

25周年記念論文

機械翻訳における訳語選択と構造変換過程†

長尾 真†† 辻井 潤 一††

自然言語処理や機械翻訳の研究が活発化するにもなって、文の解析方法については多くの論文が発表されてきている。しかしながら、機械翻訳の中心的部分である言語間の変換の部分については、詳しく議論したもののはほとんどない。本論文では、科学技術庁の機械翻訳プロジェクト（Mu-プロジェクト）の日英言語変換の過程について、その設計方針、具体的な処理手順などを明らかにするとともに、変換過程で処理すべき言語現象、そこで使われる辞書の記述内容についても詳述する。Mu-プロジェクトの変換過程は、日本語と英語の構造的な差を調整する Pre-Transfer-Loop, Post-Transfer-Loop と、局所的な単語の使用環境を検査して訳語を選択する変換部の3つの段階から構成されている。変換部では、依存構造に従った再帰的な処理によって環境に応じた適切な訳語を選択するのに対して、Pre-Transfer-Loop, Post-Transfer-Loop の2つの段階では、両言語の文体的な差を調整するために、より一般的な木構造変換の規則が順次適用される。この両過程の存在によって、日本語の統語的な特徴を残さない自然な英文の生成が可能になるとともに、解析過程と生成過程の処理の分離度が良くなり、将来の多言語間翻訳への拡張も容易になっている。また、すべての処理を多層的な規則構成により実現することによって、解析の結果得られた情報を最大限有効に活用することができるようになっている。

1. はじめに

自然言語処理や機械翻訳の研究が盛んとなり、文の解析の方法については多くの論文が発表されてきた。しかし、機械翻訳の中心的部分である言語間の変換の部分については、詳しい内容を明らかにしたものはほとんどなく、この部分がどれだけ複雑な内容を持ったものであるかは、必ずしも明らかになっているとは言えなかった。

われわれは、かねがねこの部分の明確化について種の考え方や試みを行ってきたが、今回科学技術庁の科学技術振興調整費による機械翻訳システム^{1)~4)}の開発の機会を得たので、この機会にその考え方を大規模なシステムの上に具体化することを試みた。本稿では、その考え方と実際について報告する。

2. 変換過程の設計方針

機械翻訳の方式として、原言語の解析結果を『言語に依存しない』中間言語で表現し、これから相手言語の生成を直接行う中間言語方式もあるが、これには根本的な難点がある。まず、第1に、『言語に依存しない中間言語がどのようなものであるべきか』が明らか

ではない。特に、単語の意味を言語に依存しない形式で表現することは不可能であり、単語の訳語選択は翻訳対象の言語対に依存せざるを得ない。また、両言語に見られる慣用的・準慣用的表現の間の対応も、抽象的な『意味』のレベルでは、復元できなくなる。第2に、現時点の技術では、入力文の完全な意味解析（理解）を行うことは実際上不可能であるために、実用的な機械翻訳システムは、深い意味解釈が得られた部分については、より適切な訳文を生成し、これが得られない部分についても、それなりの訳文が生成できる必要がある。このようなことから、われわれは、原言語の解析と相手言語の生成の間に言語対に依存した変換の過程を設け、この部分に単語ごとの詳しい訳語選択の情報を与えること、および、両言語の発想の相違を調整するために言語構造の変換を行うことが、翻訳の質を高めるために不可欠であると考えている。すなわち、変換過程は、次の二つの仕事を行わなければならない。

(1) 訳語選択: 各単語が使われている周囲環境を検査することにより、適切な英語での訳語を選択する。

(2) 構造調整: 日本語解析結果は、原文の『意味構造』を表現するが、同時に日本語の統語的な特質も反映している。日本語と英語のように語族が離れた言語対では、解析結果の構造自体を変換しないと自然な訳文が得られないことが多い。このような言語対依存の構造調整を行う。

† Lexical and Structural Transfer in a Machine Translation System by Makoto NAGAO and Jun-ichi TSUJII (Department of Electrical Engineering, Kyoto University).

