

Making paradigms of verbs and adjectives using a dialect corpus

CHITSUKO FUKUSHIMA

*University of Niigata Prefecture, 471 Ebigase, Higashi-ku, Niigata,
Niigata 950-8680 Japan, chitsuko@unii.ac.jp*

SCN III/1 [2010], 124–131

Razprava predstavlja izdelavo narečnega slovarja Tokunoshime (otočje Amami, Japonska), ki nastaja na podlagi narečnega korpusa. Analiza korpusnega gradiva in pogovorov z informanti omogoča oblikovanje glagolskih in pridevniških paradigem, ki bodo uvrščene v multimedijški narečni slovar. Stavki v korpusu so bili razdeljeni v sklope, na podlagi katerih so bili ugotovljeni glagoli, ti pa so bili razvrščeni v sezname – šlo je za ugotavljanje spregatvenih vzorcev. Vse spregane oblike so bile pregledane glede na oblike, ki jim sledijo, nato pa je bila na podlagi distribucije spregana oblika izbrana za geslo. V japonščini pripadajo glagoli in pridevniki isti sintaktični kategoriji (pridevniki spreminjajo svoje oblike tako kot glagoli). Enak postopek je bil ponovljen tudi pri pridevnikih – poiskani so bili vzorci in paradigme pridevniškega pregibanja.

The author has been involved in the making of a dialect dictionary of Tokunoshima, Amami, Japan, using a dialect corpus. The analysis of the dialect corpus and face-to-face interviews were combined to obtain the paradigms of verbs and adjectives to be included in the multimedia dialect dictionary. Sentences in the corpus were cut into phrases and verbs were identified and sorted into lists of verbs. The lists were examined to find patterns of verb conjugation. All conjugated forms were examined regarding succeeding forms, and, based on the distribution, a conjugated form was chosen as an entry. In Japanese, verbs and adjectives belong to the same syntactic category and adjectives change their forms as verbs do. Thus the same procedure was repeated concerning adjectives, and patterns and paradigms of adjective inflection were found.

Ključne besede: narečni korpusi, morfologija, paradigme, glagoli, pridevniki

Key words: dialect corpus, morphology, paradigms, verbs, adjectives

0 Introduction: Making a dialect dictionary using a dialect corpus

We have been making a multimedia dialect dictionary of Tokunoshima which is located in Amami, Japan. We reported the progress of our research at the fourth and fifth International Society for Dialectology and Geolinguistics congress (Sawaki, Fukushima, and Nakajima 2003, 2006, Sawaki, Nakajima, and Fukushima 2006). This is the third presentation about the research and herein the author describes her own part of the research concerning morphological analysis of the dialect, which is necessary for inclusion of the paradigms of verbs and adjectives into the dictionary. In order to obtain the paradigms, the analysis of the dialect corpus and face-to-face interviews were combined.

1 Data & Method

1.1 Dialect corpus as basic data

We used a dialect corpus as basic data to make a dictionary. The corpus is called *The Two Thousand Sentences of the Tokunoshima Dialect*. This is a free translation based on the Japanese version of *Le livre des deux milles phrases* by Henri Frei. Our co-researcher Takahiro Okamura, a native speaking dialectologist made the text data of two thousand sentences expressing everyday life in Tokunoshima. They were made into a dialect corpus.

1.2 Tokunoshima Dialect

The Japanese language is classified into two major dialect groups. One is the Mainland dialects and the other is the Ryuku dialects. The Tokunoshima dialect belongs to the Amami dialect, a sub-group of the Ryukyu dialects. The dialect has fewer and fewer speakers, so we need to make a dialect dictionary to keep a record of the dialect and help maintain it.

The following symbols are used in the transcript. When compiling the dialect corpus, we used only letters we could input using the keyboard.

1. Central Vowels: Capitalized ex. \bar{i} > I, \bar{e} > E
2. Glottalized Consonants: Capitalized ex. k' > K, t' > T
3. Glottal Stop: ʔ
4. Syllabic Nasal: N

1.3 Method

Japanese is an agglutinative language, so a word in Japanese is a linear sequence of distinct morphemes, each of which has lexical or grammatical meaning.

Independent forms such as nouns, verbs, adjectives, or adverbs are followed by dependent forms such as particles or suffixes (auxiliaries). Here is an example.

