

日本語から手話への機械翻訳における手話単語辞書の補完方法について

徳田昌晃[†] 奥村学[†]

聴覚障害者が使う言葉である「手話」は日本語とは違う体系を持つ言語であり、日本語と手話間の機械翻訳システムは有用であると考えられる。しかし、日本語と手話間の機械翻訳を実現するためには手話の計算機上での扱いと手話の言語的な資源の貧弱さが問題となる。前者は手話が視覚的な言語であるために起こる問題であり、後者は十分な語彙数を持つ機械可読手話単語辞書が現存しないために起こる問題である。本研究では手話を文字列として表現する「手話表記法」を提案する。次に手話単語辞書の語彙の少なさを補うために、日本語辞書を利用して類似語を獲得し、不足している手話単語を補う手法を提案する。これらの手法をもとに日本語から手話への機械翻訳システムを試作し、有効性を検討した。その結果、日本語ニュース文を入力した場合に約83%の形態素が手話単語に変換できることが分かった。

Automatic Complement of Sign Language Dictionary in Japanese-Sign Language Machine Translation

MASAAKI TOKUDA[†] and MANABU OKUMURA[†]

While Sign Language is widely used in the community of deaf, it is noted that the syntax and expression of Sign Language and Japanese are so different and then the Sign Language-Japanese machine translation (MT) is in demand. However, there are two main problems in realizing the Sign Language-Japanese MT. One is how to represent Sign Language which is not a literary style language. The other problem is there are very few entries in current Sign Language dictionaries. To solve these problems, we propose a Sign Language description method and a complement method to deal with input words which do not exist in a Sign Language dictionary. The system finds a similar word from a Japanese dictionary and converts it to a similar Sign Language word. Finally, we implement a prototype MT system based on this method and evaluate it using Japanese News sentences. The experiment result shows that the system succeeds in translation with the rate of nearly 83% for Japanese news sentences.

1. はじめに

聴覚障害者が使う言葉に「手話」がある。手話は手や腕、表情を駆使する視覚的な言語である。近年の言語学における研究により、日本で話されている手話は、単に表現が視覚的であるという表出上の違いだけではなく、構文などの言語面からも日本語とは違う言語であることが明らかになっている¹⁾。使う言語が違うということは、手話だけを話す日本の聾者は、ちょうど健聴者が言葉も分からず外国にいる状況と同じである。もし、日本語と手話の機械翻訳を実現すれば、言葉が通じて便利であると同時に、社会的少数者である

聾者にとって、使い慣れた言語である手話で社会に参加できるという面からも有意義であると考えている。また、現在手話通訳者が不足していること²⁾や、頸肩腕障害³⁾発生の防止面からも機械翻訳システムが貢献できることは多い。

しかし、手話を計算機で扱うためには入出力の壁がある。手話は視覚的な言語であるために画像処理などの労力がかかる。また、画像データそのままでは、計算機上で手話を自然言語として扱うことは難しい。そこで本研究では、まず手話の表記法を定義する。この表記法を用いて手話を記号として表現することにより、計算機で処理することが容易になる。本研究では、入出力となる手話は手話表記法の記号列に変換されたことを仮定して、機械翻訳システムは日本語文と手話表記法で記述した手話文の変換手法を考える。

[†] 北陸先端科学技術大学院大学情報科学研究科
School of Information Science, Japan Advanced Institute of Science and Technology, Hokuriku

機械翻訳を行うためには、変換元の言語と変換先の言語の機械可読辞書が不可欠である。日本語の大規模な機械可読辞書は存在するが、現存する手話単語辞書は小規模であり^{4),5)}、現在でも十分に実用的な手話の単語辞書は作られていない。そこで、本研究は基本的な手話単語を収録した辞書をあらかじめ作成し、収録されていない単語は日本語辞書を利用して補完する手法を提案する。そして試作した機械翻訳システムで手法の有効性を検討する。

本稿では最初に2章で日本の手話の現状について述べる。次に3章で手話を扱った研究と本研究との関係について述べる。次に4章で手話の表記法について述べ、5章で変換手法について説明する。特に概念説明文を使った類似語の獲得方法について6章で検討する。これらの手法を適用して試作したシステムの概要を7章で述べ、8章でシステムの評価を行う。最後に9章で日手機械翻訳システムを実現するための問題について考察を行う。

2. 日本の手話について

日本で話されている手話は「日本手話」と「日本語対応手話」の2つに大別できる⁶⁾。

「日本手話」は「聾者的手話」「伝統的手話」とも呼ばれ、昔から聾者が使ってきた手話である。構文などに日本語とは違う独自の体系が見られ¹⁾、単に「手話」と呼ぶことも多い。一方、「日本語対応手話」は「手指日本語」「健聴者的手話」「同時法手話」「シムコム」とも呼ばれ、手話の初学者がよく使う。特に聾学校などで日本語を教えるために発達したということもあり、構文や語義の扱いが日本語とほぼ同じである。現在では、健聴者²⁾や中途失聴者³⁾、難聴者⁴⁾に広まっており、聾者でも健聴者と話す場合には日本語対応手話で話すことが多い。これら2つの手話は混合して使われることもあり、はっきりと区別することが難しいときもある。