†† 京都大学工学部電気第2教室

われわれの変換過程は、日本語の解析結果をできるだけ中立の構造へと変換する Pre-Transfer-Loop, 大きな構造変換をとまなわない訳語選択を行う変換部, 中立的な構造を英語の言語的特徴を反映した構造に変換する Post-Transfer-Loop の3段階に分かれている。また、日本語の意味解析が完全には行われなかった場合に対応するために、変換過程での各処理は、

(1) 単語固有の規則の適用。最も優先度の高い規則で、慣用的・準慣用的な表現の対応を行う。

(2) 深層の意味に基づく規則の適用。

(3) 表層的・統語的構造に基づく規則の適用。深層の意味が解析過程で得られなかった場合にデフォルト規則として起動される。

といった多層の規則構成をとり、解析過程で得られた情報の範囲で、それを最大限活用することによって最も適切と考えられる英語での訳語や構造を計算するように設計されている。このため、日本語解析結果は、入力文の『意味構造』だけでなく、解析過程で得られた種々のレベルの情報(表層的, 形態的, 統語的, など)を残したものになっている。

以下では、まず、変換過程の中心となる変換部の処理について述べ、この前段階である Pre-Transfer-Loop の処理は、Post-Transfer-Loop と併せて7章以降で述べる。

3. 名詞中心の構造の変換

変換部の処理は、解析過程で得られた依存構造の頂点節点から出発して、その下位節点での処理を木構造にしたがって再帰的に起動することで実行される。各節点では、その節点の単語の品詞と上位節点からの拘束条件(名詞句として訳出せよ, など)から決まる部分文法(Sub-Grammar, 文法規則の集合で以下ではSGという)が起動される。図-1に、節点Xの変換処理が起動される時点での木構造を一般的な形で示す。各SGでは、次の処理を行う。

(1) 節点Xの訳語を、 W_i, Y, Z_j の各節点で与え

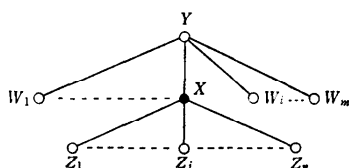


図-1 変換部での処理

Fig. 1 Data structures in the transfer phase.

表-1 名詞変換辞書の各欄の役割

Table 1 Roles of descriptions in the noun transfer dictionary.

A欄	その名詞に後接する付属語(格助詞)によって、英語での訳語が左右される場合に記述する。例に示したように、日本語での名詞が英語では副詞や形容詞に訳出される場合、あるいは、その名詞に後接する付属語(格助詞)が特定の前置詞になる場合に指定される。
B欄	その名詞に前接する付属語(格助詞)が、特定の前置詞になる場合に指定される。名詞が通常の動詞と同じように格構造を持ち、格要素の表層表現形がその名詞に固有である場合がこの欄で処理される。
C欄	D欄に記述できない、名詞と連体詞などの二語の間の共起関係による訳し分けを指定する。
D欄	名詞と動詞、名詞と形容詞などの(名詞+述語)の二語の共起関係による訳し分けを行う。日本語での二語の共起が、英語で一語の述語になる場合、特定の名詞との共起が、述語の訳語を決定する場合、特定の動詞との共起が名詞の訳語を決定する場合がこの欄で処理される。

られる統語的、意味的環境を参照して決定する。

(2) 下位節点 Z_j に訳出形などの拘束条件を与え、そこでの変換処理を起動する。

(3) 節点 X を中心とする構造の局所的な変換を行う。

節点 X が専門用語などの名詞の場合には、英語でも名詞に訳出され、なんの構造変換も必要ない。この場合には、変換辞書中の訳語を節点 X に付与し、無条件に下位節点の変換処理を起動するだけで処理は終了する。しかしながら、特定の名詞を含む文的な表現が、英語では形容詞1語で表現されたり、『名詞+格助詞』が副詞に訳出されることも多い。このような現象は、それぞれの語に依存する現象であり、個々の名詞の変換辞書のA, B, C, D欄と称する4つの欄に一定のフォーマットで指定されている(表-1)。

(1) A欄:『名詞+付属語(格助詞)』が特別な取り扱いを受ける場合に記述する。次の2つの場合がある。

(a)『名詞+付属語』が全体として形容詞・副詞のような別の品詞の語に変換される場合:

ex:『刺激性の気体』……>stimulative gas

ex:『対話形式で入力する』……>to input interactively

(b)『名詞+付属語』が『前置詞+名詞』になるが、前置詞の選択が名詞固有であり、一般の規則では誤る場合:

ex:『試みに～する』……>to (do) for trial

A欄には、次の4つの項目が記述される。

[解析結果に関する条件指定]

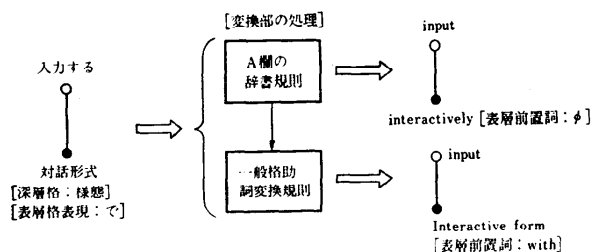


図-2 辞書規則の役割
Fig. 2 Lexical rules and general rules.

