自然言語処理　40－5
（1983. 10. 21）

機械翻訳における校正機能

岡島　惇　新田　義彦　山野　文行
（㈱）日立製作所システム開発研究所）

1．はじめに

近年，機械翻訳の研究，開発が盛んに行なわれ出して来たが，校正機能に対する考え方については，人によりかなりの差異が見られる。比較的校正が楽と考えられる欧米語間においてさえ，Martin Kay [1] のように，マンマシン・システムとしてしか機械翻訳は成功しないだろうという立場に立つ人と，Peter Toma [2] のように公表すべき文書以外は，自動翻訳がほぼ可能だろうという人もいる。

我々は，「入力に制限を加えず，直訳に近くとも良いが，意味が分るレベルの翻訳」を行なう際に，どのような校正機能が有効であるかについて考察する。又その一例として，我々が実験的にATHENE-82*) [3] に実現した機能 [4] についても述べる。

2．校正機能

機械翻訳における校正機能としては，一般的に，

（1）プレ・エディット
（2）インタ・エディット
（3）ポスト・エディット

の3つに大別できるが，Kayと Tomaの考え方の違いは，主に（2）に対する意見の相異と考えられる。人の介在なしで機械がどこまで翻訳できるかという点と，校正者が何をなすべきかにより，人と機械の役割が変ってくるため，校正機能全体を一まとめに議論するのは難しい。以下では，上記の分類に従い個々検討する。

2.1 プレ・エディット

プレエディットで処理すべきものは，

（a）個別に翻訳しても正しい結果が得られないイディオマティックなもので，汎用辞書に入れるというほどでもないもの
（b）構文解析に際して多義性が生じると考えられるもので，人間にとっては，簡単に指示できるもの

等が考えられるが，翻訳の前に人間が作業をしても見合うだけの効果が上らないと意味がない。例えば，上記（a）についても，その出現がまれであれば，わざわざ辞書（この場合，一時的な辞書であっても）に入れる手間を考えれば，その文を翻訳しないという方法もあり得る（その場合は翻訳率が下ったことになるが，本質的には変りはない）。この延長線上には，全体を訳さないで，機械でやるべきものと，人間が訳すべきものとを切り分ける作業を行なうことが考えられるがその作業が有効なのは，出力がかなり高品質のものの場合であろう。大意が分れば良い程度の出力で満足な場合は，上記の選別の工数が負担になる場合もある。

（b）は，構文解析で何が難しいかによる。一般的に考えられる多義性で構文解析の難しいものは，

*) ATHENE-82：Automatic Translator of Hitachi for English into Japanese with Editing support－'82

(i) 句(節)の範囲
(ii) 並列句
(iii) 動詞を含む格構造
(iv) 修飾とそのタイプ
(v) 訳語の多義
といったものがある。以下，議論を明確にするために，ATHENE-82の対象とする英和を例にとって話を進める。

上記の多義性の例をあげる。

(i) 句(節)の範囲

(文1) The man said (it is useful to use machine translation system) at the conference.

(ii) 並列句

(文2) John <invited Mary and Sara>, and <had a great party>.

(iii) 動詞を含む格構造

(文3) The man [ENV] elected the chairman proved to be agreeable.

(iv) 修飾とそのタイプ

(文4) They rent a house 【ADV】 to live in Tokyo.

(v) 訳語の多義

(文5) President said," Yes."

