

日本語テキストから手話単語列への機械翻訳システム

jaw/SL の試み

田中伸明，吉田鑑地，松本忠博，池田尚志
岐阜大学工学部

1 はじめに

我々は日本語テキストから手話単語列への機械翻訳システムについての検討を行っている。

まず市販の手話辞典などから日本語-手話単語対訳辞書データ，日本語例文 - 手話単語列対訳コーパスを作成した。これを用いて単純な自立語単語の置き換えによる翻訳システムを作成したが，このような単純なシステムではうまく翻訳できないケースがたくさん出てくる。本報告では，これらの問題について整理する。

一方我々の研究室では，通常 of 自然言語間の機械翻訳を実現するための翻訳エンジン jaw を開発しており，日本語からアジア諸言語（中国語，ベトナム語，シンハラ語，ミャンマー語）への翻訳システムを研究開発している。手話もこれらの言語と同じく 1 つの言語である。現在我々はこの翻訳エンジンを用いて日本語テキストから手話単語列への機械翻訳システムの構築を検討している。

2 日本語-手話対訳データの収集

2.1 『日本語-手話単語辞書』

日本語-手話機械翻訳のための基礎データとして『日本語-手話辞典』[1]，『手話・日本語大辞典』[2]，『イラスト手話辞典 1・2』[3]，『わたしたちの手話 1～10』[4] の 4 種類の手話辞典から単語を収集した。各辞書にはそれぞれ 3,217 件，2,951 件，4,612 件，4,122 件の手話単語が収録されている。辞典の例として『日本語-手話辞典』の構成を図 1 に示す。



図 1: 日本語-手話辞典の構成

これら 4 種類計 14 冊の手話辞典の手話単語データをまとめることによって，1 冊の辞書のように扱う事ができる『日本語-手話単語辞書』を作成した。

この辞典には，異なり数で 7,732 件の手話単語が収録されている。

2.2 『日本語-手話対訳コーパス』

2.1 で取り上げた辞典の一つである『日本語-手話辞典』には，日本語例文とそれを手話で表した手話表現が掲載されている（図 1）。このデータを取り出すことによって 8,972 件の『日本語-手話対訳コーパス』を得ることができた。日本語例文と手話表現との対応の例を表 1 に示す。表中“<>”で囲まれたものは手話の動作に日本語のラベルを便宜的に割り当てた，手話単語を表現するためのラベル”手話単語ラベル”である。

表 1: 『日本語-手話対訳コーパス』

日本語例文	手話表現
明るい政治	<明るい> + <政治>
空が明るい	<空> + <明るい>
テーブルマナー	<ステーキ> + <常識>
便利な所	<便利> + <場所>
日頃の行いが良い	<いつも> + <活動> + <良い>

2.3 『日本語-手話対訳コーパス』からの単語対訳データの取り出し

『日本語-手話対訳コーパス』の日本語例文を，我々の研究室で開発している文節構造解析システム ibukiB で解析した。ibukiB は以下のような解析を行うシステムである。

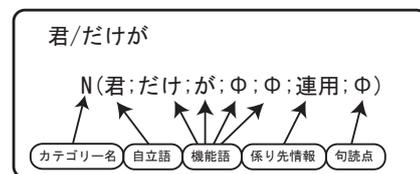


図 2: ibukiB 解析結果の例

『日本語-手話対訳コーパス』において，日本語例文中の単語がどのような手話単語へと翻訳されているのかを調べるために，解析結果の自立語と対応する手話表現を比べてみた。その結果，手話辞典に収録されていない表現を用いて翻訳されているものがあることが分かった。そこで『日本語-手話対訳コーパス』の手話表現を参考に，表 2 のような単語対訳データを『日本語-手話単語辞書』に 2,383 件追加した。