- 節点Xに付いた表層格表現
 - 節点XのYに対する深層格関係
- 【変換後の英語表現】
- 節点Xに付けられる英語の表層付属語（前置詞相当表現，空であっても良い）
 - 節点Xの日本語名詞に対する英語での訳語
- A欄の記述は，実行の時点では GRADE の辞書規則の形式に展開されており，変換部の文法は，節点XがA欄の記述を持っているかどうかに関係なく，この辞書規則を呼び出すだけで良い，辞書規則がなければ，一般の格助詞変換規則（5章）が適用されて，節点Xの英語での前置詞が選択される。図-2に，辞書規則がある場合とない場合の処理を示す。また，A欄からの辞書規則は，その名詞が修飾句をとまっているとき（ex:『機械との対話形式で入力する』）には，適用されない。

(2) B欄:『付属語+名詞』が英語で『名詞+前置詞』になるが，前置詞の選択が名詞に固有であり，深層格に基づく一般の規則では誤る場合に記述される。

ex:『教師としての資格』…>qualification for a teacher

(Cf. 「英語教師としての地位」→position as a teacher of English)

ex:『結果に対する影響』…>effect on the result

(Cf. 「家族に対する義務」→duty to one's family)

これらの例で，『として』，『に対する』が一般的な訳 'as'，'to' にならないのは，『教師』，『結果』が特殊なためではないので，これらの名詞のA欄には記述できない。B欄は，このように特定の名詞（『資格』，『影響』）が下位節点の訳出形を決める場合に記述する（図-3）。

(3) D欄: 名詞を含む文法的な構造全体が英語では1語の単語になったり，名詞が英語の動詞に語い化され

る場合をこの欄で取り扱う。

(a) 『名詞+述語』が英語で1語の形容詞になる場合:

ex:『光沢がある』…>lustrous

ex:『効率が良い』…>efficient

(b) 『名詞+述語』の結び付きが慣用的であり，英語では1語の動詞になったり，英語での慣用句になる場合，あるいは，その名詞と共起した場合には，動詞の訳語を特定の動詞にするべき場合:

ex:『～に焦点を合わせる』…>focus on ~

ex:『間隔をあける』…>leave a space

D欄の記述は，実行時には，『その名詞がどのような述語と共起した時に特別な処理を必要とするか』を示す属性（J-DIC-DERIV）と，実際に変換処理を実行する辞書規則とに変換されている（図-4）。この辞書規則は，変換処理の中心となる図-1の節点Xが述語の場合に，

(i) 従属節点Z_iで自分（節点Xの述語）との共起条件を持つ名詞があるかを属性 J-DIC-DERIV を見て検査する。

(ii) そのような名詞があれば，その辞書規則を起動する。

という手順で実行され，述語を中心とした構造全体を変換する（4および5章）。また，依存構造では，形容詞による名詞修飾も埋め込み文構造と同じ形式で表現されているので，

ex:『激しい雨』…>heavy rain

見出し語: 資格

表層形	意味関係	表層付属語	変換見出し
として	役割	FOR	qualification
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮

前接条件

条件部

変換後の構造

GRADE辞書規則

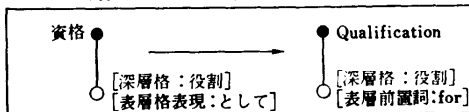


図-3 B欄の記述例と辞書規則

Fig. 3 Examples of lexical rules generated from the B-column.

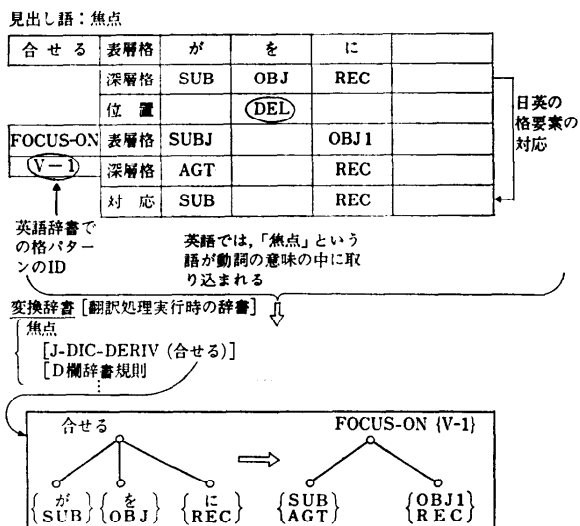


図-4 D欄の記述の例と辞書への展開

Fig. 4 Examples of lexical rules generated from the D-column.

