

日英翻訳過程での処理と その翻訳結果への反映

辻井潤一、長尾 真 (京大・工)

1 はじめに

電総研、工業技術院筑波計算センタ、日本科学技術情報センタ、京大長尾研究室の4者は、昭和57年度より科学技術庁の科学技術振興調整費の補助を受けて、科学技術論文抄録の機械翻訳システムの開発を行っている(MU-プロジェクト)。本プロジェクトの機械翻訳が、トランスファ方式によっていること等のシステムの基本方針、および、システムのソフトウェア的側面については、すでに報告している([1][2][3][4])ので、本稿では、日英機械翻訳の各ステップの文法的な側面について、詳細に述べる。

2 システムの概要

2-1 処理の流れ

MU-システムの翻訳は全体として

- (1) 日本語形態素解析過程
- (2) 日本語構文解析過程
- (3) 日英変換過程
- (4) 英語生成過程
- (5) 英語形態素生成過程

の5段階に分かれるが、(1)と(5)の過程は使われるソフトウェアも別であり、かなり独立度が高いので、本稿では(2)(3)(4)の各過程について述べることにする。この3つの過程は、それぞれ次の処理を行う。

〔日本語解析過程〕形態素解析の結果から入力文の句構造的な構造を計算し、さらに、この句構造を単語間の意味関係を表示する依存構造へと変換する。

〔日英変換過程〕入力文の依存構造表現をうけとり、英語での適切な訳語を選択し、言語対に依存する構造的な変換を行う。出力は、節点に英語の単語を持つ依存構造である。

〔英語生成過程〕英語の依存構造をうけとり、英語の句構造を作り出す。英語の文として読み易く、自然な文体を選択すると共に、英語として文法的な文の構造にする。

2-2 多重レベルの表現

日本語文を解析した結果は、基本的には格関係を表示する依存構造であるが、これを中心に、入力文の持っていた様々なレベルの情報が、(属性名 属性値)の集合の形式で依存構造中の各節点に付けられている(このように木構造中の各節点に(属性名 属性値)の集合の修飾情報の付いた構造は、翻訳の全過程を通じて使われる基本的なデータ表現であり、Annotated Tree Structure - ATS-と呼ぶ。また、属性名と属性値の対を、単に節点の属性と呼ぶ)。変換過程では、解析結果のATSについたすべてのレベルの情報を参照して、相手言語での訳語や文体を選択

する。これは、次のような基本的な考え方を実現するためである。

(1) 入力文の意味構造を唯一の解析結果とし、これをもとに相手言語の構造や訳語を選択するシステムでは、入力文の完全な意味解釈が得られることが必須となる。これは、現時点での自然言語処理技術では不可能である。実用的な機械翻訳では、1文中においても、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。

(2) 『言語に依存しない意味の表現形式がどのようなものであるべきか』は長期の研究テーマであり、この解決が得られない限り、純粋な中間言語方式はありえない。中途半端な意味や概念といった抽象レベルの構造に、文全体を写像してしまうことは、言語の表層が担っていた『意味』を失ってしまうことになる。

以下に日本語解析の結果のATS表示に使われる主要な属性を示す。

a. 深層格関係 : この属性が付けられた節点が、その上位節点に対してどのような深層格関係に立つかを示す。依存構造の中心となる属性である。通常の名詞句と述部の関係を表示する格関係の考え方を拡張し、文と文との関係や名詞(句)と名詞(句)の関係もこの格関係で表現する。現在、59個の深層格を用意している。

b. 表層格表現 : 深層格を表示していた具体的な格助詞(相当表現)や接続助詞(相当表現)が何であったかを示す。これは、

1. 同じ深層格でも、表層表現の差で訳出形にも差がでること

2. 深層格が完全に唯一に決定できない場合でも、表層の格助詞や接続助詞をもとにして、多くの場合正しいデフォルトの訳出形を決定できること

3. 例えば、『～のように～する』、『～のままで～する』、『～なしで～する』といった表現は、いずれも深層格表現では様態格になるが、変換過程では様態格一般の規則ではなく、個々の表層表現を参照して訳出形を決めなければならないこと

などから、変換過程でも重要な役割を果たす。

c. 表層副助詞 : 副助詞は、日本語にあって英語には直接対応する品詞のないものの典型であり、これを英語に訳出するには、副詞の使用や文体の変形を行う必要がある。副助詞の(言語によらない)深層の意味を表示することは、中間表現中で話者の心的態度等を表現することに繋

がり、現時点の自然言語処理の技術では不可能である。したがって、日本語解析結果のATSでは、その副助詞を伴った名詞句の中心名詞に、具体的な表層の副助詞がそのまま付けられている。変換過程では、各副助詞ごとに変換規則が用意され、直接英語の副詞や文体に変換される。

