

人工知能とは何者か

信原 幸弘

1、どんな世の中になるのだろうか

人工知能が大きなブームを迎えている。一九五〇年代後半に人工知能研究が始まって以来、三度目のブームだ。今度こそは本物だという声があちこちから聞こえる。じっさい、今度こそは人工知能が世の中を大きく変えそうな予感がする。そこでまずは、人工知能によって今後世の中がどうなっていくかを展望しながら、人工知能とはいったい何者なのかを考察するための手がかりを得ることにしたい。

アルファ碁の衝撃

いまの人工知能の威力をまざまざと見せつけ、人工知能ブームを一挙に爆発させたのは、おそらくグーグル傘下のディープマインド社が開発したアルファ碁であろう。二〇一六年三月、囲碁の人工知能ソフトであるアルファ碁が世界最強の囲碁棋士の一人である韓国の李世ドルに勝利した。チェスや将棋では、すでに人工知能がトップ棋士を打ち負かしていたが、囲碁はチェスや将棋に比べて、はるかに複雑なゲームであり、人工知能がトップ棋士に勝つのはまだかなり先のことだと見られていた。そのような大方の予想を裏切って、アルファ碁は見事トップ棋士に勝利したのである。

アルファ碁の勝利をもたらしたのは、ディープラーニングという学習法とビッグデータの活用である。アルファ碁は入力層と中間層と出力層からなるニューラルネットであるが、中間層が多く層からなっており、それゆえ入力層と出力層のいずれからも遠い「深い」層がいくつもある。このような深い層を含む階層化されたニューラルネットによって、与えられた入力にたいして適切な出力を返すように学習を行わせるのがディープラーニングである。アルファ碁はこの学習法により、囲碁のプロ棋士の膨大な数の対局データに基づいて学習を行うことで、トップ棋士に勝つことができたのである。

ビッグデータによる学習では、結局のところ、データに含まれていることしか学習できず、真に新しいことを見いだすことはできないと言われることがある。これは「真に新し

いこと」が何を意味するかにもよるが、まずまず正しいと言ってよいだろう。しかし、そうだとすると、人工知能はすでにその限界を突破しつつある。つまり、既存のデータがなくても、自分でデータを作り出して学習することが可能なのである。そのようにして誕生したのがアルファ碁ゼロである。この囲碁ソフトはプロ棋士の対局データを用いて学習するのではなく、自分で自分と対局するという自己対局を数えきれないほど繰り返すことによって学習を行い、アルファ碁を凌駕する力を達成したのである。

自動運転車

囲碁や将棋のソフトは思考領域の人工知能と言えるだろうが、それにたいして知覚・運動領域でいま大変な注目を集めているのが自動運転車である。自動運転車は自分で道路状況を知覚的に認識し、それに基づいて速度調整や衝突回避などの適切な運転操作を行う人工知能である。その運転技能はすでに人間の技能を上回っており、人間よりも事故率が低いだろうと言われている。自動車の場合、運転操作自体はそれほど複雑なものではない。アクセルやブレーキを踏み、ハンドルを回すだけである。困難なのは道路状況の把握である。どこに歩行者がいるのか、信号は青か、前方の車が急ブレーキをかけないかなど、多くの複雑な状況を瞬時に的確に把握しなければならない。そのような状況を把握できさえすれば、それに応じてどう運転すればよいかはそれほど難しいことではない。

自動運転が可能になってきたのも、人工知能がセンサーを通じて周囲の事物を知覚的に認識できるようになってきたことが大きい。人間は目や耳などの感覚器官を通じて周囲にどんな事物がどのようなあり方をしているかを知覚的に認識できるが、人工知能もそのような知覚的な認識を行うことが可能になってきたのである。道路状況の把握は知覚的な認識の一種にほかならない。それゆえ、人工知能も道路状況を把握できるようになってきたのである。

自動運転車は技術的にはもうほとんど完成していると言ってよい。それを社会に導入するうえで問題となるのは、むしろ事故が起きた場合の責任の問題である。自動運転車が事故を起こした場合、誰が責任をとるのか。自動運転車は自分で運転しているとはいえ、おそらく責任を担えるような主体的な存在とはみなされないだろうから、自動運転車に責任を問うことはできない。そうだとすれば、自動運転車の制作者か、利用者か、あるいはさらに別の者か（関係者への応分の責任ということも考えられる）。これは少し厄介な問題だが、いずれ適当な法整備がなされて解決されるだろう。そしてその暁には、自動運転車がおおいに普及することになるう。

