



Title	ユーザーとの対話を重視した英日機械翻訳システム
Author(s)	沢村, 孝至; Sawamura, Takashi; 宮永, 喜一 他
Citation	北海道大學工學部研究報告, 153, 79-89
Issue Date	1990-11-29
Doc URL	https://hdl.handle.net/2115/42253
Type	departmental bulletin paper
File Information	153_79-90.pdf



ユーザーとの対話を重視した英日機械翻訳システム

沢村 孝至 宮永 喜一 栃内 香次

(平成2年8月31日受理)

English-Japanese Machine Translation System Emphasized on User-Interface

Takashi SAWAMURA, Yoshikazu MIYANAGA, Koji TOCHINAI

(Received August 31, 1990)

Abstract

We develop an English-Japanese Machine Translation system emphasized on a User-Interface, which frequently communicates with users and performs translation in cooperation with users. Users can obtain high quality versions by using the system integrated processes of pre-editing, registration of words into dictionary, translation and post-editing, which are separated in conventional MT systems. And this system acts like "interpreter" because it interpretes input words in a sentence successively.

1. はじめに

多国語間の文書の翻訳を機械により自動で行ないたいという自動翻訳(機械翻訳)の要求は古くからあるが、コンピュータ技術の発展により実用にむけての見通しが立ってきたのはここ数年来的なことである。特に日本においては、英日および日英の文書間の翻訳の需要が高く、いくつかの翻訳システムがパーソナルコンピュータやワークステーションあるいは汎用大型機上で商品化されている。また人間が行なう翻訳作業を機械が支援する翻訳支援システムも実用化されている。

これら既存のシステムに共通する特徴は大量の文書を一括して翻訳させるのに向いていることであり、ユーザーインターフェース面は不十分で使い勝手は良くない。これに対して、われわれは例えば1編の論文などのような比較的少量の文書の高品位な翻訳を目的とした使い勝手の良い機械翻訳システムの開発を行っている。このシステムはユーザーとシステムの対話を重視し、単語を少しずつ入力しながら翻訳と修正を繰り返すことにより質の良い翻訳結果を得ようとするものである。本稿ではこのシステムの構想ならびに翻訳アルゴリズムの基本概念を述べる。以下、第2章では現在の機械翻訳システムのインターフェース上の問題点を考察し、われわれが開発しているシステムの立場を明らかにする。第3章では翻訳プロセスの視点から、本システムの基礎となっている概念と動作について述べる。

2. 機械翻訳システムのユーザーインターフェース

機械翻訳はある言語で書かれた文章を他の言語へ翻訳するという人間の知的活動をコンピュータに行なわせようとするものである。

その一般的な方法は、原文の構文を調べて各々の語の役割を明らかにし、文の意味を理解して、対応する翻訳対象言語の構文上の特徴を反映した翻訳文を作る¹⁾というものである。しかしながら現時点では意味の理解はまだ研究の途上であり、本格的な意味解析と理解を行なう機械翻訳システムは実用化されていない。なお、一般的な機械翻訳のプロセスについては3.1節で述べる。

2.1 商用機械翻訳システムのインターフェース

現在実用化されている多くの商用機械翻訳システムまたは機械翻訳支援システムは構文情報と部分的な意味情報を使って翻訳を行なっている。

ここでは主にインターフェース面から見た商用機械翻訳システムの現状と問題点を述べる。

(1) 前編集と後編集²⁾

一般に機械翻訳システムはなるべく人手を煩わせず機械（コンピュータ）だけで翻訳を済ませることを目指してきた。これは理想的ではあるが、現実にはコンピュータ翻訳の性能は低く、翻訳前に原文に手を加え、また翻訳後には訳文を修正する必要がある。このような翻訳の前後に行なわれる人手による処理には、熟練した翻訳家が作業に当らねばならない。