名詞、動詞、形容詞等の内容語（依存構造でも節点として自然に表現できる）以外で、機能語でありながら、この副助詞のように、表層語そのものを変換の対象としているものには、『～てしまう』、『～ようとする』等の法助動詞（相当表現）がある。一般に、内容語と機能語の境は連続的で、変換過程で表層の機能語そのものが必要になることが多い（例えば、『と』、『や』、『さらに』といった並列句を作る表現、『こと』、『の』、『ところ』といった補文標識、etc）。従って、解析結果のATSには、依存構造の節点にならない機能語についても、節点の属性として表層形を保存している。

d. 深層テンス、深層アスペクト： 相対テンスの日本語と絶対テンスの英語を結び付けるために、テンスの体系の変換を必要とするが、現在のシステムでは、日本語解析の最終段階で絶対テンスへの変換を行っている。また、『～ている』、『～つつある』、『～きている』等、日本語のアスペクト表現も、このままでは、完了形・進行形といった英語のアスペクト表現に移行することができない。したがって、解析の最終段階で、動詞のアスペクト素性と『ている』等のアスペクト形式素との組み合わせで、深層アスペクトを決定している。この深層アスペクトは、英語の表層アスペクト表現を選ぶだけでなく、『ため』・『のに』といった多義な接続助詞の意味を決定したり、格助詞の英語での訳出形を決めるときにも使われる。

e. 統語カテゴリ： 依存構造では、『計算機による処理』も『計算機によって処理する(こと)』も、基本的には同じになっている。しかし、相手言語での表現を選択する場合、とくに不都合がなければ、原言語での統語カテゴリをそのまま引き継いだほうが、余計な『意味の転移』を防ぐことができる。したがって、各節点より下の部分がどのような句構造的なまとり方をしていたかが、この属性を使って示されている。

f. 埋め込み文のタイプ： 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, etc.)、同格のTHAT節、ING形あるいはED形の分詞形による名詞修飾

等、様々なものに訳出される。この変換を系統的に行うために、解析時に、埋め込み文と被修飾名詞とがどのような関係にあったかを、次の4つのタイプに分類しておく。

1. 被修飾名詞が埋め込み文の格を埋めているもの
2. 被修飾名詞が埋め込み文の格を埋めている名詞句と意味的關係を持っているもの
(ex: 処理速度が速い計算機)
3. 埋め込み文が被修飾名詞の具体的な内容を示す同格節になっているもの、文と被修飾名詞の間に『という』が挿

入できるもの

(ex: 教師が生徒を殴ったという事件)

4. 埋め込み文と被修飾名詞とが何らかの意味的關係を持っているもの

(ex: 新しい計算法を理論値と比較した結果)

(ex: 誤りを適度に押える指針)

(ex: 魚を焼くけむり)

この分類は、変換過程で参照され、各タイプ別の変換規則が起動される。この種の分類は、完全な意味表現と統語カテゴリとの中間的なもので、変換過程の処理効率を高めるのに役に立つとともに、タイプ4のように完全な意味關係が復元できない場合でも、相手言語でそれに相当するあいまいな形式（例えば、of ~ing）を選択するのに役立つ。

g. トレースマーク： 埋め込み文のタイプ1・2において、被修飾名詞が文中のどの位置に入るべきかが判明すると、被修飾名詞の節点のコピーがその位置に挿入される。このコピーには、その節点がコピーであるというマークともとの節点の節点IDが付けられる。

h. 品詞細分類、名詞句意味マーカ、等： 以上のような解析段階で付与される属性の他に、入力文中の各単語の属性として日本語解析辞書に記述されていたものも、各節点の属性として付けられている。特に、品詞の細分類、名詞意味マーカ、用言の意志性にもとづく分類は、変換過程の訳語選択等で頻繁に参照される重要な属性である[5]。

3 日本語解析過程

3-1 解析過程の概要

解析文法を作成する際の基本方針を次に示す[2]。

(1) すべての解析結果を出力するのではなく、最も妥当であると思われる解析結果を1つだけ出力する。局所的なあい味さは、その部分だけの解析を複数個試みて、それらを相互に比較することによって、その段階で解消する。

(2) 慣用的表現や準慣用的表現等の個別的な現象に対処するために、GRADEの辞書規則の機能を積極的に活用する。

(3) 論文抄録という分野の特徴を積極的に活用する。ただし、文法は個々の言語現象ごとにまとめて部分文法に作成し、新しい分野に適用する場合には、関係する部分文法だけを修正すれば良いようにする。

