Case Syncretism in Kasavakan Puyuma: A Field Data Analysis of Noun Phrase Markers

Deborah Watty[1], Yung-Jui Yao[1], Jens N. Watty[2]

[1]Graduate Institute of Linguistics, National Taiwan University
[2]Department of Physics, National Taiwan University
{r11142012, r10142002, r11222082}@ntu.edu.tw

Abstract

Previous research has reported differing patterns of case syncretism across three dialects of Puyuma, an Austronesian language of Taiwan (Nanwang, Katipul, Ulivelivek). This study presents a quantitative analysis of case syncretism of noun phrase markers and disambiguation strategies in the Kasavakan dialect. Our dataset comprises 377 sentences elicited from five speakers, which we annotated for voice, potential semantic ambiguity, word order, and case marking of different semantic roles. We find evidence for a high degree of syncretism between genitive and nominative markers, alongside a decline in the use of genitive forms, particularly for common definite nouns. Some overlap with oblique markers is also attested, suggesting varying degrees of case syncretism between speakers. Topicalization appears to be the most frequent disambiguation strategy, while the order of non-topicalized noun phrases does not seem to aid disambiguation. Other factors, including age and individual experiences may contribute to inter-participant variation. These findings contribute to a more complete understanding of case marking in Puyuma by adding new empirical data from the Kasavakan dialect, where patterns of syncretism and disambiguation differ from previously described varieties.

1 Introduction

Puyuma is a Formosan language spoken primarily in Taitung County in southeastern Taiwan by the Puyuma people, whose population is approximately 13,000 (Teng, 2018). It is particularly relevant to the reconstruction of Proto-Austronesian, as it has been argued to represent one of the primary branches of the Austronesian language family (Ross, 2009). Traditionally, the Puyuma community consists of eight main villages: Puyuma (Nanwang), Katipul, Rikavung, Tamalakaw, Kasavakan, Pinaski, Alipai, and Ulivelivek (Teng, 2009, 2018).

For educational purposes, the language is classified into four dialects: Puyuma (Nanwang), Katipul, Kasavakan, and Ulivelivek (Teng, 2018).

Similar to other Austronesian languages, central features of Puyuma syntax are its use of noun phrase markers (NPMs) and the distinction between actor voice and undergoer voice. In actor voice (AV), the subject of the verb is the actor (Example 1). In undergoer voice, an argument other than the actor becomes the subject, while the actor is relegated to non-subject status. The three variants of undergoer voice are Patient Voice (PV, Example 2), Locative Voice (LV, Example 3), and Conveyance Voice (CV, Example 4):

1. (Nanwang, Teng, 2008)[1]
   tr<em>akaw dra   paisu i   Isaw
   <AV>steal   OBL.INDF money NOM.SG Isaw
   Isaw stole money.

2. (Nanwang, Teng, 2008)
   tu=trakaw-aw   na   paisu kan
   3SG.GEN=steal-PV NOM.DEF money OBL.SG
   Isaw
   NAME
   Isaw stole the money.

3. (Nanwang, Teng, 2008)
   tu=trakaw-ay=ku   dra   paisu
   3SG.GEN=steal-LV=1SG.NOM OBL.INDF money
   kan   Isaw
   OBL.SG NAME
   Isaw stole money from me.

4. (Nanwang, Teng, 2008)
   tu=trakaw-anay   i   tinataw
   3SG.GEN=steal-CV NOM.SG his.mother
   dra   paisu
   OBL.INDF money
   He stole money for his mother.

---

[1]In examples from other papers, we have changed the glossing conventions to our own for the sake of comparability. A list of abbreviations is included in Appendix A.

Note how in the above examples, all of which are from the Nanwang dialect, the non-subject actor is marked with an oblique NPM (see Examples 2 and 3). This is not always the case in the Katipul and Ulivelivek dialects, which sometimes use a genitive marker (Teng, 2009). In the Nanwang dialect, genitive and oblique NPMs are fully syncretic (see Figure 1).

While the use of NPMs has been documented for the other three dialects, the Kasavakan dialect remains understudied in this regard. To provide a more comprehensive understanding of Puyuma NPMs, it is of research interest to document the Kasavakan NPMs.

We initially collected an explorative data set from only one speaker who seemed to display a distinct pattern of case syncretism, as shown in Example 5:

5. (Kasavakan, Speaker 1)
tu=pa-ated-ay   na   pawko
3SG.GEN=CAUS-send-LV NOM.DEF package
i   Dipung *i*   Simuy
LOC Japan   NOM.SG NAME
Simuy sent a package to Japan.

Here, the non-subject actor, Simuy, is marked with a nominative NPM, suggesting that case syncretism might go in the opposite direction of Nanwang, with nominative and genitive being syncretic. However, data collected from four additional speakers showed a complex and variable pattern with notable inter-speaker variability. A quantitative approach was thus adopted to account for the nature of the distribution of these markers more effectively.

