

解 説**自然言語処理技術の応用****4. 機械翻訳における自然言語処理†**

江 原 晴 将‡ 田 中 穂 積††

1. はじめに

機械翻訳の研究は、計算機の出現と同時、あるいは、それ以前から始められ、本誌でも何回か取り上げられた。初出は、創刊号にまで、さかのばる¹⁾。特に、1985年10月号では、機械翻訳の特集が組まれており、詳細な解説がある。ちょうど、Mu プロジェクト^{*}が最終段階を迎えていたところである。このプロジェクトは、現在、市販されている日本の商用機械翻訳システムに、大きな影響を与えている。

本稿は、主として、この特集記事以降の研究開発の進展について解説し、機械翻訳における自然言語処理技術についての現状と方向性を示すことを目的としている。

本稿では、まず、翻訳処理の手順を概説し、次に、機械翻訳の最大の課題である曖昧性の問題を取り上げ、その解消方法として、規則方式、用例方式に分けて解説する。また、翻訳処理上、重要な辞書の充実と、翻訳品質の評価法についても言及する。

2. 翻訳処理の手順

機械翻訳とは計算機を用いて、ある言語で書かれた表現を他の言語に翻訳する技術である。元になる言語を原言語、翻訳される先の言語を目標言語と呼ぶ。原言語の表現を原文、目標言語の表現を訳文と呼ぶ^{**}。

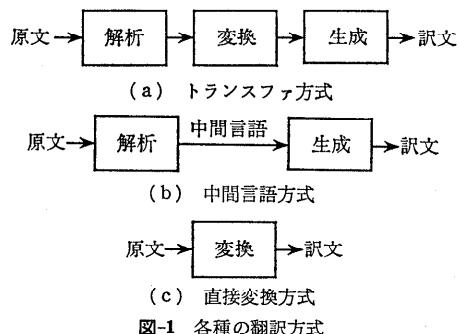
機械翻訳の方法は、図-1(a)に示すように、解

析、変換、生成の三つの過程で行われるのが一般的である。この方式はトランスファ方式と呼ばれ、現在市販されている多くの翻訳システムはこの方式である***。トランスファ方式では、まず、解析過程で、原文を（原言語の）中間表現と呼ばれるものに、変換する。次に、変換過程で、原言語の中間表現を目標言語の中間表現に変換する。最後に、目標言語の中間表現から、訳文を生成する。

この方式では、中間表現をどう定義するかが問題となる。中間表現の要件として

- 原文の表現している内容を失わないだけの豊富な表現能力がある。
- 変換・生成が簡単に実行できるように、できるだけ単純な形式である。

の2点が考えられる。このような、謂わば、あいまい矛盾する要件をできるだけ満足する表現形式として、形態素****間の依存構造が用いられることが多い。依存構造は、依存文法^{5), 6)}に基づいて定義



* 科学技術文献の日英・英日翻訳目的とした機械翻訳システム開発のプロジェクト²⁾。

** 原言語や目標言語の表現は必ずしも文ではなく、句や段落であつてもよいが、原文・訳文という用語を用いる。

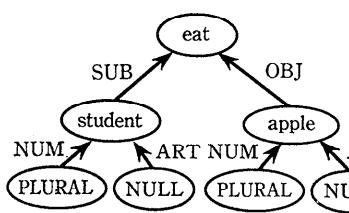
*** たとえば、文献3)の中に、各種システム事例がある。

**** 言語単位の一つであり、意味を担う最小の単位⁴⁾と定義されるが、機械処理の場合、単語あるいは、単語辞書に登録されている見出し語（複合語や慣用句も含む）を形態素と呼ぶことが多い。本文でも形態素と単語を混同して用いる。

† Natural Language Processing in Machine Translation by Terumasa EHARA (Japan Broadcasting Corporation, Science and Technical Research Laboratories, Program Production Technology Research Division) and Hozumi TANAKA (Tokyo Institute of Technology, Faculty of Engineering, Department of Computer Science).

