

方言から共通語への翻訳システムに関する基礎的研究

佐藤 守[†] 横山 晶一^{††} 西原 典孝^{††}

[†], ^{††} 山形大学工学部 〒992-8510 山形県米沢市城南 4-3-16

E-mail: [†]m01833@eiefs.yz.yamagata-u.ac.jp, ^{††}{yokoyama,nisihara}@yz.yamagata-u.ac.jp

あらまし 日本には北海道から沖縄まで様々な方言が存在する。昔に比べ人の移動が多くなり、山形にも県外や海外から移り住む人々や留学生が増加している。地方で生活するためには、必然的に地元の人々と方言によるコミュニケーションが必要になる。方言理解の補助手段として機械翻訳は有効である。ここでの機械翻訳は、多国語間の翻訳レベルではなく同じ日本語内での翻訳であるため、一般の機械翻訳に比べてそれほど困難ではないと考えられる。我々のグループは既に共通語から方言へ、品詞ごとの置き換えて、ある程度自然な訳が得られる翻訳システムを作成している。本稿ではこれとは逆の方言から共通語への翻訳というアプローチを試みる。原言語は村山方言とし、機械処理の観点から基礎的な研究を行った。

キーワード 村山方言, 共通語, 翻訳システム

Fundamental Research on the Translation System from a Dialect to Common Language

Mamoru SATO[†], Shoichi YOKOYAMA^{††}, and Noritaka NISHIHARA^{††}

[†], ^{††} Faculty of Engineering, Yamagata University Jonan 4-3-16, Yonezawa-shi, Yamagata, 992-8510 Japan

E-mail: [†]m01833@eiefs.yz.yamagata-u.ac.jp, ^{††}{yokoyama,nisihara}@yz.yamagata-u.ac.jp

Abstract There are various dialects all around in Japan. The movement of people has been increasing, and many people and foreign students move from other prefectures and around. They need communicate with native people using the dialect. Machine translation is one of the valid means to assist the understanding the dialect. Here, "machine translation" means that it is not the translation between different languages, but between the dialect and the common language inside in Japanese, and then it would be not so difficult. Our group has already made the prototype of the system from the common language Japanese to the dialect. This papaer describes the converse approach, that is, the system from the dialect to the common language. We select Murayama dialect, and have fundamental studies for the system.

Key words Murayama dialect, common language, translation system

1. ま え が き

日本には北海道から沖縄まで様々な地方の方言が存在する。昔に比べ人の移動が多くなり、地方の人が他の地方に住むというケースも増えている。当然、地方に住む人は、普段その地方の方言を用いて会話する。ここで、地方の人が使用している言葉を理解できないという事態が発生する。地方に住む若い世代の人たちが話す言葉は、方言語彙よりも共通語の語彙のほうが多く含まれるため、全く理解できないということはまずない。しかし、年輩の人たちが話す言葉は、圧倒的に方言語彙のほうが多く含まれ、発音する際のなまりも生じるため、全く理解できない場合

も多々ある。最近では、ここ山形にも海外から移り住んでくる人や、留学生が増加している。同じ日本人同士でも、これだけのギャップが生じているわけであるから、共通語しか学んできていない外国人にとっては、もっと理解が困難であるはずである。その地方で生活していくためには、その地方に住む人たちとコミュニケーションする機会が必ずある。そこで方言理解の手助けとして必要となってくるのが、機械翻訳システムである。

機械翻訳といっても、ある言語から別の言語への翻訳といったレベルではなく、同じ日本語内での翻訳であるため、品詞ごとの語彙の置き換えによってある程度自然な翻訳が得られることが報告されている [1]。これは共通語から方言への翻訳システム

