

琉球大学学術リポジトリ

日本共通語と琉球方言との間の音韻対応分析エキスパートシステム

メタデータ	言語: 出版者: 琉球大学工学部 公開日: 2007-08-23 キーワード (Ja): キーワード (En): Expert system, Ryukyu dialect, Standard Japanese, Phonetic correspondence, Word transformation 作成者: 高良, 富夫, 久場, 長司, Takara, Tomio, Kuba, Choji メールアドレス: 所属:
URL	http://hdl.handle.net/20.500.12000/1461

日本共通語と琉球方言との間の音韻対応分析

エキスパートシステム

高良 富夫* 久場 長司**

An Expert System for Analyzing the Phonetic
Correspondence between Ryukyu Dialect and Standard Japanese

Tomio TAKARA* Choji KUBA**

According to linguistic research, the mainland-Japanese dialects and the Ryukyu dialects have the same origin even though they have very different forms now. One factor which supports the thesis that they have the same origin is that a systematic law of phonetic correspondence exists between them.

In order to investigate the phonetic law more objectively, we propose a word transformation system which synthesizes words of the Ryukyu-Shuri dialect from those of the Tokyo dialect (standard Japanese). In the system, the phonetic law is embodied in the IF-THEN rules of the OPS5. Using this system, we transformed nouns of the Tokyo dialect into words of the Ryukyu-Shuri dialect. It was found that 49% of the generated basic vocabularies agree with real nouns of the Ryukyu-Shuri dialect. Because the phonetic law can be prescribed in an explicit form in the system, we can study the phonetic law more objectively by this system than by the manual research of traditional dialectology.

Key words: Expert system, Ryukyu dialect, Standard Japanese, Phonetic correspondence, Word transformation

1. まえがき

日本語は方言学上, まず本土方言と琉球方言とに大分類される。琉球方言は現代では, 本土方言とかなり異なっており, 琉球方言に属するどの方言も本土方言のどの方言ともまったく通じないほどである。しかし,

言語学的研究によれば, 琉球方言と本土方言とは同一の祖語から発したと考えられている¹⁾。

両方言が祖語を同じくすると考えられるひとつの根拠は, 両方言間に存在する整然たる音韻の対応である。例えば, 東京方言の /ame/ (雨), /kumo/ (雲) は, それぞれ琉球首里方言では /ami/, /kumu/

* 工学部電子・情報工学科 Dept. of Information Eng., Fac. of Eng.

** 沖縄日本電気ソフトウェア㈱ Okinawa NEC Software Co.

本論文の内容は, 情報処理学会知識工学と人工知能研究会 (昭和63年7月) および電気関係学会九州支部連合大会 (平成元年10月) で口頭発表された。

であり、東京方言の母音/e,o/は、それぞれ琉球方言の/i,u/に対応する。このような対応法則はいくつかあり、これらの法則を適用すれば、琉球方言の多くの単語は本土方言の単語に対応づけることができるといわれている。

音韻法則に関しては、音韻学的によく分析されている¹⁴⁾。しかし、これまでの研究方法は、研究者の方言語彙にたよる手作業的方法によるものであり、音韻対応法則の有効性を大量の語彙を用いて検証しようとする場合など、思わぬ見落としや、法則(規則)の矛盾した適用などが生じる恐れがあった。

そこで我々は、この音韻法則の有効性をより客観的に検討することを目的として、音韻対応法則を用いて共通語の単語から琉球方言の単語を生成する単語変換システムを作成した。このシステムは、エキスパートシステム構築用言語であるOPS5を主として用いて構成された。音韻対応法則は、「もし本土方言の音韻がAである場合、琉球方言の音韻はBである」という形式で書けるので、OPS5のIF-THENルールの構文が有効に利用される。

琉球方言話者の多くは、音韻法則を無意識に適用していると思われる。すなわち、元来琉球方言の語彙に属さない単語は、音韻法則が適用され琉球方言らしく発音されることがある。従って、本システムは、方言話者を模倣する方言話者エキスパートシステムと考えることができる。もちろん本システムは、音韻学者の見いだした音韻法則を知識ベースとして持っていることから、方言学エキスパートシステムである。なお本システムは、現在のところ名詞だけを交換するものである。

本システムは図1に示すように5つのサブシステムから構成されている。規則作成サブシステムは、音韻法則を規則としてコンピュータに入力するためのシステムである。これは、現在のところパソコンのエディタで代用しており、OPS5のIF-THENルートを編集するだけである。音韻変換サブシステムは、本システムの本体であり、規則作成サブシステムで作成された音韻変換規則を用いて、東京方言の単語を琉球首里方言の単語に変換する。ここでは、エキスパートシステム構築用言語OPS5を用いて音韻法則をIF-THENルールで表現している。評価サブシステムは、音韻変換サブシステムによる変換結果とそこで用いた変換規則自体の評価をするシステムである。このサブシステムの