言語理解

人間の知能のなかでも、言葉を理解する能力は、人間の社会的・文化的な生を成り立た

人工知能とは何者か

せるうえでとくに重要な役割を果たしている。一緒に仕事をするのにも、気持ちを通わせあうのにも、言葉が通じなければ、できることはきわめて限られている。母国語が別で、お互いの言語を知らない者どうしても、多少の共同作業や気持ちのやりとりはできるが、言葉が通じる者どうしに比べれば、どうにももどかしいほど、それは圧倒的に限定されている。人工知能が開発された当初から機械翻訳はきわめて重要な目標のひとつだったが、それもそのはずである。

言語理解の領域でも、現在の人工知能においては、着実に開発が進められている。日本マイクロソフトが開発した人工知能の女子高生「りんな」は、LINE やツイッター上で会話することができ、若者を中心に多くのユーザーが「りんな」との会話を楽しんだ。また、政治的発言をすることで話題になった人工知能もある。中国テンセント社の SNS「QQ」上で利用者と会話する人工知能「ベビーQ」は、ユーザーが「共産党万歳」と書き込むと、「こんなに腐敗して無能な政治に万歳できるのか」と批判的な答えを返した。これは SNS 上の利用者の発言から学習したものだろうと推測されている。また、マイクロソフト社がツイッター上で展開している人工知能「Tay」はヒトラーを礼賛する発言をし、そのため閉鎖された。これもユーザーの発言から学習したものと考えられる。

人工知能による言語理解への挑戦は、日常の会話にとどまらず、小説を書くという文芸の域にまで及んでいる。Botnik というグループは『ハリー・ポッター』シリーズから学習した予測入力プログラムを用いて『ハリー・ポッターと灰の肖像画』（二〇一七年）を書いている。また、日本でも、はこだて未来大学の松原仁教授を中心とするグループによる「きまぐれ人工知能プロジェクト 作家ですよ」は、星新一のショートショートを学習することで小説の執筆を目指している。それはいまのところはまだ、人間があらかじめ書いた文章を分解し、それを再構成して小説にするという段階にとどまっているが、それでも二〇一六年に星新一賞の一次審査に通ったことで話題となった。

東ロボの挫折

このように見てくると、日常の会話を自由に行い、知らない言語も即座に翻訳してくれる人工知能が出来上がる日も、そう遠くはないように思われる。しかし、その一方で、言語理解については、それほど楽観できないのではないかと思わせるニュースもある。国立情報学研究所が中心となって二〇一一年に開始されたプロジェクト「ロボットは東大に入れるか」の人工知能キャラクターである東ロボくんは、二〇一六年度までに大学入試センター試験で高得点をマークし、二〇二一年度に東京大学の入試に合格することを目標にしていた。しかし、東ロボくんは文章の意味が理解できず、今後の進展も見込めないということから、二〇一六年にこのプロジェクトは断念された。じっさい、東ロボくんは文章理解が大きく関係する国語や英語の試験がとくに苦手であった。

試験への合格ということ言えば、中国の iFlytek 社の開発した人工知能が中国の国家資格である医師免許試験の筆記試験に合格したという報道がなされている。しかし、このような医師試験はあまり文章の理解を必要とせず、たんに知識を問うだけの問題であったために、人工知能でも合格できたと考えられる。現在の人工知能にとっては、定型的な文でなければ、簡単な文章でも、その意味を適切に理解するのが非常に困難なのである。

なぜ人工知能は、囲碁や将棋は途轍もなく強いのに、また自動運転もできるのに、簡単な文章が理解できないのであろうか。その理由を探ることによって、人工知能とはいったい何者なのかという問いに迫っていきたい。