The goals of this study were as follows:

1. Describe the use of NPMs in Kasavakan Puyuma, an aspect not previously studied.

2. Identify preferred disambiguation strategies in cases of syncretism.

This paper is structured as follows. Section 2 reviews prior work by Teng (2009), which describes NPM use and disambiguation strategies in Nanwang, Ulivelivek, and Katipul Puyuma. Section 3 outlines the data collection and annotation process. Section 4 presents our findings, including the analytical approach and illustrative examples from our dataset. Section 5 discusses the implications of the results and offers additional observations.

Figure 1: Illustrative example of genitive-oblique case syncretism in Nanwang Puyuma (personal singular genitive marker), based on Teng (2009).

2 Literature Review

To our knowledge, Teng (2009) is the only work that covers the differences between the patterns of case syncretism between different dialects of Puyuma in detail. Citing Baerman et al. (2005), Teng distinguishes between diachronic and synchronic syncretism, the former referring to forms being merged over time so that the distinction between the two forms disappears, and the latter referring to one form covering two functions in certain cases whereas those functions have separate forms elsewhere in the language.

Teng's reconstruction of Proto-Puyuma NPMs contains two genitive markers: *ni* for personal singular nouns and *nina* for common definite nouns. In the case of Nanwang Puyuma, Teng argues that the genitive case has completely syncretized with the oblique case, resulting in a pattern where rather than a three-way distinction between subjects (nominative), non-subject actors and possessors (genitive) and other non-core arguments (oblique), the distinction is now between subjects (nominative) and non-subjects (oblique). As illustrated in Figure 1, this has resulted in genitive markers becoming obsolete.

In Katipul, which is geographically closest to Kasavakan, genitive markers are replaced by nominative markers rather than oblique markers for common definite nouns, such as in Example 6:

6. (Katipul, Teng, 2009)
tu=atek-aw   na   sa'az
3SG.GEN=hack-PV   NOM.DEF branch
na   lakak
NOM.DEF   children
The children hacked the branches.

Rather than a distinction between subject and non-subject as in Nanwang, the new distinction in

65

this case appears to be whether or not the noun phrase is a core argument, with nominative NPMs being used to mark all core arguments including subjects, possessors, and non-subject actors.

However, nominative-genitive syncretism is not complete in Katipul. Genitive markers have not been lost entirely and tend to be used for disambiguation, as in Example 7:

7. (Katipul, Teng, 2009)
tu=karatr-aw   na   suan *nina*
3SG.GEN=bite-PV NOM.DEF dog   GEN.DEF

unan
snake
The snake bit the dog.

In Example 6, the semantics of the verb alone are sufficient to disambiguate the actor, whereas in Example 7, using nominative *na* to mark both nouns would result in the sentence being semantically ambiguous. Using the genitive marker *nina* clearly marks the snake as the actor in this sentence.

The fact that genitive markers have not been entirely replaced by nominative markers in Katipul Puyuma becomes even more apparent when considering personal singular nouns, where a distinction between nominative and genitive is obligatory. Interestingly, common indefinite nouns show the same genitive-oblique syncretism as seen in Nanwang Puyuma.

The situation is similar in Ulivelivek Puyuma, with one difference being that genitive-oblique syncretism also applies to common definite non-subject actors (see Example 8), but not to possessors (see Example 9):

8. (Ulivelivek, Teng, 2009)
tu=senan-ay
3SG.GEN=sunburned-LV

nina/*kana*/**na*   kadaw
GEN/OBL/*NOM.DEF   sun
It was burned by the sun.

9. (Ulivelivek, Teng, 2009)
tu=tial   *nina*/*na*/**kana*   suan
3SG.GEN=belly GEN/NOM/*OBL.DEF dog
the dog's belly

Teng's findings can be summarized as follows:

- In Nanwang Puyuma, genitive and oblique markers are fully syncretic, i.e., oblique markers have completely replaced possessive and genitive markers, leading to ambiguities between non-subject actors and other oblique noun phrases.

- In the Katipul dialect, nominative and genitive/possessor markers are partially syncretic for common definite nouns, leading to ambiguities between actor and subject in undergoer voice. This type of ambiguity can be resolved by using the more specific genitive marker or through topicalization. The distinction between nominative and genitive is obligatory for personal nouns. For common indefinite nouns, the pattern is the same as in the Nanwang dialect.

- Only the Ulivelivek variety distinguishes between markers for non-subject actors and possessors in some cases. While a specific genitive marker exists as in the Katipul dialect, genitive NPMs are partially syncretic with oblique NPMs, and possessor NPMs are partially syncretic with nominative NPMs.

