

生成文法が生成する文の数について

外池 俊幸

概要

この短い論文では、生成文法で共通認識になっていると思われる、自然言語の文法は無限の数の文を生成できるという主張と相反する主張をする。

Pullum & Scholz (2010) で、生成される文の無限性について詳細な検討がなされている。本稿では、その中で、以下の二つを取り上げ、それぞれ少し捉え方を変えた提案をする。

1 回帰性(同じ範疇の下に同じ範疇のものを組み込める)を持つ構造を作れるようになったことで、人間は言語を獲得し、それが思考を可能にした。生成文法では、回帰的な構造を作る操作を繰り返せば、無限の数の文を作れると考えてきた。本稿では、回帰的な構造を作れるようになったことが言語の起源だと考えるが、複数回その操作を繰り返すということが可能になったのは、理論的、派生的なことだと主張する。2 俳句の文字数は、17 個、日本語のモーラの数はおよそ 100 個。すると、モーラ列の可能性を計算すると、100 の 17 乗。これは 10 の 34 乗になる。この数は、現在のスーパーコンピューターでも扱うのが大変な非常に大きな数である。このことから、俳句という有限の文字数の可能性を考えても、たとえ有限であっても大きな数は、非常に大きいということが言える。

1. はじめに

この論文では、生成文法が提案された当初から仮定されてきた、自然言語の文法によって無限の数の文が生成されるという考え方を検討し、それはあくまで理論的な可能性であり、自然言語の本質ではないという主張をする。この主張は、すでに、Pullum & Scholz (2010) が主張したことであり、この論文では、その基本的な主張を、継承する。もう一つの主張も、Pullum & Scholz (2010) が指摘したのであるが、日本語の短詩形式の俳句を取り上げ、17 文字の位置に、日本語の可能なモーラが入るとすると、組合せの可能性が、10 の 34 乗になるという。字余り、字足らずは数に入れていない。さらに、例えば、句の初めに、「ん」がくることはあまりないだろうとか、様々な、数を限定する条件が存在する可能性があるが、それも検討されてはいない。文字の数とモーラの数

掛け合わせて、いくつになるかを計算した結果だと書いている。計算は単純で、100種類のモーラがあって、それが入る可能性があるのは、17個なので、100の17乗。それを10の何乗にあたるかを計算すると、10の34乗という数が出てくる。

Pullum & Scholz (2010) は、言語の無限性について様々な観点から論じているが、本稿では、二つの問題に限定して論じる。関連する言語の無限性にかんする議論の検討は、今後の課題としたい。

筆者は、自ら10の34乗の計算をしたわけではない。実際の検討は、今後の課題としたい。しかし、これが正しいとすると、知り合いの数学者、スーパーコンピュータの専門家に尋ねたところ、その数が大きすぎて、現在のスーパーコンピュータで扱うにしても、かなり厳しい数だという指摘を受けた。

知り合いの数学者は、そういう指摘をしたイギリスの言語学者を評価すべきだという意見だった。そこで、この論文で指摘したいもう一つの点は、生成文法で伝統的に主張されてきた、自然言語の文法が無限の数の文を生成できるという主張が、いわば単なる理論的なことだということを主張したのに対して、俳句のように文字数もモーラ数も有限であるのに、その組み合わせの数が、10の34乗という、有限ではありながらとんでもない数になることを考えると、無限ではなく有限なのだが、大きな数は、とてつもなく大きいということである。日本人で俳句を作っている人は、ほとんどそのことは意識していないだろうと考えられる。しかし、それだけの組合せの可能性があるので、同じ俳句を作る可能性は、非常に低いし、新しい俳句を作ろうとする時に、同じ俳句を作ってしまう可能性があって困るということもほとんどありえない。

比喩的な捉え方になるが、むかし、*The Powers of Ten* という映像が作られた。湖のそばに寝転ぶ男の人の手から入り込み、人体のどんどん小さい部分に這いこんでいく。当時知られた最少の部分まで行くと、戻ってきて、今度は、湖の上からどんどん地球を離れていく。銀河系をはずれて、宇宙空間にまで進むと、星もなく非常に希薄な空間に達する。10の34乗という可能性は、その宇宙空間と似ていると言えるのではないだろうか。空間は広がっているのだが、星は非常に少ない。俳句の可能性は、数で言うと、非常に大きいのだが、実際に作られる俳句は、その可能性の極極一部であることになる。

