Asia Pacific Journal of Corpus Research
Vol. 1, No. 2, 41-56

http://www.icr.or.kr/apjcr
e-ISSN 2733-8096
https://doi.org/10.22925/APJCR.2020.1.2.41

Word Order and Cliticization in Sakizaya: A Corpus-based Approach

Chihkai Lin
(Tatung University)

Lin, C. (2020). Word order and cliticization in Sakizaya: A corpus-based approach. *Asia Pacific Journal of Corpus Research*, *1*(2), 41-56.

This paper aims to investigate how word order interacts with cliticization in Sakizaya, a Formosan language. This paper looks into nominative and genitive case markers from a corpus-based approach. The data are collected from an online dictionary of Sakizaya, and they are classified into two word orders: nominative case marker preceding genitive case marker and vice versa. The data are also divided into three categories, according to the demarcation of the case markers, which include right, left, or no demarcation. The corpus includes 700 sentences in the construction of predicate + noun phrase + noun phrase. The results suggest that the two case markers tend to be parsed into the preceding word and show right demarcation. The results also reveal that there are type difference and distance effect of the case markers on the cliticization. Nominative case markers show more right demarcation than genitive case markers do in the corpus. Also, the closer the case markers are to the predicate, the more possible the case markers undergo cliticization.

Keywords: Sakizaya, Formosan Language, Word Order, Cliticization, Nominative Case Marker, Genitive Case Marker

1. Introduction

This paper investigates the interaction of word order and cliticization of nominative and genitive case markers in Sakizaya, an endangered Formosan language in Hualien County, eastern Taiwan. Sakizaya is a language with three case markers and three types of nouns. The three case markers include nominative case maker, genitive case marker, and oblique case marker. Nouns are divided into proper nouns and general nouns in Sakizaya, and proper nouns can be singular and plural (Tsukida, 1993; Shen, 2016). Thus, there are nine situations, as shown in Table 1.[1]

Table 1. Different Case Markers in Sakizaya

Nouns \ Cases		Nominative	Genitive	Oblique
Proper nouns	singular	ci	ni	ci … -an
	plural	ca/cini	na/nini	ca … -an/ cini … -an
General nouns		ku	nu	tu

Except for circumfixes such as *ci … -an* and *ca … -an* in oblique case marker, other case markers precede a noun in a noun phrase, as in *ku heci nu kakulut* ‘the pulp of bitter gourd’, in which the nominative case marker *ku* precedes the noun *heci* ‘pulp’, and the genitive case marker *nu* precedes

[1] The abbreviations of the glosses in this paper are listed as follows: agent focus (AF), ergative case (ERG), exclusive (EXCL), future tense (FUT), genitive case (GEN), locative (LOC), nominalizer (NML), nominative case (NOM), oblique case (OBL), patient focus (PF), perfective (PERF), plural (PL), pronoun (PRON), proper noun (PN), reduplication (RED), transitive (TRANS), and undergoer voice locative (UVL).

the noun *kakulut* ‘bitter gourd’.

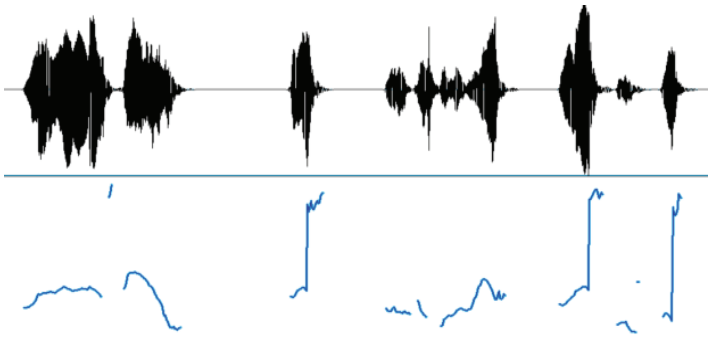
In Sakizaya, a predicate-initial language, the arguments follow the predicate, and the focus system of the predicate, agent focus vs. non-agent focus, determines the order of the core phrases. According to Tsukida (1993) and Shen (2016), nominative noun phrase precedes genitive noun phrase in agent focus, as in (1a), and genitive noun phrase precedes nominative noun phrase *in* non-agent focus, as in (1b).

(1) a. mibanic ku cucu nu katalalan.
mi-banic ku cucu nu katalalan
AF-diarrhea NOM milk GEN cow
‘(I have) diarrhea when I drink milk.’
(From Sakizaya Online Dictionary)

b. mabihkac ni Taymu ku pawalil.
ma-bihkac ni Taymu ku pawalil
PF-fall.into.trap GEN PN NOM rabbit
‘The rabbit falls into Taymu’s trap.’
(From Sakizaya Online Dictionary)

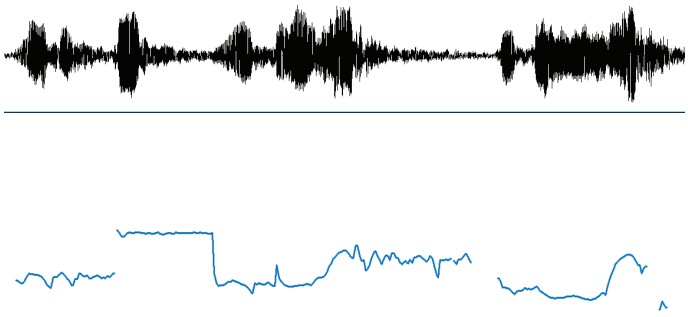
Although examples in (1) show two word orders, there is a similar mapping of the syntactic structure and phonological representation in (1a) and (1b). The syntactic structure is [ ]$_{XP}$ + [ ]$_{NP}$ + [ ]$_{NP}$, corresponding to the phonological structure [ ]$_{PW}$ + [ ]$_{PhP}$ + [ ]$_{PhP}$, as supported by the phonetic realizations in (2) (PW = prosodic word; PhP = phonological phrase).

