共通語から山形村山方言への翻訳システム

横山晶一† 尾形真美†
（山形大学工学部）
yokoyama@yz.yamagata-u.ac.jp

１．はじめに

方言は、地域に根ざした生活語として重要な役割をになってきた。一時共通語の普及をはかるために方言を圧迫した時代もあったが、今日では、地方の文化の活性化などのために、あえて方言を利用する場合も多い。

言語学的には、方言学という学問体系が確立しており[1-3]、音韻やアクセント、語彙、文法など、さまざまな面から研究が行われている。

また、地方へ定住する外国人が日本語習得をする過程で、共通語と方言の使い分けができずに混乱するのを防ぐための教材開発なども行われている[4]。

このように、言語学的、教育学的な面からの方言研究は広範囲に行われているが、機械処理の面から方言を扱った例はそれほど多くはない。

我々のグループでは、山形県内の各地の方言を対象に、共通語と方言の間の翻訳システムを作成し、発表してきた[5-7]。それは、次のような考え方に基づいたからである。

(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[1]。規則的な解明が現代語に対する示唆を与える可能性がある。
(2) 方言は主として話し言葉である。話し言葉としての方言文法を作成できれば共通語の話し言葉文法にも影響が与えられる。
(3) もともと同じ言語である共通語と方言の間の翻訳システムを作ることによって、言い換えにある程度の寄与ができる。

しかしながら、これまで作成したシステムは、語彙も余り多くなく、プロトタイプの域を出なかった。本稿では、共通語を山形県の代表的な方言である村山方言に翻訳するシステムについて述べる。このシステムは、現時点で約 3000 語の方言語彙を備え、これまでよりはより実用に近づいたシステムであると考えている。

Machine Translation System from Standard Japanese to Murayama Dialect in Yamagata Prefecture † YOKOYAMA Shoichi and OGATA Mami, Yamagata University

２．山形村山方言概説

東北地方の方言は、青森、秋田、岩手を中心とする北奥羽方言と、宮城、山形、福島を中心とする南奥羽方言に大きく分けられる[8,9]。後者の特徴としては、典型的な東北弁として知られる「ズーズー弁」がある。

山形県内を見ると、日本海に面した庄内地方は北奥羽方言に属し、しかも江戸時代に北前船が出入りした関係上、大阪弁の影響を受けていると言われている[10,11]が、ここと山を隔てた内陸部の３地方（村山地方、置賜地方、最上地方）は南奥羽方言に属している。

村山方言は、山形市を中心とする村山地方の方言で、この地域が江戸時代には天領であったり、他地方との交流が盛んに行われていたこともあって、他の地方よりやや方言色が薄まっている印象を受けるが、外部の者には典型的な東北弁の印象を受ける。以下にこの方言の特徴について簡単に述べる。

２．１．音声面での特色

音声面では、いわゆる「ズーズー弁」的な特徴を持っており、語中や語尾にある共通語のカ行やタ行の音がガ行、ダ行の濁音に変化する。

(文 1) このかぎあがいけんどもしぶい。
（この柿赤いけれども渋い）
(文 2) うだわね。（歌わない）

３音節から成る形容詞は促音化する。

(文 3) このみがん、やっすいな。
（このみかん、安いな）
(文 4) 手、いったい。 （手、痛い）

さらに、終止形の語尾が「る」となる動詞に、「べ」、「けど」、「し」などが後接する場合には、「る」が促音化し、「～っべ」、「～っけど」、「～っし」のような形になる。

(文 5) してっべ。 （してるでしょう）
(文 6) 友達とかなら、わがっけど。
（友達とかなら、分かるけど）

(文 7) ナオミのごどは、すぐわがっし。
（ナオミのことは、すぐ分かるし）

２．２．村山方言の品詞と語尾

(1) 動詞の活用形

動詞は、古い形態を残しているために、共通語では五段活用になるものが、一般的には四段活用になる。「行く」（共通語では「いく」、村山方言では「いぐ」の活用形を比較対照すると、表１のようになる。

表１　「行く（行ぐ）」の活用形比較

	行く（共通語）	いぐ（方言）
基本形	いく	いぐ
未然形	いか	いが
未然ウ接続	いこ	—
未然ン接続	—	いが
連用形	いき	いぎ
連用タ接続	いっ	いっ
終止ベ接続	—	いぐ
仮定形	いけ	いげ
仮定縮約	いきゃ	—
命令形	いけ	いげ

