

等価的類推思考の原理による機械翻訳方式

池原 悟*¹ 佐良木 昌² 宮崎 正弘 池田 尚志
新田 義彦*⁵ 白井 諭⁶ 柴田 勝征

*1 鳥取大学工学部知能情報工学科 ikehara@ike.tottori-u.ac.jp *2 長崎純心大学 saraki@st.rim.or.jp

*3 新潟大学工学部情報工学科 miyazaki@ie.niigata-u.ac.jp *4 岐阜大学工学部応用情報学科 ikeda@info.gifu-u.ac.jp

*5 日本大学経済学部・理工学部 nitta@eco.nihon-u.ac.jp *6 NTT-アドバンステクノロジ株式会社 shirai@nlp.ntt-at.co.jp

*7 福岡大学理学部応用数学科 kshibata@vin.sm.fukuoka-u.ac.jp

要素合成法を基本とする従来の自然言語処理の限界を克服することを狙って、「意味類型論（セマンティック・タイプロジ）」と「類推思考」の2つの原理からなる「言語の意味的等価変換方式」を提案し、その有効性について検討した。「意味類型論」の観点では、「人間の対象把握作用には、思考形式とも言うべきある種のフレームワークが存在し、それが言語表現に反映される」とする考えに基づき、表現構造の意味的な単位化を行う。すなわち、言語表現から、人間の持つ対象把握作用の形式と見せせるような意味的に非線形な表現構造を「意味類型パターン」として取り出し、構造的な意味の単位とする。次に「類推思考」の観点からは、「等価的類推思考の原理」を言語に適用し、原言語の表現構造を言語共軛な概念項を介して、目的言語の表現構造に写像する。なお、表現内の線形要素は、従来の要素合成法によって翻訳され目的言語に組み込まれる。

キーワード：機械翻訳、意味類型、類推思考、非線形モデル、意味解析

Analogical Mapping Method for MT based on Semantic Typology

Satoru Ikehara*¹ Masashi Saraki Masahiro Miyazaki Naosi Ikeda
Yoshihiko Nitta*⁵ Satoshi Shirai Katsumasa Shibata

*1 Engineering, Faculty of Engineering, Tottori University *2 Nagasaki Junshin Catholic University

*3 Faculty of Engineering, Niigata University *4 Faculty of Engineering, Gifu University

*5 College of Economics, College of Science and Technology, Nihon University

*6 Advanced Technology Co. Ltd *7 Faculty of Science, Fukuoka University

This paper proposes an Analogical Mapping Method (AMM) for machine translation, in order to break through the limits of conventional technology of Natural Language Processing based on Compositional Semantics. This method is constructed from the following two concepts. One is conceptualized on the basis of the Theory of Semantic Typology (TST). The TST suggests that human cognition to the world is accompanied by an epistemological framework under the influence of a mother tongue, bringing a heuristic approach that semantically non-linear expressions, which are inherent in an individual language, are extracted from the huge corpus of the language and then patterned as a unit of a meaning pattern. Another is derived from the Analogical Mapping Theory (AMT). The authors have applied the AMT to natural language processing and then invented a new method. According to the method, a set of meaning patterns is established from a source language, and then mapped to a set of those from a target language through the common concept to both.

Key Words: Machine Translation, Semantic Typology, Analogical Mapping, Non-linear Model, Semantic Analysis

1. はじめに

自然言語処理では、機械翻訳システムを中心に過去多大な研究投資が行われ、大きく発展してきた[長尾 89, 99]が、日英両言語のように言語族の異なる言語間の翻訳は難しく[中村 83, 長尾 86]、訳文品質の向上は依然として重要な課題となっている。従来の中心的課題であった形態素解析や構文解析の技術は、すでに、限界に近いレベルに達しており、今後の発展を図るには、パラダイムシフトを伴うような新技術の実現が期待される。

ところで、従来、機械翻訳システムで研究開発されてきた翻訳方式は、ほとんどがトランスファー方式である。この方式は、構文解析結果を用いて原言語の表現構造を目的言語の構造に変換したのち、原言語に使用された単語や句を目的言語の訳語に置き換えることを基本としている。文の構造は、意味と独立に扱われることから、意味の線形性を仮定した要素合成法(重ね合わせの原理)を基本とした方式と言える。

しかし、実際の言語表現では線形性は保証されないことが多く、この方式は文を単語に分解する過程で表現全体の意味が失われ易いことが問題である。言語表現の非線形性は、言語の本質的な特徴[溝口 96, 池原 01]でもあり、要素