であるが、本稿では方言から共通語への翻訳システムの構築という逆のアプローチを試みる。共通語を原言語とした場合、「茶筌 [3]」を利用すれば形態素解析は高精度で成功するため、翻訳システムの構築は、方言から共通語への翻訳システムの構築より容易であると考えられる。ここでは方言を形態素解析するために、既存の形態素解析システム「茶筌」を利用する。「茶筌」を方言語彙の登録などによりカスタマイズし、正しく形態素解析できた文章を中心に翻訳システムの構築を行う。その際に、多義語の方言語彙を翻訳可能にするため、文献 [10] で作成された辞書を利用する。

この研究の目的は、次の通りである。

(1) 方言は、語彙や文法の面で、古い日本語を保存していると言われている。もしそうであれば、系統的な規則で記述が可能ではないか。また、そのような規則で記述した文法が共通語の記述に示唆を与えるのではないか。

(2) 方言は主として話し言葉として伝えられてきた。したがって、方言文法を書くことは、話し言葉の文法を作ることになる。

(3) 方言と共通語を比較することによって、日本語の性格がより明らかになる。さらに両者の機械翻訳システム的なものが作成でき、翻訳の研究に有用である。

以上のような考えから、方言を機械処理するために、村山方言を例にとって機械処理の観点から基礎的な研究を行った。

本稿の構成は次のようになっている。まず次節では、機械翻訳の定義について若干触れる。3 節においては、村山方言の概要について述べる。4 節では、村山方言から共通語への翻訳アルゴリズムと、その翻訳システムの現段階での問題点について考察する。

2. 機械翻訳とは [5,6]

2.1 機械翻訳の立場

翻訳とは、おおよそ「ある言語 (原言語) のテキストを別の言語 (目的言語) の等価なテキストに置き換えること」として定義できる。ここで問題となるのは、「等価」というものをどう捉えるかということである。どのような等価性を重視するかによって、それぞれの翻訳の立場が生じる。

(1) 単語と構造の等価性を重視する翻訳

与えられた文の単語と構造を目的言語の対応する単語と構造に移し変えるということになる。いわゆる「逐次訳」に相当する翻訳である。

(2) 意味内容の等価性を重視する翻訳

与えられた文がもつ意味内容と同一の意味内容をもつような目的言語の文を作り出すことである。「逐次訳」を越えた「意訳」に相当する。

現在の機械翻訳が目指しているのは、(1) を翻訳の基本とし、一部 (2) の翻訳を実現しようとするものである。本研究において

も、これと同様の翻訳を目指す。

2.2 形式的対応のずれと翻訳の難しさ

共通語と方言では、語彙の対応のずれや単語と句の対応などの形式的対応のずれが見られるため、翻訳は困難になる。

● 語彙の対応のずれ

共通語で「(学校に) 入学する」、「(仕事が終わって) 帰る」、「(食べ物を) 食べる」などの動詞を村山方言では「あがる」で表す。これは共通語と村山方言の単語が 1 対 1 に対応しておらず、意味する内容にずれがあるということを示している。

● 単語と句の対応 (層のずれ)

村山方言の「うるがす」という語は共通語では「水につけて柔らかくする」という句に対応する。

3. 村山方言概説 [7,8]

山形県内は、村山、置賜、最上、庄内の四地方に大きく分かれ、各地方に特有の方言がある。庄内地方は、他の三地方とは、大きく異なり、北奥羽方言に属している。村山方言は、山形市を中心とした、内陸部の中央に位置する村山地方の方言である。この地方は、17 世紀以降小藩が分立し、変動の激しい治世が続いたこともあって、音韻や文法で南奥羽方言的ではあるが、語彙では北奥羽方言的な要素をもちあわせたりする。このような理由で村山方言は、庄内方言、置賜方言とは語彙や文法で異なる面がある。本研究ではそれらを解明する。

3.1 村山方言文法

本節では村山方言文法において特徴的なものについて説明する。

3.1.1 否定表現

共通語の否定表現「～ない」に対応する方言は「～ね」である。動詞の未然形に助動詞「ね」を接続することによって否定表現となる。しかし、五段活用動詞で、最後に「る」がつくものや、可能・受け身の「れる・られる」に接続するときは、「～らね」ではなく、撥音化して「～んね」となる。