- Preferred disambiguation strategies vary by dialect as ambiguities arise in different situations. Depending on the dialect, strategies can include topicalization, word order, verbal semantics and cross-referencing.

3 Dataset

The dataset[2] we present was selected from sentences collected during interviews with five different speakers of Kasavakan Puyuma (4 female and 1 male, born between 1953 and 1958). All interviews took place in the first half of 2024 as part of a class on field linguistics. Communication with the speakers was conducted through Mandarin. Speakers were interviewed separately so that they could not directly comment on each other's sentences. The final dataset contains 377 sentences, some of which were directly elicited, while others were presented to the speaker and rated acceptable or unacceptable (see Figure 2).

The criteria for inclusion of a sentence in the final dataset were as follows:

- The sentence contains a semantically transitive verb and/or a possessive structure.

- No pronouns except for the genitive clitic *tu* in undergoer voice sentences.

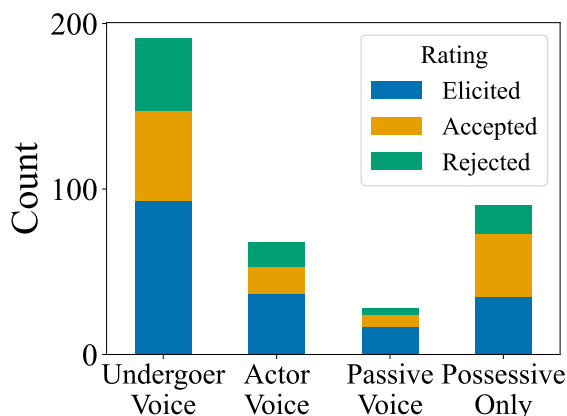[2] https://github.com/deborahwatty/kasavakan_npms.git

Figure 2: Overview of the number of sentences by voice and rating. The dataset also includes entries that only consist of a possessive construction.

- Sentence structure not too complex, no existential clauses, no copula.
- We excluded sentences where the speaker was unsure about the meaning or correctness.

Annotation was performed by consensus among the authors. Each sentence in the dataset was annotated for the following features:

Rating

We annotated whether sentences were *directly elicited*, suggested by us and rated *acceptable*, or *rejected* by the speakers.

Voice

This feature denotes whether the sentence is in *actor voice*, *passive* or one of the undergoer voices (*patient*, *locative*, or *conveyance*). When one of the semantic roles is filled by a possessed noun, this is also annotated in this column.

Semantic Roles

The possible semantic roles in the dataset are *actor*, *undergoer*, *beneficiary*, *location*, *possessor* and *theme*. For each sentence, we annotated the type (*personal* or *common* noun, *definite* or *indefinite* for common nouns, and *singular* or *plural* for personal nouns) as well as the grammatical case of the noun that fills each semantic role. Note that for common definite nouns, the annotation does not distinguish between cases where an NPM was used and cases where a demonstrative was used (such as *ini* or *kanidu*, see Example 21). These cases are counted as variations of the nominative marker *na* or oblique marker *kana*, respectively, which is the approach taken by Teng (2009).

Ambiguity

This feature describes which of the semantic roles in a given sentence may be confused for each other when given only the verb in its basic form and the words that fill the different semantic roles, but not the word order or further context. For example, given the words "to eat", "cookie" and "Tom", it would be reasonable to assume that "Tom" is the actor and the "cookie" is the undergoer with little room for ambiguity. However, when given the words "to bite", "snake" and "dog", it is unclear which of the animals is the actor and which is the undergoer. In this case, we would annotate *actor-undergoer ambiguity*.

Word Order

We also annotated the order in which the different constituents of the sentence are arranged. For example, the word order of Example 12 is annotated as *Verb-Undergoer-Beneficiary-Actor*. *Possessors* and *possessed nouns* are annotated individually; where these structures are described by a *predicate* (such as "is thick" or "are round"), the predicate is denoted as such. Where there are additional constituents that do not fit into any of the aforementioned semantic roles, the letter *X* denotes the presence of additional constituents at this position in the sentence.

Topicalization

Here, we annotated which, if any, of the constituents are topicalized (i.e., placed at the beginning of the sentence, before the verb).

4 Analysis and Results

4.1 Distribution of Noun Phrase Markers

The distribution of NPMs in our dataset is compared to the data by Teng (2009, 2018) in Table 1.

Below follows a more detailed description of the data used to fill in the fields for possessors (PSR) and non-subject actors (GEN) in the Kasavakan column. As all topicalized nouns were consistently marked as nominative (with a single exception, all sentences where this was not the case were rejected by all speakers), examples where the noun in question is topicalized are excluded from this part of the analysis. Post-hoc two-sided binomial tests (Siegel and Castellan, 1988) were conducted to examine whether speakers showed a tendency to prefer any specific marker.