近年、手話の研究はさかんに行われているが、残念ながら日本手話の構文体系は網羅的には明らかにされていない。一方、日本語対応手話は基本的に言語の特性が日本語と同じなので、計算機上に記号列として記述できれば、扱いは容易である。しかし、厳密に日本

語と比較すると、助詞の欠落や日本手話的な接頭辞の付加などがあり、単純に日本語と同じ扱いはできない。

本研究では最終的に日本手話を扱うことを目標とするが、基本的には日本語対応手話を扱う。ただし、日本手話の特徴が明確であり、計算機上で容易に日本手話の特徴が扱える場合には日本手話を扱う。

3. 関連研究

手話を計算機で扱う研究には、入出力に関するものと言語的な変換に関するものがある⁷⁾。

手話の入出力に関する研究は数多く行われている。入力にはビデオカメラなどから動画像を入力し、画像処理を行って手話を認識する方法⁸⁾と、データグローブにより入力する方法^{9),10)}の2種類がある。出力は、過去には手話単語を記録したビデオ動画像をつなぎあわせる方法が用いられてきたが、つなぎあわせた動画像が不自然なことから撮影の労力が莫大になることから、最近ではコンピュータグラフィックスによる人体モデルを使った手話画像の生成¹¹⁾が行われている。

一方、手話の言語的な変換に関する研究は少ない。西川ら¹²⁾は手話単語の表現の際に使っている腕や手の情報を単語辞書内に含めることにより、日本語文が完全に解析できて必要な情報が付加できれば、自然な手話画像を出力できることを示している。藤原ら¹³⁾は日本語文を解析した結果を意味ネットワークで記述することで、手話文が生成できることを示している。これらの研究で使用している入力文や辞書はかなり小規模であり、一般的な文章を扱うことは課題となっている。安達^{14),15)}、鎌田ら¹⁶⁾はニュース文を手話文に変換する手法を検討しているが、変換に利用する現実的な手話単語辞書やシステム全体の構築は課題となっている。

基本的にはこれらの研究をつなぎあわせることで機械翻訳システムが実現すると考えられる。しかし、現実には大規模な手話単語辞書を作るのはコストの問題から難しく、人手で作成した手話単語辞書を元に自動的に語彙を補完する必要があると考える。また、これらの研究で使用している出力形式はそれぞれ独自のものであり、これらを相互変換する必要もある。

本稿では機械翻訳で利用する辞書の補完手法について述べる。また、相互変換が容易であると考えられる手話の表記法を定義して、この表記と日本語間の変換を考える。

4. 手話表記法について

手話単語を認識しても、数値データのままで人間

★1 聴覚障害者で、日本語を第一言語として獲得できなかった、もしくは獲得が不完全だった者。

★2 聴覚障害者でない者。聾者に対する対義語。

★3 日本語を第一言語として習得し、ある程度の年齢になってから聞こえなくなった者。

★4 聴力が著しく低下しているが、多少残存聴力がある者。本稿では難聴者に中途失聴者を含める。

表1 手話表記法

Table 1 Description method of Sign Language.

< 手話表記文 >	::=	< 手話表記文 >'/'< 単語列 >
		< 単語列 >
< 単語列 >	::=	< 手話単語 >
		< 手話単語 > '*'< 数字 >
		< 手話単語 > '<'< 手話単語 >
		< 手話単語 > '<'
		'[image]'< 図像表記 >
		'['< 指文字表記 >']'

には理解できず、構文変換などの言語的な処理も難しい。そこで手話単語を意味するシンボルを定義して、記録する方法を定義する。

手話を動作まで含めて表記する方法には藤井ら¹⁷⁾、本名ら¹⁸⁾、相良¹⁹⁾、伊東²⁰⁾の研究があるが、これらの表記法は動作の完全な再現を目指しているために、絵や平面的な位置情報を含んでいるので、計算機上で自然言語として手話を処理することには向いていない。また英文字などの記号列で表記する方法^{4),21)}もあるが、人間が読み書きすることには向いていない。そこで、本研究では手話単語を意味するシンボルとして日本語を利用し^{6),22)}、その文字列で手話を表記する。

4.1 手話表記法の定義

手話表記法の定義を表1にBNFで示す。

それぞれの項の内容を次に示す。

手話単語 手話の単語を示す。手話単語辞書に登録されている日本語見出しを使用する。

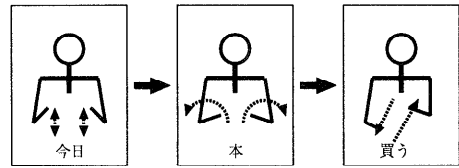
図像表記 手話特有の図的な表現(図像表記)を示す²²⁾。文字列「[image]」から続く日本語文が表現する状況を話し手自らが考えて記述する。