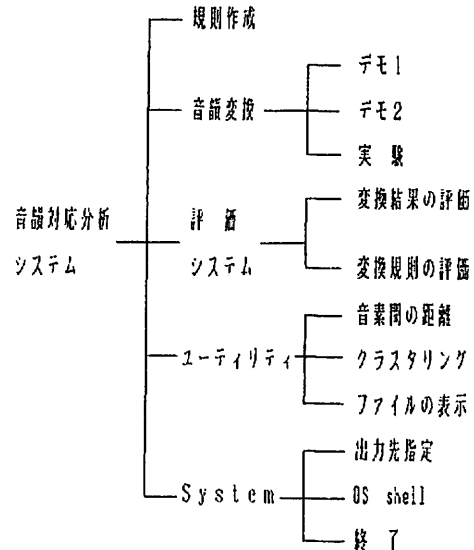


図1 音韻対応分析エキスパートシステムの構成

評価結果により、音韻法則の有効性および音韻変換規則の妥当性を検討することができる。ユーティリティは、音韻対応分析においてその補助として役にたつサブシステムをまとめたものである。Systemは、音韻分析システムに有用なOSに関するソフトウェアをまとめたものである。

以下では、各サブシステムについて説明した後、本システムを実行した結果を示し、その考察を述べる。

2. 音韻変換サブシステム

これまで我々は、Prolog-KABAによりOPS5を構成し、そのOPS5(KABA)を用いて東京方言から首里方言への音韻変換システムを構成した¹⁵⁾。しかし、Prolog-KABAでのメモリの制約のため、OPS5(KABA)には音韻変換の規則をすべてのせることができなかった。そこで今回は、OPS5をArity-Prologで構成した¹⁶⁾。また、OPS5(KABA)はPrologインタプリタ上で動作させていたが、今回、ArityコンパイラによりMS-DOS実行形式のOPS5システム(OPS5インタプリタ)を構成した。Arity-Prologを用いたことにより、大規模なOPS5プログラムが作れるようになり、また、OPS5の動作が高速化された。

今回構成した音韻変換サブシステムでは、「デモ1」

「デモ2」, 「実験」の3つの中から使用形態を選択できる。「デモ1」は従来の音韻変換システム¹⁰を使いやすくしたものであり, 画面に拡大文字を出力し, 共通語入力をひらがなで, 琉球方言の出力をカタカナで表示する。「デモ2」は, 従来の音韻変換システム¹⁰であり, OPS5のトレース機能を利用して, 単語変換に使用されたルールおよびその適用順序を確認することができる。「実験」は, 音韻変換実験を行うときに使用するシステムである。大量のデータを自動的に変換し, その結果を評価サブシステムで利用する形式でファイルに出力する。

2.1 琉球首里方言と東京方言の音韻対応

本研究では, 琉球方言の中から, かつて琉球王府のあった首里の方言を取り上げる。これは, 現代日本語における東京方言に相当し, 琉球方言の共通語と考えることができる。また, 本論文で単に共通語(または標準語)というときは概ね東京方言のことである。

首里方言と標準語との音韻対応に関しては文献(1)の解説編を参照した。文献(1)では, 首里方言で使用される音素のそれぞれについて, 標準語の音素との対応関係が述べられている。そこでは, 「〜に対応する」

「〜に対応するのが普通である」「しばしば対応する」「多く〜に対応する」と確定的な表現や, 「ただし〜」「対応する例がみられる」「〜ことがある」などの例外的な表現で述べられているものがある。ここでは, 確定的な記述だけを採用することとし, 対応法則を表1のように整理した。表1では, 対応するモーラの数で分類してあり, 「共通語の音韻Aは, 首里方言の音韻Bに」に対応するという形式で表現した。表1に現れていない音素は, 首里方言と東京方言とで一致すると考えてよい。

また, 表1を作成するに当たって, 文献(1)の表現をできるだけ忠実に表現することを基本としたが, プログラミングが容易となるように規則を整理した部分もある。表1から, 規則を組み合わせれば, より少ない規則に整理することができると思われるが, これについてはシステムの実現法と関連して今後の課題となっている。

後述するように, この音韻対応法則では説明できない単語も琉球首里方言に多数存在する。これらの単語は, 本土方言の単語とは語源を異にすると考えられるが, その由来が不明であるものも少なくない。

