

実用性の向上を目指した機械翻訳システムの改良

2F-7

宮原浩二, 鈴木克志, 丸山冬樹, 太細 孝
三菱電機 情報電子研究所

1. はじめに

我々は、現在MELCOM PSI-I上に日英機械翻訳システム MELTRAN-J/E(以下、MELTRAN) [1] を開発中であり、実文書(主として、マニュアル、仕様連絡書)を対象とした翻訳結果の評価を行ない、翻訳能力の向上を図っている。本稿では、実文書を翻訳するために、MELTRAN に対して行なった改良のうち、特徴的なものについて述べる。

2. 改良の概略

パソコンマニュアル約1800文、仕様連絡書及び議事録約150文の翻訳結果を評価して、以下の処理を施すことにした。

(a) 入力日本語の正規化

分野によっては、助詞や助動詞の省略等、独特の表現を多用するため、その解析に失敗してしまう場合があった。しかし、ある種の表現は、語の補完、置換によって解析の失敗を防ぐことが可能である。

(b) より英語らしい英語の生成

英語の生成では語順が重要である。主語や目的語等の機能要素のレベルだけではなく、訳語として選択された英単語の性質といった、より細かな情報も考慮して生成する必要がある。

これらのことを踏まえ、MELTRAN の解析 [2]、変換 [3]、生成のそれぞれの処理の前処理として改良を加えた。解析前処理は字面の上で、変換前処理は内部構造のレベルで日本語を正規化する。生成前処理は、英語の内部構造に対して、より英語らしい表現を生成するための情報を与える。

3. 解析前処理での改良点

3.1 用言の補完

特に、仕様連絡書においては、体言止めの文、接続助詞で終わる文が頻出する。これらの文の中には、文要素が欠けているため正しく解析できないものがあった。例えば例文(1)は、後置詞句(係り先として用言を要求する)の係り先がなく解析に失敗する。

(1) その機能は重要。

そこで文末に用言を補完することにした。

(1) のような体言止めの場合には、文末の文節が“サ変”ならば「する」、それ以外の場合には「である」を補い構文解析をする(図1)。

(1) その機能は重要 [である]

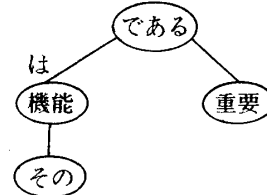


図1. 例文(1)の解析結果

その結果、(1) は、以下のように翻訳される。

(1)' That function is important.

また、接続助詞で終わる表現には次の(2)、(3)のような文がある。

(2) 主記憶について

(3) 英語を翻訳するため

(2) は、表題で多く見られ、(3) は、仕様連絡書に多く見られる表現である。この場合は、主動詞が欠けている文とみなし、文末に empty-node という仮想的な動詞を補い構文解析をする(図2)。

(3) 英語を 翻訳するため [empty-node]

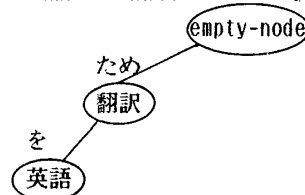


図2. 例文(3)の解析結果

この empty-node は、一般の動詞と全く同じに扱われ、通常の翻訳処理が行なわれ、生成時に空列を生成するのみである。(3) の翻訳結果は、以下ようになる。

(3)' To translate English.

3.2 特有の表現に依存した補完

(a) “サ変” + {「可能」「可」「不可」「不要」等} の表現は、“サ変”の後ろに「することが」を補完。

(4) マニュアルは参照不要。

(4) では本処理と3.1で述べた処理が適用される。し

たがって以下の文が翻訳されることになる。

(4) 'マニュアルは参照することが不要である。

(b) “サ変” + 「のこと」で終わっている文は「のこと」を除去し、“サ変”の語尾を「せよ」とする。このパターンの文はマニュアル中に多く現れる。

(5) 詳細についてはマニュアル参照のこと。

(5) の翻訳は、「詳細についてはマニュアル参照せよ。」を翻訳した結果と同じになる。

(5) ' Refer to the manual about details.

