

言語で表現される概念と翻訳の原理

池原 悟

鳥取大学工学部 〒680-8552鳥取市湖山町南4-101 Email: ikehara@ike.tottori-u.ac.jp

あらまし 概念は話者の認識を掘り取って表現するための網の目であるとするなら、言語翻訳は、原言語の概念で掘り取られた話者の認識を目的言語の網の目で掘りなおして表現することだと言える。言語過程説では、言語表現は、一般に、主体的表現と客体的表現から構成されるとされているが、時枝誠記によれば、客体的表現が概念化された話者の認識を表す表現であるのに対して、主体的表現は概念化されない話者の認識の直接的な表現であると説明されている。これは、概念化されない認識が言語で表現されることを意味し、網の目の議論と矛盾する。そこで、本報告では、三浦言語学の立場から、主体的表現と客体的表現は共に概念化された認識の表現であるとし、従来の単語レベルの概念を表現レベルの概念まで拡張する。また、それをベースに、概念を介した機械翻訳方式の基本原則を提案する。

キーワード： 話者認識、主体的表現、客体的表現、単一概念、複合概念、機械翻訳、言語過程説

Concepturization of Speaker's Cognitions and Machine Translations

Satoru IKEHARA

Faculty of Engineering, Tottori University, Minami4-101, Koyama-chou, Tottori-shi, 680-8552 Japan

Abstract : If we say that concepts are mesh to scoop out speaker's cognitions, language translations are considered to re-scoop the speaker's cognitions represented in the original language by using the concepts of target language and represent them by the target language. Motoki Tokieda pointed out that linguistic expressions are composed by Subjective Expressions and Objective Expressions. According to his Constructive Process Theory of languages, Objective Expressions are conceptualized expressions and Subjective Expressions are non-conceptualized Expressions. This means that non-conceptualized cognitions are also represented by linguistic expressions. And it contradicts the above explanation about concepts. In order to solve this problem, we consider that Subjective Expressions also represent speaker's conceptualized cognitions. Based on this idea we expand the definition of conventional concept to complex concepts and propose a new machine translation principle via concepts.

Key Words: Speaker's Cognition, Subjective Expression, Objective Expression, Simple Concept, Complex Concept, Constructive Process Theory of languages

1.はじめに

自然言語処理では、言語規範に関する知識を体系化することが重要な課題であるが、中でも「意味的約束」(表現の意味と用法に関する約束)を体系化することの重要性が指摘されている [池原 01]。これに対して、近年、伝統的な言語学の分野でも表現の構造と意味に関する種々の研究が行われている。例えば、構文と意味の関係では、南不二男による日本語表現のABC3分類 [南 93]がある。この成果は、従来の構文解析で懸案となっていた複文解析の問題を解決するための手がかりとなり、新しい構文解析アルゴリズムに応用されている [白井他 95]。寺村英夫の構文論は埋め込み文の「内の関係」と「外の関係」の解析や複文の内部構造に着目した意味類型化の方法の研究に応用されており [藤本他 01]、複文構造の研究 [益岡 97]、条件文構造の研究 [益岡 93] なども、複文を対象とした意味類型型パターン知識ベースの設計に使用されている [池原 02]。

また、日本語文型を収集したものとしては、何種類かの日本語文型辞典や感情表現を収集したもの [中村 93, 95a, 稗島 94, ジャマシイ 98, 小池他 02]、比喩表現を収集したもの [中村 95b] などが、出版されている。そのほか、最近、外国人への日本語教育用の教材 [柄柄 87, 友松 96 など] も多数見られるが、それらは従来の言語学関連の著作などとは異なり、種々の表現の用法に関する情報が説明されている点で、機械処理に役立つ。これは、日本の文化に不慣れた外国人に日本語を教えることと一般常識を持たない計算機に日本語を教えることの類似性によるものと言える。

ところで、意味解析型の機械翻訳方式の研究では、すでに、単文レベルでの意味的約束を体系化した「表現意味辞書」が開発され、それを使用した「多段翻訳方式」が実現されている [池原他 87]。しかし、まだ二つの大きな問題を残している。それは、第一に、原言語の表現構造に対する目的言語表現の対応づけが決定論的で単一の目的言語表現が生成されることである。適切な表現選択の仕組みを持たないため、文脈上不適切な表現への翻訳が防止できない。また、第二は複文(埋め込み節を持つ文)や重文(接続構造を持つ文)の表現について意味的な非線形性を扱う仕組みがなく、従来通りの要素合成法が適用されることである。

そこで、これらの問題点の解決を狙って、最近、新しい言語変換方式を実現するためのプロジェクト研究が行われている [池原他 02]。それは、多段翻訳方式で提案された「意味的構造の単位」の概念を意味類型論の観点から発展させると共に、「等価的類推思考の原理」 [市川 63] を言語に適用したものである。この方式の動作原理は、異なる言語の表現を概念を介して意味的に対応づける

ことであり、概念の持つ構造と表現形式の関係を基本としている。

ここで問題となるのは、果たして話者の認識の結果である概念を表現したものと言って良いかどうかである。この前提は、工学的立場からのアプローチで持ち出されたものであるが、言語学的に見れば、にわかには断定できない問題をはらんでいる。これは以下のよう問題である。

すなわち、言語過程説では、言語表現は、一般に、主体的表現と客体的表現から構成されるとされているが、時枝誠記によれば、客体的表現が概念化された話者の認識を表す表現であるのに対して、主体的表現は概念化されない話者の認識の直接的な表現であると説明されていることである。これは、概念化されない認識が言語で表現されることを意味し、概念を介した翻訳では、原言語の客体的表現は翻訳できるが、主体的表現の意味は翻訳できないことになる。そこで、本報告では、三浦言語学の立場から、話者の認識における「客体化」と「概念化」の過程を検討し、従来の単語レベルの概念を表現レベルの概念まで拡張する。また、それをベースに、概念を介した機械翻訳方式の基本原則を提案する。

2.概念形成過程と表現過程

本章では、「言語は話者の認識の中に形成された概念を表現する」と言って良いか、言い換えると、概念化されない話者の認識は言語では表現されないかの問題について検討する。

