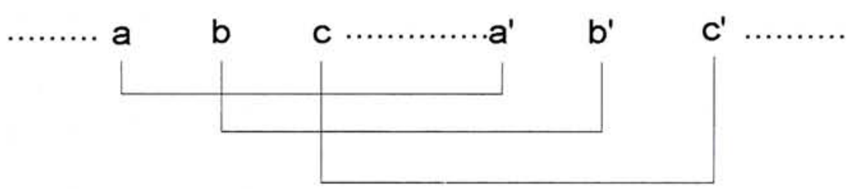
[4] C.f. Peters (1987) pp. 1 8.

rewriting a symbol. Accordingly the derivation of a context-free language (CFL)[5] sentence can be perspicuously represented by a phrase marker (tree diagram representation of the constituent structure of a sentence) but that of context-sensitive language cannot be so. CFG is familiar as its mathematical and computational properties have been thoroughly examined and there have been a great deal of feasible and efficient algorithms for parsing and generating CFL. However, CFG has been regarded by most generative linguists as an unsuitable model for natural language grammar. Let us go on to see if these arguments are persuasive.

2. Mathematical linguistics in a Nutshell

Before our examination of the arguments against the context-free character, or context-freeness (CFness), of natural language, some mathematical methods utilized by these arguments should first be outlined. The kernel of these arguments[6] is that some kinds of grammatical sentences in some language exhibit *cross-serial dependency.* Fig. 1 is a graphical representation of a symbol string[7] with cross-serial dependency pattern:

Fig.1



There are two portions in this schematic symbol string such that the first elements of them correspond to each other, and so are the second, third, ··· elements. It is shown that CFGs are unable to generate sentences with cross-serial dependency pattern. To understand this point, the notion of pushdown automaton (PDA) should be introduced, since the sentences generated by a CFG are accepted by a corresponding PDA.[8]

A PDA is an abstract machine with a reading head and a *stack*, and fed with an input tape moving in one direction. It is also endowed with a set of final states and a set of manipulation rules. At every instant a PDA is in a certain *state*. In every step

[5] The language generated by a type-n grammar is called a type-n language, and so a context-free language is the language generated by a CFG.

[6] The argument by Higginbotham (see section 5) is an exception.

[7] A terminological note: in this paper the terms *sentence* and *symbol string* are interchangeable, and so is the relationship between *word* and *symbol*. When standing alone, mathematical linguistics deals with symbols and symbol strings; when applied to natural language, its objects of study are words and sentences.

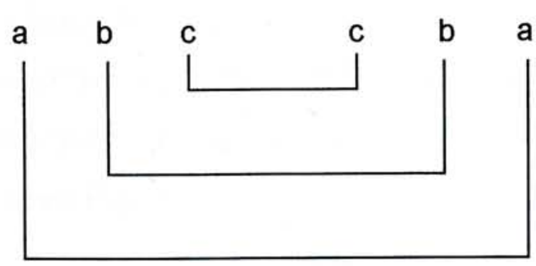
[8] The rigorous assertion is: all the sentences generated by a CFG are accepted by a corresponding *non-deterministic* PDA. As it is not our aim to investigate automata theory thoroughly, the issue of determinancy is not included in this paper.

a PDA reads a symbol on the input tape, check its current state and the topmost symbol in the stack. Then in accordance with some manipulation rule the PDA changes its state, and removes a symbol from or adds a symbol to the stack, or simply do nothing to the stack at all. In the next step it reads the next symbol in the input tape, check its new state and the new topmost symbol in the stack and operates in accordance with some manipulation rule. The operation of the stack is governed by the 'first-in, last-out' principle. That means the first symbol added in the stack is the last symbol removed from it, because whenever a new symbol is added, it pushes down the symbol(s) already in the stack and become the new topmost symbol, and a symbol cannot be removed from the stack unless there is no other symbols on its top. The PDA accepts a symbol string recorded on the input tape if and only if it is in a final state with an empty stack when it finishes reading this symbol string.

The purpose of the stack is to give a PDA some 'memory'. For example, suppose a PDA accepts only strings with even number occurrences of the symbol '*a*', then it can be so designed that whenever the first '*a*' is encountered, it puts a symbol 'A' in the stack, which is removed when the PDA comes across the second '*a*'. In this case, the stack records whether there is a single '*a*' that the PDA has read.

Consequently, a PDA accepts strings with *embedded dependency*, as the stack operates with the 'first-in, last-out' principle. For instance, the symbol strings '*abba*', '*baab*', '*abccba*', etc. are accepted by some PDA with similar design as described in the last paragraph.[9] The symbol string '*abccba*', with its structure shown in Fig.2, has the pattern that the inter-dependent '*c*'s are embedded in the inter-dependent '*b*'s, which are in turn embedded in the inter-dependent '*a*'s.

Fig. 2



[9] That is, the PDA is so designed that when it first encounters the symbols '*a*', '*b*', '*c*', it adds in the stack the symbol 'A', 'B', 'C' respectively. These symbols are removed when the PDA comes across the corresponding symbols the second times. Thus, when the symbols '*a*', '*b*' are read by the PDA, the stack have 'B' on the top of 'A'. According to the 'first-in, last-out' principle, B must be removed before A, and so the substring '*ab*' must have the substring '*ba*' follow it. That means, '*abba*', but not '*abab*', is an acceptable string.

However, the cross-serial dependency pattern cannot not be accepted by PDAs, as such strings violates the "first-in, last-out" principle of stack operation. Consider the exemplar string "*abcabc*". The first half "*abc*" adds to the stack the symbol "A", "B" and "C", with "A" at the bottom, "B" at the middle and "C" at the top. But the latter half "*abc*" requires the PDA remove the bottom "A" first, which is contradictory to the "first-in, last-out" principle. Thus, if we can show that the sentences of a language manifests cross-serial dependency, then this language cannot be accepted by any PDA, hence it cannot be generated by any CFG.

There is another mathematical tool for proving the non-CFness of a language. It is the *pumping lemma*.

The Pumping Lemma[10]

Let L be a CFL. Then there exists a constant k such that for any sentence "z" in L with $|z| \geq k$,[11] we can rewrite "z" as "*uvwxy*" such that

(i) $| vx | \neq 0$;

(ii) $k \geq | vwx |$;

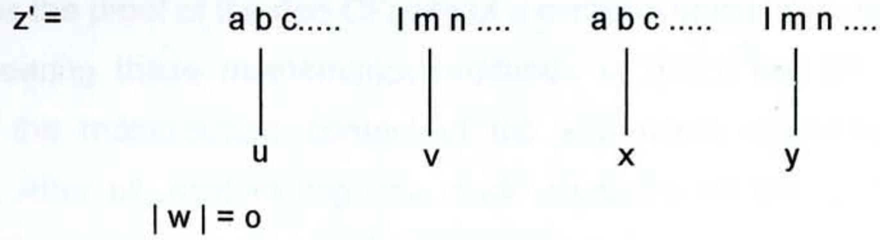
(iii) for all integers i, $uv^iwx^iy \varepsilon L$.

Let us refer a language with cross-serial dependency pattern as a *repeating language*, which is defined as the set $\{ \alpha x \beta x \gamma \mid \alpha, \beta, \gamma, x \varepsilon V^*$, V is the vocabulary$\}$. The pumping lemma in its contrapositive form is used to show that repeating languages are not CF. Suppose R is a repeating language with $|\alpha| = |\beta| = |\gamma| = 0$, i.e. $R = \{ xx \}$. For any string "z" in R with length not greater than k, it can be rewritten as "uv^iwx^iy" in the following way: take "u", "w", "y" as empty, and "v" equal to the first half of "z" while "x" the second half. But this method is not suitable for strings longer than k. Let "z'" be such a sentence, then it cannot not be rewritten as "uv^iwx^iy" in the way as described above, since it is prohibited by the condition (ii) of the pumping lemma. Nor can we rewrite "z'" such that $| vwxy | = 0$ or $| uvwx | = 0$, as it is not allowed by condition (i). The remaining possibility is that "v" being the latter portion of the first half of "z'", "x" being the first portion of the latter half, while the remaining words of "z'" reside in "u" and "y". (See Fig. 3)

[10] Adopted from Aho and Ullman (1972) p.195.

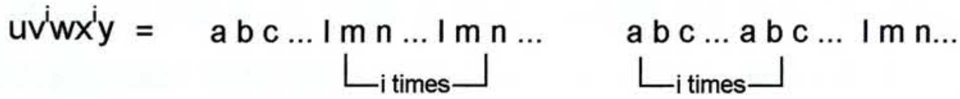
[11] |z| means the *length* of z. The length of a string (sentence) is the number of symbols (words) contained in it.

Fig. 3



But such construction would violate condition (iii), since uv^iwx^iy no longer maintain the balance of the first and latter halves, as shown by Fig. 4.

Fig. 4



Thus the sentences in R longer than k cannot be rewritten as required by the pumping lemma. Hence R is not a CFL. Repeating languages with non-empty α, β, γ are proved to be non-CF in a similar way.

Of course, not all natural language sentences exhibit cross-serial dependency. The English sentence *John loves Mary*, for instance, has no constituent which is cross-serially dependent on others. What we can prove is at most that some kinds of sentences of a certain language exhibit cross-serial dependency. Many linguists assume that

(2.1) a subset of a language being repeating implies that the language as a whole is not CF.

This assumption is wrong, as illustrated by the counter-example $\{ a^nba^nb \}$. Clearly $\{a^nba^nb\}$ is a repeating language, yet it is a subset of a CFL which is generated by a CFG with the following syntactic rules:

(i) $S \rightarrow a\,S$

(ii) $S \rightarrow b\,S\,b$

(iii) $S \rightarrow a$

Thus we need a rigorous mathematical tool to derive the non-CFness of a language from that of its subset. The required tool lies in the *closure properties* of CFL. It is shown[12] that

(2.2) the set formed by intersecting a CFL with a regular language is also a CFL.

If we find a certain set of sentences S of a natural language L not being a CFL, then we construct a corresponding regular language R such that S is the intersection of R

[12] C.f. Hopcroft and Ullman (1979) pp.130 - 136 and Partee *et.al.* (1990) pp. 497 499.

and L. The non-CFness of L is thus deduced by using (2.2) contrapositively. That completes the proof of the non-CFness of a certain natural language.

Bearing these mathematical methods in mind, we will skip as much as possible the mathematical content of the arguments examined in the following sections. After all, seldom linguists made mistakes on the formal methods; what arouses disagreement is the empirical premises.

3. Two Classical Arguments

The two arguments examined in this section were also discussed in Pullum and Gazdar (1982). In spite of that, I would still invite readers' attention to them because these arguments had been so widely believed as persuasive that any detailed research on the CFness of natural language should not neglect them.