といった、特定の名詞のための形容詞の訳し分けも、この欄からの辞書規則で取り扱われる。

(4) C欄: D欄の記述は、名詞と述語の共起関係を取り扱ったが、共起関係による訳し分けは名詞と名詞、あるいは、連体詞と名詞のような場合にも必要となる。

ex: 大きな差 ……>great difference

D欄からの辞書規則は述語中心の構造を変換する際に起動されたが、C欄の記述は、節点Xがこの欄の記述を持つ名詞の時に、辞書規則として起動される。

4. 述語を中心とした構造の変換

前述の図-1で節点Xが動詞・形容(動)詞などの述語である場合には、述語を中心とした変換処理が起動される。この処理では、

(a) 従属節点(格要素)の意味マーカによる述語の訳し分け

(b) 格要素の表層形式(名詞句・特定の前置詞をともなった前置詞句・形容詞補語・補文、など)の決定が、述語の変換辞書の記述にしたがって行われる。

日本語解析辞書では、同一の述語であっても意味の差が顕著な用法は、それぞれ別の格フレームに記述されており、変換辞書では、それぞれの格フレームに対して英語の訳語が用意される。したがって、顕著な用法の差による訳し分けの問題は変換過程では生じな

い。しかしながら、次のような場合がよく生じる。

(1) 日本語としてはそれほど顕著な差は感じられない場合でも、英語では別の語を使う。

ex: 『性能を上げる』

……>improve (performance)

ex: 『濃度を上げる』

……>increase (density)

ex: 『水面を上げる』

……>raise (surface)

これらを解析時点での多義とすることは、多言語間翻訳を考えた場合、不必要に解析段階に負担をかけることになり、変換過程で生じる訳し分けと考えた方がよい。

(2) 特定の名詞との共起が述語の訳語を決める場合がある。

したがって、変換過程においても述語の訳し分けが行われる。(2)の特定の名詞と

の共起による訳語選択(前章の名詞変換辞書のD欄からの辞書規則によって処理される)は、訳し分け条件としては最も優先度が高いため、まず最初に適用される(適用の手順は3章)。

(1)の現象に対処するために、述語変換辞書には、

(i) 深層格を埋めた名詞の意味マーカをチェックする訳し分け条件

(ii) 条件を満足した場合の英語での訳語

(iii) その訳語を選択した場合の日本語の格構造と英語の格構造の写像関係

が、図-5のフォーマットにしたがって記述されている。この記述は、実行時には、(i)を条件部、(ii)と(iii)を変換部に持つGRADE規則の集合(SG)の形に展開されている。このSGが述語の辞書規則として起動されると、それぞれの訳し分け条件に対応した規則が順次チェックされ、最初に適用可能となった規則にしたがって、訳語の選択と構造の変換が行われる。SGの最後の規則は条件部のない規則であり、その述語のデフォルト訳を与える。

述語の従属節点には、以上2種類の辞書規則によって訳出形が決まる必須の格要素のほかに、自由格、副詞句、従属節などがある。これらはそれぞれの統語カテゴリによって決まるSGを起動して変換する(5章)。