2.1 言語表現と概念の対応関係

従来、概念とは、対象の持つ特殊性を普遍性として取り上げる分割不能な認識の単位で、言語では、単語や複合語で表現されると考えられている。しかし、言語表現の側から見ると、句や節など複数の単語が組み合わされた表現は概念を表さないと良いかどうか問題となる。現実の言語では、複数の単語が組み合わせられて初めて一定の意味を表すような表現が多いことから、単一の単語で表せないような概念は、複数の単語の組み合わせによって表現されると考えることはできないか、また、その際、表現構造の持つ意味を概念とは別のものとするべきか、などの問題である。

ところで、言語表現で表されるのは概念化された話者の認識のみであると考えると同時に、構造を持った概念を考え、概念の持つ構造が句や節等の表現に対応づけられるとすれば、この問題は、概念化の方法論の問題に帰着する。

言語表現で表されるのは「概念化された話者の認識のみだ」とする考え方は、翻訳の方法を説明する上でも支持される。翻訳は、通常、原言語の持つ「概念」の網の目で掘り取って表現された話者の認識を翻訳者が理解し、理解した内容を改めて目的言語の

持つ「概念」の目で掬いなおして表現することだと言われている。この考えは、上記の前提があって成り立つ。

そこで問題となるのは、果たして、すべての言語表現が、概念化された話者の認識を表すか、逆に、概念化されない認識は表現されないのかと言う点である。句や節の表現には、名詞、動詞などの客体的表現だけでなく助詞、助動詞など主体的表現も含まれる。客体的表現が概念化された対象の表現であることについては、従来の議論に異論はないと思われるが、主体的表現を概念化された表現かどうかについては異論もある [三浦 67, 76]。

もし、主体的表現が概念化されない認識の表現だとすると、果たして、句レベルや節レベルで表されるような概念の存在を認めて良いかどうか問題となる。

2.2 言語表現と概念の対応関係の検討

さて、概念と表現の関係について、概念が単一の単語で表現される場合と、句、節、文など、通常、複数の概念の複合体と見なされるような表現の場合に分けて検討する。

(1) 単一の単語による表現の場合

まず概念が単一の単語で表される場合は、その単語を翻訳するには、両言語に共通する概念を介して、原言語と目的言語間の対応付けを考えればよく、問題は、言語間で概念が対応関係を持つかどうかとなる。

この点、原言語では単語で表されるような概念でも目的言語側と同じ概念があるとは限らないし、一見、同じような概念でも実際は異なる場合があるのが問題である。例えば、生物学的な種としての意味での「リンゴ」は、英語では apple と訳せばよいが、健康的でかわいらしい意味での「リンゴのような類」では、「リンゴ」を apple と訳してもその意味は伝えられず、むしろ誤解される恐れがある。

このような違いは、カテゴリー分類の観点から解析するだけでは説明できない。また、カテゴリー分類の観点から分類される語の概念でも、その分類法は言語で共通とは言えない、例えば、日本語で、「唇」と言えば、口紅をつける部分を指すが、英語で「lip」と言えば、鼻の下の部分」等のように、異なった部分を示す。

いずれにしても、対象言語間で同一の概念は、翻訳できるが、異なった概念は、原理的に翻訳不能で、対象言語間共通の低位概念を用いて説明するしかない。

(2) 複数単語による表現の場合

次に、句、節、文などの表現を考えると、「概念」は単語レベルの表象に対応づけられるものだと考える立場からは、句や節は、認識の中に形成された「概念」がそれぞれ、「単語レベルの字面(表象)」に媒介されて構造化され、表現として組み上げられたものと解釈される。

これを言語過程説の図式に当てはめると、「対象」「認識」(概念)「表象」「表現」となる。但し、「認識」(概念)「表象」は、概念化された認識のみが、「表象」を得ることを意味する。「認識」が「表現」に対応づけられるとする以上、表現の構造は、「表象」からではなく、認識の持つ構造から説明することが必要であるが、この図式では、構造上の情報は、表現にまで媒介されない。従って、どのようにして個々の「表象」から複雑な表現が生まれるのかについては別の方法で説明する必要がある¹⁾。

これに対して、「概念」を、言語表現上、必ずしも単語で表されるだけでなく、句や節などの表現でも表されるものと考え、句や節を単語レベルの「表象」から組み立てられた表現だと説明する必要はなくなる。この場合、「概念」と表現の間を「表象」で媒介する必要はなく、認識は、「概念」を介して直接「表現」に対応づけられる。これを図式的に書くと、「認識」(概念)「認識」(概念)「表現」である。この場合も、翻訳が可能か否かは、原言語の持つ「概念」と同じ「概念」が目的言語側に存在するか否かに依存する。言語間に同一、もしくは、ほぼ同一と見られる概念がなければ、翻訳は不可能である。

(3) 言語表現と概念についての仮説

そこで、句や節で表現される概念の存在を認めるか否かである

が、「概念」は、対象の持つ個別性を普遍性として取り上げる認識の単位である」と理解すれば、「概念」がそれぞれ固有の内部構造を持つことと矛盾しない。例えば、「比較」は一つの概念であり、言語表現では認識の単位として使用されるが、この概念は「比較するもの」、「比較されるもの」、「比較する観点」の3つの要素の複合体と考えることができる。

また、慣用的な複合語、「以心伝心」、「間一髪」なども、それぞれ独自の内部構造を持った「概念」と考えるなら、単語や複合語から表現に拡大して、同じ意味内容を表す表現、例えば、「猿も木から落ちる」、「弘法も筆の誤り」、「上手の手から水が漏れる」は、いずれも同じ「概念」を表すものと見ることができ、句や節だけでなく、文レベルの表現も単一の概念を表していると言える。

以上から、本稿では、句や節で表現される概念の存在を認め、単一の単語では表せないような概念をあらわす時は、複数の単語を組み合わせた表現が使用されると考える。

2.3 認識論から見た新解釈の問題点

前節では、言語表現上、概念は、単語だけでなく、句や節などの表現にも対応づけられるとした。そこで、本節では、果たして、句や節等の表現の構成要素である「客体的表現」と「主体的表現」が、いずれも概念化された認識の表現と言って良いかどうかについて検討する。