指文字表記 指文字*での表現を示す。「」に日本語のひらがなで指文字で表現すべき内容をそのまま記述する。

* 繰り返しを示す。直前の手話単語を指定された回数だけ繰り返し表現する。

< <の左側に代名詞となる手話単語(人, 車, 場所など)を, 右側にそこに働きかける動作となる手話単語を記述する。この場合, 代名詞となる手話単語を左手で表現し, 右手で動作となる手話単語を左手に向かって指し示すか動かすように表現する。右側の手話単語が省略されたときには指文字の「そ」を使う。

手話表記文を表現するには, < 手話単語 > や < 図像表記 > で該当する手話や表現をそのまま行い, <



今日/本/買う

図1 手話表記文の例

Fig. 1 Description of Sign Language sentence.

で示される方向性を表し, *で示される回数の表現を繰り返す。日本語文「今日, 本を買った」を手話表記法で記述した例と表現の模式図を図1に示す。

5. 日本語から手話への変換

本研究の最終目的は日本語と手話間の機械翻訳システムの実現であるが, 解析部分には従来の日本語の自然言語処理の技術が使えるため, 当面は日本語から手話への変換手法の確立を目指す。また, 本研究では, 入力した日本語文と意味が等価な手話表記文を出力することで, 変換したと見なす。

本稿で扱う機械翻訳システムはトランスファー方式によるものである²³⁾。この方式をとった理由は, 日本語と手話間の変換規則を構築すればシステムとして容易に実現できることと, 日本語側の解析は現存する技術が利用できることにある。ただし, 後者については, 本研究では日本語文の解析結果は十分に精度の高い結果が得られることを仮定して, 試作した機械翻訳システムへ日本語文を入力する際には形態素解析した結果を手作業で修正し, 語義を決定している。

日手機械翻訳システムを実現するために解決しなければならない大きな問題の1つに手話単語辞書の貧弱さがある。我々は基本的な語彙を持つ機械可読な手話単語辞書を作成したが, この辞書だけでは機械翻訳システムで利用するには語彙数が不十分である。5.1節では, 我々が作成した基本手話単語辞書について述べ, 5.2節で足りない語彙を補う手法について述べる。

5.1 基本手話単語辞書の構築

手話表記文で使われる手話単語は, 日本語見出しから手話表現が容易に想起できる必要がある。特に人, 地域によって表現の差異²⁴⁾が少ないものが望ましい。そこで現存する手話の単語辞書の見出し語を再編集して, 基本手話単語辞書を作成した。

本研究では手話の単語集として最もよく使われている「わたしたちの手話」²⁵⁾(収録語数約4000語**)を

* 日本語の五十音に相当する表現方法で, 仮名に対して手形が決まっている。

** 収録語数は日本語の見出し語の数

基本に、2種類の電子手話辞書^{4),5)}(収録語数各々約700語)と手話初級講習会で使われているテキスト²⁶⁾(収録語数342語)の収録単語から日本語見出しを抽出した。その中から明らかに石川県内で通用しない語を除き、聾者や手話通訳の経験者の意見を参考にしながら、日本語見出しを見て躊躇することなく手話が表現できる3162語の手話単語を選定した。

辞書に含まれている主な項目を次にあげる。

日本語見出し 市販されている手話単語辞書から抽出した日本語の見出し。

読み 日本語見出しの読み。

語義 日本語見出しの語義。これはEDR日本語単語辞書²⁷⁾で定義されている語義より、該当するものを選択した。

5.2 手話単語辞書の補完

今回作成した基本手話単語辞書の語彙数は一般的な日本語文を機械翻訳するには不十分である。そこで、入力された日本語単語に対して、該当する手話単語が存在しない場合には、EDR電子化辞書²⁷⁾を使って類似語を獲得し、手話単語へ変換する。類似語を獲得する3つの手法の詳細を次に述べる。

5.2.1 同じ概念識別子を持つ単語の利用

EDR日本語単語辞書²⁷⁾の見出し語には概念識別子が付加されている。これは単語の概念を記号で表したもので、その単語の語義と見なすことができる。もし、同じ概念識別子を持つ見出し語があれば、これらの単語は同じ概念を共有していることを意味するので、類似語であると見なすことができる。そこで、同じ概念識別子を持つ単語を検索し、それらの見出し語から手話単語へ直接変換する。図2に、日本語単語「今」を、概念識別子(0f1b46)によって手話単語「現在」に変換する例を示す。

もし、複数の単語が類似語として獲得された場合には出現頻度^{*}の最も多い見出し語を採用する^{**}。

5.2.2 概念説明文の利用

概念識別子には意味を把握しやすいように説明文が付加されている。つまり1つの単語には概念識別子とその説明文である概念説明文が付与されている。そこで、この説明文より単語列を取り出し、手話単語へ変換する。図3に、日本語単語「高波」を、概念説明文を使って手話単語列「高い」「波」に変換する例を示す。