表1 標準語と琉球首里方言との音韻対応法則

1 モーラが2 モーラに	(14) 2 モーラ以上のai, aeはeeに
(1) 1 モーラの語尾のaはaaに	(15) 2 モーラ以上のaotはookに
" i はiiに	(16) 2 モーラ以上のawaはaaに
" u はuuに	(17) 2 モーラ以上のui, ueはiiに
" e はiiに	(18) 2 モーラ以上のhai, haeはhweeに
" o はuuに	(19) 2 モーラ以上のkuraはqkwaに
(2) 1 モーラのhi, heはhwiに	(20) 2 モーラ以上のkureはqkwiに
(3) 1 モーラのkiはciに	(21) 2 モーラ以上のguraはngwaに
(4) 1 モーラのgiはziに	(22) 2 モーラ以上のgureはngwiに
(5) 1 モーラのcuはciに	2 モーラが1 モーラに
zuはziに	(23) 2 モーラ以上の語頭の
suはsiに	mi?a, mu?aはn?aに
1 モーラが1 モーラに	mi?e, mu?eはn?eに
(6) 3 モーラ以上の語尾のmiはnに	mi?o, mu?oはn?oに
(7) 2 モーラ以上のhi, heはhwiに	(? は子音)
(8) 2 モーラ以上のkiはciに	子音が子音に
(9) 2 モーラ以上のgiはziに	(24) 2 モーラ以上の語頭のrはdに
(10) 2 モーラ以上のcuはciに	(25) 2 モーラ以上の語頭以外のrjはjに
zuはziに	(26) 2 モーラ以上のkj, cjはcに
suはsiに	(27) 2 モーラ以上のgj, zjはzに
2 モーラが2 モーラに	(28) 2 モーラ以上のsjはsに
(11) 2 モーラ以上の語尾のmonoはmuntに	母音が母音に
(12) 2 モーラ以上のariはaiに	(29) 2 モーラ以上のeはiに
(13) 2 モーラ以上のuri, oriはuiに	(30) 2 モーラ以上のoはuに

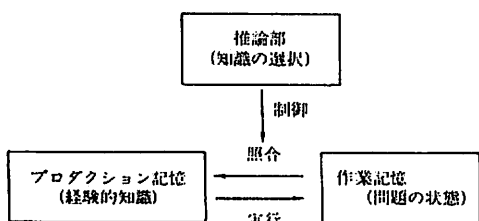


図2 エキスパートシステム構築用言語OPS5の構造

2.2 OPS5TM

本サブシステムは、プロダクションシステムをベースにしたエキスパートシステム構築用言語であるOPS5を用いて構成した。OPS5は、プロダクション記憶、作業記憶、推論部の3要素から構成されている

(図2)。プロダクション記憶は、IF-THEN型ルールの集合で記述された知識ベースである。作業記憶は、対象の状況がルールのIF(条件)部と同様の形式で表現された要素の集まりである。推論部は、実行可能なルールを、ある戦略(LEX戦略)TMに基づいて選択、実行する機構である。

[作業記憶]

作業記憶は、作業記憶要素とタイム・ダグの対で構成されている。作業記憶要素のフォーマットは以下のようにになっている。

```
(クラス名 ^属性1 値1 ^属性2 値2
      ... ^属性n 値n)
```

タイム・ダグは、その作業記憶要素がいつ作られたか、または修正されたかを表し、値が大きほどその作業記憶要素は新しい。

[プロダクション記憶]

プロダクション記憶内のひとつのIF-THEN型ルールは、図3のようなフォーマットで表される。このルールの条件部と作業記憶要素が適合すると、動作部が実行可能になる。動作部を実行することによって、作業記憶の内容が変更されると、次のルールが適用される状態になる。動作部の動作命令としては、以下のようなものなどがある。

```
make ... 新しい要素を作業記憶に加える。
modify ... 存在する作業記憶要素の値を変更する。
remove ... 作業記憶から要素を削除する。
```

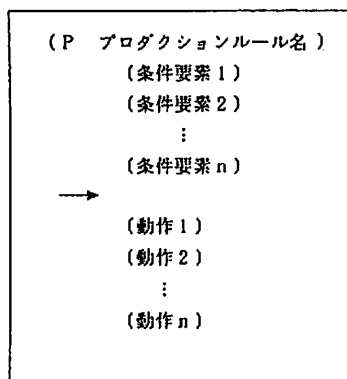


図3 プロダクションルールのフォーマット

[推論部]

推論部は、大きく分けて以下の4つのステップの動作を行う。

ステップ1: 照合

プロダクション記憶内にあるすべてのルールの条件部と作業記憶の現内容を照らし合わせ、条件部が満足されたルールを選び出す。この選び出されたルールのひとつひとつをインスタンシェーション、これらの集合を競合集合という。