3.1 The argument from the '*respectively*' construction

This argument was first proposed by Bar-Hillel and Shamir[13]. It captures the intuition that there is cross-serial dependency in English sentences involving the adverb '*respectively*'. For example, in the sentence '*The answer to question 1 and that to question 2 are 86 and 75 respectively*', the noun phrase (NP) '*the answer to question 1*' corresponds to '*86*' while the NP '*that to question 2*' corresponds to '*75*'.

The original argument as formulated by Bar-Hillel and Shamir focuses on this schematic English sentence:

(3.1.1) *John, Mary, David, ···, are a widower, a widow, a widower, ---, respectively*

where the dots are replaced by a string of personal names, and the dashes are replaced by a string of '*a widower*' and '*a widow*' such that the *n*th name has the same SEX feature as the *n*th element of the '*a widower/ a widow*' list. The argument commits the fallacy of (2.1) and makes too strong a claim about the grammaticality of certain English sentences, as criticized by Daly (1974). For this reason Langendoen gave another version of the argument[14] which focuses on the following sentence schema:

(3.1.2) *The woman and the men and the woman ··· smokes and drink and smokes --- respectively*

where the dots are replaced with a string of '*the woman*' and '*the men*', and the dashes are replaced with a string of '*smokes*' and '*drink*' such that the *n*th element of the former string has the same NUMBER feature as the *n*th element of the latter string. We need not go on complete his argument, for Pullum and Gazdar have

[13] Bar-Hillel, Y. and Shamir, E. 'Finite State Languages: Formal Representations and Adequacy Problems' in Bar-Hillel, *Language and Information.*

[14] Langendoen (1977) p.162.

already overthrown the grammaticality of the sentences outlined as (3.1.2). Pullum and Gazdar point out that whenever the subject involves two or more NPs so that its NUMBER is PLURAL, all the verbs in the predicate must be PLURAL too. That is, the grammatical counterpart of (3.1.2) should be:

(3.1.3) *The woman and the men and the woman ··· smoke and drink and smoke --- respectively*

Hence Langendoen s version could not be sound because of its false empirical premise.

Let us return to Bar-Hillel and Shamir s argument, as it can be remedied in the following way: Let L be the set of sentences depicted by the sentence schema (3.1.1), and R be the regular language

{ x are y respectively | x ε N*, N being the set of English personal names, and y ε { a widower, a widow}* }

Then L is the intersection of English with R. L is not CF since it is a repeating language. Hence English is not a CFL as CFL is closed under intersection with regular language.

Note that some of Daly s criticisms still apply to this modified argument. First, many English personal names are used by both male and female; second, (3.1.1) allows the sentence

(3.1.4) *John, John, John are a widower, a widower, a widower, respectively*

which is too weird to be considered as grammatical. But even if we assume the grammaticality of sentences like (3.1.4), and that English personal names can be completely divided into male and female, this argument is still unsound. Note that the following use of *respectively* is also possible:

(3.1.5) *The answers are 86 and 75 respectively*

The sentence (3.1.5) is true if the context implies that the subject *the answers* refers to answers for two questions, otherwise it is false. That means sentences like (3.1.5) are grammatical anyway, though there is no cross-serial dependency pattern; and their acceptability depends on non-syntactic information. (3.1.5) shows that *respectively* constructions are not bound to exhibit cross-serial dependency; the adverb *respectively* is not a syntactic element used to introduce cross-serial dependency pattern. This adverb is just a paraphrase of *in the specified order*. For example, (3.1.5) can be rewritten as *The answers, in the specified order, are 86 and 75*. Even in cases like (3.1.1) where there seems to be a cross-serial dependency pattern, the pattern is *semantic* rather than syntactic. Thus Bar-Hillel and Shamir s argument fails to prove that English is not a CFL.

3.2 The argument from Mohawk

This argument, proposed by Postal[15], makes use of a special construction in Mohawk, a Northern Iroquoian language of Quebec and upper New York state. Here is a simple Mohawk sentence:

(3.2.1) kaksa?a ka*nuhwe?*s ne- ka*nuhs*a?

the girl *likes* the *house*

'The girl likes the house'

For such SVO sentences in Mohawk, the object NP can be 'incorporated' into the verb stem to form a 'complex' verb. For example, (3.2.1) can be transformed to:

(3.2.2) kaksa?a ka*nuhs*nuhwe?s

the girl the-*house*-likes

'The girl likes the house'

What is special in Mohawk is the *double occurrences of object NP*: The object NP may appear both in the original position and inside the complex verb. (3.2.3) is a paraphrase of (3.2.2):

(3.2.3) kaksa?a ka*nuhs*nuhwe?s kik^ ka*nuhs*a?

the girl the-*house*-likes [modifier] the *house*

'The girl likes the house'

Postal remarks that in such constructions the NP inside the complex verb *must be identical to* that in ordinary object position.

Like other languages, in Mohawk the verbs, including complex verbs incorporating object NPs, can be nominalised. Thus a complex verb of the form 'the-house-finds' can be nominalised as a noun of the form 'the-house-finding' meaning 'the finding of the house'.[16] This 'complex' noun can in turn be incorporated into another verb; such as a Mohawk verb of the form 'the-house-finding-likes'. Again, this verb can be nominalised as a noun of the form 'the-house-finding-liking'.

Hence an argument against the CFness of Mohawk can be constructed. Let us adopt the following abbreviation[17]:

a = the translation into Mohawk of '*the man*'

b = the translation into Mohawk of '*admired*'

c = the translation into Mohawk of '*liking (of)*'

d = the translation into Mohawk of '*praising (of)*'

---

[15] Postal (1964).

[16] As the Mohawk verbs and NPs for this and some following examples are too long and too 'horrible' to be cited, sometimes I only write down the English equivalent (in a morpheme-by morpheme style) of Mohawk complex verbs and NPs.

[17] Adopted from Pullum and Gazdar (1982) p.163.

e = the translation into Mohawk of 'house'

Construct a regular language R: { axebye | x, y ε {c, d}* }. Then by intersecting Mohawk with R we get a set L: { axebxe | x ε {c, d}* }. L is a repeating language and is therefore not a CFL; and hence Mohawk is not CFL because of (2.2).

In response to Postal's argument, Pullum and Gazdar pointed out that in Mohawk there is another type of sentence --- the *possessed incorporation* construction. In the following example,

(3.2.4) i?i k-*nuhwe?s* ne *sawatis* hrao-*nuhs*-a?

I *like* *John('s)* *house*

'I like John's house'

the object NP involves a possessive noun *sawatis* (*John*)'. When incorporation transformation takes place in (3.2.4), the whole object NP need not be incorporated; the possessive noun may remain in the original position:

(3.2.5) i?i hrai-*nuhs*-nuhwe?s ne sawatis

I *house*-like John

'I like John's house'[18]

Sentences like (3.2.5) are called possessive incorporation construction. The existence of this construction implies that a sentence is still grammatical even if the incorporated NP inside the complex verb is not identical to the NP in the ordinary object position. Citing Pullum and Gazdar's own example, the Mohawk sentence of the form 'the man praising-of-liking-of-house admired that liking-of praising-of-house' is grammatical, despite its absurd meaning 'the man admired that liking-of praising-of-house's praising-of-liking-of-house'.

Thus L would not be the intersection of Mohawk with R, as Mohawk does not require x being identical to y in those sentences depicted by R. The non-CFness of L cannot derive the non-CFness of Mohawk.

4. The Arguments from Sluicing and Doubling Relative Constructions

Two years after Pullum and Gazdar's successful repudiations, there appeared two new arguments against the CFness of the English language. One is proposed by Postal and Langendoen.[19] This argument makes use of the *sluicing construction* in English, which is exemplified by these examples:

(4.1a) *Sarah considered some proposals but it's unknown how many.*

(4.1b) *If any books are still left on the sale table, find out which ones.*

---

[18] Readers may notice that the prefix of '*nuhs*' in (3.2.4) is '*hrao*' whereas that in (3.2.5) is '*hrai*'. But this difference would not upset the refutation. C.f. Pullum and Gazdar (1982) pp.168 - 169.

[19] Postal and Langendoen (1984).

In these constructions, the second constituent sentences seem to have some missing parts which "sluice" away. The "complete" form of (4.1a) should be "*Sarah considered some proposals but it's unknown how many proposals are considered*"; likewise, (4.1b) is a simplified form of "*If any books are still left on the sale table, find out which ones are still left*". Confined to constructions with only two constituent sentences, sluicing sentences are of the form

V Q1 X1 W Y Q2 X2 Z

where V, W, Y and Z are word strings; Q1 is an indefinite quantifier/pronoun, and X1 is the rest of the nominal quantified by Q1; Q2 is a *wh*-quantifier or pronoun anaphorically related to Q1, and X2 is the rest of the nominal quantified by Q2.[20] (4.1b) can be thus analyzed as

If *any books* are still left on the sale table, find out *which ones*
V *Q1* *X1* W Y *Q2* *X2* Z=∅

In general, the reference of X2 must be identical to or subsume that of X1, otherwise the sentences thus formed are ungrammatical, like :

(4.2a) **The warehouse will ship several machines to our office but we have no idea how many typewriters*

(4.2b) **A few physicians still use this drug and Sam can tell you how many nurses*

However, in some cases X2 must be identical to X1, such as

(4.3a) *Joe discussed some bourbon lover but it's not known which bourbon lover*

(4.3b) **Joe discussed some bourbon lover but it's not known which lover*

(4.4a) *Joe discussed some bourbon lover hater but it's not known which bourbon lover hater*

(4.4b) **Joe discussed some bourbon lover hater but it's known which hater*

(4.4c) **Joe discussed some bourbon lover hater but it's known which lover hater*

The reason for the ungrammaticality of (4.3b) is that "*lover*" has quite different a meaning with "*bourbon lover*": the former concerns love, while the latter, drinks. Nor is there any other possible pronoun for "*bourbon lover*". Hence X2 must be identical to X1 in the cases of (4.3 a-b). The cases for (4.4 a-c) are similar.

In the light of sentences like (4.3a) and (4.4a), Postal and Langendoen construct a regular language

R: { Joe discussed some bourbon x but which bourbon y is unknown | x, y ε {hater, lover}* }

The intersection of R with English is

L: { Joe discussed some bourbon x but which bourbon x is unknown | x ε

[20] This characterization of sluicing constructions is adopted from Postal and Langendoen (1984) p.178.

{hater, lover}* }

L is a repeating language and therefore not a CFL. Hence the non-CFness of English.

In response to this argument, Pullum proved that L is indeed not the intersection of English with R. He drew readers 'attention to the following discourse[21]:

(4.5)

Speaker 1: *It looks like they're going to appoint another bourbon-hater as Chair of the Liquor Purchasing Committee.*

Speaker 2: *Yes --- even though Joe discussed some bourbon-lover, but which bourbon-hater is still unknown.*

Obviously the sentence articulated by speaker 2 is a member of the regular language R with $x \neq y$, yet it is also a grammatical sentence in English. The reason is that an anaphoric element can be bound in an *inter-sentential* way. "It is permissible for the antecedent of the [anaphoric] constituent in a construction of this type to be in a previous sentence in the discourse, and for the anaphor relation to hold across an intervening conjunct with arbitrary content."[22] In (4.5), the *wh*-phrase "*which bourbon-hater*" by speaker 2 is bound by the antecedent "*another bourbon-hater*" uttered by speaker 1.