翻訳の前後に機械翻訳システムの負担を減らすために行なわれる処理を前編集、また翻訳後に不完全な訳文を正しい文にする処理を後編集と言う。前編集はシステムに入力する前に原文を修正してシステムが扱いやすい文体に変えてから入力することである。後編集はシステムが出力した訳文を読み、誤りや翻訳されなかった部分を修正して読みやすい文に変えることである。このような作業を行ないながら翻訳を進める際に、一文ごとに『前編集→入力→翻訳→出力→後編集』のプロセスを切り替えるのは煩わしい。そこで商用システムでは大量のテキストを一度にシステムに処理させ、ユーザーが前後の編集処理をまとめて行なう使い方が定着した。

そのような使い方を要請するシステムは今日いくつも実用化されており、人間が行なう翻訳を機械が支援するという意味の翻訳「支援」システムと呼ばれている。

翻訳支援システムは大量文献を扱わねばならないユーザーを想定して、翻訳を行なう際の辞書引きの省力化を狙っており、翻訳文一つ一つの質より単位時間あたりの処理量（＝速度）を優先している。

(2) 辞書登録

基本的には、翻訳したい原文の中の単語のすべてが機械翻訳システムの辞書中に存在しなければならない。それゆえあらかじめ必要な単語を単語辞書に登録しておく必要がある。商用システムではそれを作るメーカーが十分な単語数を登録して提供するが、ユーザー自身による登録の余地を残しておくこともある。

さて商用システムの場合、これから翻訳しようとする一連の文章に現れる全ての単語が辞書中にあるかどうかをユーザーが翻訳前に判断することは実際的ではない。実際には翻訳した結果の中に正しくない訳文があるとき、その原因をユーザーが解析する段階で単語の未登録を知るのである。

一般に、システムの構成上翻訳プロセスと単語登録は分離しており、翻訳作業中の状態では単語登録はできない。上述のように未登録単語は多くの場合、後編集集中に誤訳が発見されることに

より判明し、可能かつ必要ならばそこで単語登録が行なわれる。

ユーザーが辞書登録を行なうことはメーカー側からは仕様の点で、ユーザー側からは使い勝手の点で重要な問題である。未登録単語を登録する意味は、その対象となる単語が将来再び原文中に出現したときに対応できるようにするためであり、したがって出現頻度が重要となる。商用システムでは、使用頻度が高い単語はほとんど辞書中に用意されている。それゆえ辞書にない単語は、固有名詞など極めて特殊性・限定性の強い単語が主となる。

そのような未登録単語の出現頻度が十分低いときはユーザーは辞書登録を行わず、訳文の修正のみを行なう方が効率的である。その理由は、現在の機械翻訳システムでは辞書登録を行なうには専門的な文法知識が必要で、その手順も煩雑であり、それを行なうよりは、実用上は後編集で単に訳文のみを修正する方が楽だからである。一方、未登録単語でも限定的に出現頻度が高い場合（企業における自社名など）がある。その場合はユーザーは辞書登録の煩雑さを考慮しつつ登録するか否かを判断する必要がある。

ユーザー側が辞書登録の必要を感じても、メーカー側はユーザーによる登録の機能を機械翻訳システムに付与しないことがある。これには3つの理由が考えられる。

一つは、未登録単語の出現頻度は低くその処理は重要ではないという考え方に立つ場合である。これは完全に前・後編集に問題を委ねてはいるが、むしろ実際的である。

第二は未登録単語を処理するための機構を機械翻訳システムが用意している場合であるが、現在はまだ十分な能力を備えたシステムはない。例えば固有名詞の出現にはある程度のパターンが認められるので処理が可能である。しかし一般の品詞については困難な問題が多く、いまだ処理の方法は確立していない。

最後は辞書が複雑すぎてユーザーには扱えなくなっている場合である。一般に質の良い翻訳文を得ようとするほど、各単語に要求される要素が多くなる傾向にある。現在の機械翻訳システムの辞書構成はメーカーしか保守できないほど緻密で巨大になっている。そのような辞書にユーザーが操作することは実際的でなく、場合によっては危険であると考えられる。