(2a) = (1a)



mangeluay		ku	nikauzip		nu maku
[ ]$_{XP}$	+	[	]$_{NP}$	+	[ ]$_{NP}$
[ ]$_{PW}$	+	[	]$_{PhP}$	+	[ ]$_{PhP}$

(2b) = (1b)



mabihkac		ni	Taymu		ku	pawalil
[ ]$_{XP}$	+	[ ]$_{NP}$		+	[ ]$_{NP}$	
[ ]$_{PW}$	+	[ ]$_{PhP}$		+	[ ]$_{PhP}$	

Different from the phonetic realizations in (2) showing a perfect alignment of syntactic structure and phonological representation in Sakizaya, a mismatch is also observed, as in (3).

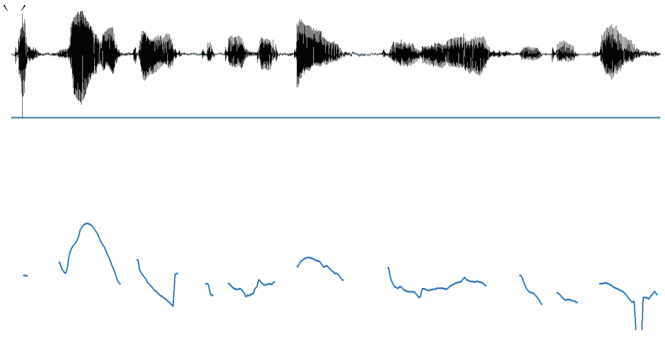
(3) ahbal ku kakabian nu luma' ni Kacaw.

ahbal	ku	kakabian	nu	luma'	ni	Kacaw
wide	NOM	room	GEN	house	GEN	PN

'Kacaw's room in his house is spacious.'
(From Sakizaya Online Dictionary)

The syntactic structure of (3) is similar to that in (1), whereas the phonetic realization of (3) shows a different phonological representation, as in (4). The phonological representation of (3), based on the phonetic realization in (4), is that the nominative case marker *ku* does not constitute a noun phrase with the noun *kakabian*. Instead, it attaches to the predicate *ahbal* 'wide', [XP] + [*ku* N] → [XP *ku*] + [N].

(4) = (3)



ahbal	ku		kakabian	nu	luma'	ni	Kacaw
[XP	+	*ku*] + [	N	+ *nu*+	N	+ *ni* +	N]

The mismatch of the syntactic and phonological phrases in (3) and (4) can be treated as cliticization (Klavans, 1985; Inkelas, 1990; Selkirk, 1996; Anderson, 2005, 2011). As there are two word orders for the noun phrases in (1a) and (1b), and the case marker may become a clitic to the preceding XP in (4), an issue arises as to whether the word order of the nominative and genitive noun phrases interacts with the cliticization of the case markers in Sakizaya.

This issue can be investigated from two aspects: type difference of the case markers and distance effect of the predicate and following noun phrases. The first aspect is to discuss whether the nominative and genitive case markers vary in cliticization. For instance, in (1a), the nominative case

marker *ku*, which precedes the genitive noun phrase, attaches to the predicate *ahbal*. The word order in (1b), on the other hand, suggests that the genitive noun phrase appears before the nominative noun phrase. In the word order like (1b), the genitive case marker should have similar behavior to the nominative case marker, which is attached to the predicate. As the phonetic realization (2b) does not support this idea, this paper aims to obtain a better understanding of the behavior of the case markers after the predicate, and therefore focuses on whether the closer the noun phrase is to the predicate, the more possible the case marker cliticizes to the predicate. Or only the nominative case markers after the predicates tend to cliticize.

The second aspect is to discuss whether the distance between the predicate and the following noun phrases affects cliticization. Suppose that the first aspect shows a positive result, the predicate-argument relationship fosters cliticization. Regardless of which case marker it is, the first noun phrase should outnumber the second noun phrase in cliticization. If the first aspect suggests a negative result, namely, a type difference between case markers being observed, it is necessary to point out which type of case marker is preferred in Sakizaya.

To explore the two aspects from a quantitative method, this paper adopts a corpus-based approach by collecting data from a Sakizaya online dictionary. This paper limits the discussion to two word orders in two case markers: predicate + NOM + GEN and predicate + GEN + NOM. The organization of this paper is as follows. Section 2 discusses clitics and cliticization. Section 3 introduces the corpus and data selection criteria, and Section 4 reports the results of the corpus data. Section 5 explores the interaction of word order and cliticization in nominative and genitive noun phrases. Section 6 concludes this paper.

2. Clitics and Cliticization

2.1. Clitics

Generally speaking, clitics are not only a morphosyntactic issue but also a phonological topic (Vogel, 2009).[2] Zwicky (1977) proposes three types of clitics: special clitics, simple clitics, and bound words. Special clitics refer to words which can be clitics as well as content words. According to Anderson (2005), French *la* has two positions in a sentence. When it is a pronoun, *la* precedes the verb *vois* 'see', as in *Je la* vois 'I see her'. When it is not a pronoun, *la* follows the verb, as in *Je vois la femme de mes rêves* 'I see the woman of my dreams.' In phonology, special clitics lack stress, in contrast to the function as content words, which carry stress.

