Journal of Liberal Arts and Sciences.
Korea Maritime University. Vol. 3, 1995

Regularity of Natural Language

Jeong - Ryeol Kim*

자연언어의 규칙성

김 정 렬

목 차

Abstract
Ⅰ. Introduction
1. Definition of Context-free Language
2. Definition of Regular Language
Ⅱ. Natural Language Is Not Context-free?
1. Long-distance Dependencies
2. Non-identity dependencies
3. Identity Dependencies
4. Cross-serial Pairship [x x] Arguments
5. Multiple Dependencies [a^n b^n c^n]
Ⅲ. Natural Language is Not Regular?
1. The Nested Pairship [x x-1] argument
2. Bi-directional Branching Argument
Ⅳ. Conclusion
References

Abstract

Natural Laguage appears to be a mass of abstract chain of sounds delivering intended meaning of a speaker. However, it is a strictly rule-governed system for any child to acquire and use to communicate with others. This learnability is a concrete piece of evidence that natural language is regular. This paper investigates all the different pieces of evidence across different languages on the issue of regularity of natural language, and further argues that natural language is regular enough for a context-free grammar to be hired to describe it.

* 한국해양대학교 교양과정부 조교수(언어학 전공)

Ⅰ. Introduction

Until the early 80's there was a tendency among linguists to consider natural language to be non-context-free. Pullum and Gazdar (1982) have cast substantial doubts on the validity of the proposed arguments in favor of adopting this position. In this paper, I summarize the arguments for the position that (1) natural language is not context-free and (2) natural language is not regular, though it may be context-free, drawing citations from the original papers or discussions where possible. Otherwise I will rely on citations in Pullum and Gazdar (1982). I also present counter-arguments by Pullum and Gazdar (1982) who take the position that natural language is context-free.

According to the Chomskian hierarchy, languages are arranged in descending order of complexity as follows :

Type 0 : transformational languages

Type 1 : context-sensitive languages

Type 2 : context-free languages

Type 3 : regular or finite state languages

The relationship among these different types of languages is such that Type 0 includes Type 1-3, and Type 1 includes Type2-3, and Type 2 includes Type 3.

The types of languages crucially related to this paper are Type 2 (context-free) languages and Type 3 (regular) languages as described below.

1. Definition of Context-free Language

Context-free languages (hereafter CFL) constitute a class of languages generated by Type 2 grammars (or context-free phrase structure grammar). Context-free phrase structure grammar (hereafter CF-PSG) consists of a finite number of context-free phrase structure rules (hereafter CF-PSR). CF-PSR's, according to Chomsky (1959 : 130), are constrained in such a way that each rule allows only one symbol to appear on the left-hand side of arrow and the symbol to be rewritten into a finite string of symbols on the right-hand side.

2. Definition of Regular Language

Regular languages constitute a class of languages generated by Type 3 gram-

mars, Finite State Automata (FSA) or others : L(G1), L(G2) and L(G3) below. Type 3 grammars, according to Chomsky (1959 : 130), are constrained in such a way that they allow only a non-terminal symbol to be rewritten either as a terminal symbol or as a terminal symbol followed by a non-terminal symbol as in G={A->a, A->aB}. FSA consists of states (initial, final and any finite number of intermediate states) and transitions between states. A sentence is generated when it reaches the final state by a coordination of appropriate transitions corresponding to terminal symbols with states corresponding to non-terminal symbols. The following kinds of grammars share the same weak generative capacity with regular grammar[1)]

L(G1) : Languages generated by grammars with rules satisfying the restriction that there is a single symbol on the left of the arrow and on the right of the arrow is either a single terminal symbol, or one non-terminal symbol followed by one terminal symbol, G1 = {A->a, A->Ba}.

L(G2) : Languages generated by grammars with rules satisfying the restriction that there is a single symbol on the left of the arrow and on the right of the arrow is either a terminal string, or a terminal string followed by one non-terminal symbol, G 2 = {A->x, A->xB}.

L(G3) : Languages generated by grammars with rules satisfying the restriction that there is a single symbol on the left of the arrow and on the right of the arrow is either a terminal string, or one non-terminal symbol followed by a terminal string, G 2 = {A->x, A->xB}.

It is important to note that grammars generating regular languages allow only one kind of branching as we can see in the given grammars, i.e. the right-hand side of arrow is either one terminal string followed by one non-terminal symbol (right-branching), or one terminal string preceded by one non-terminal symbol (left-branching).