合成法の仮定は従来の翻訳方式の最大の問題とすることができる。

この問題に対して、表現の構造と意味の関係を考慮した方法として、「多段翻訳方式」[池原 87]が提案されている。この方式では、表現の構造と意味の問題に迫るには、言語間での発想の違いを考慮すべきであるとして、話者の概念化された認識を表す「客体的表現」と概念化されない感情や意志を表す「主体的表現」を別々に変換するパスを設けている。客体的表現の翻訳では、動詞と名詞の結合した基本的な言語表現に対して、構造的な意味の単位を設け、表現構造の持つ意味を失わないように翻訳しており、要素合成法の限界に挑戦した方法の一つとすることができる。

ところで、この方式を実現するには、対象認識の形式に関する精密で膨大な規模の言語知識ベースを開発する必要がある。すでにその第1歩として、単文レベル(述部が一つの文)の表現を対象に言語知識ベース[池原 97]が開発されている。これを用いた翻訳実験では、IPAL辞書[IPAL87]の全例文約 5,000 文に対して 90%の精度で訳語が正しく決定されたことが報告されている[金出地 01]。また、誤りの原因の大半は、必要な結合パターン不足、形態素解析または構文解析の失敗、結合パターンとの照合の失敗であることから見て、単文レベルでの意味解析の問題は、方式

的に、ほぼ解決できる見通しである。

しかし、この方法は以下の問題点を持つ。一つは意味の単位と見なされる原言語の表現構造に対して、固定的に単一の目的言語表現が対応づけられるため、必ずしも適切な表現に翻訳されるとは限らないことである。もう一つは複文（埋め込み節を持つ文）や重文（接続構造を持つ文）では、単文相当の表現間の意味的関係の非線形性が問題となることである [Ikehara 01]。

そこで本研究では、上記で提案された「構造上の意味の単位」の概念を意味類型論（セマンティック・タイポロジー）[有田 87]の観点から発展させるとともに、「等価的類推思考の原理」[市川 63]を言語に適用した新しい翻訳方式を提案する。

2. 言語の個別性と表現の意味類型化

言語翻訳において問題となるのは言語表現とそれによって表される内容の関係である。本章では、言語によって異なるこの関係を意味的に類型化することを考える。

2.1 対象認識と言語の関係

(1) 対象認識と言語の個別性

言語は話者の認識を表現する手段としての慣習である。各民族の有する自然観および社会観の違いから認識の違いが生じ、これが表現の違いとなっている[渡辺 73, 池上 81]。例えば、緑なす平原に住む民族では自然との一体性を基調とする多神教の自然観が生まれ、諸物の中に調和して存在する対象の姿を描写する表現や実体より属性をクローズアップするような表現が発達し易いものに対して、砂漠に住む民族では人間と自然との戦いを基調とする一神教の自然観が生まれ、主体と客体の区別を明確にし、その関係を立体的に表すための表現が発達することが指摘されている[外山 92]。

このような民族による自然観と社会構造の違いは言語発生以前から営々と蓄積され、世代から世代へと受け継がれており、それらの構造の上に発達した言語は、必然的に個別的とならざるを得ない。

ところで、言語は、いったん、それが社会慣習として定着すると、対象認識の枠組みとしても使用され[林 60]それを母語とする話者の対象認識の形成能力を豊かにすると同時に対象認識の制約条件としても働くようになる。

このように、異なった認識を表現する手段として成長してきた言語は、再び、それを使用する人の認識の制約条件として働くことから、認識は言語集団において2重の意味で個別的なものと言える。

(2) 対象認識の共通性と言語表現の保守性

ところで、どの民族も同じ地球の重力の下で生活する人間である[レイコフ 86]、居住する地域と環境が異なっても、最低限、衣食住によって生活を営む存在であることから、異なる言語間においても対象認識の仕方には基本的な部分で多くの共通点が存在する。

この認識の共通性は、民族が文化的な共通基盤を持つほど大きくなる。例えば、漢字文化圏である中国と日本とは同時に仏教文化圏でもあり、社会制度にも歴史的に近いものが存在した。また、近代以降、西洋科学技術が普遍化し、異なる言語集団でも物質的对象世界に対して、共通する認識を持つようになってきた。社会に対する認識においても、文化的交流の進展によって、言語間で共有する認識の範囲は増大している。

ところで、このような言語間での対象認識の共通化は、必ずしも言語間での表現構造の共通化を意味しない。言語

の表現構造は、文化交流による対象認識の共通化に追従できるほど柔軟ではない。

言語は常に矛盾を内包した過渡的な存在で、時と共に変化する。語彙のレベルでは、造語や外来語の導入によって、概念化された新しい認識を比較的容易に取り込むことができるが、表現構造の場合は部分的な変更であっても多くの場合、体系自体に矛盾が生じる。