言語の起源としての回帰性と、回帰性が保証されれば、自然言語の文法は無限の数の文を生成できるのだという主張を分けて考えることを本稿では主張する。Hauser et al. (2002) が、人間の言語に特有なこととして、回帰性があるという主張をした。その点には、異論はない。外池 (2012) で論じたように、約5万年前に、人間は、回帰的な構造を作れるようになった。そして、そのことが、人間が思考することを可能にした。しかし、そこからすぐに、回帰的な構造を作れるようになったのだから、無限の回数埋め込みを行うことができるので、人間は無限の数の文を生成できるという主張につなげるか

どうかには、検討の余地がある。

筆者は、回帰的な構造を作れることが言語を獲得することであり、思考を可能にしたという点には異論はないが、回帰的な構造を作ることを繰り返すこと（一つは、文の埋め込みを繰り返すこと）につながるわけではないと考える。回帰的な構造を作れるようになったことに言語の本質があるのであって、それを繰り返して無限の数の文を生成できるようになったというのは、派生的なことであると考え。実際に、回帰的な構造を作れることに言語の本質はあるのであって、回帰的な構造を作る操作を繰り返せることができることに言語の本質があるとは考えられない。しかし、生成文法の提案された当初からチョムスキー自身が無限の数の文を生成できるということを言い続けてきた。ここでは論じないが、Pullum & Scholz (2010) は、どうして、言語学者が、無限の文が生成できるという主張に魅せられるのかということまで論じている。そして、生成文法を中心となるグループの言語学者は、無限の数の文を生成できるということを、言語学の授業でも、議論の余地のないこととして教え、世界中の国で出版されている言語学の入門書にも、自然言語の文法が無限の数の文を生成するということが書かれるようになっていく。

日本には、かなりの数の言語学者、あるいは言語に関心を持っている人がいると考えられる。そして、生成文法が現代の言語学に強い影響を与えていることは明らかである。しかし、日本で生成文法の枠組みで研究している人たちのほとんどが、生成文法のメインストリームである、チョムスキーを中心として進められてきた枠組みをもとに研究している。そういう人たちは、自然言語の文法が無限の数の文を生成するということは、既に常識になっていると思われる。それに対して疑問を呈するために、Pullum & Scholz (2010) の重要な主張を二つ取り上げ、少し違うまとめ方で、本稿で示すことにした。

2. 生成文法に関するチョムスキーの主張の特徴

生成文法に関するチョムスキーの主張に関して、次のような一般的なことが言えるかもしれない。生成文法の提案をした初期の段階から、言語能力と言語運用を区別するという提案をしてきた。つまり、本稿では、自然言語の文法が無限の数の文を生成できるということは、あくまで理論的、そして派生的なことで、回帰的な構造を作れるようになったことに言語の本質があると主張したのだが、チョムスキーにとってみると、理論的であることは、一般的な主張をできることで、それを主張することは、生成文法の中核をなすことだと考えてきたのであろう。ここにチョムスキーの主張と、本稿での主張の違い、捉え方の違いがあるのではないかと考えられる。

それから、生成文法の研究に関して、ここで簡単に整理しておこう。生成文法の歴史

をあまり知らない人は、チョムスキーが、Chomsky (1957) で、生成文法を提案してから、半世紀以上にわたって生成文法の中心にいて、研究を引っ張ってきたと聞くと、異様な印象を持たれるかもしれない。他の領域では考えられないことだろうから。しかし、それが実態で、チョムスキーの存在の大きさ、貢献の大きさをあらわしている。チョムスキーのもとで、1980年代に博士号を取得した日本人の言語学者が、チョムスキーは、生成文法を提案し、その後、ほとんど一人で、生成文法の研究を引っ張ってきたと言われた。筆者もそう思う。参考文献に、チョムスキーの生成文法に関する著書や論文をたくさん挙げるべきであるが、煩雑なので、代表的なものにとどめる。Chomsky (1957, 1967, 1980, 1981, 1995).