Therefore L is not the intersection of English with R; the argument from the English sluicing construction fails.

Langendoen and Postal are fully convinced by Pullum's criticism, which is based on the inter-sentential character of the binding of certain anaphoric *wh*-phrases. Yet they soon found[23] another construction, the *doubling relative construction*, which may be taken to establish another argument against English CFness while avoiding Pullum's attack. The doubling relative construction is exemplified as follows:

(4.6a) *The FBI arrested some senator, which senator committed suicide.*

(4.6b) *Some mammals, which mammals are now able to vote, are hostile to reptile rights.*

It is still found in such sentences the constituents Q1, X1, Q2 and X2 as defined in page 11, but this time the phrase [Q2 X2] immediately follows [Q1 X1]. Clearly [Q2 X2] initiates a relative clause, and it is called doubling relative construction because

21 Adopted from Pullum (1984) p.183, with little modification.

22 Pullum (1984) p.183.

23 Langendoen and Postal (1984).

of the *co*-occurrence of X1 and X2, where X1 and X2 are usually identical. Langendoen and Postal claim that

(4.7) When a doubling relative is adjoined to a nominal not in sentence-final position, that nominal must be the antecedent of the *wh*-phrase.

Thus we can construct a regular language

R: {Some bourbon x, which bourbon y has been nominated, merit consideration | x, y ε {hater, lover}* }

and we would get a set

L: { Some bourbon x, which bourbon x has been nominated, merit consideration | x ε {hater, lover}* }

by intersecting R with English. Thus the non-CFness of English is derived from that of L. Unlike that by sluicing construction, this new argument eludes Pullum s attack because we cannot assign an antecedent outside the sentence to the anaphoric [Q2 X2], as shown in:

(4.8)

Speaker 1: *Some <u>bourbon lover hater</u> will be nominated.*

Speaker 2: **Some bourbon hater lover, which <u>bourbon lover hater</u> merit consideration, have proposed to raise the drinking age to 85.*

No linguist gave any comment on this new argument; it is not too difficult, however, to give a repudiation. We should first note that the repudiation must not be a modification of Pullum s original point because (4.7) denies any inter-sentential anaphoric binding. In fact, if X2 is omitted from a doubling relative construction, e.g. rewrite (4.6b) as (4.6c),

(4.6c) *Some mammals, which are now able to vote, are hostile to reptile rights*

the resulting sentence involves an ordinary relative clause in which the *wh*-element must refer to an intra-sentential antecedent. The presence of X2 in doubling relative constructions is just for providing some more information about [Q1 X1]; so [Q2 X2] must refer to an intra-sentential antecedent.

My criticism begins with an example used by Langendoen and Postal themselves[24] to illustrate (4.7):

(4.9)

Speaker 1: *Some mammals are hostile to reptile rights.*

Speaker 2: **Some Boeing 747s, which mammals are now able to vote, are being sold to the Saudis.*

---

[24] Langendoen and Postal (1984) p.187.

Langendoen and Postal regard speaker 2's utterance as ungrammatical because here neither X2 is identical to X1 nor does the reference of X2 subsume that of X1. But this judgment of grammaticality is simply wrong. Suppose a child walks by the airport and is told that the "big birds" over there are Boeing 747s. So the child thinks, "The big fish whale is a mammal, so the big bird Boeing 747 must be a mammal too!" Under this context speaker 2's utterance is acceptable. That means this sentence is grammatical anyway; its unacceptability in "normal" contexts is merely due to pragmatic reasons.

Recall that the issue of weak generative capacity is a matter of syntax rather than semantics or pragmatics, so even we admit the truth of (4.7), the sentences in R with $x \neq y$ are still grammatical. For example, the sentence

(4.10) *Some bourbon lover hater, which bourbon hater lover has been nominated, merit consideration*

is grammatical, despite that it is usually wrong because bourbon lover haters are generally not bourbon hater lovers. Thus the intersection of R with English is not L; the argument from the doubling relative construction fails.[25]

5. The Argument from the English *such that* constructions

Just a few months before Langendoen and Postal proposed the argument from the sluicing construction, Higginbotham suggested another way to prove the non-CF character of English. This argument deserves more elaborate explication because it is quite different from other arguments against CFness: it does not make use of the concept of cross-serial dependency and does not utilize the pumping lemma.

He began his argument with a regular language

R: { the woman such that X* she Y* left is here}, where

X = { the man such that }; and

Y = [gave him to him] $\cup$ {gave him to this} $\cup$ {gave this to him} $\cup$ {gave this to this}

And he claimed that the intersection of R with English is:

L : { the woman such that X^n she Y^n left is here | no initial segment of Y^n has more occurrences of *this* than those of *him* }[26]

The condition specifying L may be difficult to understand at a glance. Let us first look at an exemplar sentence in L:

[25] Note that this criticism of the argument from the doubling relative construction also applies to the argument from the sluicing construction.

[26] "x^n" means *n* occurrences of x.

(5.1) *The woman such that the man such that the man such that she gave this to him gave him to him*

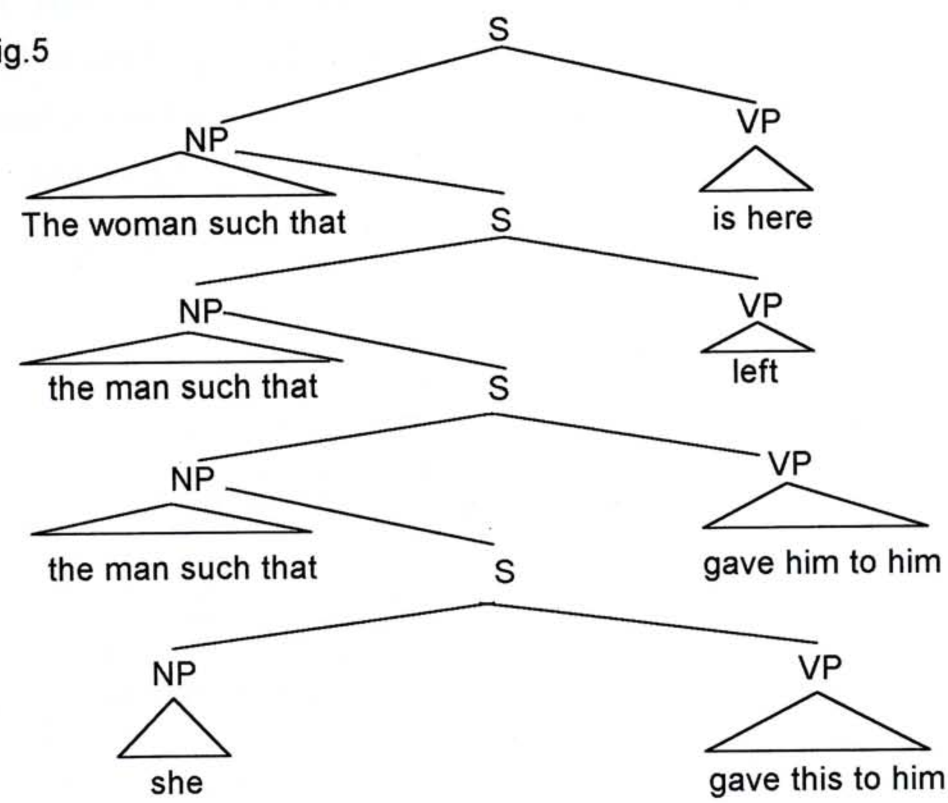
which structure is shown in Fig.5.

In sentences in L the reference of every "*the man*" is specified by a "*him*"; that is, every "*the man*" in an X must be assigned a "*him*" in an Y. Moreover, a "*him*" can be assigned to a "*the man*" only once. In example (5.1) the first "*the man*" corresponds to either a "*him*" in "*gave him to him*", and the second "*the man*" corresponds to the "*him*" in "*gave this to him*". Therefore one requirement of sentences in L is:

(5.2) the number of "*him*" is not less than the number of "*the man*"[27].

This condition is satisfied when the sentence contains no "*gave this to this*". When some "*gave this to this*" appear, there must be at least an equal number of "*gave him to him*" so as to provide a sufficient amount of "*him*". However, just the balance

Fig.5

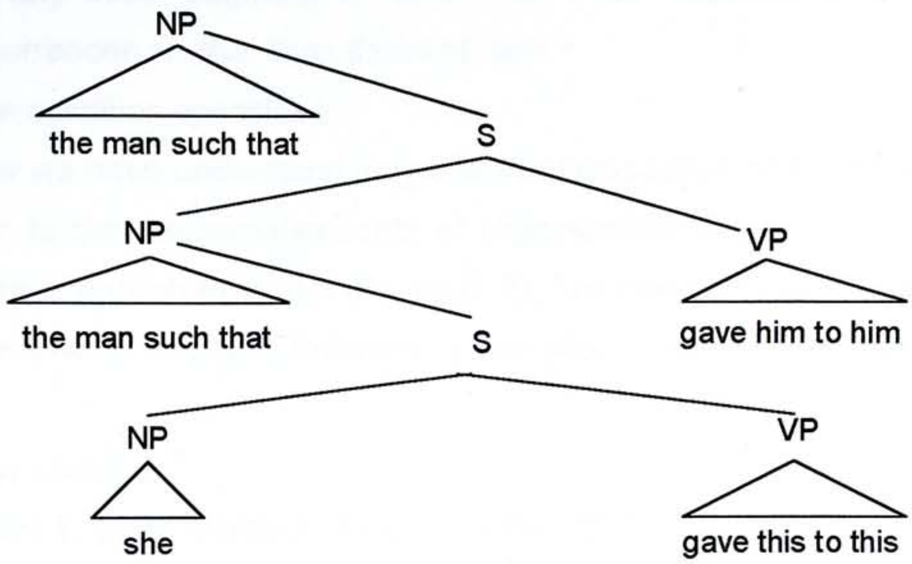


of these two VPs is not enough; the position of "*gave him to him*" also matters to grammaticality; as shown by the ungrammatical (5.3) (its structure is shown in Fig.6):

(5.3) **The woman such that the man such that the man such that she gave this to this gave him to him left is here*

[27] It is not a problem if the number of "*him*" exceeds the number of "*the man*", because the extra "*him*" may refer to NPs outside the sentence.