その理由は、言語が多次元的な対象の認識を一次元的な文字の列、または、空気の粗密波で表現せざるを得ないという宿命をもつことである。言語は、立体的構造を持った対象認識を一次元的な文字列上に模写する方法として、それぞれ独自の表現体系を発達させてきた。そのため、新しい概念を輸入しようとして、他の言語の表現構造をそのまま借りてくると、せっかく構築された枠組みに矛盾が生じてしまう。

以上のように、時代と共に言語間で認識の共通化が進むのに対して、それを表すための表現構造は頑固で保守的であることは、機械翻訳方式を考える際の第1の検討条件といえる。

2.2 言語表現の意味的非線形性

言語は、他の自然現象や社会現象と同様、複雑系の特徴を持つことが指摘できる[池原 01]。

その第1の点は、自己組織化された表現規則の存在である。言語規範が自然発生的な社会規範の一つであると言うことは、それが人間同士の相互作用を介して自己組織化されたことを意味する。現存する言語族間での表現構造の大きな違いは、発生段階での小さな違いから成長してきたものと考えられるが、これも複雑系の特徴の一つと見ることができる。

第2は、表現と意味の関係の非線形性である。これは、言語が非平衡な開放系として進化の途上にあることと関係している。言語は、常に既存の枠組みで表現できないような新しい経験と認識にさらされており、そのような場合、比喩[山梨 88]や類推の仕組みなどが使用されるが、同じ比喩や類推の表現が繰り返されると、それが習慣となり、新しい表現規則が成長する。このような規則は既存の枠組みと調和しにくい場合が多いこと、また、その場合は、非線形な構造として取り込まれることにより、言語表現の意味的な非線形性は、拡大再生産されることになる。

このような表現の意味に関する非線形性は言語の本質で、これを扱うための枠組みを考案することが新しい翻訳方式に課せられた第2の検討条件と言える。

2.3 言語表現の「意味類型」

本節では、前節で示した2つの条件に適した方式を実現するための「意味類型」なる概念について述べる。

(1) 言語表現の形式と意味類型

各言語の語によって表される概念は、単なる名辞ではなく、話者が母語をもって対象を把握するための網の目だと言われている。これに着目すると、翻訳は、原言語の網の目（概念）で掬い取って表現された話者の認識を目的言語の網の目で掬い直し、改めて表現することだと言える。

従って、翻訳が可能であるためには、対象とする言語間で網の目を共有するか、もしくは類似の網の目が存在することが必要となる。網の目が共通しない場合は、より低位の概念を組み合わせることによって対象概念を説明することになる。

ところで、網の目となる「概念」は、特殊性と普遍性の統一体として、それ以上分割できない認識の単位だとされている[三浦 67]が、その内部は必ずしも単純ではなく、

様々な構造を持つ。例えば、「山」や「川」のような実体概念は、そのまま、それを象表する単語に対応づけられるのに対して、「因果関係」のような関係概念は、原因と結果になる事象の関係が様々な形式を持った表現に対応づけられる。また、「比較」の概念では、「誰が」、「何を」、「何と」、「どんな点で」比較するかなどの要素が各言語の表現に対応づけられる。

自然言語において、このような概念を表すために使用される形式は多彩である。どの言語にも多数の形式があり、認識の微妙な違いによってそれを使い分けている。

ところで、従来、人間の対象認識では、概念形成の過程で何らかの形式が伴うことについて、また、それは母語の表現の枠組みであろうとすることについて種々の指摘が行われてきた。ヘーゲルは、「直感的に統括される類型」として、概念把握で使用される形式の存在を示唆している[ヘーゲル 98,87]。また、関口文法は、「意味形態論」において「世界の把握作用が形式を伴う」としている[関口 80]。これに対して、時枝誠記は、対象認識のあり方が言語表現の構造に反映することに着目して言語過程説を提唱した[時枝 41]。ドイツ語学者の有田潤は、「意味類型は具体的言語表現の一段奥にある思考形式のごときもの」として「意味類型」を提案し、「その性質上、各自の母語をもってこれを考え、理解し、また整理することができる」としている[有田 87]。また、さらに、「自動翻訳機がもし本当に自然言語を翻訳しようになれば、その原理もおそらくここに求められるに違いない」とも言っている。