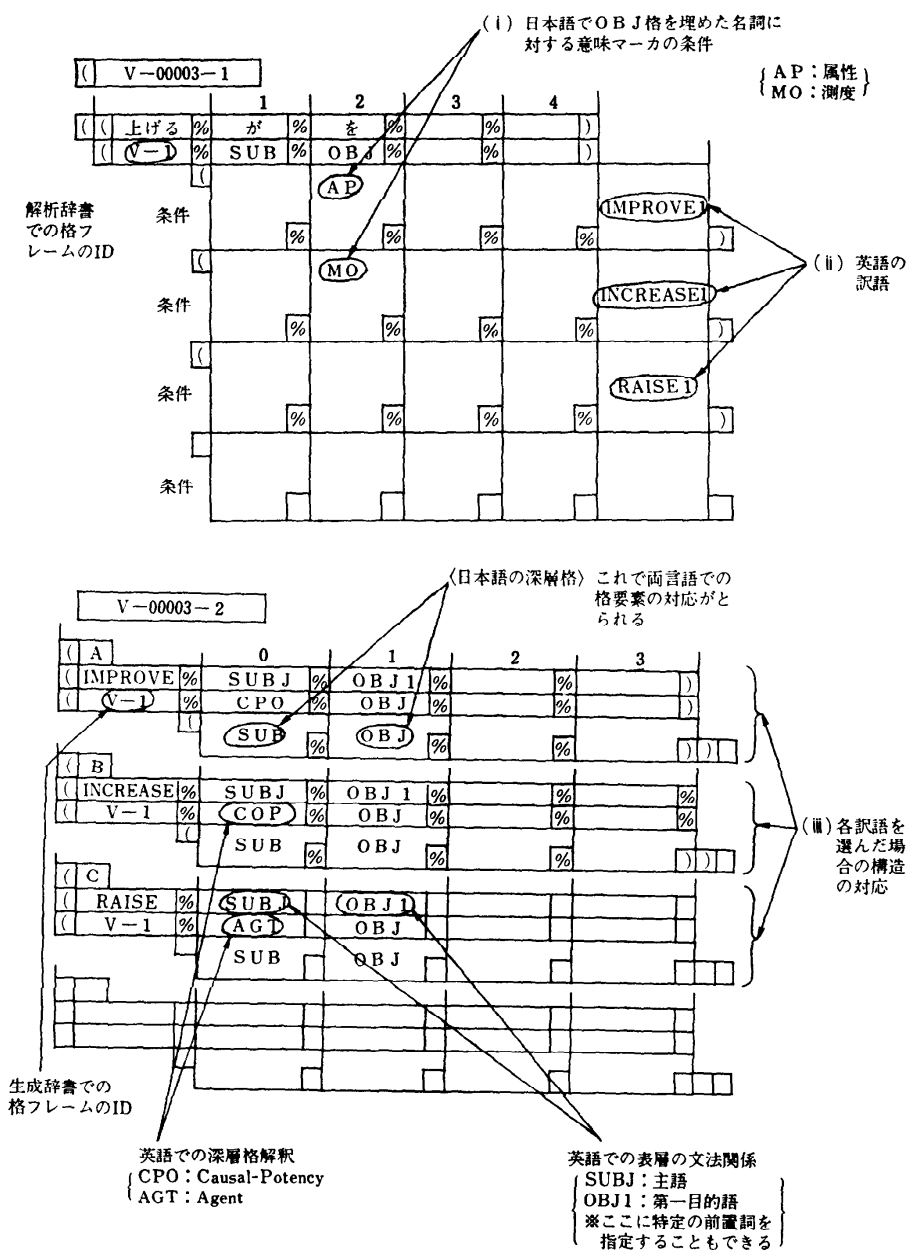


図-5 動詞フォーマット辞書の記述
Fig. 5 Description fformat for the verb dictionary.

選択する。

(ii) 深層格が一意に決定されなかった場合の規則: 表層格表現を参照し, その表層格表現に対応する最も意味の範囲の広い前置詞をデフォルト訳とする。

この2段階の構成は一般的な処理手順であり, 実際には, 個々の格助詞ごとに個別の変換規則が書かれている。特に, (ii)では, 解析過程での深層格解釈の限定の程度に応じてデフォルト訳が決定される⁹⁾。

以上のように, 連用格助詞をともなった名詞句の訳出形は, 優先度の異なる種々の規則の多層的な適用によって, 『解析結果の持つ情報を最大限活用して, 最も適切な英語表現を選択する』という原則で選択される。同様の手法は, 従属節・副詞句などの訳出形の決定にも使われている。

6. 個別な辞書規則による処理

われわれが変換辞書に一定の記述フォーマットを設定したのは, 大量の単語に対して能率良く辞書記述を行うためであって, この固定的な記述だけで変換過程のすべての単語個別な現象を扱う必要はない。記述フォーマットでは表現できない条件をチェックする必要のある単語については, 文法記述者が GRADE の規則記述の形式で直接辞書規則として定義しておけば, その規則が5章で述べた処理の各段階で起動される。

この単語個別の辞書規則の機能は, 非常に広い意味と用法を持つ動詞(『する』・『みる』・『行う』, など)の訳語を選択するときに使われている。また, 品詞は同じでも単語ごとの独自性が高く, 処理が一般化できない品詞の訳し分けにも活用される。

(a) 接頭語や接尾語の訳語選択: 『用』・『中』・『全体』・『製』といった単語は, 英語では, 前置詞・限定詞・形容詞・再帰代名詞などさまざまな品詞の語に対応し, 訳し分け条件も個別(例: 金属製 → made of ~, 米国製 → made in ~, IBM 製 → made by ~)である。