(1) 主体的表現に関する従来の説明

三浦言語学[三浦 67]は、時枝誠記の言語過程説[時枝 41]を受け継いでおり、言語表現に「主体的表現」と「客体的表現」の区別を設けているが、それらの表現と概念との関係については、解釈に大きな違いがある。それは、時枝言語過程説が「主体的表現」は「概念化」されない話者の感情や意志を表現するとしているのに対して、三浦言語過程説では、「主体的表現」と「客体的表現」のいずれも「概念化」された表現だとしている点である。

三浦は、「主体的表現」における「概念化」を「客体的表現」の場合と区別して「特殊な概念化」であるとしつつ、その詳細については述べていないが、句や節などの表現では、客体的表現としての自立語だけでなく、助詞、助動詞などによる主体的表現も使用される。従って、句や節の表現と概念の関係を説明するには、いずれの立場を取るかを明確にしておく必要がある。そこで、まず、「客体化」とは何か、「概念化」とは何かを定義し、その違いを考察する。

(2) 客体化と概念化

さて、「客体化」と「概念化」の意味の定義であるが、いずれも行為の一種であり、行為者を必要とする。

そこで、まず「対象の客体化」について、「客体化」は、対象が主体でないこと、すなわち話者自身でないことを意識する行為で、対象とそれに立ち向かう話者との対峙関係は、「客体化」によって成立する。」と定義する。

行為者は、通常、「主体」としての現実の話者であるが、話者自身を客体化する場合の行為者は、現実の話者が自己分裂することによって生じた観念的話者である。これに対して、「概念化」とは、話者が、対象の持つ個別的特徴を捨象し必要不可欠の要素を統一体で捉えたことを意味するものとする。行為者は、「客体化」の場合と同じ話者である。

(3) 主体的表現解釈の問題点

次に、「概念化」は、対峙関係があつて初めて可能になると考えると、「概念化」を行うには対峙関係を生じさせるため、あらかじめ「対象」が「客体化」されている(すなわち対象が話者以外のひとまりの存在であると認識されている)必要があることになる。

これによって「客体的表現」は、対象が「客体化」された後「概念化」された表現だと説明される。これに対して「主体的表現」は「客体化」されない表現であるから時枝文法の言うように「概念化」されない表現だということになる。

しかし、この解釈では、それではなぜ「主体的表現」では「概念化」なしに、それを表現するための適切なラベル(単語)が選べるのかが問題になる。「認識」は、確かに「概念化された部分」だけで構

¹⁾ まさにこの考えに基づく機械翻訳方式がトランスファー方式である。トランスファー方式は、単語単位の意味を運ぶ仕組みと表現構造を変換する仕組みの2つから構成されている。

²⁾ 人間は、通常、対象が話者以外のひとまりの存在であることを簡単に認識する。しかし、これが単純なものではないことは、下記の例からも指摘できる。それは、生まれつき目の見えなかった人が、成人してから角膜移植の手術を受けて外界が見えるようになり、ベッドで窓から見える外の絵を書いてもらったところ、その絵は、窓枠と窓枠から見える外の世界との区別が全くなく、また、窓枠が四角形をしていると言う意識も感じられないものだったと言う例である。目が開いた直後は、事物の物理的形狀に関する知識がないため、対象世界に存在する個々の対象物を相対的に独立したものとして捉えることができないが、点や線の形状に関する概念を通じて窓の形状についての知識を獲得するにつれて、次第に対象物を識別することができるようになることを意味している。

成されているという必然性はなく、概念化された部分」と概念化されていない部分」があっても不都合はない。しかし、助詞や助動詞といえども、それらは、単語辞書で定義されているような語義を持つ。これは、認識の形態として、これらの語で表されるような概念が存在することを意味していると考えるのが素直である。

2.4 「客体化」と「概念化」の関係

前節で述べたように、「客体化」が話者と対象の間の「対峙関係」の成立を意味し、「概念化」は対峙関係を必要とする」と考えると、「概念化」するには前もって対象が「客体化」されていなければならないから、「客体化」されない「主観的表現」は「概念化」されない表現と解釈され時枝説と同じ結論となる。これに対して、「主観的表現」も「概念化」された認識の表現であると解釈する方法は3つある。

(1)主観的表現を概念化された表現とする3つの方法

第1の方法は、「概念化」において「対峙関係」は不要とし、「客体化」の定義を、「客体化」は認識の過程で対象が話者（主体）以外に存在であると認定する行為である」とするにとどめる方法である。

第2の方法は、「対峙関係」は、「客体化」以外の方法でも発生すると考え、定義の後半を「客体化」では「対峙関係」が発生する（しかし、「対峙関係」ができるのは、「客体化」だけとは限らない。）と修正する方法である。この場合は、「主観的表現」の概念化では、どのようにして「対峙関係」が生まれるかの説明が必要になる。

これに対して、第3の方法は、「対峙関係」を必要としない概念化が存在すると考える方法である。すなわち「主体」の感情や意志を対象とする場合、「話者」と「対象」との間に「対峙関係」が形成されていなくても、「概念化」は行えるとする方法である。確かに、話者が観念的自己分裂によって観念的話者を発生させるには、心理的なコストがかかると同時に時間的ロスが発生すると思われる。そこで、話者自身の感情や意志は主体の属性であるから、わざわざ話者と対象との「対峙関係」を発生させるような面倒なことをせず、直接的に表現できるような仕組みが発達したと考えるとそれなりに納得のいく説明となる。

(2)3つの方法の検討

そこで、2.1節の検討結果と矛盾しないようにするため、どの方法を採用するかについて検討する。

まず、話者自身の「客体化」で観念的話者が出現するのはなぜかを考えると、それは、話者と対象の間に「対峙関係」を発生させるためだと考えられるから、少なくとも「客体化」において「対峙関係」は必須である。よって、それを考えない第3の方法は否定される。