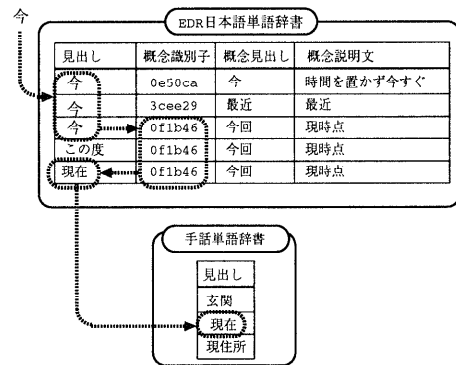


図2 同じ概念識別子を持つ単語を利用する

Fig. 2 Using the word has same identifier of original word.

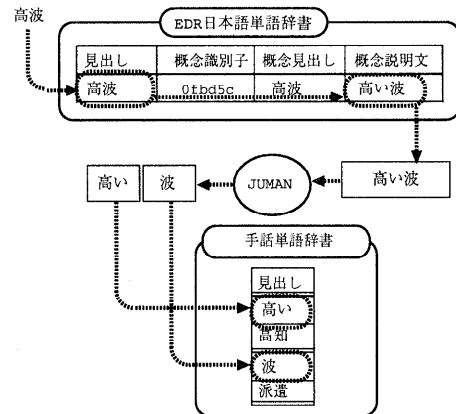


図3 概念説明文を利用する

Fig. 3 Using the definition sentence of a concept.

本手法は概念説明文を処理し、単語列を抽出するものであり、意味のある単語列をいかにして取り出すかが問題となる。その手法を6章で検討する。

5.2.3 上位概念の利用

EDR概念辞書²⁷⁾は概念識別子を階層構造として体系化した辞書である。この概念体系の中で、ある単語の上位階層にある単語は、元の単語を抽象化した単語であると考えられる。そこで見出し語の上位概念を類似語として、手話単語へ直接変換を行う。日本語単語「決勝戦」を、上位概念である「試合」に変換する例を図4に示す。

6. 概念説明文から類似語を獲得する手法について

類似語を獲得する方法を3つ述べたが、ここで概念説明文から類似語を獲得する手法について検討する。概念説明文から類似語を獲得するには、説明文を形

* 見出し語がEDRコーパスに出現する回数。EDR日本語単語辞書に記録されている。

** 8章の実験の結果では、この頻度による選択手法が原因となる失敗は起こらなかったため、妥当な方法であると考えている。

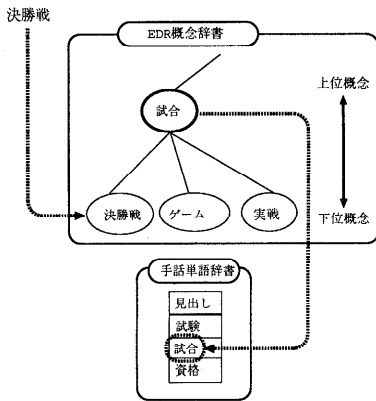


図4 上位概念を利用する
Fig. 4 Using the concept hierarchy.

態素解析し、自立語の単語列を抽出し、手話単語への直接変換を行う。概念説明文は、ほとんどの場合、文なので、1つの単語を複数の単語で説明することになる。つまりこの変換方法では意味のある形態素「列」を取り出す必要がある、複数の説明文の統合や有用な単語を残す手法が問題となる。そこで、この手法を検討するために2つの予備実験を行った。

6.1 予備実験1 (概念説明文の1本化)

手話単語と日本語単語の対応は1対1ではない。ある日本語単語を手話単語へ変換する際に、その手話単語が複数の日本語単語と対応しているために、それぞれの単語から異なった概念説明文が獲得できることがある。この場合には説明文を統合するか、1つの説明文を選択する必要がある。

たとえば「穏やかなだ」は EDR 日本語単語辞書では3つ以上の語義^{*}に分類されているが、そのうち2つ^{**}は手話では同じ語義と見なすことが適当である。この場合に、それぞれの語義に対応する概念説明文である「物事が静かで穏やかなさま」と「人の性格や事物が穏やかなさま」の2つの文から、有益な1つの単語列を生成する必要がある。

そこで EDR 日本語単語辞書で語義が複数個付与されている単語のうち、手話単語辞書に載っている80個の手話単語について、次の2つの方法で概念説明文の1本化を行い、手法の比較を行った。

- (1) 最も出現頻度の多い単語見出しの概念説明文を使う。
- (2) 複数の説明文で使われている単語を抽出して使う。たとえば、見出し語「治安」の概念説明は

^{*} 「物事が穏やか」、「性格が穏やか」、「気候が穏やか」などの語義が付与されている。

^{**} 「物事」と「性格」の穏やかな様子は手話では同じ表現をとる。

「国や社会の秩序が保たれて平穏であること」と「国や社会を平穏に治めること」の2つであるが、ここから「国」「社会」「平穏」を取り出す。

この予備実験の結果、後者の手法で手話単語が獲得できたのは、80個の見出し語の概念説明文に対してわずか24例であった。そのうち複数の形態素が取り出したのは3例であった。また、後者の手法で形態素を獲得した場合、前者の手法でもほぼ同様の結果が得られており、後者の手法が確実に前者の手法より良かった場合は1例しかなかった。

