

6E-2 日英機械翻訳における数量表現処理の拡張

高山泰博 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

日英機械翻訳における数量表現の取扱いは、日本語と英語の言語構造の違いを反映した、複雑な処理の一つである。数量表現の分析や処理方法に関する研究には、従来も種々の議論がある(文献2~5)。我々も日英機械翻訳システムMELTRAN-J/Eにおける数量表現の基本処理について文献1に述べた。ここでは、技術論文(文献8)中の約1700文の分析とマニュアル中の約850文の翻訳テストに基づく数量表現処理の拡張について述べる。

2. 数量表現を中心とする翻訳過程

図1に数量表現の処理に着目した翻訳過程を示す。

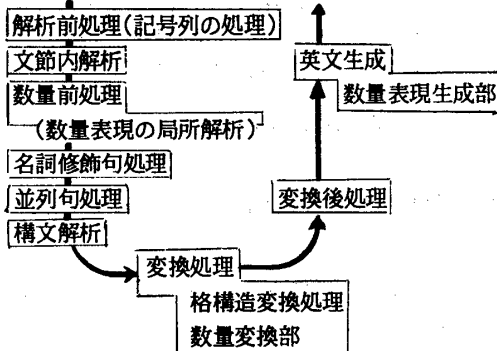


図1 数量表現という観点からみた翻訳過程

文献1では、数量表現の形態的な分析と基本的な数量表現処理の実装について述べた。形態的な処理では、助数詞の繰り返しを考慮した。数量表現の処理は、その局所性から数量詞(Qで表す)と隣接する名詞(Nで表す)との関係にのみ着目しがちになる。ところが、実際の文書中の文で用いられる数量を分析すると、QやNの部分が複合構造である場合がある。

(1)Nが並列している場合の例

「定格速度が60m/minの乗用エレベーターと住宅用エレベーター」

(2)Qが並列している場合の例

「定格速度を45, 60m/minとする。」

上記のように数量表現の解析は名詞修飾句や並列句の解析とも関連する。そのため、日本語文における数量表現を英語に翻訳するためには、図1に示すように複数の処理フェーズによる取扱いが必要となる。

例えば、「減速する」「短縮する」など量の増減を表現する用言の副詞となる数量表現については変換処理で扱う必要がある。また、「たくさん」「すこし」などの数詞を伴わない程度表現(文献6)に関しては、英語表現の構造がある程度決定してから変換処理の後処理で処理している。

3. 数量数量詞の処理

数量数量詞は、修飾する名詞の数量を表す。数量数量詞の処理では、以下の2つの情報を使用する。

(1)名詞の一般的な意味分類と、その意味分類に属する名詞に係り得る助数詞との組の表

(2)名詞と助数詞の辞書項目に設定した数量表現情報
名詞の辞書項目に設定する数量表現情報を以下に示す。

```
jdict(エレベーター,名詞,<乗物>,...,拡張項目)
    数量数量詞 一般的な意味分類
    拡張項目=[...,qq([台,基]),...]
```

辞書記述は上記のようにProlog項の形式をしており、拡張項目と呼ぶリストに付加情報を設定できる。

(1),(2)の名詞と数量詞との関係は数量表現の解析部で数量詞の係り先の名詞を決定する際に利用する。

助数詞の拡張項目には英語表現での訳出要・不要などの情報を設定できる。

```
jdict(台,助数詞,...,[...,josuushi(訳出不要),...])
```

助数詞情報は主として数量表現の変換部で使用される。実際にはすべての名詞に上記の情報が設定してあるとは限らない。また、現時点では辞書中の各単語に数量表現の扱いを考慮したシソーラスコード(文献1)を設定していないため(1)のデフォルト処理が必要となる。

4. 属性数量詞の処理

属性数量詞は、修飾する名詞の属性の値を表す。属性数量詞を解析するためには数量数量詞に比べて詳しい情報が必要となる。

4.1 属性数量詞の情報

属性数量詞に対しても数量数量詞と同様に、一般の意味分類を用いたデフォルト処理と個別の辞書項目に設定した情報を解析に利用する。属性数量詞の情報の記述例を以下に示す。

```
jdict(検出器,<計測器>,...,拡張項目)
    拡張項目=[...,qq([台]),
               aq([voltage('V'),
                  属性数量詞 e_current('A')]),...]
```

4.2 一つの名詞を二つ以上の数量表現が修飾する場合

数量数量詞と属性数量詞の両方が一つの名詞句を修飾したり、複数の属性数量詞が一つの名詞句を修飾する場合がある。

例1 「30V以上の検出器を3台」

3 detectors of 30V or above.

