

徳之島方言の音節頻度表から何が分かるか

沢 木 幹 栄

要 約 『徳之島方言二千文辞典』のテキストから音節を切り出し、それぞれの頻度を産出した。その結果に基づき考察を行った。ほとんど自明に近いことが多いが、実際の使用に基づいて頻度を調べたことに意味がある。

徳之島浅間方言の音節ごとの実使用頻度の統計資料を基に考察を行った。

音節の使用頻度の統計はありそうでないものである。音韻論では「機能負担量」という術語があるが、それは使用頻度に根拠を置いているものではなく、きわめて概念的で具体性を欠いていると言われても仕方のないものである。

では、実際に音節ごとの実使用頻度がどうなっているかは、日本語については、各種のコーパスが公開されるようになってやっと研究が行われるようになったが、音韻論との関係での重要性にはあまり着目されていないようである。

我々（沢木、中島、福嶋、岡村）は『日本語二千文』の翻訳（実際の翻訳は徳之島浅間方言の話者である岡村が行い、中島をはじめとする他の3人は編集を行った）により、徳之島浅間方言の二千文のテキストを確定し、それを材料に研究を進めてきた。なお、『日本語二千文』は川本茂雄がアンリ・フレの“Le livre de deux milles phrase”を日本語に翻訳したものである。

2009年には『徳之島方言二千文辞典』（以下、『二千文辞典』と呼ぶ）として、浅間方言（以下、徳之島浅間方言を浅間方言と呼ぶ）テキストと標準語テキストを対比させたものを主体にしたものを出版したが、ここで特に革新的であったのは付属のDVDで、その内容は、文節によるKWICなどをブラウザで見ることができるハイパーテキストであった。方言テキストのKWICは全く新しい試みだったが、そのほかの内容のなかには、『二千文辞典』のテキストを音節に分解し、それぞれの頻度を測定して表にしたものがある。この表については、今まで紹介も解説も行われていないが、方言のテキストをもとに音節の実使用頻度を計算したのは、『二千文辞典』のDVDが日本では初めてだったと考えられる。

『二千文辞典』のテキストの分量は決して多いとは言えないが、それを音節に分解し数え上げることを人手で行うのはかなり困難だと言える。こういうことはコンピューターで行うのが早いし、ミスも防げるのでそのようにした。具体的な手順を述べるまえに、浅間方言をどのような表記で記録したかについて述べたい。

柴田1960は琉球方言の音韻論的研究としては初期に属するものであるが、今でも価値を失っていない。ここで示された音韻分析はそのまま、現在でも使えるものである。なお、柴田1960のインフォーマントは『徳之島方言二千文辞典』の翻訳者と同じ岡村隆博氏である。『二千文辞典』の浅間方言はもちろん柴田1960に基づいた記録を行った。ただし、柴田は特

殊記号を使ったりしたので、その方式そのままではコンピューターに簡便に入力できない。そこで大文字と小文字を用いた簡略な方法で表記を行った。

浅間方言の音素として柴田1960では子音 k, t, m, n, c について喉頭化していない系列と喉頭化しているものがあるとしている。また、母音には a, e, i, o, u のほかに i と u の中舌化したものがあり、母音も喉頭化することがあるとしている。

ここでは、喉頭化した子音は K, T, M, N, C のように大文字で表した。喉頭化した (glottal stop に先立たれる) 母音は 'i のように' をつけて表した。j, w についても同様で 'j と' w, j と w で喉頭化の有無を表している。

中舌母音は I, E のように大文字で表した。

モーラ音素（にあたるもの。柴田によれば徳之島浅間方言はシラビーム方言であり、モーラ音素という言葉を使うのは適切ではないが、ほかの方言であればモーラ音素に該当するものという意味でここでモーラ音素という言葉を使う）のうち、母音の引き音は柴田は R で表記しているが、我々は：を用いた。二重母音の後半部分は柴田が J を使っているのに対して、我々は i を用い、撥音は柴田も我々も N を用いている。促音は我々は特にそのための記号を作らず、tta のように同じ子音を重ねることによって表記している。