現実のシステム設計ではこれら3つの理由と、逆にユーザーによる登録の必要性を総合的に判断し、登録機能を組み込むか否かを決定している。

(3) 再翻訳

機械翻訳システムが出力した翻訳文により未登録単語が発見され、それを辞書に登録したとする。この登録された情報を使ってもう一度該当するテキストを翻訳し直すことは可能であるが、翻訳し直しても正確な訳文が得られる保証はない。そうやって新たに翻訳しなおすより、既に出力されている文を後編集で正しい訳文に修正する方が、手間と時間の点でメリットが大きい。

既に述べたように、商用の翻訳支援システムは大量の文書の一括処理を前提にした、前・後編集重視のシステムになっている。したがって同一の文を修正しながら何度もシステムに通すことは翻訳の質を評価する上では興味深いことであるが、始めからそのような使い方には向いていない。

2.2 機械翻訳の研究動向

ここで商用レベルでなく研究レベルでの機械翻訳システムの一般的な方向性について少し述べておく。現在広く興味を持たれている研究課題は、文章の解析と生成をより精緻に行なうことと、単語及び文法辞書を広範囲かつ詳細に整備することである。特に機械翻訳を自然言語処理研究の一環という面で見ると、自然言語の理解という大目標があり、機械翻訳はその応用として実

現の目標を立てやすい分野である。このような面に関連して機械翻訳の立場から、意味理解・文脈理解・知識獲得の研究が進められており、その成果はいずれ商用システムにも取り入れられて行くであろう。一方、主に企業側で大規模な辞書の拡充が行なわれており、より広範囲の分野での良質な翻訳が可能なシステムを目指している。

2.3 対話的機械翻訳システムのインターフェース

上述のように、現在の実用機械翻訳システムは前編集・翻訳・後編集を分離した一括型のものが多い。一方で、少量の丁寧な訳文を得ることを重視したいという個人的用途をもつユーザーは潜在的に少なくない。このような目的のためにわれわれは即時応答性を持ち、かつユーザーインターフェースを重視した英日機械翻訳システムを開発している。これはパーソナルコンピュータやワークステーション上でユーザーとシステムが頻繁に対話しながら訳文を生成するものである。ここで即時応答性とは、ユーザーが翻訳したい英文を入力していくと、一つの単語を入力することにすでにそこまで入力されている未完成の英文の訳がシステムの応答として即座に出力されることを意味する。このとき、ユーザーは単語入力だけでなく、訳文の整形、すなわち正しい訳語の選択や位置の補正を同時に行なう。このようにして単語の入力と翻訳と修正を繰り返し、結果を確認しながら訳文を完成させていくので、ユーザーとシステムの間には頻繁に情報のやり取りが行なわれる。そのためシステムはユーザーに煩わしさを感じさせないように、優れたユーザーインターフェースをもつことが必要である。

具体的には、原文の入力における変更追加が即座に訳文の出力に反映され、かつそれを支えるべく、両者が容易に編集可能でなければならない。さらに重要なこととして、ユーザーの使用状況に合せた自由度の高い辞書登録の方法を持つ必要がある。この場合も、新たに登録された訳語はその場で出力文に反映されねばならない。辞書登録がある程度自動で行なわれることも使いやすさの向上に必要である。このため、辞書は構造が簡単で個人ユーザーにも保守できるものでなければならない。すなわち、対話的機械翻訳システムでは入力・出力・編集並びに辞書登録が統合化されたユーザーインターフェースを持つことが必要とされる。

以下、われわれが現在開発中の翻訳システムの基本仕様を述べる。

(1) 入出力

システムはパーソナルコンピュータまたはワークステーション上において実現され、ユーザーインターフェースとして2つの画面、すなわち英文（原文）表示画面と和文（訳文）表示画面を持つ。これらは常にディスプレイ上に同時に表示される。2つの画面に対応して、入力は英文入力モードと和文編集モードの2つのモードを持つ。英文表示画面では入力された英文がそのまま表示される。和文表示画面では翻訳された文または編集により変更を加えられた文が表示される。