複数の概念説明文を1つの文に統合するのは人間でも難しいということもあり、後者の手法が特別に良いとは断定できなかった。よって、試作したシステムでは、前者の手法をとることにした。

6.2 予備実験2 (有効な形態素の抽出と主部の抽出)

概念説明文からは自立語を取り出し変換に利用するが、概念説明文には冗長なものがあり、的確な単語列を抽出する方法が必要である。

概念説明文は、ある単語の語釈文と見なすことができる。ここでは主部が単語の説明に重要な役割を果たしていると考え、フィルタにより主部のみを取り出した場合の効果を調べた。

フィルタなし すべての形態素を使用する。

フィルタあり 最後の名詞句のみを使用する。つまり、名詞か形容詞か副詞が0個以上続き、その後動詞が0個以上続く文字列を使用する^{***}。

また、このフィルタによりいくつかの形態素を除去した際に、説明文は単語を説明しているのだから、主部が存在しなければ意味のある説明文になっていないのではないかと考えた。そこで単語の最後の形態素を主部と見なし、主部がある場合とない場合の効果を調べた。

主部は必須 最後の形態素(主部)が手話単語に変換できた場合のみ使用する。

主部は任意 1つでも形態素が手話単語に変換ができたら使用する。

実験の結果を表2に示す。対象とした形態素は347個であり、「成功」は有意義な形態素列が獲得できた見出し語の数、「その他」はそれ以外の場合の見出し語の数を示している。

実験の結果からフィルタをかけたり主部を必須にした場合には、制約が厳しくなって変換される形態素が少なくなり、結果が悪くなることが分かった。特に主部を必須にした場合は、フィルタの効果がなくなって

^{***} サ変名詞は動詞として扱う。

表2 フィルタの効果
Table 2 Effect of filter.

	フィルタなし		フィルタあり	
	成功	その他	成功	その他
主部は必須	68	279	68	279
主部は任意	91	256	86	261

しまうほど厳しい制約となり、有用な形態素列を得ることが難しい。

よって、今回のシステムではフィルタは使わず主部は任意として概念説明文を処理している。

6.3 概念説明文からの類似語の獲得の流れ

予備実験の結果より、概念説明文を次の手順で自動的に処理して1つの単語列を生成する。

- (1) 概念説明文の1本化
複数の概念説明文が得られた場合、最も出現頻度の多い単語見出しの概念説明文を使う。
- (2) 抽出規則の適用
概念説明文は一般の文よりも形式が類似しているので、抽出規則により有用な単語を取り出しやすい形に加工する。抽出規則は、実験で使用した入力文に関係している概念説明文(309文)から、手作業により15種類を作成した。抽出規則は次の2つに大別できる。
 という規則 「(Xにおける)AというB」や「(Xで,)AというB」のような「という」の形式の文に対して、Aを固有名詞として指文字表記に変換し、Bの部分のみを説明文として処理する規則
 語尾規則 「Aすること」「Bという程度」の「こと」や「程度」を除去したり、「Aができる」を「A/可能*」、「Bという様」を「B/様子**」に変換するなどの語尾を処理する規則
- (3) 形態素解析
抽出規則で処理した概念説明文を形態素解析する。形態素解析にはJUMAN3.1²⁸⁾を利用して、出力結果の手作業による修正は行っていない。
- (4) 自立語の抽出
名詞、副詞、形容詞、動詞を取り出す。活用のある単語は終止形を使う。
- (5) 有効な形態素の抽出