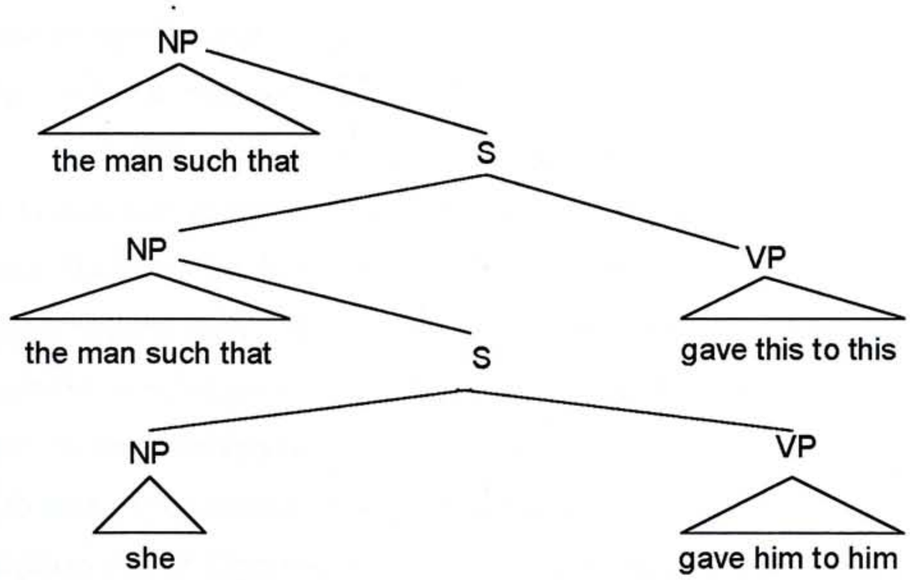
Fig.6



It is ungrammatical because the reference of the second *the man* is not restricted by the *such that* clause attached to it, as there is no *him* available. When *gave him to him* and *gave this to this* inter-change their positions, as illustrated in (5.4), the result is grammatical (c.f. Fig.7):

(5.4) *The woman such that the man such that the man such that she gave him to him gave this to this left is here*

Fig. 7



(5.4) is grammatical because both *the man* are assigned some *him* in the *such that* clauses attached. From these two examples we see that

(5.5) for any initial segment X^k of X^n, the corresponding Y^k of Y^n must have a sufficient number of *him* to be assigned to *the man* in X^k.

which can be read as

(5.6) for any initial segment X^k of X^n, the number of *gave this to this* in the corresponding Y^k of Y^n must not exceed that of *gave him to him* .

In turn, (5.6) can be paraphrased as

(5.7) for any initial segment X^k of X^n, no initial segment Y^k of Y^n has more occurrences of *this* than those of *him*.

which is the condition specifies L.

Now we have understood why L is the intersection of R with English and we should turn to the mathematical side of Higginbotham's argument. He derives the non-CFness of English from that of L by (2.2). The non-CFness of L is proven by the *Ogden's lemma*, a stronger theorem in the sense that the pumping lemma is its corollary.

The Ogden's lemma[28]

For each CFL L, there exists an integer k such that if

(a) $z \varepsilon L$;

(b) $|z| \geq k$; and

(c) there are k or more symbols in z being called as *in distinguished positions*;

then z can be written as uvwxy such that

(i) w contains at least one of the distinguished positions;

(ii) either u and v both contain distinguished positions, or x and y both contain distinguished positions;

(iii) vwx has at most k distinguished positions;

(iv) for all integer i, $uv^iwx^iy \varepsilon L$.

Higginbotham's proof chooses z as

The woman such that X^{2n} she A^n B^n left is here

where A is { gave him to him }, B is { gave this to this } and n > k. The distinguished positions are taken to be the positions for the words of B^n. By condition (i), 'w' contains some words of B^n. By condition (ii), there are two alternatives. The first is that both 'u' and 'v' contain words of B^n. In this case X^{2n} and A^n are also contained in 'u', but then 'uv^iwx^iy' would have more Bs than As, i.e. more '*gave this to this*' than '*gave him to him*'. That would violate the specifying condition (5.7) of L and therefore violate condition (iv) of Ogden's lemma. Thus it is not 'u' and 'v' which contain the distinguished positions.

The remaining alternative is that both 'x' and 'y' contain some words of B^n. In this case there are three options regarding the position of '*she*'.

1) '*she*' is inside 'v'. Then 'uv^iwx^iy' would have more than one '*she*' and thus it is not a sentence in L.

[28] Adopted from Aho and Ullman (1972) p.193, with some modification.

2) *she* is inside u. In this case *the man such that* is also inside u. Then uv^iwx^iy would have more *gave this to this/ gave him to him* than *the man such that* and therefore it is not a sentence in L.
3) *she* is inside w. In this case A^n is also contained in w. Then uv^iwx^iy would have more B than A and therefore it is not a sentence in L.

Hence x and y cannot contain the distinguished positions either. As condition (ii) is not satisfied, L is not a CFL, and neither is English.

Similar to those arguments that we have evaluated, the problem of Higginbotham's one lies in its empirical assumption rather than its mathematical proof. The non-CF character of L, as Higginbotham conceived, lies in that sentences in L need to provide sufficient *gave him to him* preceding *gave this to this* as well as to provide sufficient *him*; this is due to, as Pullum points out, Higginbotham's basic assumption that the English relative clause construction with *such that* is constrained to contain a pronoun anaphorically bound to the head.[29] This claim, however, is too problematic to be considered as likely to be true, for there exists many counter examples:

(5.8) *every triangle such that two sides are equal is called an isosceles triangle*

(5.9) *the number system such that 2 and 3 make 5 is taught in kindergarten*

In fact, Higginbotham also noted such examples, but he insisted that any sentences like (5.8) and (5.9) are ungrammatical, a judgment which I think simply false. Apart from this disagreement in grammaticality judgment, the problem of Higginbotham's argument includes his two assertions about these sentences:

(5.10) they are acceptable and able to be assigned an interpretation, in spite of their ungrammaticality;

(5.11) they are interpreted, where possible, as elliptical for sentences that are not merely relevant to the content of the head nouns, but further supply a place into which binding is possible.[30]

(5.10) is weird enough; how can an ungrammatical sentence be acceptable? How can an interpretation be assigned to it? We have seen in section 1 that the point of grammar is to compose the meaning of a whole sentence from the meanings of the lexical items. The necessary condition of assigning an interpretation to a sentence is that the sentence is itself well-formed, i.e. grammatical. Thus (5.10) is highly implausible.

Nor would (5.11) make (5.10) any more reasonable. What (5.11) claims is, say, that (5.8), though ungrammatical, can be interpreted as an ellipsis of:

---

[29] Pullum (1985) p.291.

(5.12) *every triangle such that two sides of it are equal*

Were (5.11) true, it is still unclear why (5.10) may be true. An elliptical sentence is grammatical as well as its "complete" counterpart. The sentence "*Mary can't speak English, but John can*" is elliptical since the second clause lacks a VP, yet it is grammatical for itself, even if its interpretation depends on a "complete" sentence.

Moreover, (5.11) itself is not correct. Pullum gave many counter examples in which there are no place for a pronoun to be added in, such as:

(5.13) *Over many years, it had become clear that Lee and Sandy were just one of those couples such that people always reported loving her but hating him*[31]

To these examples Higginbotham simply replied that there is still pronominal binding, the *split* pronominal binding. In (5.13) it is both "*her*" and "*him*" which are bound to the head *couples*. It seems that Higginbotham did not examine Pullum's examples carefully enough, since there are some of them which cannot be explained by split pronominal binding, such as:

(5.14) *The old crone had a manner such that even the children who saw her pass in the street would shudder and turn away*[32]

Higginbotham might rebut the grammaticality of sentences like (5.14); again, this would be an incorrect judgment. Thus (5.11) is simply false and even it were true it could not prove (5.10).

Owing to Higginbotham's false judgment of the grammaticality of sentences like (5.8) and (5.9), and his false assertions (5.10) and (5.11), the argument from the English *such that* construction is unsound.

6. The Argument from German constructions

The first argument showing that Dutch is not a CFL is proposed by Huybregts[33]. He discovered that there are infinitely many Dutch subordinate clauses exhibiting such cross-serial dependency pattern like

(6.1) dat N_1 N_2 ⋯ N_n V_1 V_2 ⋯ V_n

where N_i corresponds to V_i . Examples are[34]:

(6.2) ⋯ dat Jan *Piet* Marie *de kinderen* zag *helpen* laten *zwemmen*

⋯ that Jan *Piet* Marie *the children* saw *to help* to make *to swim*

'… that Jan saw Piet help Mary make the children swim'

(6.3)⋯ dat de leraar *Jan* Marie *de kinderen* leerde *laten* leren *zwemmen*

---

[30] Higginbotham (1984) p. 347.

[31] Pullum (1985) p.292.

[32] Pullum (1985) p.292.

[33] Huybregts, M. "Overlapping Dependencies in Dutch" in *Utrecht Working Papers in Linguistics.*

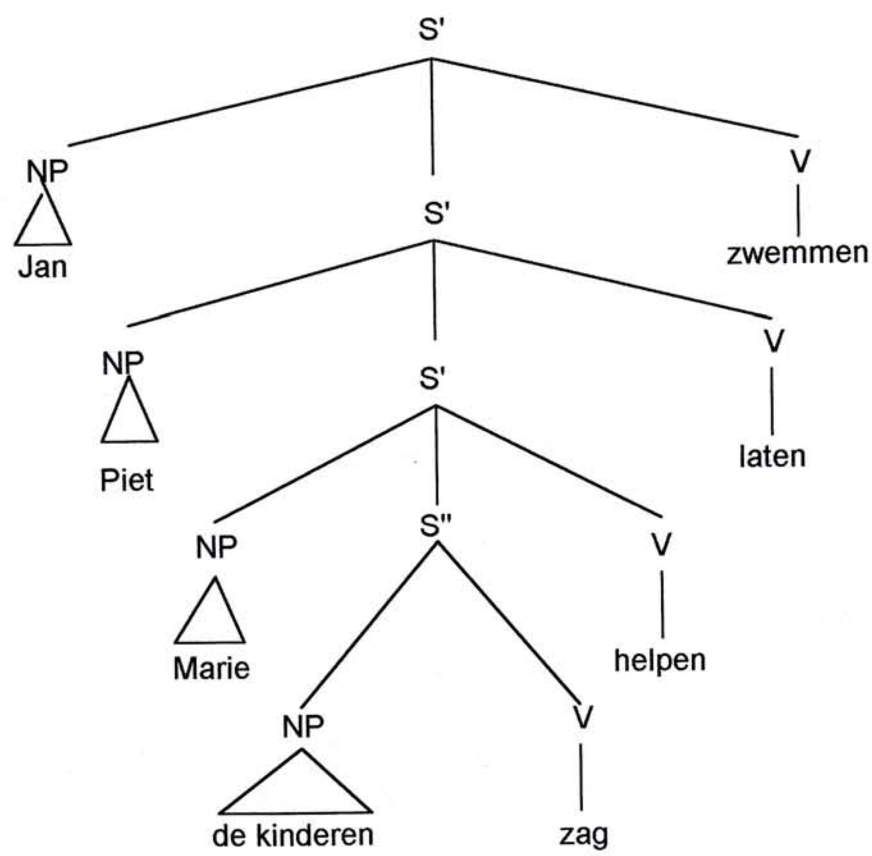
[34] Cited from Bresnan *et.al.* (1982) p.288.