次に、「対峙関係」は話者が対象に立ち向かう関係にあることを意味するから、「主体」でない対象をひとまとまりのものとして意識した時点で発生すると考えられる。従って、このような意識を持つ行為を「客体化」と呼ぶ以上、「客体化」以外の方法で「対峙関係」が発生するとは言えない。これより、第2の方法も否定される。

以上から、本稿では、第3の方法を採用し、主観的表現によって表現される概念は、話者との「対峙関係」なしに、「概念化」された認識だと解釈する。これは以下のことを意味する。

話者は、対象を捉えるため、「客体化」によって対象と自分との間に「対峙関係」の認識を成立させるが、その結果得られた対象に関する認識を話者が自分とどのような関係で捉えたかを表すのが辞である。すなわち、辞で表現される認識は、対象と立ち向かって認識された内容に対する話者自身の認識で、それを表すための辞の選択は話者自身の内的な行為であるから、話者と対象との間の「対峙関係」はもはや必要性はない。それでも助詞、助動詞と言ったラベル（表象）への対応づけられるのは、対象の普遍的特徴が「概念」として認識されたと言うことである。

(3)本稿の解釈のまとめ

以上をまとめると、以下の通りである。

「客体化」は、対象が話者以外であることを認識し、話者と対象との間に「対峙関係」を発生させる行為であり、「対峙関係」は「客体化」によってのみ発生する。

「主体」の感情、意志等の場合は、「対峙関係」なしに「概念化」することができる¹⁾が、それ以外の対象を「概念化」する場合は、「対峙関係」を必要とするため、あらかじめ対象を「客体

化」することが必要である。

なお、言語では、このように「客体化」されずに「概念化」されることもあるが、逆に、「概念化」してから「客体化」することはない。

(4)客体化と実体化の違い

ところで、「客体」は「実体」と混同されやすい。しかし、本検討では、「対象」のうち「実体」、「属性」、「関係」は「客体」として「客体化」される対象には、名詞で表現される「実体」と「関係」の他、形容詞や動詞で表現される「属性」も含まれる。例えば、「属性概念 重い（形容詞）」、「帰る（動詞）」を「重さ（名詞）」、「帰る（名詞）」のように「名詞化」する場合である。そこで、「実体化」とは、本来は「実体」ではない「性質」、「状態」、「作用」を固定化して捉える行為であると定義する。

「実体化」も話者の認識の一形態である。「概念化」と「実体化」の関係について言えば、一度「概念化」された対象が「実体化」されることはある。例えば、一度、助詞、助動詞で表現された話者の感情や意志を、次の文では、「実体化」して表現されるような場合である。よって「実体化」も「概念化」の前提とはならない。

なお、本検討では「対象化」と言っ言葉は使用しない。もし使うとすると話者と区別するという意味の「客体化」と同一の意味である。

3.複合概念と言語表現

前節では、「言語表現を構成する「主観的表現」と「客観的表現」は、いずれも概念化された話者の認識を表すものとした。言い換えれば、言語表現を構成するすべての単語は、何らかの概念を表すと言うことである。

そこで、本章では、「概念を「単一概念」と「複合」概念に分類することを提案し、これらの二つの概念の相互関係と言語表現との関係について検討する。

3.1 単一概念と複合概念

(1)単一概念と複合概念の分類

話者の認識が形成される過程では、言語を手がかりに様々な概念化が行われる。得られた概念には、単語や複合語などの表象への結びつきを得たものや、そのような表象を得ることができず、いくつかの単語が複合した表現により、はじめて対応づけが可能となるような概念などがある。さらには、適切な言語表現が見当たらないような概念もある。

このうち、単語レベルの表象を得た概念は、通常、内部構造を意識することなく言語で表現されるが、それ以外の概念は、内部構造を意識した複合的な表現として、句や節に対応づける方法が模索される。前者は、認識の過程で運用上の自由を獲得し、新たな概念形成に使用されるのに対して、後者の概念は、自由を獲得する過程にあると見なすことができる。

本章では、話者の認識におけるこのような概念の表現過程に着目して、概念を「単一概念」と「複合概念」の二つに分類する。「単一概念」は、話者によってその内部構造が意識されない概念であり、「複合概念」は内部構造が意識される概念である。

(2)複合概念を単一概念化する仕組み

概念の表現過程から見て、「単一概念」と「複合概念」の分類は、相対的なものである。故事に基づく様々な概念などは、定義なしに使用され、単一の複合語として複雑な内容を表しているが、これは、一単語で表現できる程度に概念化が進んだためと言える。このように言語には、概念形成を支援する仕組みとして、「単一概念」から「複合概念」を形成する仕組みがあるが、これは、「複合概念」を「単一概念」に移行させるための仕組みでもある。

例えば、言語規範にない名詞や動詞などの述語の意味を文中で定義して使用する仕組みや、一度、個別的な概念を組み合わせて表現された内容を「それ」や「これ」と言った代名詞で表現する仕組みなどがある。日本語で、「…すること」等のように前文の内容を代名詞で参照する仕組みや、英語で、複雑な内容をthat節を使用して先行詞に縮退させる仕組みも同様である。

このような捉え直しによって概念の複合体を一つの概念にまとめ上げる仕組みは、自然言語だけでなく人工言語にも組み込まれている。プログラミング言語における「関数」や「手続き」の定義も、単

¹⁾ これは「辞」(主観的表現)の役割と矛盾しない解釈である。すなわち、対象を捉えるために、話者は「客体化」によって対象と自分との間に「対峙関係」の認識を成立させるが、その結果得られた対象に関する認識を話者が自分とどのような関係で捉えたかを表すのが辞だと言うことである。言い換えれば、辞で表現される認識は、対象と立ち向かって認識された内容に対する話者自身の認識で、それを表すための辞の選択は話者自身の内的な行為であるから、話者と対象との間の「対峙関係」はもはや必要性はないと言うことである。