一例として英文入力モードにおいて一つの文を入力することを考える。その文の初めの単語をキーボードから入力すると英文表示画面に単語が表示される。ここでシステムに翻訳を指示すると、または設定によっては自動的に、その訳語の第一候補が和文表示画面に表示される。ここで単語の訳語が不適當な場合はいくつかの候補を表示させて選択し直すことができる。次の単語を入力するとその訳語が表示されるが、既に表示されている1番目の単語との関係によって、その表示される位置は1番目の単語の訳語の前か後かになる。たとえば「1番目・2番目」の単語がそれぞれ「名詞・動詞」の組み合わせの場合は訳文の表示は「名詞・動詞」の順になるが、組み合わせとして「前置詞・名詞」の場合は訳語の表示は「名詞・前置詞」の順になる。この場合、2番目に入力された名詞の訳語が、1番目に入力され既に表示されている前置詞の訳語の前に挿

入される。(ここで文頭方向を前、文末方向を後と表現する。)以下単語を入力するごとにその訳語が適切な位置に挿入される。単語によっては複数個の位置の候補が存在することもある。これも優先順位の高い位置にまず表示され、不適當ならばユーザーが次候補の位置表示を促すことで適切な位置を選択する。

和文編集モードでは訳文を直接修正できる。これは後編集と等価である。ただしここで行なわれる位置や訳語の修正結果をシステムが判断して自動的に辞書登録することが可能な点が通常の後編集とは異なる。

(2) 単語登録

単語の訳語候補選択時に必要な訳語が見つからないときや、単語そのものが辞書にないときは、自動的に単語登録モードになる。これは和文表示画面上で行なわれ、単語の訳語と位置をユーザーがシステムに教えるという状態である。登録された情報は和文表示画面に反映してただちに表示される。

3. 翻訳プロセス

本章では、従来の機械翻訳や人間の行なう翻訳と比較しながら、本システムの文法と内部動作のあらましを述べる。

3.1 従来の機械翻訳と人間の翻訳

従来方式の機械翻訳では原言語文は図1に示すように形態素解析・構文解析・意味解析の過程を経て内部表現に展開される。以下、各々の過程を簡単に説明する。形態素解析では入力された文を品詞など基本的な構成要素に分解する。これに構文解析を行ない形態素どうしのつながりを構造化する。次に単語の意味を調べ、後の訳語選択に必要な情報を抽出する。文によっては形態素解析や構文解析結果が一意的に決まらず、単語の意味情報を利用して候補を絞る必要があることがある。これらの解析によって原言語文は機械が利用しやすい固有のデータ、すなわち内部表現に展開される。次いでこれが原言語の構造に依存しているときには目的言語の内部表現へと変換される。このため一般にこのような翻訳方式はトランスファ方式と言われる。次に目的言語の

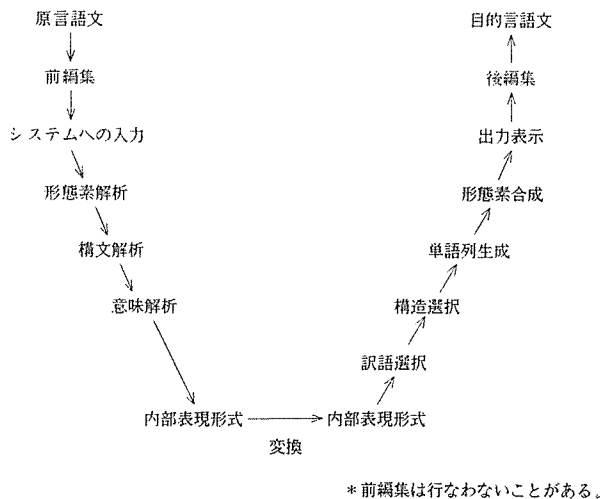


図1 一般の機械翻訳のプロセス。

内部表現は訳語選択・構造選択・単語列生成・形態素合成の過程を経て目的言語の自然言語文に構成される。訳語選択では単語の訳語を求め、構造選択では目的言語の文の構造をできるかぎり絞り込む。単語列合成で目的言語の内部表現を単語列に置き換え、さらに形態素合成で格変化や助詞など語尾・語頭の整形を行なって自然な文を得る。