表 2: 単語対訳データ

自立語	手話単語
近所	<家> + <短い>
大食い	<たくさん> + <食べる>
地面	<土>
急行列車	<はやい> + <電車>

3 単純単語置き換え法による日本語-手話機械翻訳の試み

3.1 単純単語置き換え法

日本語-手話機械翻訳への足がかりとして、『日本語-手話対訳コーパス』の日本語例文に対して、例文の語順のまま単純に手話単語へと置き換える、という方法で翻訳を行うことにした。そして、得られた翻訳結果と対訳コーパスの手話表現を比較することで問題点をリストアップし、検討することにした。この翻訳の例を図 5 に示す。

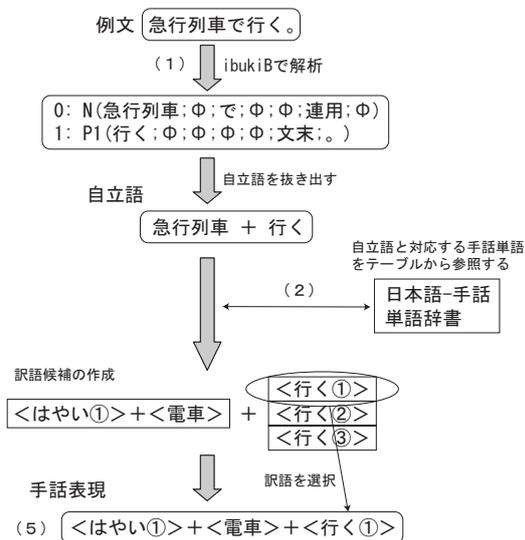


図 3: 日本語-手話機械翻訳の例

また、以下でこの翻訳の流れを簡単に説明する。

1. 日本語例文を ibukiB で解析する。
2. 解析結果の自立語と対応する手話単語を『日本語-手話単語辞書』から検索する。
3. 見つからない場合は、その自立語の翻訳結果を空白にしておいて、次の文節の自立語について処理を繰り返す。
4. 最終的に見つかった手話単語を順番に並べて手話表現として出力する。

3.2 翻訳結果および考察

『日本語-手話対訳コーパス』の例文 8,972 例文に対して単純単語置き換え法によって翻訳を行った。その結果、うまく翻訳できたものと、うまく翻訳できなかったものの翻訳結果を表 3, 4 に示す。

およそ半分程度は表 3 のようにうまく翻訳できた。但し、複数の訳語候補が得られた場合は、訳語候補の中に対訳コーパスの手話表現と一致する手話単語が含まれていれはうまく翻訳できたとした。また、「漢字を覚える。」などのように、『日本語-手話対訳コーパス』の手話表現とは異なるが意味的には合っていると考えられるものもこちらに含めた。

表 4 を見てみると、意味が文脈に依存するような単語(例:「～進む」)を含む文の場合は、翻訳結果が違ってしまっている。

手話では<雨>という手話単語 1 語で「雨が降る。」という意味も表すが、今回のシステムでは<雨> + <降る>と翻訳してしまい、余分な手話単語が入ってしまっている。

「日が落ちる。」のように意識をしているものもうまく翻訳できなかった。

表 3: うまく翻訳できた翻訳結果の例

例文	
対訳コーパスの手話表現	翻訳結果の手話表現
空が明るい。	
<空> + <明るい>	<空> + <明るい>
漢字を覚える。	
<複雑> + <文字> + <覚える>	<漢字> + <覚える>
状況が逆転する。	
<状態> + <逆転>	<状況> + <逆転>

表 4: うまく翻訳できなかった翻訳結果の例

例文	
対訳コーパスの手話表現	翻訳結果の手話表現
高校に進む。	
<高校> + <入る>	<高校> + <進む>
雨が降る。	
<雨>	<雨> + <降る>
日が落ちる。	
<夕方>	<日> + <落ちる>

単純単語置き換え法では、例文中の単語をそのまま対応する手話単語に置き換えているため、原文と大幅に意味がかけ離れてしまうことはあまり無く、約半分程度は意味が通じる翻訳ができたと思われる。しかし、対訳コーパスの手話表現は文脈依存や意識をしているものが多数あり、単純単語置き換え法で精度の良い翻訳を行うのは難しい。