··· that the teacher *Jan* Marie *the children* taught *to make* to teach *to swim*

···that the teacher taught Jan to make Marie teach the children to swim ·

In general, an infinite number of such subordinate clauses can be generated by choosing a finite verb as the first verb, an infinitival verb not taking VP complement as the last verb, infinitival verbs taking VP complement as the remaining verbs, and a number of NPs equal to that of the verbs. It seems then the Dutch sentences containing these clauses can be extracted to form a non-CFL, which is in turn used to prove the non-CFness of Dutch by (2.2).

This kind of proof would not be successful, since there exist some CFGs which can generate such Dutch clauses. Bresnan *et.al.* (1982) gave one CFG which produces the clauses like (6.2) and (6.3), and Pullum and Gazdar (1982) gave many more CFGs which generate wider classes of Dutch subordinate clauses exhibiting the pattern (6.1). These CFGs generate a phrase structure tree for, say, (6.2), as follows:



From this graph, it is easily seen that the phrase structures assigned by those CFGs are by no means adequate, nor are they suitable for semantic interpretation. But our problem in hand is about the *weak* generative capacity of CFG; the adequacy of phrase structure representation is not our main concern. Thus the CFGs proposed by Pullum and Gazdar and Bresnan *et.al.* are valid disproof of Huybregts 'argument.

The effort to prove that Germanic languages are weakly non-CF continued, however; and eventually Shieber succeeded in showing the non-CFness of Swiss German[35]. In Swiss German there are also subordinate clauses exhibiting cross-serial dependency. What is important is that in such clauses different verbs subcategorize for object nouns of different cases (dative and accusative). Examples are[36]:

(6.4) das mer em Hans es huus h÷lfed aastriiche

that we Hans-DAT the house-ACC helped-DAT paint-ACC

that we helped Hans paint the house ,

(6.5) das mer d chind em Hans es huus haend

that we the children-ACC Hans-DAT the house-ACC have

wele laa hølfe aastriiche

wanted let-ACC help-DAT paint-ACC

that we have wanted to let the children help Hans paint the house ,

Accordingly we can construct a set of sentences containing subordinate clauses with n dative NPs preceding m accusative NPs, followed by n verbs requiring dative objects and m verbs requiring accusative objects. Let us call this set L, which is defined as:

L: $\{ \alpha (N_{DAT})^n (N_{ACC})^m \beta (V_{DAT})^n (V_{ACC})^m \gamma \mid \alpha, \beta, \gamma$ are Swiss German word strings, $N_{DAT}, N_{ACC}, V_{DAT}, V_{ACC}$ are the sets of dative NPs, accusative NPs, verbs requiring dative objects, and verbs requiring accusative objects respectively}

L is the intersection of Swiss German with a regular set

R: $\{ \alpha (N_{DAT})^* (N_{ACC})^* \beta (V_{DAT})^* (V_{ACC})^* \gamma \mid$ all the symbols are defined as in L}

L is a repeating language and therefore not a CFL. By (2.2) the non-CFness of L derives the non-CFness of Swiss German.

As seen in previous sections, the most probable way to disprove arguments against CFness is to show that the cross-serial dependency pattern in question is not purely syntactic but semantic or even pragmatic. This strategy, however, does not work here, as most grammarians would agree that case-marking is a syntactic phenomenon.

So we finally find out a natural language which is not a CFL, and that is sufficient to show that CFG is not appropriate enough to generate all well-formed sentence structures of all natural languages. Yet we must beware that this result still does not cater for the curiosity of philosophers of language. After all, the non-CF

[35] Shieber (1985).

[36] Cited from Shieber (1985) pp.321 323.

character of natural language is shown in only some constructions of just one language; it seems that the cross-serial dependency pattern of some subordinate clauses in Swiss German is just an accident. Philosophically, Shieber's discovery does not give any deep insight in the structure of natural language sentences; practically, any AI researchers could ignore this finding and carry on natural language processing with CFG in most cases.

Bearing this remark in mind, let us leave the arguments from cross-serial dependency as examined above and turn to two arguments against CFG from some syntactic phenomena such as subject-auxiliary agreement, anaphoric binding, question formation. Readers should notice that these grammatical phenomena appear in (almost) all natural languages, otherwise they would give rise to arguments as unsatisfactory as the Swiss German example even if the arguments were successful.

7. The Argument from Feature Agreement

During the heyday of transformational grammar, many theoretical linguists believed that the *subject verb agreement* pattern exhibiting in most human languages is itself a direct proof of the inadequacy of CFGs.[37] In many languages nouns and verbs bear certain syntactic features such as PERSON (1^{st}, 2^{nd}, 3^{rd}), GENDER (male, female, etc) and NUMBER (singular, plural). The features of the subject must match with those of the verb in any finite clause. English examples:

(7.1a) *The boy comes* (singular subject with singular verb)

(7.1b) *The boys come* (plural subject with plural verb)

Similar examples can be found in Latin (7.2) and in French (7.3).

(7.2a) *puer* *venit*
the boy comes

(7.2b) *pueri* *veniunt*
the boys come

(7.3a) *le garçon vient*
the boy comes

(7.3b) *les garçons viennent*
the boys come

In fact, nearly all the verbs and nouns of most inflectional languages illustrate a variety of morphological forms with respect to different possible syntactic feature combinations. For languages with seldom or even no morphological inflection,

[37] See, for example, Grinder & Elgin (1975) pp. 57-59.

linguists hypothesize that there are still subject verb agreements of syntactic features, though they are not "realized" in most cases. This is a very tricky point, yet a point with some empirical evidence. For example, in Chinese the adverb "*dou*^{1}" is used when the subject is plural, as shown by the grammatical sentence "*Tamen doushi hanren*" and the ungrammatical sentence "**Ta doushi hanren*".

Advertisements of transformational grammar, which overwhelm linguistic literature in 1960s and early 1970s, claimed that CFGs are incapable of generating sentences in which subjects and verbs agree. Take the generation of (7.1) as an illustration. By the use of some familiar grammar rules we first get the structure $[_S\ [_{NP}$ Art N] $[_{VP}$ V]]. Then there are two alternatives regarding the lexical insertion (i.e. the rewriting of non-terminal symbols into terminal symbols): either insert the verb first or insert the words in the noun phrase first. Suppose the latter is the case. We can then arbitrarily choose between "*boy*" and "*boys*" for rewriting N; however, the insertion of verb is by no means so liberal. V can be rewritten as "*comes*" only if the N is rewritten as "*boy*", as "*come*" only if the N is "*boys*". The "only if" here implies that the applications of rules "$V \rightarrow$ *comes*" and "$V \rightarrow$ *come*" are not unconditional, i.e. they are employed under certain contexts. As its name and definition suggest, rules of CFGs rewrite a syntactic category regardless of its "context" i.e. other constituents in its vicinity. That means the lexical insertion rules for V are not CF. Similarly, lexical insertion rules for N are not CF if V is rewritten before N.

With such examples in hand, those transformational linguists announced that in order to generate a sentence containing two or more interdependent words/ constituents, the grammar must contain some context-sensitive rules. Thus agreement is beyond the scope of CFG. This completes their proof of CFG inadequacy and that of the non-CFness of natural languages.

It is not too difficult to repudiate this argument from subject verb agreement, as it is merely the consequence of ignorance on the generative power of CFGs. According to the argument, CFGs are incapable of all kinds of dependency between any two or more words. This is definitely an absurd proposition, as we have seen in section 2 that CFGs are able to generate strings exhibiting embedded dependency, let alone the simple single dependency between a subject and a verb. In fact, the English subject verb agreement pattern can be generated by the following CF grammar rules[38]:

(7.4a) $S \rightarrow NP_{[SINGULAR,\ 1ST\ PERSON]}\ VP_{[SINGULAR,\ 1ST\ PERSON]}$

[38] Notational note: the square brackets specify the syntactic feature combinations. Note also that (7.5) and (7.6) are schemata for grammar rules. The dots mean possible constituents accompanying the N or V.

(7.4b) $S \rightarrow NP_{[SINGULAR,\ 2ND\ PERSON]}\ VP_{[SINGULAR,\ 2ND\ PERSON]}$

(7.4c) $S \rightarrow NP_{[SINGULAR,\ 3RD\ PERSON]}\ VP_{[SINGULAR,\ 3RD\ PERSON]}$

(7.4d) $S \rightarrow NP_{[PLURAL]}\ VP_{[PLURAL]}$

(7.5a) $NP_{[SINGULAR,\ 1ST\ PERSON]} \rightarrow \cdots\ldots N_{[SINGULAR,\ 1ST\ PERSON]} \cdots\cdots$

(7.5b) $NP_{[SINGULAR,\ 2ND\ PERSON]} \rightarrow \cdots\cdots N_{[SINGULAR,\ 2ND\ PERSON]} \cdots\cdots$

(7.5c) $NP_{[SINGULAR,\ 3RD\ PERSON]} \rightarrow \cdots\cdots N_{[SINGULAR,\ 3RD\ PERSON]} \cdots\cdots$

(7.5d) $NP_{[PLURAL]} \rightarrow \cdots\ldots N_{[PLURAL]} \cdots\cdots$

(7.6a) $VP_{[SINGULAR,\ 1ST\ PERSON]} \rightarrow \cdots\cdots V_{[SINGULAR,\ 1ST\ PERSON]} \cdots\cdots$

(7.6b) $VP_{[SINGULAR,\ 2ND\ PERSON]} \rightarrow \cdots\cdots V_{[SINGULAR,\ 2ND\ PERSON]} \cdots\cdots$

(7.6c) $VP_{[SINGULAR,\ 3RD\ PERSON]} \rightarrow \cdots\cdots V_{[SINGULAR,\ 3RD\ PERSON]} \cdots\cdots$

(7.6d) $VP_{[PLURAL]} \rightarrow \cdots\cdots V_{[PLURAL]} \cdots\cdots$

(7.7a) $N_{[SINGULAR,\ 1ST\ PERSON]} \rightarrow$ *I*

(7.7b) $N_{[SINGULAR,\ 2ND\ PERSON]} \rightarrow$ *you*

(7.7c) $N_{[SINGULAR,\ 3RD\ PERSON]} \rightarrow$ *he/she/boy/*……

(7.7d) $N_{[PLURAL]} \rightarrow$ *they/boys/*……

(7.8a) $V_{[SINGULAR,\ 1ST\ PERSON]} \rightarrow$ *am/come/*……

(7.8b) $V_{[SINGULAR,\ 2ND\ PERSON]} \rightarrow$ *are/come/*……

(7.8c) $V_{[SINGULAR,\ 3RD\ PERSON]} \rightarrow$ *is/comes/*……

(7.8d) $V_{[PLURAL]} \rightarrow$ *are/come/*……

These CF grammar rules generate sentences perfectly obeying the subject verb agreement requirement, and therefore overthrow the argument from this syntactic phenomenon. The basic mistake of this argument is that it takes, without any ground, the syntactic categories N, V, etc as *atomic* constituents which cannot be further analyzed into other non-terminal symbols. If we expand N and V into categories each of which corresponds to a N or V with certain feature combination, then the grammar rules can be designed as (7.4) through (7.8) such that the rules themselves automatically guarantee subject verb agreement.