<i>watasi-wa</i>	<i>ik-anakat-ta.</i>
Pronoun+Particle	Verb+Aux1+Aux2
“I” + topic	“go” + negative + past
“I didn’t go.”	

Because of this characteristic, verbs and adjectives change their forms according to succeeding forms. In Indo-European languages, the inflection of adjectives is called declension as that of nouns. In Japanese, however, verbs and adjectives belong to the same syntactic category and adjectives change their forms as verbs do. For example, adjective stems can be followed by suffixes meaning “past”. Thus the author does not call the inflection of Japanese adjectives declension in this paper.

In order to obtain paradigms, sentences in the dialect corpus are cut into phrases, and they are sorted into alphabetical order. Then the list of independent forms plus dependent forms in a row is obtained. In the case of nouns, we can obtain a list of nouns with succeeding particles. In the case of verbs, we can get a list of verb stems with succeeding suffixes. Here are examples of a verb meaning “eat”. The verb has verb stems such as *kam-* the basic verb stem and *kad-* the euphonic verb stem. They are followed by auxiliaries meaning “negative” or “past” etc.

“eat”		
<i>kam-</i>		basic verb stem
<i>kad-</i>		euphonic verb stem
<i>kamaN</i>	“don’t eat”	basic verb stem + negative
<i>kamada:tl</i>	“didn’t eat”	basic verb stem + negative + past
<i>ka:dI</i>	“ate”	euphonic verb stem + past

2 Three steps to obtain the paradigms

2.1 Obtaining patterns of verb conjugation

There are three steps to obtaining paradigms. On the first step, basic patterns of verb conjugation were obtained as follows. First, sentences in the corpus were cut into phrases. Verbs were identified and sorted into a list of verbs. Interviews were conducted to make lists of basic conjugated forms of all verbs used in the corpus. The number of the verbs obtained was close to 500. The lists were examined to find patterns of verb conjugation. Verbs were classified based on the patterns.

As a result, three patterns of verb conjugation were obtained.

- Pattern I: Regular Conjugation I (Consonant-Ending Verb Stem type)
'asIb-: Standard Japanese *asobu* “play”
'asIbjui, *'asIbi:*, *'asIbaN*, *'asI:dI*
 (Basic, Preceding predicate, Negative, Past, respectively)
 [made up of verb stems: *'asIbj-/asIb-/asId-*]
- Pattern II: Regular Conjugation II (Vowel-Ending Verb Stem type)
'wI:-: Standard Japanese *okiru* “get up”
'wI:jui, *'wI:*, *'wI:raN*, *'wI:tI*
 [made up of verb stems: *'wI:j-/wI:-/'wI:t-*]
- Pattern III: Irregular Conjugation
s-: Standard Japanese *suru* “do”
sjui, *sI:*, *sjaN*, *sjI:*
 [made up of verb stems: *sj-/s-/sj-*]

2.2 Obtaining more detailed paradigms of verb conjugation

The second step is to make more detailed paradigms of verb conjugation. In order to decide entry forms for the dictionary, all conjugated forms of the verbs were examined regarding the succeeding forms. There are two basic forms as possible entry forms: *-jui* and *-juN* (Ex. *'asIbjui*, *'asIbjuiN*, “play”). Another basic form, *-ju:/ju*, was added through the research. Based on the distribution, a conjugated form, specifically the *-jui* form, was chosen as the entry form. The *-jui* form can be used without succeeding forms while both the *-juN* form and the *-ju:/ju* form are always used with succeeding forms.

Verb basic forms

- jui* *jamjui* “hurt”
 Can be used without succeeding forms.
 Or are followed by forms such as *ja:*, *jo:*, *sjarE:*, etc.
- juN* *sjuNda:* “do”
 Always followed by other forms such as *da:*, *cjI*, *do:*, *ga*, *du(ka)*, *kja*,
gadaN, etc.
- ju:/ju* *cIkju:mI* “[Do you] touch?”
 Always followed by other forms such as *mI*, *sI*, *wa:*, etc.

In this way, more detailed paradigms of verb conjugation were obtained.

2.3 Obtaining patterns and paradigms of adjective inflection

The third step is to obtain patterns and paradigms of adjective inflection. The same procedure was repeated concerning adjectives, and patterns and paradigms of adjective inflection were found.

The following illustrates three patterns of adjective inflection.