以下では、有田潤に倣って、概念を対応づけるための表現の形式を「意味類型」と呼ぶことにする。

(2) 非線形な表現要素としての意味類型

「意味類型」は、言語表現をその表す概念を変えない範囲で抽象化したものと考えることができる。すなわち、言語表現から線形要素（置き換え可能な要素）を捨て、非線形な部分を取り出したものと見ることができる。そこで、本研究では、「意味類型」を文法的、意味的な制約付きの変数と字面からなるパターンで記述することにする。

ところで、このような「意味類型」は言語に個別的である。日本語には日本語の「意味類型」があり、英語には、英語の「意味類型」が存在する。一般に、表現は無限に存在するが、人間の記憶能力からみて、それを生成するための言語規範の数は有限と考えられる点から見て「意味類型」の数も有限と考えられる。一般に、異なる言語間の個別的で多様な表現を意味的に対応づけることは、工学的に困難であるが、有限の数の「意味類型」を対応づけることは可能性がある。

3. 言語表現の意味的等価変換方式

前章では言語表現の意味類型化に至る考え方について述べた。本章では「等価的類推思考の原理」[市川 63]を応用した「言語の意味的等価変換」の概要について説明する。

3.1 意味的等価変換方式の原理

(1) 等価的類推思考の原理

かつて、人工知能の分野では、人間の知能の本質は、汎用的な思考規則とそれを用いた推論能力にあると考えられていたのに対して、1980年代以降、類推能力の役割が注目されるようになってきた[鈴木 96]。類推は、あるものを他の似たものに例えることによって思考する形式を持っている。一般的な規則を用いない点に特徴があり、「覚える」、「思い出す」、「一般化する」、「特殊化する」等の機能が有機的に結合したものとすることができる。

ところで、市川亀久彌は、「創造的研究」の中で、「人間の独創性は類推思考の能力にある」として、「等価的類推思考の原理」を提案した[市川 63]。これは創造的問題解決プロセスのモデルと言えるもので、式(1)に示すように、「異なる系での異なる現象が、ある一定の条件Cの下で共通点を持つこと」を前提としている。

$$C(A = B) \quad (1)$$

但し、「C」は条件、「」は共通見地、「A」は系上の事象A、「B」系上の事象Bを表す。

ここで、類推とは、ある系上の事象A（ソース）が与えられたとき、条件Cの下でなる共通性を持った事象系上の事象B（ターゲット）を思い浮かべることである。

ところで、翻訳では原言語表現の意味を理解したら、それを表すのに適した目的言語の表現を思い浮かべ、もっとも適切な表現を選ぶことが行われる。従って、翻訳も類推であり、式(1)を前提としていると言える。

式(1)を言語に適用した場合、その意味は以下の通りである。すなわち、言語上の表現Aが言語上の表現Bに翻訳できるには、言語上の表現の中に、同じ概念を表すような表現Bが存在しなければならない。これは、異なる言語間での翻訳が成り立つための条件で、を言語間で共通する概念と考えれば、前章で述べた考えと同じである。

(2) 意味的等価変換方式の原理

すでに述べたように、異なる言語間で、無限大の多様性を持つ表現を直接、意味的に対応づけることは工学的に困難である。そこで、これを有限の数の対応関係に縮退させる方法として、前章で定義した「意味類型」を使用した等価変換方式を提案する。本方式の原理を式(2)に示す。

$$A \quad C(A) \quad C(B) \quad B \quad (2)$$

但し、「」は写像、「」は「論理の意味範疇」つまり、「概念(真理項)」の集合、「C」は言語表現を意味類型化するための関数を表す。

式(2)は、の場合には翻訳方式となり、=の場合には同一言語内での言い換え方式となる。また、この式は「意味類型」間での変換、すなわち言語表現の非線形部分の変換を示しているが、線形な言語表現への適用することも何ら差し支えはない。

等価変換方式の処理ブロック図を図1に示す。図1では、非線形な言語の表現構造を対象とした「意味類型知識ベース」による変換パスと併せて、従来の語彙変換と同様の方法による線形要素（代入可能な要素として変数で表現される）の変換のパスを示している。