次に、英文和訳を例にとって人間の行なう翻訳のプロセスを考えてみると以下の様になるであろう。

まず与えられた英文に対して、文頭から単語を眺めていく。意味がわかる単語は訳語を記憶し、わからない単語は辞書を引きつつ、逐語訳で日本語にしていき、訳語が決まらないときは保留にして後で考えることにする。このとき頭の中にある文は英語と日本語の混合文になっている。基本的に先に出てきた単語ほど日本語文でも先になるが当然語順が異なる場合も多い。例えば動詞がきたらそれは訳文では後ろの方にもっていく必要がある。このように後ろに回す単語の訳は一時棚に上げておいて後で引き出すわけで、機械の用語でいえばスタック処理に相当する。文頭から文末に向かって単語を読み出していき、語順を換えたり語尾の整形などを行なって途中までできた訳文をできるだけ整形しながら訳出し、最後の単語まできたら棚上げしておいた訳語をおろして一応訳文が完成する。未定であった訳語は訳文の生成の途中で判明するはずであるが、最後まで未定なら情報が不十分であるのもう一度辞書を引くことになる。訳す人の熟練度が高いほどこの方式で訳せる英文は多くなる。図2に翻訳の一例を示す。

0. The value nn can be any number from 00 through 99.

 (頭から順序を考えながら読んでいく)
1. the value nn(値nn)...
 2. 値nnは...できる
 3. 値nnは...beできる
 4. 値nnは...任意の数を...とることができる
 5. 値nnは...from 00 through 99 の...任意の数をとることができる
 6. 値nnは00から99までの任意の数をとることができる
- 0は原文、1～5は途中経過、6が翻訳結果を表わす。

図2 人間が英文を訳すときの例。

難しくて訳しにくい英文の場合は人間でも解析的な手法を用いる。この場合、人によって、また文によって一文全体か部分かの違いはあろうが、わからなかったところを詳細に解析し構文構造を生成してから訳出を試みる。

このように考えると、従来の機械翻訳方式を人間の行なう翻訳と対比させると、後者の解析的手法に対応している。前者に対応する機械翻訳方式はこれまで存在していない。

3.2 対話的翻訳システム

本システムは、内部的には前節で述べた人間が英文を訳すときの方法に対応している。すなわち、文の最初の単語が入力されたらその単語を辞書から探して訳語を表示する。次の単語が入力されたら同じように辞書引きを行なって、既に入力されている(表示されている)単語との位置関係を調べ、整形しつつ前後に配置する。入力された単語の訳語や位置が一意に決定できないときはユーザーに問い合わせる。これを一文の最後の単語が入力されるまで繰り返す。

このようなシステムを構築するために、われわれは単語同士の位置関係を解析・処理する独自

の手法を開発している。

以下では、本手法の基本概念と内部プロセスについて述べる。

(1) 基本概念^{3)~6)}

本システムは従来の機械翻訳において必要な構文解析部や文法辞書を持たないが、それはユーザーとの対話によって必要な情報を得ようになっているためである。この対話をユーザーが負担に感じないようにするために、われわれは二つの点を重視している。ひとつは使いやすいインターフェースを持つことである。もうひとつはユーザーに高度な文法知識を要求しないということである。これらのことが相乗して、辞書登録を容易にし、かつシステムの使いやすさを向上させる。

本手法で用いられる文法は英文法を基礎にしてそれを機械翻訳向けに改良した独自のものであり、自然言語処理でよく使われる格文法やCFGなどとはあまり関連がない。文法と翻訳手法は不可分であり、強いて言えば文法構造を単語ごとの振る舞いに割りあてたものと言える。これだけでは構文解析の手法のひとつにすぎないが、本システムとの全体的な環境の中では、位置情報(訳順)を簡潔に表わすのに適している。

この文法で基本となるのは語順の並べ換え規則である。以下に本システムの内部動作の基本となる概念を説明する。

① 機能

各単語がどのくらいの範囲にたいして語順をどのような順に並べ換えるかという情報をその単語の『機能』と呼ぶ。名詞以外のほとんどの単語が機能を持っている。機能のいくつかの例を図3に示す。