ステップ2: 競合解消

照合時に選び出された競合集合の中から指定された戦略に基づいて、ひとつのルールを選択する。

ステップ3: 動作

競合解消によって選ばれたルールの動作部に示された動作を実行する。

ステップ4: 繰り返し

ステップ1に戻る。

本システムで用いたOPS5TMは、Prolog上に実現されたものであり、IF-THENルールの動作部は、Prologの述語により実行することもできる。また作業記憶要素の属性の値がベクトル量である場合、これらをリストで表すことができる。

OPS5では、図2のように、推論部とプロダクション記憶(OPS5のプログラミング部分)とが独立しているので、Cなどの手続き型言語に比べて知識が浮き彫りにされており、知識の追加や変更が行いやすくなっている。従って、OPS5を用いると知識ベース型プログラミングの開発がより容易に行える。

2.3 音韻変換サブシステムの構成⁴⁾

本サブシステムは、共通語と琉球方言の間に存在する音韻法則を、エキスパートシステム構築用言語であるOPS5上にIF-THENルールとして作成したものであり、これにより共通語の単語を琉球方言の単語へ変換する。今回構成したシステムでは、変換されるモーラ数に着目して、変換規則を表1のように分類し、それぞれをIF-THENルールで表現した。

共通語を琉球方言に変換するためにOPS5のIF-THENルールで必要な情報は、以下のものである。

入力段階における単語の①「文字集合」（ひとつひとつの文字の集まり、[リスト]）、②「文字数」（未変換の文字数；繰り返しの条件に使用される）、③「モーラ数」（ルールはその単語のモーラ数で異なる）、現在の変換タスクの対象である文字の④「位置」、④の位置における入力情報である⑤「1文字目から最大5文字目まで」。

これらの情報は、文字列の入力の際または変換（変換は規則がないときの無変換も含む）の際、認識され、IF-THENルールの条件部との照合、ルールの動作部の実行のために使用される。

本サブシステムは、「入力段階」、「音韻変換段階」、「出力段階」の3段階からなる。すなわち、本サブシステムのすべてのIF-THENルールは、この3段階のいずれかに分類される。以下にそれぞれの段階について簡単に説明する。

〔入力段階〕

使用者がキーボードから入力した共通語のローマ字の文字列を“文字集合”に変換し、その“文字数”、“モーラ数”、注目する文字の“位置”、“1文字目から5文字目までの文字”を抽出し、次の段階に受け渡す。ただし、入力文字列が“end”である場合は終了する。

〔音韻変換段階〕

この段階は、本サブシステムの本体であり、表1の6種のIF-THENルール群より成る。IF-THENルールの条件部では、入力段階から受け取った“文字集合”、“文字数”、“モーラ数”および文字の“位置”、“1文字目から5文字目までの文字”を認識する。

IF-THENルールの動作部では、まずルールと一致した条件部の注目している文字を変換する。次に“文字数”から動作を行った文字の数を減じ、次の位置の5文字を読み込む。6種のどのIF-THENルールも適

用されない場合は、単に、1文字だけシフトする（“文字数”から1だけ減じ、次の5文字を読み込む）。

音韻変換の終了条件は、“文字数”がゼロ（0）になった場合である。この時、変換した文字集合を出力段階へ受け渡す。

〔出力段階〕

変換された“文字集合”を文字列としてディスプレイに表示する。その後、作業記憶に残っている情報を消去し“位置”を1とし、タスクを入力段階へ受け渡し、上記の動作を繰り返す。

なお、入力段階、出力段階ともOPS5ではIF-THENルールで表現される。

3. 評価サブシステム

本サブシステムは、音韻変換結果の評価と、音韻変換に用いる規則の評価を行う2つのシステムから構成されている。

「変換結果の評価システム」は、これまで手作業で行っていた音韻変換実験結果の評価⁴⁾を自動化し、判定基準を明確にするために構成した。このシステムでは、音韻変換システムで東京方言を変換した結果と実際の首里方言との音素一致率、単語一致率を求める。また、基礎語彙の単語の一致率に言語年代学⁵⁾を適用して、東京方言と首里方言とが何年前に分岐したかを推定することができる。音素の一致の判定には音素の弁別的素性を用いた距離を利用するDPマッチング⁶⁾を用いている。

「規則の評価システム」は、音韻変換に用いた規則、すなわち音韻法則には、どのような対応の特徴があるかを、音素の弁別的素性を用いてベクトル量で表現し評価する。

3.1 弁別的素性

言語学は大きくわけて、音韻論と構文論とにわけられる。音韻論には、音声の基本単位は何かという立場から、音素という概念があり、母音（/a,i,u,e,o/）、子音（/p,b,m,.../）などのように区別されている。音素はさらにその発声という立場からみて、弁別的素性（distinctive feature）⁷⁾といういくつかの要因と結びつけられている。