N は喉頭化した子音と、モーラ音素と両方の意味を持つことになるが、どちらであるかは位置で分かる。

このように、『二千文辞典』の表記は厳密な音韻表記ではないが、柴田式の表記にまぎれがない形で変換可能なものとなっている。

つぎに、音節を切り出すアルゴリズムについて説明する。

『二千文辞典』の表記を音節に分解するにあたって、音節と音節の境界をまず見つけることにした。境界と境界のあいだの文字列が音節になるはずである。

具体的には以下の手順に従った。

1. 母音に子音が続く場合、子音がモーラ音素でなければ、母音と子音の間が境界となる。
2. 母音に N が続いたとき、N の後が境界となる。
3. 母音に：あるいは母音が続いたときは、境界はまだ先にある。：あるいは 2 番目の母音を母音として 1. 2. 3. を適用し、母音の境界を定める。
4. 以上の手順で境界を定めたあと、音節の頭に tt や cc のような同一子音の繰り返しがあ
る場合は、前の子音を先行する音節の末尾に移動させ、Q に変える。促音は徳之島方言
ではかならず音節末にあるからである。

以上の手順をコンピューターのプログラムとして書き、実行して『二千文辞典』のテキストすべてを音節に分解することができた。このようにして切り出した音節は全部で 21841 個だった。

しかるのちに、音節の異なり別に頻度を調べ、表にして最後に示した（徳之島方言音節頻度表）。

それぞれの音節の使用頻度がわかると、実際にどのような文節のなかで使われたかを知りたくなる。たとえば、英語では th の子音を使う語は少ないが、実際の使用は決して他の子

音に劣るということはない。それは、th が the あるいは they, them のような語で使われるからである。だから、実際の使用の状況を知るとはとても重要である。

音節がどの文の文節にあったかを知るために、音節の KWIC とも言うべきものをプログラミングで作ってみた。これを「音節の KWIC」と呼ぶことにする。これは非常に分量が多いので、一部だけをここに示す。

'wIN

1494 'wIN

'wa:

1190 'wa:nu 1253 'wa:ckija 1254 'wa:ckki 1255 'wa:ckija 1256 'wa:ckija:

1265 'wa:ckija 169 'wa:slja 186 'wa:sl 189 'wa:slja 1919 'wa:ckija:

1947 'wa:ckisIdE: 220 'wa:slbEN 33 'wa:slbatu 56 'wa:nu'i:ja:tIN

823 'wa:naisjuNda:

'waN

1542 'waNhu:nI 190 'waNhu:nIja

CjE:

464 CjE:

音節を見出しに、その音節が現れた文の番号と、文節を記録している。

音節頻度表を見ると bjo : iN のように徳之島方言としては異例（柴田1960の音節構造に当てはまらない）な音節があるが、それは『二千文辞典』のどこにあるのだろうか。徳之島方言本来の語では、母音のあとに母音が続く場合には二番目の母音のまえに ' がつき、母音のあとに直接母音が続くことはない。調べてみたところ bjo : iN は「病院」に対応し、kue は「机」にあたる文節で出現していた。徳之島方言として異例な音節はすべて浅間方言本来のものでない語のなかにあったことが確認できる。このような音節は表には残すが、分析の対象から除外することにする。

それでは、音節頻度表からほかにどんなことが分かるか、「音節の KWIC」を参照しながら考察してみたい。

c, C の破擦音では喉頭化音の C は Cj のように j を後続させた形でしか存在しない。柴田1960では CINGjo（深い井戸）のような語例があるので、これはたまたま二千文でこの音節を含む語を使う場面がなかったのだと考えられる。

ただし、今までの研究で存在が確認されていない音節はこの調査でも出現しないと考えられる。その点も考慮に入れて音節頻度表を見る必要がある。

浅間方言で「人」は Cju であり、『二千文辞典』で「人」に関係する表現が多出することから、Cju の頻度は高い。ところで、Cju が文節の内部にあるときは、'azjamaccju（浅間の人間）のように ccju の形で表れる。この場合は cju は付属語的な扱いである。

そこで、cju がどこに表れたかを「音節の KWIC」で調べてみると「人」に対応するのは5例である。同じ「人」に対応する Cju と cju では、前者の音節が135例あり、そのほとんどが「人」を表すのに対し、後者は、5例のみということになる。