Zwicky's (1977) simple clitics are morphemes that must attach to other words when they are phonologically weakened. For example, English pronoun her can be [hɜ˞] as a content word or [ə˞] as a clitic (Zwicky, 1977, p. 5). Zwicky's (1977) bound words are morphemes that cannot be content words. Besides, bound words are syntactically free and phonologically weakened. For example, English possessive *'s* is semantically part of a noun phrase, whereas phonologically it is affiliated with the last noun of the noun phrase, as in 'the Queen of England's hat'.

Zwicky's (1977) clitics have been modified by Anderson (2005, 2011), who proposes only phonological clitics and morphosyntactic clitics. The former refers to a component without any phonological status in a prosodic word; the latter refers to a component that is different from free morpheme and has to be governed in a phrase or a clause.[3] Furthermore, Anderson (2005, p. 79) divides the morphosyntactic clitics into three subtypes: (a) clitics related to IP or CP, such as auxiliary

[2] In Zwicky and Pullum (1983), a simple way of differentiating clitics from affixes is that clitics can attach to different part of speech, whereas affixes have to combine with the same part of speech.

[3] Anderson (2005, p. 23) suggests that the phonological clitics correspond to Zwicky's (1977) simple clitics, and the morphosyntactic clitics to Zwicky's (1977) special clitics. However, bound words have been forsaken in Anderson (2005).

and discourse particles; (b) clitics related to NP or DP, such as case markers and determiners; (c) clitics related to other XP, such as negation or emphatic particles.

2.2. Cliticization

When clitics undergo phonological processes, there are three phonological representations (Selkirk, 1996), as in (5) (FW = function word, CW = content word).

(5)	Syntax		[function word	content word]
	Phonology	(i)	(FW	$(CW)_{PW})_{PPh}$
		(ii)	((FW	$CW)_{PW})_{PPh}$
		(iii)	((FW	$(CW)_{PW})_{PW})_{PPh}$

In (5i), the function word is not a prosodic word, but it belongs to a phonological phrase with the content word. In the situation, the function word is a *free clitic*. In (5ii), the function and content words constitute a prosodic word, and they are in the same phonological phrase. The function word in (5ii) is an *internal clitic*. When a content word, also as a prosodic word, constitutes another prosodic word with the function word, as in (5iii), the function word is an *affixal clitic*.[4]

2.3. Clitics and Cliticization in Formosan Languages

Clitics are common in Formosan languages, as they have been observed in Paiwan (Ferrell, 1972), Seediq (Sung, 2016), Atayal (Liao, 2005), Puyuma (Ross and Teng, 2005), and Kavalan (Hsieh, 2016). Most clitics in Formosan languages are found in pronouns and tense-aspect-modality (TAM) markers, as in (6) for Kavalan pronoun *iku* and (7) for Kavalan perfective marker *ti*.

(6) mai=iku m-ala.
mai=**iku** m-ala
NEG=1st.PL.PRON AF-take
'I do not take (it).'
(From Hsieh 2016, p. 60)

(7) teqas ti benina zau.
teqas=**ti** benina zau
ripe=PERF banana this
'This banana has ripened.'
(From Hsieh 2016, p. 64)

In addition to the post-verbal clitics in Kavalan, pre-verbal clitics are also observed in Formosan languages. According to Tsukida (1993) and Shen (2016), TAM markers in Sakizaya can be pre-verbal and post-verbal clitics, as in (8) and (9). In Sakizaya, the past tense marker *na* and the future tense marker *a* precede the verb, and the perfective marker *tu* follows the verb.

(8) na mukan tu kiza tatayan tu kama.
na=mu-kan=**tu** kiza tatayan tu kama
PAST=AF-eat=PERF that.ACC woman OBL tangerine
'That woman ate a tangerine.
(From Shen 2016, p. 83)

[4] There is a mirror image of (5) when the function word follows the content word.

(9) a mabi' tu ci ina.
a=mabi'=**tu** ci ina
FUT=AF.sleep=PERF ACC mother
'Mother is going to sleep.'
(From Shen 2016, p. 84)

Following Selkirk (1996), the TAM markers in Sakizaya, as in (8) and (9), are affixal clitics, that is, a prosodic word and a clitic constituting another prosodic word. On the other hand, the case markers in (1) and (3) should be like free clitics. The cliticization in Sakizaya is summarized in Table 2.

Table 2. Cliticization in Sakizaya

	Prosodic word	Phonological phrase
Pre-XP	Tense markers	
Post-XP	Aspect markers, Modality markers	Case markers

In Sakizaya, the cliticization is obligatory for TAM markers within a prosodic word, whereas the cliticization is optional for case markers. As an optional process, the case markers may or may not become clitics. To better understand the cliticization of the case markers in Sakizaya, this paper adopts a corpus-based approach and probes into this issue from type difference and distance effect.