図1に日本語解析の概略のフローを示す[6]。

3-2 日本語解析過程のステップ

(A) 形態素解析の後処理

この処理段階では、後の構文解析の規則を簡潔にするために、次のような処理を行う。

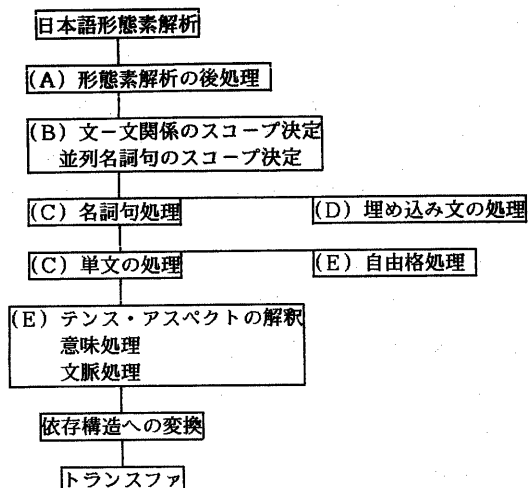


図1 解析過程の流れ

1. 1語としておいた方が良い語の連続をまとめる:

『150分の1』・『数表現+単位』等の数量表現、『高速計算機データ処理』等の名詞・形容動詞語幹の連続、格助詞・副助詞等の連続、が1語相当としてまとめられる。まとめられた表現は、以後の処理では、全体として1語のように取り扱われるので、相当する品詞やその細分類を与えておく。

2. 残しておくとの後の処理で不要なあい昧さを生じる語の連続を1つにまとめる:『美しく、知的な』のように述部が連続している場合、それが文並列の1部ではなく、述部並列になっていることの方がはるかに多い。したがって、この種の並列を解析の早い段階で1つの述部にまとめておく。これは、

『花子は美しく、知的な妹が好きだ』――>
『(花子は美しい) (知的な妹が好きだ)』

といった文並列の解釈を阻止することになる。

3. 周囲の環境を参照して、語レベルのあい昧さを解消したり、品詞の解釈を変えたりする:

a. 形態素解析では解消できないあい昧さ:英語ほどではないが、例えば、助動詞『だ』の連用形と格助詞の『で』、など日本語においても形態素解析において多品詞の解釈を受ける語がある。また、

『電総研で開発、京都大学で実験した』

のように活用語尾の省略があると、サ変名詞か動詞かのあい昧さが残る。これらの品詞のあい昧さを周囲環境を参照して解消する。

b. 品詞の付け替え:形容(動)詞の連用形は、副詞として使われることが多い。特に、形容詞の連用形は、連用

中止法と連用修飾句とで活用形に差がないので、普通に処理すると、例えば、次のような場合には不必要なあい昧さの原因となる。

雨が激しく屋根をたたく--->(雨が激しい)(屋根をたたく))

従って、状態形容詞の連用形について、句点の存在や後続文の形等によって、連用中止法の可能性が強いもの、および、格要素に入ると推定されるもの以外は、副詞とする。

『各種』・『全体』等の単語、および、『5回』・『10個』・『3分間』等の数量表現は、助詞なしで連用句を作る。したがって、これらが助詞を伴わずに生起している場合には、副詞相当の表現とし、『の』を伴っている場合は、全体を連体詞扱いにする。

また、名詞であるが接続助詞と同じように使われる『場合』・『時』・『結果』等の名詞を、周囲環境を見て、本来の名詞用法か、接続助詞の用法かを定める。

4. 各種の省略を復元する:格要素の省略のように、全体の解析が終了した段階で処理すべきものもあるが、述部の省略等は本体の処理の動きに大きな影響を与えるため、あらかじめ省略を補っておく必要がある。次のような処理を行っている。

『直径が5cmの円版』のように述部『である』が『の』に縮退する場合がある。『だ』の連体形活用は『の』を認めることもできるが、非常に多くのあい昧さの原因となる。したがって、

『～が<形容動詞語幹>の場合』(ex:パラメータが未知の場合)、
『～で<動作性名詞>中の<名詞>』(ex:本論文で考慮中の問題)、
『<属性の名詞>が<属性値の名詞>の<名詞>』(ex:長さが5cmの棒)

のように、局所的な環境から『の』を『だ』の連体形と解釈した方が正しい解釈になることの多い場合を検出している。また、

『～として～を、～として～を<動詞>』

『～は～、～は～である』

『～には～、～には～を<動詞>』

のように、周囲環境からみて述部省略が明らかなのを検出し、これを補う。

(B) 連用中止法、名詞並列句のスコープの決定

論文の抄録では、文の並列である連用中止法や並列の名詞句が非常に多くあらわれる。我々は、[7]で

1. 並列名詞句、連用中止法、埋め込み文のスコープが相互に重なり合って、組み合わせ的に解釈の可能性が爆発すること

2. 意味情報は、より妥当性の高い解釈を選択するには有効であるが、並列表現の解釈の可能性を制限するのには

使えないこと

3. すべての解釈を出力しそれを相互に比較するアプローチは、もう一度解釈を行うのと同じ手間がかかり、非現実的であること

を指摘し、

1. 名詞並列句では、同じ名詞や語構成中に同じ語基を含む名詞が並列されやすいこと
2. 『～と～と』、『～、～などの』、『～と～の関係の』のようにスコープの終端を示す典型的な表現があること
3. 連用中止法のスコープ決定には、意志・準意志・無意志・記述動詞などの動詞の分類が有効であること