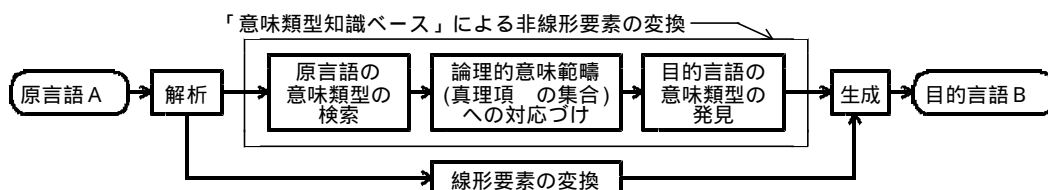
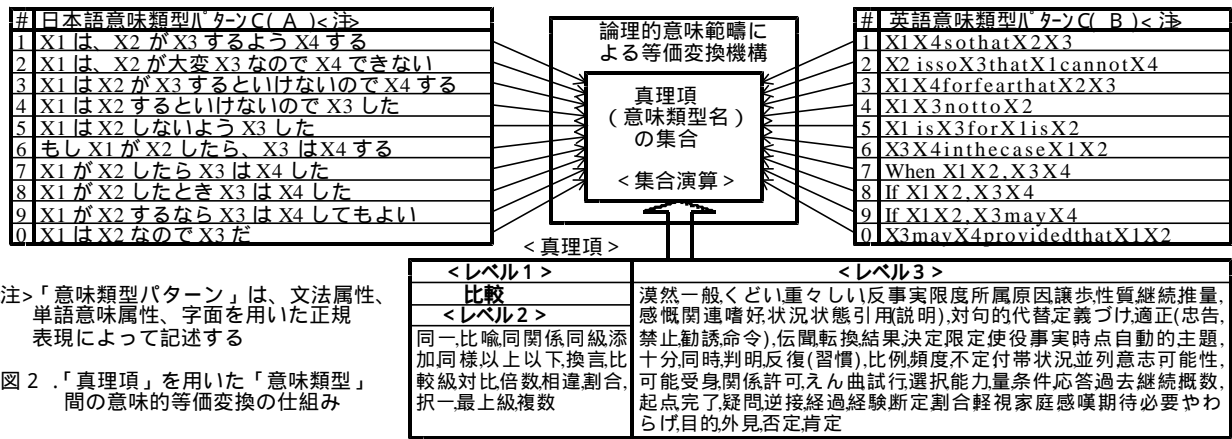


図1. 言語の等価変換方式の構成



(3) 「意味類型」と「論理的意味範疇」について

上記の方式で、異なる言語の「意味類型」の写像を媒介するのは「論理的意味範疇」の仕組みである。「論理的意味範疇」は、それぞれの「意味類型」によって表現される「概念」の集合である。一つの「意味類型」が複数の個別的概念を組み合わせた複合的な概念に対応する場合もあるので、ここでは、概念の真偽値によって複合的な概念を表現することを考えて、個々の概念を「真理項」と呼ぶ。

3.2 意味類型間の写像の方法

(1) 言語間のパターンの対応関係

原言語の「意味類型パターン」とそれに意味的に対応する目的言語の「意味類型パターン」は、あらかじめ両言語に共通の「論理的意味範疇」の「真理項」毎に集められ、グループ化される。これによって、両言語の「意味類型パターン」は、意味的に同等なグループ毎に対応づけられる。但し、複合的な概念に対応する「意味類型パターン」は、必要に応じて複数のグループのメンバーとする。

以上により、翻訳では、原文解析によって得られた原言語の「意味類型パターン」に対応するn個の目的言語のパターンの中から、最も適切なパターンが選択される。次に、その方法を述べる。

(2) 静的選択と動的選択

最適なパターンは原則として動的に選択する。そのため、意味類型知識データベースには、パターンを選択するための文脈上の条件を記載し、翻訳実行時に得られた情報によって、適切なパターンが選択できるようにする。しかし、動的に決定できない時は、あらかじめ決定しておいたデフォルトパターンを選択する。

動的選択で使用する情報は以下のようなものである。

文内情報：副詞により補足された意味、形容詞修飾による名詞の意味制約など、パターン要素に含まれない文要素の情報や疑問文、命令文、否定文、受動態文など文型に関する情報

文間情報：文脈による代名詞の意味や接続詞で結合された文の情報など

文外情報：フォーマル/インフォーマルの区別、書き言葉、話し言葉の区別、場面情報など

これらの情報は、従来の機械翻訳でゼロ代名詞の補完処理で使用されている情報とほぼ同様である[村木 89、中岩 93、堂坂 94、村田 96]

3.3 複文・重文翻訳への適用例

日英機械翻訳では、「日本語語彙大系」を用いることにより、単文レベルでの訳文品質は大幅に向上してきたのに対して、複文と重文の翻訳精度の向上は、緊急の課題である。そこで、本方式を複文と重文の翻訳に適用する例を図3に

示す。

図3では、入力された日本文は、非線形要素を取り出すため、「意味類型パターン辞書」と照合され、文中で使用された「意味類型パターン」が抽出される。その際、パターンに使用された変数に対応する日本文中の表現(「単語」、「句」、「節」)は線形要素リストとして保存される。