3. Corpus and Data Selection Criteria

This paper adopts a corpus-based approach and collects data and sound files from an online Sakizaya dictionary, "https://e-dictionary.apc.gov.tw/ais/Search.htm", which is established by the Council of Indigenous Peoples in Taiwan. Although there are nine situations in Table 1, this paper does not investigate all the three case markers. As the oblique case marker for a proper noun is represented by a circumfix *ci* … *-an*, it would be difficult to observe cliticization. Thus, this paper focuses on the nominative and genitive case markers. The corpus uses case markers not only for general nouns *ku* and *nu*, but also for proper nouns *ci* and *ni*. Besides, this paper only explores monosyllabic case markers. The plural forms for pronouns in nominative and genitive case markers, *ca/cini* and *na/nini*, are not analyzed. In other words, disyllabic case markers such as *cini* and *nini* are not included in the corpus. Finally, to restrict the number of noun phrases in a sentence, this paper uses sentences with only two noun phrases after the predicate, [XP] + [NP] + [NP].

The online Sakizaya dictionary provides a quick survey by using keywords. With a keyword, all the relevant sentences can be quickly collected. In this paper, there are four keywords:

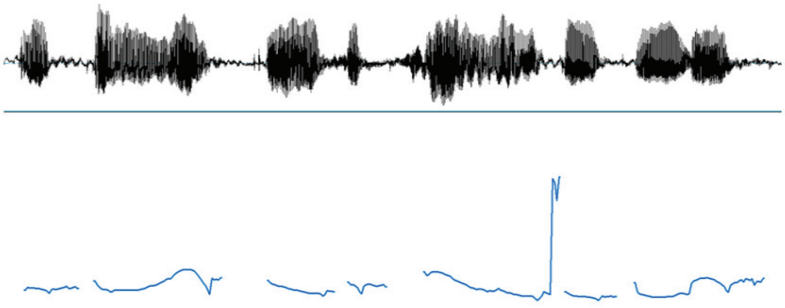
ku 'nominative case marker (general noun)'
ci 'nominative case marker (proper noun)'
nu 'genitive case marker (general noun)'
ni 'genitive case marker (proper noun)'

The collected sentences are scrutinized, and only those with the nominative and genitive noun phrases are analyzed. Based on the word order, the selected sentences are divided into two groups: [XP] + $[NP]_{NOM}$ + $[NP]_{GEN}$ and [XP] + $[NP]_{GEN}$ + $[NP]_{NOM}$. In each group, the data are further classified into three subgroups depending on how the case markers are phonetically demarcated. The criteria for determining the demarcation are based on pauses or pitch changes before or after the case markers. There are three patterns of demarcating the case markers. First, the case marker is part of the preceding word, [XP + ku/nu/ci/ni] + [N], when there is a pause or pitch change after the case marker.

The case marker may stay within its original noun phrase, [XP] + [ku/nu/ci/ni N]. In this case, there is a pause before the case marker. For example, the nominative case marker *ku* in (2a) and that in (4) have shown different positions of pauses. The pause appears before the case marker in (2a); the pause appears after the case marker in (4). When the case marker is not demarcated at all, there is no salient pause before or after the case marker, [XP + ku/nu/ci/ni + N], as shown in (10) and (11).

(10) akenek ku heci nu kakulut.
akenek ku heci nu kakulut
bitter NOM frui GEN bitter.melon
'Bitter melon is bitter.'
(From Sakizaya Online Dictionary)

(11)



akenek ku heci nu kakulut
[XP NOM N GEN N]

According to the above criteria, there are nine categories of phonetic realizations for each word order in the corpus, as explicitly listed in (12) (C = case marker).

(12) a. [XP + C] + [N + C] + [N]
b. [XP + C] + [N] + [C + N]
c. [XP + C] + [N + C + N]
d. [XP] + [C + N + C] + N]
e. [XP] + [C + N] + [C + N]
f. [XP] + [C + N + C + N]
g. [XP + C + N + C] + [N]
h. [XP + C + N] + [C + N]
i. [XP + C + N + C + N]

With two word orders ([XP] + $[NP]_{NOM}$ + $[NP]_{GEN}$ and [XP] + $[NP]_{GEN}$ + $[NP]_{NOM}$) and the nine categories in (12), there are 18 outputs for the corpus instances. After all the data are carefully sorted, the distribution is reported in Section 4.

4. Results

The corpus includes 700 sentences, which are presented in two viewpoints by focusing on the noun phrases in two word orders. The first viewpoint is to zoom in the first noun phrase, as shown in Tables 3 and 4. Table 3 shows the distribution in the order of [XP] + $[NP]_{NOM}$ + $[NP]_{GEN}$.

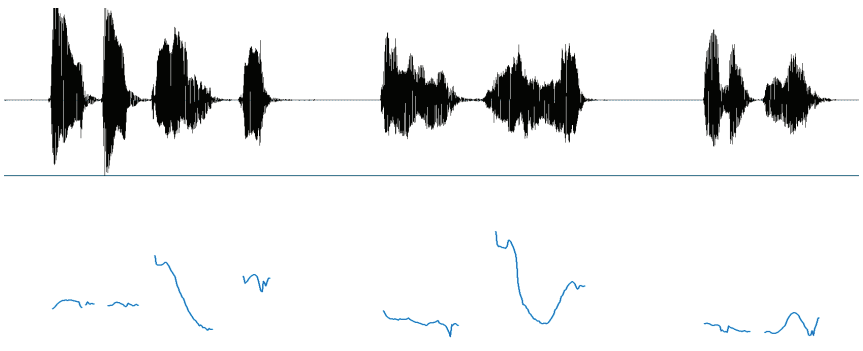
Table 3. The Distribution of [XP] + $[NP]_{NOM}$ + $[NP]_{GEN}$ (Focusing on NP1)