(1) 山田さん、今日、休みだがもす んね。(山田さん、今日、休みかもしれない)

(2) 人がら笑わんねようにさんな ね。(人から笑われないようにしなければなら ない)

3.1.2 意志・推量・可能表現

共通語の意志・推量・可能表現「～う、～よう、～かしら」に対応する方言は助動詞「べ」である。動詞の終止形に、この助動詞「べ」を接続すると勧誘表現になる。なお、最後に「る」のつく動詞の場合は、促音化が起こり、「っべ」となることが多い。

(1) 七日町さ行ぐんなら、一緒に行ぐべ。(七日町に行くなら、一緒に行こう。)

(2) さあ、ごはん食べっべ。(さあ、ごはんを食べよう。)

3.1.3 禁止表現

共通語の禁止表現「～な」に対応する方言も、同様に終助詞「な」である。最後に「る」のつく動詞に接続した場合、その動詞は「～るな」ではなく、撥音化して「～んな」になる。

(1) こっちさ来んな!(こっちに来るな!)

(2) そだなごどすんな!(そんなことするな!)

3.1.4 強意表現

方言では「はー」を文末に接続すると意味が強くなる。また、「ずー」を文末に接続すると、より強くなる。「はー」は自分の行為について使用することが多いが、「ずー」は相手の行為・言動について使用することが多い。

(1) きょうは具合わるいがら休むはー。(きょうは具合がわるいから休みます。)

(2) そだなごどすんな ずー。(そんなことするなよ。)

3.1.5 形容詞の促音化

村山方言では、3音節からなる形容詞は促音化する。

(1) いっだい。(いたい。)

(2) ほっそい。(ほそい。)

3.1.6 有声化

有声化は東北のことば(ズーズー弁)の特徴として最も顕著なものである。有声化するのはカ行とタ行の音である。

(1) ほんとは、今週3回あったんだけど、バイトなぐなって(ほんとは、今週3回あったんだけど、バイトなくなって)

(2) これ、借りてんだけど、これ使えないと、バイトんどき困る。(これ、借りてんだけど、これ使えないと、バイトのどき困る。)

3.2 研究手法

本研究では以下のような手順でシステム構築から評価まで行う。

- (1) 茶筌を村山方言用にカスタマイズする
- (2) 翻訳システム・翻訳辞書を作成する
- (3) 例文収集と翻訳実験を行う
- (4) システムの評価関数を作成し、評価する

4. 村山方言から共通語への翻訳アルゴリズム

4.1 翻訳処理の流れ

村山方言から共通語への翻訳処理の流れとしては、まず村山方言を茶筌で形態素解析し、その形態素解析された品詞ごとに翻訳辞書を参照して語彙を検索する。登録されてある場合は、方言の語彙を共通語に置き換える(方言の種類によって異なる)。全ての形態素について、この処理を施したあとに置き換えた語彙を含む全てを連結して共通語の文章として出力する。文章中に動詞が含まれる場合は、まず動詞の結合価を読み込む。図1にその流れ図を示す。

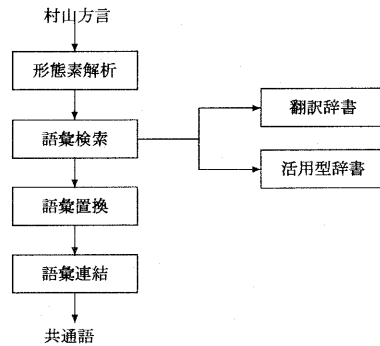


図1 村山方言から共通語への翻訳処理の流れ

Fig.1 Flow of translation processing from Murayama dialect to common language

4.1.1 本研究システムの仕様

本システムは、Perlで記述されており、内部で茶筌を実行している。結果は標準出力に村山方言と、それを翻訳した共通語を表示する。

4.1.2 前処理

茶筌の村山方言辞書を作成する際に、名詞に意味素性の情報を付加するために、品詞分類辞書(grammar.cha)に追加書き込みを行った。具体的には、形態素解析したときに「名詞-一般-(*)」というように(*)の部分に意味素性が付加されるように、一般の下位層に各意味素性名を追加した。カスタマイズした茶筌による形態素解析結果を以下に示す。