2、なぜ文章が理解できないのか

人と会話したり、メールや本を読んだりするなど、私たちは日々、文章の理解を行っている。ときには理解に苦しむこともあるが、たいていの場合、簡単に理解できる。しかし、文章を理解するというのはどういうことであろうか。人工知能にとって文章の理解がどうして困難なのかを探るために、そもそも文章を理解するとはどのようなことかを考察してみよう。

中国語の部屋

文章理解をめぐるのは、サール（1984）の有名な議論がある。まずは、それを検討してみよう。サールはつぎのような状況を想像してみよと言う。いま、ひとつの部屋があり、そのなかにトムがいるとする。トムは英語を母国語とし、中国語はまったく知らない。トムがいる部屋には、中国語の文章をその字形に基づいて操作することを可能にする膨大なマニュアルがある。このマニュアルは英語で書かれているため、中国語を知らないトムでも、中国語の文章をその字形に基づいて操作することができる。

トムのいる部屋には、外部の人と文章をやりとりするための受取口と差出口がある。中国語で書かれた質問が受取口に差し込まれると、トムはそれを受け取って、マニュアルに従って質問の文章を操作し、答えとなる中国語の文章を作成して、差出口から返す。マニュアルがとても良くできているので、トムの作成する答えは中国語を理解する人が作成する答えと何ら遜色がない。部屋の外にいる人から見れば、部屋のなかで中国語の理解がまさに生じているかのようである。

しかし、部屋のなかで中国語の理解が生じていないのは明らかだ、とサールは主張する。トムは質問の文章を理解せずに、ただその字形に基づいて操作しているだけであり、そうやって答えを導き出しても、その答えの文章が何を意味するかをまったく理解できない。それはトムにとっては意味不明の中国語の文章である。ここからサールは、たんなる記号

の操作だけでは、文章の理解は生じないと結論づける。

チューリングテスト

サールの「中国語の部屋」の思考実験は一見、説得的に見えるが、本当に記号操作だけでは理解は生じないのだろうか。たしかに部屋のなかにいるトムは中国語の質問も答えも理解できない。しかし、トムとマニュアルを含む中国語の部屋全体は、中国語の質問にたいして適切に中国語で答えを返すことができる。そうであれば、それは中国語を理解できると言ってよいのではないだろうか。人間がある言語を理解できるかどうかを問題にする場合でも、その人が適切な質疑応答をすることができれば、それで十分である。質疑応答をするさいに、その人の心や頭のなかでどんなことが起こっているかは問題ではない。どんなことが起こっていようと、質問にたいして適切に答えられれば、それで十分なはずである。

チューリング（1984）はコンピュータが考えているかどうかを判定するためのテストとして「チューリングテスト」を考案したが、それはまさにそのような考え方に基づくテストである。いま、壁の向こう側に一台のコンピュータと一人の人間がおり、それぞれにたいしていろいろなことがらについて文章でやりとりをする。そのやりとりから、どちらがコンピュータでどちらが人間かがわかれば、コンピュータは考えていないと判定され、そうでなければ、考えていると判定される。つまり、人間と区別できないような適切な応答ができれば、コンピュータも人間と同じく考えているとされるのである。

チューリングテストの根底にあるのは、心のあり方はすべて行動に現れ出るという心にかんする「現出原理」である。心のあり方の違いは必ず行動の違いとして現れ、それゆえ行動になんの違いもなければ、心のあり方にも違いはない。そうだとすれば、コンピュータが人間と同じ行動をするなら、コンピュータも人間と同じ心のあり方をするということになる。こうして人間が考えるなら、人間と同じ行動をするコンピュータも考えるとされるのである。このようなチューリングテストの考え方によれば、中国語の部屋は中国語を理解できる人と同じ応答をすることができるので、中国語を理解できると言ってよいはずなのである。

記号操作だけで適切な応答ができるのか

サールの思考実験で問題とすべきなのは、中国語の部屋で理解が生じているのかどうかということではなく、そもそも中国語の部屋が適切な答えを出せるのかということであるように思われる。つまり、記号操作を行うだけで、本当に適切な答えを出すことができるのだろうか。サールはこの点をまったく問題にしていないが、ちょっと考えてみれば、そこに大きな疑問があることがすぐわかろう。私たちが自分の理解できる言語で書かれた質