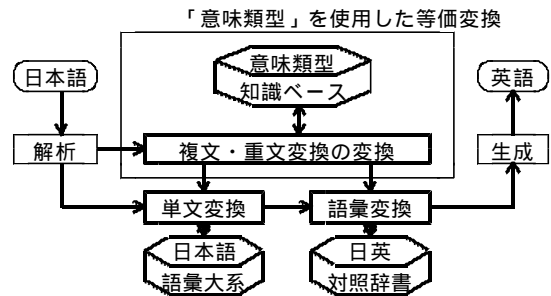


図3. 言語の意味的等価変換システム

次に、得られたパターンは、「論理的意味範疇」による等価変換機構を通じて英語の「意味類型パターン」に写像される。最後に、英語の「意味類型パターン」から英文が生成される。このとき、線形要素リストに保存されていた日本語表現は、既存の翻訳機構によって翻訳され、英語「意味類型パターン」の該当する変数の値として使用される。なお、複数の意味類型パターンを組み合わせた表現の場合は、より大きな構造から順に、本方式が繰り返し適用される。

4. 意味類型知識ベースの構築

前章では、意味類型を用いた等価的言語変換の原理とそれによる機械翻訳方式の構成概念を述べた。この方式を実現するには、(1)原言語と目的言語に対して網羅性の高い意味類型パターン集を構築すること、および、(2)各パターン間を意味的等価性に基づいてマッピングするための論理的意味範疇を体系化することが必要である。そこで本章では、その方法について述べる。

4.1 網羅性の高い文型パターンの収集

本研究では大量の対訳文をパターン化して汎化すること、また、得られたパターンの包含関係を調べ、同形式のパターンを縮退させることによって網羅性の高い日英文型パターン集を作成する。得られたパターン集の被覆率を向上させるには、汎化の方法が重要である。そこで対訳文から以下に示す3段階の版化を行う。

(1) 単語レベルの汎化

対訳文の要素のうち、線形要素と判断される自立語をN(名詞)、V(動詞)、AJ(形容詞)などの変数に置き換え

る。線形要素とは、対訳関係が決定できる要素を言う。なお、活用のある用言は、原型を変数とする。また、線形要素と見なすことのできる時制、相、様相に関する情報は、形式指定関数によって表現するほか、線形要素としては、省略可能な要素や選択の使用の可能な要素も指定する。

< 文型パターン化の例 >

和文 P : #1 [N1 は] /V2 て /N3 を /N4 に /tekita (V5)。
 英文 P : It was so AJ (V2) as to V5 poss (N1) N3 at N4.
 和文) うっかりして定期券を家に忘れてきた。
 英文) It was so care less as to leave my season ticket at home.

(2) 句レベルの汎化

上記で得られた文型パターンを対象に、線形要素と見られる名詞句を NP、動詞句を VP に置き換える。但し、変数 VP の述部は原型とし、それに付随する助詞、助動詞の情報は関数を使用して表現する。

< 句レベルの汎化の例 >

和文 P : VP1 て /#1 [N2 の] /N3 は、past (VP4)。
 英文 P : When N2 past (VP1), #1 [poss (N2)] N3 past (VP4).
 和文) 合格の知らせを聞いて彼女の顔は明るくなった。
 英文) When she heard she had passed the examination her face brightened up.

(3) 節レベルの汎化

上記で得られた文型パターンを対象に、線形要素と見られる節を変数 CL に書き換える。

< 節レベルの汎化の例 >

和文 P : CL1 ので、N1 にあたっては /must (VP2)
 英文 P : so + that (CL1, passive (must (VP2))) with poss (subj (CL1) N1)
 和文) これは極めて有毒であるので、使用に当たっては十二分に注意しなくてはならない。
 英文) It is significantly toxic so that great caution must be taken with its use.

4.2 論理的意味範疇を記述する体系

個々の文型パターンの属す意味類型が、論理的意味範疇であり、真理項の組によって表現される。すでに述べたように真理項は、各文型パターンの表す概念であるから、この概念をどのように体系化するかが問題である。ここでは、文型パターンが表す概念は多面的であること、また、重文、複文では、内部に含まれる単位文の概念から 1 文全体の概念が構成されることに着目して、真理項は「複合概念」と「単体概念」の 2 種類で構成する。

表 1 . 「複合的概念」の真理項

#	大分類	中分類
1	時間	期間、合間、時節、同時、毎時、以前、以後、以内、過程、契機、継起、など
2	空間	範囲、大きさ、方向、距離、高さ、など
3	論理	以外、同一、近似、差異、仮定、代替、基準、対照、原因、付加、結果、理由、譲歩、条件、根拠、手段、程度、比例、様態、範囲、など
4	心理	意図、目的、受動、動機、感情、など