>びつき

>びつき びつき びつき 名詞-一般-動物

>EOS

4.1.3 結合価と意味素性

本研究で使用する結合価と意味素性について説明する。

- 結合価

ある動詞を基準とした場合に、共に出現する名詞・格助詞の情報をまとめたもの。

例

N[HUM].+N[CON](は)+V

これは、動詞「かます」を結合価で表現したものである。ここで、N は名詞を表し、[] の中は意味素性を表す。() の中に記述してあるのは格助詞であり、は主格の助詞「が」が省略されていることを表す。V は動詞を表す。

• 意味素性

名詞の持つ意味を表したもの。以下に名詞の意味素性の一部を示す。

表 1 村山方言で用いる名詞の意味素性の一部

Table 1 Examples of semantic features of noun used in Murayama dialect

略号	Full Name	素性名
ANI	Animal	動物
HUM	Human	人間
ORG	Organization	組織・機関
LOC	Location	場所
DIR	Direction	方角

4.2 茶釜辞書作成支援ツール

周知のとおり、茶釜は日本語の共通語を形態素解析するために開発されたフリーソフトである。本研究では、原言語が村山方言であるため、茶釜で形態素解析するには村山方言語彙の辞書を作成して再コンパイルする必要がある。そこで、その各品詞ごとの方言語彙辞書を手動で作成するには時間がかかり過ぎるので、手動で辞書作成する際に、この手間を少しでも省くことが必要となってくる。

(品詞 (名詞 一般)) ((見出し語 (蛙 3755)) (読み カエル))

具体的には、上記の下線部をユーザーが入力し、それ以外の部分を自動生成するものである。補助機能として、品詞名、見出し語、読み、とそれに対応する共通語を入力すると、その品詞の共通語辞書を読み込んで対応する共通語のコストを検索するコスト検索機能を付加した。もちろん新規語彙の登録だけでなく、既存の辞書ファイルの編集や削除も行なえる。

4.3 語彙データベース

4.3.1 語句の定義

• 訳語数

本システムで使用する翻訳辞書に登録されている、ある方言に対応する共通語訳の数を表す。この値が 2 以上のときは次節で述べるアルゴリズム、動詞の結合価を利用して適切な共通語を選択する。

• 方言の種類

次節で述べる 3 つの種類に分類される。翻訳辞書には、共通語が

規則的に変化した方言を 1、共通語と全く異なる語彙に変化した方言を 2、多数の共通語に対応した方言を 3 として登録する。

4.3.2 データベースの特徴

村山方言から共通語へ翻訳する際に参照する語彙データベースについて説明する。データベースのタイプとして、リレーショナル (関係) データベースという、データを関係と呼ぶ一種の表で表現する形式を採用した [9]。

このデータベースには、翻訳辞書から読み込まれた各品詞ごとの、方言、共通語、訳語数、方言の種類などの語彙情報が格納されている。翻訳辞書を作成する際には、文献 [2,7] を参考にした。翻訳辞書の登録形式で、共通項目は方言語彙、訳語数、共通語、読みの 4 項目である。その他の項目として、名詞の場合は“意味素性”、動詞の場合は“意味・用法”、活用する品詞の場合は“活用型”、“語幹”、動詞の場合は“結合価”の項目が追加される。以下に名詞の語彙データベースの一部を示す。

表 2 名詞の語彙データベース一部

Table 2 Some examples in noun database

方言	方言の種類	訳語数	共通語 1	意味素性 1	読み 1
おなご	2	1	女	HUM	オンナ
こだぶ	1	0	-	PRO	-
じゃね	2	1	お金	PRO	オカネ
びつき	2	1	蛙	ANI	カエル