NP1	NP2	*ku*	*ci*	Total
ku/ci]	nu/ni]	204	1	205
	[nu/ni	38	0	38
	NO	98	3	101
[ku/ci	nu/ni]	3	0	3
	[nu/ni	2	0	2
	NO	6	1	7
NO	nu/ni]	45	0	45
	[nu/ni	11	1	12
	NO	46	2	48
Total		453	8	461

There are 461 corpus examples in nine categories in Table 3. About 75% of the corpus examples (344 tokens) are in the category where the first noun phrase demarcates on the right edge. In Table 3, two categories exceed 100 tokens: 205 tokens for the second noun phrase which is demarcated on the right edge, and 101 tokens for the second noun phrase without demarcation. (13) and (14) show examples of the right demarcation of the two case markers.

(13) auaukan ku kanamuhan nu padakuwan.
au-auk-an ku ka-namuh-an nu badakuwan
RED-bamboo-LOC NOM KA-like-NML GEN egret
'Egrets like a place with many bamboos.'
(From Sakizaya Online Dictionary)

(14)

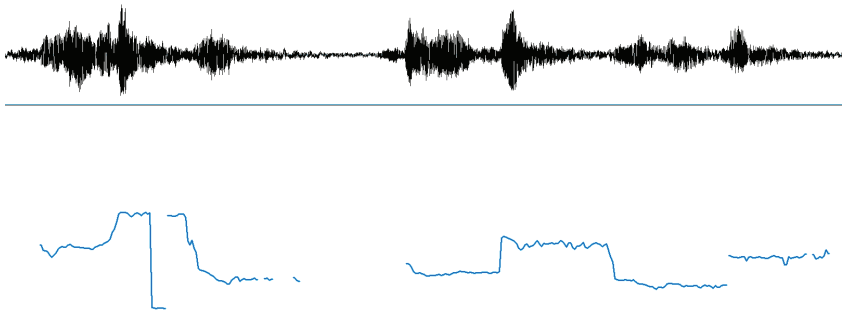


auaukan ku kanamuhan nu padakuwan
[XP + NOM] [N + GEN] [N]

(15) and (16) show examples when there is only a pause after the nominative case marker *ku*.

(15) sumanah ku sangutuc nu bintic.
Sumanah ku sangutuc nu bintic
red NOM mouth GEN king.fisher
'King fisher's beak is red.'
(From Sakizaya Online Dictionary)

(16)

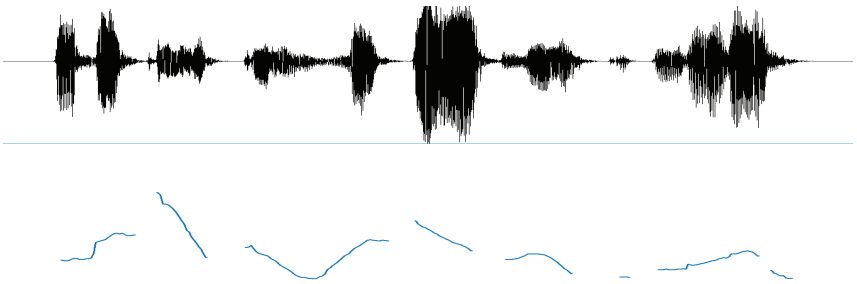


sumanah	ku		sangutuc	nu		bintic
[XP	+	NOM]	[N	+ GEN	+	N]

About 23% of the corpus instances (105 tokens) are found in the category when there is no demarcation on the first noun phrase. 48 tokens in this category lack any parsing of the case markers in the sentences. In the corpus, only 12 tokens show the left edge on the case marker in the first noun phrase. Two examples are shown in (17) – (20). (17) and (18) show an example when the genitive case marker is parsed into the preceding word. (19) and (20) show an example when there is no demarcation of the case markers.

(17) tabaki ku mata nu katalalan.
tabaki ku mata nu katalalan
big NOM eye GEN cow
'The cow's eyes are big.'
(From Sakizaya Online Dictionary)

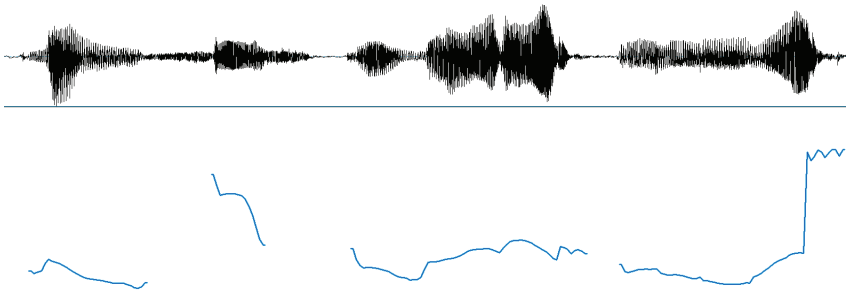
(18)



tabaki	ku	mata	nu	katalalan
[XP +	NOM +	N +	GEN]	[N]

(19) bangsis ku balu nu mami'.
bangsis ku balu nu mami'
fragrant NOM flower GEN pomelo
'The flower of pomelo is fragrant.'
(From Sakizaya Online Dictionary)

(20)



bangsis ku balu nu mami'
[XP + NOM + N + GEN + N]

There are 240 tokens when the word order is $[NP]_{GEN}$ + $[NP]_{NOM}$, and Table 4 shows the distribution in the order of [XP] + $[NP]_{GEN}$ + $[NP]_{NOM}$.