一の概念」から「複合的な概念」を導く仕組みである。

例えば、三角関数の定義の中では、その計算手続きが意識され、手順を表現するための種々の単一概念が組み合わせて使用され、複合概念が形成される。しかし、一度定義された関数は自由を獲得し、その後、他の箇所では、単一のラベル名で参照され、いちいち計算手続きは意識されない。

(3)概念と言語表現の階層性

単語が言語規範として持つ概念の間には、包含関係(isa関係)から見た階層構造が存在することはよく知られている。上位概念の持つ性質が下位概念に伝搬される性質は、言語処理を行う上で大変重要である。

単一概念」と「複合概念」は、言語表現上から見た分類であり、相対的な関係にある。両者は共に階層関係を持つ。これは、「複合概念」に対応づけられた2単語以上の言語表現も階層構造を持つことを意味する¹⁾。

3.2 言語表現から見た意味の単位

(1)言語表現における意味の単位

すでに述べたように、言語表現上、「単一概念」は、単語レベルの表現で表され、「複合概念」は、句や節によって表される。一方、言語表現を構成する「生体的表現」と「客体的表現」は、いずれも「概念化」された話者の認識の表現であることから、言語表現は、話者の認識を「概念」を介して文字列もしくは音素列に結びつけたものである。

これは、概念化されていない認識は言語では表現されないと言うことであるが、逆に、概念化された認識は必ず言語表現で表されることを意味しない。また同様、言語表現のどの部分を取り出して、それが一つの概念を表していると言うことを意味しない。

そこで問題となるのは、言語表現のどの部分が概念を表しているかと言う問題である。概念を意味の単位(対象認識の単位)と考える以上、意味処理においては、言語表現の中からどの部分が概念に対応する「意味の単位」を構成しているかを調べ、概念と表現の対応関係を決める必要がある。

(2)入れ子構造から見た意味の単位

上記の問題は、意味的な約束との関係で考えることが必要である。言語表現の形成過程を見ると、「客体的表現」が「生体的表現」に包まれ、外側に向かって入れ子が形成されていく。句や節はこのようにして形成された意味の単位であるから概念の単位だと言うことができる。

また、文レベルの表現は、話者の判断や推論の内容を含む単位である。前述の南不二男によれば、表現は「描叙」、「判断」、「提出」、「表出」の四段階の内容を含むとされているが、これらも入れ子の内側を一つの単位として見ると、いずれの段階の表現も概念を表すものということができる。

3.3 概念形成の場としての話者の認識

(1)話者認識と概念の関係

ところで、「認識」と「概念」の関係であるが、「概念」は「認識」の形態である。絵画や音楽などの表現によって表されるような認識もあれば、言語によって表される認識もある。言語による認識は、超感性的な認識であり、社会規範としての言語規範(概念)を媒介に認識が形成されることに特徴がある。

話者の認識としての概念が形成される段階には、外界の対象が五感を通して捉えられた対象をイメージとして捉える段階があり、それらが、「客体化」された後、「実体」、「関係」、「属性」などのカテゴリー分類され、上り、下りの仮説と検証の過程を経てそれぞれの概念化が行われるレベルなど様々なレベルが存在する。

その中で、「概念」が形成されたと言うことは、対象の持つ特殊性を類としての普遍性の側面から捉えることに成功したと言うことであり、そのような普遍性の発見に至らず、イメージのレベルにとどまっているような認識も存在する。このように、概念の中で言語表現との結びつきに成功した概念が、はじめて、現実の言語で表現可能となる。但し、既に述べたように、概念化された認識がすべて言語表現で表現できるわけではない²⁾。

(2)概念と表象の問題

ところで、通常、「概念」と言えば、感性的なイメージ(音声、文字)に結びつけて考えることが多い。このことから、「認識」は「感性的なイメージ」に結びつくことができ初めて、「概念」としての地位を獲得するという考え方がある。

確かに、概念はそれを表すための物理的表象を獲得することで、より大きな自由を獲得し、認識は深まる。しかし、「認識」は、「音声」や「文字」と言った物理的表象と結びつかなければ「概念化」されないとするれば、目や耳の不自由な人は「概念」が獲得できず、言語が理解できないと言う重大な欠陥を持ち込むことになる。

4.言語表現の部分的線形性と文法規則

前章では、概念と言語表現との対応関係について検討した。本章では、概念の階層性と言語表現の部分的非線形性に着目する立場から、プラトンの問題として知られている「言語表現の無限性と人間の記憶容量の有限性」の問題を考察し、それを解決する手段としての文法規則の役割を論じる。

4.1 言語表現の非線形性

(1)言語表現の意味的非線形性

前章の検討によれば、言語表現は、いずれも概念化された話者の認識を表すもので、単語は単一概念を表し、複数単語からなる表現は複合概念を表す。

ところで、概念が「必要不可欠の要素を統一体として捉えた認識の単位」である以上、「複合概念」の構成要素も概念形成上不可欠の要素であることから、「複合概念」を個々の要素に分解することは不可能である。このことから、「複合概念」の表現である言語表現も非線形性だと言うことになる。

(2)言語表現の無限性と人間の記憶の有限性

そこで、言語表現の無限性と人間の記憶容量の有限性の問題について考える。

話者の認識として形成される概念の数は無限であり、それを表現する言語表現の数も無限である。このような言語を人間はいかに獲得して使いこなしているかと言う問題であるが、比較的限られた数の構成規則と表現要素からなる線形な言語表現なら、表現要素と構成規則を覚えれば使用できるようになる。

反対に、これらの言語表現がすべて意味的に非線形だとすると、分割した場合は元の意味が失われるから、それを使用するには、表現そのものをすべて覚えておく必要がある。しかし、人間の記憶容量は有限であるから、現実には、無限に存在する言語表現のすべてを覚えておくことは不可能である。

従って、言語表現のすべてが非線形で、要素合成法が完全に否定されるようであれば、言語表現の無限性から言ってその機械処理は完全に否定されてしまうだけでなく、人間がそれを獲得することも不可能となってしまう。

このような「有限の記憶容量の人間がいかにして無限と見られる言語表現を使いこなすことができるか」と言う問題は、古くから「プラトンの問題」として知られている。チョムスキーは、深層構造を導入し、深層構造からの変形操作によって無限の表層構造が生成されると説明した。