---

1) These grammars were presented as examples of regular languages by Dr. Gregory Lee in his Mathematical Linguistics Course.

Ⅱ. Natural Language Is Not Context-free?

1. Long-distance Dependencies

The belief that long-distant dependencies cannot be treated by CF-PSG has been expressed by Grinder and Elgin(1973) and Bach (1974). This view is well-represented in Bresnan (1978 : 38) in the following claim :

… the distant type of agreement … cannot be adequately described even by context-sensitive phrase structure rules (hereafter CS-PSG) for the possible context is not correctly describable as a finite string of phrases.

Bresnan (1978) assumes that natural languages have unbounded dependencies which cannot be adequately described by CS-PSG, but can be described by a transformational grammar. Consider the interaction of wh-extraction and number agreement in (1) :

(1)
Which problem/problems did your professor say (she thought)* was/were unsolvable?[2)]

In (1), there is a dependency between *problem/problems* and *was/were* across an unbounded number of strings. However, this type of dependency can be accommodated even by G2 described above which has the same generative capacity of a regular grammar as shown in Pullum and Gazdar (1982 : 474) :

(2)
S -> Which problem did your professor say T
S -> Which problems did your professor day U
T -> she thought T | you thought T | was unsolvable?
U -> she thought U | you thought U | were unsolvable?

The most important criticism against Bresnan (1978) is in Gazdar (1982) in which he claims, contrary to Bresnan, that natural languages do not allow dependencies whose domain is not describable by CF-PSG. In fact, Ross (1967) demonstrates that a

2) x* stands for any number of string x.

transformational grammar is too powerful an apparatus and needs a package of constraints along with it for dealing with movement transformations in English. Gazdar (1982), and Pullum and Gazdar (1982) argue further that only the kinds of long-distance dependencies and syntactic concord which can be described by CF-PSG appear in natural languages.

2. Non-identity dependencies

Chomsky (1963 : 378-79) argues that a grammar must be able to describe non-identity dependencies between substrings in a sentence. The examples shown in Chomsky (1963) are English comparatives :

(3)

a. That one is wider than this one is DEEP.

b. That one is wider than this one is (*WIDE).[3)]

In (3) a repeated element is deleted and a non-repeated element receives heavy stress. We find similar situation when noun phrases are involved :

(4)

a. John is more successful as a painter than Bill is as a SCULPTOR.

b. John is more successful as a painter than Bill is (*as a PAINTER).

The constructions in (4) show that comparative constructions have a non-identity dependency between two elements to be compared. Chomsky thus claims that this type of dependency is beyond the range of the theory of context-free grammars.

Gazdar and Pullum (1982) criticize Chomsky's premise that non-identity dependency is inherently non-context-free by providing a set of CF-PSR's to deal with a sen-

3) Chomsky changed his grammatical judgment on this sentence in Chomsky (1977:122) in the course of arguing against Bresnan's analysis of comparative clauses. The following sentence is grammatical according to Chomsky's later judgment:

What is more, this desk is higher than that one is HIGH.

Pullum and Gazdar (1982:476-79) disagree with Chomsky's judgment. They claim the ungrammaticality of the above sentence is in parallel with the sentences below:

John is as fat as Bill is (*obese).

John is fatter than Bill is (*obese).

The ungrammaticality of these sentences are due to the fact that the two adjectives employed are synonymous.

tence such as (4a) as in the following manner :

(5)

{uxkyw | x,y L((a|b)*) & x =\= y } where {u,x,k,y,w} is a terminal vocabulary.

where u : John is more successful

a : as a painter

b : as a sculptor

k : than Bill is

w : null

Note that x and y surrounding k are not identical in language (uxkyw) in (5).

(6)

(a) S -> uS'w | uS"w

(b) S' -> CS'C | Dk | kD

(c) S" -> AB' | BA'

(d) A -> CAC | a(D)k

(e) B -> CBC | b(D)k

(f) A' -> a(D)

(g) B' -> b(D)

(h) C -> a | b

(i) D -> C(D)

The final derivations by CF-PSG in (6) represent the fact that string 'x' between 'u' and 'k' is not identical to the string 'y' between 'k' and 'w' as follows :

(7)

S, uS'w, uDkw, uCkw, uakw (x=a, y=null)

S, uS'w, uDkw, uCkw, ubkw (x=b, y=null)

S, uS'w, ukDw, ukCw, ukaw (x=null, y=a)

S, uS'w, ukDw, ukCw, ukbw (x=null, y=b)

S, uS"w, uAB'w, uakB'w, uakbw (x=a, y=b)

(Notice that uakbw = sentence (4a))

S, uS"w, uBA'w, ubkA'w, ubkaw (x=b, y=a)

S, uS'w, uCS'Cw, uaDkCw, uaCkCw, uaakaw (x=aa, y=a) ···

Therefore, the non-identity dependencies such as the one in English comparatives can still be derived by the CF-PSG given in (6). Chomsky's claim that non-identity dependency is inherently non-context-free fails to be supported.

