

2013 年 11 月 18 日

言語記述の意味と知識

小山照夫 国立情報学研究所

1. はじめに

自然言語の主要な働きは、人間の間で有用な情報の伝達を行うことであり、この背景には言語記述の意味というものの存在が仮定されている。実際、言語記述の送り手は、発話ないしは文字列を受け手に提示することによって、意味のある情報を受け手に伝え得ると信じているし、一方で言語記述の受け手から見た場合、受け手は送り手の話した(書いた)言語記述を観察し、それが文と認められるならば、そこに意味を読みとることができると思っている。

自然言語を用いた情報交換は、人間にとって日常的なあたりまえのものと考えられており、そこで実際に何が起きているのかに関して疑問が持たれることはほとんどない。しかし、このことは果たして自明のこととして良いのだろうか。特に「意味」というものがどのような本質を持ち、また、どのような前提に支えられているのかは自明のこととは言えないように思われる。

以下では自然言語記述によって「意味」が提示でき、また読み取れると信じる根拠は何か。また、ここで問題とされる意味とは何かについて考察した後、関連する問題として併せて知識処理に関わるいくつかの側面について考察する。

自然言語記述とその意味に関する問題は二つ存在する。一つはここで言われる意味とはどのようなものであり、またどのような条件によって構成されているのかである。言語記述は記号であると考えられており、なんらかの対象を参照していると考えられるから、このことは言語記述に現れる記号をその参照対象と結びつける根拠は何かという問題と言ってもよい。また、もう一つの問題は、意味を支える根拠として、受け手の知識が重要であり、その限りでは意味は受け手の知識に依存するように見えるのに反して、言語記述にはその受け手に依存しない一般的な意味があると広く信じられているのだが、このことは何を意味しているのかである。

2. 受け手から見た言語記述の意味

2.1. 受け手の知識と意味

主として受け手の側から言語記述の意味を問題とすると、記述の対象となる世界について受け手の持っている知識が重要であると考えられる。たとえば「犬が歩く」という記述が与えられたとき、この記述列が各記号の文法的分類に基づいて文としての構造を持つことを確認するだけでは、記述の意味を議論するには不十分である。例えばここで「犬」という名詞が、どのような世界の、どのような対象を指示することができるか、あるいは少なくとも対象のあり方を十分に限定できる制約が示されているか明らかにできなければならない。ここでは「犬」と呼ばれる対象が世界の中でどのような存在であるかに関して受け手の持っている知識が加わって、初めて記述の意味が現れてくることができると言えるであろう。意味処理は既に知識処理であると言っても良い。

ここで問題は、受け手の知識はどのような形で記述の意味理解に関与するかである。言語記述とはそもそも何であるかを考えるなら、それは他の主体である送り手から受け手に対して提示される文字列ないしは音韻列である。提示された列は一般には文の形をとると考えられるのであり、文の種類にはいくつかのものが考えられるのだが、以下では簡単のために平叙文ないしはそれに相当する文の構造を持つものが与えられた場合に限定して考察することとしよう。

2.2. 自然言語処理の概要

最初に自然言語記述を解析する方法を考えよう。ここでは文の構造と、その文によって示される意味の決定が基本的な問題とされる。人間にせよ機械にせよ、言語処理では一般に形態素解析、構文解析、意味解析が行われると考えられている。人間は実際に、文字列ないしは言語記号列が提示された場合、それが単語列として解釈できるかを判断し、また、単語並びが各単語の文法的属性から、文としての正当な構造を持つかどうかを判断するであろう。これは、基本的に形態素解析と構文解析を行っていると言える。

形態素解析と構文解析が終了し、与えられた記述が文としての構造を持っていることが確認されれば、引き続き意味解析が行われることになるが、それでは意味解析では何が行われているのだろうか。人間の場合、記述の適用される特定の世界を想定し、記述中に認められる各記号がその世界でどのような物事と対応しているかを決定していると考えてよい。これは結果として記述に現れている限りの事態が再構成されることになると思うことができよう。人間の場合、意味は単一の文についてのみならず、一連の、関連する文の集合についても解析されると考えられるが、ここでは単一文の意味に限定する。

この、記述に対応する、実際になんらかの世界で成立している事態を再構成するということは、言語理解の問題では古くから指摘されていたことで、例えば Frege の *Bedeutung* はその代表的なものであったといえる。平叙文で記述されていると想定できる事態の再現は、記述の意味の主要な部分を占めているとも言えるであろう。しかし、記述に現れた限りの物事の間の関係のみを再現することをもって記述の意味が決定されたとして良いのだろうか。