概念と言語表現の関係で見ると、チョムスキーに倣って、「1つの概念が複数の言語表現に対応する」と説明する方法も考えられるが、その方法は適切でない。すなわち、前章の議論によれば、厳密に言って、概念が対応する言語表現は1つである。さらに、話者の形成する概念は、言語表現の数以上に無限に存在することを考えると、現実の概念と言語表現の対応関係は、むしろ、1対1である(多義もここから発生する)と言えるからである。そこで、以下の節では、この問題に対して独自の解決を試みる。

4.2 言語表現の無限性と文法的規則

(1)文法規則による言語表現の部分的線形化

社会的慣習としての言語が歴史的に形成されてきた過程に着目して、「人間の記憶容量の有限性と言語表現の無限性」の問題について考える。

¹⁾ このことに着目すれば言語表現を意味類型化することが可能となる。意味類型化に基づく翻訳の原理については、5章で述べる。

²⁾ 言語上の表象を獲得できなければ、概念化できたとはいえないとする説もある。しかし、言語では、直喩「-のようなもの」や例示「-や-など」によって間接的に表現する方法があり、話者の認識の中に形成された概念を表すための直接的な表現が悪い浮かばないときなどに使用される。もし、話者の認識の中に概念が形成されていなければ、このような喩や例示はできないことから考えても、概念の中には、表象を得る前に形成されるものがあるだけでなく、形成された後も表象を獲得できないものも存在する。

まず、言語規範を「文法的な慣習」と「意味的な慣習」に分けて考えると、言語において社会的慣習としてまず重要なのは「意味的な慣習」である。これは、単語や表現とそれによって表される概念との対応関係に関する約束であり、「このことを表現するにはどのような表現が使えるか」、また、「この表現はどんなことを表すための表現か」といった社会的な約束の体系のことである。話者が認識の内容を表現するとき、この約束の体系に従ってどの単語や表現を使えばそれが表現できるかを考え、使用する約束を選択する。

人間が単純で有限な概念しか獲得しておらず、それがすべて、単一の単語で表現できる時代には、複数の単語から形成されるような表現は不要である。しかし、次第に、単語のレベルで表現できないような複雑な概念を獲得するようになると、複合的な表現が必要となる。

複合的な表現といえども、その数が人間が容易に記憶できる範囲にあれば、合成規則などは不要であり、単語と同様、記憶していれば済む話であるが、人間の認識に形成される概念は記憶しておくような数ではない。そこで、言語表現を部分的に線形化することによって、複合的で多彩な表現を有限の規則で形成するための仕組みとして「文法的な慣習」を発達させたと考えられる。

すなわち、文法規則は、言語が表現の無限性を有限の手段で実現させるために発達したもので、線形性と非線形性と言う2つの対立物を相互浸透させるための手段であると考えられる¹。

(2) 子供に見られる線形規則の獲得

同様のことが、子供の言語獲得の過程にも見られる。子供が初めに使用できるようになるのは単語である。たまたま複数単語の表現を使ったとしても、それは単語の組み合わせが固定した表現で、単語単位への分節化が意識されたものとは言えない。そのうち、表現内に複数の単語が存在することに気づき、その間に一定の結合規則（線形規則）が存在することに気がつく、表現能力は急速に拡大する。

しかし、むやみに線形規則を使用し、組み合わせてはいけなような表現を使い始めるが、「そんな言い方はしない」といった周辺からの指導によって、表現の背後に存在する個々の結合規則を習得していく。また、結合規則には線形規則だけでなく、この場合はこうとしか言えないような非線形な規則もあることに気がついていく。

以上から、言語表現の無限性は、言語表現の部分的線形性に基いて説明されるものであり、それを支える規範が文法規則だと言ふことになる²。

5. 意味類型と機械翻訳

前章までの検討では、複数単語からなる言語表現も概念（複合概念）を表すとした。本章では、これを前提に機械翻訳の方法について考える。

5.1 概念を介した翻訳の原理と概念スキーマ

(1) 概念を介した翻訳の原理

言語表現は概念を表現する手段であるという考えからみて、翻訳は、異なる言語の言語表現を概念を介して対応づけることだと言ふことができる。

既に述べたように、現実の言語には、それぞれ、規範としての概念体系があり、言語が異なれば、「概念」の体系も異なるが、現代社会では、ほぼ共通しているとみなせるような概念が沢山存在する。そのような共通的な概念は、言語上の表現の仕方（表現構造）は異なっても、その内部の論理的な構造は共通していると予想されるから、機械翻訳を実現するには、異なる言語の表現を言語間に共通した「概念」を介して対応づければよい。

例えば、「因果関係」、「比較」、「親子関係」などの概念が持つ論理的構造は言語間で共通しているから、これらの表現は概念を介して対応づければ良い。単語の意味的対応づけである訳語選択の課題は「単一概念」を介して解決が図られるのに対して統語構造を持つ句や節のレベルでの意味的対応づけの問題は「複合概念」を介した対応づけの方法によって解決が図られる。

(2) 概念スキーマによる共通概念の表現

さて、言語間に共通する「概念」が持つ独自の内部構造を認知言語学 [山梨 95, 00, Langacker 87] の「イメージスキーマ」に做って「概念スキーマ」と呼ぶ。「スキーマ」は、通常、「データの論理的構造を表現する枠組み」と言った意味で使用される言葉である。Minsky の「フレーム」や Shank の「スクリプト」などの「知識構造」と同様の形式 [ウインストン 79, オールウッド 79] で表現することも考えられるが、現実の言語表現で表現される概念は多種多様であるので、より汎用的な用語を使用する。

例えば、「因果関係」や「比較」等の名詞は、「単一概念」を表すが、これらの概念は、原因と結果や誰が、何を、何とどんな点で比較するかなどの内部構造を持つ。「概念スキーマ」は、このような内部構造を表す。また、「美しい山」、「流れの速い川」等の表現は、単一の単語で表すのが困難な概念を表しているが、これらも「概念スキーマ」を持つ。