3. Identity Dependencies

Elster (1978) touches on the issue of the non-context-free nature of natural language and attempts to show that English is not a CFL. In order to show the inadequacy of CF-PSG he illustrates with a sentence that cannot be extended by indefinite repetition of one subpart.

(8)

a. The first two million numbers in the decimal expansion of pi are a1 a2 ⋯ a2,000,000.

b. The first (two million)^2 numbers in the decimal expansion of pi are a1 a2 ⋯ a (2,000,000)^2. ⋯

k. The first (two million)^k numbers in the decimal expansion of pi are a1 a2 ⋯ a(2, 000,000)^k.

k'. *The first (two million)$^{k+q}$ numbers in the decimal expansion of pi are a1 a2 ⋯ a(2,000,000)^k ⋯

Notice the identity dependency between the mentioned number in the subject position and the actual repetition of digits in the complement position which exists in the pattern, (the first)(two million)^m (numbers in the decimal expansion of pi are)(a1 a2 ⋯ a(2,000,000)^m) or else the resulting sentence will be ungrammatical, as shown in (8k'). Elster (1978)claims that this dependency cannot be resolved by CF-PSG, and shows the following additional example which parallels (8) :

(9)

*The two largest animals in the zoo are an elephant.

The point of Elster (1978) is that in the construction 'The W1 are W2', the number of entities listed in W2 must correspond to the number named in W1.

Pullum and Gazdar (1982) do not completely agree withthe grammatical judgments on the sentences in (8) made by Elster (1978), though they agree with the ungrammaticality of (9). The ungrammaticality of (9) is explained by the rule in En-

glish that requires predicates to agree with their subjects in number between singular and plural. However, according to this rule, the following sentences should be grammatical and in fact Pullum and Gazdar claim that they are.

(10)

a. The two largest animals in the zoo are Mickey, Minnie and Donald.

b. Here are six random integers : 3, 17, 9, 5, 8.

c. Our three main weapons are fear, surprise, ruthless,efficiency, and a fanatical devotion to the Pope.

They recognize the oddity of the sentences in (10) and claim that his oddity does not lie in syntax, but has to do with pragmatic felicity instead.

4. Cross-serial Pairship [x x] Arguments

1) Sentences with *respectively*

Bar-Hillel and Shamir (1960 : 96) present the earliest argument that English is an [x x] language. An [x x] language represents dependencies between the first member, the second member, the third member, ··· the nth member ··· of the first x and the first member, the second member, the third member, ··· the nth member ··· of the second x respectively, and thus is not a CFL. These dependency pairs often occur with the word *respectively* such as in example(11) :

(11)

John, Mary, David, ··· are a widower, a widow, a widower, ···, respectively.

(11) is a grammatical sentence if and only if the three dots are replaced by a string of any number of proper names and the same number of parallel complements.

Langendoen (1977 : 4-5) attempts to reconstruct the *respectively* argument with a different example :

(12)

The woman and the men smokes and drink respectively.

Langendoen defines the *respectively* construction in this way : {x1 x2 respectively | x1 is a member of L (the woman | the man)$^+$ and x2 is the corresponding string that has *smokes* in place of *the woman* and *drink* in place of *the men* from (smokes |

drink)$^{+}$} where (x^{+} stands for one or more x's).

English requires that in addition to constraint that dependent subparts contain the same number of elements, there be number agreement between an element in (the woman |the men) and its corresponding element (smokes | drink). Based on the type of sentence found in (12), Langendoen (1977) concludes that English is not a CFL.

Pullum and Gazdar (1982) claim the requirement in Langendoen (1977 : 4-5) for number agreement between elements in dependent subparts in a sentence containing *respectively* does not stand up any more, if we consider the following examples :

(13)

a. *The woman and the men smokes and drinks respectively.

b. ?The woman and the men smoke and drink respectively.