等のような発見的規則を使うことによって、個々の部分の構造を解析する以前に、これらの並列構造のスコープを決定しておく手法を提案した。その後、このスコープ決定の部分文法は実験結果をフィードバックすることにより、つぎの各点が改良された。

1. 並列名詞句は、連用中止法と比べると狭いスコープしか持たないのが普通である。したがって、まず文レベルの並列のスコープを決定し、各スコープの中で並列名詞句のスコープ決定を行うことにより、無意味な解釈の可能性を制限していた。しかし、

『～する方法、～し、～する方法、～する方法を～』

のような箇条書き的な名詞句並列は、非常に大きなスコープを持つことがある。したがって、同じ名詞が繰返されるなどの非常に強い発見の手がりがある名詞句並列については、連用中止法のスコープ決定より先に処理することにした。

2. 純粋に対等な並列だけでなく、

『時間高調波級数ではなく固有関数級数で～』
『ライフサイクルだけでなく出荷数の予測も～』
『～をするばかりでなく、～もする』

といった各種の並列的表現（疑似並列と呼ぶ）も処理できるようにした。

2の処理結果は、通常の並列句と同じ依存構造になるが、接続表現が何であったかが節点の属性として記述されており、変換過程でこれを参照することにより、対応する英語での文体（ex: not only A but also B, not A but B, etc.）が選択される。

（C）単文、名詞句の処理

（B）のスコープの決定が終わると、各スコープの右側には主節あるいは従属節の述部があり、その左側には名詞句（とそれを修飾する埋め込み文）がある。各スコープ内の構造は、次の4つの部分文法ネットの再帰的な呼び出しにより決定される。

【メインの部分文法ネット】 左からスキャンして、最初の述部を見つける。それが埋め込み文の述部ならば、埋め

込み文処理の部分文法ネットを、主節の述部なら単文処理の部分文法ネットを呼び出す。述部がなければ、このスコープをぬけて、次のスコープ内の処理に進む。

【埋め込み文の部分文法ネット】 その処理の詳細は、次の（D）の項で述べるが、このネットは、連体形の述部より前の埋め込み文の部分と、被修飾名詞句の部分とをうけとり、それぞれ単文処理の部分文法ネットと名詞句処理のネットを呼び出して、

1. 埋め込み文によって修飾されている名詞
2. 埋め込み文のタイプ
3. 埋め込み文内の文構造
4. 被修飾名詞句内の構造

を決定し、全体を名詞句として呼び出した上位のネットに戻る。

【単文処理の部分文法ネット】 1つの述部のみを持ったスコープが、上位のネットから渡される。スコープ内を右からスキャンし、連用格助詞に囲まれた部分を切り出して、順次名詞句処理のネットへ引き渡す。述部より前の名詞句がすべて処理された時点で、述部の辞書を参照し、単文内の格関係を決定する。

【名詞句処理の部分文法ネット】 与えられた名詞句の構造を決定する。『の』を伴う修飾句はなるべく近い名詞を修飾することを原則とし、『動作性』の名詞は多くの句を支配する』等の種々の発見的規則によって、名詞句内での構造を決定する〔8〕。

この流れの主要部である単文処理の内容を次に示す。用言（動詞、形容詞、etc）の解析辞書には、標準形の場合の格フレームが、

1. 表層格マーカ（格助詞）
2. その格を埋める句の統語的制限（補文をとる場合の補文標識、形容詞の連用形が許されるかどうか、etc.）
3. その格を埋める句の意味的制限（主名詞の意味マーカ、あるいは、共起する名詞の範囲が非常に限定されている時は、表層の名詞そのもの）
4. 深層格の解釈

の4字組の集合の形式で与えられている。単文処理は、この辞書記述と実際の入力文とのパターン照合により行われるが、次の2つの問題がある。

1. 述部に含まれる助動詞的表現（使役態、受動態、可能態、難易態、etc）により、標準形の表層格マーカが変化すること
2. 1つの用言には複数の意味と用法があり、それに応じて、複数の格フレームがあること、さらに、日本語では格の省略が自由であり、連用句が現在処理中の述部を飛び越して、より後の述部に係る可能性を考えると、ほとんどすべての用法について複数個の解釈が局所的には可能であること

