無限の長さの文

上林洋二

本稿は，自然言語は，原理的に生成文法で扱い得ないという証明をした Langendoen and Postal (1984)〔以下 L & P と略記〕の好意的な書評である。「好意的」とわざわざことわったのは，筆者が目にしたこの本の書評の二つ (Thompson (1986), Abbott (1986)) ともが，あまり好意的と言えるものではないこと，そして，この本があまり多くの人に好意的に受け入れられることはないであろうという予測を筆者がたてていることによる。

ただ，好意的な書評を始める前に，ことわっておかねばならないことがある。それは，この本に対して予想される次の反論（それは Thompson にも Abbott にも見られないが）には，今回筆者は直接立ち入らないということである。

それは，Chomsky 側からなされるであろう反論だが，L & P の議論は E 言語に関するものであり，E 言語の存在を前提としているが，E 言語などというものは，その存在が疑わしい，Chomsky の生成文法は E 言語ではなく，I 言語，即ち人間の頭の中の文法的知識の研究なのだから，L & P の議論は Chomsky にとっては，痛くもかゆくもない，という論である。(E 言語と I 言語の区別は Chomsky (1985) 参照)

L & P は実はこの反論に対しても一章を設けている (Chap. 6. Ontological Escape Hatches) のだが，そこでの議論は，結局，文法あるいは言語といったものに関する存在論的問いにゆきつくことになる。L & P は自分たちの主張は Katz (1981, 1983) の立場と結びつきやすくはあるが，必ずしもそれを前提とはしないと述べているのだが，Chomsky 側の上のような反論に答えるためには，結局，言語という抽象的な実在が存在する，そして言語学とはその抽象的な実在を対象とした研究であって，Chomsky の説くように人間の知識の研究，即ち心理学，ひいては生物学となるものではないという，Katz の議論に依存しなければならないと思われる。

現在の筆者は，この言語学の研究対象は何かという問いに対する Chomsky と Katz の論争に判定をくだすつもりはない。ただ，Chomsky の立場に立ってしまえば，L & P の議論はまったく空振りに終わってしまうから，以下では一応 Katz の立場に立って，L & P の議論を見ていくことにしたい。

と言っても，L & P の「自然言語は原理的に生成文法では扱い得ない」という主張のために，認めるべき前提は，初期の生成文法の文献にはいくらでも認められたことだけである。即ち，「生成文法とは，当該言語に属する文すべてを，そしてそれだけを生成する文法である」。

そして，ひとたびこの前提を認めれば，論理的にゆるがない L & P の議論にのることになる。

*　　　　*　　　　*

自然言語に属する文の長さ，及び文の集合の濃度について，広く受け入れられている仮定は，

(1) 個々の文の長さには有限数の上限はない

(2) 個々の文の長さは，aleph_0 より小である（有限である）

(3) 文の集合の濃度は aleph_0 である

であるが，L & P はまず (1) を裏付ける従来の議論のほとんどが健全でないことを指摘する。例を一つあげると，今までに作られた英語の最も長い文の前に，I know that …，とか，I know that I know that …，といったものをつけ加えれば，さらに長い英語の文を作ることができるという事実を (1) の裏付けとする議論があった。しかし，この事実は，個々の文の長さにある有限の上限（たとえば，1 億 8 千万）が存在するという仮定と両立可能である。なぜなら，いまだかつて（文の長さを単語の数で数えるにせよ，音素の数で数えるにせよ）1 億 7 千万を越える長さの英語の文は作られておらず，従ってその文に，I know that を付加してもなお英語の文であるということが確かめられていないからである。

さて，L & P は (1) を正当化する根拠はないのかというと，そうではなく Katz (1966, 122) に見られる議論が説得的であると説く。その議論は，ある記号列が文であるか非文であるかの決定が，その記号列の長さがある特定の数を越えるか越えないかによってなされるような，そういう特定の数を措定するのには，正当な根拠がない，記号列の長さという性質は，それが文であるか非文であるかを決定するのには関らないという議論である。

そして，L & P はこの議論を認めるのなら，全く同様の議論で (2) は否定

されると説く。文の長さの上限をたとえば 1 億 8 千万とするのが恣意的ならば，**aleph$_0$** とするのも同様に恣意的だというのである。

Abbot は，ここで L & P は特定の上限を定めることと，単に文を(無限の長さのものは排除して）有限の長さのものに限ることとの相違を見のがしている，と評している。これはたしかにポイントとなる問題で，本稿の題を「無限の長さの文」としたのも，この問題点の故であるが，この点は後で触れる。