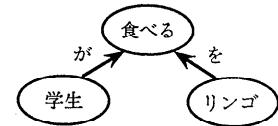
‡ 日本放送協会放送技術研究所先端制作技術研究部

†† 東京工業大学工学部情報工学科



SUB: 主格 OBJ: 目的格
NUM: 数 ART: 冠詞

図-2 依存構造の例



されるものであり、日本語の場合は係り受け⁷⁾とも呼ばれる。依存文法は、種々の言語に共通な枠組みとして知られているため⁸⁾、依存構造は中間表現として適している。図-2に依存構造の例を示す。英語の“students eat apples”と日本語の「学生がリンゴを食べる」の依存構造である。図-2から分かるように、依存構造は、形式的には、節点や弧にラベルが付けられ、弧に方向性をもつグラフ（ラベル付き有向グラフ）である。通常は、ループ（方向性を無視しての）をもたない木構造が用いられる。内容語を節点に、機能語あるいは文法的機能（SUB: 主格など）を弧に対応させるものが多い^{*}。また、ラベルには単語の語形だけでなく、品詞や意味的特徴などの情報も記述できる。依存構造以外の中間表現⁹⁾を用いているシステムもあるが^{10), 11)}、その数は少数である。

トランスファ方式以外に二つの方式が考えられる。第1は、図-1(b)のように、変換過程がまったく存在しない方式である。この場合、原言語の中間表現と目標言語の中間表現が一致する。この共通な中間表現は中間言語と呼ばれ、この方式は中間言語方式と呼ばれる^{12)~14)}。多言語間翻訳の場合、変換過程が言語対の数だけ必要となるが、中間言語方式を採用すれば、この部分が不要である。アジア5カ国語（中国語、タイ語、インドネシア語、マレーシア語、日本語）を対象にした。多言語間翻訳プロジェクト（CICCプロジェクト）は、この方式を採用している¹⁵⁾。

他のは一つは、図-1(c)のように、変換過程のみ存在する方式である。あらかじめ、原文と、それに対応する訳文の対訳パターンを、大量に用意しておき、翻訳対象の入力原文を、対訳パターンの原文部分とマッチングさせることによって直接翻訳する方式である^{16), 17)}。

最もよく使われるトランスファ方式に戻って、三つの過程をもう少し詳細にみてみよう。原文を中間表現に変換する解析過程では、まず、形態素解析が行われる。ここでは、原言語の辞書を用いて、原文に含まれる単語が抽出されるとともに、品詞や意味的特徴などの情報が得られる。特に、日本語は単語単位に分かち書きされないため、単語境界の検出も行われる¹⁸⁾。次に構文規則を用いて、構文解析が行われる。ここでは、文脈自由文法¹⁹⁾に代表される構成文法⁶⁾が用いられることが多い。構文解析によって、構文木が得られる。構文木からは、句や節の範囲やそれらの構成関係など、構文的情報が得られる。次に、構文木を依存構造に変換する。この変換には、たとえば、構文解析のための文法規則に付随する変換規則が用いられる²⁰⁾。構文解析に、依存文法（係り受け解析）を用い、構文木を経ることなく依存構造を直接求めるやり方もある^{21), 22)}。

変換過程では、構造と語彙の変換が行われる。原言語の依存構造と目標言語の依存構造が、グラフの構造としてみたとき、異なる場合に、構造変換が行われる。依存構造は、多くの言語で類似の構造をもつので、小さな構造変換で済む場合が多い。たとえば、図-2では、英日ともに、類似の構造である。図-3には、大きな構造変換が必要な例を示す。

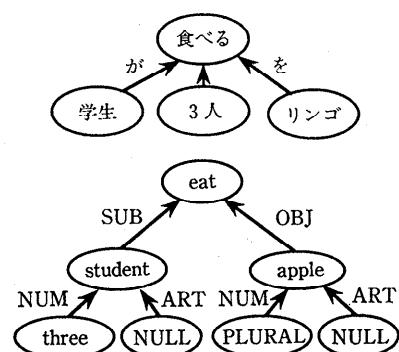


図-3 大きな構造変換の必要な例

* 名詞、動詞、形容詞、副詞など実世界で具体的な内容をもつものを表現する語を内容語、前置詞、後置詞（たとえば、日本語の助詞）など文法的機能を主に表現する語を機能語と言う。

一方、語彙変換は、依存構造の弧や節点に付けられたラベルを原言語のものから、目標言語のものに変換する過程である。たとえば、図-2の英語の節点ラベル“student”は日本語の節点ラベル「学生」に変換され、英語の弧ラベル“SUB”（主格）は日本語の弧ラベル「が」（主格助詞）に変換される。

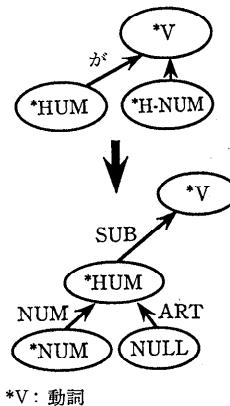
最後の過程は、目標言語の中間表現から目標言語の文字列である訳文を生成する過程である。ここでは、語順の決定、語の屈折処理などが行われる。たとえば、図-3の英語中間表現からは、名詞が複数形に屈折され、語順が決定され、文頭の文字が大文字にされ、“Three students eat apples”が生成される。