*「Aができる」に相当する手話表記文。「可能」という見出し語は基本手話単語辞書に存在する。

**「Bという様」に相当する手話表記文。「様子」という見出し語は基本手話単語辞書に存在する。

残った形態素に対して、手話単語に変換できた形態素を使う。

7. システムの試作

本システムが日本語から手話表記文を生成する過程を順に述べる。また、その流れを図5に示す。

- (1) JUMAN3.1を利用して入力文の形態素解析を行う。
- (2) 形態素列にEDR電子化辞書の概念識別子を手作業で付与する。
入力された形態素は語義が決まっていないため、表層情報だけでは違う概念の単語に変換されてしまう可能性がある。そこで形態素に概念識別子を付与する。これで単語の語義を決定したことになる。この作業は理想的には自動化されるべきであるが、今回は類似語による変換を主に評価したいので、確実に手作業で処理した。
- (3) 自立語の抽出を行う。
特殊(句読点など)、助詞、助動詞や接辞の一部を変換しない形態素として除去する。
この後の処理では、変換できた形態素は処理が終わったと見なし、それ以降の処理は適用されない。
- (4) 句の変換規則を適用する。
今回のシステムでは一般に手話通訳や講習会のテキストでよく使われる表現から、日付と語尾に関する規則を手作業でまとめ、句の変換規則として使用した。
日付変換規則 「??月??日」の形式の形態素列を取り出して手話単語としてひとまとめにする変換規則にかける。
語尾変換規則 手話単語に変換すべき有用な情報を担っていると考えられる動詞の活用語尾(助動詞)と接尾辞、接頭辞の一部について、あらかじめリストに登録しており、該当する形態素が出てきたときには変換規則で処理する。
- (5) 手話単語への直接変換
入力文と手話単語辞書の見出し語にはEDR電子化辞書の概念識別子が付与されているので、単純な一致をとる。概念識別子が付与されていない形態素どうしは同じ語義であると見なしている***。数字は手話単語と見なす。

***つまり、入力された日本語単語側と手話単語側の両方に概念識別子が付与されていない場合には、EDR電子化辞書に存在しない語義が付与されていると見なしている。

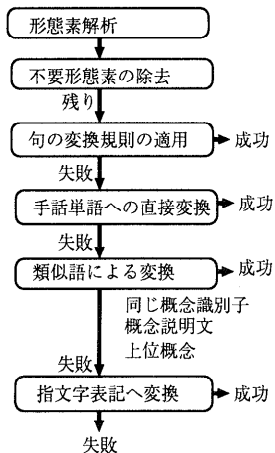


図5 変換過程の全体の流れ

Fig. 5 Processing flow of translation.

(6) 類似語による手話単語への変換

前述した3つの手法により、直接手話単語に変換できない形態素を、類似語を使って手話単語へ変換する。

- 同じ概念識別子を利用する。
- 概念説明文を利用する。
- 上位概念の見出し語や概念説明文を利用する。

(7) 指文字表記への変換

残りの変換できない形態素のうち、名詞である形態素は指文字表記^{☆1}に変換する。一般に指文字で表す単語は固有名詞や外来語が多いので、これらの単語はカタカナや漢字のみで表記されたものが多いと考え、平仮名が含まれている形態素は除外している。

以上の処理で、変換できない形態素は、このシステムでは変換できない(失敗)とする。

8. 実験と結果

入力文としてNHKニュース²⁹⁾の音声を書き下したものを91文用意した。形態素解析結果^{☆2}を修正し、語義を付与して、延べ1272個の形態素をシステムに入力した。

表3に結果を示す。この結果はシステムの出力結果を手作業で確認し、正しく手話単語に変換されている形態素を「正解」、間違った手話単語に変換されるか変換できなかった形態素を「失敗」として示している。

表3 実験結果

Table 3 Result of experiment.

変換処理	成功	失敗	その他
不要形態素			546個
手話単語へ直接変換	252個	0個	
句の変換規則で変換	59個	0個	
同じ概念識別子で変換	64個	4個	
概念説明文で変換	91個	137個	
上位概念で変換	2個	26個	
指文字表記に変換	37個	35個	
システムが変換できない			19個
計	505個	221個	

総形態素数1272個に対して約83%が変換されている。ただし、不要形態素の除去は実質的には例外処理なので、これを除くと約70%の形態素が変換されることになる。また、類似語による変換の対象となる形態素は415個であるが、そのうち約38%が正しく変換されている。

同じ概念識別子を使った変換では、「伝える」を「コミュニケーション」に変換する誤りが4回起きている。手話単語での「コミュニケーション」の語義が日本語見出しと微妙に違うためのもので、語義の付け間違いとも考えられるが、これを防止するのはかなり難しい。

概念説明文を使った変換の失敗のうち、最も多い原因が、説明文が不適当であることがあげられる。明らかな悪文として、見出し語がそのまま繰り返されるものがある。また逆の意味や単語を含んでいるために、逆の意味になってしまうもの^{☆3}やかなり冗長なもの^{☆4}がある。

しかし、一般の国語辞典の語釈文は複文であることが多く、「という」形式のような定型的な文であることは少ない。このような場合には有効な抽出規則や不要語を除去するフィルタの作成が困難であり、表層関係だけから、複文を統合して必要とする単語列を得るのは難しい。一方、EDR日本語単語辞書の概念説明文は明らかな悪文を除いては、比較的短くまとまっておき、文の形態も定型のものも多く、本システムには適当であると思われる。よって、利用しようとする説明文が悪文の場合には、次の段階へ進ませる手法の開発が必要であると考えられる。