機能は訳語や変換規則を混然一体化したものである。一つの機能はそれぞれ一つの訳語・単語列の区切り方・訳順の変換規則に対応し、それ以上細かくできない単位である。

機能は品詞と関連するが、依存はしない。一つの機能が複数の品詞の性質を兼ねることもあれば、逆にある品詞の限られた一性質を一つの機能で表すこともある。

並べ換えの機能がない名詞についても、「並べ換えを行わない」という機能を与えることにより、文中の全ての単語を『機能子』としてとらえることにする。また、各々の単語はもとよりカンマやピリオドなどの特殊記号も機能子に含める。以上のように機能子は周辺の機能子の機能を直接制御し訳語や訳順(次節)を決定させるもの、と定義される。

② 訳順

まず『訳順』と「あと・さき」について説明する。

ある英文 E が機能子列 $\{E_1 \cdots E_n\}$ からなっているとす。この中の任意の2つの機能子 E_j と E_k を選んだとき、日本語訳文 J の中で各々の訳語 J_j および J_k がどのような前後関係を持つかで訳順の「さき・あと」を決める。すなわち、 J_j が J_k より文頭に近いとき、 E_j の訳順は「さき」で E_k の訳順は「あと」と呼ぶ。呼び方自体は E の中で E_j と E_k の前後関係がどうであるかには関係ない。またこの「さき・あと」を決定することを英文中の2つの機能子における『訳順の決定』という。この決定には個々の機能子だけでなく E の中で E_j と E_k の前後関係がどうであるかが関係している。

いま、注目する機能子(単語)を E_0 、影響される機能子群を E_G とする。さらに E_G を E_0 よりも前(文頭)の機能子群 E_{Gb} 、 E_0 よりも後ろ(文末)の機能子群 E_{Ga} にわける。訳順の決定とは $E_{Gb} \cdot E_0 \cdot E_{Ga}$ の3者の順序の並べ換えがおこなわれることである。正しい英文では少なくとも $E_{Gb} \cdot E_{Ga}$ それぞれのうちの一つの機能子は E_0 にとって不可欠でなければならず、それらを $E_b \cdot E_a$ と

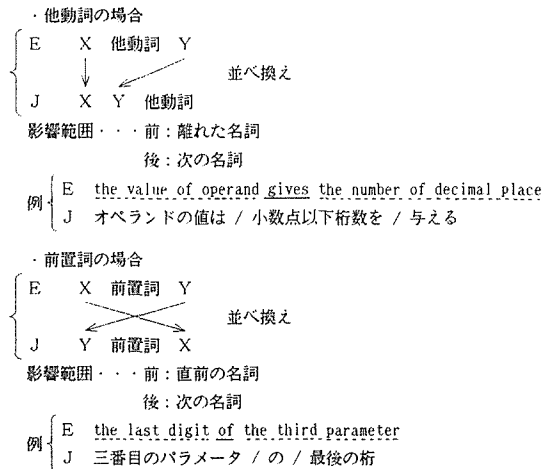
する。ただし E_0 の属するグループによっては $E_{Gb} \cdot E_{Ga}$ のどちらかがないことがある。

$E_b \cdot E_0 \cdot E_a$ の訳順のあと・さきは E_0 のグループと $E_b \cdot E_a$ のグループ（品詞でもよい）の組合せによって決定する。ある E_0 のグループに対する $E_b \cdot E_a$ のグループの組合せは複数ありうるが、訳順は E_0 のみに依存し、かつ一意的である。

たとえば、限定用法の形容詞を E_0 とすると必要な機能子は名詞であり、 E_a を E_{Ga} に属する名詞の中から選ばねばならない。また選ばれた名詞 E_a と形容詞 E_0 との訳順は E_0 がさきで E_a があと、と一意に決まる。

また別の例として、 E_0 に叙述用法の形容詞をとれば必要な機能子は be 動詞またはそれに準ずる動詞であり、 E_b は E_{Gb} の中から選ばれる。そして E_b は E_0 より前にあるが訳順は E_0 がさきで E_b はあとになる。これも一意に決定する。