表１に示すように、共通語では「カ行五段活用」である動詞「行く（いく）」が、村山方言では濁音化して、「ガ行四段活用」の動詞「行ぐ（いぐ）」になる。

(2) 形容詞の活用形

形容詞は、特に共通語と変わらないが、語尾に共通語にはない打消しの「ね」（共通語の「ない」に相当する）などが続く。

(3) 助詞

方言の特徴が最もよく現れるのが助詞である。ここでは、一部の特徴のみを簡単に述べる。

(a) 主格

村山方言では、共通語で主格を表す「は」、「が」が省略されることが多い。

(文 8) 桜さいだよ。　（桜が咲いたよ）

(b) 目的の「ば」

主格と同じく省略されることがあるが、強調したいときや、選択の結果を言う場合には省略しない。

(文 9) 本買いに行った。（本を買いに行った）
(文 10) 本ば買いに行ったのよー。
（本を買いに行ったのよー）

(c) 場所・方向の「さ」

共通語の場所、方向を表す「に」、「へ」は、村山方言では、いずれも「さ」で表される。

(文 11) 病院さ行ぐ。　（病院に行く）
(文 12) 北の方さ行ぐ。　（北の方へ行く）

(d) 受け手「がら」

受身表現の受け手は、共通語では「に」で導かれるが、村山方言では「がら」という助詞を用いる。

(文 13) 店長がらごしゃがっだっけよ。
（店長に怒られたよ）

(4) 文末表現

文末につくさまざまな表現（主として終助詞や助動詞）も、方言らしさを表現する重要な要素の一つである。一部の特徴的な表現について述べる。

(a) 否定表現

村山方言の否定形は、助動詞「ない」の音がなまって変化した「ね」を用いる。「ない」と同様に未然形に接続するが、ラ行四段動詞や、受身の「れる」、「られる」につくときは、「んね」という形になる。

(文 14) 行がね。　（行かない）
(文 15) 見らんね。（見られない）

(b) 希望表現

共通語「たい」の変化形である「だい」を用いる。３音節以下で、「る」で終わる動詞につく場合は促音化する。

(文 16) 見っだい。　（見たい）

(c) 意志・推量・勧誘表現

意志・推量・勧誘・同意を求める場合は、終助詞「べ」を用いる。

(文 17) 一緒に行ぐべ。　（一緒に行こう）

(d) 依頼表現

共通語の「てくれる？」に当るのは、「てける？」である。一段動詞と、カ行四段動詞の場合には、「でける？」と、「て」が濁音化する。共通語の「てくれない？」と同様のバリエーションも存在する。

(文 18) 明日お弁当作ってける？
（明日お弁当作ってくれる？）
(文 19) 明日お弁当作ってけねが？
（明日お弁当作ってくれない？）

３．実装システム[12]と評価

すでに、村山方言から共通語へのプロトタイプシステム[7]がある。また、ほぼ同様の手法を用いて作成した共通語から最上方言への翻訳システムプロトタイプ[13]も存在する。基本的には、これらのシステムで用いた手法を踏襲するが、いくつかの異なる点も存在する。以下ではシステムの内容について述べる。

３．１．翻訳手順

語順や構文の変形はほとんど必要ないので、基本的には形態素解析と語彙トランスファのみで翻訳が可能である。したがって、システムは次のようなステップで構成される。

【ステップ１】形態素解析

茶筌[14]をカスタマイズしたものを用いて共通語を形態素解析する。このカスタマイズは、名詞に簡単な意味素性を付加したことと、動詞の活用形に特殊なものを加えたものである。

【ステップ２】語彙トランスファ

辞書引きにより語彙変換する。意味素性まで参照して、一意の訳語を持つ語を変換している。活用語については、活用型定義辞書を参照して語尾変換を行う。

【ステップ３】例外処理

文末部分の変換や、句が単語になっているものの変換を行う。辞書を参照しても単純に変換できないものや、変換の規則性が見つからないものをここで処理する。

【ステップ４】形態素構造生成

変換された単語を連結し、文を生成する。

３．２．語彙・辞書

庄内方言についてはまとまった辞書が存在する[10]が、村山方言について単独にまとめられた辞書はない。そこで、やや古いが、山形県方言辞典[15]の中から、村山方言であると明記されている語（村山方言でも、ごく一部地域で用いられている語を除く）を抜粋した。方言としては、比較的若い世代が日常使用しているものを目指したため、この辞書の中の表現で使われないものも多い。そのような語彙については、作成者の判断で採用しなかった。

また、比較的多くの入力に対して柔軟に対応できるように、語彙変換の部分については、内部的には CSV 形式で構成されている。現在、名詞、助詞、助動詞、動詞、形容詞、形容動詞、連体詞、副詞、感動詞、接続詞を合わせて約 3000 語が入力されている。

表２　名詞変換辞書の一部

共通語 見出し語	共通語 読み	方言 見出し語	方言 読み
あいつ	アイツ	あいづ	アイヅ
山菜	サンサイ	あおもの	アオモノ
蛙	カエル	びっき	ビッキ

表３　動詞・形容詞変換辞書の一部

共通語 見出し語	共通語 読み	翻訳後 活用形	方言 語幹	方言 読み
食べる	タベル	ラ四段	あが	アガ
あやす	アヤス	サ四段	あえ	アエ
疲れる	ツカレル	形容ク	こわ	コワ

```
（四段・ガ行促音便
    （　；（語幹　＊　）
        （基本形　　ぐ　　グ）
        （未然形　　が　　ガ）
        （未然ン接続　がん　ガン）
        （連用形　　ぎ　　ギ）
        （連用タ接続　っ　　ッ）
        （終止ベ接続　ぐ　　グ）
        （仮定形　　げ　　ゲ）
        （命令形　　げ　　ゲ））
）
```

