

6J-2

日英翻訳システムALT-J/Eにおける
日本語解析技術

宮崎正弘 白井 諭 林 良彦

NTT電気通信研究所

1. はじめに

現在、開発中の日英翻訳システムALT-J/Eにおける日本語解析技術について報告する。本システムの日本語解析処理は、解析の各段階で生じる多義を効率的に絞り込むこと、多様な語いと表現をもつ文を取り扱えることに重点をおいて設計されている。以下、本稿では、上記のような点をいかにして達成するかということに的を絞り、本システムの日本語解析手法について述べる。

2. 解析の多義とその解消策

すべての自然言語処理は、与えられた表現から目的とする情報を抽出する過程もしくはそれらへの変換の過程で生ずる解釈の多義性の中から、いかにして正しい解釈を選び出すことができるかによってその解析精度が決まると言ってもよい。特に、多様な語いや表現を含む文を高い精度で解析し、きめ細かい訳文を生成するためには、文解析を行うための規則や辞書の収録語数が増大し、形態素解析や構文解析など解析の各段階で生ずる多義が急増する。このことは、処理時間を増大させる。一般に、自然言語処理においては、精度向上を狙いとした処理の精密化、辞書規模の増大と、処理の高速化は相反する面をもっており、この両者の矛盾をいかに解決するかという点が実用上、重要である。本システムの日本語解析においては、解析の各段階で生じる多義を以下のように効率的に絞り込んでいる。

- (1) 解析の各段階で生ずる多義をできるだけ早期に絞り込むため、形態素解析などの浅いレベルの解析にも、必要ならば部分的に意味を取り入れた解析を導入する(構文・意味融合型の解析)。
- (2) 多義を絞り込むために効果的なヒューリスティックルールを導入する。
- (3) 文を単文、句、文節など階層化された小さな処理単位に分割して、各処理単位内でローカルな解析を行い、その結果を上部の処理単位に引き継ぎ、よりグローバルな解析を行う。
- (4) 構造のもつ意味¹⁾などでローカルな多義を成長させないため、ローカルな解析では解消されない単語の多義(意味属性など)、構文上の多義(用言の係り先など)などは、できるだけ個別

に展開せず、縮約して文構造を生成し、よりグローバルな解析でこれらの多義を絞り込む。

3. 多様な語いと表現をもつ文を扱うための二段階の標準変換

(1) J-J変換による等価的変換

日本語文は、明確な正書法が確立されていないため、表記上のゆれがある。また、同様の意味を異なった表現で表すことも多い。そこで、本システムでは、英語に訳し分けられる範囲の表現として標準表記を設定し、標準表記以外の形式で書かれた表現を、辞書、ロジックによって、標準的な表現に縮退させるJ-J(日本語-日本語)変換を行う。以上により、表記上、表現上の多様なゆれを英語に対応させて吸収することができる。

(2) 主体的表現と客体的表現の分離による表現の簡略化¹⁾

助詞、助動詞、補助用言、および、それらの組み合わせで表現される主体的表現を文中から抽出し、これらの表す様相、時制の属性を決定し、主体的表現を陽には含まず、これらの表す属性のみ付加された文に変換する。これにより、文を客体的表現を表す単文の組み合わせに分割し、単文解析や日本語→英語への変換を行うパターン変換規則²⁾の適用を容易に行うことが可能となる。

4. 日本語解析処理の概要

従来、トランスファー方式など、多くの日英翻訳システムでは、解析と変換は独立しており、入力文は解析部で均一な深さの解析を行った後、変換が行われる。これに対し、本システムでは図1に示すように解析と変換が融合しており、単位文内で用言とその修飾要素を認定するための解析と同時に、単位文の変換を行い、英語構造の骨格部を決定する²⁾。以下、日本語解析処理部の概要を述べる。

(1) 分ち書き処理

最長一致法における解析精度の問題と総当たり法における処理速度の問題を克服するため、文を階層化された小さな処理単位に分割し、各処理単位で総当たり法の解析を行う局所的総当たり法³⁾による解析を行う。なお、複合語の分割には、係り受け解析を導入し、単語間の意味的結合関係を