さて，L & P は特定の有限数と同様に，**aleph$_0$** という上限も恣意的であるというのだから，即ち，(2)ではなく，

(4) 個々の文の長さに上限は全くない (**aleph$_0$** でも **aleph$_1$** でも **aleph$_2$** でも…でもない)

を認めるべきだとする。(2)を仮定すると，単に長さが **aleph$_0$** である，あるいは **aleph$_0$** を越えるというだけで，他は文法的な文と同じ構造を持つ記号列を非文としてしまうことになるからである。

(4) を仮定すれば，必然的に，

(5) 自然言語に属する文の collection は，set ではない (濃度をもたない)

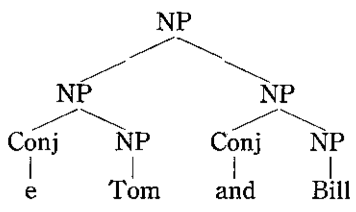
が得られる。これは論理的帰結であるが，今このことをカントールの論法と並行的に証明すると以下のようになる。

自然言語においては，普通，同じ構成素どうしを等位接続してできあがったものは，もとのものと同じ構成素に属する。たとえば，

(6) Tom and Bill

は，L & P の記法によると

(7)



のようになるが，Tom も Bill も NP であればそれらを等位接続してできあがったもの (これを coordinate projection と呼ぶ) も NP に属する。このような Closure Principle がすべてのカテゴリーにおいて成り立つかどうかは疑問だとしながらも，少なくとも文というカテゴリーには認められると，L & P は述べる。

(8) L をある自然言語におけるカテゴリー S のすべてのメンバーの col-

lection とし，CP(U) を文の Set U の coordinate projection とすれば，次の式が成り立つ。

$$(\forall U)(U \subseteq L \rightarrow CP(U) \in L)$$

次に，ある自然言語 L において等位構造を持たぬ文の Set を S_0 とする。S_0 の濃度を $aleph_0$ と仮定する。S_0 は例えば次のようである。

(9) So＝{Babar is happy, I know that Babar is happy, I know that I know that Babar is happy, I know that I know that I know that Babar is happy …}

今，(8) が認められているのだから，L は S_0 の subset のうち，2 個以上の要素を含むものすべての coordinate projections も含んでいるはずである。この coordinate projections と S_0 の和を S_1 とすると，たとえば

(10) S_1＝{Babar is happy, I know that Babar is happy, I know that I know that Babar is happy, …, Babar is happy and I know that Babar is happy, Babar is happy and I know that I know that Babar is happy, …}

のようになる。ここで S_1 は S_0 の power set のうち，空集合だけを除いたものと一対一の対応をしていることになる。従って S_1 の濃度は $aleph_1$ である。

同様に，S_1 の subset のうち，2 個以上の要素を含むものすべての coordinate projections も Lに含まれるはずである。この新たにできた coordinate projections と S_1 の和を S_2 とすると，全く同様に S_2 は S_1 の power set から空集合だけを除いたものと一対一の対応をする。従って S_2 の濃度は $aleph_2$ である。

全く同様の議論が S_3, S_4 … と無限に読いていくわけであるが，これで自然言語に属する文の collection が set ではないということが次のように証明される。

今，ある自然言語 L の文の collection が set である，即ち濃度を持つと仮定してみよう。その濃度を #L とする。L は (8) に従っているのだから，L はその subset のうち，2 固以上の要素を含むものすべての coordinate projections からなる subset（これを Z とする）を含んでいるはずである。

(11) $(\exists Z)(Z \subseteq L)$

$Z = \{x:(\exists y)(y \subseteq L \wedge x$ は y のcoordinate projection$)\}$

そして，L の中には全く等位構造を持たぬ文も含まれているから，Z は L の proper subset である。即ち Z の濃度を #Z とすると，

(12) $\# Z < \# L$

が成り立つ。ところが Z は (11) の定義からして，L の power set と対応するものである。従って

(13) $\# L < \# Z$

(12) かつ (13) という矛盾が生じてしまった。背理法により，仮定は誤りである，即ち，自然言語の文の collection は set ではない，ということが証明された。

そして，set ではない collection (L & P の用語に従えば，megacollection) を（数学的な意味で）生成する装置はあり得ない。従って

(14) 自然言語は原理的に，生成文法では扱えない

という結論が導き出される。

この，カントール的な証明において，L & P は (4) を仮定する必要はなく，ただ Closure Principle を認めればよい，(4) はむしろ Closure Principle の系であったと述べているが，実はここでも無限の長さの文を認めるかどうかということが決定的である。無限の長さの文を認めないのなら，L の subset のうち，その濃度が aleph_0 のもの，または aleph_0 を越すものの coordinate projections は文にならない，即ち (8) はそのままでは認められないからである。

つまり，(4) という仮定を認めなければ，今までの議論は全く成り立たないが，(4) を認めれば，そこからは (14) が必然的に導かれるのであるから，ポイントは (4) を認めるか否かに絞られるわけである。

L & P は (4) を認めるべき理由として，有限の長さを持つ文と無限の長さを持つ文とは，性質が等しいということをあげている。無限の長さを持つ文を言語学から排除すると，次のような奇妙な結果が生じる，と L & P は言う。

(15) (a) * John, I think they visited and his father_1.