3. 曖昧性の問題

このように述べてくると、機械翻訳には、基本的に何の問題もないようと思われるかもしれない。解析、変換、生成の処理を順次、原文に適用し、出力の訳文を得ればよい。しかし、実際には大きな問題がある。それを述べる前に、上記の処理過程を若干抽象化して考えよう。

依存構造を用いた中間表現は、ラベル付き有向グラフであることはすでに述べたが、原文から訳文に至る、機械翻訳の処理対象は、すべて、そうみることができる。たとえば、日本語の原文は、原文に出現した各文字を節点で表し、各文字の出現順に弧で結んだグラフ構造で表現できる。このように考えると、すべての処理は、ラベル付き有向グラフ上の書換えシステムと形式化できる。つまり、原文のグラフから始めて、書換え規則を順次適用して、訳文のグラフに至る。書換え規則は、グラフ上のパターンと、その書換えを表すアクションからなっている。パターンが入力グラフにマッチすれば、アクションが実行されてグラフが書き換えられる。たとえば、図-3の日英変換のうちの構造変換の部分は、図-4のような書換え規則の適用によって行われる。

グラフ書換えシステムの一般的な問題として、規則集合の被覆問題と適合問題がある。被覆問題とは、入力グラフを正しい出力グラフに導く規則の系列が存在するかどうかの問題である^{*}。入力グ



*V: 動詞
*HUM: 名詞(人間)
*H-NUM: 数量詞(人間)
*NUM: 数詞

図-4 書換え規則の例

ラフのうち、そのような系列が存在するものの割合を規則集合の被覆率と呼ぶ。

被覆問題の例として、原言語の辞書に登録されていない語が処理対象の原文に出現する場合がある。原始的な形態素解析処理では、このような場合、解析に失敗する。つまり、グラフ書換えシステムの上で正しい出力に向かって進むことができない。規則集合がこの原文を被覆していないことになる。そこで、実際の形態素解析処理では、なんらかの未登録語処理規則を用いて、この問題を逃れている^{23)~25)}。

一方、正しい出力に導く規則系列が存在したとしても、必ずしも、正しい出力が得られるわけではない。なぜならば、ある処理対象に、複数の規則が適用可能な場合、正しい出力に導く規則が適用されず、誤った出力に導く規則が適用されてしまうかもしれないからである。これが適合問題である。正しい出力が得られる可能性のある入力グラフ、つまり、規則集合によって被覆されている入力グラフのうち、規則適用が正しく行われて、出力として、実際に正しいグラフが得られるものの割合を適合率と呼ぶ。適合問題は、従来から、規則適用の曖昧性の問題と言われてきた。本文でも、この呼び名に従う。

曖昧性の問題の例として、語彙変換規則で生ずるものがある。原言語の単語に対して、対応する目標言語の単語は、通常、複数存在する。たとえば、“make”的訳語には「作る」「する」など多数ある。そのような、複数の訳語の中から原文の内

^{*} 機械翻訳では正しい出力とは、つまり、正しい訳文である。何が正しい訳文であるかは、必ずしも明確ではないが、ここでは、すべての入力原文に対して正しい訳文が定義できると仮定する。

容を正しく表現するものを選択して、訳文を作らなければならない。これは、訳語選択の問題と呼ばれ、典型的な曖昧性の問題である^{26), 27)}。

近年の翻訳システムは書換え規則が大規模になり、被覆問題は、かなり解決してきている。つまり、多くの原文に対して、その正しい訳文に至る書換え規則の系列が存在するようになってい。逆に、曖昧性の問題は、ますます重大になってきた。規則の数が多くなるほど、適用可能な規則の数も増え、曖昧性が増えるからである。

曖昧性の問題は、自然言語処理一般の問題であり、すでに、優れた解説がある²⁸⁾。本文では、特に、機械翻訳の観点から、この問題をみてみよう。まず、どのような曖昧性が存在するのであろうか。

ラベル付き有向グラフ上の書換え規則は、グラフとしての構造を変更する部分と、ラベルを変更または新たに付与する部分に分けられる。それに応じて、曖昧性の問題も

- 構造的曖昧性（グラフ構造変更の曖昧性）
- 語彙的曖昧性（ラベル変更・付与の曖昧性）

の2種類がある。解析、変換、生成の各過程でこの2種類の曖昧性が生ずるが、特に、機械翻訳で特徴的なものは変換過程における曖昧性である。

変換過程の語彙的曖昧性としては、前述した訳語選択の問題のほかに次のようなものがある。それは、目標言語の中間表現に新たに設けられた節点や弧にラベルを付与する場合に生ずる。この場合、原言語の中間表現には陽に存在しない情報を付与しなければならず、ゼロ要素の補完と呼ばれる。具体的には以下のようなものがある。日英翻訳における冠詞や数の選択^{29), 30)}（「学生」を“a student”, “students”, “the student”, “the students”のいずれにするか）や、英日翻訳における助数詞の選択³¹⁾（“three students”を「三つの学生」ではなく、「3人の学生」とする）、主語など英語では必須の要素が欠けている日本語原文へ、それらの要素を補完すること^{32)~34)}（「さっそく送らせていただきます」を“I will send it off to you pretty quick”と“I” “it” “to you”を補完する）などがある。ゼロ要素の補完は訳語選択にも増して困難な問題である。

