ゲーデルの不完全性定理とタルスキの定理

飯田 隆

「この文は証明できない」

すでにこのシリーズの第 11 巻『論理・数学・言語』でも何回か出てきたが、古代から知られている有名なパラドクスに、嘘つきのパラドクスというものがある。そのいちばん簡単な形は、「嘘つき文」と呼ばれる文、すなわち、

この文は偽である。

によって表現される。普通に考えれば、この文は真であって、かつ、偽である文となってしまう。だが、これは、どんな文も真であって、かつ、偽であることはないという矛盾律に反する。矛盾律は論理的原理のなかでも、もっとも基本的なものであり、また、嘘つき文に含まれるのは真偽という、これもまた、われわれの思考にとって基礎的な概念である。それゆえ、嘘つき文が真剣な議論の対象となってきたことに不思議はない。

さて、嘘つき文とよく似た文として、つぎの文を考えよう。(この文は、論理パズルの大家としても有名な論理学者レイモン・スマリヤンに由来する。)

この文は証明できない。

この文も嘘つき文と同じパラドクスを生み出すようにみえる。

> この文が偽だと仮定してみよう。つまり、この文は証明できないということは偽であると仮定する。ということは、この文は証明できるということである。ところで、証明できるものはすべて真であるから、この文は真である。だが、それは、この文が偽だとした仮定に反する。よって、もとの仮定が否定されなければならない。つまり、この文は真である。
>
> だが、翻って考えれば、いまわれわれがしたことは、この文が真であるということを示したこと、つまり、この文を証明したことではないだろうか。実際にこの文が証明された以上、それが証明できないということはない。つまり、この文は偽である。

しかし、嘘つき文の場合と違って、これは本当のパラドクスとはみなされない。その理由は、ここに現れる「証明できる」という概念の不明確さが問題の元凶であると考えられるからである。したがって、まずは「証明できる」とはどういうことなのかから考え直してみる必要がある。

「証明できる」ということ (1)—公理化

よく言われるように、すべてを証明することはできない。すでに正しいとされた前提から新たな結論を引き出すのが証明である。したがって、前提の証明に必要な前提をつぎつぎと遡って行けば、どこかで前提なしに正しいとされる命題に突き当たるのでなくてはならない。このことは少なくとも古代ギリシア以来知られていた事柄である。すべての証明の出発点には、証明なしで受け入れなければならない命題がある。こうした命題のことを「公理」と呼ぶ。

証明は何のために必要なのか。この問いには二通りの答えがある。まず、結論の正しさを納得するために証明はなされるという答えがある。その前提の正しさがすでに納得済みであるときにのみ、証明はこうした目的を果たすことができる。この考え方に基づけば、すべての証明の出発点となる公理は、それ自体でその正しさが納得できるようなものでなくてはならないということになろう。他方、証明の目的とは、何らかの主題に関してわれわれがもっている知識を体系化するためであるという答えもある。決まった主題に関する知識を構成する命題は相互に論理的に結び付いている。たとえば、古代のギリシア人は図形についてのさまざまな知識をもっていたが、そうした知識がたがいにどのような関係にあるのかにも気付いていた。A が成り立つことから他の B や C が成り立つことが出てくるといった事柄である。証明とは、こうした論理的関係を利用した知識の体系化であるとみなすことができる。その際の理想は、ごく少数の命題から出発して、その主題に属する知識のちょうど全体を導き出すことである。この少数の命題が公理であり、他の命題はそれから証明によって定理として導き出される。知識のこうした体系化を「公理化」、その結果生み出されるものを「公理的理論」と呼ぶ。

どちらの見方を取ろうとも、公理が満たさなければならない最低限の条件は、そこから矛盾が出てこないという意味で、矛盾を含まないことである。一方で、矛盾を含む理論は、A とその否定 $\neg A$ の両方を主張するのであるから、間違った主張を含んでいることは明らかである。他方で、通常の論理に従う限り、矛盾からはどんな命題も帰結するので、矛盾を含む理論は、真偽にかかわりなく何でも認めてしまうことになるため、知識の体系化の役に立たない。

