

文法が生成する文の数が無限だという主張について

外池 俊幸

1. まえがき

この短い論文では、生成文法が提案された初期の頃から、多くの生成文法学者が共有してきたと考えられる主張に対して、それには問題があることを論じる。具体的には、多くの生成文法学者が、自然言語の文法が生成する文の数は無限であると考えてきた。しかし、言語が回帰性という特性、言い換えると、無限回繰り返し埋め込みを行える特性を持つことを、言語が生成する文の数が無限だという根拠にすることには無理があると主張する。ただし、この言語についての回帰性に関する議論は、注意して行う必要がある。言語が誕生して以降、言語が言わば引きずってきている問題である。つまり、例えば、文の下に文を埋め込めるという構造を作れるようになったことが言語の起源であるが、理論的にはその埋め込みを何度も繰り返せるが、それはあくまで理論的な話で、実際には、処理に関する制約が存在し、簡単に処理出来なくなって、無限回の繰り返しは我々人間には出来ない。処理に関する議論を始める前に、無限は数学的にどう規定されるかを確認する：

無限集合

加算無限集合 数え上げ可能な無限集合。整数の全体など、自然数のように一つ一つ数え上げていける。

不可算無限集合 数え上げが不可能な無限集合。実数の全体。実数は、稠密性と言って、どんな二つの実数の間にも必ず他の実数が存在する。数直線上の数の全体からなる集合。

無限には、大きく二つあることは、以前から知られている。一つは、自然数が、集合の要素になる無限集合である。こちらは、数直線で考えると、離散的な自然数が、数直線上で点として捉えられ、自然数の間の部分には、何も無いことになる。これに対して、実数を要素とする集合は、同じく要素の数は無限だと言われるが、自然数の集合よりも、無限ではあるが、濃度が大きいと考えられる。そして、実数の稠密性と呼ばれる特徴を実数は持っていて、いかなる二つの実数の間にも、必ずその二つの実数とは違う実数が

存在する。自然言語の文の集合は、基本的に、文として認定されるものと、文として認定されていないものがあるが、離散的だとすると、自然数の集合の濃度と同じ無限の集合になると考えられる。しかし、本稿では、回帰的な構造を繰り返し構築する操作を、処理に関する制約から、無限に繰り返すと考えるのは、私たち人間が行っていることを反映していないので、実際には無限の集合にはならないと考える。

次のような等位接続の起こっている文を考えてみる。

I went to Tokyo from Nagoya yesterday morning and came back to Nagoya yesterday evening.

そして数直線を時間軸だと考えると、二つの文の間に、例えば、and went to Yokohama and という文が作れる。そして、went to Yokohama のすぐ前にも、すぐ後ろにも、別の動詞句を入れてやるのが可能だ。しかし、これは、実数の緻密性と似たところがあるように思えるかも知れないが、違ふと考へた方がよい。ただ、この問題を詳しく論じるのは、本稿の目標を大きく逸脱するので、ここではこれ以上論じない。

2. 生成文法と問題となる現象

生成文法の歴史では、提案された1950年代から、60年以上、ほとんどチョムスキー一人が引っ張ってきたと言っているだろうと、よく知られた日本人の生成文法学者が発言されたのを聞いたことがある。これは、やはり否定しようのないことだろう。チョムスキーに影響を与えた言語学者もいたと考えられるが、結果としては、生成文法のメインストリーム（と呼ばれることがある）を構成してきた動きはチョムスキー一人が引っ張ってきた。そして、メインストリームを構成する多くの研究者が、現在でも、自然言語の文法の生成する文の数は無限だと考えていると言っているだろう。なかには、文法が無限の数の文を生成するということが、自然言語に普遍的にみられる特徴だと考え、universal だという主張をしている研究者もいる。具体的にそういう仮定を考えていると思われる論文や著書は、枚挙にいとまがない。

生成文法のメインストリームの研究に対して、筆者は、1980年代に出てきた句構造文法の枠組みで、基本的には生成文法の一様であるが、言語研究を続けてきた。

生成文法の研究との関係で興味深いのは、Geoffrey K. Pullum の存在である。もちろん、変形規則を仮定しない文法の枠組みとしては、Joan Bresnan を中心に研究が進められたLFG(Lexical-Function al grammar)も重要な存在であるが、本稿では取り上げない。