ある名詞句に対して数量表現を解析する際に、一つの数量表現に対する処理が終了しても、更に他の数量表現が修飾する可能性を考慮して処理を継続する。

一方、以下の例2では、名詞「マイコン」を数量数量詞「128個」が修飾しており、「一つ」は数量数量詞であることと「ずつ」という副助詞とで判断して「マイコン」に「一つ」を係けない。

例2 「128個のマイコンを一つずつポーリングする」

4.3 属性名-属性値型の数量表現

以下の例は、文献4において値-関数用法と呼んでいる属性名-属性値型の数量表現である。

- (a) 「45°の角度」 (Q/N)
 (b) 「変調度25%」 (NQ)
 (c) 「周波数は25MHzである」 (N/Qデアル)

対訳程度の情報しか設定していないユーザ辞書や専門語辞書から得られた語では、Nに<物理量の名称>としての意味マーカが付与されていなかったり、Qにその物理量の単位を含むことの情報が付与されていない場合がある。その場合にはNの末尾の語「~度」「~率」を手がかりに、この用法の数量表現として処理する。

(b)では、NQという形態で名詞と数量表現が文中に隣接して出現する。この表現に対しては、文節内解析時に複合表現としての構造を作成しておく。

4.4 係り先が曖昧な属性数量詞

数量表現の情報を詳細に設定しても係り先の曖昧さを解消できない場合がある。

例 「レーザを10mW出力で動作させる。」

上記の例では、「10mW」係り先として3通りある。

- (1) 「レーザ」に係る。
 (2) 「出力」と複合名詞を作る。
 (3) 「動作」を副詞として修飾する。

5. 数量表現に関連する表現

技術文書を初めとして機械翻訳の対象となる文書中には、種々の記号表現が用いられる。形態素解析を補助するための記号表現の処理については、文献7において述べた。ここでは、数量表現と関連する記号表現について述べる。

5.1 時間表現や頻度表現

時間表現(年月日の表示)や頻度表現には数量表現と同様に数詞が使用される。頻度表現は英語生成時の順序制御にも影響する。(文末の前置詞句の順序は<様態><場所・方向><確定的頻度><時間>の順である。)頻度表現では、頻度の基準が隣接している場合がある。

(a) QQ 「1日3回」

(b) QあたりQ

「1時間当たり10人の車椅子使用者を輸送する。」

(c) QにつきQ

「1極につき2点切りの並行反発導体構造」

一般の数量表現でも「の」以外の付属語表現を持つ名詞が隣接し数量表現の内容を限定する場合がある。

(d) 「需要は、合計で14万台程度」

(e) 「面積比で135%の大型化を図った」

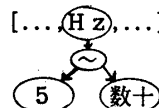
この場合にも「合計で」、「面積比で」を数量表現の一部として処理しなければならない。

5.2 概数を含む数量表現

数量表現中の数量詞の部分に「数十」、「十数」、「数万」などの概数を含む場合がある。以下の例は数詞と概数表現が並列している。

例 「5~数十Hz」

[... (5) (数十) (Hz) ...]



<文節内解析の出力>

<数量前処理の出力>

上記の例を英語に翻訳する際には名詞frequencyを挿入し以下のように訳出する必要がある。

a frequency from 5 to tens of Hz

数量詞の形態が複雑な例に以下の表現がある。

「2~3」「2、3」「2、3」「2つか3つ」

上記の日本語表現を英語側では a few の形に訳出する必要がある。

数量詞に数字が用いられない例に以下の表現がある。

「約 α/β 」「約半数」

5.3 数量表現の情報の利用

これまで述べてきたように数量表現の処理は複雑であるが、逆に、数量表現の情報を他の解析に利用できる。

(1) 「である」が「の」に縮退した場合

「内部電源が5V以下のテスト」

==> 「内部電源が5V以下であるテスト」

(2) 並列表現のスキープの決定

「筐体の高さ69mm、重量8.4kgという薄型」

(3) 省略された述部の推定

「電力は15%、容量は15~20%低減した。」

6. おわりに

従来は、数量表現に関して、局所的な処理について議論されてきたが、本稿では文全体で数量表現を考慮する必要性を指摘した。また、一般的な処理に用いる意味分類を数量表現処理に利用し、その意味分類と数量に関する分類のズレを補正するために、個々辞書の辞書項目に数量表現情報を設定する方法について述べた。更に、現時点での問題点を例示した。今後、翻訳対象となる文書分野毎での数量表現の違いなどを考察し、改良を進める予定である。

参考文献

- (1) 吉武ほか: 日本語の数量表現とその英語への機械翻訳に関する一考察, 情処研報NL74-7(1989).
 (2) 飯野, 亀井: 英語数量表現の解析, 第38回情処全大6E-6(1989).
 (3) 旭岡ほか: 日本語の数字表現の意味的な分類について, 第40回情処全大5F-2(1990).
 (4) 旭岡ほか: 日本語の数量的表現の意味処理について, 情処研報NL78-17(1990).
 (5) 亀井, 村木: 日本語助数詞の分析, 第41回情処全大4S-3(1990).
 (6) 亀井, 村木: 程度表現の形式化, 信学研報NLC88-6(1988).
 (7) 高山ほか: 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5F-2(1990).
 (8) 三菱電機技報Vol.64, No.10 「昇降機特集」(1990).