二つの見方は、必ずしも相反するわけではない。理想的なのは、証明の二通りの目的を同時に実現することであろう。つまり、それ自体でその正しさが確信できるような公理から出発して、その分野に属するすべての真理を引き出すことができることである。しかし、二十世紀以降の数学においては、体系化の手段としての証明という見方の方が優勢であると言ってよいように思われる。そこで公理に求められることは、その「自明さ」であるよりは、その主題に属する知識の全体をそこから引き出せることだからである。

この要件を公理が実際に満たし、意図されていた知識のちょうど全体を導き出すことに成功した場合、その公理による公理化は「完全である」と言われる。どのような主題を取っても、それに関する知識全体の完全な公理化があるという保証はない。また、異なる範囲の知識ごとに、そこで公理としてとるべき命題が何であるかが一通りに決まっているわけでもない。複数のたがいに異なる完全な公理化がある場合もあれば、完全な公理化がない場合もあるかもしれない。しかしながら、完全な公理化がある、また、それが複数ある場合には、その間に何らかの仕方で優劣をつけて最良のものを選び出すことができるという信念が、知識の体系化への努力を導いてきたと推測することは許されよう。

「証明できる」ということ (2)—形式化

証明の出発点にとられる公理について、これまでもっぱら述べてきたが、公理だけでは定理を証明することはできない。もうひとつ必要なのは、推論のための規則である。

定理はすべて最終的には公理から論理的に導かれるものでなくてはならない。証明とは、それが実際にどのようにしてなされるかを示すものである。この際の手順が正確に何であるかがはっきりと述べられたのは、意外にも遅く十九世紀の後半になってからのことである。この遅れの原因は、何が正しい論理的推論であるかは、改めて述べるまでもない自明なことと考えられていたからである。

何が正しい論理的推論であるかを、史上初めて明示的に述べたのは、フレーゲである。かれには、算術的真理は論理的真理にほかならないという主張を立証するために、論理的真理の範囲を正確に特徴づける必要があった。この特徴づけにおいて最大限の厳密さを確保するために、フレーゲは二つの段階を踏んだ。第一に、ドイツ語や日本語のような自然言語で証明が表現されている限り、その証明の各段階で何が生じているかが明示されないおそれがあるため、かれは、証明を表現するための特別の言語を考案し、その言語の基本的な語彙の表と、そこに属する語をどのように組み合わせれば有意味な表現が得られるかを示す明確な規則を与えた。第二に、このような言語で表現された命題を用いて正しい推論を構成するために従うべき規則を与えた。しかも、これはフレーゲより後にわかったことであるが[1]、ごく少数の基本的な規則とその組み合わせで、数学に必要な推論はすべてまかなえるのである。

フレーゲが先鞭をつけたのは、不完全性定理が発表される直前に絶頂に達した「数学の形式化」と呼ばれる流れである。形式言語と呼ばれる、明確な文法を備えた人工言語で表現され、そこにおける証明も、あらかじめ明示さ

[1] これはゲーデルが、その不完全性定理に先立って証明した完全性定理によって示された。完全性定理の「完全性」は、不完全性定理の「完全性」とは違う概念なので、たいへんまぎらわしいが、ここではこの点について深入りしない。

れた公理と推論規則に従ってなされるように定式化された数学的理論は、形式化された理論と呼ばれ、そこにおける証明は、形式化された証明と呼ばれる。形式化された証明は、それ自体、数学的探究の対象となりうる。この事実は後に重要になるが、とりあえずいまは、二つのことを指摘しておきたい。

証明が形式化されることによってはっきりしたことは、ある命題が証明できるか、それとも、できないかは、どのような理論のなかでの証明が問題になっているかによって異なりうるということである。よって、証明できるかどうかという問いは、命題 P 単独について問うべきではなく、P と何らかの理論 θ について「P は θ で証明できるか」という形で問われなくてはならない。

もうひとつ注意すべきなのは、証明の形式化によって、証明の概念は新たに定義し直されるという点である。この定義は、次のようなものになる。

> 理論 θ における証明とは、θ の言語の文から成る列であり、その列のひとつひとつは、(1) θ の公理であるか、あるいは、(2) その列において先立つ文から、 θ の推論規則の適用によって得られる文である。