文法が生成する文の数が無限だという主張について

Hauser, Chomsky and Fitch (2002) は、人間の言語に特徴的、従って他の動物には見られない特徴として、回帰的 (recursive) な構造を作れることがあると主張した。この点についてはその後かなりの議論があった。人間以外の動物が、回帰的な構造を作れるかどうかの問題になった。筆者は、言語学者で、人間以外の動物についての研究を行っているわけではない。人間の話している言語に関しては、回帰的な構造を作れるようになったことが言語の本質であり、それが言語の起源であるという考え方を、現在では、多くの言語学者が共有していると考えられる。ここでは触れないが、鳥の中に、回帰的な構造を持つ鳴き方ができるものがあるという研究がある。また、言語の起源を考える場合に、鳥の鳴き方は重要な意味を持つと考え研究を進めている研究者がいる。重要な研究動向のように思えるが、この論文では議論しない。

Pullum and Scholz (2012) は、The No Maximal Length Claim (NML) と呼べる主張があると指摘する。

The No Maximal Length Claim (NML)

どのような英語の表現に対しても、さらに長い英語の表現が存在する。そして、結果として、以下のような、The Infinitude Claim を主張することになる。

The Infinitude Claim

すべての英語の表現を集めたものは、無限の集合になる。

Pullum and Scholz (2012) は、これらの議論が成り立たないことを示した。いくつかの問題を取り上げて論じているが、本稿では、特に回帰性の問題を取り上げる。回帰性とは、あるカテゴリーの下に同じカテゴリーを埋め込むことができるという特性のことである。英語の例で言えば、例えば、文の下に文を埋め込む、あるいは、名詞句の下に名詞句を埋め込めるということである。回帰的な構造を作れるということは、何度もその操作を繰り返すことが可能で、言語が生成する文の数が無限だと考える根拠にされる可能性がある。本稿では、次のようなことを主張する。回帰的な構造を作れるようになったことが言語にとって本質的なことであって、それが言語の起源である。しかし、それを繰り返し適応することを生成される文の数が無限大だということの根拠にはしない方がよい。その理由は、生成文法の初期の頃から議論されたことであるが、言語能力と言語運用あるいは言語処理とも言われるが、その二つを峻別しようという主張がなされた。本稿では、その二つを峻別することに反対はしないが、それでも明らかに人間の広く言えば情報処理能力に対して、処理に対しての制約があって、それをはっきり認めて、処理に対しての制約が働く場合を想定するとどうなるかを検討する。

この問題は以前から気になっていた問題である。そして、Pullum & Scholtz (2012) が、文法が生成する文の数を無限大だと考える問題点を丁寧に検討し、すべてなりたたないと論じている。

生成文法学者は、規則が繰り返し適用されることを、生成文法の基本的な仮定として初期の頃から受け入れてきた。初期のモデルでは、規則の集合があって、規則が適用された結果出来上がるのが文法的な文の集合だと考えられた。色々なところに、繰り返し適用される可能性が広がっている。例えば、次のような日本語の名詞句の埋め込みを例に挙げることがもできる。

健の兄の友達の子 健の兄の友達の子の友達

すでに発話された文、書かれた文の数は、大きな数ではあっても、有限だと考えざるを得ないだろう。そして、ある時点を想定して、この論文を誰かが、今読んで下さっているあなたが読んでいる時点を想定して、その時点から先に発話される、書かれる可能性がある文の数は、理論的に、可能性として、加算無限だと考えることはできるだろう。しかし、それは可能性としての話で、あらゆる方向に、加算無限な集合の要素が増えていく可能性をはらんでいる。しかし、実際に発話されなければ、書かれなければ、それを集合の要素として認定するのは難しいだろう。そして、結果は、ある発話が行われ、集合の要素としての具体的な要素が一つ増える。そしてそれを足し合わせる操作が続けば、要素の数が一つずつ増えていく。埋め込みという操作が繰り返し適用される場合とは本質的に違うことがここでは起こる。しかし、あくまで可能性としての不可算無限の集合が形成されるということで、常にある時点を想定し、それまでに発話された文の数を数えると、その集合は有限集合であるということを忘れるわけにはいかない。

2. 1 「とても」や very の修飾の繰り返し

以下のような日本語の「とても」を繰り返し使う例を考えてみよう。

とても大きな建物、とても、とても大きな建物、とても、とても、とても大きな建物、とても、とても、とても、とても大きな建物

回帰的な構造を作る操作を複数回繰り返す場合と比較すると、次の二つの大きな違いがあるように思える。実際に口に出してみると、回帰的な操作を何度も繰り返した例は、処理に対する制約で簡単に解釈不能に陥る。それに対し、「とても」を繰り返す例では、