変換過程の構造的曖昧性としては、たとえば、

「学生が3人リンゴを食べる」と「学生が3個リンゴを食べる」の遊離数量詞構造を英語に変換するとき、前者は「学生」を、後者は「リンゴ」を修飾させなければならないことなどがある。

このように、曖昧性の問題は機械翻訳を含む自然言語処理の最大の問題であり、今日でも十分に解決されているとは言いがたい。機械翻訳研究開発の歴史は曖昧性との闘いの歴史であると言っても過言ではない。

4. 規則方式による曖昧性の解消

曖昧性の問題を解決するには、二つの方法が考えられる。第1は、規則を細分化して、複数の規則が適用されることがないようにする方法である。これによって、曖昧性を減らすことができる。第2の方法は、規則適用に曖昧性が生じた場合、複数の適用可能な規則をなんらかの手段で評価して、評価値の高い規則を選択するものである。

規則細分化の例としては、品詞の細分類がある。英語の動詞を、文型パターンによって細分類し、その結果に基づいて、構文解析規則を作成すれば、構文解析における構造的曖昧性を少なくすることができる²⁷⁾。

規則評価の例として、他動詞の訳語選択で、目的語の意味的な特徴をとらえる方法がある。たとえば、“make”的訳語選択で、“make a cake”的ように、目的語が具体物ならば、「ケーキを作る」のように「作る」と訳し、“make an address”的ように、動作名詞ならば、「演説をする」のように「する」と翻訳する。これは、“make”がどのような意味特徴をもつ目的語と共に起したかという狭い意味での文脈を用いて語彙変換規則を評価していくことになる。別の例として、分野別辞書の利用がある。分野別辞書は、翻訳対象が属している分野に適応させた辞書であり、このような辞書を利用することによって、その分野で、よく用いられる訳語が優先して選択されるようになる。この場合は、分野という広い意味の文脈が利用されている。

このように、規則評価は、各種の文脈を利用して行われることが分かる。利用される文脈の範囲は、動詞と目的語の2語の共起という狭いものから、動詞、目的語、主語の3語の共起、1文内の

共起、文間にまたがる共起、さらには、入力原文には表現されていない人間のもっている知識や、原文の用いられた状況などといった広い意味での文脈³⁵⁾まで、さまざまなものが考えられる。もし、文脈の把握が機械で正確にできるならば、できるだけ広範囲の文脈を用いたほうがより正確な規則評価が行えるはずである。

現在の商用システムを含む、従来の機械翻訳システムの多くは、翻訳処理の各過程で用いられる規則の作成および、規則評価のための関数の作成を、人間の直感に頼って行ってきた。このような方式は規則方式と呼ばれる。たとえば、Mu プロジェクトのシステムはこの方式である。しかし、曖昧性の問題に対処するためには、規則自身や、規則評価関数を精密にしなければならず、そうすると、人手による規則の作成や管理が困難になってくる。

たとえば、“make”を含む文 1008 文を人手で翻訳し、その訳語を調べたところ、110 語の異なる訳語が得られた。このように多くの訳語の中から目的語や主語などとの共起によって、適切な選択を行う規則を記述するには、きめ細かく意味特徴をとらえることが必要である。そのためには、詳細な意味分類辞書を作成し、目的語や主語あるいは前置詞句内の目的語として、どのような意味分類項目が出現したときには、どのような訳語を用いるかということを、逐一記述する必要がある。しかし、このような作業は、手間がかかるばかりでなく、そもそも、その基準を判断することも難しい。この問題に対処するために、用例の利用が考えられた。