			Proto-Puyuma	Nanwang	Katipul	Ulivelivek	Kasavakan
Personal	Singular	NOM	*i	i	i	i	i
		PSR	*ni	kan	ni	ni	i / ni / kani
		GEN	*ni	kan	ni	ni	i / ni / (kani)
		OBL	*ka-ni		kani	kani	kani / i?
	Plural	NOM	*na	na	na	na	na
		PSR	(unknown)	kana	nina / na	(unknown)	na / kana
		GEN	(unknown)	kana	nina / na	(unknown)	na / (kana?)
		OBL	*ka-na		kana	kana	kana
Common	Definite	NOM	*na	na	na	na	na
		PSR	*ni-na	kana	na / nina	nina / na	na / kana
		GEN	*ni-na	kana	na / nina	nina / kana	na / kana
		OBL	*ka-na		kana	kana	kana
	Indefinite	NOM	*a	a	a	a	a
		PSR					(da) / (a)
		GEN	*dra	dra	za	za	a / (da)
		OBL					da

Table 1: Comparison of the distribution of NPMs across dialects. The data on the Nanwang, Katipul and Ulivelivek variants as well as the tentative reconstruction of Proto-Puyuma is taken from Teng (2009), the personal plural NPMs for Katipul come from Teng (2018), and the Kasavakan column is based on our dataset. Parentheses indicate that instances of the form were rated acceptable by speakers but not actively used by any speaker. Where multiple markers were acceptable, they are sorted by the perceived preference of the speakers overall. NPMs that only occurred once are marked with a question mark. The case column denotes the function that the NPMs take on in a sentence, whereas the glossing in examples in the main text denotes the form of the NPM – not the function.

4.1.1 Personal Singular Nouns

There is some flexibility in how personal singular possessors are marked. While the nominative marker was used most often in elicited sentences, the difference is not statistically significant (two-sided binomial test, nominative vs. non-nominative, $N = 19$, $p_0 = 1/3$, $p = 0.089$; Figure 3). Instances of all three markers are found in the data, although there seem to be differences between individual speakers. For example, Speaker 5 pointed out that she differentiates between possessive noun phrases to be used in the context of a longer sentence (such as Example 10) and the same phrase being used in isolation to express that an item belongs to the possessor (such as Example 11). In the former, any of the three markers are acceptable to her, but she would not use a nominative marker in the latter. Speaker 2 disagrees and used the nominative version *i Lutan* when asked how she would express "This car is Lutan's" (as in Example 11). In the following, directly elicited NPMs are bolded.

10. (Kasavakan, Speaker 5)
    tu=paliding   **ni**/i/kani   Lutan
    3SG.GEN=car   **GEN**/NOM/OBL.**SG**   NAME
    Lutan's car

11. (Kasavakan, Speaker 5)
    tu=paliding   **ni**/kani/*i   Lutan
    3SG.GEN=car   **GEN**/OBL/*NOM.**SG**   NAME
    This car is Lutan's.

Personal singular non-subject actors also show flexible case marking, although not to the same extent as possessors. We observe a clear preference for nominative markers (two-sided binomial test, nominative vs. non-nominative, elicited only, $N = 18$, $p_0 = 1/3$, $p < 0.001$; see Figure 4). The dataset only contains one elicited sentence with a genitive non-subject actor (Example 12), which was provided by Speaker 5, who also rated the version with the nominative marker as acceptable:

12. (Kasavakan, Speaker 5)
    tu=veray-ay   na   liwu   kani
    3SG.GEN=give-LV   NOM.DEF   gift   OBL.SG
    Lutan   **ni**/i   Pusang
    NAME   **GEN**/NOM.**SG**   NAME
    Pusang gives the gift to Lutan.