While (13a) is ungrammatical, the grammaticality of (13b) is of uncertain even though number agreement between *the woman* and *smokes* is violated. Also, the following sentence shows that the requirement of numerical matching between subparts is not rigid :

(14)

They are dating Mary, Carol, and Lisa respectively.

Note in (14) the number of elements in the subparts do not match. Thus, it is not well supported to conclude that English is not a CFL based on *respectively* construction.

2) Dutch

Huybregts (1976) proposes that Dutch is not a CFL on the following grounds :

When transitive infinitival VP's are nested, VP's will generally be of the type NP1 NP2 ··· NPn V1 V2 ··· Vn.V1···Vn-1 are taken to be verbs that select a direct object NP and a complement VP and Vn is some transitive verb. For all i (where 1<=i<=n), NPi is the direct object of Vi, and is present because Vi is subcategorized to require it. Dutch has an infinite subset with an unbounded cross-serial dependency of the type [a 1 a2 ··· an b1 b2··· bn] which is not context-free.

Pullum and Gazdar (1982 : 485-6) make an additional observation about Dutch.

Consider the following sentence :

(15)
dat Jan [Marie1 Pieter2 Arabisch3 laat1 zien2 schrijven3]
that Jan Marie Peter Arabic let see write
'that Jan let Marie see Peter write Arabic.'

The bracketed portion contains three verbs and three NP's. The first NP is the direct object of the first V and the subject of the second; the second NP is the object of the second V and the subject of the third V; and the third NP is the object of the third V.

An additional verb can be inserted between the third NP and the first V if it is an intransitive VP-complement taking verb :

(16)
dat Jan [Marie1 Pieter2 Arabisch3 wil laten1 zien2 schrijven3]
that Jan Marie Peter Arabic will let see write
'that Jan let Marie see Peter will write Arabic.'

Based on (16), Pullum and Gazdar (1982) provide a context-free grammar to generate the subset of Dutch discussed above as follows :

(17)
(a) Syntactic rules
A->BCD | CE
C-> BCF | CG | BH | I
(b) Lexicon
B : Marie, Pieter, other personal names
D : schrijven, other transitive infinitives
E : liegen, other intransitive verbs
F : laten, other transitive VP-complement-taking infinitives
G : willen, other intransitive VP-complement-taking infinitives
H : laat, other finite transitive VP-complement-taking verbs
I : wil, other finite intransitive VP-complement-taking verbs

Rules in (17) successfully generate the bracketed parts of both (15) and (16), thus

Huybregts's conclusion based on the sentences (15) or (16) fails to be supported.

3) Mohawk[4)]

Postal (1964) argues that the interaction of the processes of nominalization and incorporation in Mohawk indicate the existence of a property that places Mohawk outside the CFL's.

Mohawk is a Subject-Verb-Object order (SVO) language, and the object can be incorporated into verb as follows :

Subject incorporation
N-subj V
1 2 => 1+2 [pfx N-stem Base]IV

Object incorporation
N-subj V N-obj
1 2 3 => 1 3+2 [pfx N-stem Base]TV

The rules indicate that a verb incorporates the noun-stem of its subject if it is intransitive, or the noun-stem of its direct object if it is transitive. I cite a transitive case in (18) from Postal (1964 : 147) :

(18)
ka-ksa?a ka-nuhwe?-s ne-ka-nuhs-a?
the-girl pfx-like -sfx prt-pfx-house-sfx
'The girl like the house.'

ka-ksa?a ka-nuhs-nuhwe?-s
the-girl pfx-house-like-sfx
'The girl likes the house.'

In addition, in Mohawk the following nominalization rule exists :

Nominalization
[pfx-VS-sfx]V
1 - 2 - 3 => [1 2 -hsra/tsra 0]N

4) Mohawk is a Northern Iroquoian language of Quebec and upper New York state.

(19)

ka-nuhs-nuhwe?-s

pfx-house-like-sfx

'likes the house.'

ka-nuhs-nuhwe?-tsra

pfx-house-like-NMLZR

'liking the house.'