5. 用例の利用

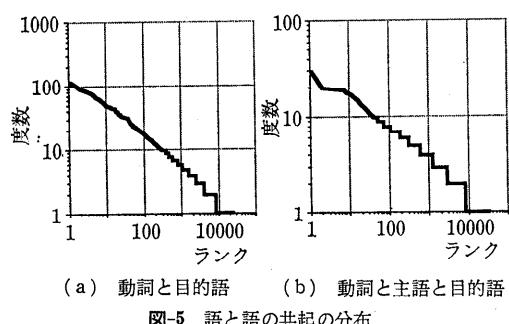
人間の翻訳は、機械のように、解析、変換、生成といった複雑な処理過程を経て行っているのだろうか。これまでの自分の翻訳経験の中から、原文に類似する翻訳例を利用して直接翻訳しているのではないだろうか。このアイデアに基づいて、類推による翻訳が提案された³⁶⁾。この方式は、数多くの原文とその対訳を、用例としてあらかじめ用意しておき、翻訳対象の原文に類似の用例を検索し、それを用いて直接翻訳するものである。このため、用例方式とも呼ばれる。用例方式は規則方式の限界を越えるものとして期待され、近年、

盛んに研究されており、部分的ではあるが、適合率の向上が得られている³⁷⁾。用例方式に関して、詳しい解説が、文献 38) にあるので、ここでは、訳語選択を例に、規則方式との比較をみてみよう。

目的語との共起を用いて、“make”の訳語選択をする場合を再び考える。規則方式では、目的語の意味分類項目に従って選択規則を人間が作成していた。一方、用例方式では、“make a cake”なら「作る」，“make an address”なら、「する」と目的語の語形そのものに従って選択規則が書かれる。そして、意味分類辞書は用例と入力原文との類似性の計算に用いられる。“make a speech”的 “speech”は“cake”よりも“address”に意味的に類似しているから、「スピーチをする」と訳される。

このように用例を利用することは、規則方式と比較して、より細かいレベルの規則の利用、あるいは、より精密な規則の評価をしていることになる。この意味で、「用例も規則である」と言える。しかも、人間による抽象化、理論化を経ていない規則である。用例に理論的な考察を加え、できるだけ一般化することが、今後の発展には、必要であろう³⁸⁾。

また、用例方式を実用するには、大規模な用例ベースが作成できるかどうかが問題となる。AP 通信社から 1989 年 6 月に配信された外電、14 万文余りから、構文解析を用いて、動詞と目的語、主語の共起を抽出した。その結果得られた動詞と目的語および、動詞と目的語、主語の共起データに対する度数分布を図-5 に示す。度数 1 の用例が非常に多い。これは、ここに示す用例ベースを用いたのでは、原文と完全に一致する用例が見つかる率が小さいことを意味している。しかも、適合率を上げようとして、文脈の範囲を広くすれば



するほど、一致する用例が見つかる率は下がってくる。そこで、類似性を利用するわけであるが、類似性の範囲を広くとると適合率が下がってしまう。実用的な翻訳システムにて用例方式を利用するためには、相当大規模な用例ベースが必要になると予想される。

最近、大規模なテキストが機械可読の状態で入手可能であるが、これらの多くは、フラットなテキスト（単なる文字列）であり、用例方式で用い得るようなデータ形式にはなっていない。テキストから用例ベースをいかにして効率よく構築できるかが、この方式を実用する上での大きなポイントとなるであろう。対訳テキストから、原文と訳文の対応する部分を統計的に抽出する手法⁴⁰⁾や原文と訳文の構文構造や依存構造から変換規則を抽出する手法^{41), 42)}は、大規模な用例ベースの作成に利用できる可能性がある。

また、フラットなテキストであっても、原言語と目標言語の対訳テキストは得にくい。両言語の必ずしも対訳でないテキストを用いて、規則を抽出する方法が提案されており^{43)~45)}、有望な手段として期待される。

6. 辞書の充実

用例方式での類似性の計算をきめ細かく行うには詳細な意味分類辞書が必要である。また、未登録語による被覆率の低下を防ぐには大規模な単語辞書が必要である。このように、機械翻訳における辞書の重要性は高く、各メーカーは、その充実に努力してきた。その結果、我が国では通産省の主導の下に、日本電子化辞書研究所（EDR）が発足

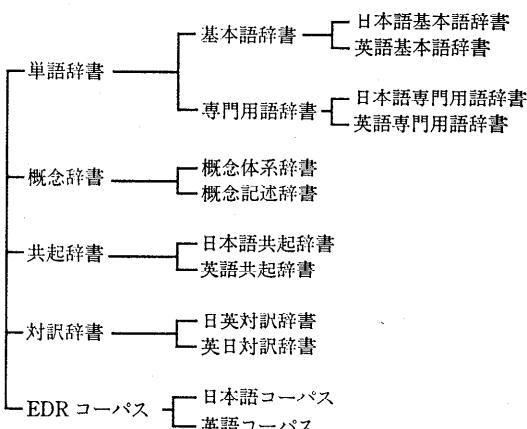


図-6 EDR 電子化辞書の構成⁴⁶⁾

し、現在、図-6に示すような各種の辞書類が出来上りつつある⁴⁶⁾。

図-6には、単語辞書ばかりでなく、単語のもつ概念と概念間の種々の関係を記述した概念辞書がある。これは、意味分類辞書としても利用可能である^{*}。また、共起辞書やEDR コーパスは用例ベースとして、利用可能である。