4.1.2 Personal Plural Nouns

Personal plural NPMs indicate that the referent of the personal noun includes not only the individual
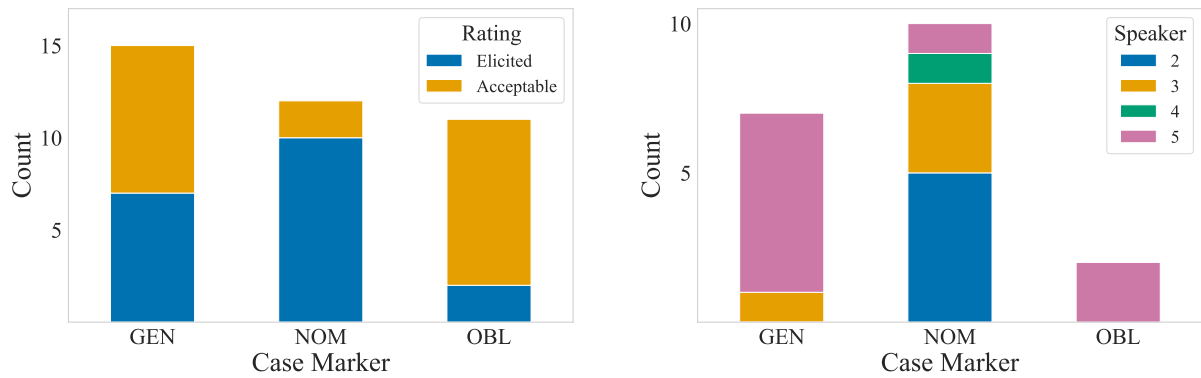
Figure 3: Distribution of NPMs for personal singular possessors, excluding topicalized noun phrases. **Left:** All acceptable sentences, divided by elicited vs. rated acceptable. **Right:** Only directly elicited sentences, divided by speakers.
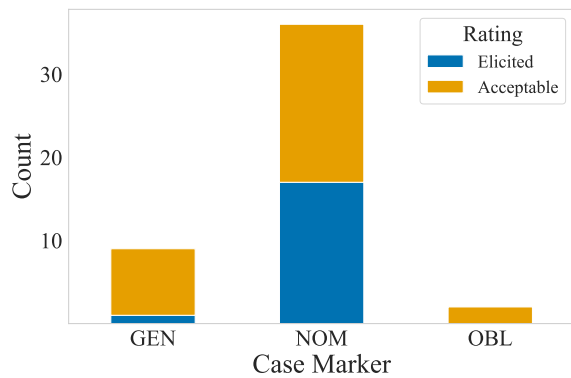

Figure 4: Distribution of NPMs for personal singular non-subject actors, excluding those in sentences where the actor is topicalized.
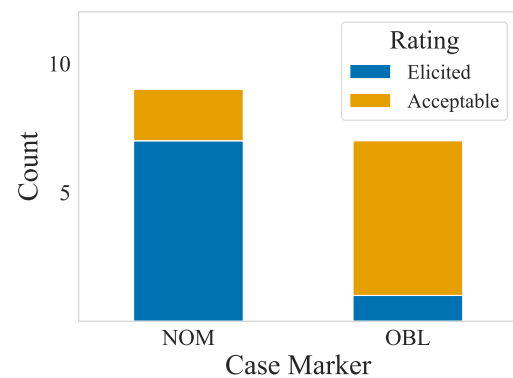
Figure 5: Distribution of NPMs for personal plural possessors, excluding topicalized noun phrases.

denoted by the noun, but also a broader group associated with that person. According to Speaker 3, the group may be their family, their friends or any other group of people they may be associated with.

All plural markers in the dataset were either nominative or oblique. Figure 5 shows the distribution of NPMs for personal plural possessors. Nominative markers are used most often but our data is too sparse to establish a clear preference (two-sided binomial test, nominative vs. oblique, elicited only, $N = 8$, $p_0 = 1/2$, $p = 0.070$). The lone actively used oblique marker (Example 13) was given as an alternative to the nominative marker:

13. (Kasavakan, Speaker 3)

nantu	paliding	**na/kana**
3SG.POSS.NOM	car	**NOM/OBL.PL**

Lutan
NAME
Lutan et al.'s car

As for non-subject actors, the data is even sparser, but both instances of non-topicalized non-subject actors were marked with the nominative marker *na*, as in Example 14:

14. (Kasavakan, Speaker 5)

na	vulraw	tu=akan-aw	**na**
NOM.DEF	fish	3SG.GEN=eat-PV	**NOM.SG**

Umang
NAME
Umang et al. ate fish.

It is difficult to draw any conclusions about the acceptability of the oblique *kana* for non-subject actors as the dataset does not contain any non-topicalized occurrences. In topicalized sentences, we only have two contradicting examples of accepted (not elicited) sentences, Examples 15 and 16:

15. (Kasavakan, Speaker 3)

na/*kana Lutan (mu),
NOM/*OBL.**PL** NAME (TOP)

tu=veranay na kavang
3SG.GEN=gift.LV NOM.DEF clothes

kana lralrak
OBL.DEF children

Lutan et al. give the clothes to the children.

16. (Kasavakan, Speaker 2)

kana Umang mu tu=pa-akan-anay
OBL.PL NAME TOP 3SG.GEN=CAUS-eat-CV

idu na vulraw
that.NOM LNK fish

Umang et al. are feeding that/those fish.

We did not encounter a specific genitive marker in this category, and as Teng (2009) did not include data on a potential personal plural genitive marker in Katipul or Ulivelivek, there was no basis for us to ask about the acceptability of such a hypothetical marker.[3]

4.1.3 Common Definite Nouns

Speakers prefer to mark common definite possessors as nominative (two-sided binomial test, nominative vs. oblique, $N = 12$, $p_0 = 1/2$, $p = 0.006$), but often also accept oblique markers (same test, elicited and accepted combined, $N = 21$, $p = 0.189$; see Figure 6).