図１　活用型定義ファイルの一例

表２に名詞変換辞書の一部、表３に動詞・形容詞変換辞書の一部を示す。いずれも、共通語と方言とが対照される形式になっている。名詞の場合には、共通語彙を方言語彙に置き換える形式であるが、用言の場合には、動詞が形容詞に変わる例（表３の「疲れる」→「こわい」）がある。すでに述べたように、共通語の動詞から方言の別の動詞に変換される場合にも、多くの動詞が四段活用になっている。

図１には、表１で述べた特殊な活用形を扱う定義ファイルを示す。この図から分るように、「行く」→「行ぐ」に対応した活用形を扱えるようになっている。

また、茶筌をカスタマイズして意味素性を付加することによって、次のような同音異義語の判別が可能になっている。

(文 20) りんごを<u>切る</u>。→ りんごば<u>はやす</u>。
(文 21) 紙を<u>切る</u>。→ 紙ば<u>ひちゃばぐ</u>。

共通語の「切る」は、村山方言では、「（食

物）を切る」場合は「はやす」、「（物）を切る」場合には、「ひちゃばぐ」という動詞を用いる。助詞はいずれも「ば」を用いる。このように、意味素性によって動詞が変化する場合に対応するため、「りんご」には、従来の「名詞－一般」という品詞分類をカスタマイズして、「名詞－一般－食物」という意味素性を、「紙」には、「名詞－一般－道具」といった意味素性を加えている。それほど細かい区分は必要ないので、ここで扱っているのはごく簡単なものである。

３．３．実行例

＜共通語＞[16]
私達の向かっている“偉大なる航路”は、世界で最も危険な場所なのよ。その上ワンピースを求める強力な海賊たちがうごめいている。当然強力な船に乗ってね。船員の頭数にしても、この船の装備のなさにしても、とても無事でいられるとは思えないわ。

＜村山方言への自動翻訳結果＞
わだしらの向かっていだ“偉大なる航路”は、世界で最も危険な場所なんだず。そのうえかっつワンピースば求める強力な海賊らがうごめいていだ。当然強力な船さ乗ってな。船員の頭数さしても、この船の装備のなささしても、とても無事でいられるどは思えねは。

＜手作業による翻訳結果＞
わだしらの向かっていだ“すんばらしい航路”は、世界で最も危険な場所なんだず。そのうえかっつワンピースば求める強力な海賊らがうごめいていだ。当然強力な船さ乗ってな。船員の頭数さしても、この船の装備のなささしても、とっても無事でいられっどは思えねは。

図２　システムの実行例

システムの実行例を図２に示す。下線部が誤った箇所である。品詞の違い（「偉大なる」→「すんばらしい」）や促音化（「とても」→「とっても」）などがまだ完全には扱えていないが、おおむね良好な結果が出ている。

３．４．評価

現在のところ、村山方言ネイティブスピーカーによって、300 文を評価してもらい、そのうち 228 文（76%）については正しいとの評価を得ている。今後さらに文の数やジャンルを増やす予定である。

４．問題点と今後の検討

語数については、約 4000 語程度あればほぼカバーできそうなので、今後は語彙を増やしたい。また、共通語よりも細かい区別（たとえばニシンは、生のものを「かど」、干したものを「にしん」と言う）や、独特の文化によって共通語が存在しない語（雪や農業に関するものが多い）をどう扱うかも今後の検討課題である。

方言は、放送で行われる[17]など、広がりを見せている。今後音声合成と結びつけるなどの拡張に努めたい。

参考文献

[1] 藤原与一：方言学の方法、大修館書店(1977)
[2] 北原保雄監修、江端義夫編：方言、朝倉日本語講座 10、朝倉書店(2002)
[3] 佐々木冠他：方言の文法、シリーズ方言学２、岩波書店(2006)
[4] 髙木裕子：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、科研費研究成果報告書(2001)
[5] 横山晶一、安野克彦：方言の機械処理に関する予備的考察、電子情報通信学会技術報告 NLC95-45 (1995)
[6] 佐藤守、横山晶一、西原典孝：方言から共通語への翻訳システムに関する基礎的研究、情報処理学会第 150 回自然言語処理研究会 NLC2002-22 (2002)
[7] 佐藤守、横山晶一、西原典孝：山形方言から共通語への翻訳システム、言語処理学会第９回年次大会(2003) pp.401-404
[8] 森下喜一：標準語引東北地方方言辞典、桜楓社(1987)
[9] 平山輝男編：山形県のことば、日本のことばシリーズ６、明治書院(1997)
[10] 佐藤雪雄：庄内方言辞典、東京堂出版(1992)
[11] 国立国語研究所編：鶴岡方言の記述的研究―第３次鶴岡調査報告１―、国立国語研究所報告 109-1、秀英出版(1994)
[12] 尾形真美：共通語から村山方言への機械翻訳システム、山形大学卒業論文(2007)
[13] 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)
[14] 奈良先端科学技術大学院大学：茶筌
[15] 山形県方言研究会：山形県方言辞典(1970)
[16] 尾田 栄一郎：ONE PIECE、集英社(2006)
[17] 今夜はなまらナイト、NHK 山形第一放送(2007.1.24 放送)