1に対処するために、単文処理の部分文法は、述部の形から入力文にどのような態の変換が生じているかを推定し、これを元に戻す逆変換操作を行ってから、辞書内の格

フレームと照合する。この逆変形操作は、受動態の場合には、一般に

『によって(に)』を『が』、『が』を『を』

なる格助詞の変換を行うことであるが、実際には、次に示すように、動詞によってこの逆変形の操作が変わる。

『業者が客から信頼される』

---->『客が業者を信頼する』

『財団が市から運営を委託される』

---->『市が財団に運営を委託する』

『彼が警官に免許を取り上げられる』

---->『警官が彼から免許を取り上げる』

『彼が幹部から実力を買われる』

---->『幹部が(彼の)実力を買う』

したがって、用言の解析辞書には、それぞれの用言がどのような態の変換をうけるかが、例えば、受動態の場合には14の種類に分けて指定されている。

2については、単文処理ではとりあえず、すべての用法を可能性を出力し、これらの中から、なるべく多くの格要素を支配する解釈を採用することを基本原則として、いくつかの発見的評価基準を使って、この段階で最も妥当と考えられる解釈を1つ選んで、次の処理へと進む。図2に単文処理の流れを示す。

1つの用法についての処理

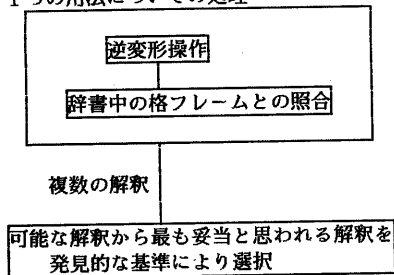


図2 単文の処理

(D) 埋め込み文の処理

2-2で述べたように、日本語の埋め込み文には4つのタイプがある。これを分離し、タイプ1・2については、被修飾名詞が文中でどの位置をしめるかを決定しなければならない。英語の関係節が統語構造上の明らかなギャップを持っているのに対して、格要素の省略が自由な日本語では、タイプ1に限っても、埋め込み文中で欠けている格の中で被修飾名詞によって埋められるべき格を決定するのは容易ではない。単純な処理では、非常に多くのあいまいさを生じることになる。しかしながら、実際には、次に示すように、各タイプごとに様々な発見の手がかりを使うことができる[9]。

[タイプ1]

1. 被修飾名詞が表層上『が』・『を』・『に』でマ-

クされる格を埋めることが多い

2. 述部が受動態の場合には、表層上『が』でマークされる格(深層の目的格)を埋めることが多い

3. それ以外の格を埋める場合は、被修飾名詞自体が格の名称を示すような特殊な名詞(ex: 原因、目的、動機、条件、方法etcを含む語)であることが多い(ex: この装置が正常に動作する環境条件)

4. 3の特殊な名詞の場合には、『～ための方法』といった補助的表現が挿入されることがある。

[タイプ2]

このタイプは、原理的には様々な形を採りえるが、最も多いのは、

<部分、属性、動作>が<状態性述部>[被修飾名詞]

の形式である。これ以外の場合には、このタイプの解釈を許すことは、あいまいさを増やすだけで、全体としての正答率を悪くする(ただし、<部分、属性、動作>は、この種の意味マーカを持つ名詞を示す)。

[タイプ3]

1. 『～という』・『～などの』等の補助的表現が埋め込み文と被修飾名詞の間に挿入される

2. このタイプを採りやすい一連の被修飾名詞がある(ex: こと、場合、点、必要性、状況、事実、方法、etc)

[タイプ4]

1. このタイプもタイプ2と同様に様々なものがあり得るが、良く生起するパターンのみに限定した方がよい。例えば、次のようなパターンがある。

『～を用いた<方法、行為、生産物>』

『～を考慮した<方法、行為、測度>』

2. このタイプを採りやすい一連の被修飾名詞がある(ex: 音、匂い、結果、etc.)

上に示した発見の手がかりは、常に正しい結果を与えるとは限らないし、また、手がかりにもかなり正答率の良いものから、低いものまで種々のものがある。また、

『温度を上げる方法は、うまく行かない』(タイプ3)

『温度を上げる一つの方法は、～である』(タイプ1)

のように、発見の手がかりが異なった解釈を与えることもある。従って、名詞句並列のスコア決定の場合と同様に、なるべく多くの場合に正しい解釈が出るように、制御の効いた規則適用を行っている。

E. 意味の解析: 以上では、語の係受け関係の解析を中心にして述べてきたが、これ以外に語や文の意味関係、テンス・アスペクト、省略の処理等が行われる。

1. 自由格: 動詞格フレーム中の格要素については、各動詞の辞書に深層格での解釈が記述されている。しかし、場所格・時間格・随伴格・様態格等の自由格については、各動詞の辞書には記述できないので、それぞれの格助詞ごとに、名詞意味マーカ・述部の素性(アスペクト、意志性、

