



#### 4.1 形態素解析

最初に日本語文を JUMAN[4] により形態素で区切る。この段階で手話表記文には明らかに表れない単語(助詞、助動詞、接尾辞の一部など)を取り除き、動詞、形容詞など活用形があるものは終始形に変換される。

#### 4.2 手話単語への変換

形態素解析が終了したら、形態素が手話単語辞書に登録されているかを調べ、その結果に応じて次のように処理する。

1. 単語が手話単語辞書に存在する場合 その単語を採用する。
2. 単語が手話単語辞書に複数存在する場合 EDR 日本語単語辞書 [5] の使用頻度情報を参考にして単語の語義を決定し、それに見合う手話単語を採用する。
3. 単語が手話単語辞書に存在しない場合 EDR 日本語単語辞書より単語の概念識別子と日本語概念説明を抽出し、単語と類似した意味を持つ単語を選び出す。それらの単語について再び手話単語辞書を調べ、登録されていたら採用する。また、名詞の場合は指文字表記に変換する。
4. 1~3 の方法がすべて失敗した場合 図像表現に変換する。形態素解析段階で取り除かれた単語を復元し、「[image]」記号を付加する。

#### 4.3 変換規則の適用

最後に構文的な変換や単語の挿入処理が行われる。この時に使われる変換規則は入門用テキストなどで多く見受けられるものに限っており、現在次の2つの規則が使われている。

**挿入規則** 固有名詞で場所の名前や氏名の前には手話単語の「場所」「名前」を挿入する。

**入れ換え規則** 数、年齢、時間など数値に関する表現の場合には数値の前後の語順を入れ換える。

変換処理までが終了すると、整形して区切り文字を挿入し、手話表記文を出力する。

### 5 システムの試作

データとして「NHK ニュース聴力障害者のみなさんへ(以下、手話ニュース)」の1994年9月18日(夜)、9月20日(夜)、9月22日(昼)放送分を元にして対訳コーパスを73文作成した。手話ニュースをデータに選んだのは以下の理由による。

- 口話を併用するため日本語文と手話の対訳が得られやすい。

- ニュース文であるので、日本語文は一般的な日本語文法に沿っている可能性が高いため、既存の自然言語処理技術で扱いやすい。
- 日常会話ではあまり使われないような用語、抽象概念の手話表現が得られる。

このデータを使い手話単語変換段階の語義情報を使用しないシステムを試作した。入力73文について、形態素解析が終了した段階で残った形態素が611個あった。そのうち、手話単語辞書に存在するものが179個、EDR日本語単語辞書で検索可能なものが241個、すべてに失敗したものが191個であった。EDR日本語単語辞書で検索可能なもののうち、約半数は手話単語辞書に存在する類語に到達できる見込みがあったが、残りは不完全な情報が多く、別の手法が必要と思われる。

### 6 おわりに

手話表記文では手話単語辞書に登録されている手話単語のみを使用する必要があるが、人間が記述すると同じ手話表現でも個人の知識に依存して違う日本語見出しで記述する可能性がある。そのため記述した手話表記文は、変換と同じ手法により手話単語辞書で登録されている手話単語のみで清書する処理が必要だと考えている。

また、現在の変換方法では日本語見出しが手話単語辞書と一致したら、変換が終了のもののみとしているが、この場合、「奴は警察の犬だ」という文を処理した場合「彼/警察/犬」という結果を出力する。しかし、ここでの「犬」というのは「スパイ」の意味であり「彼/警察/スパイ」という手話表記文を出力するようになりたい。

#### 参考文献

- [1] 徳田, 奥村. 日本語から手話への機械翻訳に関する検討. 手話学会第21回大会, 1995.
- [2] 安達. 手話通訳のためのニュース文の話しコトバへの変換処理. 電子情報通信学会研究会資料 NLC92-47, pp.17-24, 1992.
- [3] 安達, 吉澤, 藤田, 松本, 鎌田. 手話ニュースの分析と手話変換処理. 情報処理学会研究会資料 92-NL-91, pp.17-23, 1992.
- [4] 松本, 黒橋, 宇津呂, 妙木, 長尾. 日本語形態素解析システム JUMAN 使用説明書 version 2.0. 1994.
- [5] 日本電子化辞書研究所. EDR 電子化辞書 Version 1.0. 1995.