文法が生成する文の数が無限だという主張について

そういうことはない。英語の例を考えてみよう：

very hot, very, very hot, very, very, very hot

この例でも、最後の very の音調が、それまでのものを発音する場合と変わるように感じられる。ただ、これが最後だと言うことを示した方が相手に自分の思うことを伝えやすいと考えれば最後のものに違った音調を使えるので、同じ音調でも言えるかもしれない。この点も、本稿ではこれ以上論じない。

3. 言語の成立する基盤である回帰性と言語の生成する文の数が無限であるという主張との関係

言語の成立する基盤が回帰性にあると考えることに異論はない。妥当な考え方だと認める言語学者、あるいは関連する領域の研究者が多いだろう。問題は、回帰的な構造を作る規則は、繰り返し適用が可能で、理論的には、その操作を無限回繰り返すことができる。そして、その点を考えて、無限の数の文が生成できると、生成文法の初期の段階から考えられてきて、その主張を、世界中の多くの生成文法学者と関連する領域の研究者が、かなり前から共有しているが、そう考えることには疑問を差し挟む余地があると本稿では主張する。Pullum & Scholtz (2012) は、生成される文の数が無限だという主張の問題点を丁寧に論じて、そうではないと主張をしていることになる。

3. 1 無限と大きな数との関係

生成文法が生成する文の数が、回帰性を根拠に無限だと考えることの問題について論じてきた。続いて、ここでは、無限でなくても大きな数は、どこまで大きいかを考えると、現在のスーパーコンピューターでも処理することが難しいくらい大きな数が問題になることが分かってきた。

Pullum and Scholz が指摘していることだが、日本語の俳句になる可能性のある日本語のモーラ列の数が問題になる。日本語にはおおよそ 100 個のモーラがあって、5 - 7 - 5 に、それぞれ 100 個のモーラを入れて、可能な組み合わせ、それが日本語として意味を成すかどうかを判断せず、その組み合わせの数だけ考えると、10 の 36 乗位の数になるという。そして、この数は、現在のスーパーコンピューターで処理することを考えると、処理の負担がかなり大きなものになる数だということである。スーパーコンピューターの専門家に直接質問して、その答えである。ここで数えているのは、17 個のモーラ

列として可能なものの数であって、日本語として意味をなすかどうかを判断してはいない。例えば、次のような日本語として意味をなさないモーラ列も数えていることになる。

ままままーままままままーまままま

10の36乗というのは、とんでもない大きな数なので、俳句をよむ人が、他の人のよんだ俳句と同じ俳句をよんでしまうことをあまり心配する必要がないということになる。

大きな数がとんでもなく大きいという例として、先日のテレビで将棋のプロの試合を取り上げていて、可能な数の手の数を問題にしている、おおよそ10の200乗だということであった。

これは、もう想像を絶する大きな数で、その数の可能性をすべて検討することなど、人間には不可能であるし、プロの棋士とコンピュータのプログラムの対戦が行われたが、5人のプロの棋士がコンピュータと対戦し、コンピュータが3勝して、優勢だった。筆者は、過去のある時期、将棋や囲碁やチェスでの、人とコンピュータとの対戦について、メモリーはどんどん増えて、処理速度もどんどん速くなるので、それでコンピュータを強くする過程が進むのではないかと考えていたが、現在では、将棋で言えば過去のプロ棋士の棋譜を取り込み、高速でその棋譜のデータベースを検索し、良い手を探すことで勝負をしようとしている。そう考えると、言語に関して、可能な音連続をすべて検討するのは、人間がある言語を処理する実態を反映していないので、その問題を認識し、自然な分析を行うためにどうすべきか考える必要が出てくるだろう。

日本語のモーラの数がおおよそ100として、モーラ数が、例えば5のストリングの可能性は、100の5乗、6のストリングの可能性は、100の6乗、これを足し合わせていくと、可能なモーラ列の数は計算できることになるが、想定する例の数が、直観的な表現ではあるが、まさに天文学的数字で、しかもその天文学的数字のほとんどすべてと言っているほどの大部分が、日本語として意味をなさないものであると考えられる。

そう考えると、可能性の数を考えることの意味を再検討する必要が出てくるだろう。ここまで来ると、将棋の手の数を引き合いに出して論じるのは、あくまで便宜的なものであるが、検討するものから不必要なものを排除する方法を考えないといけないのだろう。

無限という数の大きさと、可能性としての有限の数でも大きいということの問題は、無限の数を回帰性を考えることで認めることは実際的でないと考え、さらに、可能性としての大きな数の大きさも、あくまで可能性としての数の大きさであって、実際に問題になるのは、そのごくごく一部であることを認めると、実際には、制限された数を問題にすることになると考えるのが妥当であろう。