(b) * John, I think they visited and his father_1 and his father's father_2.

(c) * John, I think they visited and his father_1 and his father's father_2 … and … and … father_k.

(d) * John, I think they visited and his father_1 and his father's father_2 …and … and … $\text{father}_{\mathrm{aleph}_0}$.

(15) は等位構造制約を破っているために非文法的になる文だが，無限の長さの文を言語学から排除する立場に立つと，英文法は (15a)—(15c) の非文法性は教えてくれるが，(15d) の非文法性は示せないということになる。しかし，言

語学の中で (15d) の非文法性が示せないのなら，どこで示したらよいのか，と L & P は述べている。

(4) から (14) へ至る推論の厳密さに比べ，この (4) を認めるべきだという L & P の主張には，たしかに若干の疑問は残ろう。しかし，これもやはり首尾一貫性ということを考えれば，L & P の主張を認めてよいと思われる。1 億 8 千万の長さを持つ文は認め，無限の長さを持つ文は認めないというのは，片手落ちなのである。

* * *

以上見てきたように L & P の証明はまったく妥当なものである。L & P の主張に対する可能な反論は，彼ら自身が preface で述べているが，無限の長さの文を排除するか，最初に述べたように，言語という実在の存在を否定するかしかない。その点で Thompson の書評はまったく的はずれであるとしか言えない。

さて，L & P は (14) を主張して，それでは自然言語の妥当な理論はどのようなものかという問いに対して，それは（比喩的に使われていると思うが）Model-theoretic grammar であると答えている。つまり，生成文法，比喩的に言えば Proof-theoretic grammar は原理的に認められない，当該言語に属する文すべてを，そしてそれだけを生成することは不可能であって，できることは，当該言語の文はこのような条件を満たすものであるという，その条件を提示することだというのである。そして，この観点から，現行の文法理論のほとんどが原理的に認められず，ただ一つ，Arc Pair Grammar (Johnson and Postal (1974) 参照) だけが，若干の修正を加えれば生き残るとしている。

* * *

最後に，もう一つ予想される意見，即ち，このように極度に抽象的・形式的な議論をして，それが実際の言語研究に何の役に立つかという意見について考えてみよう。

すでに見たように，L & P が正しければ，現行の文法理論のほとんどは退けられる。それも，技術的にうまくないというのではなく，原理的に認められないというのだから，これは言語研究に与えるインパクトははかり知れないほど大きい。

だが，もし L & P が正しくなかったとしても，と言うのは，この議論の前提が誤っているとしても，この研究の意義は大いにあると思われる。というのは，彼らは，最近の言語研究の一つの流れ，抽象化・形式化の傾向を，極度ま

でおし進めればどこまで行くかという，一つの見本を提示してくれたからである。これ以後の言語研究が，さらに抽象化の道をたどるにせよ，その反対の方向をとるにせよ，L & P は一つの道しるべになると思われる。

そういった意味からも，(4) の仮定から (14) の結論に至る論理の厳密さは高く評価すべきである。

REFERENCES

Abbott, B. (1986) "Review of *the Vastness of Natural Languages*," *Language*, 62, pp. 154–157.

Chomsky, N. (1986) *Knowledge of Language: Its Nature, Origin, and Use*, Praeger, New York.

Johnson, D. and P. Postal (1980) *Arc Pair Grammar*, Princeton University Press, Princeton, New Jersey.

Katz, J. (1966) *The Philosophy of Language*, Harper & Row, New York.

——— (1981) *Language and Other Abstract Objects*, Rowman and Littlefield, Totowa, New Jersey/Basil Blackwell, Oxford.

——— (1983) "An Outline of Platonist Grammar," in T. Bever, J. Carroll and L. Miller (eds.) *Talking Minds: The Study of Language in Cognitive Science*, The MIT Press, Cambridge, Massachusetts.

Langendoen, T. and P. Postal (1984) *The Vastness of Natural Languages*. Basil Blackwell, Oxford.

Thompson, H. (1986) "Shorter Notice of *The Vastness of Natural Languages*," *Journal of Linguistics*, 22, pp. 241–242.