4 機械翻訳システム jaw による日本語-手話機械翻訳の検討

前述したように、単純な自立語の置き換えでは、文脈の意味が吸収できない、冗長な表現が含まれてしまう、などの問題点がある。

そこで、我々の研究室で開発している機械翻訳システム jaw による日本語-手話機械翻訳を検討することにした。

4.1 機械翻訳システム jaw

Jaw (from Japanese to Asian and World languages) はパターン転写方式による機械翻訳システムであり、原言語である日本語の表現パターン (係り受け構造のパターン) とそれに対する翻訳規則を与えることにより、任意の目的言語に対応することが可能である [5]。現在は、中国語、シンハラ語、ベトナム語、ミャンマー語を目的言語としてシステムの開発を進めている。

表現パターンは、図 4 (a)–(c) のような 3 種類に分類される。

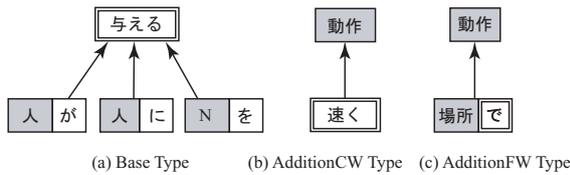


図 4: 表現パターンの例: 2 重の四角はキーワードを、グレーの四角は意味属性を表している

(a):Base タイプ キーワード (表現パターンの中心となる語) が自立語であり、それがどのような文節を受けるかを記述したパターン

(b):Addition (CW) タイプ キーワードが自立語であり、それがどのような文節に係るかを記述したパターン

(c):Addition (FW) タイプ キーワードが機能語 (付属語) であり、その機能語がどのような自立語に後接するか、また、その機能語を含む文節がどのような文節に係るかを記述したパターン

なお、用言に後接する機能語 (過去を表す「た」など) については、別途処理される。

図 5 に翻訳処理の流れを示す。

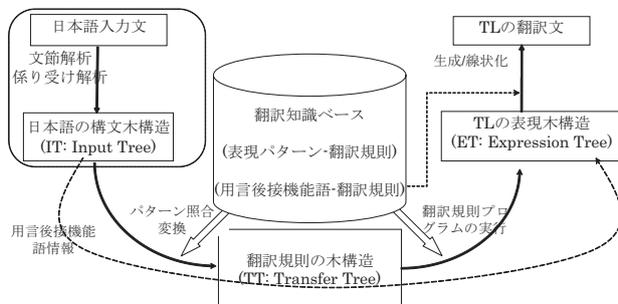


図 5: 機械翻訳システム jaw の処理の流れ

入力日本語文を構文解析することによって得られた係り受け構造 IT (Input Tree) は、翻訳規則パターン辞書との照合により日本語パターンと対応する翻

訳規則の木構造 (TT: Transfer Tree) に変換される。次に TT 中の翻訳規則に対応づけられた翻訳プログラムを実行することで目的言語の表現構造木 (ET: Expression Tree) が生成される。最後に、各 ET に対応した生成関数を実行することで文が生成される。

このシステムは、翻訳知識の多くを翻訳規則パターンとして辞書に表現しているため知識の追加が容易である。また、パターン辞書との照合による TT の作成、TT に対応づけられた翻訳規則の実行による ET の生成、ET に対応づけられた生成関数の実行による目的言語の生成という方式は、翻訳知識データとその処理を完全に分離しており、任意の言語を目的言語として対象とすることができる。

4.2 jaw を利用した日本語-手話機械翻訳の試み

3.2 で述べたように、自立語を単純に置き換える方法ではうまく翻訳できない原因として以下のようなものが挙げられる。

- (1) 日本語の単語と手話単語の意味が完全には一致しておらず、用いられ方によって別の単語に置き換える必要がある。
例: 「高校に進む」 <高校> + <入る>
- (2) 手話動詞がその動作の主体や対象などの意味を含んでいる場合がある。
例: 「雨が降る」 <雨>
「お茶を飲む」 <(お茶を)飲む>
- (3) 日本語と手話とで語順が異なる場合がある。
例: 「私は 22 歳 です」
<私> + <年齢> + <22>
「あの人は とても美しい」
<あの人>
+ <美しい> + <きわめて>
- (4) 自立語以外の語も手話単語で表現しなければならない場合がある。
例: 「大阪 で 生まれる」
<大阪> + <場所> + <生まれる>
- (5) 過去、可能、推測などの意味を持つ用言に後接する機能語も翻訳しなければならない。
例: 「大阪に行った」
<大阪> + <行く> + <過去>
「大阪に行き たい」
<大阪> + <行く> + <好き>
「大阪に行き たくない」
<大阪> + <行く> + <嫌い>