上記の内，プレエディットで処理が可能であると考えられるのは，(i)～(iv)であり，その入力方法の一例は各文例に示してある。自然語文には，構文的に見れば「本質的にあいまいな」文が存在する為，プレエディットが有効とも考えられる。又，局所的な解析のみでは，バックトラックが必要な場合が多く，かつ，その個数は文が長くなれば，それに比例して増え，さらに，ある時点で解析が間違っていると判定すること自体難しい場合もあることを考えるとプレエディットの必要性も否めない。

しかし，ATHENE-82に上記の機能をインプリメントした結果では，あまり有効とはいえないという結論が出た。これは，次のような理由による。

① どんなことが機械にとって難しいかをユーザが認識するのが難しい。
(何を指定するか)
② 指定の方法が複数個あったり，あいまいになってしまうことがある。
(どうやって指定するか)
③ 指定しても，うまく解析してくれるとは限らない。
(システムのレベルとの相関)
④ 必要な時，機械が聞いてきてくれる方が楽そうである。
(インタエディット，ポストエディットとの比較)

2.2 インタ・エディット

インタ・エディット方式は，機械翻訳システムを作る側から見ると考えやすい。ただし，この場合，ユーザとの関わりからは，次の2つに大別される。

(a) 構文解析の中間結果を表示して，ユーザが用意されたコマンド群を使って構文解析をコントロールする(ユーザ主導型)。
(b) 構文解析部(パーサ)が解析中に不明の点が出て来たら，ユーザに質問を発する(パーサ主導型)。

(a) の方式で，中間結果として考えられるのは，構文解析木が適当である。英語と日本語との間で品詞のずれが生じたり，イディオム的表現のある場合には，構文解析木を表示するのみでは不十分の場合もあろうし，逆に構文解析木に多義性があっても，和文は同じということもあり得る（副詞や，前置詞の場合）が，一般的には，木が正しくできれば，その後は，問題は和文生成側にあり，和文生成におけるエラーは，和文エディタによるのが素直である。

構文解析木の表示には，いくつかの問題がある。インタエディットを行なう場合，CRT に表示をし，これを修正する方法が望しいが，その際には，画面の表示可能サイズも問題になる。例えば，図1に示す例文は，短い部類に属するが，これを一画面で表示するには，65行程度は必要であり，さらに長い文では，スクロールも必須となる。図1は，我々が提案している擬似句構造表現 QPS (Quasi Phrase Structure) [5] による表示であるため，木の段数が少なくてすんでいるが，通常の句構造表示をすれば，6段にもなり，左右のスクロールも必要となる。

木の多義性を解消するには，図1の表示後，修正用のコマンドをユーザが入れることになる。例えば，名詞の同格とされた "the idiosyncracies of speech" が，もし "grammar" でなく，"the principle" にかかるとすれば，「LINK 18, 16」（18番目のノードを現在の修飾先から16番目のノードへ修飾するようにつなぎ変える）といったコマンドとなる。