上位概念を利用した手法ではほとんど類似語を獲得できなかった。上位概念の概念見出しはもともとEDR電子化辞書内に付与されていないことが多く、また概

☆1 日本語の五十音1文字1文字に相当する表現方法

☆2 「アルバイトスチュワーデス」が1つの形態素になるため、「アルバイト」と「スチュワーデス」に区切った。

☆3 例、打者：野球で、投手の投げた球を打つ人

☆4 例、土俵：相撲で、まわりを土の入ったたわらでまわす囲み、中に砂をしきつめた、力士が勝負する場所

念説明文もかなり抽象的で元の語の類似語が含まれていないことが多かった。

指文字表記での失敗の大部分は漢字熟語で、これらの単語は本来、類似語を使って変換されるべきものである。

システムが明らかに失敗とした形態素は感動詞（「ごきげんよう」など）と接尾辞のような名詞（「角さんばり☆」の「ばり」など）で、これらは辞書に登録するか、構文処理の段階で変換規則を適用して処理すべきであると考ええる。

9. 考 察

9.1 変換精度の向上

今回の実験では、概念説明文の形態素解析結果には語義を割り当てていないが、もし、語義を割り当てても、精度がほとんど上がらないことは確認している。語義を割り振って再度類似語による変換を行っても、獲得される単語の系列が循環してしまい、有用な手話単語へ結び付けることができないためである。類似語による手法では現在の変換精度が限界と考える。

類似語の獲得に失敗して変換できない形態素は別の方法を考えなければならない。実際の手話通訳において、日本語単語に直接対応する単語がなく、類似した単語も思い当たらない場合には、単純に文字の似ている単語への置換え（「政策」を「政治」にするなど）や、図像的に説明されることが多い。現在の基本手話単語辞書には漢字手話☆☆と呼ばれる手話は登録していない。日本語から手話への機械翻訳に使用するだけの辞書ならば漢字手話を登録することにより、単純に文字列の似ている単語へ置き換えることができると考える。また図像的な表現は概念説明文をそのまま出力する処理が考えられる。

9.2 抽出規則の作成

今回試作したシステムでは概念説明文を処理するために15種類の抽出規則☆☆☆を作り、309文を扱った。EDR日本語単語辞書全体では20万の概念説明文があるので、すべてを扱おうとすれば抽出規則を増やさなくてはならない可能性がある。

しかし、「という」形式の文を処理している抽出規則は、単純な構文的な情報を利用しているだけなので、ほとんど増えることはない。一方、残りの12種類は語尾に関する抽出規則である。そのうちの8種類は自立語（名詞、動詞、形容詞）以外を対象にしており、

残りの4種類は「できる（可能である）」「様子（状態である）」など手話単語に依存した表現を処理するものである。後者の規則は日本手話自身に依存しており、これ以上増やす必要はないと考える。よって、増やさなければならないと考えられる規則は語尾が自立語以外の文を処理する抽出規則である。

EDR電子化辞書全体の20万の概念説明文の語尾の品詞を調べると、自立語でない形態素は約150種類であるので、手作業で抽出規則が作成できない量ではないと思われる。

10. おわりに

本稿では日本語から手話への機械翻訳システムで問題になる辞書の貧弱さを解決する手法として、既存の大規模な日本語辞書を使い、類似語を使って手話単語辞書を補完する方法を提案した。本手法により単純な直接変換では手話単語に変換できなかった単語のうち約4割の単語を変換することができた。また、試作した日手機械翻訳システム全体では約7割の単語を変換することができた。

現在は、入力文の語義を手作業で付与しているが、システムとして実用化するためには自動的に語義決定まで行う必要があると考える。また、現在のシステムは単語単位での変換しかできないので、出力される手話文は日本語対応手話であり、自然な日本手話とはほど遠い状況にある。さらに、変換規則により強引に日本手話らしい文を生成している。また、形態素単位での変換なので繰返しも多い。今後は省略や照応表現を利用した、より自然な文を生成できるような方法も考えていきたい。

謝辞 手話単語の選定、記録などに関して、野々市町聴覚障害者協会の皆さんと手話サークル「てのひら」の皆さん、大家智子さんに多大な協力をいただきました。深く感謝します。

参 考 文 献

- 1) 市田泰弘：日本語の文法と語彙，日本語学，Vol.13, No.2, pp.25-35 (1994).
- 2) 鳥越隆士：手話通訳養成の現状と課題，日本手話研究所所報，No.9, pp.14-20 (1991).
- 3) 全国手話通訳問題研究会：手話通訳者の頸腕情報縮刷版，全国手話通訳問題研究会 (1995).
- 4) 神田和幸：日本語電子辞書，アルファメディア (1994).
- 5) 日本アイ・ビー・エム：はじめての手話，日本アイ・ビー・エム (1995).
- 6) 木村晴美，市田泰弘：はじめての手話，日本文