等)を参照して深層格の解釈を絞る[10]。

2. 文と文との意味関係: 連用中止、接続助詞等の表す文の意味関係を、22の深層格関係に解釈する。比較的意思の明確な接続助詞も、

『のに』→逆接、目的; 『ように』→目的、様態; 『ながら』→時間の同時、逆接; 『ほど』→度合、比較

のように意味的には多義であり、それに応じて英語での訳出形が変化する。この解釈には、接続される2文の述部のアスペクトと意志性の有無が有効な手がかりとして使われている[11]。

3. テンス・アスペクトの処理: 動詞を状態・準状態・継続・瞬間・推移の5つに分類し、この分類とアスペクト形式素の組み合わせで述部全体のアスペクトを決定する。

4. 省略の処理: 格要素の省略、『は』の係先の決定等を行っている。

4 変換・生成過程

4-1 変換過程の概要

日本語解析結果として2-2で述べた日本語の依存構造が得られているので、これを出発点として対応する英語の依存構造を計算するのが、変換過程である。変換過程は、次の2つの仕事を行なう。

(1) 日本語の各単語がどのような統語的・意味的環境で使われているか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) 依存構造は、原言語の文の意味構造を表現するが、原言語の統語的な特質も反映しており、日本語と英語のように語族が離れている言語対においては、こ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も多い。このような言語対に依存した構造の変換もこの変換過程で行なう。

変換過程は、次の3段階に分かれる。

1. Pre-Transfer-loop: 解析で使われた属性のうちで、以降の処理で不要なものを消去したり、変換メインの処理を容易にするための属性の付与を行ったりするほか、日本語独特の構造で、英語にはそれに直接対応するような構造がないものを、より中立的な構造に変換する(ex: 『Xが多い』→『多くのXが存在する』)。また、

a. 『注意して～する』(cf: 『～に注意して～する』)

のように、日本語では述部として独立し従属節を作っているが、英語では副詞表現に縮退する構造、

b. 『～を中心(目的、手段、etc)にして』
『～が原因(動機、etc)となって』

のように、そのまま英語にすると冗長となる構造、

c. 『Xが～する傾向にある』

のように、英語では助動詞相当の表現に変換する必要がある構造

などが処理され、英語の依存構造に近い形に変換される。

2. 変換メイン: 依存構造全体の節点について、訳語選択とそれに伴う局所的な構造の変換を行なう。依存構造の頂点節点から出発して、その下位節点の変換を木構造に従って再帰的に起動してゆく。

3. Post-Transfer-loop: 英語らしい表現を作り出すために、英語の依存構造を多少手直しする。例えば、

『Aを回転させる』→ 'make A rotate'

では、使役態が使われているが、変換メインで選択された英語動詞(rotate)が自他両用の場合には、使役を外して直接他動詞の表現にする。

上記の1と3は、こなれた訳文を出すために、各種の準慣用的な文体を処理する段階である。ただ、この両段階は、かなり個別的な現象を取り扱うものであり、必ずしも必須ではないため、現時点ではあまり活用されていない。今後、より多くの実験を繰返すことにより充実させてゆく必要がある。以下では、2の変換メインについて述べる。

4-2 変換メインの処理

A. 名詞辞書を使った訳語の選択: 日本語の名詞の多く(特に、専門用語)は、英語においてもそのまま名詞に訳出される。しかしながら、日本語の名詞をその中に含む文的な表現が、英語では形容詞1語で表現されることもよくある。このような現象は、語に依存する規則なので、それぞれの名詞の変換辞書に登録されている。次のような情報が、名詞の変換辞書に一定のフォーマットで書かれている[12]。

1. 『名詞+付属語』が形容詞や副詞になる場合:

ex: 『刺激性の気体』→ stimulative gas
『木の椅子』→ wooden chair
『対話形式で入力する』→ to input interactively

2. 『名詞+付属語』が英語でも『前置詞+名詞』になるが、前置詞の選択が語固有であり、深層格にもとづく一般の規則では誤る場合:

ex: 『低温で殺菌する』
----> to sterilize at a low temperature

3. 『付属語+名詞』が英語でも『名詞+前置詞』になるが、前置詞の選択が語固有であり、深層格にもとづく一般の規則では誤る場合:

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

4.『名詞・動詞、形容詞』が英語で1語の形容詞になる場合:

ex:『光沢がある』→ lustrous
『効率が良い』→ efficient

5.『名詞・動詞、形容詞』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、その名詞と共に起した場合には、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 ~
『間隔をあける』→ leave a space