さらに E_C から $E_b \cdot E_a$ を選び出すときの方法は単純で、 $E_{Gb} \cdot E_{Ga}$ それぞれの中の $E_b \cdot E_a$ になりうる機能子のうちもっとも E_0 に近いものが $E_b \cdot E_a$ になる場合と、もっとも遠いものが $E_b \cdot E_a$ になる場合があり、どちらであるかは E_0 の属するグループごとに決まっている。



Eが原文の機能子列全体、Jが並べ換えられた訳文で、X、Yは部分の機能子列である。
下線実線部が注目している単語で、点線部が直接影響される機能子の範囲である。

図3 機能の例。

図3にあげた例文を用いて説明しよう。

E the value of operand gives the number of decimal place

この例では E_0 として“gives”（他動詞）を選んでいる（下線実線部）。すると E_C は E_0 を除くこの機能子列全体であり、 E_b および E_a として“value”と“number”が選ばれる（下線点線部）。 E_0 として三番目の単語“of”（前置詞）を選んだ場合は次のようになる。

E the value of operand

E_C は上行に表示された機能子列のうち E_0 （下線実線部）を除いたものであり、 E_b および E_a と

して“value”と“operand”が決まる（下線点線部）。

同様にもう1つの“of”についても同様に処理できる。

E the number of decimal place

E_cは上行に表示された機能子列のうちE_o(下線実線部)を除いたものであり、E_bおよびE_aとして“number”と“place”が決まる（下線点線部）。

③ 分類

一般に、英文法において定められている品詞分類は、機械翻訳には必ずしも適当とは言えないことが多い。

たとえば、いくつかの名詞・形容詞が集まって一つの名詞句を成す場合、訳語としてはそれぞれの構成要素の訳の連続として与えられる。この場合、それぞれが名詞か形容詞かのどちらかであればよく、一方に決定する必要はない。特に技術系の文に多くみられる複合名詞句においては一方に決めにくいものも多いので、決定せずに済めば解析の負担を減らすことができる。

また別の例としては、単語“as”があげられる。“as”は直後に名詞がくるときは接続詞と前置詞の場合があり、直後に節がくるときは接続詞と関係代名詞の場合が考えられるが、いずれの場合も訳語は共通して「～のように(ような)」となることがある。このように、いくつかの品詞に分類される可能性があっても、機械翻訳の上ではある一つのグループにまとめられるという例が存在する。

一方、例えば形容詞というひとつの品詞に数えられていても限定形容詞と叙述形容詞では文全体に対する働きが異なる。したがってこの場合は機械翻訳の立場からは、品詞より細かな分類を行なわねばならない。

以上により、機械翻訳においては機械翻訳に向けた単語の分類が存在するとわれわれは考える。そして、本システムでは機能ごとに処理を行うので、単語を機能別にグループ分けしてそのグループに対してグループ名を与えている。機能で単語を分類すると、従来の品詞分類を組み替えたような分類が得られる。現在までに得られたグループ分けの一部を図4に示す。本システムの辞書では各単語はグループの名前で登録される。

- ◎従属接続詞群：従属接続詞
- ◎動詞群：助動詞・動詞
- ◎関係代名詞群：関係代名詞
- ◎修飾詞群1：現在分詞・過去分詞・形容詞（限定用法）
- ◎修飾詞群2：動名詞・前置詞
- ◎叙述詞群：現在分詞・過去分詞・形容詞（叙述用法）
- ⋮
- ⋮
- ◎名詞群：名詞

各行の「・・・群」とあるのが各グループ名を表わす。

その後ろの品詞は属するものを表わしている。

また機能子として機能の及ぶ範囲の広いものを上から並べてある。

図4 機能別グループを品詞名で表現したものの一部。

(2) 訳順確定

単語が入力されるたびに辞書引きしてその機能を調べる。既に入力されている単語（単語列）との訳順を検討し、矛盾がないように配置し表示する。先述のように複数の配置があり得る時は優先順位を調べ、高順位配置から表示してユーザーに選択を求める。