この新たに定義し直された証明の概念によれば、理論 θ の公理は必ず θ で証明されることになる。これは、公理は証明できないという、公理についての一般的理解とは反するものであるが、公理は公理自身によって証明されると考えることは、数学ではよく出会う一般化の戦略の表れのひとつである。この点に関連して、どんな理論によっても証明不可能な命題というものはないことを指摘しておこう。なぜならば、どんな命題 P も、それを公理として採用する理論において証明されるからである。

ただし、このことは、どんな命題でも、それが形式化された言語で表現できるほど明確なものであれば、それを公理として採用する理論を考えることができるというだけであって、そうした理論が意味ある理論であるということではない。しかし、これは同時に先の論点をさらに補強する材料でもある。すなわち、どんな命題に対してもそれがそこで証明可能である理論を見つけることは容易にできるのだから、ある命題が証明できるかという問いは、それだけでは実質のある問いにはならないということである。

「この文は証明できない」ふたたび

いま見たように、ある命題が証明できるかどうかという問いは、何らかの理論 θ と相対的にしか意味をもたない。よって、「この文は証明できない」ではなく、ある理論 θ に相対化された証明概念を用いた文

> この文は θ で証明できない。

を考えてみよう。今度は、先ほどのような「パラドクス」は生じない。

> 前と同じように、まずこの文が偽であると仮定してみよう。つまり、この文は θ で証明できると仮定しよう。θ が「正しい」理論、つまり、そこで証明されるものはすべて真であるとすれば、この文は真でなくてはならない。つまり、この文は θ で証明できない。だが、それは、この文が θ で証明できるとした最初の仮定に反する。よって、この仮定が否定され、この文は真である、つまり、この文は θ で証明できないと結論される。
>
> さて、以上のことは、この文が真であることを示したことである。ルーズな言い方をすれば、この文を「証明した」と言ってもよい。だが、それは、θ での形式的証明を与えたことではない。つまり、「この文は θ で証明できない」を理論 θ の中で証明したのではない。したがって、それは、この文が述べていることそのこと、すなわち、それが θ で証明できないことがまちがいだと示すものではない。

結局、わかったことは、「この文は θ で証明できない」という文は、θ が正しい理論であれば、真であるが、θ では証明できないということである。これはこれで興味深い結果かもしれないが、それほどたいしたことども思えない。もちろん、ここから、すべての真理を、そして真理だけを証明するような、真に包括的な形式的理論などというものはありえないという結論を引き出すことはできよう。しかし、ある文がある体系で証明できるかどうかは、体系についての事柄であって、その体系自身の中に含まれるべき事柄のように思えないというのが自然だろう。

だが、こうした思い込みは危険である。ひとつには、形式化された理論には、もともとそれが意図されていた解釈以外のさまざまな解釈がありうるという事実がある。したがって、形式化された理論に属する文が、何を主張し、何を主張しないかは、一概に言えることではない。一見したところ、また、本来も、証明可能性について云々する文は含まないと思われた言語において定式化された理論 θ があったとしよう。ひょっとすると、θ の言語に対して新しい解釈を与えて、証明可能性について語る文があるようにすることもできるかもしれない。そうした文の中に「この文は θ で証明できない」と解釈できる文があったらどうだろうか。そのときは、先のような推論により、θ の言語で表現できる真理でありながら、θ では証明できないものがあるということになろう。体系化の理想が、体系化されるべき真理のすべてを網羅することにある以上、こうした θ は「不完全」であると言ってよいだろう。

さらに、こうした再解釈が可能であり、しかも、そうした再解釈によって、自身の証明不可能性を主張することになる文を含む理論といったものが、案外ありふれた理論であり、われわれが必要とする理論の大部分がそうした理論だとしたらどうだろう。そのときには、われわれが出会う理論の大部分が「不完全な」理論であり、その理論が扱うべき真理の全体を包括していないと

いうことになるだろう。

そして、こう聞いただけでは信じられないようにみえるが、実際のところ、まさにこれこそが、ゲーデルの不完全性定理、より正確には、第一不完全性定理によって示されることにほかならないのである。