Pullum and Scholz (2012) は生成文法学者が、無限の文を生成できるという考え方に魅せられてしまうということも論じている。歴史的には、コントロールによって始められた集合論が、物事に対しての新しい考え方を導入することになり、人々を魅了した。その考え方の転換が、生成文法での自然言語の文法をどういふものだと考えるかに色濃く残っている。文法的な文を生成する装置としての文法、そして生成された結果としての文法的な文の集合という捉え方に、無限の数の文を集合の要素として捉える可能性を最初から想定していたと考えることができるのではないのか。

3. 2 埋め込み構文を持たない言語

Everett (2005) などで示されたことだが、南米に住んでいる部族で話されている言語に回帰性を持たない言語があるという研究が出てきて、注目された。そういう言語に対して、一時期かなりの議論が行われた。現在では、一応落ち着いていると考えられる。どう落ち着いているのかと言えば、調べられた言語で、一見回帰性を欠いていると思われる言語にも、ある特徴があって、回帰性を持つ多くの言語とは違ったタイプの言語だという議論がある。これに関しては、違ったタイプの言語だという主張を認めるか認めないかで議論が分かれる。しかし、一見回帰性を欠いていると考えられる言語を話している部族が存在しても、赤ちゃんが、回帰性を持つブラジルポルトガル語が話される社会で育てば、回帰性を持つブラジルポルトガル語を習得できることが予想される。つまり、たまたま回帰性を欠いた言語が存在しても、人は回帰的な構造を作れるという能力を持っていることには変わりはないということである。

4. 結論

Pullum and Scholz (2012) の論じた問題を検討していく過程で、二人の結論とは、かなり違った結論に至った。この論文を書こうと考えた動機は、言語の本質は、回帰的な構造を作れるということにあるということ認め、しかし、それが何度も適用されることが無限の数の文を生成することを保証することはならないと考えることにあった。

他方で、Pullum and Scholz (2012) の指摘した、例えば俳句の 17 文字に、おおよそ 100 個の日本語のモーラの可能性を入れてみると、10 の 36 乗というとんでもなく大きな数のモーラ列を考えることが可能になると言っているのか、考える必要が出てくると言うべきなのか。そこで、無限ではないけれど、大きな数は、とんでもなく大きいということ結論にしようと考えていたが、それを結論にするには無理があると思うようになった。比喩的な言い方をすると、俳句として認められる可能性があるモーラ列の

うち、実際の俳句と認められる可能性があるものは、数からするとごくごく限られたもので、例えてみれば、星がほとんどない宇宙空間を進んで行き、すかすかの宇宙を進んで、やっと一つ星にぶつかるというような感じである。つまり可能なモーラ列の広がりをはほとんど星がない宇宙の空間のようであると捉えることが可能だと考えられる。

対象の数をうまく限定して、こういう限定を加えて、サンプルを取って、基本的には手仕事で、日本語の俳句として可能かどうかを調べて、可能性の全体の中で、日本語の俳句として認められそうな数が、全体の中でどれくらいの割合かを調べることが考えられる。その割合が極端に低いということが予想される。しかし、俳句の構成を考えてみると、日本語であるので、その構成単位をモーラで考えることには問題はないが、実際には、モーラがまとまって、もう一つ大きな構成要素（ここでは深く議論はしないが、語や句のような単位）を形成して、それらが相互に構造を作り、全体としての俳句が構成されることになるので、分析する対象、検討する単位は、本来はモーラではない。何が分析の対象なのかを考えて、具体的な分析を行う必要があることを忘れないようにしないといけない。

さらに、俳句では一般的には季語が必要であると考えられているが、それをどう考えるかという問題がある。実際の俳句を想定し、季語がなくてはならないと考え、モーラ列の可能性から、俳句として認められるものの数はさらに限定されることになる。ここまで考えて来ると、17のモーラの可能な組み合わせの数を数えることに果たしてどういう意味があるのかという疑問を感じざるを得ない。既に、指摘したが、5モーラの可能な組み合わせの数、100の5乗、6モーラの可能な結び付きの数、モーラ数を一つずつ増やして行って、17モーラくらいまで行って、それらをすべて足し合わせると、またさらに、有限ではあるが、とんでもない大きさの数になる。しかし、それらのうち日本語として普通に意味を持つモーラ列は、ほんとうにごくごく限られたものであるはずである。