これらに対しては、表5に示すような、適切な構文パターンとそれに対する翻訳規則を登録することで正しく翻訳できた。

また、(5)のような用語に後接する機能語を含む文は、機能語部に対する翻訳規則を別途記述することで翻訳できる。

構文パターンと翻訳規則	翻訳例
	<p>「彼は高校に進む」</p> <p>↓</p> <p><彼> + <高校> + <入る></p>
	<p>「私はお茶を飲む」</p> <p>↓</p> <p><私> + <(お茶を) 飲む></p>
	<p>「彼女はとても美しい」</p> <p>↓</p> <p><彼女> + <美しい> + <きわめて></p>
	<p>大阪で生まれる</p> <p>↓</p> <p><大阪> + <場所> + <生まれる></p>
	<p>「大阪に行った」</p> <p>↓</p> <p><大阪> + <行く> + <過去></p>

4.3 手話特有の動作方向の変化

手話では対象の位置や状態により動作の方向が変化する。特に受動表現ではそれらの情報が重要になってくる。



図 6: 日本語-手話の能動と受動表現

図6は能動態と受動態の手話表現である。能動態での「助ける」は自分から相手の方向に手を押し出すようにする。一方受動態での「助けられる」は相手

から自分の方向に手を引き寄せるようにする。これは「助ける」という手話単語が助ける対象に向かって表現する手指動作であるからである。このため、能動文「あなたが私を助ける」、受動文「私はあなたに助けられる」は、どちらも助けられる対象が「私」であるので「助ける」の動作は「私」に向かって行われる。

以上のことから、日本語の受動表現を手話で表現するには、行動主と受身側を判断し、動作の方向を明確にする必要がある。そして、より正確な翻訳のためには動作方向などの情報を翻訳文に付与する必要があると考え、jawでの翻訳文の表現法について以下のようなものを検討している。

例：あなたが私を助ける

{2: <あなた>} + {1: <私>}
+ {2 1: <助ける>}

例：私はあなたに助けられる

{1: <私>} + {2: <あなた>}
+ {2 1: <助ける>}

5 おわりに

日本語-手話機械翻訳へ向けて、日本語-手話対訳データの収集を行った。これらのデータを用いて、日本語文の語順のまま単純に手話単語に置き換える日本語-手話機械翻訳システムを作成した。約半分程度は意味の通じる翻訳が得られたが、文脈依存や意識など、この方式では対応できない問題がでてきた。

そこで、機械翻訳システム jaw を利用した日本語-手話機械翻訳の検討を行った。これにより、単純置き換え翻訳でできなかったものも、うまく翻訳できる見通しを得た。

今後、これらの試みを踏まえて、手話単語の表現方法についてさらに考察を進め、翻訳システムでの日本語-手話翻訳規則の表現法・翻訳方法について検討を進めてゆく予定である。

参考文献

- [1] 財団法人全日本聾唖連盟日本手話研究所, 日本語-手話辞典, 財団法人全日本聾唖連盟出版局, 1997
- [2] 竹村 茂, 手話・日本語大辞典, 廣済堂出版, 1999
- [3] 丸山 浩路, イラスト手話辞典 1・2, ダイナミックセラーズ出版, 1984-1998
- [4] 全日本聾唖連盟日本手話研究会, わたしたちの手話 1~10, 財団法人全日本聾唖連盟, 1969-1986
- [5] 今井 啓允, 日本語からアジア諸言語への機械翻訳の試み, 情報処理学会 第 65 回全国大会, May.2003