

2C-8

実用性の向上を目指した 機械翻訳システムの改良(2)

宮原浩二, 鈴木克志, 丸山冬樹
三菱電機(株) 情報電子研究所

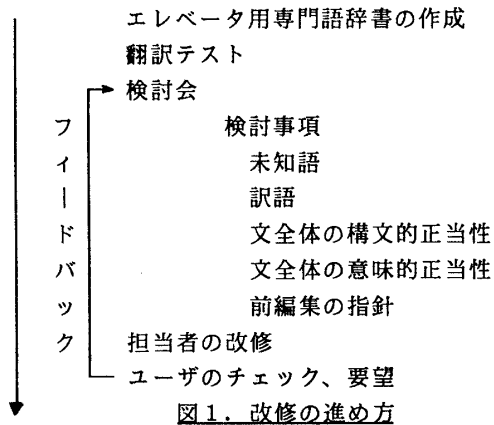
1. はじめに

我々は、日英機械翻訳システムMELTRAN-J/Eを開発中である。機械翻訳システムをユーザに提供し、実運用してもらうためには、翻訳の分野を限定し、辞書や文法の整備を行なうことが不可欠である。そこで我々は特定の分野に焦点を絞って、翻訳結果の評価を行ない、その結果をシステムに反映させるという作業を行なっている。前回の報告[1]に引き続き、本論文ではその過程で発生している問題点について考察する。

2. 改良過程の概略

今回の改良の最終目的は、エレベータのマニュアル類(据え付け、調整等)を機械翻訳システムを用いて日本語から英語に翻訳し、翻訳のコストを低減することである。そこで、まず1冊の日本語マニュアル(約850文、既に英語版が存在している。)を取り上げ、そのマニュアル中に含まれる文章を正しく翻訳できるよう改良を進めることにした。

改良過程を図1に示す。



以下では、これまでに翻訳結果の検討を行なった約450文(これらの内150文は名詞句)について抽出した問題点について述べる。

3. 改修事項

3.1 改修事項の概要

検討会を行なって抽出した主たる改修事項は表1のよ

Improvement of Machine Translation System
for practical use (2)
Koji MIYAHARA, Katsushi SUZUKI, Fuyuki MARUYAMA
Mitsubishi Electric Corp.

うに大別できる。

表1. 改修事項

(a)	辞書の見出しの追加 及び訳語の変更	172語
(b)	不適切な動詞、前置詞の選択 (格パターンの不備を含む)	約80文
	名詞句の並列のスコープの誤り	約10文
	数量の係り先の誤り	約10文
(c)	ユーザに前編集を要求したい文	約30文
(d)	その他 ・文体制御スイッチの導入	

以下、(a)については3.2で、(b)については3.3で、(c)については3.4で、(d)については、3.5で考察する。

3.2 辞書見出しの追加及び訳語の変更

単語については、多くの専門語があることが予想されたため、予めユーザに用語集の提示を求め、エレベータ用の専門語辞書(約2000語)を作成した。以後の辞書登録はこの辞書に対して行なった。

辞書見出しの追加及び訳語の変更については、大きく以下の4つの作業があった。

- (1) 未知語であるため見出し、訳語の登録を要する。
例: 消勢(drop out), 付勢(pick up)
- (2) エレベータ用語としては、不適切であるので訳語を変更する。
例: かご(basket から carに)
- (3) 異表記であるので新たに見出しを登録する。
例: カゴ(「かご」の異表記)
タッチ釦(「タッチボタン」の異表記)
- (4) 複合語であるが、まとめて一単語にした方がよい単語を登録する。
例: 加速丸め点(transition from constant acceleration to maximum speed)
未着床階(floor at which the car has not landed)

3.3 意味的な情報の付与

表1で示した、動詞・前置詞の選択、並列のスコープ、数量の係り先の誤りは、辞書中の意味的な情報(意味マーカ、共起情報、属性情報)が、欠けていたために生じたものである。以下に例を示す。

(1)動詞の誤り

ブレーキを開放する。 Open a brake.
 (cf. 接点を開放する。 Open a contact.)
 正しくは、 Release a brake.