ここまで考えてくると、日本語の処理の単位をモーラ一つ一つに切って考えるのか、もう少し大きな分節を考え、その分節を構成する単位としてモーラを考えるのかという問題が出てくる。実際には、文を構成する単位としてモーラを想定するのは、過去の日本語研究の成果から考えても不適切なところがあり、モーラではなくて、モーラが集まった単位が、文法上重要であることにはほとんど議論の余地がないので、今後はそこをつなく必要が出てくる。

本稿の暫定的な結論は以下のようにまとめられる。

多くの生成文法学者が想定する、生成文法では、無限の数の文を生成するという主張

は、少なくとも、回帰的な構造を繰り返して作ることで保証すると考えるのは処理の観点から不自然である。

日本語の可能な表現の数を数える場合に、モーラを単位に、可能なモーラ列を考えれば、その数を問題にするのも、不自然で、実際に自然な日本語のモーラ列は、その中のごくごく限られた一部なので、それを取り出して問題にすべきである。

これは言い換えると、検討の対象を適切に限定するには、どうすればいいのかという問題だということになる。検討してきた例で言えば、日本語の俳句のモーラ列の可能性からは、文頭に「ん」が来る組み合わせの多くは、日本語として不自然なものとして、検討すべきモーラ列から除外される可能性があるだろう。

残された検討すべき問題の中にも、重要なものが多い：

理論的な可能性と実際に発話されたもの、コーパスなどで確認が可能なものとの区別をどう考えるのかという問題がある。ここにも、生成文法が登場する以前の言語研究の方法と、生成文法で提案された方法論との間はかなり大きな方法論の違いが問題になる。

重要な違いを一つ挙げると、生成文法は、当該の言語を母語とする話者の言語直観を言語研究の重要な手段だと考えた。それまでの、過去に書かれた書記データを言語研究の主な資料として使ってきた言語研究との間には、大きな、本質的な違いがある。この点が、生成文法が新しい言語研究の手法として出てきて、注目を集めた重要な点である。

実際にうまくいくかどうか現段階でははっきりしないが、可能な場合をすべて問題にして、それを研究の対象にするのではなくて、人間が処理を実際に行っている、その処理の様態を反映したデータの取り込みを行って、それに対して、人間が行っている処理を想定した研究を押し進め、人間の言語処理の実態を反映した実験を行えないかどうかを検討するのは魅力的な課題だろう。

*

長年生成文法の研究を引っ張ってきた、スタンフォード大学の言語学科にいた Ivan A. Sag ががんで亡くなった。長い間の研究仲間であるエディンバラ大学にいる Geoffrey K. Pullum が自らのページに、Ivan A. Sag 氏に対する追悼文を公表している。自分がよって立つ言語研究の枠組みにこだわらず、世界中の言語学者、特に若手の研究者とも親しく関わってきたので、亡くなられたのを残念だと思っている人が世界中に少なくとも何千人という言語学習者がいると思うと書いている。その通りだ。Ivan A. Sag 氏ともう会えない、氏が言語の関しての研究をもう行えないことを考えると、残念でならない。謹んで、Ivan A. Sag 氏のご冥福を祈る。

本稿では、言語の文法の生成する文の数が無限であるかどうかを検討し、無限であると考えない方がよいということを論じた。

最後に、Pullum and Scholz (2010) では、日本語のモーラの数とその組み合わせの数は論じてはいるが、具体的にどういうモーラがあって、その組み合わせを問題にはしていない。日本語の文法も辞書も持っていないので、そういうことを具体的に論じてはいない。主に日本人の研究者による日本語に関する関連する先行研究を踏まえた上で、研究を進めることが今後の課題である。

参考文献

- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1980. *Rules and Representations*. Oxford: Basil Blackwell.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. 1986. *Knowledge of Language: Its Origins, Nature, and Use*. New York: Prager.
- Epstein Sam and Norbert Hornstein. 2005. Lecture on 'The future of language', *Language* 81: 3-6.
- Everett, Daniel L., 2005. Cultural constraints on grammar and cognition in Piraha: Another look at the design features of human language. *Current Anthropology* 46: 621-646.
- Hauser, Marc D., Noam, Chomsky, and Warren Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569-1579.
- Lasnik, Howard. 2000. *Syntactic Structures Revisited: Contemporary Lectures on Classical Transformational Theory*. Cambridge, MA: MIT Press/.
- Pullum, Geoffrey, K. and Barbara C. Scholz. 2012. Recursion and the infinitude claim, in H. van der Hulst ed. *Recursion in Human Language*, Berlin, De Gruiter Mouton.
- 外池俊幸 (準備中) 「生成される文の数と回帰性について」
- Yang, Charles. 2006. *The Infinite Gift*. New York: Scribner.