2.3. 記述と事態の関係

言語記述と記述されている事態との関係を考えるなら、言語記述中に陽に現れるのは、対応する事態の

ほんの一部に過ぎない。記述の中に例えば「犬」という記号が現れたとして、この字面だけを見たのでは、様々な種類のいる犬という動物の中で、どのような個体なのか判断することはできない。

しかし、実際の事態としては、どのような世界で起こるにせよ、そこに出現するのは抽象的な犬ではなく、世界内の具体的存在としての犬であるはずである。この、実際に世界内に出現している実体は多くの属性を持つものであり、そのすべてを言語的に記述することは一般には不可能である。

それでは記述の意味を考える際に、記述に現れない事態の側面は受け手にとって問題とならないのであろうか。記述に陽に現れていない事態の側面であっても、もしそれが一定以上の確実さで推定できるなら、より価値の高い情報となる可能性があるのではないか。実際には受け手は、記述に現れる以上の、事態のより詳細な側面を推定していると考えられる。

2.4. 知識による事態の補足

一般に言語記述そのものを見るとき、それが特定の事態を部分的に記述するものであったとしても、本来生起している事態と比較すれば、そこには極端に縮退された情報のみが示されていることになる。しかし記述の受け手が人間である場合、自身の持つ知識を活用することにより、単に記述に現れた以上の事態の詳細を把握できることがある。

例として「作る」という動詞を考えてみよう。この動詞は基本的には、人間がその行動を通じて、なにか特定の材料から、それまでに存在しなかった物を新しく出現させることを示す。この基本構造は項構造に相当するものであり、したがって文法的な規則であると考えられる。動詞単独として見た場合、この語で記述される事象はそれ以上のものではありえない。しかしこの語が用いられる文中で同時に作られるもののクラスが指定されるとき、受け手のうち少なくとも一部の者、作成対象に関する十分な知識を持っている者は、その作成される過程や状況までをかなり詳細に推測することが可能となる。特にその対象を過去に作成した経験のある人間であれば、ほとんどすべてのプロセスまでを推定することが可能となるであろう。ここでは受け手が過去に経験した類似事態の記憶が知識として働いてくることになる。

「作る」のような、複雑なプロセスを含む多様な事態を記述できる動詞の場合、たとえ項構造としては類似していても、その働きかける対象、この場合は作成される対象によって、事態の様相がまったく異なる可能性がある。野菜を作る場合と模型飛行機を作る場合とでは、そこで使用される材料や道具、必要な環境や操作手順などがまったく異なるのであり、実際の事象としてみた場合、作るという動詞が多義性を持つように見えるほどその様相が異なっている。このような場合には受け手の知識を適用した結果までが、受け手の捉える意味に含まれることになるのではないだろうか。

ここで対象の作成ということに知識を持つ受け手は、その違いを明確に想定できるのに対して、知識を持たない受け手は、むしろ漠然とした違いがあるかもしれないという程度にしか、記述される事象の相違を想定できないかもしれない。このことはとりもなおさず、記述の意味は受け手ごとに変わってくる可能性を示している。

2.5. 文脈知識の働き

受け手が把握できる意味に影響する他の知識として、記述の現れる文脈がどのようなものであるかに関する知識もまた、対応する事態の、記述自体には現れない詳細を想定する上で大きな違いをもたらすことがある。「犬が病気になった」という記述は、これ自体ではどの「犬」の状態に言及しているのか決定できないが、しかし記述の受け手が送り手を良く知っており、送り手が現在飼っている犬の話をしているのだと前後の会話の文脈から理解できる場合には、ここでの「犬」について、個体決定が可能であり、それに基づいてさらに詳細な情報を補完することができるであろう。ここでの「犬」はもはや、どこかの世界の詳細不明な対象ではなく、大きさや色、犬種や年齢までもが特定できる存在ということになるであろう。このような場合にも知識の有無によって受け手が再現できる事態の詳細さが異なってくる。

さらに極端な場合には、受け手の知識と文脈理解は、しばしば受け手に非文からの意味読み取りを可能とさせる。極端な場合には、親しい間同士では「あれ」という記述だけできわめて複雑な内容が伝わってしまうことさえある。

3. 共通の意味は存在するか

3.1. 共通意味の仮定

これまでに考察してきた限りでは、受け手が言語記述から読み取る意味は、受け手の持つ知識に依存すると考えられる。しかしそうすると、受け手の持つ知識は受け手ごとにそれぞれに異なったものと考えられるから、受け手の把握する意味もそれぞれに異なることになる。

一方で言語研究ではむしろ、言語記述には受け手とは独立して確定した意味があると考えらる傾向がある。記述に対する、このような共通の意味は、意味と知識の関係を考えるなら、本質的に仮想的なものと考えられるのだが、それではそれは一体どのようなものなのか、また、それを仮定することにはどのような意味があるのであろうか。

