7H-8 手話表記法の提案と日本語から手話への変換方法の検討

徳田昌晃　奥村学
北陸先端科学技術大学院大学 情報科学研究科

1　はじめに

近年、聴覚障害者の社会参加が進むにつれて、手話通訳者の不足、健康問題が注目されているが、日本語と手話間の機械翻訳システムを開発することにより、このような問題の解決が期待できる。

手話は表現が立体的な言語であり、また記述文字が確定していないために、計算機で扱うためには画像処理などの一次元的な記号に変換する処理が必要であるが、そのために費やす労力は莫大である。本研究では手話を表記する方法を提案し、さらに日本語から手話への変換手法を検討した。

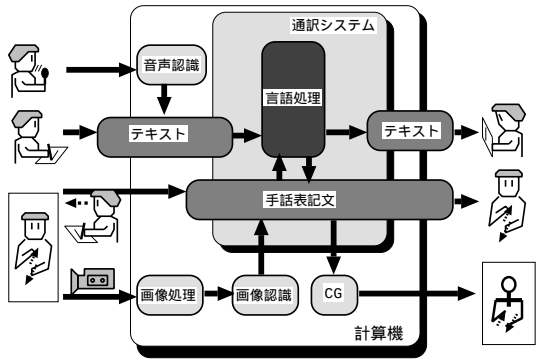


図 1. 日本語から手話への機械翻訳システムの概観

2　手話表記法 (文)

本研究では手話の一次元表記として、手話表記法 (文) を提案し使用する。これは手話を日本語文字列をラベルに用いて記述したものである。手話表記文は人間が使うことも想定し、直観的にわかりやすく容易に読み書きができるように特殊記号の使用は少なくしている。手話表記文は BNF 記法により次のように定義されている。

< 手話表記文 > ::= < 手話表記文 >< 文節 > |
< 文節 >< 文節 >
< 文節 > ::= < 単語列 >< 区切り文字 >
< 区切り文字 > ::= ′/′|′ ′|′,′
< 単語列 > ::= < 手話単語 > |
< 手話単語 >′∗′ < 数 > |
< 手話単語 >′<′< 手話単語 >
| < 手話単語 >′<′ |
< 指示詞 >< 図像表記 > |
′『′ < 指文字表記 >′』′
< 数 > ::= 任意の数字
< 指示詞 > ::= ′[image]′

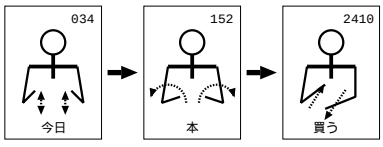
手話単語　手話の単語を示す。

区切り文字　「/」、「,」、空白のいずれかを使い、項の区切りを示す。

指示詞　特殊な表現が次に続くことを指示する。現在は [image](図像表現) のみが定義されている。

<　左右の方向性を示す。

手話表記文の例を次に示す。() に理想的な日本語訳を示す。



今日/本/買う (今日、本を買った)

手話表記文では強さ、速度、位置関係、表情等のノンバーバルな表現などの要素は明確に記述されていないが、これらは話手が表現する際に自然にそれらの情報が読みとれると考える。

3　手話単語辞書

手話表記文で使われる手話単語は、手話の表現と 1 対 1 で対応し、手話表現がわかりやすいものである必要がある。特に人、地域によって表現の差異が少ないものが望ましい。そこで聾唖者や手話通訳経験者の協力を得て、市販されている手話辞典やテキストから表現しやすい単語、約 3000 語を選定した。そして、これらの単語に読み、分類、用例、手話表現に対応する番号 (ID) などの情報を付加して手話単語辞書を作成した。

4　変換方法

本稿では日本語文の入力文を手話表記文に変換する場合のみを考える。以下、変換手法を順に従って述べる。

Orthography of a Japanese Sign Language and Translation of Japanese into JSL

TOKUDA Masaaki, OKUMURA Manabu
Japan Advanced Institute of Science and Technology
15 Asahidai, Tatsunokuchi, Ishikawa 923-12, Japan

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、時間など数値に関係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、形態素解析が終っ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が可能なものが 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、変換が終ったものとみなしているが、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したい。

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.