重要なことは、ある時点で入力している単語（機能子）の訳順を求めるために必要な機能子は、一般にその単語の前と後にあるということである。ここで「前」の単語は既に入力されているからその情報を用いることができるが、「後」の単語はまだ入力されていないので何の情報も得られない。したがって文末方向に機能子を求める機能は保留され、続けて単語が入力される毎に機能処理を行なうことができるかどうかを検討される。ここで、入力の途中で訳順が未確定であれば、単純に単語の出現順に訳語を並べておけば実用上は十分である。必要な単語が入力された時点で機能が働き、訳順が確定する。しかし最後の単語を入力し終わるまでは訳順が変わる可能性があるため、システム内部には入力中の文のそれまでに入力された単語の機能情報の一切が一文を入力し終わるまで保持される。しかしまた一方で、それまでにはユーザーからの修正を何回か受け入れることにより、構文上・意味上のあいまい性は排除されているので、最後の方で訳順が大きく変化する可能性は小さい。

最後の単語を入力して訳文が完成したとき、保持されていた機能情報のうち、ユーザーによって修正された情報の中で辞書に登録されていないものを登録した後、消去される。

(3) 辞書情報

本システムは単語辞書しか持たない。その辞書内の各単語のもつ情報は機能情報（訳語候補の一群とそれらの位置情報）であり、さらに各々は履歴情報を持つ。単語候補と履歴とは、一つの単語についていくつかの訳語の候補があるときは使用頻度が高い訳語から優先順位が高いものとして順番に表示され、それらの中からユーザーが選択できるようにするために使われるものである。この位置情報も履歴を持ち、ユーザーが訳語を選択するために用いられる。

このような単語の各々の情報はコンパクトにコード化されて辞書中に記録される。辞書からの読み出しは各単語の入力時であり、システムはその単語を見出し語として辞書を検索しコードを解読して訳語と位置情報および履歴情報を得る。辞書への書き込みは、後に述べる単語登録モードのときと、各単語の訳語とその位置がユーザーによって選択確定されたときであり、その情報が履歴とともに辞書に記録される。

(4) ユーザーの登録

ユーザーが行なう訳語選択・位置選択は辞書に登録されている機能情報そのものであるから、それを行なうことで辞書が変化を受けるのは履歴のみであり、既に登録されている選択候補の頻度が増えるだけである。

一方、ユーザーが訳語や位置情報を登録しようとするときは、それは機能情報に変換されてから辞書中に登録されなければならない。しかしユーザーが機能そのものについて知っている必要はなく、対話によりシステム側でユーザーの知識（品詞など）を引き出して判断し、機能に変換する。

4. おわりに

上述の構想に基づき、現在システムの構築を進めている。

さらに、このシステムで十分高速な処理が可能になれば、原文の入力と訳文の出力との間にほとんど時間差のないリアルタイム翻訳が可能であり、入出力の形式に文字データでなく音声を用

いることにより、マシンに通訳を行なわせることができる。このようにリアルタイム機械翻訳システムと音声認識及び音声合成とを結びつけることで自動機械通訳に発展ことを将来の課題としたい。

謝辞

種々討論していただいた講座各位に感謝します。また北海道工業大学の鈴木講師には研究を進める上での重要な示唆をいただき、当講座修士二年の吉川君からは商用システムに関する貴重なデータを提供していただきました。ここに厚く感謝の意を表します。

参考文献

- 1) 野武則『機械翻訳』, オーム社(1989).
- 2) 長尾真『機械翻訳はどこまで可能か』, 岩波書店(1986).
- 3) 沢村「形態素の連結による英日機械翻訳の試み」, 卒業論文(1987).
- 4) 沢村・宮永・栃内「品詞間の結びつきを用いた機械翻訳方式の提案」, 昭和62年電気関係学会北海道支部連合大会, 283 (1987).
- 5) 沢村「英日機械翻訳システム“TEMPLE”における訳順決定プロセス」, 修士論文(1989).
- 6) 沢村・宮永・栃内「形態素の多義性に注目した機械翻訳方式」, 昭和63年電気関係学会北海道支部連合大会, 270 (1988).