The interaction of nominalization and object incorporation may result in the case in which a verb with an incorporated subject (or object) noun-stem occurs with an overt subject (or object) NP. Consider the schematic example supposing that the English words are the translation into Mohawk of the corresponding English word :

(20)

[[The man][N[[[[house-praise]V-ing]N-like]V-ing]N-admired]V[[[[house-praise] V-ing]N-like]V-ing]N]S

In the case (20) the incorporated noun-stem in the verb exactly matches the noun-stem in the external NP in the object position. This is string-copying over an infinite set of strings (the set of noun-stems), hence Postal's claim is that Mohawk is an [X X] language and is not a CFL.

Langendoen (1977) introduces a formal definition of Mohawk :

A regular language : $Lf = a(c|d)^*eb(c|d)^*e$

$Lm = \{axebxe|\ x \text{ is a member of } L((c|d)^*)\}$

where

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)'

e = the translation into Mohawk of 'house'

Although Mohawk, which is described by Lm, is similar in structure to a regular language, described above by Lf, the addition of the [X X] cross-serial dependency adds to its complexity such that Mohawk can no longer be described as a CFL. The most important premise here, according to Pullum and Gazdar (1982 : 492), is that

no Mohawk sentences are of the type in which the incorporated noun-stems in a verb and the noun-stems in an external noun phrase are not identical. This is an empirical issue which Pullum and Gazdar (1982 : 494-95) further look into. They find counter-examples such as :

(21)

wa?khnekahniu? ne otsi?tsa?

I-liquid-bought flower/wine

'I bought the wine.'

In (21) the meaning of the incorporated stem resolves the ambiguity of the external stem *-tsi?ts-*, which means both 'flower' and 'wine'. (22) also furnishes another example in which the external NP and the incorporated noun-stem are not paired :

(22)

wakeselehtahni : nu?se? ne? 'bike'

I her-vehicle-bought bike

'I bought her a bike.'

It is evident that Mohawk has sentences in which an incorporated noun-stem fails to match the noun-stem of the direct object of its host verb. Consider the following example as one more piece of evidence :

(23)

a. i?i k-nuhwe?s ne sawatis hra-o-nuhsa?

I like John('s) house

'I like John's house.'

b. i?i k-nuhs-nuhwe?s ne sawatis

I house-like John('s)

'I like John's house.'

(23b) shows that when the direct object NP of a verb contains a possessive NP modifier, it is possible to incorporate the noun-stem denoting the possessed entity, keeping the external NP separate. Thus, Pullum and Gazdar(1982 : 497) argues that Mohawk is of the form {axebye | x, y are drawn from L((c | d)*)}, which is a regular language, instead of {axebxe | x is a member of L((c | d)*)} which is not. Note that the

change of representation from 'axebxe' to 'axebye' shows that the cross-serial identity dependency is not necessary in Mohawk.

5. Multiple Dependencies [a^n b^n c^n]

1) English is Not a CFL

Higginbotham (1984) supports his claim that English is not a context-free language by comparing its complexity with that of a regular language which contains the same terminal vocabulary. Higginbotham (1984 : 225) defines the following regular language L :

L = the woman such that (the man such that)* she (gave (this| him) to (this | him))* left is here

English = {the woman such that (the man such that)^n she (gave (this |him) to (this | him))^n left is here : n >= 0, and, reading from left to right, the number of occurrences of 'this' never exceeds by more than 1 the number of occurrences of 'him'}

Possible sentences of English : [5]

the woman such that she left is here

the woman such that [the man such that] she [gave this to him] left is here

the woman such that [the man such that the man such that] she [gave this to him gave this to him] left is here···

We see that English requires parity in the number of occurrences between [*the man such that*]^n and [*gave* (*this* |*him*)]^n, and 0 or more center-embedded 'such that' – relatives, so that any deletions or additions of [*the man such that*] or [*gave* (*this* | *him*) to (*this* | *him*)] will unbalance the parity of occurrences, consequently the sentence becomes ungrammatical. Also, for each of the *n* occurrences of [*the man such that*], and there must be an occurrence of *him* somewhere among the n occurrences of [*gave* (*this* | *him*) to (*this* | *him*)]. If this is violated, an ungrammatical sentence arises as follows :

(24)

*The woman such that [the man such that the man such that] she gave [this

5) Most people I have talked to increasingly disagreed with the grammatical judgment of Higginbotham's as the number of *such that* embeddings increases.

to him gave this to this].