Transformational linguists may reply that these rules are not brief and elegant enough to be the correct description of human linguistic competence. Yet we should note that elegance is just an aesthetic issue, which has nothing to do with the problem of generative capacity. A complicated set of grammar rules would not generate more, or less, types of sentences just by being reformulated in a more elegant way. Moreover, the rules (7.4) through (7.6) can indeed be formulated in a simple and thus graceful way:

(7.9) $S \rightarrow NP{:}[\alpha]\ \ VP{:}[\alpha]$

(7.10) NP:[α] $\rightarrow$ ······ N:[α] ······

(7.11) VP:[α] $\rightarrow$ ······ V:[α] ······

where α ranges over the set of possible syntactic feature combinations. The set of rules (7.9) through (7.11) is just an abbreviation for rules (7.4) through (7.6). That is, (7.9) is a rule schema standing for rules (7.4a d), (7.10) a schema for rule schemata (7.5a-d), etc. It provides a simpler format to describe our syntactic knowledge yet without any influence on the CF character of the grammar. This reaffirms the point that the elegance of grammatical description is, after all, merely a notational issue.

So far we have arrived at such a proposal:

(7.12) whenever there is a CSG for a language in which sentences a word agrees with another word with respect to certain syntactic features, we can reformulate the grammar as another set of CF grammar rules consisting a larger set of non-terminal vocabulary.[39]

Not only suitable for explaining subject verb agreement, this method refutes all sorts of arguments against CFness from feature agreement between a pair of words/ constituents, such as feature agreement between *anaphors* and their *antecedents*. Yet we should note that the application of (7.12) is not always as straightforward as in the case of subject verb agreement; let us stop for a while to consider this point.

Remark on Anaphoric Binding

In the tradition of generative grammar, especially the GB (Government and Binding) approach, the phenomenon of anaphoric binding receives much attention. *Anaphor* consists of two kinds of linguistic elements: *reflexives* (e.g. *himself*, *herself*, *themselves*) and *pronouns* (e.g. *him*, *her*, *them*). Such elements must be bounded by their *antecedents*; that is, an anaphor refers to some constituent existing in the preceding portion of the sentence. Moreover, there are some rules governing the distance between anaphors and their antecedents.

(7.13a) John_i hurts himself_i.[40]

(7.13b) *Mary_i hurts himself_i.

(7.13c) Mary said [$_S$ John_i hurt himself_i].

(7.13d) *Mary_i said [$_S$ John hurt herself_i].

(7.14a) John_i hurts him_j/her_j.

(7.14b) *John_i hurts him_i.

[39] To avoid forming a grammar with an infinite number of CF rules, it is assumed, in accordance with the existing data, that for any language there are only finite kinds of syntactic features and for each kind of feature there are only a finite number of values.

[40] Here the subscripts *i*, *j* indicate the identities of anaphors and other nouns. Two nouns having the same subscript means that both of them refer to the same entity, i.e. one is the antecedent of the another; and two words having different subscripts means that they refer to different entities.

(7.14c) $Mary_i$ said [$_S$ John hurt her_i].

(7.14d)*Mary said [$_S$ $John_i$ hurt him_i].

Examples (7.13) suggest that reflexives should have their antecedents within the same clause. On the contrary, examples (7.14) suggest that pronouns should be bound by antecedents outside the clauses that the pronouns appear. In fact, these are not necessarily so. In the terminology of GB theory, reflexives must be bound by antecedents within the same *governing category*, and pronouns must *not* be bound by such antecedents. Here we need not bother about the technical definition of the notion governing category ; what we need to know is that anaphors bear certain distance limitations to their antecedents.

Since an anaphor refers to the same entity as its antecedent, they must share the same syntactic features. An argument against CFG from feature agreement can thus be made. There is no problem regarding the feature agreement between pronouns and their antecedents because, as far as weak generative capacity is concerned, a sentence comprising pronouns but no reflexives is always grammatical. (Whenever a governing category consists both of a pronoun and a noun eligible for being an antecedent, it can always be interpreted as that two different subscripts are assigned to the noun and the pronoun. The resulting sentence is therefore grammatical.) The focus is the argument from feature agreement between reflexives and their antecedents.

The strategy of the rebuttal is to list all possible types of constructions containing reflexives and then capture them by a set of grammar rules designed in the light of method (7.12). One typical situation is, as shown in (7.13), that anaphors and antecedents coexist in the same clause, be it a matrix clause or a subordinate clause. Such construction can be managed by rule schemata (7.9), (7.10) and (7.15) with lexical insertion rules (7.16).

(7.9) $S \rightarrow NP{:}[\alpha]\ \ VP{:}[\alpha]$

(7.10) $NP{:}[\alpha] \rightarrow \cdots\cdots\ N{:}[\alpha]\ \cdots\cdots$

(7.15) $VP{:}[\alpha] \rightarrow \cdots\ V \cdots\ \ NP{:}[\alpha]\ \cdots$

(7.16a) $N_{[1ST,\ SINGULAR]} \rightarrow$ *myself*/······

(7.16b) $N_{[2ND,\ SINGULAR]} \rightarrow$ *yourself*/······

(7.16c) $N_{[3RD,\ SINGULAR,\ MALE]} \rightarrow$ *himself*/······

(7.16d) $N_{[3RD,\ SINGULAR,\ FEMALE]} \rightarrow$ *herself*/······

(and so on for the insertions of plural reflexives)

Note that (7.10) is applied twice so as to maintain the feature agreement between the subject NP (antecedent) and the object NP (reflexive). However, these rules lead to ungrammatical sentences such as (7.17).

(7.17) **himself hurt John.*

To avoid such undesirable consequences, we could rewrite (7.15) and (7.10) as:

(7.18) VP:[α] $\rightarrow \cdots$ V$\cdots$ NP_{OBJECT}:[α] $\cdots$

(7.19) NP_{OBJECT}:[α] $\rightarrow \cdots$ N_{OBJECT}:[α] $\cdots$

and state precisely that only N_{OBJECT} can be rewritten into reflexives. Yet this proposal is still problematic for it cannot generate grammatical sentences like (7.20) in which reflexives occur in the subject position of non-finite subordinate clauses.

(7.20) $John_i$ persuaded [$_S$ $himself_i$ to do the job].

Again, we can solve this problem by differentiating N into $N_{SUBJECT\ OF\ FINITE\ CLAUSE}$, $N_{SUBJECT\ OF\ NONFINITE\ CLAUSE}$, and N_{OBJECT}, and add corresponding grammar rules. We need not go on explore more constructions including reflexives because the above instances are sufficient to show that (7.12), though able to counter-attack all arguments from feature agreement, the application of which may be very complicated.

8. The Argument from Unbounded Dependency

Now let us turn to another argument which is considered by many linguists the most fatal assault on CFG. This argument focuses on the phenomena of unbounded dependency. In linguistics *dependency* refers to any relation between two elements or positions in a sentence by which the presence, absence or form of an element in one position is correlated with the presence, absence or form of another element in another position.[41] Subject verb agreement, the topic of the last section, is clearly an example of dependency, for the form of a subject depends on the form of the verb accompanying the subject. Yet it is a kind of *local* dependency as the interdependent subject and verb exist in the same clause; and we have seen in the last section that local dependency phenomena give no real challenge to CFG. *Unbounded dependency* phenomena are those in which interdependent elements/positions may be distant from each other at an unbounded length. *Topicalization* is one kind of unbounded dependency phenomena:

(8.1a) *Dr. Chan, MediCenter had employed.*

(8.1b) *Dr. Chan, nurses thought MediCenter had employed.*

(8.1c) *Dr. Chan, I think nurses thought MediCenter had employed.*

---

[41] Trask (1993) p.77.

According to linguistic tradition the topic of (8.1a-c), viz. *Dr. Chan* , is considered correlating with the empty object position following the transitive verb *employed* . For the verb *employ* selects an object with the HUMAN feature and so the topicalized NP should have this feature as well. Otherwise ungrammatical sentence would be formed:

(8.1d) **The knife, MediCenter had employed.*

Such dependency is unbounded because we can go on construct more topicalized sentences with an unbounded number of subordinate clauses so that the topicalized NP may be kept unboundedly distant from the empty object position.[42] Note that topicalization is not a phenomenon specific only to English; we can translate (8.1 a-c) directly into, say Chinese, topicalized sentences.

In this section attention is paid to another unbounded dependency phenomenon, *question formation*, which can be illustrated by:

(8.2a) *John bought a truck.*

(8.2b) *What did John buy?*

Apart from a few languages like Chinese, in many languages the NPs being asked about, or *wh*-phrases, appear in the initial positions of questions[43], and therefore questions have a word order different from the SVO order of statements. As in the case of topicalization, linguists have traditionally conceived the *wh*-phrases at initial positions as dependent on an empty position from where it is moved; moreover, the distance between *wh*-phrase and empty position may be unboundedly long. Below we focus on an argument, proposed by linguist Culicover, which is based on the unbounded dependency as shown in English questions. Yet we should beware that the discussion about it applies to all kinds of arguments from all sorts of unbounded dependency existing in all languages.

8.1 Culicover's agrument

Culicover (1982) claims that the English questions as illustrated in (8.2b) can be captured by the following rule:

(8.1.1) $S \rightarrow NP\ (V_{AUX})\ NP\ V_T$ [44]

However, rule (8.1.1) cannot generate the question (8.1.2).

---

[42] Of course, this is a point made on *linguistic competence*. Due to factors such as limitation of memory, the actual human *linguistic performance* would not allow sentence of unbounded length.

[43] In traditional terminology wh-phrases are said to be *moved* to the initial positions. Note that PPs can also be wh-phrases. For example, '*In what manner did John carry out his presentation*?' In the following we concentrate on NPs only, yet the readers should bear in mind that our discussion also applies to PPs.

[44] Here 'V_{AUX}' stands for auxiliary verb and 'V_T' stands for transitive verb. In the following passage we use 'V_I' standing for intransitive verb.

(8.1.2) *Which beach did Mary believe that the ocean damaged?*

The sentence (8.1.2) is captured by rule (8.1.3).