問を読むとき、心にさまざまな思考やイメージ、情動が湧き起こる。これらは質問への答えを導き出すうえでおおいに関係する。どんな思考、イメージ、情動が湧き起こるかで、答えは変わってくる。思考は心のなかの語り（つまり内語）とみなせるかもしれない、それゆえ記号の組み合わせ（すなわち文章）にほかならないかもしれないから、思考は記号操作に組み込むことが可能かもしれない。つまり、質問の文章にたいしてどんな思考の文章を形成し、それをさらにどう変形していけばよいかがマニュアルに書かれており、それに従って思考に関わる記号操作を行うことができるかもしれない。しかし、イメージや情動はどのようにして記号操作に組み込むことができるだろうか。ある文章が現れたときに、どんなイメージや情動を抱くか、そしてそのようなイメージや情動を抱いたときに、どんな文章を生み出すかをどのようにしてマニュアルのなかに書き込むことができるだろうか。

記号を記号だけではなく、イメージや情動とも関連づけるような規則を定式化することは非常に困難であろう。しかし、そのような規則をマニュアルのなかに書き込まなければ、質問にたいして適切な答えを返すことはできないだろう。記号を直接的に記号と関連づけるような操作を行うだけでは、適切な答えを導き出すことはできない。適切な答えを導き出すためには、イメージや情動を介して記号を操作しなければならない。しかし、そのようにしてイメージや情動も加わって適切な答えが導き出されるのであれば、そこには記号の理解が生じているという感じがするのではないだろうか。「中国語の部屋」の思考実験で理解が生じていないように思われるのは、中国語の文章をただひたすらその字形に基づいて操作するという様子を想像しているからであろう。しかし、そうではなく、イメージや情動が適切に関与するあり様を想像すれば、むしろまさに理解が生じているという感じがするのではないだろうか。

意味の状況依存性

文章は文章の内部だけで閉じているわけではない。それは文章の外の物理的な事物および心のなかのイメージや情動とも繋がっている。文章を理解するには、そのような全体的な連関を把握しなければならない。文章の理解が困難なのは、ひとつには、この連関がきわめて広範だからである。

しかし、現在の人工知能は事物を知覚的に認識したり、顔の表情を捉えたりすることもできる。つまり、周囲の物理的な事物も、心的なイメージや情動も、かなり取り扱うことができる。そうだとすれば、文章と物的なものや心的なものとの連関がどれほど広範でも、人工知能はその連関を把握して、文章を理解することができるのではないだろうか。

たしかに文章どうしの繋がりや、文章と他の物的および心的なものとの繋がりが一義的に確定したものであれば、そのような繋がりやがどれほど広範であっても、それを把握することは、現在の人工知能でも可能かもしれない。しかし、同じ文章でも、それが他のどの

文章と繋がり、どの物的および心的なものと繋がるかは、状況に応じて変わってくる。「窓が開いているね」という発言にたいして「そうですね」と応じることが適切な状況もあれば、そうではなく「窓を閉めましょう」と応じなければならない状況もある。前者の状況では、「窓が開いているね」という文は、じっさいに窓が開いているという物的状態や、発言者がそれを見ているという心的状態と繋がっている。また、後者の状況では、窓が開いているという物的状態だけではなく、窓から風が吹き込んでいるという物的状態や、窓を閉めてほしいという発言者の意図とも繋がっている。

文章はさまざまな他の文章や物的および心的なものと繋がるのが可能であり、何と繋がるかは状況に応じて異なってくる。文章を理解するには、その文章が発された状況において、それが他のどんなものと繋がっているかを理解する必要がある。それぞれの状況において文章が他のどんなものと繋がっているかが、その状況における文章の意味にほかならない。したがって、文章の意味は状況に相対的であり、文章を理解するには、そのような状況に応じた意味を把握しなければならない。しかし、可能な状況は無数にあり、状況に応じた文章の意味を理解することは、きわめて膨大かつ複雑な作業である。人間はたいいの場合、そのような作業を難なく行っているが、それはけっして簡単な作業だというわけではない。意味の状況依存性こそ、人工知能にとって文章の理解が困難な真の理由である（この点については、川添 2017 がじつに楽しい寓話的な物語を交えながら、平易かつ丁寧な説明を行っている）。