Similarly, for non-subject actors, speakers strongly prefer the nominative (same test, elicited only, $N = 17$, $p = 0.013$; see Figure 7). While oblique markers were actively used in three instances, it is worth noting that two of these examples were given by Speaker 5, whose preferences also seem to deviate from the other speakers in other cases (such as for personal singular nouns, see Figure 3).

The most striking finding is that the genitive marker *nina* was universally rejected, indicating that it is not used at all in Kasavakan.

4.1.4 Common Indefinite Nouns

NPMs for common indefinite nouns are some of the most difficult to elicit because Mandarin does not have a definite-indefinite distinction. Indefinite possessors were rated acceptable with both nominative and oblique markers (with two and three exam-

---

[3]The personal plural use of *nina*, as attested for Katipul by Teng (2018), was not known to us at the time of the interviews.
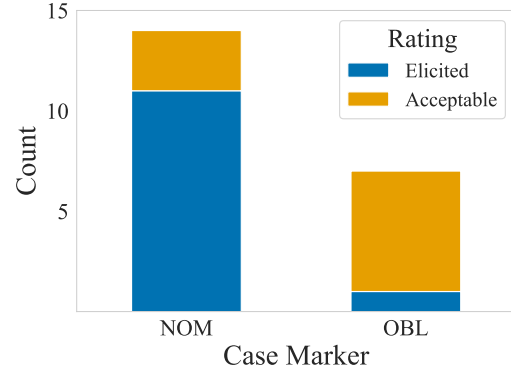

Figure 6: Distribution of NPMs for common definite possessors, excluding topicalized noun phrases.
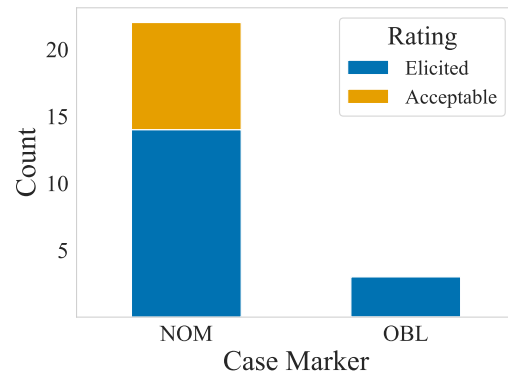

Figure 7: Distribution of NPMs for common definite non-subject actors, excluding those in sentences where the actor is topicalized.

ples, respectively), but there were no instances of indefinite possessive markers in elicited sentences.

For non-subject actors, there was only one case of an indefinite marker being actively used (Example 17), and it was one of several options given:

17. (Kasavakan, Speaker 5)

tu=aday-aw na kuce
3SG.GEN=take-PV NOM.DEF shoes

na/kana/a/da suwan
{**NOM/OBL**}.**DEF**/{**NOM/OBL**}.**INDF** dog

The shoes were moved by the/a dog.

The lack of elicited sentences containing indefinite NPMs makes it difficult to make definitive statements about the preferred marking of common indefinite nouns.

4.2 Disambiguation Strategies

4.2.1 Word Order

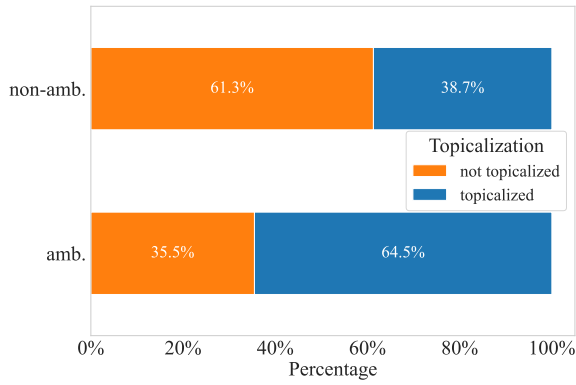Topicalization is the only obvious and consistent disambiguation strategy in our data. An analysis

Figure 8: Topicalization by ambiguity status in elicited undergoer voice sentences (N=93). *amb.* refers to ambiguous and *non-amb.* refers to non-ambiguous.

of the order of non-topicalized noun phrases in elicited ambiguous undergoer voice sentences revealed no significant difference between the orders *verb-undergoer-agent* ($N = 4$) and *verb-agent-undergoer* ($N = 6$). Ambiguity can arise when two noun phrases marked with the same NPM appear on the same side of the verb (see Example 18).[4] Topicalizing the agent resolves the ambiguity (see Example 19).

18. (Kasavakan Speaker 5)
tu=karac-aw	**na**	unan
3SG.GEN=bite-PV	**NOM.DEF**	snake

kana	suwan
OBL.DEF	dog

    The dog bit the snake./The snake bit the dog.