算術化と対角化

第一不完全性定理を証明するためにゲーデルは、二つのことをした。一つは、形式化された理論の中の文について、それが証明できるとか証明できないといった事柄について述べる文を、自然数のあいだの関係を表す文—算術文—に翻訳することである。これを「構文論の算術化」、もっと簡単に「算術化」と呼ぶ。これによって、自然数について述べるものとして作られた言語を、証明できるとか証明できないといった事柄について述べるものとして再解釈することが可能になる。もう一つは、「この文は θ で証明できない」という文に端的に現れているような、自分自身について何かを述べる文—「自己言及的文」と呼ばれる—を作り出す仕組みを用意したことである。これを「対角化」と呼ぶ。

θ を算術（自然数論）をその一部とする形式化された理論であるとしよう。θ は、ある形式言語 L を用いて表現されている。L の語彙を構成する語の一つ一つに、ある自然数を指定してやる。次に、こうした語から作られる表現の各々に、それを構成している語に指定された数から決定される一つの数をあてがう。こうした数を、その表現の「ゲーデル数」と呼ぶ。文もまた表現の一種であるから、L の文の各々にはそのゲーデル数がある。θ における証明とは、L の文がある順序でつらなっているものにほかならないから、それにもそのゲーデル数をあてがうことができる。そうすると、文 S の θ での証明（のひとつ）にあてがわれたゲーデル数 n と、S のゲーデル数 m のあいだには、

> n は、ゲーデル数 m をもつ文 S の、θ での証明のゲーデル数である。

という関係が成り立つ。これを

$$Pr_{\theta}(n, m)$$

と書こう。このように、形式的証明という概念は、語から文、文から文の列という仕方で定義されるものとして、言語の構文論（シンタクス）の一部とされるのである。

さらに、文 S が θ で証明できるということは、S のゲーデル数とのあいだに Pr_{θ} という関係が成り立つ自然数、つまり、S の証明のゲーデル数となっ

ている数が存在するということ、あるいは、論理記号を使って書けば、

$$\exists x Pr_{\theta}(x,m)$$

ということである。これを、

$$Prov_{\theta}(m)$$

と書こう。ゲーデルは、$Prov_{\theta}(m)$ が、θ の言語 L の中で表現できるだけでなく、真であるならば θ で証明できることを示した。他方、S が θ で証明**できない**ということは、

$$\neg Prov_{\theta}(m)$$

であるから、これもまた θ の言語の中で表現できる。

次に必要なことは、自分自身が θ で証明できないと述べる文を、θ の言語の中で構成することである。おどろくべきことに、θ のような言語には、自身について何かを述べる自己言及的文を作る資源が備わっている。「対角化定理」と呼ばれるものがそれである[2]。

> $\phi(x)$ を、x だけを自由変項としてもつ L の式とする。また、$\overline{E}$ を、表現 E のゲーデル数とする。このとき、ある式 D が存在して、
>
> $$D \leftrightarrow \phi(\overline{D})$$
>
> が θ で証明できる。

この定理に現われる式 D が何を言うものであるかを考えよう。それは、$\phi(\overline{D})$ と同値であることが θ の中で証明できるのであるから、θ を信じる限り、D が「ϕ」によって表される性質をもつことを言うものであると考えてよい。つまり、理論 θ の正しさを仮定するならば、D は、自分自身が性質 ϕ をもつことを言う文である[3]。ϕ は、L で表現できる任意の性質であるから、この定理は、L で表現できるどんな性質についても、自分という文はその性質をもつ文だと述べる自己言及的な文が L の中に存在することを示すものである。

第一不完全性定理

ここまで来れば、ゲーデルの第一不完全性定理を示すことはむずかしくない。θ で証明できないという性質、つまり、$\neg Prov_{\theta}(x)$ は、L で表現できる

[2] 「対角化補題 diagonal lemma」と呼ばれることの方が多いかもしれない。また、「不動点定理 fixed point theorem」とも呼ばれる。この定理をゲーデルは、カルナップ（『言語の論理的構文論』）に帰している。

[3] ただし、ここには注意すべき問題がいろいろある。第一に、たとえ θ が正しい理論であったとしても、二つの文 A と B の同値性が θ の中で証明できることだけでは、A と B が「同じことを言っている」とするためには十分ではない。第二に、D が何を言うかは、D が具体的にどのような仕方で作られたかに依存する。したがって、本文のすぐ後で出てくる G（これは「ゲーデル文」とも呼ばれる）が「自分は θ で証明できない」ということを「言う」とみなせるためには、理論 θ と文 G がいくつかの条件を満たしていなくてはならない。この点については、次を参照。P.Milne, “On Gödel sentences and what they say” *Philosophica Mathematicae* (III) 15 (2007) 193–226.