(b) 副助詞の訳語選択: 多くは英語の副詞に対応するが, やはり個別的な取り扱いが必要である(例: 『A, BなどのC』→ C such as A and B, 『A, Bなどを』→ A, B and so on)。

7. 構造の調整

変換部は, 木構造を頂点からたどりながら変換対象の節点の属性に応じた SG を再帰的に起動していく。したがって, 処理の順序は日本語の依存構造によって

あらかじめ決まっており, 再帰処理の途中で下位の構造が大きく変化することがあると, 上位の節点では対応ができなくなる。あるいは, 対応できるとしても, 再帰の結果得られる下位の構造を検査して, 適切な修復を行う規則をすべて用意しておかねばならず, 文法の構成を不必要に複雑にする。また,

『～だけを～するのではなく, ～も～する』

……> not only ~ but also ~

のような両言語での表現の対応を行うためには, 依存構造中での深い構造を参照する必要があるが, このことは, 各レベルの処理はそのレベルの処理に任せるという再帰処理の原則に反する(再帰処理が簡潔な記述になるのは, この原則に従って処理が記述できる場合である)。

このような再帰的な処理では対処しにくい大きな構造変換は, 変換部の前後に実行される Pre-Transfer-Loop と Post-Transfer-Loop で行われる。Pre-Transfer-Loop は, 日本語の依存構造だけで実行できる構造の調整を行い, Post-Transfer-Loop は英語での訳語などが判明してから構造の調整を行う。

この2つの段階では, 木構造中のどの節点に規則が適用できるかを判断するための木構造の走査は, 変換部のような再帰処理による走査ではなく, 個々の規則ごとに独立に行われる。ただし, 木構造の走査は, 規則ごとに規則適用の制御モードが指定できる GRADE の機能を使うことによって, GRADE 実行系が自動的に行ってくれるので, 文法規則の可読性を損なうことなく実現されている⁹⁾。

前述の例のように深い構造を参照しなければならない表現間の対応は, Pre-Transfer-Loop の段階であらかじめ見つけておき, これを木構造中に特別な節点でマークしておくことによって処理される(8章)。同様の手法は, 日本語の埋め込み文と英語表現を対応づける場合にも使われる(10章)。

8. Pre-Transfer-Loop での処理

この段階での処理は,

『Xが多い』…>『多くのXが存在する』

のように, 対応する英語の構造がない日本語の依存構造を整形するだけでなく,

(a) 英語の慣用的な表現と対応する日本語の構造をあらかじめマークしておく

(b) そのまま変換すると分かりにくくなる構造をより中立的な構造に変換する

「数値だけでなく文字列も処理する」

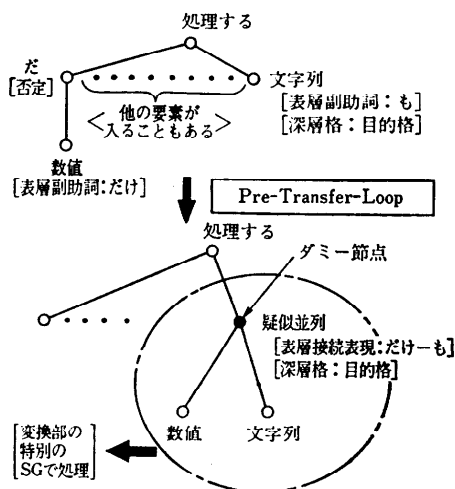


図-7 英語の慣用的表現に対応する日本語表現の処理
Fig. 7 Processing of idiomatic expressions.

ことによって、変換部の処理が複雑になるのを防ぐ。次に代表的な処理を示す。

(1) 両言語の慣用的表現の対応：前章で述べたように、変換部では、構造的に分散した表現間の対応を行うことは難しい。したがって、Pre-Transfer-Loopでは、変換部での処理を簡単にするために、英語の慣用的表現に対応する日本語依存構造の部分、ダミーの節点を挿入してマークしておき、変換部ではそれぞれの節点に応じた変換処理用のSGを起動する(図-7)。

(2) 日本語従属節の簡単化：日本語の従属節で、そのまま英語の従属節にすると分かりにくくなるものを簡単化する。次のような処理が行われる。

(a) 日本語の慣用的表現で英語にすると冗長、あるいは、意味不明となるもの

ex: 『～が原因(動機, etc)となって』…>by

変換部での名詞D欄の処理では、変換結果の英語の構造も節になることが前提である。したがって、上の例のように前置詞や前置詞相当の表現に縮退する現象は処理できない。この現象を引き起こす名詞の変換辞書にそのことを示す属性と英語での表現とが指定されており、変換部処理の前に、一般の格助詞と同じ構造に変換しておく。