このように文節の内部で喉頭化音が非喉頭化音として発音される現象は Kw, kw を含む音節でも起きている。「与える」は KwIjuN と喉頭化音で始まるが、「読んでやった」の「やった」のように補助動詞的に使われた場合は喉頭化音ではない。「音節の KWIC」で kw を含む音節を調べると、「与える」のいろいろな活用形が多く使われているのがわかる。

このように喉頭化音は文節の頭で出現する。喉頭化音で始まる語であっても、それが文節の内部で使われたときは非喉頭化音で始まることになる。柴田1960では naCI : (夏) のような語例があるが、『二千文辞典』のテキストのように個々の単語を意識しない、流れの中での発話では文節の内部では喉頭化音は発音しにくいようである。

したがって、喉頭化音を含む音節の頻度はそれに対応する非喉頭化音の音節のそれにくらべて少ないことが多い。

助詞 ga (主格, 所有格を表す格助詞, あるいは疑問詞疑問の終助詞), nu (主格, 所有格を表す格助詞), du (とりたての助詞), ba (Nba とも, 「も」にあたる助詞), ja (本土方言の「は」にあたる助詞) は非常に多く使われるので, 音節としても頻度が高くなることが予想される。実際, 音節の ga, nu, du, ba は100回以上, ja に至っては963回も使用されており, 助詞として使われた比率が高いことは「音節の KWIC」から確認できる。ある音節がこのように使用頻度が高い語のなかにあったとき, たとえその音節を持つ語の異なりが少ししなくても, 音節としては使用頻度が高くなるのが分かる。

このほかにも使用頻度の高い音節として mI, tI があるが, これらは動詞の活用形の一部に nI:mI (無いか), 'ikada:tI (行かなかった) のような形で表れ, しかもこれらの活用形がよく使われるものであるために音節としても頻度が高くなるのである。

p を持つ音節の頻度はどれも高くない。本土方言の h は浅間では h で対応し, 浅間の p に対応する音は本土方言にはないからである。語頭の音素 p はむしろ, 本土方言本来の語を浅間方言で使ったときに現れるものである。

ti, di は音素の組み合わせとして理論上は可能であるが, 頻度はゼロである。本土方言の「チ」は浅間方言では cI または CI が対応し, 「テ」「デ」はそれぞれ tI, dI が対応するので, ti, di は本土方言に対応する音節を持たないのである。浅間方言は音節 ti, di を最初から持たない可能性がある。実際に柴田1960でも存在しない組み合わせとされている。

以上をまとめると, 次のようになる。

1. 浅間方言に存在するのにたまたま『二千文辞典』のなかに使用例がないために頻度ゼロになっているものがある。
2. 喉頭化音の子音は対応する非喉頭化音に比べて頻度が低い。
3. 本土方言に対応する音節がないものは, 浅間方言に音節として存在しない可能性がある。また, あっても頻度が低い。
4. よく使われる語や活用形の一部に現れる音節は頻度が高い

以上のように浅間方言の音節の頻度について考察した。

いままでの音韻論では, 質的な考察しかできなかったが, 量的な考察を試みたのが最大の

成果だと考える。『二千文辞典』のテキストという決して量が多くはない資料を対象にしたが、確定されたテキストがあれば、同じようなことが他の方言でも可能である。コンピューターを利用することによってこのようなことができることを示したのが、もう一つの成果だと言えよう。

参考文献

- Frei, Henri (1953) *Le livre des deux milles phrases* Genève
川本茂雄 (1971) 『日本語二千文』早稲田大学語学教育研究所
服部四郎 (1955) 音韻論『国語学』22
岡本隆博, 沢木幹栄, 中島由実, 福嶋秩子, 菊池聡 (2006) 『徳之島方言二千文辞典』 自家版
岡本隆博, 沢木幹栄, 中島由実, 福嶋秩子, 菊池聡 (2009) 『徳之島方言二千文辞典 改訂版』 自家版
柴田武 (1960) 「徳之島方言の音韻」『国語学』41