弁別的素性とは、表2のようなものである。この表は本システムで実際に用いたものであり、Chomsky

表2 音素とその弁別的素性の表

		a	i	u	e	o	j	w	r	p	b	m	l	d	n	s	z	c	k	g	h	q				
vocalic	(母音性)	+	+	+	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	0			
consonantal	(子音性)	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	0		
high	(高性)	-	+	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	0	
back	(後性)	+	-	+	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	0	
low	(低性)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	0	
anterior	(前方性)	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	0	
coronal	(舌頂性)	-	-	-	-	-	-	-	+	-	-	-	+	+	+	+	+	+	+	+	+	-	-	-	0	
round	(円唇性)	-	-	+	-	+	-	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tense	(緊張性)	+	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
voice	(有声性)	0	0	0	0	0	0	0	+	-	+	+	-	+	+	-	+	-	+	-	+	-	+	-	0	
continuant	(持続音性)	0	0	0	0	0	0	0	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	+	0	
nasal	(鼻音性)	0	0	0	0	0	0	0	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	0	
strident	(粗擦音性)	0	0	0	0	0	0	0	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	0	

& Halleの弁別的素性を参考にしている。なお、Chomsky & Halleの表では、この表の0のところには、なにも書かれていないが、プログラミングの簡単化のために0を記入した。またqはChomsky & Halleの表にはないが本システムで使用するため、上の表に示した素性を持つものとして付け加えた。

本システムでは、弁別的素性の各素性に次のような値を対応させ、音素を弁別的素性を成分とするベクトルとみなして、弁別的素性を利用している。

記号が+のときの値	: 1
記号が-のときの値	: -1
記号が0のときの値	: 0

[変換結果評価システムにおける弁別的素性の利用]

変換結果の評価システムでは、弁別的素性を用いて音素間に次のような距離の概念を導入している。ある2つの音素について、音素ベクトルの各成分の差を求め、その各成分の「絶対値」の和を音素間の距離とする。

例えば、

音素 a (1, -1, -1, 1, 1, -1, -1, -1, 1, 0, 0, 0, 0) と
音素 i (1, -1, 1, -1, -1, -1, -1, -1, -1, 0, 0, 0, 0) と
の距離は次のようになる。

$$0+0+2+2+2+0+0+0+2+0+0+0+0=8$$

変換結果の評価システムでは、この距離を用いてDPマッチングでの一致基準を緩めることができるようになってい

[規則評価システムにおける弁別的素性の利用]

規則の評価システムでの弁別的素性の利用法を、音韻変換規則(17)、すなわち

2モーラ以上のui,ueはiiに

を例に用いて説明する。

音素列uiのベクトルを、音素ベクトルuと音素ベクトルiの和を「音素の数」2で割ったものとして定義する。

$$\begin{aligned} & u(1, -1, 1, 1, -1, -1, -1, 1, -1, 0, 0, 0, 0) \\ & +) i(1, -1, 1, -1, -1, -1, -1, -1, -1, 0, 0, 0, 0) \\ & =) (2, -2, 2, 0, -2, -2, -2, 0, -2, 0, 0, 0, 0) \\ & \therefore ui(1, -1, 1, 0, -1, -1, -1, 0, -1, 0, 0, 0, 0) \end{aligned}$$

音素列ue,音素列iiのベクトルも同様にして求める。

$$\begin{aligned} & ue(1, -1, 0, 0, -1, -1, -1, 0, -1, 0, 0, 0, 0) \\ & ii(1, -1, 1, -1, -1, -1, -1, -1, -1, 0, 0, 0, 0) \end{aligned}$$

音素列uiを音素列iiに変換する規則を表現するベクトルを、(ベクトルii) - (ベクトルui)で定義する。

$$\begin{aligned} & ii(1, -1, 1, -1, -1, -1, -1, -1, -1, 0, 0, 0, 0) \\ & -) ui(1, -1, 1, 0, -1, -1, -1, 0, -1, 0, 0, 0, 0) \\ & ui \rightarrow ii(0, 0, 0, -1, 0, 0, 0, -1, 0, 0, 0, 0, 0) \end{aligned}$$