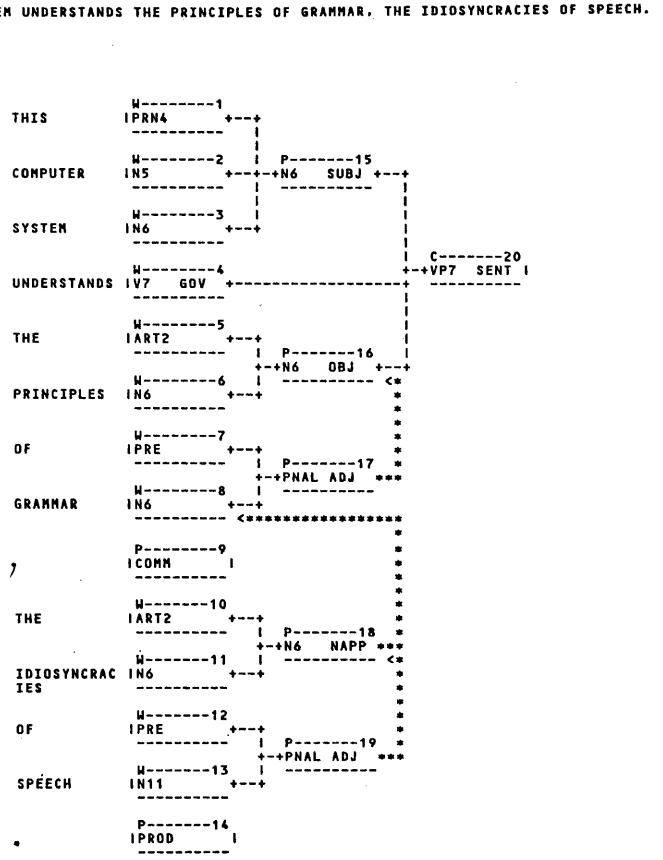

図1. ATHENE-82 における解析木表示例

解析木変更に関係するコマンドには，次のようなものがある。

(1) 親子関係の変更：必須格的木の変更
(2) 修飾関係の変更：任意格的修飾関係の変更
(3) ロール（格，修飾）の変更：主語，目的語や従属文の意味変更
(4) 構文情報の変更：特に品詞の転用がある場合の品詞の変更等

これらは，独立に行なわれるのではなく，例えば，「yesterday」といった名詞が，目的語に取り込まれているような場合には，これを親から切り，修飾語として，

動詞等にかけ，ロールと品詞も変更するすることになる。その内のいくつかは，システムが，めんどうを見ても良いが，その場合は，実質的に再翻訳となる。（4）の品詞等の変更の際には，他の品詞を表示するには，ウィンド機能を持った端末が望まれる。さらに，それらの情報は，できれば，端末側に置くのが良いが，品詞や属性の変更は，基本的に再翻訳になることを考えると，端末単独で翻訳ができるようになるまでは，機能の分散は，制限されていても良いかもしれない。

ATHENE-82では，各ノードの情報をCRTに表示し，品詞の変更をユーザが指示すると再翻訳を行なう方式を取っていたが，手間がかかりすぎることが分った。再翻訳が必要なくらい解析木がくずれている場合は，原文に戻って，プレ・エディット的処理をするのが得策かも知れない。

（b）のパーサ主導型のインタ・エディットでは，かなりの注意が必要である。それは，パーサが，構文解析を行なう場合，各種の判定を行なう回数は，制限文法による入力でない限り，かなりの数に登るからである。インタ・エディットをすることにより，かえって工数が増すことのないようにする為には，ユーザに対して発する質問の回数を極力減らす必要がある。その為には，システム自体の多義解消能力を高めておく必要がある。例えば，Winograd[6]には，前置詞7つで，429個の構文的多義がある例があげられているが，この内いくつをシステムが解消できるかが実用化の大きな鍵となる。

インタ・エディットの質問の種類は，プレ・エディットと同様のものとなるが，プレ・エディットに比べて，指定方法が限定されるため，ユーザにとっては楽であり，かつ，パーサが解析できない文についてのみシステムが質問を発することになるため，作業量も減ると考えられる。

2.3 ポスト・エディット

ポスト・エディット機能は，機械翻訳の利用のされ方によって異なる。Machine Aided Human Translationならば，ほとんどワード・プロセッサに辞書引き機能を付加した程度のものでも許されようし，Human Aided Machine Translationならば，直訳といわないまでも，原文側の構文解析結果を反映した校正方式が一般的である。

ポスト・エディターは，翻訳者か校正者かの議論は，さておくとしても，ポスト・エディットには，翻訳とは異なる面が多いようである[7]。
その原因の1つは，原文の構文解析結果を尊重するためであり，その中間語が，原文側に傾きがちであるため，直訳調になりやすく，人間の翻訳と離れてしまうためと考えられる。これを救う方法は，言語に依存しない汎用的中間語で原文を表現することだと考えられるが，言語表現を完全に論理のみで表わすことができないことと，イディオムのように，言語対で変換規則を書いた方が楽なものもあることを考えると，我々は当面，「イディオム＋直訳調」を基本に校正機能を検討するのが良いと考える。

原文の構文解析結果を尊重する立場に立つ時，注意を要するのは，校正を解析木のレベルで捉える場合と，一次元の文字の並び（字づら）で捉えるかという2つのフェイズがあることである。