德之島方言音節頻度表

'a 291	bo: 38	cjaN 45	do: 62	gjoN 1	hiN 7
'a: 53	boi 2	cjaQ 3	doN 13		hiQ 1
'aN 198	boN 4			gju: 9	
'aQ 54		cjE 1	du 135	gjuN 2	hja 3
	bu 31	cjE: 6	du: 51		hja: 3
'e 1	bu: 16		dui 16	go 23	
'e: 1	bui 8	cjI 230	duN 9	go: 20	hjo: 3
	buN 47	cjI: 51	duQ 1	goN 30	
ba 343	buQ 1	cjIN 22		goQ 1	ho 1
ba: 37			ga 437		ho: 21
bai 6	Cja: 8	cjo 2	ga: 104	gu 73	hoN 19
baN 26	CjaN 2	cjo: 27	gai 8	gu: 25	hoQ 1
baQ 1	CjaQ 1	cjo:i 1	gaN 4	gui 6	
		cjoQ 1	gaQ 9	guN 34	hu 88
be 4	CjE: 1				hu: 53
be: 1		cju 42	ge 1	gwa 17	hui 1
beN 8	CjI: 15	cju: 29	ge: 5	gwa: 102	huN 15
	CjIN 1	cjui 16	gei 1	gwai 3	huQ 4
bE 2	CjIQ 2	cjuN 37	geN 1	gwaN 19	
bE: 63				gwaQ 1	hwE: 10
bEN 19	Cjo: 5	da 148	gE 12		hwEi 1
bE:N 1		da: 652	gE: 60	gwI 1	hwEQ 11
bEQ 1	Cju: 135	dai 19	gEN 17	gwI: 2	
	Cjui 6	daN 85			hwI 6
bi 29	CjuN 34	daQ 2	gi 44	ha 203	hwI: 5
bi: 9			gi: 22	ha: 21	hwIN 4
biN 8	cji 1	de 2	giN 24	hai 108	
biQ 2		deN 12	giQ 1	haN 92	'i 364
	cI 460			haQ 13	'i: 32
bI 18	cI: 82	dE 9	gI 19		'iN 12
bI: 12	cI:i 2	dE: 6	gI: 6	he 2	'iQ 97
bIN 1	cIiN 1	dEN 2	gIN 1	hei 4	
	cIN 49			heN 1	ji 11
bja 1	cIo: 2	dI 23	gja 3		ji: 28
bjo: 4	cIQ 12	dI: 9	gja: 3	hE: 4	jiN 28
bjoQ 1		dIN 9		hE:i 1	jiQ 1
bjo:iN 3	cja 21	dIQ 1	gjo 12		
	cja: 51		gjo: 1	hi 11	'ja 2
bo 3	cjai 4	do 15	gjo: 1	hi: 3	'ja: 19