4.4 翻訳アルゴリズム

村山方言を大きく分類すると、以下の 3 つのグループに分けられる。

- (1) 共通語が規則的に変化した方言
- (2) 共通語と全く違う語彙に変化した方言
- (3) 多数の共通語に対応した方言

本節では、この 3 つのグループの方言について、各々の翻訳アルゴリズムを説明する。以下、方言が正確に形態素解析されてあることを前提に議論をすすめていく。

4.4.1 Perl 組み込み関数を用いた翻訳

まず始めに、Perl の組み込み関数 pack,unpack を使用して文字コードを取得し、そのコード操作によって、規則的に共通語が方言に変化した語彙を共通語に変換する。

文字列の文字コードを取得する際には、unpack 関数を用い、文字コードから文字列を取得する際には pack 関数を用いる。

• 共通語が有声化した方言

日本語を文字コードで表すと、濁音は清音より 2 バイト目が 1 多い値となっている。つまり、濁音の 2 バイト目から 1 を減らしてやれば、清音の文字コードと一致する。以下に日本語 (EUC) 文字コード表の一部を示す。

この規則性を利用して、が行、だ行の濁音を清音に浄化することで方言を共通語に翻訳する。

表3 文字コード表(EUC)

Table 3 Character-code table(EUC)

か	a4ab	が	a4ac	た	a4bf	だ	a4c0
き	a4ad	ぎ	a4ae	ち	a4cl	ぢ	a4c2
く	a4af	ぐ	a4b0	つ	a4cd	づ	a4c5
け	a4b1	げ	a4b2	て	a4c6	で	a4c7
こ	a4b3	ご	a4b4	と	a4c8	ど	a4c9

清音 = pack("H*", 1.(r-1))

1:濁音の1バイト目の文字コード

r:濁音の2バイト目の文字コード

● 共通語が促音化した方言

3音節(読みの文字列長が3)からなる共通語の形容詞は、促音化して方言になっている。つまり、読みの文字列長を取得し、それが4のときは「っ」を削除して文字列を切り詰めることによって翻訳可能となる。

これによって、規則的に変化した語彙を共通語に翻訳する際に、perlの組み込み関数を使用しているため、翻訳辞書に訳語を登録する必要がなくなる。

4.4.2 単純置換による翻訳

村山方言と共通語が1対1に対応している場合、村山方言をその共通語で置換する。活用しない品詞の場合は、そのまま置換するだけでいいが、活用する品詞の場合は、まず語幹を置換し、村山方言の活用形に対応した共通語の活用語尾を付加する。

● 活用しない品詞の場合(例:名詞「びつき」)

びつき → 蛙(かえる)

● 活用する品詞の場合(例:動詞「ごしゃぐ」)

ごしゃ +	未然形	が	⇒ 怒 +	未然形	ら
	連用形	ぎ		連用形	り
	終止形	ぐ		終止形	る
	連体形	ぐ		連体形	る
	仮定形	げ		仮定形	れ
	命令形	げ		命令形	れ

4.4.3 動詞の結合値を利用した適切な置換による翻訳

2.2で説明したように、村山方言と共通語が1対1に対応していない場合もある。このような場合は、単純置換では適切な翻訳はできない。そこで、本アルゴリズムでは動詞の結合値を利用して適切な翻訳を試みる。ここでは、共通語の訳を複数持つ、助詞「さ」と名詞「いぎ」を例に説明する。