3.2. 共通意味と共通知識

言語記述の意味確定には知識が必要であることを前提とするなら、共通の意味が存在するという考え方の背景には、当該言語を母語とする成人ならばほぼすべてが共通に持っており、言語記述に対して適用可能な知識が存在することを想定していることになる。これは、おそらくは漠然と常識と呼ばれている知識が相当するのであろう。

ただ、常識として認められる知識の範囲を具体的に決定することには大きな困難が予想される。たとえば千人というサンプルに対して95%以上が認める知識を常識と考えるとき、主要な部分はともかく、細かな

周辺の知識を常識に含めるかどうかはあいまいであり、かつ、全体を知識集合と仮定するなら、一部の周辺知識を入れ替えてもおよその条件を満たすことはできると考えられるから、常識としての条件を満たす知識集合が複数通り存在することになるであろう。要するに常識というものも、具体的な実在として決定できるわけではなく、単に理念上／便宜上のものであると言えるであろう。

3.3. 共通意味の働き

すると、このような常識を前提とした記述の共通意味もまた理念的／便宜的なものだということになる。それではなぜこのような共通意味が問題となるのであろうか。これはおそらく、最初に出会う主体の間での言語使用の効果を確信させるためのものではあるまいか。

それまで交渉のなかった二つの主体が出会い、何らかの情報交換の必要が生じることは、例えば旅先で買い物をするなどの状況を考えれば、日常常に生じうる状況であると言えよう。ここで、交換される言語記述に最低限共通に認識できる意味が存在すると信じることは、むしろそこで相互の情報交換が成立する必要条件であると考えても良い。この場合、送り手と受け手の常識は厳密に一致している必要はなく、およそのところが似通っていればよい。最初の情報交換が成立したなら、その後は相互の反応を見ながら、状況の文脈構築を行いつつ、相手の知識レベルを推定することによって、より円滑な情報交換が可能となっていくであろう。記述の共通意味もまた、言語の社会的有効性を確信させるための仮想的な構築物であると考えることができる。

また、別の側面として、言語学研究で一般的な議論をするとき、共通意味を仮定しないと議論自体が成立しづらくなることも考えられるであろう。一つの記述について、複数の対応する事態がありうることを前提にするなら、議論その物が複雑なものとならざるを得ない。言語研究ではしたがって、主要な常識の範囲で取り扱え、文脈に影響されない例文を検討対象とすることになる。このような例文は、ある意味不自然なものとなることが避けられない。たとえば「川を渡る」という例文について議論されることはありうるが、しかし、現実の世界の言語使用の中でこのような、帰属する世界も、その世界の中で対象を特定することも省略した文が出現するとは考えにくい。文脈を考慮したとしてもすくなくとも「その川を渡った」と言う必要があろう。

4. 知識と言語

4.1. 知識の非言語的側面

これまで知識というものを漠然と、人間であるなら当然持っている何らかの情報として考えてきたが、しかしこの情報が実際にどのようなものであるかについて、完全に知られているわけではない。

一般に知識というと、教科書にかかれたテキストや、法則を記述する数式などを思い浮かべる。しかし、これらはどのような理由から知識と呼ばれているのであろうか。また、知識と呼ばれる情報はすべてがテキストや数式で記述できるものなのだろうか。

実際のところ、知識の中には、自然言語にせよ、数式などの人工言語にせよ、外部的に記述する事自体が困難なものがある。対象認識に関する知識はその代表的なものである。我々は日常生活で猫と呼ばれる動物や犬と呼ばれる動物に出会うが、初めて見た犬を猫と誤ることはほとんどない。特に犬はその体の大きさや顔の形、体型などに幅広いバリエーションが存在するにもかかわらずである。

世界内の対象を見てその種類を区別できるのは、とりもなおさず世界に関する知識の働きであり、対象の識別ができる事自体、対象識別に関する知識が関与していると考えられるのだが、この知識を、例えば自然言語で記述しようとする、極めて大きな困難に遭遇する。この場合、十分に識別力のある定義を決定する事自体ほとんど不可能であろう。

画像識別知識を自然言語的な記述などで表すことと比べれば、多くの画像サンプルを集めて、画像認識の基本パラメータを学習させる方がいくらかはましではあるが、これもまた、人間の識別能力に匹敵するほどの性能を出すことは容易ではない。

この種の識別に当たっては、人間は過去に体験した感覚の記憶に基づいて、何らかの基準による類似性尺度を用いているように思われるが、それは言語的に記述できるというよりは、個体内の身体感覚に近いもののように思われる。