表 2 . 「単体的概念」の真理項

#	大分類	中分類
1	状態	抽象的關係 存在、属性、所有、相対關係、因果關係
2		精神的關係 知覚状態、感情状態、思考状態、心的状態、身体状態
3		自然現象 - - -
4	動作	場所变化 物理的移動、所有的移動
5		状態变化 属性变化、身体变化
6		身体的行動 身体動作、利用、統合動作
7		精神的行動 精神的移動、知覚動作、感情動作、思考動作
8	結果	結果、生成、消滅・破壊

「複合的概念」と「単体的概念」の例をそれぞれ、表 1、

表 2 に示す。このうち、表 1 は、言語の表現対象は、「時間」、「空間」、「論理」から構成される考え方 [有田 87] に「話者自身」を加え、全体を 4 種類に分類している。

また、表 2 の分類は、「日本語語彙大系」で定義された動詞 (結合個パターン) の意味属性分類であり、複文、重文では、これらを組み合わせて使用する。

5. 新方式の特徴と限界

5.1 新方式の特徴と有効性

(1) 意味処理基本方式としての意義

従来、「計算機が、与えられた表現の意味を理解したかどうか」を検証する方法として、「意味を変えないで言い換えができるか」、もしくは、「与えられた複数表現の意味的同等性が判断できるか」のテストを用いる方法が考えられている。

これに対して、異なる言語間の翻訳機能と同一言語内での言い換え機能の実現を目指した本方式は、まさに、これらの質問に答えるためのものであり、意味処理の基本技術と言える。

(2) 柔軟で適切な翻訳の可能性

厳密に言って、100 % 正確な翻訳は存在しない。翻訳は近似であり、近似精度をいかに上げるかが問題である。この点から見ると、本方式は、原言語と目的言語の概念把握の共通性の程度に応じた近似度での翻訳を可能としている。

ところで、一般に、原言語の意味に相当する意味を持つ目的言語の表現は複数存在する。従来翻訳方式では、同一の表現は同一の変換パスを通り、同一の翻訳結果を生成するのが基本であったが、本方式は、目的言語の「意味類型」の中から、文脈に応じて適切な表現形式を選択することを基本としており、柔軟な翻訳方式と言える。

(3) 従来翻訳方式との違い

従来トランスファー方式が、言語表現の意味の線形性を仮定した方法を基本としているのに対して、本方式は、意味的に非線形な表現の形式を「意味類型」として抽出し、それを目的言語に対応づける (Mapping であり、Transfer ではない) ことを基本としている。すなわち、要素合成法を超えた方式である。

また、本方式は、言語間に「論理的意味範疇」を設ける点で、一見して、従来インターリング方式と類似している。しかし、インターリング方式では、言語共通の意味を表現するためのピボットを設け、原言語表現の意味のすべてをピボットを介して目的言語に翻訳しようとしていたのに対して、本方式の「論理的意味範疇」は、類型化された概念レベルで両言語の表現構造を対応づけるためのものであり、表現の持つ詳細な情報を変換するためのパスは別に存在する。

5.2 新方式の限界について

本方式は、原理的に「意味類型」と「類推思考」の考えから成り立っている。単純に言えば、翻訳は、原言語の表現と目的言語の表現を意味的に対応づけることである。しかし、計算機の場合、原言語、目的言語共に無限に存在する表現を意味的に対応づけるのは工学的に困難である。そこで、本方式では、言語表現を意味類型化することによって、無限の対象を有限な領域に持ち込むことと、類推原理によって意味類型間の対応づけることで工学的な解決を図っている。

このような工学的調和の方法には必然的に限界が生じる。本方式の場合、あらかじめ言語表現に使用される対象認識の形式を網羅的に収集し体系化することが必要である。従

って、適用対象は言語規範に直接媒介された表現、すなわち「直接的言語表現」に限定されます。比喩（直喩を除く）などの間接的な言語表現への適用はできない。

しかし、比喩理解の問題も、「類推思考」考えが手がかりとなりそうである。すなわち、比喩が認知される枠組みは、「例えるもの」、「例えられるもの」、「例えの根拠」の3つの基本要素から構成されるが、これらの要素の間には、「類推思考の前提」として3章で示した(1)式と同じ関係が成り立つ。この場合、比喩理解は、与えられた同一言語の表現AとBに対して、共通見地を求めるとなる。