4. 変換前処理での改良点

4. 1 副助詞「は」の扱い

副助詞「は」を主格を示す『が』格として解釈されることが多い。しかし、マニュアルでは(6)のように目的格を示す『を』格の代用として、「は」を用いた文が多く現れる。

(6) その機能は第3章で説明します。

そこで、日本語の内部構造が持っている助詞「は」の情報を「を」に書き換える処理を導入した。以下の条件を満たした時に、助詞「は」を「を」に置き換える。

- ・受動態でないこと。
- ・「は」が接続して名詞の意味分類が、人間でないこと。
- ・他に『を』格がないこと。(一文一格)
- ・他に『が』格がないこと。(「は」「が」構文等があるので、格の認定は変換処理に委ねる。)
- ・係っている動詞の格パターンが全て『を』格を必須格として要求すること。

この結果、例文(4)の内部構造は、(4)の文に本処理が適用されて図3のようになる。

(4) マニュアルは 参照 [することが] 不要 [である]

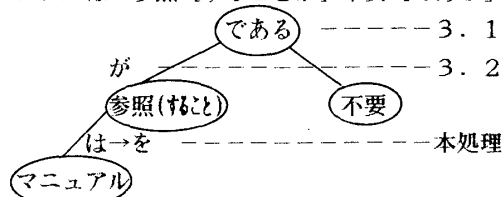


図3. 例文(4)の内部構造

(4) は、以下のように翻訳される。

(4) " It is unnecessary to refer to the manual.

5. 生成前処理での改良点

5. 1 動詞修飾句の生成順序制御

一般に英語は日本語に比べ語順の制約が厳しい。特に任意格、副詞の語順にその差が顕著に現れる。そこで、変換後の英語の内部構造に対して、任意格(前置詞句)

と副詞の生成順序を指示する処理を導入した。

(a) 副詞の情報

各々の英語の単語自身が持つ生起位置(主語と動詞句の間、文尾etc.)と意味(場所、時間、頻度etc.)

(b) 任意格の情報

任意格を構成している中心名詞の意味と日本語の格助詞を基に与えた格マーカ(道具、場所、起点etc.)

これら2つの情報を基に、ある動詞に係っている全ての任意格と副詞を以下に述べるような規則に従って順序付けし、その順序に従い英語を生成することにした。

・ mid-position (主語と動詞句の間に生成する。)

主語と動詞句の間に生起すべき副詞

・ end-position (以下の順序で文尾に生成する。)

- ① 様態を示す副詞
- ② 場所格、時間格以外の任意格表現
- ③ 場所を示す副詞
- ④ 場所を示す任意格表現
- ⑤ 時間を示す副詞
- ⑥ 時間を示す任意格表現

また、動詞修飾句の生成順序制御と同様に名詞修飾語(形容詞、限定詞)の意味情報(色、形状、新旧etc.)を利用した生成順序制御の処理も導入した。

6. まとめ

実文書を対象とした機械翻訳の翻訳能力向上のために行なった改良の中で、解析、変換、生成のフェーズのそれぞれの前処理として導入した処理について述べた。本稿で述べた規則は、既に我々のシステムに組み込まれており、有効であることを確認している。実用システムとするためには、これら3つの前処理をより充実させる必要があるのは言うまでもない。今後、さらに多くの実文書の翻訳評価を行ない、有効な規則を実装していきたいと考えている。この他にも、特に現在の機械翻訳システムの一般的弱点とされている長文の翻訳の扱いについて、改良策を検討し、翻訳能力の向上を図っていく予定である。

参考文献

- [1] 鈴木: 日英機械翻訳システム MELTRAN-J/E, bit別冊, 機械翻訳, 共立出版(1988).
- [2] 高山他: 日英機械翻訳における日本語解析前半部の一構成, 第39回情処全大(1989).
- [3] 鈴木他: 日本語の表層格構造の記述について(2), 第39回情処全大(1989).