同様にして、

$$ue \rightarrow ii(0, 0, 1, -1, 0, 0, 0, -1, 0, 0, 0, 0, 0)$$

となる。

このようにして各規則を表現するベクトルを求め、それらのベクトルの平均を音韻変換規則全体を表現するベクトルとする。このベクトルは、音韻変換規則の

特徴を弁別的素性により表現するものであるといえる。

3.2 DPマッチング

文字列のパターンマッチングにおけるDPマッチング法¹⁰⁾について簡単に説明する。文字列 $s_1s_2 \dots s_m$ と $t_1t_2 \dots t_n$ との照合を行うものとしよう。文字 s_i と t_j との近さ(距離)という概念を導入し、これを $d(s_i, t_j)$ とする。 $s_1s_2 \dots s_i$ と $t_1t_2 \dots t_j$ とが最もよくマッチしたという時のずれの量 $f(i, j)$ を導入し、これを次のような漸化式によって計算する。

$$f(0, 0) = 0$$

$$f(i, j) = \min \{ f(i-1, j) + 1, f(i, j-1) + 1, f(i-1, j-1) + d(s_i, t_j) \}$$

$$d(s_i, t_j) = \begin{cases} 0 & (s_i = t_j \text{ の場合}) \\ 1 & (s_i \neq t_j \text{ の場合}) \end{cases}$$

以上は一般的な方法であるが、変換結果の評価システムで用いている方法は次のようになっている。

上記の関数 d を文字 s_i と t_j との一致度とみなし、 $s_1s_2 \dots s_i$ と $t_1t_2 \dots t_j$ とが最もよくマッチしたという時の一致した文字数を関数 f とし、次のような漸化式によって計算する。

$$f(0, 0) = 0$$

$$f(i, j) = \max \{ f(i-1, j), f(i, j-1), f(i-1, j-1) + d(s_i, t_j) \}$$

$$d(s_i, t_j) = \begin{cases} 1 & (\text{dis}(s_i, t_j) \leq \theta \text{ の場合}) \\ 0 & (\text{dis}(s_i, t_j) > \theta \text{ の場合}) \end{cases}$$

ここで関数 dis は音素間の弁別的素性を用いた距離である。 θ は音素一致の基準を定める値であり、この値を可変値にして一致の基準をゆるめることができるようにしている。すなわち、 $\theta = 0$ のときは完全一致であり、 θ を大きくすると、弁別的素性のパターン(ベクトル)が似ている音素間では一致とみなされるようになる。

3.3 言語年代学

言語年代学¹¹⁾は、おもに、アメリカの言語学者M. Swadeshによってとえられたものである。言語年代学では次のように考える。

ある1つの言語の語彙は時代が変わるとともに変化していく。しかし、どの言語にも必ず存在する基本的語彙(基礎語彙)の変化は比較的安定であり、その変

化は1000年を単位にとってみても非常に定常的なものである。同一の祖語から分裂した2つの言語の場合、共通な基礎語彙がどれだけ残っているのかを調べることによって、2つの言語が分裂したのが何千年前であるかを推定することができる。と考える。

「変換結果評価システム」では、この言語年代学により、求めた単語一致率から東京方言と首里方言とが何年前に分岐したかを推定することができる。ここでは、「基礎語彙の単語一致率」を言語年代学における「共通な基礎語彙の残存率」とみなしている。ここで意味する「一致」は言語年代学での「共通」とは異なり、それは、音韻変換すなわち音韻法則を含めた「一致」のことである。従って、「一致率」は、ゆるい意味での残存率となっている。

分岐年代を求める式としては、Swadeshの考案した次式を用いる。

$$t = \log p / \{ 2 \log(0.805 \pm 0.0176) \}$$

p: 基礎語彙の残存率

4. ユーティリティ, System

ユーティリティは、音韻対応分析においてその補助として役に立つサブシステムをまとめたものであり、「音素間の距離、一致度」、「クラスタリング」、「ファイルの表示」の3つの下位システムから構成されている。

「音素間の距離、一致度」では、音素間の距離(3.1参照)を音素の全組合せについて求める。また、音素間の距離がいくらまでを音素の一致と見なすかによって、音素の一致、不一致を一覧表にして示す。「クラスタリング」は、音素間の距離が θ 以下の音素同士をまとめて、いくつかのグループを作る。ここでは、 θ は可変とし、キー入力できるようになっている。グループの音素同士は似た音素であるといえる。クラスタリングには階層クラスタリングの最大法¹²⁾を用いた。「ファイルの表示」は、MS-DOSテキストファイルを表示する。ここでは、音韻対応表や弁別的素性の表などを表示することができる。

「System」は、音韻対応分析システムに有用なOSに関連するソフトウェアをまとめたものであり、「出力先設定」、「OS Shell」、「終了」からなる。

「出力先設定」は、評価サブシステム、ユーティリティにおける出力先を画面、プリンター、ファイルのい

れかに設定するものである。「OS Shell」は、音韻対応分析システムから一時的にMS-DOSに抜けるためのものである。音韻対応分析システムに戻るときには「exit」とタイプする。「終了」は、音韻対応分析システムを終了する。