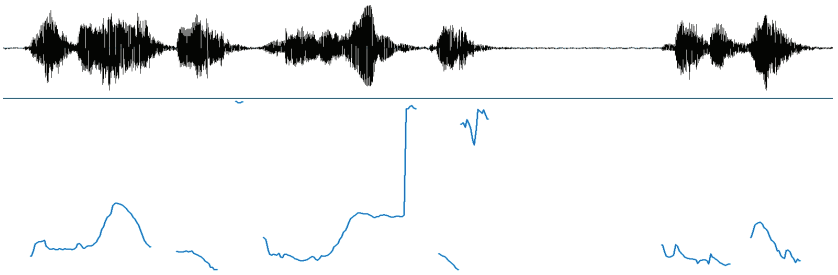
Table 4. The Distribution of [XP] + $[NP]_{GEN}$ + $[NP]_{NOM}$ (Focusing on NP1)

NP1	NP2	*nu*	*ni*	Total
nu/ni]	ku/ci]	56	36	92
	[ku/ci	8	1	9
	NO	14	3	17
[nu/ni	ku/ci]	1	0	1
	[ku/ci	0	2	2
	NO	1	0	1
NO	ku/ci]	42	18	60
	[ku/ci	5	2	7
	NO	37	13	50
Total		164	75	239

In Table 4, about half of the examples (118 tokens) are in the category where the genitive case marker is affiliated with the predicate. In this category, 92 tokens are attested in the cell in which the case markers in the two noun phrases are parsed into the preceding words. An example of this category is shown in (21) and (22).

(21) mahuwad nu baliyus ku badahung.
ma-huwad nu baliyus ku badahung
PF-move GEN typhoon NOM roof
'The typhoon blew away the roof.'
(From Sakizaya Online Dictionary)

(22)

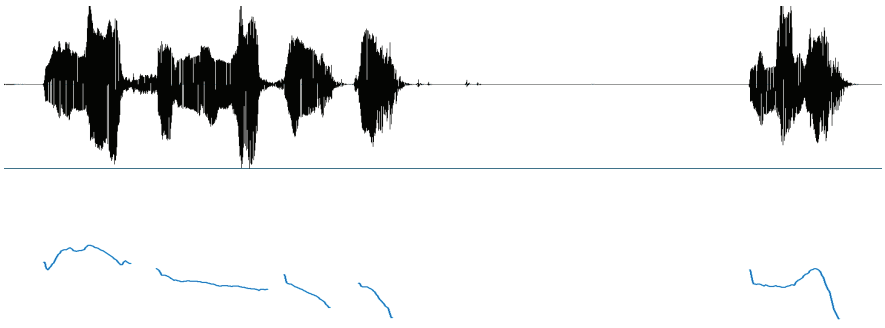


mahuwad nu baliyus ku badahung
[XP + GEN + N + NOM] [N]

A quarter of the corpus instances (60 tokens) are observed in the situation when the genitive case marker is not parsed, but the nominative case marker is part of the preceding noun, as shown in (23) and (24).

(23) ninac han nu wacu ku pinanum.
ninac=han nu wacu ku pi-nanum
sip=PERF GEN dog NOM PI-water
'The dog sipped the water.'
(From Sakizaya Online Dictionary)

(24)



ninac han nu wacu ku pinanum
[XP + GEN + N + NOM] [N]

The second viewpoint of presenting the corpus data is to focus on the second noun phrase, as in Tables 5 and 6, respectively.

Table 5. The Distribution of [XP] + $[NP]_{NOM}$ + $[NP]_{GEN}$ (Focusing on NP2)

NP1	NP2	*ku*	*ci*	Total
ku/ci]	nu/ni]	204	1	205
[ku/ci		3	0	3
NO		45	0	45
ku/ci]	[nu/ni	38	0	38
[ku/ci		2	0	2
NO		11	1	12
ku/ci]	NO	98	3	101
[ku/ci		6	1	7
NO		46	2	48
Total		453	8	461

In Table 5, the majority (205 tokens) goes to the cell where the case markers have the right demarcation. More than 20% of the corpus examples (101 tokens) are found in the cell in which the first case marker belongs to the predicate, but the second case marker lacks any demarcation. Also, more than 10% of the corpus examples (48 tokens) are in the cell where the case markers are not parsed.

In Table 6, approximately 85 % of the corpus examples are found in three cells.

Table 6. The Distribution of [XP] + $[NP]_{GEN}$ + $[NP]_{NOM}$ (Focusing on NP2)

NP1	NP2	*nu*	*ni*	Total
nu/ni]	ku/ci]	56	36	92
[nu/ni		1	0	1
NO		42	18	60
nu/ni]	[ku/ci	8	1	9
[nu/ni		0	2	2
NO		5	2	7
nu/ni]	NO	14	3	17
[nu/ni		1	0	1
NO		37	13	50
Total		164	75	239

The first cell (92 tokens) is that the two case markers are parsed into the preceding words. The second cell shows that there are 60 tokens in the situation when there is no demarcation on the first case marker, but there is a right demarcation on the second case marker. 50 tokens are attested in the cell when the case markers are not parsed at all.

