

むかし王国は、言語の発展により、言葉と知識があふれていた。が、その豊かさは同時に混沌ももたらした。各地の学者や論者は、自らの解釈と信念に基づき、あらゆる“命題”を口にした。その中には、互いの矛盾も多く、国民はどの言葉を信じるべきか迷っていた。

事態を憂慮した賢者は、すべての命題が記された「真理の書」を記した。さらに内部計算装置タルスを開発し、以下の機能を組み込んだ
「もし命題Pが正しいならば、タルスはPに『真』のラベルを付ける」

きだ』

各々は自己言及文ではないが、二つ合わさると矛盾が生じる。これも排除するが、自己言及的状況の命題は、他にも容易に見つかった。そこで、自己言及しないように、言葉や文法に制限を加えたが、真理の書は薄くなった。こうして、タルスは限界に達し、賢者は次の洞察を得た。

「どれほど高度なシステムでも、内部だけで真理を定義し、評価する試みには必ず矛盾が含まれ、失敗する」
これは「タルスキーの定義不可能性定理」*と呼ばれ

数 | 理 | の | 窓

すべての真理を記述する
「真理の書」は完成するか



当初、タルスは見事に機能し、ほとんどの命題が評価された。しかし、ある時、真理の書の奥深くから、謎めいた命題が見つかった。

『この文は、タルスによって『真』とラベルづけされるべきでない』

タルスは計算を始めた。この命題に「真」のラベルを付けるならば、命題の内容に反し「真」ではない。逆に、タルスが命題に「真」を付けなければ、命題の主張より「真」となり、矛盾が生じる。

そこで賢者は、真理の書から自己言及文だけを排除しようとした。が、次の議事を見つけた。

アリス：『私はボブを信じない』

ボブ：『アリスが私を信じなければ、アリスは嘘つ

る。十分な表現力を持つ言語において、その言語内では「真理」を表す真理述語 $T(x)$ は定義できず、かつ $T(x)$ の定義には“メタ言語”が必要であることを示している。定義不可能性定理は、ただの計算上の不備ではなく、知識体系自身に根源的な制約をもたらす。そして、確率統計的ではない“記号推論型AI”が、そのアプトプットに対して自身では真であることを評価できない、理論的背景である。

王国では、外部との対話によって知識体系はより豊かで複雑になったが、そこでも矛盾や限界がみえた。いつしか人々は探究をやめ、不完全と分かりつつ、すべての決定を神託するのであった。

(外園 康智)

* 証明は、ゲーデルの不完全性定理と同じ構造である。