対象認識の問題と並んで、言語記述が困難なのは身体運動に関わる知識である。ここでは例えば柵の上の物をとるときにどのような身体の動かし方をすれば良いかなどが問題となるが、自然言語で記述された内容だけを元にして正しく動くことは難しく、実際に体験することによって初めて身体の動かし方が理解できるようになる。

一方で、一見自然言語による記述が可能に見える知識であっても、実際には純粋に言語記述の範囲に収まりきらないものも存在する。演繹推論規則などは、一見言語的に記述されているように見える。しかし、一般の言語記述と同様に、ここでも事態のすべてが語られているわけではなく、過剰に単純化された形で記述されているという側面は残っている。したがってそこで使用される記号の参照先としての物事を考えるなら、結局世界内の存在に関わる判断を避けるわけにはいかない。言語的に参照される事態が、実際に起こっているかどうか、同じことだが規則を適用することができるかどうかは、結局は人間の感覚的識別にかかっているとも言える。ここで感覚的な識別のための知識が必要とされるなら、それは結局は言語記述のみで完結しているとは言えないことになる。

4.2. 知識処理の可能性

知識の少なくとも一部については非言語的な側面があるとすれば、知識の機械処理を目指す知識ベースシステムでは、知識が言語記述だけでは完結した形で定義できないことが、システム構築上の大きな問題となる。

初期のエキスパートシステムでは、知識は工夫さえすれば忠実に言語的に記述できるという前提から、

単純に専門家が自然言語を用いて提示する情報を知識として機械利用可能な形に変換して蓄積することで知的システムが実現できると考えてきた。

しかし言語記述が実際には、対象とする事態のごく一部しか記述していないことを考えると、与えられた言語記述の背景にはさまざまな隠れた側面が存在しているのであり、実際の問題に知識を適用するためには、背景となる知識までを記述する必要があることが指摘されてきた。

しかし、問題はおそらくさらに複雑で、上述のように表面的に提示された知識の背景を考察していくと、そこには言語的ではない知識の存在が明かとなってくる。このような知識を、近似的に言語記述で表すことはある程度可能であろうが、本来言語的でない知識を無理に言語的に表現することは、知識を定義する人間にとって大きな負担となる。しかも、記述された知識の正当性や有効性は必ずしも保証されず、実際に様々な場合に適用することを通じて、実用的な定義となっているかどうかを検証する必要がある。これではシステムを構築する労力は膨大なものとならざるを得ないし、ある程度動くシステムができたとして、後から明かとなった問題を解決するためにシステムを修正する労力も大変なものとならざるを得ない。

この問題に対しては、対象とする事象について、カバーすべき範囲で全体を見通すことを可能とするモデルをまず構築すべきであると考えられる。すなわち、専門家の記述する内容をそのまま知識として採用するのではなく、専門家の記述を参考にしながら、むしろ対象とする事象がどのようなモデルの下に理解されているかを推定する手がかりとし、具体的にモデルを明かにするのが適切であろう。こうして決定されたモデルの下に、実際に対処に必要な事態がどのようなものであり、どのようにして事態の発生を認識する方法を考えるべきである。

人工的なシステムを対象とする場合には、個別の要素に関するモデルは存在していると考えてよいから、それらを組み合わせた全体システムとして、どのようなモデルを設定すれば必要な事態に対処できるかを検討すべきであろう。

一方で医療など、人工的ではないシステムが主要な問題となる場合には、既存の知識から、心身状況や生理学的メカニズムについてどのようなモデルが想定され、どのような問題に対して適用されるかから検討する必要がある。この場合には個々のモデルの有効範囲に限界があり、また、視点に応じて同じ対象についても複数のモデルを使い分ける必要があるなど、より複雑なものにならざるを得ないであろう。

このようなモデルを設定した上で、その分野の専門家が問題とする言語記述が、どのようなモデルのどのような状態に関するものか、その時観測される、症状などの感覚的現象はどのような言語表現をとり得るかを検討し、必要とされる知識を、現実に応用可能な形で記述することが有効と考えられる。

場合によっては既存の知識だけでは、要求されるモデル構築が困難な場合も考えられるが、このような場合には例えば大量データからの学習等により、領域の問題構造を推定するなどを通して、有効なモデル自体を新しく構築していく必要が生じるかもしれない。

言語による知識記述が、実際に適用されるべき知識のごく一部しか記述できていないこと、また、感覚情報に基づいて状況を判断する知識など、本質的に言語的とは言えない知識も存在するところから、記述可能な知識は結局は近似的なものに留まらざるを得ない。このことから、知識処理システムの構築にはやはり多大な労力を費やさざるを得ないことになる。