19. (Kasavakan Speaker 3)
na	unan	tu=karac-aw
NOM.DEF	snake	3SG.GEN=bite-PV

na	suwan
NOM.DEF	dog

    The snake bit the dog.

The ratio of topicalized vs. non-topicalized sentences in ambiguous and non-ambiguous sentences is visualized in Figure 8. A Fisher's exact test (Fisher, 1922) revealed that ambiguous sentences were significantly more likely to be topicalized than unambiguous ones (odds ratio = 2.88, $p = 0.026$).

Further evidence that topicalization is used as a disambiguation strategy comes from the fact that in ambiguous and unambiguous sentences alike, it is almost universally the actor (and only rarely the undergoer) that is topicalized. Actor topicalization

---

occurred in 36/40 ambiguous and 10/12 unambiguous topicalized undergoer voice sentences. In both cases, this preference was significant (two-sided binomial tests, $p_0 = 1/2$, $N = 40$, $p < 0.001$; $N = 12$, $p = 0.039$).

4.2.2 Avoiding Undergoer Voice

In some cases where we suggested ambiguous undergoer voice sentences, speakers actively gave alternative sentences in actor voice or passive, such as Examples 20 and 21:

20. (Kasavakan, Speaker 3)
t<em>enged	**i**	Pusang	**kani**
<AV>beat	**NOM.SG**	NAME	**OBL.SG**

Umang
NAME

    Pusang beat Umang.

21. (Kasavakan, Speaker 3)
ini	na	suwan	ki-karac	**kana**
this.NOM	LNK	dog	PASS-bite	**OBL.DEF**

unan
snake

    This dog was bitten by the snake.

In these versions, ambiguity is resolved because the actor is necessarily marked nominative in actor voice and necessarily marked oblique in passive constructions, making it unnecessary to use further disambiguation strategies.

5 Discussion

5.1 Implications

Our data reveals that non-subject actors and possessors tend to be marked nominative in Kasavakan Puyuma, but there are cases where genitive markers are still used, and even oblique markers are frequently found to be acceptable by speakers.

Nominative-genitive syncretism appears to be stronger in Kasavakan than in Ulivelivek and Katipul, as evidenced by the fact that genitive markers seem to not be used at all for common definite nouns. Additionally, since not all speakers use genitive markers actively, the use of genitive markers does not appear to be a preferred disambiguation strategy.

Speakers tend to disambiguate the semantic roles in ambiguous sentences through the use of topicalization. Alternatively, they may avoid using undergoer voice altogether and opt for actor voice or a passive construction instead.

---

[4]This sentence was elicited twice in different sessions for opposite elicitation prompts.

5.2 Potential Reasons for Variation

A potential explanation for the fact that oblique markers are sometimes accepted for non-subject actors may be that the speakers have had contact with other villages. We directly asked Speaker 5 about the use of *ni* (since she prefers to mark personal possessors with a genitive marker, see Figure 3). She mentioned that she was aware that *ni* is commonly used in Ulivelivek. Furthermore, she noted that the marker used to be more common in Kasavakan and is now generally used less frequently by those under the age of 80. If her observation is accurate, it may also be evidence for diachronic nominative-genitive syncretism in Kasavakan – a complete syncretism between nominative and genitive could be a possibility in the future.

Besides nominative-genitive syncretism, there are hints that the distribution of case markers may be even less constrained in Kasavakan. While there is no elicited sentence with the nominative marker *na* in the oblique sense, Speaker 4 accepted such a use in Example 22.[5] In addition, the personal singular nominative marker *i* was marked on the undergoer in a monotransitive causative sentence (see Example 23, compared to Example 24), a position where oblique is expected in the Katipul dialect (see Example 25). Additionally, we also observe possessive and genitive usage of *i* (see Examples 10 and 12, respectively) and *na* (see Examples 26 and 17, respectively).

22. (Kasavakan, Speaker 4)
    na   suwan   '<em>a~'evang
    NOM.DEF   dog   <AV>RED~chase

    na   tutus
    NOM.DEF   mouse
    The dog is chasing the mouse.

23. (Kasavakan, Speaker 1)
    pa-uwa   i   Valraka   t<em>akesi-a
    CAUS-go LOC   USA   <AV>study-PJ

    i   Lutan
    NOM.SG   NAME
    (He) let Lutan study in the US.

24. (Kasavakan, Speaker 1)
    pa-uwa   i   Valraka t<em>akesi-a *kani*
    CAUS-go LOC USA   <AV>study-PJ OBL.SG

    Lutan   i   malri
    NAME   NOM.SG father
    The father let Lutan study in the US.

---
[5]The undergoer in actor voice sentences is usually oblique.

