

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 multimedijski narečni slovar. Stavki v korpusu so bili razdeljeni v sklope, na podlagi katerih so bili ugotovljeni glagoli, ti pa so bili razvrščeni v seznane – 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 Ryuky 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. $\text{ĩ} > \text{I}$, $\text{ě} > \text{E}$
2. Glottalized Consonants: Capitalized ex. $\text{k}' > \text{K}$, $\text{t}' > \text{T}$
3. Glottal Stop: '
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> Pronoun+Particle “I” + topic	<i>ik-anakat-ta.</i> Verb+Aux1+Aux2 “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:tI</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:ju:, *'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: *-ju:* 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 *-ju:* form, was chosen as the entry form. The *-ju:* 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

- ju:* *jamju:* “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 全動詞の基本活用形の例 動詞五十個例

活用形	動詞	意味	基本形(原形)	活用形	活用形		活用形			
					基本形	活用形	活用形	活用形	活用形	
1	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
2	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
3	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
4	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
5	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
6	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
7	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
8	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
9	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
10	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
11	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
12	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
13	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
14	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
15	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
16	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
17	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
18	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
19	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)
20	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)	食(ク)

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

◀ マウを一個クリックすると音声が発生します

表 2A 後之島方言における動詞活用形の生成
活用集(1) 子音群

		I 組1 -a 食べら	I 組1 -a 遊ぶ	I 組1 -a 行く	
		kamju	'aɸju	'ajju	
基本語幹		kam-	aɸ-	'aj-	
融合語幹		kamj-	'aɸj-	'ajj-	
目使語幹		kaɸ-	'aɸɸ-	'ajɸ-	
		kamjɸ-	'aɸjɸ-	'ajjɸ-	
活用語幹	活用形	意味 [活用]			
I 組1 終止形	-mi	～る [言ひ切り、+格、(た) ja (よ) jaE (どうう)]	kamju[mɪ]	'aɸju[mɪ]	'ajju[mɪ]
I 組1 終止形	-miN	～る [+da (よ) eɸ (って) da (ぞ) ja (た) jaE (よ) ME (らしい)]	kamju[mɪN]	'aɸju[mɪN]	'ajju[mɪN]
連体形	-mi	名詞が後接	kamju[mɪ]	'aɸju[mɪ]	'ajju[mɪ]
禁止形 1	-miNa	～るな	kamju[mɪNa]	'aɸju[mɪNa]	'ajju[mɪNa]
I 組1 終止形	-u	～る [+ad (のた) じ (の、もの、こと) wa (だた)]	kamju[mɪ]	'aɸju[mɪ]	'ajju[mɪ]
I 組1 終止形	-u	～る [+me (まい) baki (-べき) ge:d (よくに) mka (-したて)]	kamju[mɪ]	'aɸju[mɪ]	'ajju[mɪ]
係助形	-uNa	muɸ (どうして) や da (ぞ) に対する結び	kamju[mɪNa]	'aɸju[mɪNa]	'ajju[mɪNa]
			kam-	aɸ-	'aj-
連体形	-i	～る [言ひ切り、-助詞、動詞/補助動詞]	kam[i]	'aɸ[i]	'aj[i]
禁止形 2	-iNa	～るな	kam[iNa]	'aɸ[iNa]	'aj[iNa]
否定形	-iN	～ない	kam[iN]	'aɸ[iN]	'aj[iN]
否定過去形 1	-iNaE	～なかった [言ひ切り、+助詞]	kam[iNaE]	'aɸ[iNaE]	'aj[iNaE]
否定過去形 2	-iNaN	～なかった [+助詞]	kam[iNaN]	'aɸ[iNaN]	'aj[iNaN]
使役形	-iNa*	～させる	kam[iNa]*	'aɸ[iNa]*	'aj[iNa]*
受身・可動形	-iNa*	～られる	kam[iNa]*	'aɸ[iNa]* O(能)	'aj[iNa]*
意向形	-i	～よう	kam[i]	'aɸ[i]	'aj[i]
謙敬の意向形	-i	～ろ	kam[i]	'aɸ[i]	'aj[i]
謙敬の意向形	-i	～ろ (と)	ka[mɪ]	'aɸ[mɪ]	'aj[mɪ]
丁寧形	-iNa*	～ます	kam[iNa]*	'aɸ[iNa]*	'aj[iNa]*
			kaɸ-	'aɸɸ-	'ajɸ-
子用	-i	～て	ka[i]	'aɸ[i]	'aj[i]
過去形 1	-i	～た [言ひ切り、+助詞]	ka[i]	'aɸ[i]	'aj[i]
過去形 2	-iN	～た [-助詞]	ka[iN]	'aɸ[iN]	'aj[iN]
タリ形	-i	～たり	ka[i]	'aɸ[i]	'aj[i]
係助形	-ika	～れば	ka[i]ika	'aɸ[i]ika	'aj[i]ika
テイ形	-iNa*	～ている	ka[iNa]*	'aɸ[iNa]*	'aj[iNa]*

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 multimedijški 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).