本方式で使用する「意味類型知識ベース」が表現と意味に関する言語規範を体系化したものであるのに対して、比喩理解では、表現対象そのものに関する世界知識を必要とする。しかし、比喩で使用される世界知識は比較的限定されたものかもしれない。

6. むすび

「等価的類推思考の原理」に基づく機械翻訳方式を提案した。この方式は、数学的に見れば、言語表現を非線形領域と線形領域を構造的に分離すること、また、非線形領域は、分解不能部分として文型パターン化することによって全体を線形近似に持ち込むもので、「局所的写像を介して大局写像を得る方法」と言える。また、目的言語の複数の表現構造の中から、原言語の表現の意味を表すのに最も適した表現構造を類推し、その構造を使った訳文を作成しようとする点では、人間の翻訳の過程をシミュレートした方法である。

いずれにしても、この方式は、まだアイデアの段階である。実現する過程では、多くの難問が予想され、実際に実現可能かどうかは、これからの課題である。

< 謝辞 >

この研究は、科学技術振興事業団（JST）の戦略的基礎研究事業（CREST）の支援と科学研究費補助金 基盤研究（B）（課題番号 13480091）を受けています。

< 参考文献 >

- [有田 87] 有田潤：「ドイツ語講座」南江堂（1987）、pp.48-56
- [堂坂 94] 堂坂浩二：語用論的条件の解釈に基づく日本語ゼロ代名詞の指示対象同定、情報処理学会論文誌、Vol.35、No.5、pp.768-778（1994）
- [林 60] 林語堂：「開明英文法 = 表現の科学 =」、文建書房、（1960）pp.25-32（第1章表現の科学）
- [ヘーゲル 87] ヘーゲル、榎山欽四郎訳：「エンチュクロペディ」河出書房新社（1987）
- [ヘーゲル 98] ヘーゲル、長谷川宏訳：「精神現象学」作品社（1998）
- [IPAL87] IPAL：「計算機用日本語基本動詞辞書 IPAL」情報処理振興事業協会（1987）
- [池上 81] 池上嘉彦：「「する」と「なる」の言語学」、大修館書店（1981）
- [池原 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 97] 池原悟、宮崎正弘、白井諭、横尾昭男、中岩浩巳、小倉健太郎、大山芳史、林良彦：「日本語語彙大系」岩波書店（1997）
- [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）
- [Ikehara 01] Satoru Ikehara: Meaning Comprehension Using Semantic Patterns in a Large Scale Knowledge-Base, Proceedings of the PACLING'01, pp.26-35（2001）
- [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
- [金出地 01] 金出地真人、池原悟、村上仁一：結合価文法による動詞の訳語選択能力の評価、情報処理学会第63回全国大会、6Y-04、2-267-268（2001）
- [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）
- [村木 89] 村木一至：文脈理解—文脈処理の効果情報処理、Vol.30、No.10、1207-1215（1989）
- [村田 96] 村田真樹、長尾真：名詞の指示性を利用した日本語文章における名詞の指示対象の推定、自然言語処理、Vol.3、No.1、pp.67-81（1996）
- [溝口 96] 溝口理一郎：A I マップ - 形式と内容 - 内容指向人工知能研究の進め -、人工知能学会誌、Vol.10、No.1、pp.50-59（1996）
- [長尾 96] 長尾真：「自然言語処理」岩波書店（1996）
- [長尾 98] 長尾真、黒橋禎夫、佐藤理史、池原悟、中野洋：岩波講座「言語の科学」第9巻「言語情報処理」、岩波書店（1998）
- [中村 83] 中村保男：「翻訳はどこまで可能か」ジャパントイムズ（1983）
- [成田 94] 成田一：「こうすれば使える機械翻訳」、バベル・プレス社（1994）
- [中岩 92] 中岩浩巳、池原悟：日英翻訳システムにおける用言意味属性を用いたゼロ代名詞照応解析、情報処理論文誌、Vol.34、No.8、pp.1705-1715（1993）
- [鈴木 96] 鈴木宏明：「類推と思考」共立出版 1996
- [レイコフ 86] レイコフ著、渡辺昇一、楠瀬淳三、下谷和幸訳：「レトリックと人生」大修館書店（1986）
- [関口 80] 関口 存男：「新ドイツ語文法教程 第4版」三省堂（1980）
- [外山 92] 外山滋比古：「英語の発想・日本語の発想」日本放送出版協会（1992）
- [時枝 41] 時枝誠記「国語学原論」岩波書店（1941）
- [渡部 73] 渡部昇一：「言語と民族の起源について」大修館書店（1973）
- [山梨 88] 山梨正明：「比喩と理解」東京大学出版会（1988）