5. システムの実行結果

本システムを用いて、表1に示した音韻対応法則を共通語（東京方言）の単語に適用して琉球首里方言の単語を生成した結果について述べる。

入力単語としては以下の2種について検討した。

【ランダム抽出単語】

沖縄語辞典¹⁰⁾の標準語索引の奇数ページの右上から1個ずつ抜きだした100個の名詞。対応する首里方言としては、その索引に示されている首里方言のうち音韻的に標準語に近いものを選んだ。

【基礎語彙】

文献(9)に示されている東京方言と首里方言の基礎語彙対応表の200語のうち名詞102語。基礎語彙とは、日常よく使われる基本的な語彙である。

変換結果の評価システムの実行結果を表3および表4に示す。

表3は、DPマッチングによる音素の一致判定で、距離0だけを一致と見なしたものである（ $\theta=0$ ：完全一致）。表4は、DPマッチングによる音素の一致判定で、距離2以下を一致と見なしたものである（ $\theta=2$ ）。表4の結果は、一致基準を少し緩めたものであり、完全一致の場合より一致率が高くなっている。

表3 変換結果の評価システムの実行結果
（ $\theta=0$ ：完全一致）

	ランダム抽出単語	基礎語彙
音素一致率（％）	60	71
単語一致率（％）	26	49
東京方言と首里方言との分岐年代		
1489年～1823年前（1962年基準） 「2世紀なかごろから5世紀おわり」		

表4 変換結果の評価システムの実行結果（ $\theta=2$ ）

	ランダム抽出単語	基礎語彙
音素一致率（％）	68	76
単語一致率（％）	29	50
東京方言と首里方言との分岐年代		
1448～1772年前（1962年基準） 「2世紀おわりから6世紀はじめ」		

表3の音素一致率から、音韻対応法則の有効性は、ランダム抽出単語で60％、基礎語彙で71％であることが分かる。また表3および表4を比較すると、一致の基準を緩めた場合、音素一致率は高くなるが、単語一致率は、特に基礎語彙では、さほど変化しないことが分かる。

基礎語彙の一致率がランダム抽出単語の一致率より高いことから、「基礎語彙の方が他の語彙より時間に対する抵抗が強く、それほど変化せずに残る傾向が強い」という言語学的知見¹¹⁾が、共通語と首里方言の間でも成立することが確認できる。このことは、すなわち共通語と首里方言とが祖語をおなじくすることの傍証となる。

表5 規則の評価システムの実行結果

順位	弁別 的 素 性		傾 向
- 4	vocalic	(母音性)	-0.22
	consonantal	(子音性)	0.09
	high	(高 性)	0.00
- 3	back	(後 性)	-0.26
	low	(低 性)	-0.08
+ 1	anterior	(前方性)	0.35
+ 2	conronal	(舌頂性)	0.31
	round	(円唇性)	-0.00
	tense	(緊張性)	-0.02
	voice	(有声性)	-0.00
	continuant	(持続音性)	-0.08
	nasal	(鼻音性)	0.06
+ 5	strident	(粗擦音性)	0.19

東京方言と首里方言の分岐年代の推定結果は、従来の言語学的な定説⁹⁾と良く一致している。しかし、単語の一致判定基準および用いた公式が両者では異なるので、これについては今後さらに検討する必要がある。

次に、規則の評価システムの実行結果を表5に示す。表5において傾向とは、前述の変換規則を表現するベクトルであり、各弁別的素性に対応する成分を示してある。また、「傾向」の絶対値が0.10以上の素性には大きな順に順位をつけ、その「傾向」の符号をつけてある。

この表より、音韻変換に用いた規則には次のように音を変化させる特徴があることが分かる。特徴の顕著な順にならべると、

- 音を
1. 前方性にする、
 2. 舌頂性にする、
 3. 後性でないようにする、
 4. 母音性でないようにする、
 5. 粗擦音性にする。

もし、音韻変換規則が現実の法則をうまく表現しているならば、以上の特徴は現実の音韻法則にも存在することになり、それは東京方言から首里方言への音の変化の傾向であるといえる。しかし、上記の方法では、全ての規則を同じ重みで評価しているので、現実の音韻法則全体としての傾向を表現しているとは必ずしも言えない。規則は、多くの単語に当てはまることもあるが、少数の単語に当てはまることもあるからである。従って、音韻法則の特徴をよりよく表現するためには、規則の評価方法をさらに検討する必要がある。

6. 検 討

音韻対応分析システムは、方言研究の知的支援システムとして、特に、音韻法則の有効性をより客観的に検討できるようにすることを目的として提案された。コンピュータ上に本システムを構成したことにより、音韻法則の有効性をより客観的に検討できるようになった。方言研究の知的支援システムとして利用できるようにするため、我々がこれまで作成してきたいくつかのシステムを統合した。統合化したことにより、システム間でのデータの受渡しが容易になり、より容易に分析が行えるようになった。また知的支援システムには不可欠であるユーザインターフェースの充実を計り、システムをより使いやすいものとした。