性質であるから、対角化定理により、θ で

$$G \leftrightarrow \neg Prov_\theta(\overline{G})$$

が証明されるような L の文 G が存在する。G は「この文は θ で証明できない」の L への翻訳にあたると言ってよい。

さて、

(*)　θ が無矛盾であるならば、G は θ で証明できない。

を証明してみよう。

> θ が無矛盾であると仮定し、さらに、G が θ で証明できると仮定する。そうすると、G の θ での証明があるはずであるから、その証明のゲーデル数 n がある。n が G の θ での証明のゲーデル数であるということが真であるから、
>
> $$Pr_\theta(n, \overline{G})$$
>
> が、θ で証明できる。それゆえ、$\exists x P(x, \overline{G})$、すなわち、$Prov_\theta(\overline{G})$ も θ で証明できる。だが、θ において
>
> $$G \leftrightarrow \neg Prov_\theta(\overline{G})$$
>
> が証明できるのだから、
>
> $$Prov_\theta(\overline{G}) \leftrightarrow \neg G$$
>
> も θ で証明できる。よって、$Prov_\theta(\overline{G})$ が証明できたのだから、$\neg G$ が θ で証明される。結局、G と $\neg G$ の両方が θ で証明されたことになる。これは θ が無矛盾であるという仮定に反する。よって、G は θ で証明できるとした仮定を否定して、G は θ で証明できないと結論しなければならない。よって、(*) が証明された。

ところで、θ が正しい理論である、すなわち、θ の公理がすべて真であるならば、G は真でなければならない。なぜならば、G が θ で証明できないと述べる文と G 自身とが同値であることが、θ で証明でき、θ で証明できることはすべて真なのであるから、実際、G は、G が θ で証明できないということと同値である。そして、まさに、後者が正しいことがいま示されたのであるから、G は真である。

G が証明できないのであれば、θ に G を新たに公理として付け加えたらよいだろうと考えても無駄である。なぜならば、この新しい理論 $\theta + G$ に関して、いまとまったく同様の議論を繰り返すことができるからである。つまり、この新しい理論での証明可能性を表す述語

$$Prov_{\theta+G}(m)$$

に対角化定理を適用することによって、理論 $\theta + G$ で

$$G' \leftrightarrow \neg Prov_{\theta+G}(\overline{G'})$$

が証明されるような文 G' が存在することが示され、それは、先とまったく同じ議論により、真でありながら $\theta + G$ では証明できないことがわかる。

この節を終わるにあたって、第一不完全性定理を、現在しばしば見られるような仕方で述べておこう。後にも触れるように、第一不完全性定理のこの述べ方は、ゲーデルの原論文にあるものとは違うが、次の節で述べるタルスキの定理との関連をみるには、より適切である。

> **第一不完全性定理**　十分な強さをもつ形式的体系のいずれにおいても、そこで証明できない真である算術の文が存在する。

ここで「十分な強さをもつ形式的体系」であるために必要な条件は、(i) その言語で算術が表現でき、(ii) ペアノ算術と呼ばれる自然数論の公理系を含み[4]、そして、(iii) 矛盾を含まないことである。

ペアノ算術のような理論に矛盾が含まれている可能性があると考えるひとは、まずいない。したがって、第一不完全性定理の要件を満たす形式的体系は、たしかに存在していると考えてよい。そして、そうした体系にどれだけ多くの公理を付け加えても、それによって新たに矛盾が生み出されない限り、証明できない算術的真理は残るのである。このことは、自然数の領域において何が成り立つかという、ごく基本的な事柄についてさえ、われわれの知識は常に不完全なままにとどまるということである。もっとも、別の見方—そして、それはゲーデル自身が好んだ見方である—からすれば、それは数学の無尽蔵性を意味するということでもある。