翻訳の過程が，木構造の変換をベースにしている間は，基本的には，木の各ノード内の字づらの修正は，木構造に影響を及ぼさない。しかし，いくつかの木を

合せた範囲での字づらの修正を行なった場合は，木構造は保たれなくなる。この性質が一般のワードプロセッシングと機械翻訳等構文解析処理を含むものとの差異である。その意味からは，木を意識した処理と字づらの校正処理とをうまく分離することが必要となる。

3. ATHENE-82におけるポストエディット機能

図2にポストエディットの一般的手順と校正機能との関係を示す。

又，表1には，ATHENE-82の校正機能[4]の一覧を示す。

図2を見ると分るように，ATHENE-82では，最後の語尾等の校正機能を除いて，ほとんど構文解析木と原文側で持っている構文的又は意味的情報を利用して，校正処理を行なっている。例えば，図3に示すように，和文の表示においては，英文の単語に対応する形で各々が区切られ，例えば，和文を見て，「ので」が，おかしいと思った場合は，その部分をカーソルで指定して，送信キーを押せば，その原文である「AS」が点滅し，それと同時に多義表示がウィンド機能を使って表示される。

ここで一番目をキーで選択すれば，訳語が変わる。この表示・校正法の問題は，2.3にも述べた如く，一旦，木構造を壊すような字づらの変更があると，木構造が回復しにくいことであるが，木構造が壊される前の時点の情報を退避しておくことで実用上はあまり問題なく処理できると考える。

句単位の文字修正

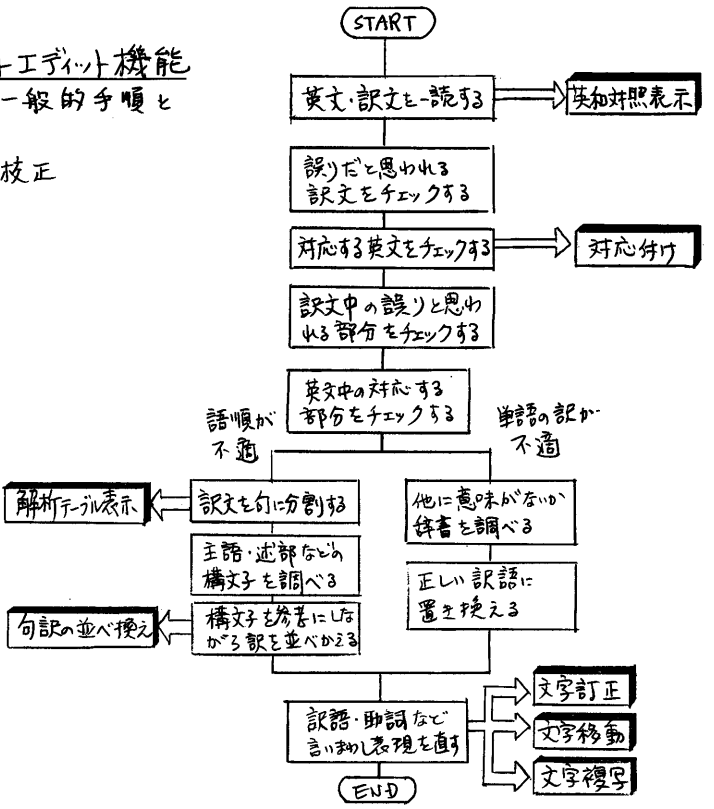

図2. ポストエディットの手順と校正機能との関係

表. 1 ATHENEの校正機能一覧

表示機能	英和対照表示機能	英和対照表示 英和対照スクロール
	辞書表示機能	ウィンド・イメージの表示
校正機能	多義表示・選択機能	英単語と訳語の対応付け 多義表示・選択
	句訳の並び換え機能	解析テーブル表示 句訳の並び換え
	文字訂正機能	文字訂正・挿入・削除 翻訳用ローマ字漢字変換 文字移動・複写 文字サーチ
	同意語表示・選択機能（ウィンド・イメージ）	
	入力チェック・出力チェック機能	
	文型スタイル・言い回し統一機能	
清書機能	レイアウト機能	