We can see a type of multiple dependencies occurring here. There exists a parity dependency between [*the man such that*]^n and [*gave* (*this* | *him*) *to* (*this* | *him*)]^n, and additionally a pronominal reference dependency between (*the man such that*)^n and (*him*)^n. It is well known that CF-PSG can easily express single dependencies such as {a^n b^n}, but it cannot express a multiple dependency such as {a^n b^n a^n}where a sequence of a's on the right cannot overbalance the sequence of a's on the left.

Ⅲ. Natural Language is Not Regular?[6]

1. The Nested Pairship [x x^{-1}] argument

1) Bracket Notation

The bracket notation in a syntactic analysis shows a type of nested pairs as follows :

(25)

[s John [vp loves [np a dog]np]vp]s

The example (25) shows [X X^{-1}] pattern between each paired bracket labels, and [X X^{-1}] is not regular. However, bracket notation was developed as a tool for the convenience of linguists and does not necessarily have any phonetic reality.

2) Hypothetical language

We know that the prototypical SVO language has the characteristics that a preposition precedes an NP, and a complementizer precedes S, while the prototypical SOV language has the characteristics that a postposition follows an NP, and a complementizer follows S. However, if some natural language shows that PP starts with a preposition and ends with a postposition, and/or a sentence starts with a complementizer and ends with a complementizer, then the language is not regular, since it shows a nested pairship which cannot be resolved by a regular grammar or a FSA.

2. Bi-directional Branching Argument

As shown earlier in section 1, a regular language allows only one type of branching, either left-branching or right-branching. An SOV language such as Korean or

6) The discussion under this heading is based on the presentations by Dr. Gregory Lee in his mathematical linguistics course.

Japanese is not regular, if we accept the conventional view of branching which most linguists adopt. A simple PSG for Korean may be represented as follows :

(26)

S -> NP-ka/-i VP

VP -> NP-ul V

NP -> (S-nun) N

Lexicon : N=Mary, John, -kes 'fact', casin 'self'

V=anta 'know'

The PSG in (26) generates the following sentences :

(27)

Mary-ka John-ul anta

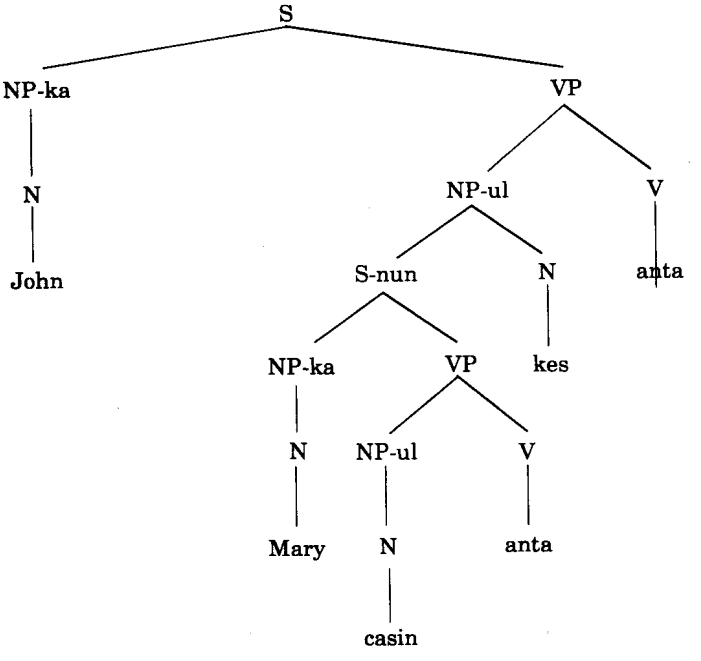
'Mary knows John.'

John-i Mary-ka casin-ul anta-nun kes-ul anta.

'John know Mary knows himself/herself.'…

The second sentence in (27) is represented in the following tree diagram :

(28)



We can see that the sentence requires both types of branching. If we can show that a language is not confined to either left-branching or right-branching, and then we say it is bi-directional branching; thus the language is not regular, since a regular grammar, FSA or others which have the same generative capacity (G1, G2 and G3 in section 1) can only deal with unidirectional branching.

Ⅳ. Conclusion

The discussed arguments for the non-context-free nature of a natural language was classified into five different types :

(1) the long-distance dependency argument :
This argument is supported by Bresnan based on the interaction of wh-extraction and number agreement.
(2) the identity dependency argument [X Y] (X=\= Y) argument :
Chomsky argued for this based on comparative constructions.
(3) the identity dependency argument :
Support for this argument was supplied by Elster's pi-expansion.
(4) the cross-serial pairship [x x] argument :
This argument finds support in the *respectively*
construction, a subset of Dutch and Mohawk.
(5) the multiple dependency [an bn cn] argument :
Higginbotham's *such that* relative constructions is cited for this case.