(b) 日本語の従属節が英語では副詞や前置詞句に縮退する構造の変換

ex: 『注意して～する』…>carefully

このように、全く格要素を支配しない従属節で、かつ、従属節の主動詞が、『繰り返す』、『安定する』、『徹底する』など、英語で副詞、前置詞句で表現できる特定の動詞である場合に、副詞句の構造に組み替える。このような動詞は現在45語あり、そのことを示す属性と対応する英語の表現が変換辞書に記述されている。

(3) 文末にくる埋め込み文で、英語では助動詞相当や文接続の表現になる構造

ex: 『Xが～する傾向にある』…> tend to

ex: 『～する目的で～した』…> in order to

■特定の属性が指定された名詞が埋め込み文の被修飾名詞の位置にあることを見つけ、それぞれの語ごとのSGを起動して処理する。

9. Post-Transfer-Loop での処理

この段階の処理では、日本語の特徴を残した部分を英語的な構造に変える。次のような処理が行われる。

(1) 使役構造の削除：日本語では、

『Aを回転させる』…>‘make A rotate’

のように『自動詞+使役の助動詞』で他動詞的な概念を表現することが多い。‘make’の補文中に深層の目的格以外の格要素がなく、かつ、補文の述語に選択された英語動詞が自他両用である場合に、‘make’を削除する。ただし、‘make’の属性は、補文の述語(例では、‘rotate’)が引き継ぐので、『Aを回転させるために、～する』を‘in order to rotate A’に変換する場合のように、この構造がより大きな構造の一部でも良い。

(2) 自動詞構文から他動詞構文への変換：日本語に多い自動詞構文(『Aによって、BがCになった』)を、英語的な他動詞構文(‘A makes B C’)に訳出するために、英語動詞が自他両用の場合には、日本語で原因格や道具格の句の深層格解釈を英語特有の格Causal-Potencyに付け替え他動詞構文にする。

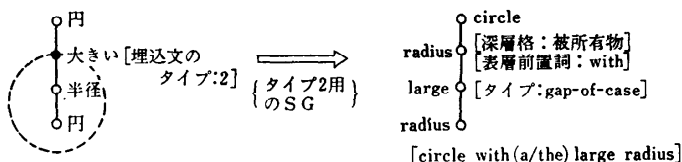
ex: 『データ処理の進歩で自動化船が増加した』

…>Advances in data processing increased the number of automated ships.

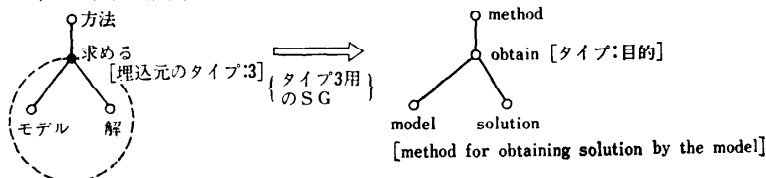
10. その他の構造調整

日本語の埋め込み文は、英語では、関係代名詞節、関係副詞節、名詞修飾の不定詞句、名詞修飾の前置詞句(for ~ing, of ~ing, など)、同格のThat節、分詞形(ING, ED)による名詞修飾、THATやWHATをともなった補文、などのさまざまな形式に訳出され

①「半径が大きい円」



②「このモデルによって解を求める方法」



③「長さを測定した結果を」

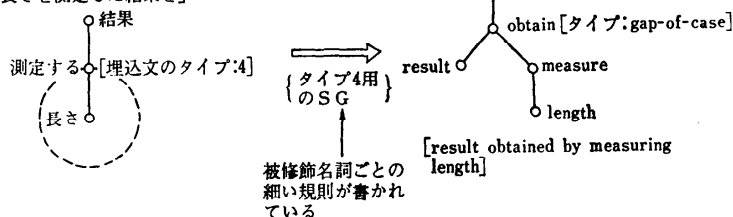


図-8 構造の調整一埋め込み文

Fig. 8 Structural adjustment for relative clauses.