タルスキの定理

対角化定理は、「タルスキの定理」と呼ばれる重要な定理を示すのにも使える。

すでに見たように、ゲーデルは、形式化された理論のもつさまざまな構文論的性質についての主張を、自然数に関する主張に翻訳する体系的方法を与えた。理論の表現に用いられる言語の文法に属する事柄だけでなく、証明に関する事柄もまた、言語の構文論に属するのであった。言語や理論がもつ性質は構文論的なものにとどまらない。そこに現われる表現が何かを指示したり、それに属する文が真であったり偽であったりということがなければ、言語も理論もわれわれの役には立たないからである。言語や理論のもつこうした性質のことを、その意味論的性質と呼ぶ。ここで当然出てくる疑問は、意

[4] 実際には、これよりもずっと弱い算術の公理系を含んでいればよいことが知られているが、立ち入らない。

味論的性質についても、構文論的性質の場合と同様の「算術化」は可能だろうかという問いである。タルスキの定理は、この問いに対して否定的な答を与える。

意味論的概念のうちで中心的位置を占めるのは、真理の概念である。これまでと同様、θ を算術（自然数論）をその一部とする形式化された理論、L を θ の言語としよう。L の文はすべて、真もしくは偽であるとする。ゲーデルは、θ で証明可能な L の文を特徴づける述語 $Prov_\theta$ を L 自身の中で構成できることを示した。では、これと同じような仕方で、真である L の文を特徴づける述語—「真理述語」と呼ぼう—を L の中で構成することは可能だろうか。真理述語 Tr については、次が成り立たなければならない。

(T)　L の任意の文 S について、$S \leftrightarrow Tr(\overline{S})$

すなわち、ある文を主張することと、その文が真であると主張することとは、同じでなければならないということである。したがって、L の真理述語が L に属するために必要なことは、(T) が θ で証明可能であることである。

しかし、Tr が L の述語であるならば、$\neg Tr(x)$ は、x だけを自由変項としてもつ L の式となるから、対角化定理をこれに適用すれば、

$$D \leftrightarrow \neg Tr(\overline{D})$$

が θ で証明できるような L の文 D が存在する。この右辺は、D は真ではないということを言い、それはまさに D と同値だと言うのだから、D は、自分は真ではないと述べる嘘つき文の一種である。ところで、先に真理述語について θ で証明されるのでなければならないとされた (T) から、

$$D \leftrightarrow Tr(\overline{D})$$

が θ で帰結する。ここから矛盾が出て来ることは明らかである[5]。

つまり、以上より、次の定理が成り立つことがわかる。

> **タルスキの定理**　θ を自然数論を含む形式化された理論、L をその言語とする。θ が無矛盾であるならば、L の真理述語—すなわち、それに関して (T) が θ で証明可能であるような述語—が L に属することはできない。

θ の不完全性はタルスキの定理から直ちに帰結することに注意しよう。まず、θ が正しい理論であると仮定しよう。つまり、L のすべての文 S について、S が証明可能であれば、S は真であると仮定する。これに加えて、L の真理のすべてが θ で証明可能であるという意味で θ が完全であると仮定する。

[5] 対角化定理から出てくる式は $A \leftrightarrow \neg B$ という形の式であるのに対して、(T) から出てくる式は $A \leftrightarrow B$ という形の式である。この二つから $B \leftrightarrow \neg B$ が帰結し、これは、明らかな矛盾 $B \wedge \neg B$ と同値である。

すなわち、L のすべての文 S について、S が真であれば、S は θ で証明可能であると仮定する。この二つの仮定から、L のすべての文 S について、その真理性は、その θ での証明可能性と一致することが帰結する。これはつまり、L の述語である $Prov_\theta$ が、L の真理述語の役を果たすということを意味する。これは、L の真理述語が L に属することはありえないというタルスキの定理に反する。それゆえ、θ は完全ではありえない。

ゲーデルとタルスキ

ところで、タルスキの定理は、かれの有名な論文「形式化された言語における真理概念」（ポーランド語版一九三三年、ドイツ語版一九三五年）に現れたものであるが、すでに一九三〇年にゲーデル自身がこの結果を得ていたことがわかっている。他方、タルスキの論文の主要部分は一九二九年のものであるにしても、真理述語に関する結果は、算術化の方法をゲーデルから聞いた後でタルスキがゲーデルとは独立に得たものである[6]。