5. Discussion

Based on the data in Section 3, this section discusses three issues of the cliticization of the nominative and genitive case markers in Sakizaya. As edges and positions are indicators for cliticization, Table 7 shows the sum of the tokens in the three types of demarcation of the case markers in the two NPs.

Table 7. Demarcation of case markers in different positions

NP1 \ NP2	Right edge	Left edge	NO	Total
Right edge	297	47	118	462
Left edge	4	4	8	16
NO	105	19	98	222
Total	406	70	224	700

The first issue is that cliticization of the case markers is a pervasive but not an obligatory process in Sakizaya. In Table 7, half of the tokens show that the two case markers are parsed into the preceding words (50%, 352/700). In the other half of the corpus examples, 36% of the corpus examples show that one of the two case markers is parsed into the preceding word, and 14% show that the two case markers maintain in the same noun phrase. Examples of no demarcation are given in (25) and (26).

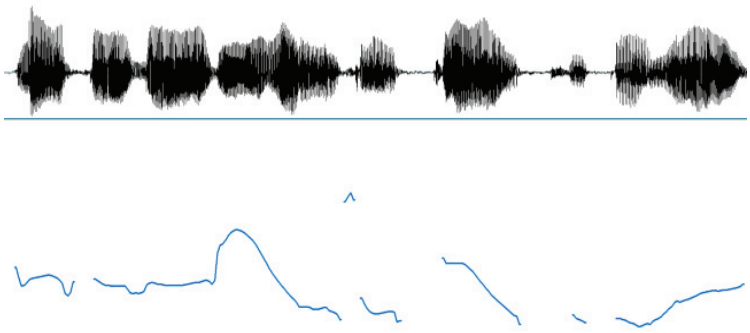
(25) mapabaliw ni Kacaw ci Panay.

ma-pabaliw	ni	Kacaw	ci	Panay
PF-doubt	GEN	PN	NOM	PN

'Kacaw doubts Panay.'

(From Sakizaya Online Dictionary)

(26)



mapabaliw	ni	Kacaw	ci	Panay
[XP	+ GEN	+ N	+ NOM	+ N]

The second issue is that the right edge of the case markers is more frequently demarcated than the left edge of the case markers is in Sakizaya. Focusing on the first NP, about 66% of the corpus examples (462 tokens) show that the case markers have right demarcation, as they are parsed into the predicate. With regard to the second NP, about 58% of the corpus instances (406 tokens) show right demarcation, as they become part of the first NP.

There is a low frequency in Table 7 when the case markers have left demarcation in the first NP (16 tokens, about 2 %). This fact suggests that having a boundary between the predicate and the first noun phrase is not preferred in Sakizaya. However, the percentage increases in the second NP (70 tokens, 10%), showing that the case markers are more likely to stay in the second NP. This discrepancy can be interpreted as that demarcating on the left edge of the second noun phrase is more acceptable than that of the first noun phrase in Sakizaya.

The third issue is concerned with different positions between the nominative and genitive case markers. In the corpus, there are two word orders: predicate + NOM + GEN and predicate + GEN + NOM. Table 8 shows the percentage of the two case markers when the demarcation is on the right edge of the case markers.

Table 8. Percentage of the Case Markers in Different Orders (Right Demarcation Only)

Case \ NP	NP1	NP2
NOM	75% (344/461)	73% (153/239)
GEN	49% (118/239)	55% (253/461)

In Table 8, 75% of the examples of the nominative case marker are demarcated in the first NP, and 73% are in the second NP. The genitive case marker shows a lower percentage, 55% in the first NP, and 53% in the second NP.

A Chi-square test based on Table 8 is conducted to check if there is a significant difference between the positions and the types of case markers. The result suggests that there is a significant difference between the positions and the types of case makers ($p < 0.01$). Regardless of the positions, the nominative case marker *ku/ci* is more likely to undergo cliticization than the genitive case marker *nu/ni* is in Sakizaya.

6. Conclusion

This paper adopts a corpus-based approach to the cliticization of the nominative and genitive case markers in Sakizaya from a phonological perspective. With 700 sentences, the data show that half of the corpus examples undergo cliticization, and the right edge after the case markers is more frequently observed than the left edge in the corpus. It is also found that the nominative case markers are more prevalent than the genitive case markers in demarcation, regardless of the positions where the nominative case marker appears in a sentence.

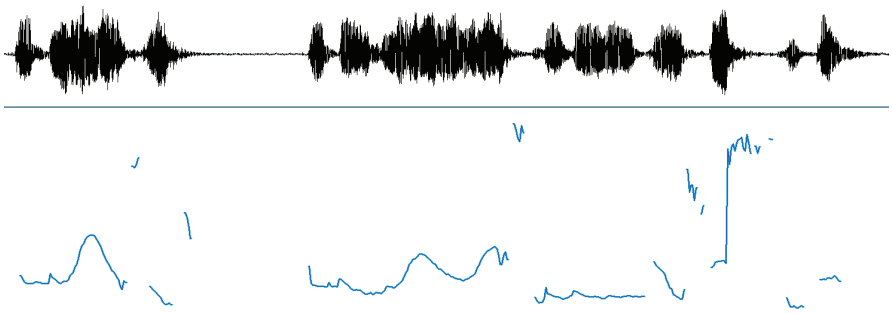
This paper has put forward two aspects of the cliticization in Sakizaya: type difference and distance effect. First, there is a type difference between the two case markers. When there is a right demarcation, more than 70% of the nominative case markers cliticize to the predicate or the preceding word, while only half of the genitive case markers undergo cliticization. Frequency could be the factor, as the nominative case markers outnumber the genitive case markers in the corpus (461 vs. 239).