- Pattern I: Regular Inflection I
naga:ha- Standard Japanese *nagai* “long”
naga:hai, naga:ku, naga:hatI
 (Basic, Preceding predicate, Past, respectively)
 [made up of adjective stems: *naga:ha-/ naga:k-/ naga:hat-*]
- Pattern II: Regular Inflection II
wassja- Standard Japanese *warui* “bad”
wassjai, wassjaku, wassjatI
 [made up of adjective stems: *wassja-/ wassjak-/ wassjat-*]
- Pattern III: Irregular Inflection
ni- Standard Japanese *nai* “not exist”
niN, ni:, ni:da:tI
 [made up of adjective stems: *ni-/ ni-/ ni:da:t-*]

Adjectives also have similar suffixes parallel to verb forms. There are *hai-*, *haN-*, and *ha:/ha-* forms (Ex. *naga:hai / naga:haN / naga:ha* “long”). The distribution is almost the same as with the verbs, so the *-hai* form was similarly chosen as the entry form. The *-hai* form can be used without succeeding forms. But the *-haN* form can be used without succeeding forms when the sentence has the causative meaning while the *-ha:/ha* form is always used with succeeding forms. Thus the distinction of usage applies to not only verbs but also other categories.

Adjective basic forms

- hai* *Ma:hai* “delicious”
 Can be used without succeeding forms.
 Or are followed by forms such as *ja:, jo:, sjarE:, etc.*
- haN* *'iba:haN* “ [because it is] narrow”
 Can be used without succeeding forms when the sentence has the causative meaning.
 Or are followed by forms such as *da:, cJI, do:, ga, du(ka), kja, gadaN, etc.*
- ha:/ha* *juta:hamI* “[Is it] good?”
 Always followed by any other forms such as *mI, sI, wa:, etc.*

3 Conclusion

As a conclusion, inflectional patterns and paradigms of verbs and adjectives of the Tokunoshima dialect were obtained using a dialect corpus. They were included in the multimedia DVD dictionary of the Tokunoshima dialect (T. Okamura et al. 2009).

The following tables are part of the results of the analysis. Figure 1 is the list of four basic conjugated forms of all the verbs used in the *Two Thousand Sentences of the Tokunoshima Dialect*. Figure 2 is a detailed paradigm of main verbs.

表4 全動詞の基本活用後の例 索引五十音順

活用分類	索引	意味	本方言言対応形	辞書は(1)の活用形	辞書は(1)の活用形		辞書は(1)の活用形			マークをクリックすると音声が再生されます
					見出し語	基本辞幹	連用形	否定形	て形	
I	あう	会う、合う	あう	ワ行五段	【o:ju】	o:r-	【o】	【o:raN】	【o:ji】	<
I	あがる	上がる	あがる	ワ行五段	【aga:ju】	aga:r-	【aga】	【aga:raN】	【aga:ji】	
I	あがるくなる	明るくなる	(あがる)	ワ行五段	【ha:ga:ju】	ha:ga:r-	【ha:ga】	【ha:ga:raN】	【ha:ga:ji】	
II	あきらめる	諦める	あきらめる	下一段	【akira:mi:ju】	akira:mi-	【akira:mi】	【akira:mi:raN】	【akira:mi:ji】	
I	あく	明るくなる開く	あく	カ行五段	【aku:ju】	ak-	【aki】	【aka:raN】	【aci:ji】	<
II	あける	開ける	あける	下一段	【he:ju】	he-	【he】	【he:raN】	【he:ji】	
I	あずかる	預かる	あずかる	ワ行五段	【acika:ju】	acika:r-	【acika】	【acika:raN】	【acika:ji】	
I	あそぶ	遊ぶ	あそぶ	バ行五段	【aso:bu:ju】	aso:bu-	【aso:bi】	【aso:bu:raN】	【aso:bu:ji】	
I	あたたまる	温まる	ぬくまる	ワ行五段	【mukama:ju】	mukama:r-	【mukama】	【mukama:raN】	【mukama:ji】	
II	あたたまる1	暖める	ぬくめる	下一段	【mukama:mi:ju】	mukama:mi-	【mukama:mi】	【mukama:mi:raN】	【mukama:mi:ji】	
II	あたたまる2	暖める	ぬくめる	下一段	【mukumi:ju】	mukumi-	【mukumi】	【mukumi:raN】	【mukumi:ji】	
I	あたる	当たる	あたる	ワ行五段	【ata:ju】	ata:r-	【ata】	【ata:raN】	【ata:ji】	
I	あつまる	集まる	あつまる	ワ行五段	【acima:ju】	acima:r-	【acima】	【acima:raN】	【acima:ji】	
II	あつめる	集める	あつめる	下一段	【acimi:ju】	acimi-	【acimi】	【acimi:raN】	【acimi:ji】	
II	あてる	当てる	あてる	下一段	【ate:ju】	ate-	【ate】	【ate:raN】	【ate:ji】	
II	あびる	あびる	あびる	上一段	【ami:ju】	ami-	【ami】	【ami:raN】	【ami:ji】	
I	あまる	余る	あまる	ワ行五段	【ama:ju】	ama:r-	【ama】	【ama:raN】	【ama:ji】	
I	あむ	噛む	あむ	マ行五段	【amu:ju】	amu-	【amu】	【amu:raN】	【amu:ji】	
I	あらう	洗う	あらう	ワ行五段	【arou:ju】	arou:r-	【arou】	【arou:raN】	【arou:ji】	<
I	あらす	染らす	あらす	ワ行五段	【aras:ju】	aras-	【aras】	【aras:raN】	【aras:ji】	
III	ある	ある	ある	ワ行五段	【a:ju】	a-	【a】	【a:raN】	【a:ji】	<