25. (Katipul Teng, 2018)
    pa-uwa=ku   *kana*   alrak   i
    CAUS-go=1SG.NOM   OBL.DEF   child   LOC

    palakuan
    palakuan
    I asked the child to go to the palakuan (adult assembly hall).

26. (Kasavakan Speaker 5)
    tu=sa'ad   **na**   kawi tatelraw
    3SG.GEN=branch   NOM.DEF   tree   long
    The branches of the tree are long.

These phenomena point to the possibility that Kasavakan case markers have become more syncretic and that these markers are gradually losing their case-marking abilities. This would require greater use of additional disambiguation strategies, which may include topicalization, word order and verbal semantics. Further research is needed to confirm the oblique uses of these markers.

5.3 Conclusion

The first goal of this study was to describe the distribution of NPMs in Kasavakan Puyuma and identify patterns of case syncretism. The results, which are based on data collected from five speakers, are shown in Table 1.

Looking at the most preferred markers, nominative, genitive and possessive markers appear to be partially syncretic in the Kasavakan dialect. Common indefinite nouns may be an exception, but our data is inconclusive due to the limited number of such examples. It can be concluded that there is some flexibility in the use of NPMs and that individual preference plays a significant role. Overall, the distribution is closer to the Katipul and Ulivelivek varieties than to the Nanwang variety of Puyuma. While the genitive marker *nina* for common nouns seems to have been lost, some speakers still use the genitive marker *ni* for personal nouns.

The second objective was to identify the preferred disambiguation strategies for sentences where case syncretism causes ambiguities. Topicalization was shown to be a frequent strategy, with the use of actor voice or passive rather than undergoer voice being an alternative. The data was inconclusive on the role of the order of non-topicalized noun phrases in undergoer voice sentences.

In summary, this study provides an insight into the distribution of NPMs in Kasavakan Puyuma, the patterns of case syncretism, and the preferred disambiguation strategies.

Limitations

The results of this study are subject to a number of limitations that need to be addressed.

First, the current study uses sentences elicited from a limited number of speakers, all of whom belong to a very limited age group. This is due to the relatively small number of speakers and the lack of available data. The elicited sentences were designed to answer the research questions. Hence, the dataset may deviate from spontaneous speech.

Second, the English translations in our dataset typically correspond to the original Chinese prompts; however, in some cases where speakers retranslated a sentence they had produced, the translation was adjusted to reflect their retranslation. This cases are not explicitly marked as such in the dataset.

Third, as mentioned in the results section, the dataset has very few examples of some types of NPMs, making some of the entries in Table 1 less conclusive than others. This is in part due to the difficulty of eliciting some of the rarer NPMs.

Fourth, while we report uncorrected p-values here, we note that multiple comparisons were conducted, and results should be interpreted with appropriate caution. Additionally, as all annotations were reviewed collaboratively and ambiguous cases were resolved through discussion among the authors, we were unable to compute formal inter-annotator agreement, which would have yielded an objective estimate of annotation reliability.

Finally, some of our annotations may be open to debate, as we worked with a limited set of semantic roles that occasionally required applying role definitions somewhat broadly.

Acknowledgments

References

Matthew Baerman, Dunstan Brown, and Greville G Corbett. 2005. *The syntax-morphology interface: A study of syncretism*. 109. Cambridge University Press.

Ronald A. Fisher. 1922. On the interpretation of χ^2 from contingency tables, and the calculation of p. *Journal of the royal statistical society*, 85(1):87–94.

Malcolm Ross. 2009. Proto Austronesian verbal morphology: A reappraisal. In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*. Asia-Pacific Linguistics, College of Asia and the Pacific, The Australian National University.

Sidney Siegel and N. John Castellan. 1988. *Nonparametric statistics for the behavioral sciences*, second edition. McGraw–Hill, Inc.

Stacy Fang-ching Teng. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Stacy Fang-ching Teng. 2009. Case syncretism in Puyuma. *Language and Linguistics*, 10(4):819–844.

Stacy Fang-ching Teng. 2018. *Beinan Yu Yufa Gailun*, 2nd edition. Taiwan Nandao Yu Yuyan Congshu. Council of Indigenous Peoples, New Taipei City.

A Abbreviations used in Glossing

Abbreviation	Meaning
1SG	first person singular
3SG	third person singular
AV	actor voice
CAUS	causative
CV	conveyance voice
DEF	definite
GEN	genitive
INDF	indefinite
LNK	linker
LOC	locative (marker)
LV	locative voice
NAME	personal name
NOM	nominative
OBL	oblique
PASS	passive
PJ	projective
PL	plural
POSS	possessive
PV	patient voice
RED	reduplication
SG	singular
TOP	topic

Table 2: Abbreviations used in glossing. With the exception of NAME, all are based on Teng (2008). Some abbreviations were expanded to make the meaning more transparent.