ここで、「概念スキーマ」で表される概念の内部構造は、論理的なものであり、構造上、実際の言語表現にそのまま対応づけられるものではない点に注意を要する。

(3) 意味類型とその言語依存性

ところで、言語では、「複合概念」の構成要素は、言語それぞれの規範に従って順序づけられ、表現が形成されるが、どの言語でも、同一の「複合概念」に対して、多数の形式があり、認識の微妙な違いによってそれを使い分けている。

そこで、一つの方法（複合概念）を対応づけるための表現の形式を有田潤 [有田 87] に做って「意味類型」と呼ぶ。「意味類型」は、言語に依存しており、言語毎に意味類型の体系が存在する。

従って、対象言語間に共通する概念を表す「概念スキーマ」に対しても、言語毎に異なる「意味類型」が存在し、各「意味類型」には、それぞれの言語に依存した複数の表現形式が含まれることになる。

5.2 意味類型パターンとその非線形性

(1) 意味類型パターンによる複合概念の表現

ここで言語の表現形式をパターンで表現することを考え、同一の意味類型に属するパターンを「意味類型パターン」と呼ぶ。

これは、言語表現の構造を線形要素と非線形要素との組み合わせによって表現するものである。一般の表現では、非線形構造はネストするため、各レベルに応じた意味類型パターンの組み合わせで表現される。例えば非線形な表現が複数組み合わせられた表現を考えると、個々の表現は線形でも、個々の表現から見れば、非線形だと言ふことになる。このような表現全体は、全体の非線形構造を表すパターンと個々の非線形要素を表すパターンとの組み合わせで表現される。

この方法は、線形要素を取り込むことによって無限の表現のパリエーションを有限の数のパターンに縮退させる方法であり、対象の無限性と手段の有限性の矛盾を工学的に調和させる方法である。

(2) 意味類型パターンの例

「比較」の概念を表す表現の例として、「富士山は大山より高い」、「リンゴはバナナよりおいしい」の文を考えると、共通する表現形式（すなわち「意味類型」）として、「名詞Aは名詞Bより形容詞」の形式が取り出される。この場合、名詞A、名詞B、形容詞の部分を他の語に置き換えても、それらが「比較」の概念を表していることに変わりはない（但し、名詞A、Bは、比較対象となり得るものである必要がある）、しかし、字面部分の「は」「より」を他の語に置き換えた場合は、「比較」の概念を表さなくなる。

これに対して、慣用表現である「猿も木から落ちる」と弘法も筆の誤り」は、いずれも「上手な人も時には失敗する」と言う意味（複合概念）を表す。このような表現では、どの部分を置き換えても表される概念は変わってしまうため、表現全体が、「意味類型パターン」となる。

このように、「意味類型パターン」は、言語表現から、線形要素（置き換え可能な要素）を抽象化し、上位概念に縮退させたものと見ることができる。

¹ ここで注意しなければならないのは、言語表現の「部分的線形性」である。この複合概念と言語表現の関係から見ると、言語表現は元々非線形であるが、概念の階層的構造から見て、着目した複合概念にもさらに上位の複合概念が存在する。上位の概念から見たとき、その構造上の特徴が下位概念に継承されることから、下位概念は共通の構造を持つことになる。従って、上位概念に対応する言語表現は、線形要素のある表現形式を持つことになる。

² このことから、文法規則は意味的な約束に制約された範囲で有効であり、文法的規則から、表現の意味的な単位が取り出せることを仮定した従来の言語処理の方法は見直すことが必要である。

5.3 意味類型パターンの意義とメリット

(1)人間の思考形式としてのパターン

人間は、思いついた事柄について思考を巡らすとき、言語の枠組みを借りて、それはどんな言い方ができるかを考え、認識を形成しているように思われる¹。人間が考えをまとめるのに使用する枠組みが「意味を単位とした言語表現のパターン」だとすると、それは既に述べた「意味類型パターン」そのものと言うことになる²。

母国語以外の言語を獲得する際も同様である。外国語を話すとき句構造で考えるのではなく、どんなことを言うときはどんな言い方があるかをパターンとして習得するようになって初めて、スムーズな表現ができるようになると思われる。英会話の授業で行われるさまざまなパターンプラクティスからも、この考えが裏付けされる。

(2)多義解消から見たメリット

言語表現の意味解析の最大の問題は、単語や表現が意味的に多義であることである。このような多義は、判断知識を持った人間では余り問題とならないが、判断知識を持たない計算機では大きな問題となる。この問題を解決するには、単語や表現とそれが表す概念の「対応関係の知識」だけでなく、どんな環境下で使用されたらどの約束が使われたかを示す「意味的用法」に関する知識が不可欠である。

意味類型パターンは意味の単位に定義されるものであり、それらの中で意味的排他性を保証する仕組みとして、線形要素に対する値域が定義される。このことから、意味類型パターンは、各表現要素の「意味的用法」の知識を持つため、意味的な曖昧性解消の仕組みとしての効果が期待できる。

(3)意味類型パターン方式の意義

以上で述べてきたことから文型パターン方式の持つ意義をまとめると以下のようになる。

言語の意味的約束として、「語や表現の意味と用法との関係」を同時に表現でき、文法的曖昧性と意味的曖昧性の双方の解消に役立つと期待できる。

言語表現を構成する線形要素と非線形要素について、それらの組み合わせ方が表現できる。

人間の思考パターンに近い構造を持つと予想される。

5.4 意味類型を介した機械翻訳

(1)翻訳の可能性

各言語が持つ概念の集合³を表す「意味類型」とそれに対する表現形式の集合である「意味類型パターン」は、いずれも言語毎に異なる。日本語には日本語の「意味類型」と「意味類型パターン」があり、英語には、英語の「意味類型」と「意味類型パターン」がある。

機械翻訳において、無限とも言える多様性を持った言語表現を異なる言語間で意味的に正確に対応づけることは、技術的に困難である。しかし、概念の階層構造に着目して、一定の粒度以下の概念をグループ化すれば、「意味類型」と「意味類型パターン」の数は有限に押さえることができると考えられる。従って、これらを介して対象言語間の表現に対応づけることは、工学的に可能性があると期待される。