Figure 1. List of four basic conjugated forms of all the verbs in the *Two Thousand Sentences of the Tokunoshima Dialect*

マーカーを一回クリックすると音声再生されます

表 2A 徳之島浅間方言における動詞活用形の生成規則動詞 (1) 子音語幹

			I 規則 -i 食べる	I 規則 -d 遊ぶ	I 規則 -s 行く
			kamju:	'asɪbjui	'ikju:
基本語幹			kam-	asɪb-	'ik-
融合語幹			kamj-	'asɪbj-	'ikj-
音便語幹			kad-	'asɪd-	'izj-
融合語幹	活用語尾	意味 [接続]	kamj-	'asɪbj-	'ikj-
ju:	-ui	～る [言い切り] ja: (わ) ja: (よ) sjɛɛ: (だろう)	kamju:[j]	'[asɪbjui]	'[ikju:]
ju:	-uN-	～る [+da: (よ) eɪ (って) do: (ぞ) ɛa (か) sjɛ: (よ) bɛ: (らしい)]	kamju[Nda:]	'[asɪbjuNda:]	'[ikjuNda:]
連体形	-uN	名詞が後接	kamju[N]	'[asɪbjuN]	'[ikjuN]
禁止形 1	-uNna	～るな	kamjuN[na]	'[asɪbjuNna]	'[ikjuNna]
ju:	-u:	～る [+mɪ (の) aɪ (の、もの、こと) wɔ: (だわ)]	kamju:[mɪ]	'[asɪbju:mɪ]	'[ikju:mɪ]
ju:	-u	～る [+mɛ: (まい) bi:ki (べき) gɛ:ɛɪ (ように) baka (～したて)]	kamjumɛ:[i]	'[asɪbjumɛ:]	'[ikjumɛ:]
ju:	-u	mɛɛɔ (どうして) や da (ぞ) に対する結び	kamju:[ɔjka]	'[asɪbju:ɔjka]	'[ikju:ɔjka]
係結形	-u:ru		kamju:[ru]	'[asɪbju:ru]	'[ikju:ru]
基本語幹			kam-	asɪb-	'ik-
連用形	-i	～る [言い切り、+助詞、動詞/補助動詞]	kamj[:]	'[asɪb:]	'[iki:]
禁止形 2	-iNna	～るな	kamjN[na]	'[asɪbNna]	'[ikNna]
否定形	-aN	～ない	kama[N]	'[asɪbaN]	'[ikaN]
否定過去形 1	-ada:ɸi	～なかった [言い切り、+助詞]	kamada:ɸi]	'[asɪbada:ɸi]	'[ikada:ɸi]
否定過去形 2	-adataN	～なかった [+助詞]	kamadaɸiN]	'[asɪbadataN]	'[ikadataN]
使役形	-asju:*	～させる	kamasju:[j]*	'[asɪbasju:]*	'[ikasju:]*
受身・可能形	-ajui:*	～られる	kamaju:[j]*	'[asɪbasju:]* (可能)	'[ikaju:]*
意向形	-a:	～よう	kama[:]	'[asɪba:]	'[ika:]
直接命令形	-o:	～ろ	kamo[:]	'[asɪbo:]	'[iko:]
間接命令形	-i:	～ろ (と)	ka:[mɪ]	'[asɪ:ɸi]	'[ike:]
丁寧形	-e:jui:*	～ます	kame:[ju:]*	'[asɪbe:jui:]*	'[ikeju:]*
音便語幹			kad-	'asɪd-	'izj-
テ形	-ɪ	～て	ka:[dɪ]	'[asɪ:dɪ]	'[izjɪ:]
過去形 1	-ɪ	～た [言い切り、+助詞]	ka:[dɪ]	'[asɪ:dɪ]	'[izjɪ:]
過去形 2	-aN	～た [+助詞]	kada[N]	'[asɪdaN]	'[izjaN]
タリ形	-ai	～たり	kada:[j]	'[asɪdaj]	'[izjai]
条件形	-ika	～れば	ka:[dɪ]ka	'[asɪ:dɪ]ka	'[izjɪ:]ka
テイル形	-u:*	～ている	ka:[du:]*	'[asɪdu:]*	'[izju:]*