(8.1.3) S $\rightarrow$ NP (V_{AUX}) NP V_T *that* NP V_T

Again, both rules (8.1.1) and (8.1.3) cannot generate a question containing two subordinate clauses like (8.1.4).

(8.1.4) *Which beach did Mary believe that Sam believed that the ocean damaged?*

Adding the rule (8.1.5):

(8.1.5) S $\rightarrow$ NP (V_{AUX}) NP V_T *that* V_T NP *that* NP V_T

is not a real remedy because then the grammar is still unable to generate questions with three or even more subordinate clauses. Eventually Culicover came to the conclusion in general, for any (CF) phrase structure grammar containing a finite number of rules like (8.1.1), (8.1.3) and (8.1.5) it will always be possible to construct a sentence that the grammar will not generate. [45] That is, since the length of question (= the number of subordinate clauses) is unbounded, there are an infinite number of questions forms and an infinite number of rules are needed to generate them. Recall that a CFG contains, by definition, only a *finite* number of grammar rules, so it is impotent in generating questions.

The key of Culicover argument and other arguments from unbounded dependency, is the infiniteness of grammar rules. The point of proposing CFG formalism is to find some way constraining the generative capacity of unrestricted rewriting systems. If CFGs are permitted to have an infinite number of rules, then for any symbol string we can construct a set of corresponding grammar rules to generate it. Obviously, this would make CFGs indistinct from unrestricted rewriting systems with respect to generative capacity. Thus, CFGs must not have infinitely many rules.

At first glance Culicover s argument is undoubtedly wrong. There is no problem with its first premise that an infinite number of question forms follow from an unbounded number of subordinate clauses. The fault is in its second premise, namely, that an infinite number of grammar rules follow from an infinite number of question forms. Surely there are infinitely many statement forms, such as:

John said that Mary loves Jim;

Peter said that John said that Mary loves Jim;

Linda claimed that Peter said that John said that Mary loves Jim;

Francis told me Linda claimed that Peter said that John said that Mary loves Jim;

---

[45] Culicover (1982) p.32. Note that the object of Culicover s argument is the inadequacy of *phrase structure grammar*. However, as he conceived phrase structure rules have CF character, there is no problem including him as an antagonist to CFGs.

and so on. But they can be captured by a finite set of rules, the most important of which are:

(8.1.6a) S $\rightarrow$ NP VP

(8.1.6b) VP $\rightarrow$ V_I *that* S

(8.1.6c) VP $\rightarrow$ V_T NP *that* S

Note that there is an S inside VP and a VP inside S. By repeated applications of rules (8.1.6a) and (8.1.6b/c) an infinite number of statement forms are obtained. It seems, therefore, that the infinitely many question forms can also be generated by a finite set of recursive grammar rules.

However, there is some regularity in statements that questions do not possess. In statements, all transitive verbs are followed by object NPs, but this is not necessarily so in the case of questions. Taking (8.1.2) as example, the transitive verb *damaged* is not followed by any object NP (in fact, the object is *which beach*, which has been moved to the initial position). In examples (8.2b), (8.1.2) and (8.1.4) all the wh-phrases correspond to the object NP positions. If this is the unique case of question formation, then questions can be generated by the following rule:

(8.1.7) S $\rightarrow$ NP (S) NP V.

Unfortunately, wh-phrases may correspond to subject NP positions; moreover, any NP in any subordinate clause can be moved to the initial position and become the wh-phrase. For example:

(8.1.8a) *Who bought the truck*?

(8.1.8b) *What did Mary believe that damaged the beach*?

(8.1.8c) *Who did Mary believe that believed that the ocean damaged the beach*?

Thus, even if there is a set of recursive grammar rules which generates the various sorts of questions, it is not as simple as that generates statements.

Maybe it would help if we postulate an empty category *e* which corresponds to a null sound in speaking and a blank space in writing. Then we can construct rules:

(8.1.9) N $\rightarrow$ *e*.

(8.1.10) S $\rightarrow$ NP S.

With the ordinary grammar rules expanding S, (8.1.9) and (8.1.10) together generate questions like *Which beach did Mary believe that the ocean damaged e*? which reads just the same as *Which beach did Mary believe that the ocean damaged* ?

But there must be some restriction on the application of (8.1.9), namely, that it can be applied only once; otherwise the grammar would generate many ungrammatical sentences like *"Which beach did e believe that e damaged e"*. A naïve modification of (8.1.9) is:

(8.1.11) N $\rightarrow$ *e* only if there is no other *e* exists in previous derivations.

Certainly the "only if" here implies that (8.1.11) is not a CF rule. In order to maintain the CFness of natural language, we should look for some mechanism to keep track of the number of 'e' without exploiting non-CF tools.

8.2 Gazdar's GPSG formalism

Now it is time our linguistic genius, Gerald Gazdar, comes into play. He proposed a grammar formalism called Generalized Phrase Structure Grammar (GPSG) which is claimed to be able to deal with all sorts of linguistic phenomena in a solely CF manner. Before introducing the various technical notions and mathematical tools in overcoming the challenge of unbounded dependency, let us see how GPSG represents the syntactic structure of, say, the question *"What did John buy"*:

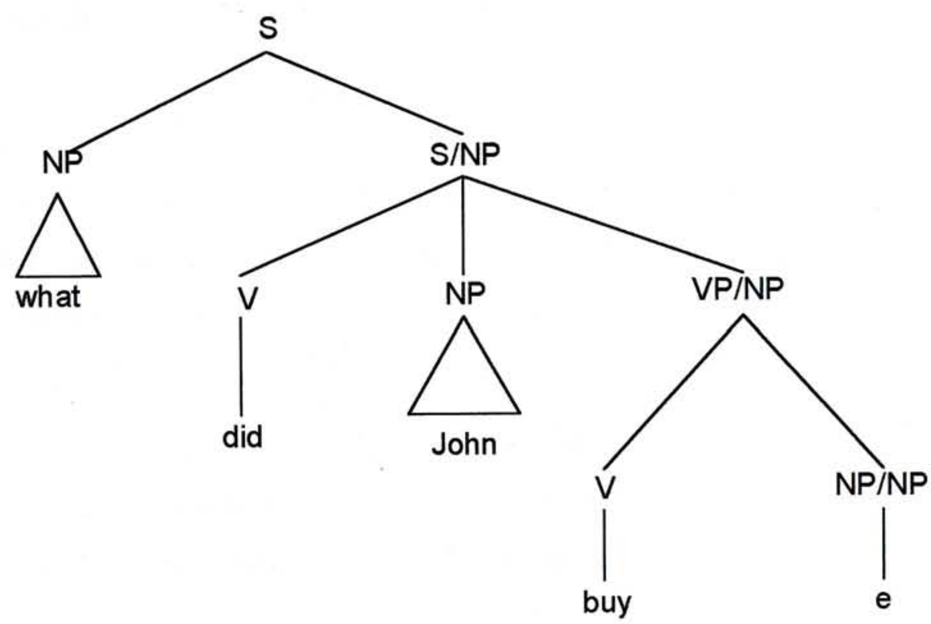


Fig.8

In this example the empty category *e* rewrites the node NP/NP, which is a notation with the first 'NP' meaning an NP position while the second 'NP' implying that an NP is missing. The exemplar question is generated first by the rule

(8.2.1) S $\rightarrow$ NP S/NP.

This rule means that an NP is moved to the sentence initial position and that it is followed by a sentence missing an NP. Then the node S/NP is further rewritten in accordance with

(8.2.2) S/NP $\rightarrow$ V_{AUX} NP VP/NP.

And the VP/NP node, meaning that a VP node missing an NP somewhere, is rewritten by the rule

(8.2.3) VP/NP $\rightarrow$ V NP/NP.

Finally, the NP/NP node, which means an NP position missing an NP, is rewritten into the empty category *e*. In general, whenever the mother node carries a '/NP' component, there is *one and only one* daughter node also carries, or inherits, the '/NP' notation. This requirement is very important because it ensures that there is one and only one 'missing NP' in each step of derivation. Recall that we are looking for a mechanism generating sentences with equal number of moved constituents and empty categories *e*. Such mechanism consists of three things: first, a set of rules like (8.2.1) which introduce the moved constituents and the '/α' notation (α stands for any syntactic category); second, a set of rules transferring each '/α' notation from some mother node to exactly one daughter node; finally, a set of rules eliminating such notation.

In details, we call the categories NP, VP etc. the *basic categories* and VP/NP, VP/PP etc. the *derived categories*. The basic categories constitute the basic non-terminal vocabulary V_N, and the set of derived categories, $D(V_N)$, is defined as follows:

(8.2.4) $D(V_N) = \{ \alpha/\beta\colon\ \alpha, \beta \in V_N \}$.

The set of rules introducing derived categories such as (8.2.1) are individually specified. The set of rules transferring derived categories from mother nodes to daughter nodes is specified by the following condition:

(8.2.5) The rule $\alpha/\beta \rightarrow \sigma_1 \cdots \sigma_i/\beta \cdots \sigma_n$ 'is a legitimate grammar rule if and only if

(i) '$\alpha \rightarrow \sigma_1 \cdots \sigma_i \cdots \sigma_n$ 'is itself a rule of the grammar in hand; and

(ii) α, σ_i dominate β according to some rules in the grammar in hand[46].

Finally, there is the rule schema replacing derived categories with the empty category *e*:

(8.2.6) $\alpha/\alpha \rightarrow e$, where α is any basic category.

Back to the example in Fig.8. Provided with the GPSG formalism, what we need to generate this question is just some basic rules 'S $\rightarrow$ NP VP', 'VP $\rightarrow$ V NP' and 'S $\rightarrow$ V_{AUX} NP VP'. Nearly all kinds of questions can be generated by the GPSG formalism associated with sufficient basic rules.[47] We would not bother with this point here as it has been fully elaborated by those textbooks on GPSG. Our issue is to

[46] The point of condition (ii) is to rule out rules like 'V $\rightarrow$ PP'.

[47] It is the questions with the wh-phrases in the matrix subject position, like '*Who saw the man?*', '*Who believed that the ocean damaged the beach?*', which cannot be generated by that method. Such questions can be generated by employing 'S $\rightarrow$ NP VP' and rewriting the NP into a wh-phrase. Gazdar (1981) proposes another set of rules for this kind of questions.

consider whether GPSG is a genuine CFG, i.e. whether the introduction of derived categories would make the resulting grammar beyond the scope of CFG.