真理述語に関する自身の結果をゲーデルが公表しなかった理由については、フェファーマンが述べていることが、きわめて説得力がある。かれによれば、それは、真理の概念について当時広く存在していた疑念のせいである[7]。証明に関する事柄は、形式化された証明が問題になっている限り、まったく危ないところがないと考えられていたのに対し、真偽の概念は嘘つきのパラドクスに象徴されるような困難をはらむと考えられていた。慎重なゲーデルは、そうした概念への明示的言及を避けたのである。

したがって、ゲーデルの一九三一年の論文に現れる完全性・不完全性の概念は、われわれがこれまで扱ってきたような真理の概念を含むものではない。ゲーデルの原論文の用語法—それは、いま一般にみられる用語法でもあるが—によれば、形式化された理論 θ が完全であると言われるのは、θ の言語に属する任意の文 S について、S もしくはその否定 $\neg S$ が証明できるときである。真である文がすべて証明できるという意味での完全性が成り立てば、L の任意の文 S について、S が真であれば S は証明でき、S が偽であれば、その否定が真であるから $\neg S$ が証明できる。どんな文も真もしくは偽であるから、どの文 S についても、それ自身、あるいは、その否定が証明できるということが出て来る。このように、証明概念だけを用いて定義される完全性は、真理概念を用いて定義される完全性から帰結する。ただし、その逆は成り立たない。なぜならば、L の任意の文 S について、それ自身もしくはその否定が θ で証明できても、θ が偽である文をも証明するような理論であることが可能

[6] A.B.Feferman and S.Feferman, *Alfred Tarski: Life and Logic*, 2004, Cambridge University Press, pp.83–85. 次をも参照。M.Gómez-Torrente, “The indefinability of truth in the ‘Wahrheitsbegriff’ ” *Annals of Pure and Applied Logic* 126 (2004) 27–37.

[7] 次を参照。S.Feferman, “Kurt Gödel: conviction and caution”, *Philosophia Naturalis*, vol. 21, pp. 546-562, 1984; Reprinted in S.Feferman, *In the Light of Logic*. Oxford University Press, 1998.

だからである。よって、ゲーデルがもともと定義した形の（第一）不完全性定理は、ここで述べたものよりも論理的に強い定理である。

しかしながら、真理の概念を持ち出さない限り、理論体系の完全性・不完全性がなぜ重要な問題になるかはわからない。先に挙げた論文においてタルスキは、少なくとも形式化された理論に関して真理の概念は明確なものでありうることを示した。ただし、明確化には、その代償がある。まず、タルスキによれば、日常の言語において用いられている真理の概念は、結局のところ、嘘つきのパラドクスのような矛盾を生み出さざるをえない。矛盾の出現の心配がないのは、真理の概念を、一定の条件を満足することが知られている形式化された言語に適用するときだけであり、しかも、そうした言語における真理の概念を定義するためには、それとは別の、より強力な言語—メタ言語—を用いる必要がある。

現在では、タルスキの方法とは違う真理概念の扱い方があることが知られている[8]。しかしながら、現代真理論の古典としてのタルスキの理論の位置はゆるがない。それゆえ、タルスキの仕事は二十世紀の論理学および哲学の歴史のなかで、ゲーデルの仕事とならぶ重要性をもつものである[9]。

第二不完全性定理

最後に、第二不完全性定理についても、ごく簡単に見ておこう。

θ を、これまで考察してきたのと同様の条件を満たす、形式化された理論であるとしよう。θ が無矛盾であるということは、$S \wedge \neg S$ という形の文が θ では決して証明されないということである。S としてどんな文を取っても構わないので、どれか特定の文 S_0 を S として取ったとして、θ の無矛盾性もまた

$$\neg Prov_\theta(\overline{S_0 \wedge \neg S_0})$$

という θ の文によって θ の中で表現できる。この文を Cons_θ と書こう。ゲーデルの第二不完全性定理が言うところは、θ が無矛盾であるならば、その無矛盾性を表現する文 Cons_θ は θ で証明できないということである。