7. 翻訳品質の評価法

完全な曖昧性の解消には、完全な原文の理解が必要である。これは、人間による翻訳の場合でも、そうであると言われている⁵¹⁾。計算機が原文を完全に理解できるためには、人間の五感に相当するような、計算機と実世界とのリンクが必要であり⁵²⁾。現在の計算機では、そのことは不可能である。そこで、被覆率、適合率ともに1である完全な自動翻訳機は、当分の間、出現せず、1未満のシステムを考えることになる。この値がどれだけあれば、実用できるかどうかは、利用する形態によって異なる⁵³⁾。

いずれにしても、被覆率や適合率の計算には、出力が正しいかどうかの判断が必要であるが、正しいかどうかは、2値ではなく、連続性をもつ。このような、翻訳品質の評価法として、文献54)で忠実度と理解容易性が提案されている。人間による翻訳に関する翻訳品質については、文献51)で、ノックスからの引用として、

- 正確であること
- 理解できること
- 読みやすいこと、読むにたえること

の3点をあげている。そして、これらのなかで、第3の点が最も重要であると、指摘している。機械翻訳に置き換えていると、「正確で理解も容易であるが、読むにたえない」という状況は考えにくい。これは、現在の機械翻訳は第3の観点からの評価を行えるほど品質が高くないことを示している。忠実度、理解容易性の基準は、第1、第2の点に対応する。

以上は訳文の品質に関するものであるが、原文の難しさの評価も必要である。簡単な原文ばかりを対象にして、高い品質の訳文が得られたとしても、必ずしも、高精度なシステムとは言えない。

* 一般に利用できる。その他のこの種の辞書としては、文献47)~49)がある。英語に対しては、文献50)などがある。

原文の評価については、次のような項目が考えられる⁵⁵⁾。

- 原文の分野（マニュアル、技術論文、報道ニュース、手紙文、対話文など）
- 原文の長さ
- 構文や文体の複雑さ

また、第3の点に関して、具体的評価項目も提案されている^{56), 57)}。しかし、これらの要素を総合して、難しさを定量的に評価する方法は得られていない。

原文の難しさが評価しにくいので、評価試験用の文例集が提案されている^{58), 59)}。しかし、機械翻訳の利用者にとって、自分の目的に合う文例集はなかなか見つからず、結局、自分のもっている翻訳対象文を用いて試験する場合が多い。

また、自然言語処理上からの評価以外に、翻訳システムとしては、翻訳速度や使いやすさなど、種々の評価基準が考えられる⁵⁹⁾。

また、不完全な機械翻訳技術を使いこなすために、人間のほうで、曖昧性の少ない原文を作成することが考えられる。そのような原文表現法として、制限言語が提案されている⁶⁰⁾。翻訳システムの評価法の研究と制限言語の研究は、表裏一休のものであろう。

8. おわりに

機械翻訳における自然言語処理技術の現状を曖昧性の解消を中心に述べた。曖昧性のきめ細かな解消のためには、用例の利用が有効であること、および、このような用例方式を実用化するためには、大規模な用例ベースを効率良く作成する方法の開発が課題であることを示した。

社会の多方面で国際化が進む今日、機械翻訳の必要性は、ますます高まっている。しかし、現在の翻訳システムは、マニュアルの翻訳など、ごく限られた分野で実用されているにすぎない。一層の研究努力の結果、幅広い分野での実用化が一日も早く実現することを期待したい。本稿が、そのための一助となれば幸いである。

参考文献

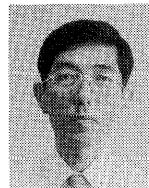
- 1) 和田 弘：計算をしない計算機、情報処理、Vol. 1, No. 1, pp. 11-15 (1960).
- 2) 長尾 真他：科学技術庁機械翻訳プロジェクトの概要、情報処理、Vol. 26, No. 10, pp. 1203-1213

(Oct. 1985).