☆ 角さんのような

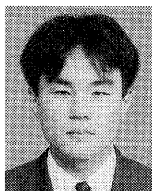
☆☆ 漢字1文字に対応する手話のこと。田，山，川など

☆☆☆ 「という規則」が3種類，「語尾規則」が12種類

- 芸社 (1995).
- 7) 加藤雄士, 神田和幸, 長嶋祐二, 市川 熹, 黒川隆夫: 手話工学の現状と将来の研究課題, 計測自動制御学会ヒューマン・インタフェース部会 Human Interface News and Report, Vol.12, No.1, pp.37-44 (1997).
 - 8) 長嶋祐二, 小野寺卓, 長嶋秀世, 寺内美奈, 大和玄一: 指文字認識に関する基礎的検討, 信学技報, HCS92-41, pp.23-29 (1992).
 - 9) Ohira, E., Sagawa, H., Sakiyama, T. and Ohki, M.: A Segmentation Method for Sign Language Recognition, *IEICE Trans. Inf. & Syst.*, Vol.E78 D, No.1, pp.49-57 (1995).
 - 10) 萩原芳彦, 堀口 進: 指手骨格モデルを用いた指文字認識, 日本手話学会第 21 回大会予稿集, 日本手話学会, pp.32-35 (1995).
 - 11) Lee, J. and Kunii, T.L.: Generation and Recognition of Sign Language Using Graphic Models, *IISF/ACM Japan International Symposium: Computers as Our Better Partners*, pp.96-103 (1994).
 - 12) 西川 誠, 寺内陸博: 手話の空間配置を考慮した日本語・手話翻訳に関する検討, 信学技報, HC94-94, pp.55-62 (1995).
 - 13) 藤重栄一, 黒川隆夫: 意味ネットワークを媒介とする日本語・手話翻訳のための日本語処理, 計測自動制御学会ヒューマン・インタフェース部会 Human Interface News and Report, Vol.12, No.1, pp.45-50 (1997).
 - 14) 安達久博: 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, 信学技報, NLC92-47, pp.17-24 (1992).
 - 15) 安達久博, 吉澤昌三, 藤田均子, 松本 崇, 鎌田一雄: 手話ニュース文の分析と手話変換処理, 自然言語処理研究会報告, NL91-3, pp.17-24 (1992).
 - 16) 鎌田一雄, 松本 崇, 山上淳二: 日本語・手話変換の基礎的検討, 信学技報, HC93-75, pp.69-74 (1994).
 - 17) 藤井昌紀, 岩波太郎, 亀井 了, 長嶋祐二: 2次元の時空間画像による手話の大局的な調動認識, 第 11 回ヒューマン・インタフェース・シンポジウム論文集, pp.179-184 (1995).
 - 18) 本名信行, 加藤三保子: 手話の表記法について, 日本手話研究所所報, No.4, pp.2-9 (1990).
 - 19) 相良かおる, 木村晋二, 渡邊勝正: 手話単語の日本語による記述とその応用について, 信学技報, HCS95-16, pp.59-66 (1995).
 - 20) 伊東雋祐: 手話研究草創期, 日本手話研究所所報, No.13, pp.23-34 (1993).
 - 21) 神田和幸: 手話学講義, 福村出版 (1994).
 - 22) 全日本ろうあ連盟: 新中級手話教室, 全日本ろうあ連盟 (1991).
 - 23) 江原暉将, 田中穂積: 機械翻訳における自然言語処理, 情報処理, Vol.34, No.10, pp.1266-1273 (1993).
 - 24) 高田英一: 手話地域格差調査報告書を考える, 日本手話研究所所報, No.6, pp.6-12 (1990).
 - 25) 全日本ろうあ連盟: わたしたちの手話 1-10, 全日本ろうあ連盟 (1987).
 - 26) 石川県聴覚言語障害者福祉協会: 手話テキスト (入門編), 石川県聴覚言語障害者福祉協会 (1995).
 - 27) 日本電子化辞書研究所: EDR 日本語単語辞書・概念辞書第 2 版, 日本電子化辞書研究所 (1995).
 - 28) 松本裕治, 黒橋禎夫, 山地 治, 妙木 裕, 長尾 真: 日本語形態素解析システム JUMAN 使用説明書 version 3.1, 京都大学工学部・奈良先端科学技術大学院大学 (1996).
 - 29) 日本放送協会: NHK ニュース「聴力障害者のみなさんへ」(1994). テレビニュース番組 NHK 教育 19:50~20:00.

(平成 9 年 6 月 30 日受付)

(平成 10 年 1 月 16 日採録)



徳田 昌晃 (学生会員)

1971 年生. 1994 年金沢工業大学情報工学科卒業. 1996 年北陸先端科学技術大学院大学情報科学研究科博士前期課程修了. 現在, 同大学博士後期課程在学. 自然言語処理, 特に手話の言語処理の研究に従事. 人工知能学会, 言語処理学会, 日本手話学会, 全国手話通訳問題研究会各会員.



奥村 学 (正会員)

1962 年生. 1984 年東京工業大学工学部情報工学科卒業. 1989 年同大学院博士課程修了. 同年, 東京工業大学工学部情報工学科助手. 1992 年北陸先端科学技術大学院大学情報科学研究科助教授, 現在に至る. 工学博士. 自然言語理解, 知識表現, 語彙知識獲得に関する研究に従事. 人工知能学会, AAI, 言語処理学会, ACL, 認知科学会, 計量国語学会各会員.