このことは、第一不完全性定理の証明をすべて θ の中での形式的証明に書き直すことによって得られる[10]。第一不完全性定理は、θ が無矛盾であるな

[8] 津留竜馬「嘘つきのパラドクス」（飯田隆編『論理の哲学』二〇〇五、講談社選書メチエ、所収）を参照されたい。

[9] タルスキ、および、かれの仕事の背景となっているポーランドの論理学的=哲学的伝統については、本シリーズ第 11 巻のコラム「両大戦間のポーランドにおける論理学と哲学」をも参照されたい。

[10] これは、言うのは簡単だが、実際には結構大変な仕事である。ゲーデルの一九三一年の論文のタイトルには「I」という数字があるが、そこでは第二不完全性定理の証明方針が述べられているだけであり、その実際の証明は「II」で述べられるはずであった。しかし、「II」は結局書かれずじまいになった。第二不完全性定理の証明がほぼ完全な仕方で述べられたのは、ヒルベルトとベルナイスの『数学の基礎』の第二巻（一九三九年）が初めてである。

らば、G は θ で証明できないことを言うものであるから、それの形式化された証明の結論は、

$$\mathrm{Cons}_\theta \to \neg Prov_\theta(\overline{G})$$

となる。そして、第一不完全性定理の証明そのものの形式化によって、これは θ の中で証明できるのであった。いま仮にこの前件 Cons_θ もまた θ の中で証明できるとすれば、後件 $\neg Prov_\theta(\overline{G})$ が θ の中で証明できることになる。しかし、これは G と同値であることが θ の中で証明できる文であった。よって、以上を合わせると、そもそも θ では証明できないはずであった G が θ で証明できることになる。これは矛盾である。よって、θ が無矛盾であるならば、θ の無矛盾性 Cons_θ を θ の中で証明することは不可能である。つまり、次の定理が成り立つ。

> **第二不完全性定理**　十分な強さをもつ形式的体系のいずれにおいても、その無矛盾性をその体系において証明することはできない。

形式的体系がここで言う「十分な強さ」をもつための条件は、先に第一不完全性定理に関して述べられたものと同じである。

ゲーデルの結果が公表された当時、より大きなインパクトを与えたのは、第一不完全性定理であるよりは、第二不完全性定理の方である[11]。その理由は、第二不完全性定理が、数学の無矛盾性を証明するというヒルベルトのプログラムが実現不可能であることを示したと考えられたためである。

ヒルベルトのプログラムについては、本シリーズ第 11 巻の第 IV 章を見ていただきたいが、それは、簡単に言って、数学理論を形式化し、ついで、「有限の立場」と呼ばれる方法論的立場のもとで、その無矛盾性を証明しようというものである。無矛盾性証明に対するごく素朴な考え方は、理論 θ の無矛盾性を証明するのに、θ で許されるのよりも制限された手段を使うべきだというものである。無矛盾性証明の第一歩は、ペアノ算術のような形式化された自然数論の無矛盾性を証明することであるが、その証明に自然数論の範囲を超えるような手段を使ったのでは、その信頼性は高くなく、自然数論の範囲に入っていても、その中でもごく基本的な部分だけを用いて証明を行うべきだといった考え方である。しかし、第二不完全性定理によれば、ペアノ算術の無矛盾性は、ペアノ算術の部分体系はおろか、ペアノ算術自身においても証明できない。もちろん、これはペアノ算術が無矛盾だという想定のもとであるが、このことを疑う数学者はまずいない。こうして、第二不完全性定理の出現は、ヒルベルトのプログラムの根本的な再考を促したのである[12]。

---

[11] そのごく間接的な証拠として、戦前のわが国での不完全性定理をめぐる議論をみるとき、第二不完全性定理はひんぱんに言及されるのに対して、第一不完全性定理への言及はまれであることを挙げることができる。拙稿「付論 ゲーデルと第二次大戦前後の日本の哲学」（田中一之編『ゲーデルと 20 世紀の論理学 1 ゲーデルの 20 世紀』二〇〇六、東京大学出版会、所収）参照。

[12] 本稿に目を通して有益な助言を寄せてくれた秋吉亮太君（慶応義塾大学博士課程）に謝意を表する。