Let us use the term "*basic rules*" to denote those rules containing only basic categories, "*derived rules*" those grammar rules transferring derived categories from mother nodes to daughter nodes, "*introduction rules*" the rules introducing derived categories, and "*elimination rules*" those rules schematized as (8.2.6). Clearly the format of all these rules are of a CF nature; that is, they share the form $A \to \psi$. The problem is whether there are infinitely many rules. We should first notice that the set of basic rules is finite; moreover, there is a limit concerning the *length* (number of daughter nodes in a grammar rule) of grammar rules, for unbounded length would lead to an infinite number of rules. Note also that the set of basic categories V_N is finite; and if it contains n elements, then $D(V_N)$ contains at most n^2 derived categories. That is, the set of derived categories is also finite. These remarks imply that the set of elimination rules is finite (for only a finite number of derived categories have the form α/α), and it makes no sense to posit an infinite number of introduction rules. Then the proof of the finiteness of the set of derived rules runs as follows:

(i) There are some basic rules which do not give rise to derived rules, namely, those rules that do not satisfy the condition (ii) of (8.2.5). Remove these rules from the grammar in hand.

(ii) Consider just one rule in the remaining basic rules. Let us say this rule be ' VP → V NP PP ' and assume counterfactually we have only NP, VP, PP and V as the basic categories. The possible derived rules are:
VP/NP → V/VP NP PP; VP/NP → V NP/NP PP; VP/NP → V NP PP/NP;
VP/PP → V/PP NP PP; VP/PP → V NP/PP PP; VP/PP → V NP PP/PP;
VP/V → V/V NP PP; VP/V → V NP/V PP; VP/V → V NP PP/V.
There are no rules for VP/VP, for all categories having the form α/α lead to elimination rules $\alpha/\alpha \to e$. Note that among these 9 rules there are many ones unlikely to be real grammar rules. Such rules include 1) rules rewriting VP/V, for there seems no VP missing a verb, and 2) rules at the leftmost column, for a V does not dominate VP, PP and NP and hence violates the condition (ii) of (8.2.5).

(iii) Anyway the above example shows that if there are n basic categories, then for any basic rule of length l there are at most $l(n-1)$ corresponding derived rules. In other words, each basic rule leads to at most $k(n-1)$ derived rules, where $p \geq k \geq 1$ and p is the maximum length of grammar rule.

(iv) As there are only finitely many basic rules and each basic rule lead to a finite number of derived rules, the summation of all the derived rules is still finite.

So far we have proved that GPSG is a genuine CFG and it is able to generate English questions which exhibit unbounded dependency. In fact, the GPSG formalism can be applied to all sorts of unbounded dependency phenomena. Taking topicalization as an instance we can generate all topicalized sentences by the same method as that generates questions except this time the daughter node NP in (8.2.1) is rewritten into a topicalized NP rather than a *wh*-phrase. This completes our repudiations of all kinds of arguments against CFG from unbounded dependency.

9. Conclusion

So far we have seen that

(i) there is no common feature among various natural languages which is beyond the scope of CFG; and

(ii) the unique evidence for the inadequacy of CFG is found in Swiss German.

By no means the story ends here; in the future linguists may discover that some more constructions of certain languages have non-CF nature. But further discovery[48] can at most add a few words to conclusion (ii) without changing (i).

A practical implication is that AI researchers are fully justified to use CFG in parsing and generating most natural languages. This is a favourable result as we have seen that CFG is simpler for processing and its mathematical and computational properties are well explored. Moreover, CFG operation is fast: in the worst cases it takes time proportional to mere the cube of the length of the sentence under processing.

However, the Swiss German case is, after all, a genuine example against the weak adequacy of CFG, and hence any linguist must look for another grammar formalism with a greater generative capacity. CSG was considered and has been rejected since there is some other formalisms, such as *indexed grammar*, *tree adjoining grammar*, *head grammar*,[49] whose generative capacity lies between those of CFG and CSG, and which has been shown that it is able to cope with the Swiss German case. Apart from pursuing a formalism generous enough to cope with all natural languages phenomena, we should also restrict the generative capacity of syntactic rules/principles so that the grammar does not generate something not of natural languages. If there are two theories which are equally good in coping with empirical data, then it is preferable to adopt the one with a smaller generative capacity.

---

[48] There are reports suggesting that Swedish and an African language Engenni *may* contain non-CF constructions, but no rigorous proof has been put forth.

[49] For the definitions and explications of these grammar formalisms, refer to Partee (1990).

There is still ongoing debate about which of these formalisms is the most suitable one for weakly generating all natural language sentences. But we need not bother about it because the whole discussion of weak generative capacity would not have a satisfactory result. A successful scientific theory has to meet two conditions: 1) it explains the phenomena; 2) it *barely* explains the phenomena, i.e. its explanatory power is not so strong that it can deduce something that is not the fact. Applied to the science of linguistics, these two requirements mean that a grammar formalism should generate all and only those grammatical sentences. The Swiss German example thus gives us a dilemma: on one hand, CFG is inadequate because it is unable to manage this counter-example; on the other hand, any formalism stronger than CFG would be too strong as they are able to generate sentences exhibiting cross-serial dependency pattern, which does not appear in most languages. It is "too ad hoc" to stipulate that there is some linguistic/cognitive mechanism making the full power of a correct grammar formalism not being realized in most cases. After all, the aim of science is to discover some regularity among the phenomena, and now it is obvious that there is no regularity can be found in the study of weak generative capacity: either the formalism is too weak (CFG), or it is too strong (indexed grammar, etc.).

Hence our most important conclusion: Chomsky is wrong in saying that "the study of weak generative capacity is of rather *marginal* linguistic interest", because it is of no genuine theoretical interest at all.

Glossary

[Symbols]

A*	the Kleene star operation of set A, i.e. the set formed by concatenating the members of A any number of times in any order
x^i	the symbol string formed by concatenating i tokens of the symbol x.
$\|z\|$	the length of the symbol string z; i.e. the number of symbols contained in z.
e	the empty category, i.e. the linguistic element with no phonetic or calligraphic realization

[Abbreviations]

CFG	context-free grammar
CFL	context-free language
CFness	context-freeness

[Key Concepts]

context-free grammar

a grammar formalism or a natural language grammar with each of its rules rewrites a non-terminal symbol (say, A) into a set of terminal and/or non-terminal symbols without considering other symbols in A's vicinity

context-free language

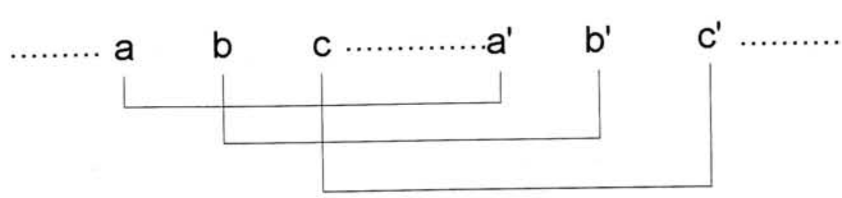
a language generated by a context-free grammar

context-freeness

the property of a natural language which is itself a context-free language

cross-serial dependency

a sentence is said to exhibit cross-serial dependency if it has pairs of elements with the following pattern:



generative capacity

the range of sentences which can be generated by a type of grammar formalisms

grammar

(a description of) the system in the human linguistic competence which combines words and morphemes into sentences; it is described as a set of rules.

grammar formalism

a formal system which specifies the formal conditions on a grammar

grammatical

a sentence is grammatical if it is (judged by native speakers to be) well-formed. Note the distinction between grammatical and **meaningful / interpretable**. A classic example is *a colourless green idea sleeps furiously*, which is grammatical but not meaningful.

syntax

in this master thesis it is used as the synonym of grammar.

Bibliography

Aho, A. and Ullman, J. (1972)

The Theory of Parsing, Translation, and Compiling

Bresnan, Kaplan, Peters, and Zaenen (1982)

"Cross-serial Dependencies in Dutch"

in Savitch et.al. (eds.) *The Formal Complexity of Natural Language*

Chomsky, N. (1965)

Aspects of the Theory of Syntax

Culicover, P.W. (1982)

Syntax

Daly, R. (1974)

Applications of the Mathematical Theory of Linguistics

Gazdar, G. (1981)

"Unbounded Dependencies and Coordinate Structure"

in *Linguistic Inquiry* Vol12 No2 pp.155-184

Gazdar, G. (1982)

"Phrase Structure Grammar"

in Jacobson & Pullum (eds.) *The Nature of Syntactic Representation* pp.131-186

Gazdar, G. & Pullum, G. (1985)

"Computationally Relevant Properties of Natural Languages and Their Grammars"

in Savitch et.al. (eds.) *The Formal Complexity of Natural Language*

Grinder & Elgin (1975)

Guide to Transformational Grammar: history, theory and practice

Harrison, M. (1978)

Introduction to Formal Language Theory

Higginbotham, J. (1984)

"English is not a Context-Free Language"

in Savitch et.al. (eds.) *The Formal Complexity of Natural Language*

Higginbotham, J. (1985)

"Reply to Pullum"

in *Linguistic Inquiry* Volume 16, Number 2, Spring 1985

Hopcroft, J. and Ullman, J. (1979)

Introduction to Automata Theory, languages, and Computation

Langendoen, D. (1977)

"On the Inadequacy of Type-3 and Type-2 Grammars for Human Languages"

in Hopper (eds.) Studies in Descriptive and Historical Linguistics

Langendoen, D. and Postal, P. (1984)

"Comments on Pullum's Criticisms"

in *Computational Linguistics*, Volume 10, Numbers 3 - 4

Levelt, W. (1974)

Formal Grammars in Linguistics and Psycholinguistics Volume 2

Partee, B. *et.al.* (1990)

Mathematical Methods in Linguistics

Peters, S. (1987)

"What is Mathematical Linguistics?"

in Savitch et.al. (eds.) *The Formal Complexity of Natural Language*

Postal, P.(1964)

"Limitations of Phrase Structure Grammars"

in Fodor and Katz (eds.) *The Structure of Language*

Postal, P. and Langendoen, D. (1984)

"English and the Class of Context-Free Languages"

in *Computational Linguistics*, Volume 10, Numbers 3 - 4

Pullum, G. (1984)

"On Two Recent Attempts to Show that English is not a CFL"

in *Computational Linguistics*, Volume 10, Numbers 3 - 4

Pullum, G. (1985)

"*Such That* Clauses and the Context-Freeness of English"

in *Linguistic Inquiry* Volume 16, Number 2

Pullum, G. and Gazdar, G. (1982)

"Natural Languages and Context-Free Languages"

in Savitch et.al. (eds.) *The Formal Complexity of Natural Language*

Savitch, W. (1987)

"Context-Sensitive Grammar and Natural Language Syntax"

in Savitch el.al. (eds.) *The Formal Complexity of Natural Language*

Savitch, W. (1989)

"A Formal Model for Context-Free Languages Augmented with Reduplication"

in *Computational Linguistics*, Volume 15, Number 4

Sells, P. (1982)

Lectures on Contemporary Syntactic Theories

Shieber, S. (1985)

"Evidence against the Context-Freeness of Natural Language"

in Savitch et.al. (eds.) *The Formal Complexity of Natural Language*

Trask, R.L. (1993)

A Dictionary of Grammatical Terms in Linguistics