● 助詞「さ」の適切な翻訳

村山方言の助詞「さ」は、意味・用法によって異なった共通語に翻訳される。以下の二文を考える。

例文1: 東京さ行く。

例文2: 北さ行く。

動詞「行く」の結合値は以下の2式で表せる。

$N[ANI/HUM/ORG]_{-}(+N[LOC] \text{ さ})(+N[PRO] \text{ で})+V$

$N[ANI/HUM/ORG]_{-}(+N[DIR] \text{ さ})(+N[PRO] \text{ で})+V$

例文1は、格助詞「さ」が付属している名詞「東京」の意味素性が場所(LOC)であるから、意味・用法が場所である共通語の格助詞「に」に翻訳される。例文2も同様に名詞「北」の意味素性を見てみると、方角(DIR)であるから、意味・用法が方角である共通語の格助詞「へ」に翻訳される。つまり結合値に含まれる名詞と格助詞の組を調べて、その名詞の意味素性と一致した場合に、それに付随する格助詞に置換するわけである。

● 名詞「いぎ」の適切な翻訳

村山方言の名詞「いぎ」は、文脈によって異なった共通語へ翻訳される。以下の二文を考える。

例文3: いぎが降る。

例文4: いぎば吐ぐ。

動詞「降る」、「吐ぐ」の結合値を以下に示す。

－ 動詞「降る」の結合値

$N[NAT/PLA/PRO]_{-}+V$

－ 動詞「吐ぐ」の結合値

$N[ANI/HUM]_{-}(+N[PHE] \text{ を})+V$

例文3には動詞「降る」が含まれており、その結合値を参照すると助詞「が」(で省略されている)が付属する名詞の意味素性は、自然物(NAT)か植物(PLA)か生産物(PRO)である。したがって、この場合の名詞「いぎ」は、意味素性が自然物である共通語の「雪」に翻訳される。例文4には動詞「吐ぐ」が含まれており、その結合値を参照すると助詞「を」が附属する名詞の意味素性は、現象名詞(生理)である。したがって、この場合の名詞「いぎ」は、意味素性が現象名詞である共通語の「息」に翻訳される。

このように、他の語彙の場合も翻訳辞書に訳語数、複数の共通語、意味素性(名詞の場合)または意味・用法(助詞の場合)という形式で登録しておけば、動詞の結合値によって適切な翻訳が可能となるはずである。

4.5 本システムによる実行例

実際に計算機上での実行結果を以下に示す。ここで、inputはシステムに入力した村山方言、outputはシステムが出力した共通語を表す。実行例1には正確な共通語訳が得られた結果を示

し、実行例 2 には正確な共通語訳が得られなかった結果を示す。
 例文は文献 [4,8] にあるものを用いた。

表 4 本システムによる実行例 1

Table 4 Examples of machine translation 1

例 1	input: 明日東京に行く。 output: 明日東京に行く。
例 2	input: 北さ行く。 output: 北へ行く。
例 3	input: このかきあがいがいげんども洗い。 output: このかきあかいかいげんども洗い。
例 4	input: いぎば吐く。 output: 息を吐く。
例 5	input: いぎが降る。 output: 雪が降る。
例 6	input: あのおなごめんこい。 output: あの子かわいい。
例 7	input: このみがんつたい。 output: このみかん冷たい。
例 8	input: 高校さあづまらね。 output: 高校に集まらない。
例 9	input: あいづじえねねおどご。 output: あいつお金なき男。

表 5 本システムによる実行例 2

Table 5 Examples of machine translation 2

例 10	input: 明日家さ行く。 output: 明日家に行く。
例 11	input: 北さ行くべ。 output: 北へ行くべ。
例 12	input: んだず。 output: そうですず。
例 13	input: んねが? output: 違いますが?
例 14	input: いぎ。 output: 雪。