(2)前置詞の誤り

巻上機端子箱で線番B4を取り外す。
 Remove wire No. B4 with the terminal box of
 the traction motor.
 正しくは、 at

(3)数量表現の係り先の誤り

500Vメーターテストを行なう。(500Vが「行なう」に係る)
 (cf. 30A電流が流れる。)
 500Vは、「メーター」に係らなくてはならない。

MELTRAN-J/E では、上記の例を扱うための枠組みは既に存在する[2][3]。例えば、動詞の訳し分けについては、動詞とそれに係る格要素の意味マーカあるいはその単語自身の関係によって適切な動詞を選択するようにしている。従って、これらの問題点を解決するためには、辞書に対して共起情報、属性情報、意味マーカ等を登録すればよいということになるが、この作業には専門的な知識を必要とする場合が多い。そのため、マニュアル中に出現して初めて辞書情報を追加するという方法になり、作業を効率的に行なうことができない。この意味的な情報を辞書に付加する手間が、実用システムとして運用するための大きなボトルネックの一つである。

3. 4 前編集とユーザの要求

機械翻訳システムの弱点として以下の3点がある。

- (1)省略に弱い。
- (2)長文の翻訳の精度が非常に悪い。
 - ・文節の係り受けの誤り
 - ・長文を構成する単文間の係り受けの誤り
 - ・生成時の戦略の誤り
- (3)複雑な並列構造の解析が弱い。

これらの弱点は、現状の機械翻訳システムでは、解決が容易でない。実際に、我々がユーザに対して前編集を要求したい文のほとんどは、上記の3つの点で問題のある文である。

そこで前編集の指針として、開発者側からユーザに

- (1)省略は、できるだけ使わない。
- (2)1文中の動詞の数は3つ程度にする。
- (3)複雑な並列は避ける。

を要求している。

しかし、文によっては、「このような表現は頻繁に使うため、前編集は避けたい。」というユーザからの要求を受けることがある。例えば以下のような表現である。

(a)右勝手巻上機の場合はMJ側が、左勝手巻上機の場合はMK側が(+)電源側になるようにする。

このような表現を解析するには、既存の文法の範囲では、処理が複雑になってしまう。そこで、(a)の文章を表層あるいは形態素の段階で、

(b)右勝手巻上機の場合はMJ側が(+)電源側になるようにする。左勝手巻上機の場合はMK側が(+)電源側になるようにする。

のように入力文を日本語側の解析中に書き換える処理を導入することを考えた。この処理については、ユーザのマニュアル執筆の自由度とのバランスをどのように取るかという点においてユーザ側との調整が難しい。

3. 5 生成する文体の制御

通常、マニュアルでは、日本語の主語無し平叙文は、命令形に訳した方が適切であることが多い。この種の規則は、システムに静的に埋め込んでしまうと他の分野の翻訳に対して悪影響を及ぼすことが考えられる。そこで、ユーザが文のスタイルを設定できるように、生成する文体を制御するスイッチを導入した。

現在、このスイッチを用いて以下のことが出来る。

- (1)主語無し平叙文を命令文として訳すかどうか。
- (2)文が名詞句だけから成り立っている時、冠詞を訳出するかどうか。
- (3)英語の他動詞を訳出し、目的語がない場合ダミーの目的語を訳出するかどうか。

今後は、このスイッチをさらに充実したものにする必要がある。

4. まとめ

機械翻訳を実運用するために抽出した改修事項について述べた。問題点を検討してみて、改めて辞書の重要性を再認識した。専門家の持っている知識(個々の単語の用法、意味)をいかに効率的に辞書作成に反映させるかが実用化の鍵になる。枠組みを持っているということとそれが実際に有効に働いているということは全く別問題である。効率的な辞書作成の手段として、例えば、共起情報については、既に存在している日英両方のマニュアルから自動的に抽出し学習することが必要である。

また、ユーザの理解を得て、ユーザの要求とシステムの性能の一致を計ることが重要である。そのためには、ユーザとシステム開発者の十分な意志疎通ができなくてはならない。ユーザがシステムを評価できるような客観的な評価基準を確立することが今後の課題である。

参考文献

- [1] 宮原他：実用性の向上を目指した機械翻訳システムの改良，第40回情処全大(1990)。
- [2] 鈴木他：日本語の表層格の記述について (2)，第39回情処全大(1989)。
- [3] 高山他：日英機械翻訳における数量表現処理の拡張，第42回情処全大(1991)。