The corpus data also show a distance effect in the cliticization of the two case markers. Table 7 has shown that when there is cliticization, 97% of the case markers in the first noun phrase cliticize to the predicate, while 85% of the case markers in the second noun phrase undergo cliticization. It is apparent that the closer the case markers are to the predicate, the more possible cliticization takes place.

On the basis of the cliticization of the two case markers, other case markers can be included in the corpus for future research. First, as discussed in Table 1, Sakizaya has three case markers, but this paper only deals with nominative and genitive case markers. The oblique case marker *tu* has not yet been investigated. When the oblique noun phrase is taken into account, a sentence consists of three arguments, as shown in (27) and (28).

(27) kapuwa ku pipazengan nu babalaki tu cikah.
kapuwa ku pi-pazeng-an nu babalaki tu cikah
clay.pot NOM PI-put-NML GEN the.elder OBL salt
'The old man puts salt in a clay pot.'
(From Sakizaya Online Dictionary)

(28)



kapuwa ku pipazengan nu babalaki tu cikah
[XP + NOM] [N + GEN] [N + OBL] + [N]

The word order in (27) is that the predicate is in the initial position, and the three noun phrases are in the order, NOM > GEN > OBL. The phonological representation in (28) reveals that the three case markers are all parsed into the preceding XP or nouns. Whether there are type difference and distance effect in the complex structure is left for future research.

In addition to the three case markers, locative case marker i is also an intriguing issue. As shown in (29) and (30), the locative case marker *i* is parsed into the preceding word *husing* 'balloon'.

(29) yadah ku husing i pikikungan.
yadah ku husing **i** pikikungan
many NOM balloon **LO**C wedding
'There are many balloons in the wedding.'
(From Sakizaya Online Dictionary)

(30)

yadah ku husing i pikikungan
[XP + NOM] [N + LOC] [N]

If the distribution of the oblique and locative case markers is available in the future, a complete picture of the cliticization in Sakizaya will be obtained.

Finally, the corpus-based approach to the cliticization of the two case markers in Sakizaya has an implication. The results have provided solid evidence for the cliticization of the Sakizaya case markers at the phonological phrase. Only when the cliticization takes place at the phonological phrase can the case markers be free to move from one phrase to the other phrase.

References

Anderson, S. (2005). *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.

Anderson, S. (2011). Clitics. In Oostendorp, M., Ewen, C., Hume, E., & Rice, K. (Eds.), *The Blackwell Companion to Phonology*. Malden, MA: Wiley-Blackwell.

Ferrell, R. (1972). Verb systems in Formosan languages. In Thomas, J., & Bernot, L. (Eds.), *Langues et Techniques* (pp. 121-128). Paris: Klincksieck.

Hsieh, F. H. (2016). *An Introduction to Kavalan Grammar*. Taipei: Council of Indigenous Peoples.

Inkelas S. (1990). *Prosodic Constituent in the Lexicon*. New York, NY: Garland Publishing Co.

Klavans, J. L. (1985). The independence of syntax and phonology in cliticization. *Language*, *61*(1), 95-120.

Liao, H. C. (2005). Another look at the order of clitic pronouns in Wulai Atayal. *Concentric: Studies in Linguistics*, *31*(1), 47-63.

Ross, M. D., & Teng, F. C. (2005). Formosan languages and linguistic typology. *Language and Linguistics*, *6*, 739-781.

Selkirk, E. (1996). The prosodic structure of function words. In Demuth, K., & Morgan J. (Eds.), *Signal to Syntax: Bootstrapping speech to grammar in early acquisition* (pp. 187-213). New York, NY: Lawrence Erlbaum.

Shen, W. C. (2016). *An Introduction to Sakizaya Grammar*. Taipei: Council of Indigenous Peoples.

Sung, L. M. (2016). *An Introduction to Seediq Grammar*. Taipei: Council of Indigenous Peoples.

Tsukida, N. (1993). A brief sketch of the Sakizaya dialect of Amis. *Tokyo University Linguistic Paper*, *13*, 372-413.

Vogel, I. (2009). Universal of prosodic structure. In Scalise, S., Magni, E., & Bisetto, A. (Eds.), *Universals of Language Today*, (pp. 59-82). Amsterdam: Springer.

Zwicky, A. (1977). *On Clitics*. Bloomington, IN: Indiana University Press.

Zwicky, A., & Pullum, G. (1983). Cliticization vs. inflection: English *n*'t. *Language*, *59*, 502-513.

THE AUTHOR

Chihkai Lin is an Assistant Professor at the Department of Applied Foreign Languages, Tatung University (Taiwan). He specializes in phonology and historical linguistics focusing on East Asian languages. His research interests include historical changes in Chinese dialects, Japanese and Korean using philological data, and synchronic phonology in Formosan languages, Chinese, Japanese, and Korean.

THE AUTHOR'S ADDRESS

First and Corresponding Author
Chihkai Lin
Professor
Department of Applied Foreign Languages
Tatung University
No.40, Sec. 3, Zhongshan N. Rd., Taipei City 104, TAIWAN
E-mail: linchihkai@gmail.com

Received: 9 September 2020
Received in Revised Form: 9 December 2020
Accepted: 19 December 2020