(2)言語表現構造の保守性と翻訳の役割

翻訳の可能性について言えば、既に述べたように、対象言語間で共通概念を表す表現が存在することが条件となる。

地域交流の進んだ現代社会では、共通する概念が増えてきた。ある言語が他言語から新しい概念を輸入する際、構造を意識せずに(単語で)表現できる概念は、新しい語彙規則を導入することで容易に表現能力を拡大することができるが、表現構造の変化を伴うような複合概念の場合は、簡単ではない。

これは、表現内容と表現媒体の持つ矛盾によるもので、多次元的な認識構造に対応づけるために発達してきた表現構造を変えようとすると、それまでの仕組みが壊れてしまうような大きな矛盾が発生するためである。このような、概念の社会的な共通化と表現構造の保守性に関する対立は、機械翻訳技術の重要性を示している。

6. あとがき

意味類型に基づいた翻訳方式の原理を明らかにするため、本稿では、従来の言語理論で問題となっていた主体的表現と概念化の問題を検討し、客体的表現と主体的表現は共に概念化された話者の認識を表すことを示した。またこのことから、従来、単語で表現されると考えられていた「概念」を句や文などで表現される「複合概念」にまで拡張し、概念を介した翻訳方式の原理を提案した。

その議論をまとめると以下のようになる。すなわち、言語表現は、概念化された話者の認識を表現する。すなわち、言語表現を構成するすべての単語は「単一概念」を表し、句、節などの表現は、「複合概念」を表す。また、意味的に非線形な構造を持った「意味的構造の単位」は、「複合概念」を表す。そこで、「複合概念」を表現するための表現形式の集合を「意味類型」と呼べば、同一の「複合概念」であっても、それを表現するための「意味類型」は、言語によって異なるから、機械翻訳は、原言語の「意味類型」を原言語と目的言語に共通する「複合概念」を介して目的言語の「意味類型」に対応づけることによって実現される。

参考文献

- [有田 87] 有田潤：「ドイツ語講座」南江堂(1987), pp.48-56
[池原他 87]池原悟,宮崎正弘,白井諭,林良彦:言語における話者の認識と多段翻訳方式,情報処理学会論文誌, Vol.28, No.12, pp.1269-1279 (1987)
[池原 01] 池原悟:自然言語処理の基本問題への挑戦,人工知能学会誌, Vol.16, No.3, pp.522-430 (2001)
[池原 02]池原悟:究極の翻訳方式の実現に向けて = 類推思考の原理に基づく翻訳方式 =, AAMT Journal, アジア太平洋機械翻訳協会, No.33, pp.1-7 (2002.3)
[池原他 02]池原悟,佐良木昌,宮崎正弘,池田尚志,新田義彦,白井諭,柴田 勝征:等価的類推思考の原理による機械翻訳方式,電子情報通信学会,思考と言語研究会, TL2002-34, pp.7-12, 2002
[市川 63] 市川亀久彌:創造的研究の方法論(増補版),三和書房, 1963
[ウインストン 79]P.H.ウインストン著,長尾真,白井良明訳:「人工知能」培風館,1979
[オールウッド 79]オールウッド,アンデソン,ダール著,公平,野家訳:「日常言語の論理学」産業図書,1979
[小池他 02]小池,小林,細川,山口共編:「日本語表現文型辞典」,朝倉書店,2002
[白井他 95]白井諭,池原悟,横尾昭男,木村淳子:階層的認識構造に着目した日本語従属説問の係り受け解析の方法とその精度,情報処理学会論文誌, Vol.36, No.10, pp.2353-2361 (1995)
[ジャマシイ 98]グループ・ジャマシイ編著:「日本語文型辞典」,くろしお出版,1998
[時枝 41]時枝誠記:国語学原論,岩波書店,1941
[友松他 96]友松,宮本,和栗:「どんな時どう使う日本語表現文型 500」,アルク,1996
[名柄 87]名柄監修:外国人のための日本語例文・問題シリーズ 全18巻,荒竹出版,1987
[中村 93]中村明:感情表現辞典,東京堂出版,1993
[中村 95a]中村明:感覚表現辞典,東京堂出版,1995
[中村 95b]中村明:比喩表現辞典,角川書店,1995
[髙島 94]髙島一郎:日英対照感情表現辞典,東京堂出版,1994
[藤本他 01]藤本敬史,表克次,池原悟,村上仁一:埋め込み文の日英翻訳方式について - 内と外の関係の判断方法 -, 情報処理学会第63回全国大会, 6Y-03-2-265-266(2001)
[益岡 93]益岡隆志:「日本語の条件表現」くろしお出版,1993
[益岡 97]益岡隆志:「複文」くろしお出版,1997
[三浦 67]三浦つとむ:「言語と認識の理論」第1~3巻,勁草書房(1967)
[三浦 76]三浦つとむ:「日本語はどんな言語か」講談社学術文庫(1976)
[南 93]南不二男:現代日本語文法の輪郭,大修館書店,1993
[山梨 95]山梨正明:認知文法論,羊ひつじ書房 1995
[山梨 00]山梨正明:認知言語学原理「くろしお出版 2000
[Langacker 87]Ronald W. ngacker: Foundations of Cognitive Grammar, Stanford University Press,1987

¹ 「人間の思考が何らかの形式に支えられていること」、また、その形式は「母国語の言語表現の持つ基本的な構造らしい」と言うことについてはさまざまな研究者が指摘している。

² 人間の思考形式が、句構造文法のような階層的な仕組みだとは思われない。なぜなら、句構造文法は、言語表現は意味的に線形であることを仮定しており、文法的規則を用いて表現を階層化しているが、階層の一つ一つが意味の単位とはみなせないからである。

³ 5. 節では、一つ概念(複合概念)に対応づけるための表現の形式を「意味類型」と定義したが、概念は無限に存在し、それが階層構造を持つことから、どの概念も下位概念を持つことになり、意味類型には複数の下位概念が対応する。