この5には、一般の木構造変換規則が書けるので、特定の名詞をキーにして起動したい任意の『日本語→英語』の木構造変換を指定できる。

B. 動詞、形容(動)詞辞書を使った訳語の選択: 日本語解析辞書では、たとえば表層上の格助詞の支配パターンが同じであっても、

『米国の大統領が日本の首相にあたる』
『車が壁にあたる』
『全員でこの困難な事業にあたる』

のように意味の差が顕著な動詞や形容詞の用法は、異なった格フレームとして分離されている。したがって、上のような用法のあい味さは、日本語解析の段階で解消されている(変換辞書は、それぞれの格フレームに対して用意される)。また、一連の特定の名詞に結び付くことによって、動詞の訳が決まる場合については、前述の名詞変換辞書の5の記述を使って処理される。しかしながら、

1. 日本語の意味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ことがよくある。例えば、

『性能を上げる』→ improve (performance)
『濃度を上げる』→ increase (density)
『水面を上げる』→ raise (surface)。

これらをすべて解析時点でのあい味さとするのは、不必要に解析段階に負担をかけることになる。特に、多言語間の翻訳を考える場合には、言語対を意識する変換段階のあい味さと考えた方がよい。

2. 上記のような訳し分けは、名詞のreplacabilityが高く名詞の辞書に登録しておくことは不可能である。

という理由から、動詞や形容詞の変換辞書には、解析辞書の各用法に対して、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とそれを満たした場合の英語での訳語の対が指定されている。

また、動詞・形容詞の変換辞書には、必須格に関して、日本語と英語の深層格間の写像関係が指定されている。各動詞・形容詞の英語生成辞書には、その語に支配される必須格がどのような統語的な形(主語、目的語、補語、特定の前置詞に導かれる前置詞句、etc.)を取るかが指定され

ているので、この2つの辞書によって、動詞・形容詞を中心とする単文の訳出形が定まる[12]。

C. 用法の広い動詞(形容詞)の訳し分け: Bの訳し分けは、一定のフォーマットで記述された辞書の情報を使って行われる。しかし、『する』『なる』『みる』のような、非常に多くの用法を持ち、用法に応じて異なった英語動詞に翻訳しなければならない動詞については、このような単純な辞書記述だけでは不十分である。これらの語に対しては、それぞれ個別の部分文法ネットを作成し、変換過程において、それを起動することにより、訳し分けしている。

D. 前置詞、従属接続詞等の選択: 英語動詞の必須格については、各格要素がどのような前置詞を伴うかが英語の生成辞書に指定されているので、変換過程では前置詞選択を行う必要はない。また、Aで述べたように名詞によって前置詞が決まることもある。これ以外の、いわゆる自由格の前置詞選択が、変換過程で行われる。ここでは、解析時に得られた深層格解釈・表層格表現・述部の素性・名詞意味マーカ・名詞自体(～の上で、など)を参照して決定する。深層格解釈が絞れていない場合にも、その他の情報から最も妥当と考えられる前置詞が選択される。文を接続する従属接続詞も、同様に選択される。ここで選択された単語や形式(分詞構文、等)は、生成過程ではデフォルトとして使われただけであり、生成時には、英語としての文体等を考慮し、他の単語や形式に変換されることもある。

E. 接辞、副助詞等の処理: 動詞、名詞といった主要な品詞の語は、両言語でも比較的同様な品詞の語で表現される。しかし、『用』『中』『全体』といった接辞や特殊な名詞などは、英語に翻訳する際に、前置詞、限定詞、再帰代名詞、形容詞等様々な品詞の語に変換する必要がある、個々の語に対して個別規則を用意している。

4-3 英語生成過程

英語生成過程は、変換過程から出力された英語依存構造を句構造表現に作り直すとともに、複数/単数の情報の付与、tense/aspect/modal表現の助動詞による具体化、主語と述部の数の一致、時制の一致、等の処理を行ない、文法的に正しい英語文を作る。また、文体面を考慮して、It-that構文やThere構文等を使用することにより、より理解の容易な英語文を生成する。この過程は、次の3段階に分けられる。

1. Pre-Generation段階: 依存構造の段階で行なった方が見通しの良い構造変換を行なう。例えば、日本語に多い自動詞構文

ex:『Aによって、BがCになった』

を、英語的な他動詞構文

ex: A makes B C

に訳出するために、原因格や道具格の位置にある句の深層格解釈を英語特有の格Causal-Actantに付け替える。ここで、『なる』『make』は自動詞・他動詞の代表であり、実際には、より具体的な動詞が入る。

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.