や，移動に対しては，図4に示すような表示を行ない，構文解析の結果を利用できるようにしている。この場合にも，句の移動により，木構造が保たれなくなるが，本来そのような機能が使われる時は，字づら処理に入っている時であると考えられる。あくまで木構造を保とうとすれば，図1の解析木そのものを表示，利用することも可能であるが，我々の経験からは，特にユーザが解析テーブル表示に慣れていない場合や，校正ということで既に和文の世界で物事を考えているという状況からは，図4の表示の方が有効なようである。

☆☆☆ 訳文校正 ☆☆☆

入力英文
AS I SAID IN MY REPORT , THE DEMAND FOR SOLAR PANELS IS GOING UP RAPIDLY , BECAUSE THE COST IS GOING DOWN .

翻訳文
私 が 私の 報告 に 言った ので 、 太陽電池用パネル のための 要求 は 急速に 上っ ている その理由は 、 値段 は 下っ ている 。

NO	多義表示
1	ので
2	ように
3	として

（PA1＝MOVE，PA2＝校正終了，PA3＝途中終了）

図3. 多義表示・選択の表示例

1 入力英文

THIS COMPUTER SYSTEM UNDERSTANDS THE PRINCIPLES OF GRAMMAR, THE IDIOSYNCRACIES OF SPEECH.

2 和文生成

英字句	構文子	訳語	助詞
THIS COMPUTER SYSTEM	主語	この計算機システム	は
UNDERSTANDS	主動詞	理解する	
THE PRINCIPLES	目的語	原理	を
OF GRAMMAR	形容	文法の	
, THE IDIOSYNCRACIES	名同格	特質	、すなわち
OF SPEECH	形容	言語の	

3 訳文

この計算機システムは言語の特質、すなわち文法の原理を理解する。

図4. 解析テーブルの表示例

4. おわりに

機械翻訳における3つのタイプの校正方式（プレ・エディット，イン・エディット，ポスト・エディット）について，人手介入のしやすさとインプリメントのしやすさの2点から考察した。また，これらの考察を通じて，現在，我々がポスト・エディット方式の校正を重視している理由についても論じたつもりである。

しかし，本稿で考察した事柄が，機械翻訳における校正の全ての局面をカバーしているわけではない。ユーザによるあいまいな校正指示を正しく解釈するdefault reasoning機能や，一度指示された校正要領を以後反復できるような学習機能など，人工知能的な側面からの研究開発も必要と考えている。

謝辞

最後に，本研究の機会を与えて下さり，日頃，御指導頂く，コンピュータ事業部三浦武雄本部長，システム開発研究所川崎淳所長，石原彦一郎部長に感謝します。

参考文献

[1] Kay, M.: The Proper Place of Men and Machine in Language Translation, CSL-80-11 Xerox, 1980.
[2] Toma, P.: 多言語間翻訳をめざすシストラン・システムその背景と将来，日経コンピュータ(1983.7.25), PP.103-114
[3] Nitta, Y., Okajima, A. et al.: A Hewristic Approach to English-into-Japanese Machine Translation, Proc. COLING82, North-Holland, 1982.
[4] 山野他：英和機械翻訳システム ATHENE-82 (1) 校正処理について，情報処理第27回全国大会 PP. 1087-1088.
[5] 岡島他：自然言語処理における中間語表現と多義解消のしやすさとの関係，情報処理「自然言語処理技術」シンポジウム，PP. 85-90.
[6] Winograd, T: Language as a Cognitive Process Vol. 1 Syntax P. 522, Addison Wesley (1983).
[7] Lawson, V. (ed.): Practical Experience of Machine Translation, North-Holland (1981).