'jaN 6	KioN 1	ki: 51		mI: 70	ne 1
'jaQ 3	KiQ 8	kIN 5	kwE: 1	mIN 3	neN 13
		kiQ 2	kwEQ 1	mIQ 6	
'jE: 32	Kja: 1				nE 15
	Kju 1	kja 49	kwI 1	mja: 3	nE: 38
'jI: 8		kja: 36	kwI: 31		nEQ 13
	Ku 55	kjai 1	kwIN 2	mju: 1	
'ju 1	Ku: 15	kjaN 3		mju: 2	ni 48
'ju: 9	Kui 8		Ma: 20	mjuN 11	ni: 17
'juN 19	KuN 4	kjo 6	MaQ 1		niN 21
		kjo: 14		mo 40	niQ 1
ja 963	Kwa: 16	kjoN 1	Me: 1	mo: 31	
ja: 372	KwaN 4			moi 15	nI 50
jai 2	KwaQ 1	kju 13	MI: 1	moN 8	nI: 63
jaN 25		kju: 30			nIN 159
jaQ 24	KwE: 3	kjui 9	Mja: 1	mu 71	nIQ 2
		kjuN 35		mu: 44	nIu 1
je: 4	KwI: 7	kjuQ 2	Mo: 13	mui 3	
jei 23			Moi 2	muN 188	nja 2
jeN 7	ka 491	ko 33		muQ 22	nja: 21
	ka: 172	ko: 89	ma 181		njaQ 3
jE 2	kai 18	ko:eN 1	ma: 91	'N: 10	
jE: 11	kaN 90	koi 9	mai 33		nje: 4
	kaQ 10	koN 10	maN 39	N 28	
jI 18		koQ 5	maQ 27	N: 20	njE: 2
jI: 6	ke 12				
	ke: 21	ku 418	me 2	Na: 94	njo: 32
jo 5	kei 12	ku: 63	me: 19	Nai 1	
jo: 101	keN 10	ku:e 6	mei 6	NaN 87	nju: 5
joi 6	keQ 1	kueN 5	meN 2	NaQ 1	njuN 1
joN 4		kuQ 3			
	kE 8		mE 10	NeN 3	no 6
ju 182	kE: 33	kui 15	mE: 39		no: 17
ju: 88	kEi 14	kuN 183	mEN 3	NE: 3	no:N 1
jui 94		kuoN 1			
juN 118	ki 131	kuQ 20	mi 84	na 281	nu 559
juQ 4	ki: 57		mi: 56	na: 168	nu: 73
	kiN 45	kwa 6	miN 10	nai 11	nui 4
Ki 122	kiQ 6	kwa: 12	miQ 2	NaN 251	nuN 29
Ki: 26		kwai 14		naQ 3	nuQ 23
KiN 8	ki 53	kwaN 7	mI 136		

'o 50	rE 3	sIai 1	TaN 1	tui 73	zI:o 1
'o: 53	rE: 56	sIN 43		tuN 76	zIQ 12
'oi 4	rEN 2	sIQ 20	TI: 22	tuQ 5	zI:Q 2
'oN 4			TIN 1		
	ri 15	sja 135		'u 187	zja 85
wo 3	ri: 6	sja: 107	To: 1	'u: 30	zja: 27
wo: 6	riN 1	sjai 15		'ui 12	zjai 8
	riQ 1	sjaN 111	Tu: 4	'uN 49	zjaN 8
pa 3		sjaq 19		'uQ 74	zjaQ 1
pa: 4	rI 110	sja:Q 1	ta 228		
pai 19	rI: 163		ta: 171	wu 57	zje 1
paN 6	rIN 2	sje 8	tai 29	wu: 10	zjei 2
paQ 1	rIQ 6	sje: 19	tai: 1	wui 10	zjeQ 1
		sjei 12	taN 116	wuN 11	
pe: 2	rjo 1	sjeN 14	taQ 18	wuQ 2	zjI 33
peN 2	rjo: 2	sjeQ 1			zjI: 57
			te 28	'wa: 15	zjIN 5
pi 4	rju: 1	sjE: 8	te: 2	'waN 2	
pi: 1			tei 3		zjo 6
piN 1	ro 70	sjI 144	teN 8	'wEN 2	zjo: 31
	ro: 65	sjI: 135	teQ 3	'wEQ 1	
pja 2	roi 8	sjIN 13			zju 10
	roN 2	sjIQ 4	tE 12	'wI: 25	zju: 35
po 3	roQ 3		tE: 37	'wIN 1	zjuN 4
po: 6		sjo 5	tEN 2		zjuQ 1
poN 1	ru 108	sjo: 44		wa 161	
	ru: 51	sjoN 3	ti:e: 1	wa: 70	zo 1
pu 9	rui 5			wai 53	
puN 3	ruN 11	sju 43	tI 644	waN 39	
	ruQ 4	sju: 53	tI: 75	waQ 43	
ra 181		sjui 13	tIN 71		
ra: 94	sa 4	sjuN 96	tIQ 4	wE: 3	
rai 3	sai 1	sjuQ 7			
raN 240	saN 2		to 37	zjiN 1	
raQ 13	saQ 1	so 2	to: 78		
		so: 1	toi 1	ziN 1	
re 80	se: 2		toN 6		
re: 99	seN 1	su: 1	toQ 1	zI 155	
reN 96				zI: 61	
reQ 4	sI 559	Ta: 14	tu 276	zIe: 1	
	sI: 78	Tai 8	tu: 138	zIN 26	