- 3) 野村浩郷、田中穂積編：機械翻訳、bit 別冊、共立出版 (1988).
- 4) Nida, E. A.: Morphology (second edition), The University of Michigan Press, p. 6 (1974).
- 5) Hays, D. G.: Dependency Theory: A Formalism and Some Observations, Language, Vol. 40, pp. 511-525 (1964).
- 6) 児玉徳美：依存文法の研究、研究社出版 (1987).
- 7) 首藤公昭、吉村賢治：日本語の構造とその解析、情報処理、Vol. 27, No. 8, pp. 947-954 (Aug. 1986).
- 8) 児玉徳美：語順の普遍性、山口書店 (1987).
- 9) 野村浩郷、内藤昭三：自然言語理解における意味表現、情報処理、Vol. 27, No. 8, pp. 906-914 (Aug. 1986).
- 10) 飯田 仁：対話翻訳と高度自然言語処理、人工知能学会誌、Vol. 6, No. 3, pp. 328-337 (1991).
- 11) 西田豊明他：モンテギューワークに基づく英文和訳システムの試作、情報処理学会論文誌、Vol. 23, No. 2, pp. 107-115 (Feb. 1982).
- 12) Hutchins, W. J.: Progress in Documentation: Machine Translation and Machine-Aided Translation, Journal of Documentation, Vol. 34, No. 2, pp. 119-159 (1978).
- 13) 田中穂積他：機械翻訳における中間言語方式をめぐって、人工知能学会誌、Vol. 4, No. 6, pp. 671-680 (1989).
- 14) 村木一至他：機械翻訳システムの中間言語、情報処理学会研究会資料、NL-73-13 (1989).
- 15) 辻 良英：近隣諸国間の機械翻訳システムの研究協力プロジェクト、電子情報通信学会誌、Vol. 72, No. 6, pp. 679-681 (1989).
- 16) 古瀬 蔵、飯田 仁：変換と解析の協調的処理による翻訳手法、情報処理学会研究会資料、NL-87-4 (1992).
- 17) 加藤直人、相沢輝昭：外電ニュースの定型文抽出とその英日機械翻訳、情報処理学会研究会資料、NL-93-2 (1993).
- 18) 池原 哲也：文書校正支援システムにおける自然言語処理、情報処理、Vol. 34, No. 10, pp. 1249-1258 (Oct. 1993).
- 19) Hopcroft, J. E. and Ullman, J. D. (野崎昭弘他訳)：オートマトン、言語理論、計算論 (I, II), サイエンス社 (1984).
- 20) 中瀬純夫：英日機械翻訳システムにおける解析手法について、情報処理学会研究会資料、NL-69-7 (1988).
- 21) 林 良彦：結合価構造に基づく日本文解析、情報処理学会研究会資料、NL-62-6 (1987).
- 22) 高木 朗、吉田 豊：次期機械翻訳システムにおける中間表現と構文解析の基本的な考え方、CSK技術通信、Vol. 8-2, No. 16, pp. 3-8 (1988).
- 23) 吉村賢治他：未登録語を含む日本語文の形態素解析、情報処理学会論文誌、Vol. 30, No. 3, pp. 294-301 (Mar. 1989).
- 24) 中村 俊：和文形態素解析における異なった送り仮名の自動検出と自動修正、第34回情報処理学会全国大会講演論文集、Vol. 2, pp. 1297-1298 (1987).
- 25) 島津美和子他：カタカナ異形表記・誤記修正機能