る。日本語の解析過程では、日本語埋め込み文を、被修飾名詞句と埋め込み文との意味的關係に基づき4つのタイプに分類し、解析終了後の依存構造では、埋め込み文の中心となっている述語の属性として表示している。変換過程の変換部の処理において、変換対象の節点Xが埋め込み文の場合には、この属性によって別のSGを起動する。図-8に日英で対応のズレの大きい例を示す。この場合には、変換部で両言語の構造調整が行われることになる。このほか、Pre-Transfer-Loop, Post-Transfer-Loopの機能を活用すれば、たとえば、『～すれば、使用者は…することができる』を『～allows the user to do ...』に変換するといったように、両言語で大きく構造の異なる場合の構造調整も可能であるが、Mu-プロジェクトの対象とする論文抄録での使用頻度が少ないことなどの理由から、現在は規則としては入れられていない。

11. おわりに

現在(昭和60年6月)、Mu-プロジェクトの日英翻訳システムは一応完成し、大量テキストの翻訳実験を行っている。この実験では、本稿で述べたような詳細な記述を持つ変換辞書が、名詞13000語、動詞1600語、形容(動)詞700語について作られ使われている。

翻訳実験の結果、本稿で述べた変換過程の処理は、所期の目的をほぼ満足する形で動作している。特に、現在の変換過程の設計は、

(1) GRADE 辞書規則の機能を活用した規則の多層的な構成によって、新たな現象に対応して処理を追加したり、個々の現象についての処理をいくらでも精密化できる。

(2) 変換部の前後に一般的な構造変換を行うPre-Transfer-LoopとPost-Transfer-Loopの段階を

設けたことによって、両言語間の大きな構造の差をも調整し、自然な訳文にすることができる。

(3) (2)の両過程により、原言語の解析と目的言語の生成の段階の独立性が確保され、多言語間翻訳への移行が容易である。

という優れた特徴を持っている。今後、これらの特徴を活かして変換過程をさらに充実させるためには、以下のような研究を行ってゆく必要がある。

(1) 変換過程には、単語個別に対処しなければならない問題が、解析・生成の両過程よりもはるかに多い。実験を繰り返しながら個別現象に対応するとともに、より系統的で網羅的な辞書記述の作成が必要である。

(2) Pre-Transfer-Loop・Post-Transfer-Loopでの規則は、論文抄録に現われる顕著な現象に対処する規則のみが書かれている。この両段階での処理を充実させるためには、機械翻訳からみた日英両言語の対照研究を系統的に進めなければならない。

(3) 現在のシステムでは、単語の訳し分けに語の共起関係と名詞の意味マーカが使用されている。現在の意味マーカがどの程度有効に使えるかを実験的に検証するとともに、分野コードやシソーラスのような別種の情報がどのような場面で有効かを検討する必要がある。

(4) 単一言語での処理(解析, 生成)と言語対を意識した変換過程で行うべき処理の境界は必ずしも明確ではない。システムを多言語間翻訳に拡張することによって、これを明らかにする必要がある。

本稿のシステムは、Muプロジェクトに関係する多

くの人びとの不断的努力によって実現したものである。特に、システムのソフトウェアの実現は、中村順一氏(京大)、変換過程の基本設計は、西田豊明氏(京大)、実際の辞書作成は、坂本義行氏(ETL)、島海剛、佐藤雅之氏(JICST)にそれぞれ負うところが大きい。また、その他多くの人びとが京大でシステム開発を担当してこられた。記して、感謝する。

参 考 文 献

- 1) 長尾 真: 科技庁機械翻訳プロジェクトの概要, 情報処理学会自然言語処理研究会, 38-2 (1983).
- 2) 辻井潤一: 日本語構文解析, 情報処理学会自然言語処理研究会, 38-4 (1983).
- 3) 中村順一: 文法記述用ソフトウェア GRADE, 情報処理学会自然言語処理研究会, 38-3 (1983).
- 4) 坂本義行: 格構造を中心とした用言と付属語辞書, 情報処理学会自然言語処理研究会, 38-7 (1983).
- 5) 隅田英一郎他: 日英機械翻訳システムにおける日本語解析, 情報処理学会全国大会, 昭和59年前期論文集 (1984).
- 6) 谷口 敏他: 単純な名詞句及び並列名詞句の解析, 情報処理学会全国大会, 昭和59年前期論文集 (1984).
- 7) 日本科学技術情報センター他: 日英科学技術用語辞書データベースの開発, 昭和58年度科学技術振興調整費報告書 (1984).
- 8) 辻井潤一他: 日本語の自由格の解析とその英語へのトランスファ, 情報処理学会全国大会, 昭和59年前期論文集 (1984).

(昭和60年7月1日受付)