しかし本システムはまだ構成されたばかりであり、実際の場面においては使用されていないので、方言学者などが実際に使用した上で、ユーザインターフェースについてさらに検討を行う必要がある。また、変換結果および変換規則自体をより客観的に評価するため、評価サブシステムでは以下のことを検討する必要がある。

変換結果の評価システムでは、単語の一致判定方法にDPマッチングを用いている。しかし、DPマッチングを用いた場合と他のパターンマッチング法を用いた場合との比較はまだ行っていない。判定方法についての検討を行う必要がある。また、DPマッチングでの判定基準を弁別的素性を用いて規定できるようになっており、今回は $\theta = 0$ 、 $\theta = 2$ としたが、これらの値は十分に検討された値ではない。この基準値の検討も今後行う必要がある。

言語年代学では、基礎語彙における共通の単語の残存率を単語の一致率とすべきであるが、本システムでは、変換した単語と実際の単語の一致率をそれとしている。また、変換した単語と実際の単語の一致の判定法は、上述したように十分検討されていない。従って、本方法で東京方言と首里方言との分岐年代を高い確度で推定するには、さらに検討が必要である。

規則の評価システムは、弁別的素性を用いて音韻法則の特徴を求めることを目的としている。そして、今回用いた方法によって、音韻法則の大まかな特徴を表現することができる。しかし、この方法によって音韻法則の特徴が十分表現できたとは言いがたい。その理由は、音韻変換規則が音韻法則を十分に表現しているか否かについての検討が十分されていないためである。しかし、音韻変換規則が音韻法則を十分に表現しているかどうかを検討するためにも、本システムは利用できる。

本方法では、音素の弁別的素性を用いているが、音素ではなく、音声（記号）の弁別的素性を用いたほうがより良く音の変化の傾向を表現できるのではないかと思われる。しかし、音素対音声の関係は一对多の関係にあり、音素を音声に置き換えるにはいくとおりかの置き換えがあって、音声を用いるのは容易でない。従って音声を用いるには、音素から音声への置き換えを全ての場合について行わなければならない。

7. むすび

琉球方言と本土方言との間に存在する音韻対応法則をOPS5のIF-THENルールとして記述し、共通語の単語から首里方言の単語を生成する単語変換システムを構成した。またこのシステムで用いた音韻対応法則を東京方言の名詞に適用し、生成した単語を実際の首里方言の単語と比較した結果、音韻法則の有効性は基礎語彙で71%であることが分かった。また基礎語彙の単語一致率は49%であった。

本システムでは、OPS5のIF-THENルールの構文を利用することにより、音韻法則(規則)を明確に規定することができ、これにより音韻法則の有効性をより客観的に検討することができる。また、本システムのルールを整理・実行する過程を通して、法則の整理・簡略化を実験的に検討することができる。

今後の課題としては、音韻対応の例外的規則も適用できるようにすること、および規則間の競合関係を説明する機能を付加すること、さらにルールをより自然言語に近い形式で挿入、削除できるようにし、一般の言語学者にも容易に利用できるシステムにすることがあげられる。

謝辞：この研究の一部は文部省科学研究費奨励研究(A)の助成による。

文 献

- (1) 国立国語研究所編：“沖縄語辞典”，大蔵省印刷局(昭58-04)。
- (2) 新垣 薫：“OPS5を用いる音韻変換エキスパート・システム”，琉球大学工学部卒業論文(昭63-03)。
- (3) 久場長司：“琉球方言の音韻対応分析システム”，琉球大学工学部卒業論文(平元-03)。
- (4) 高良，新垣，志喜屋：“音韻法則を用いる共通語から琉球方言への単語変換システム”，情報処理学会知識工学と人工知能研究会資料，59-10(1988-07)。
- (5) 鈴木宣夫：“OPS5文法入門”，Computer Today(サイエンス社)，No13，pp.11-19(昭61-05)。
- (6) 小林，小沢：“OPS5 on Prolog”，Computer Today(サイエンス社)，No13，pp.54-60(昭61-05)。
- (7) 長尾 真：“言語工学”，昭晃堂(昭62-04)。
- (8) 長尾 真：“パターン情報処理”，コロナ社(昭58-10)。
- (9) 安本美典：“日本語の成立”，講談社(昭57-01)。
- (10) 服部四郎：“「言語年代学」すなわち「語彙統計学」の方法について”，言語研究，第26/27号，pp.29-77(昭29-12)。