3. 結論

人間の話している自然言語の文法が生成する文の数は、理論的には無限大であるが、それが自然言語の本質だとは考えられない。言い換えると、Hauser et al. (2002) が主張した、回帰性が文法の起源であり、それが人間の思考を可能にしたという主張には、賛成する。しかし、例えば、文をどんどん埋め込むことで、無限の数の文が作れるというのは、あくまで理論的な話である。回帰的な構造を作れるようになったということ、無限の数の文を作れるようになったということは分けて考えるべきなのである。

そして、俳句の例を考えると、数が限定されている短詩形の数の可能性を計算すると、大きな数は、とんでもない大きさの数になる。大きな数は、無限でなく有限であっても非常に大きいということである。(回帰性、そして言語の起源などに関しては、外池 (2012, 2013) で論じたので、そちらを参照。)

もう一つ、Pullum & Scholz (2010) の指摘したこととして、文の創造的使用であるかないかという問題がある。例えば、次のような例を挙げている、

The very big, The very, very big, The very, very, very big, …

この例のように、very を無限に繰り返して、新しい英語の表現を作ることはできるが、それも理論的な可能性の話であって、実質的な意味を持つかには疑問が残る。同様のことは、当然、日本語についても考えられる。

とても暑い、とてもとても暑い、とてもとてもとても暑い、…

生成文法の研究の歴史では、理論的な生成の可能性が重視されてきた。この論文では、

その研究の流れに揺り戻しを掛けることになるが、無限の文を生成可能であるという主張に対して、俳句のような文字数もモーラ数も限られている短詩形でさえも、無限ではないが、とんでもなく大きな可能性の広がりを持っているということがあるということである。生成文法が提唱されて以来、主張されてきて、現在は、世界中の言語学の教科書にも、自然言語は無限の数の文が生成できると書かれるようになってきた。本稿では、その主張を弱めた方がいいということと、有限でも大きな数の可能性は、非常に大きいという主張をした。

本稿では、Pullum & Scholtz (2002) が論じた問題の中から、二つを取り上げ、その捉え方に関しての提案をただけであって、残された課題は多い。一つ目の問題に関しては、自然言語の文法が、回帰的な構図を作ることを繰り返して、無限の文を生成できるというのは、理論的な可能性にすぎないという主張をしたが、理路的な可能性という捉え方の妥当性の検討をほとんど行っていない。その検討が必要になるだろう。

もう一つは、俳句に関してだが、日本語のモーラの数、約 100 ということの確認が必要だろう。モーラの出現頻度から考えて、何か違いがあって、それが組合せの数の解釈に影響を与えるのか検討すべきであろう。さらに、サンプリングとか、少数のデータを作ってみて、検討することも考えられるが、結び付きの可能性の数が大き過ぎて、少し取り出して検討することにどういう意味があるのか、現状では予測がつかない。自然言語処理の研究に早い段階から関わっている知り合いが何人かいるので、そういう人の意見を聞いて、どういうことが可能か検討したい。

参考文献

- Chomsky, Noam (1957) *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam (1969) *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Chomsky, Noam (1980) *Rules and Representations*. Oxford: Basil Blackwell.
- Chomsky, Noam (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam (1995) *The Minimalist Program*. Cambridge: MIT Press.
- Hauser, Mark D., Noam Chomsky, and W. Tecumsh, Fitch (2002) "The faculty of language: what is it, and did it evolve? *Science* 298, pp. 1569-1579.
- Pullum, Geoffrey K. and Barbara C. Scholz (2010) "Recursion and the infinitude claim," In Harry van der Hulst (ed.), *Recursion in Human Language* (Studies in Generative Grammar 104), 113-138. Berlin: Mouton de Gruyter. (Link is to typescript, not printed version. One typo is corrected: the published version gives the date of Epstein & Hornstein 2005 incorrectly as 2004.)
- 外池俊幸 (2012) 「言語の起源：自分自身との対話としての思考—人工生命の観点から—」『言語文化論集』第 34 巻、第 1 号、pp. 79-89. 名古屋大学国際言語文化研究科.

言語文化論集 第XXXV 卷 第1号

外池俊幸 (2013) 「人間精神にとっての言語の重要性」『言語文化論集』第34 卷、第2号、pp. 83-93.
名古屋大学国際言語文化研究科.