実行例 2 のほうは正しく翻訳されなかった場合である。例 10 については、「家」という名詞について意味素性が付加されていないため、結合価を利用して適切な訳語を選択できず、一番最初に登録されてある訳語「に」を選択してしまったのである。例 11 については、助動詞「べ」を茶釜の辞書ファイルに登録していないために、「べ」が未知語として認識され、「べ」の部分がそのまま出力されてしまったために、不完全な訳になっている。例 12 については、例 11 と同様に、「ず」が「助動詞 特殊・ヌ」というように誤認識されて、かつ翻訳辞書に未登録な語彙であるために、そのまま出力され不自然な訳になっている。例 13 についても、方言語彙として未登録だったために、「が」が助詞ではなく接続詞として認識され、そのまま出力され不自然な訳になっている。例 14 については、複数の訳語を持つ方言語彙「いぎ」だけ入力されていて、動詞が含まれていないために適切な訳語を選択できず、例 10 と同様に翻訳辞書に一番最初に登録さ

れている訳語「雪」を選択してしまったのである。

5. む す び

語順の入れ替えがなく、品詞ごとの変換・置換によって、村山方言から共通語への翻訳がある程度可能なシステムを作成した。共通語が有声化・促音化・撥音化して方言に変化したものは、文字コードという観点から Perl の組み込み関数を利用して、その文字コードを操作することによって、他の文字に変換することで翻訳を行なうことができた。これは、翻訳辞書に訳語を登録する手間とメモリが節約されるというメリットがある。また、ある程度方言を系統的に記述したことになる。

本システムでは、村山方言用にカスタマイズした茶釜で正しく形態素解析できた文章を中心に翻訳実験を行ってきた。どんな場合でも正しい共通語が得られるのは、現段階では共通語が規則的に変化した方言と、一意の共通語訳を持つ方言のみである。複数の共通語訳を持つ方言をその文章に適切な訳語を選択するには、動詞が必要となってくる。本研究では、動詞の結合価を利用してその文章に適切な訳語を選択しているため、動詞が含まれない場合は適切な訳語が選択されない。また、動詞が 2 つ以上含まれた場合にも適切な対応がまだできない。

今後の方針として、まず茶釜を、村山方言を共通語の解析精度と同程度に解析できるようにカスタマイズすることが必要である。そのために、茶釜の辞書作成の際に、村山方言だけでなく、他地方の方言の辞書も作成できるように辞書作成支援ツールの開発についての研究も同時に行っている。翻訳システムで精度を上げるには、まず翻訳辞書の語彙数を増加させなければならない。そのためには、形態素解析の精度を上げる必要があり、茶釜用の辞書に方言語彙を大量に登録しなければならない。人手で登録する際に、少しでも速度を上げるためのツールである。このツールを完成させ、語彙数を増やす予定である。

今回の翻訳に用いた文章は文献 [7,14] から引用したものであるが、方言と共通語が全く異なった文章になったものは、ほとんどない。これは方言語彙が衰退していることを示している。衰退する原因としては、親から子に言葉を教える際に方言を使用しなかったためと考えられる。機械処理という観点から方言を見直したときに、新たな展望が開けることが望まれる。

文 献

- [1] 池田麻生: 方言機械翻訳システムの構築, 山形大学卒業論文 (2002)
- [2] 尾花沢市「昔を語る会」: 尾花沢の方言 (1992)
- [3] 形態素解析システム「茶釜」, 奈良先端科学技術大学院大学
- [4] 高木裕子: 定住型外国人を対象にした「地域共通語」教材開発に関する研究, 平成 10 年度～平成 12 年度科学研究費補助金基盤研究 (B)(2) 研究成果報告書 (2002)
- [5] 田中穂積監修: 自然言語処理: 基礎と応用, 電子情報通信学会 (1999) pp248-280
- [6] 長尾真編: 自然言語処理, 岩波書店 (1996) pp457-485
- [7] 平山輝男他: 山形県のことば, 明治書院 (1997)
- [8] 山形地域語研究会: 山形ことばを学ぼう, 山形地域語研究会 (2001)
- [9] 横田一正・宮崎収兄: 新データベース論, 共立出版株式会社 (1994) pp35-40
- [10] 我妻信博: 結合価と意味素性を考慮した置換方言の機械処理用辞書, 山形大学卒業論文 (1998)