The discussions about non-regular nature of natural language based on :

(1) the center-embedding (or the nested pairship [x x-1]) which was supported by arguments based on bracket notation and hypothetical language.
(2) the structure of SOV languages which appear to be bi-directional.

The criticism made mostly by Pullum and Gazdar (1982) is based on a very fundamental premise which I also agree with : there is the need for both empirical validity and formal validity of the proposed arguments. In the above arguments empirical validity was violated in two different ways : first, disagreements on grammatical judgments such as those found in Chomsky's comparatives and in Higginbotham's

such that relatives weakens their arguments. Second, an argument is weakened by overgeneralizations based on a narrow representation of the linguistic facts. Examples of this are the arguments based on *respectively* construction, Mohawk and Elster's pi-expansion.

Pullum and Gazdar, by showing that even an FSA can deal with a long-distance dependency bring into question the formal validity of the arguments by Bresnan. Also they point out that CF-PSG can treat the non-identity dependency and thus Chomsky's premise that a language with the non-identity dependency is not context-free is not well-supported. Additionally, they argue that CF-PSG can enhance the power to permit some [X X] languages to be generated by revising Huybregts's treatment of a subset of Dutch.

As we can see, most arguments for non-context-free nature of natural language are not satisfactory; whether or not a natural language is regular remains to be seen. The possible necessity of bi-directional branching of SOV languages seems to indicate that natural language is not regular, however, the discussions about the right-headedness of Korean and Japanese still raise some doubts about the necessity of bi-directional branching.

References

1) Bach, E. (1974) *Syntactic Theory*. Holt Rinehart and Winston,New York.
2) Bar-Hillel, Y. and E. Shamir (1960) "Finite State Languages : Formal Representations and Adequacy Problems." reprinted in Y. Bar-Hillel (1964), *Language and Information*, 87-98. Addison-Wesley, Reading, Mass.
3) Bresnan, J. (1978) "A Realistic Transformational Grammar", M. Halle, J. Bresnan and G. Miller, eds. *Linguistic Theoryand Psychological Reality*. Cambridge, Mass. : MIT Press.
4) Chomsky, N. (1959) "On Certain Formal Properties of Grammars", I*nformation and Control* 1, 91-112.
5) Chomsky, N. (1963) "Formal Properties of Grammars", *Handbook of Mathematical Psychology* vol. II. R. Luce, R. Bush and E. Galanter, eds. John Wiley, New York.
6) Chomsky, N. (1977) "On wh-Movement", P. Culicover, T. Wasow and A. Akmajian. eds. *Formal Syntax*. MIT Press, Cambridge, Mass.
7) Elster, J. (1978) *Logic and Society* : *Contradictions andPossible Worlds*. New York : John Wiley.
8) Gazdar, G. (1982) "Phrase Structure Grammar", P. Jacobson and G. Pullum. eds. *The Nature of Syntactic Representation*, 131-186. D. Reidel Publishing Co., Dordrecht, Holland : Grinder, J. and S. Elgin (1973) *Guide to Transformational Grammar*. Holt Rinehart and Winston, New

York.

9) Higginbotham, J. (1984) "English is Not a Context-free Language", *Linguistic Inquiry* 15-2, 225-34.

10) Huybregts, M. (1976) "Overlapping Dependencies in Dutch", *Utrecht Working Papers in Linguistics* 1, 24-65.

11) Langendoen, D. (1977) "On the Inadequacy of Type 3 and Type 2 Grammars for Human Languages", P. Hopper, ed. *Studies inDescriptive and Historical Linguistics : Festschrift forWinfred P. Lehmann*, 159-171.

12) Postal, P. (1964) "Limitations of Phrase StructureGrammars", J. Fodor and J. Kats, eds. *The Structure of Language : Readings in the Philosophy of Language*, 137-151. Englewood Cliffs, New Jersey : Prentice-Hall.

13) Pullum, Geoffrey and Gazdar (1982) "Natural and Context-freeLanguages", *Linguistics and Philosophy* 4, 471-504. D.Reidel Publishing Co., Dordrecht, Holland, and Boston.

14) Ross, J. (1967) "Constraints on Variables in Syntax", *Ph.D.dissertation*. MIT.