Figure 2. Detailed paradigm of main verbs in the Tokunoshima dialect

REFERENCES

Takahiro OKAMURA, Motoei SAWAKI, Yumi NAKAJIMA, Chitsuko FUKUSHIMA and Satoshi KIKUCHI, 2009: *The Dictionary of Two Thousand Sentences of the Tokunoshima Dialect, Revised Version*. Matsumoto: Association of Tokunoshima Dialect.

Motoei SAWAKI, Chitsuko FUKUSHIMA and Yumi NAKAJIMA, 2003: "Dialect Corpus as a Resource for Dialect Dictionary" A paper presented at 4th International Congress of Dialectologists and Geolinguists, Riga, Latvia.

Motoei SAWAKI, Chitsuko FUKUSHIMA and Yumi NAKAJIMA, 2006: "Dialect Corpus as a Resource for Dialect Dictionary" In: *Proceedings of 4th International Congress of Dialectologists and Geolinguists, Riga*. Riga: Latvian Language Institute, University of Latvia. 431–438.

Motoei SAWAKI, Yuki NAKAJIMA and Chitsuko FUKUSHIMA, 2006: "Making Multimedia Dialect Dictionary as a Database with Indexes and Cross-references" At 5th International Congress of Dialectologists and Geolinguists, Minho, Portugal.

OBLIKOVANJE GLAGOLSKIH IN PRIDEVNIŠKIH PARADIGEM OB UPORABI NAREČNEGA KORPUSA

Narečni slovar Tokunoshime (otočje Amami, Japonska) je nastajal s pomočjo narečnega korpusa. Ta je bil, nadgrajen s pogovori z informanti, vir za morfološki oris narečja Tokunoshime. Paradigme glagolov in pridevnikov, ki so bili uvrščeni v multimedijski narečni slovar, so bile določene glede na: (1) Stavki iz korpusnega gradiva so bili razdeljeni v sklope; glagoli so bili identificirani in razvrščeni v sezname; opravljeni so bili pogovori, da bi lahko ugotovili seznam osnovnih pregibnih oblik vseh uporabljenih glagolov; sezname so bili pregledani in ugotovljeni so bili vzorci glagolske pregibnosti; na osnovi teh vzorcev so bili glagoli razvrščeni. (2) Pred odločitvijo o vključitvi so bile pregledane vse pregibne glagolske oblike glede na oblike, ki jim sledijo; na osnovi porazdelitve je bila pregibna oblika izbrana za geslo. (3) Enak postopek je bil uporabljen tudi pri pridevniki; poiskani so bili vzorci pridevniškega pregibanja (gl. M. Sawaki, C. Fukushima in Y. Nakajima. 2003. »Dialect Corpus as a Resource for Dialect Dictionary« At 4th ICDG. M. Sawaki, Y. Nakajima, and C. Fukushima. 2006. »Making Multimedia Dialect Dictionary as a Database with Indexes and Cross-references« At 5th ICDG).