明らかにするための解析⁴⁾を行う。以上により、一般語の他に固有名詞や数詞を含んだ漢字かな混り文を高い精度で高速に自動分割することを実現している。

(2) 係り受け解析

文節の文法的機能に着目して、マトリックス型で書かれた係り受けルールを基に文節間の係り受けを定める。ここで、係り受けの多義が爆発的に増えることを抑止するため、用言の必須格となりうる助詞（格助詞相当の連語を含む）、陳述副詞の係り先の属性上の制約などの情報を用いる。なお、係り受けの多義はヒューリスティックルールを用いて縮約を行うとともに、最尤候補から順に出力する。

(3) 単文抽出処理

係り受け解析結果に基づき、文から用言を一つだけ含む部分を単文として抽出する。

(4) 単文解析処理

単文中の名詞句の構造を解析するとともに、述部の様相、時制の属性を決定する。なお、用言とその修飾要素（格要素、副詞要素）の認定は、単位文変換処理²⁾で行う。

(5) 埋め込み解析処理

用言（連体修飾）の係り先を決定すると共に、受けの文節との関連を解析する。

(6) 接続解析

用言（連用修飾）の係り先を決定すると共に、単位文間の接続の属性を決める。

なお、日本語解析の構文上の多義は、最適結果選択処理²⁾で解消される。

図2に日本語解析部の処理例を示す。

5. おわりに

今後、日本語解析処理の機能、性能の評価、ルールのチューンを翻訳実験をとおして行っていく予定である。

<参考文献>

- [1] 池原他：日英翻訳システムALT-J/Eの設計思想、第33回情処全大(1986)
- [2] 林他：日英翻訳システムALT-J/Eにおける日英変換技術、第33回情処全大(1986)
- [3] 宮崎、大山：日本文音声出力システムの言語処理、通研実報、Vol.35, No.2, PP.157-167 (1986)
- [4] 宮崎：係り受け解析を用いた複合語の自動分割法、情処論、Vol.25, No.6, PP.970-979(1984)

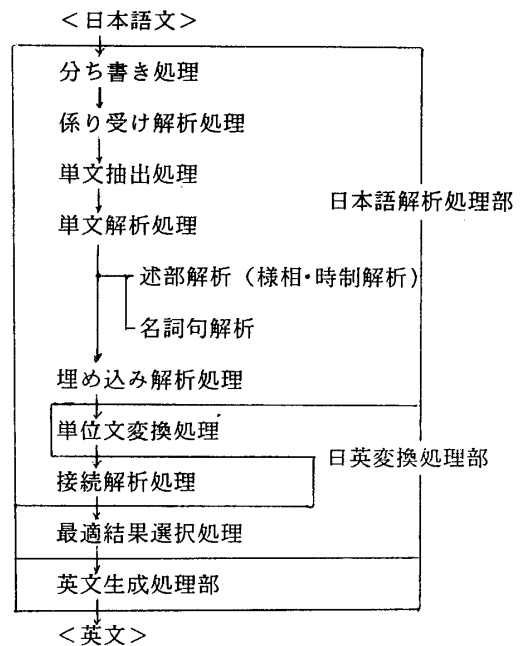


図1. 日英翻訳システムALT-J/Eの処理概要

翻訳例文

その技術をものにしたい鈴木君がああ難しい仕事を
請け負わないことは問題だ。

分か書き処理 /形態素境界 //文節境界

その//技術//を//もの//に//し//たい//鈴木//君//が//あ//
難しい//仕事//を//請け負わ//ない//こと//は//問題//だ//。

係り受け解析 (最尤候補)

その 技術を ものに したい 鈴木君が ああ 難しい 仕事を 請け負わない ことは 問題だ

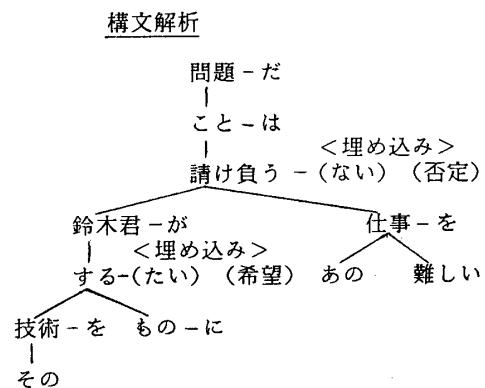


図2. 日本語解析部の処理例