ただし、このPre-Generationの段階では深層格の解釈だけが変更され、実際に統語上の主語の位置に移動する等の構造的な変換は、つぎのGeneration段階で行われる。この種の変換の延長には、

ex: 『～すれば、使用者は～することができる』

--> ~ allows the user to do ~

のように、両言語で大きく構造を変えることもできるが、この種の規則の適用条件がまだ未調査であること、かなり分野依存の規則（例えば、マニュアル）で

抄録文には出現しないこと、等の理由から現時点では組み込まれていない。また、このPre-Generation段階の処理と前述のPost-Transfer-Loopでの処理は、いずれも同じ依存構造を対象にした規則であり、どちらの段階で適用するかに任意性がある。今後、日英以外の言語対に拡張することにより、最適な規則の配置を決定してゆく必要がある。

2. Generation段階 : 依存構造から句構造への変換を行なう。依存構造の頂点節点から下位の節点に向かって再帰的に変換の部分文法を適用してゆく。この段階は、ほとんど直線的な処理である。次の処理が行われる。

- a. delimiter の生成
- b. 生成辞書引き
- c. 必須格の生成と句構造での位置の決定
- d. 自由格の生成と句構造での位置の決定
- e. delimiter の消去とS-nodeの生成
- f. 受身変形、あるいは、代動詞の使用の決定
- g. 動詞句の処理

ここで、fの処理は、日本語での省略に対処するためのものである。すなわち、日本語で主語が省略されており、解析段階でそれが補われていない場合、英語で受動態を選択する。また、主語も目的語も省略されている場合には、述部の名詞派生形を主語として、代動詞を述部とする構造に変形する。

ex: 『ヘリウムで調べたことがわかった』

--> The examination is made by the helium and the following fact is found.

3. Post-Generation段階 : これまでの処理で一応英語の構文が生成されているが、このままでは、非常に読み難い英語になっていることがあるため、変形規則を適用してより自然な文体に変換する。主な規則を次に示す。

a. It-that変形 : 主語の位置の補文が長い場合に、it-that構文を使用することによって、読み易い英文にする。

b. 並列名詞句、動詞の等位接続の縮約 : 重文における重複する格要素や、並列名詞句における重複する修飾句を削除したり、代名詞化する。

ex:

uniform components and non-uniform components

--> uniform and non-uniform components

c. 関係節の外置変形 : 主語を修飾する関係節は、主語をやたら長くし、読み難い英文となる。このため、関係節の外置変形を行なう。

ex:

The parameter range in which ~ is determined

--> The parameter range is determined in which ~

d. 関係節の縮約 : 完全な関係節の中で、含まれる格要素の数が少なく、縮約しても判りにくくならないものを、名詞修飾の分詞句 (ING, ED) にする。

ex:

The result which was obtained by ~

--> The result obtained by ~

e. There構文の使用 : existを使った存在文をより自然なthere構文に変える。

f. 副詞、前置詞の位置の決定 : 2のGenerationの段階で前置詞句や副詞の標準的な位置は決定されているが、関係節で修飾されているような長い前置詞句を後置したり、あるいは、接続詞の副詞 (however, 等のConjunct) や名詞修飾の副詞 (even, only, etc.) の位置を最終的に決定する。

g. 比較級、時制等の調整 : この段階になると、finite形で出力すべき述部が決まっているので、重文、複文の時制の一致をとる。

h. 数、冠詞、人称の決定 : まだ決まっていない冠詞や数を決める。ここでは、前置詞句や関係節で修飾されている名詞の場合は、theをつける等の発見的な基準を使う。

i. インターフェースのための変換 : 形態素出力のために木構造を線形化し、英語での活用形に関する情報を付与する。

5 おわりに

現在、Muプロジェクトの日英翻訳システムは、京都大学での1000文実験とその評価を終了し、サンプルを3000文に拡張しての実験に入っている。3000文の実験は、工業技術院筑波計算センタに移植されたシステムで行われており、現在その第1回目が終了した段階である。その結果をフィードバックするのが、今後の課題である。

なお、本稿の日英システムは、Muプロジェクトに従事している多くの人々の不断努力のたまものである。記して、感謝する。

[1] 長尾: 『科技庁機械翻訳プロジェクトの概要』、情報処理学会、NL研、38-2, 1983

[2] 辻井: 『日本語構文解析』、同上、38-4, 1983

[3] 西山: 『変換過程の基本設計』、同上、38-5, 1983

[4] 中村: 『文法記述用ソフトウェアGRADE』、同上、38-3, 1983

[5] 坂本: 『格構造を中心とした用言と付属辞書』、同上、38-7, 1983

[6] 梶田、他: 『日英機械翻訳システムにおける日本語解析』、情報全大昭和59年前期、1984

[7] 石川、他: 『科学技術論文における並列句とその解析』、情報処理学会、NL研、38-4, 1983

[8] 谷口、他: 『単純な名詞句及び並列名詞句の解析』、情報全大昭和59年前期、1984

[9] 小坂、他: 『日本語理め込み文の解析とその英語へのトランスファ』、同上、1984

[10] 辻井、他: 『日本語の自由格の解析とその英語へのトランスファ』、同上、1984

[11] 加藤、他: 『日本語単文間の関係の解析とその英語へのトランスファ』、同上、1984

[12] 日本科学技術情報センタ、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