3、空気を読むのは難しい

人間にとって文章の理解がそれほど困難でないのに、現在の人工知能にとってはそれがきわめて困難なのは、文章の意味が状況依存的であり、そのような状況依存的な意味を把握するのが人間にはふつう容易なのに、現在の人工知能にはきわめて困難だからである。しかし、なぜ人間には容易なのに、現在の人工知能には困難なのだろうか。この点を考察することで、現在の人工知能がどういうものかの理解をさらに深めよう。

人工知能はフレーム問題に苦しむ

意味の状況依存性と深い関わりのある問題として「フレーム問題」がある。フレーム問題は「関連性の状況依存性」に関わる問題であり、人工知能研究の初期のころから議論されてきた。しかし、人間にはその問題がたいい容易に解決できるのに、人工知能にはいまだにそれが解決できていない。

いま、カップに入ったコーヒーを台所からリビングに運ぶとしよう。この課題を首尾よく遂行するためには、それに関連することがらとそうでないことがらを区別して、関連す

ることがらを漏れなく考慮するようにしなければならない。たとえば、カップが傾くと、コーヒーがこぼれるから、カップが傾いていないかどうかは関連することがらである。また、床に障害物があるかどうかもそうである。しかし、壁が何色か、気温が何度か、といったことは、関連することがらではなく、考慮する必要はない。

では、課題に関連することがらとそうでないことがらをどうやって区別すればよいだろうか。すぐ思い浮かぶのは、その場のさまざまなことがらについて、ひとつずつ関連性を調べていくという方法であろう。しかし、その場で成立していることがらは無数といってよいほどたくさんある。カップの傾きや壁の色などのほかにも、窓が開いているかどうか、テレビがついているかどうか、天井がどれくらいの高さかなど、膨大な数のことがらがある。それらをひとつずつ調べていたのでは、途方もない時間がかかってしまい、コーヒーを運び始めることすらできないだろう。課題に関連することがらとそうでないことがらは、もっと効率的に区別できなければならない。では、どうすれば、効率的に区別できるだろうか。これがフレーム問題である。(デネット 1990 には、関連性を調べるのに忙殺されて、いっこうに課題に着手できないロボットの悲劇がユーモラスに描かれていて、フレーム問題がじつに明快に解説されている。)

フレーム問題を解こうとするときにまず注意を引くのは、課題に関連することがらがふつうごく少数であり、その他の膨大なことがらは関連しないということであろう。そこで、課題ごとに、それに関連することがらをあらかじめリストにして取りまとめておくという方法が考えられる。そうしておけば、課題を遂行するさいに、その課題用のリストを参照すれば、ただちに関連することがらがわかる。

この方法は非常にもっともらしく、人間もそれを用いているのではないかと思われるかもしれないが、じつはそれには致命的な欠陥がある。というのも、課題に関係することがらが状況に応じて変化するからである。たとえば、雨が降っているかどうかは、ふつうコーヒーを運ぶ課題にとって関連しないことがらであるが、雨漏りのする家では、それは関連することがらになる。課題に関連することがらがこのように状況依存のだということになると、関連することがらをリストにするにしても、リストは状況ごとに作成しなければならない。しかし、可能な状況はほとんど無数と言ってよいほどある。それゆえ、状況ごとにリストを作成するのは実際上不可能である。

このように課題に関連することがらが状況依存であるために、フレーム問題を解くことがきわめて困難になる。フレーム問題の困難さは関連性の状況依存性にあるのである。

鍵は情動にある

フレーム問題は現在の人工知能にとって非常に困難だが、人間にとってはたいい容易な問題である。では、人間はどのようにしてフレーム問題を解いているのであろうか。人

人工知能とは何者か

間がフレーム問題を解くときの鍵となっているのは、情動だと思われる。人間がある課題をある状況で遂行しようとする、その状況でその課題に関連することがらがおのずと際立って立ち現われてくる。コーヒーの入ったカップを運ぼうとすると、カップの傾きや床の障害物などがおのずと前景に浮かび上がり、その他のことがらは背景に沈む。つまり、関連することがらだけが興味・関心を引くものとしておのずと際立ってくるのである。