- の開発・評価、第44回情報処理学会全国大会講演論文集、Vol. 3, pp. 249-250 (1992).
- 26) Delavenay, E.: An Introduction to Machine Translation, Thames and Hudson, p. 90 (1960).
- 27) 田中穂積、糸井利人：基本動詞「MAKE」を含む文の日本語への訳し分け、情報処理学会研究会資料、NL-43-3 (1984).
- 28) 長尾 確、丸山 宏：自然言語処理における曖昧さとその解消、情報処理、Vol. 33, No. 7, pp. 746-756 (July 1992).
- 29) 野垣内出、飯田 仁：英語定名詞句生成のための日本語名詞句の理解、電子情報通信学会春季全国大会資料、Vol. 6, p. 83 (1990).
- 30) Francis Bond, 小倉健太郎：六種類の可算性を用いた英語名詞句の可算／不可算の決定、第46回情報処理学会全国大会講演論文集、Vol. 3, pp. 107-108 (1993).
- 31) 高山泰博他：日英機械翻訳における数量表現処理の拡張、第42回情報処理学会全国大会講演論文集、Vol. 3, pp. 174-175 (1991).
- 32) 鈴木雅実：対話翻訳における領域知識による補完手法の検討、第45回情報処理学会全国大会講演論文集、Vol. 3, pp. 89-90 (1992).
- 33) 工藤育男、友清睦子：日本語の述部の特性を用いた省略の補完機構について、電子情報通信学会論文誌、Vol. J76-D-II, No. 3, pp. 624-635 (1993).
- 34) 金 淵培、江原暉将：日英機械翻訳のための日本語ニュース文自動短文分割と主語の補完、情報処理学会研究会資料、NL-93-3 (1993).
- 35) 池上嘉彦：記号論への招待、pp. 35-64, 岩波新書 (1984).
- 36) Nagao, M.: A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, Artificial and Human Intelligence, pp. 173-180, North-Holland (1984).
- 37) 隅田英一郎、飯田 仁：用例主導型機械翻訳、情報処理学会研究会資料、NL-82-5 (1991).
- 38) 佐藤理史：実例に基づく翻訳、情報処理、Vol. 33, No. 6, pp. 673-681 (June 1992).
- 39) 田中英輝、相沢輝昭：基本動詞と動作名詞の組合せ表現の英日機械翻訳手法、情報処理学会論文誌、Vol. 34, No. 2, pp. 196-205 (Feb. 1993).
- 40) Brown, P. et al.: A Statistical Approach to Machine Translation, Computational Linguistics, Vol. 16, No. 2, pp. 79-85 (1990).
- 41) 佐藤理史、長尾 真：文法推論に基づいた翻訳文法の学習方式、情報処理学会研究会資料、AI-46-11 (1986).
- 42) Kaji, H. et al.: Learning Translation Templates from Bilingual Text, Proc. of the COLING '92, pp. 672-678 (1992).
- 43) Dagan, I. et al.: Two Languages Are More Informative Than One, Proc. of the 29th Annual Meeting of the ACL, pp. 130-137 (1991).
- 44) 野美山浩：目的言語の知識を用いた訳語選択とその学習性、情報処理学会研究会資料、NL-86-8 (1991).
- 45) 加藤直人：対象言語のフリーテキストを用いた複合名詞の訳語選択、第46回情報処理学会全国大会講演論文集、Vol. 3, pp. 141-142 (1993).
- 46) 日本電子化辞書研究所：EDR 電子化辞書仕様説明書、日本電子化辞書研究所 (1993).
- 47) 国立国語研究所：分類語彙表、秀英出版 (1964).
- 48) 大野 普、浜西正人：類語国語辞典、角川書店 (1985).
- 49) 荻野綱男：現代日本語名詞シソーラスから見た語彙の意味分類、平成4年度科学研究費補助金研究成果報告書 (1993).
- 50) Kirkpatrick, B.: Roget's Thesaurus of English Words and Phrases, Longman (1987).
- 51) 別宮貞徳：翻訳読本、講談社現代新書 (1979).
- 52) Winograd, T.: Understanding Natural Language, Edinburgh University Press (1972).
- 53) 江原暉将：機械翻訳の正翻訳率と翻訳に要する作業時間の関係、第45回情報処理学会全国大会講演論文集、Vol. 3, pp. 115-116 (1992).
- 54) 長尾 真、辻井潤一：Mu プロジェクトにおける日英翻訳結果の評価、情報処理学会研究会資料、NL-47-11 (1985).
- 55) 牧野武則：評価技術、bit 別冊 機械翻訳、pp. 128-134, 共立出版 (1988).
- 56) 池原 悟、白井 諭：日英機械翻訳機能試験項目の体系化、電子情報通信学会研究会資料、NLC-90-43 (1990).
- 57) 浦谷則好他：話し言葉の日英機械翻訳システムの評価法、第46回情報処理学会全国大会講演論文集、Vol. 3, pp. 159-160 (1993).
- 58) 成田 一：機械翻訳における構造処理能力の評価、情報処理学会研究会資料、NL-69-1 (1988).
- 59) 野村浩郷、井佐原均：機械翻訳の評価基準について、情報処理学会研究会資料、NL-89-9 (1992).
- 60) 野村浩郷他：制限言語、bit 別冊 機械翻訳、pp. 135-142, 共立出版 (1988).

(平成5年5月7日受付)



江原暉将（正会員）

昭和42年早稲田大学理工学部電子通信学科卒業。同年日本放送協会に入局。昭和45年より放送技術研究所に勤務。かな漢字変換、機械翻訳などの研究に従事。現在、先端制作技術研究部主任研究員。電子情報通信学会、Association for Computational Linguistics各会員。



田中 穂積（正会員）

昭和39年東京工業大学理工学部制御工学科卒業。昭和41年同大学院修士課程修了。同年電気試験所(現、電子技術総合研究所)入所。昭和58年東京工業大学工学部情報工学科助教授。昭和61年同教授となり現在に至る。工学博士。人工知能、自然言語処理の研究に従事。電子情報通信学会、認知科学会、日本ソフトウェア科学会、人工知能学会、計量国語学会各会員。