このように状況に応じて課題に関連することがらが情動的に際立ってくるとすると、その状況で課題に関連することがらをそうでないことから区別するのは、もはや簡単なことである。それはいわば関連することがらに赤で印が付けられているようなものである。人間は情動的に際立つことがらだけを考慮することにより、課題を首尾よく遂行できるのである。(フレーム問題を解く鍵が情動にあることの解説として、信原 2017 を参照。)

文章理解はフレーム問題の一種だ

文章を理解することは、関連性という観点から言えば、その文章と関連するさまざまなことがらを把握することだと言えよう。文章は状況に応じてさまざまな他の文章や物的および心的なものと繋がっており、そのような繋がりがその状況における文章の意味である。文章が他の文章や物的および心的なものと繋がるということは、その文章がそれらのものと関連するということである。それゆえ、文章の意味を把握することは、その文章と関連することがらを把握することにほかならない。こうして文章を理解することは、その文章と関連することがらを把握することなのである。

そうだとすれば、意味の状況依存性は関連性の状況依存性の一種だということになる。文章の意味が状況に応じて変化するということは、文章と関連することがらが状況に応じて変化するということである。したがって、文章を理解するという問題はフレーム問題の一種であり、その困難さはフレーム問題の困難さにほかならないということになる。

フレーム問題を解く鍵は情動にあった。そうだとすれば、文章を理解する鍵も情動にあることになる。状況に応じて適切に文章を理解するためには、その状況においてその文章に関連することがら(すなわちさまざまな他の文章や心的および物的なもの)がおのずと興味・関心を引くものとして情動的に際立ってこなければならない。このような際立ちを可能にする情動能力がなければ、文章を理解することはおそらく不可能であろう。

場の空気を読む

現在の人工知能にとって文章理解が困難なのは、それが興味・関心の情動を欠いているからであろう。現在の人工知能でも、顔の表情から情動を判定したり、「バカ」と言われると、悲しそうな表情を浮かべたりすることはできる。しかし、このような単純な情動能力だけでは、文章の理解に必要な情動をもつことはできない。状況に応じて文章を適切に理

解するためには、その状況のもとで文章と関連することがらを適切に際立たせるような情動が必要である。現在の人工知能にはこのように状況に応じて柔軟に変化する情動を形成する能力が欠けているのである。

状況依存的な文章の意味を把握することは、その文章が発された状況がそもそもどんな状況かを把握することであり、それはようするに、その場の空気を読むことにほかならない。空気を適切に読むことができないければ、文章を適切に理解することができないのである。「ぜひ頼む」とは言わないが、あれやこれやと言う人がいたときに、その人の発言を依頼として理解しなければ、それはおそらく不適切な理解となろう。つまり、空気が読めていないために、適切な理解ができていないのである。空気を読むことは、人間の場合、情動による。空気が一瞬にして変化することもあるが、それもふつう情動によって瞬時に感じ取られる。情動によって適切に空気を読むことによって、状況に応じた適切な対処が可能になるのである。人工知能は人間のような繊細で柔軟な情動を欠くために、状況に応じた適切な対処ができない。それゆえ、文章を適切に理解することもできないのである。

参考文献

- 川添愛(2017)『働きたくないイタチと言葉がわかるロボット——人工知能から考える「人と言葉」』朝日出版社
- サール、ジョン(1984)「心・脳・プログラム」(久慈要訳) D・R・ホフスタッター&D・C・デネット編著『マインズアイ——コンピュータ時代の「心」と「私」』下巻、TBSブリタニカ、所収(原著論文は一九八〇年に刊行)
- デネット、ダニエル・C(1990)「コグニティヴ・ホイール——人工知能におけるフレーム問題」(信原幸弘訳)『現代思想』一九九〇年五月号(原著論文は一九八四年に刊行)
- テューリング、アラン(1984)「計算機械と知能」(藤村龍雄訳) D・R・ホフスタッター&D・C・デネット編著『マインズアイ——コンピュータ時代の「心」と「私」』上巻、TBSブリタニカ、所収(原著論文は一九五〇年に刊行)
- 信原幸弘(2017)「フレーム問題と情動」信原幸弘編『ワードマップ 心の哲学——新時代の心の科学をめぐる哲学の問い』新曜社