A corpus-based analysis of word order variation in Yami relative clause construction

Hui-Huan Chang and D. Victoria Rau
National Chung Cheng University

Yami relative clauses (RCs) can either precede the head noun, for example, *kanakan* ‘child,’ as in *ko ni-ma-cita o [ji yákneng] a kanakan* ‘I saw the child who cannot hold still,’ functioning as restrictive RCs ([RC] + a + Head NP), or follow it as in *ko ni-ma-cita o kanakan a [ji yákneng]* ‘I saw that child, who cannot hold still,’ functioning as nonrestrictive RCs for complementation strategy (Head NP + a + [RC]). The VARBRUL results demonstrate that head final RCs are predominant in Yami, and Yami speakers use them to connect the given referent with the previous discourse to convey given information. The study found that Subject head nouns outnumber other grammatical roles of head NPs, and that Subject head noun with Subject RC construction is produced more than any other RC constructions, which indicates that Yami RCs are used to modify the Subject for topic continuity.

**Keywords:** Yami, relative clause construction, word order, variation, information flow

1. **Introduction**

The importance of bridging the gap between micro-variation and macro-variation was raised recently by Meyerhoff and Evans (2016). Although it is reasonable to assume that the macro-variants of word order studied by typologists, for example, OV ~ VO, originate as the micro-variation studied by variationists, there has not been enough effort to bring variationist theory to bear on typological theory.

The order of relative clause (RC) is a case in point in the study of Austronesian languages. According to Keenan (1985, p. 144), there is a general tendency across languages to favor postnominal (or head-initial) relative clauses, especially in verb-initial languages such as the Philippine (type) languages. However, some studies report that in some Philippine and Formosan languages the Head NP may either precede or follow its modifying clause (Dixon, 1988; Himmelmann, 2005;
Reid & Liao, 2004). Although there is macro-variation among seven types of relative clause (henceforth RC) among 705 languages, Dryer (2005) finds that the two basic types are RC following the noun (postnominal RC) and RC preceding the noun (prenominal RC); he also claims that prenominal RCs are much rarer than postnominal RCs, especially combining with verb-object RCs. To probe this claim, Comrie (2008) examined grammars, texts and other sources of Formosan and Malayo-Polynesian languages, and found that the macro-variation pattern occurs not only in Amis but also in Pazih (Li & Tsuchida, 2002), the dominant order of which is prenominal, and in Rukai (Zeitoun, 2007), Paiwa, (Tang, 2002), and Tagalog (Schachter & Otanes, 1972), where both orders occur. In addition to the two dominant types – prenominal and postnominal RCs – some scholars claim the heads may also occur within the RCs and thus the construction can be analyzed as head-internal relatives in Tagalog (Aldridge, 2004), Seediq (Aldridge, 2002, 2004), and Squilq Atayal (Liu, 2005, 2015). Therefore, Aldridge (2002, 2004) argues that Seediq exhibits three types of order: postnominal, prenominal, and internally headed RC.

These typologists’ studies have reported the distinctive position of nominal heads in different languages, and describe variation in RC among Austronesian languages. However, the possible pragmatic, discourse, and sociolinguistic motivations behind the word order variation have not been sufficiently explored using variationist methods. Particularly, if the predicate initial Philippine-type languages are found to favor postnominal (or head-initial) relative clauses (Keenan, 1985, p. 144), a fitting example of typologists’ correlations (Greenberg, 1963), what is the motivation for a Philippine-type language that favors prenominal (or head-final) relative clauses, an example that does not fit?

This paper aims to demonstrate that Yami, a Philippine-type language in the Austronesian family, is a case that favors prenominal (or head-final) relative clauses. By using variationist methods to study micro-variation of prenominal and postnominal relative clauses, one can understand the motivation of variation and change, bridging the gap between typology and Rau and Dong’s (2006, p. 125) claim that there is word order variation between head-initial and head-final relative clauses. They showed that when the RC precedes the head noun, it restricts the head noun, as shown in (1a); when the RC follows the head noun, it describes the characteristics of the head noun, as shown in (1b). Throughout this paper, the RC

---

1. The distribution of the 7 types is as follows: RC follows noun (507 languages), RC precedes noun (117 languages), internally headed RC (18 languages), correlative RC (7 languages), adjoined RC (5 languages), double-headed RC (1 language), and mixed types of RC with none dominant (50 languages) (Dryer, 2005).
A corpus-based analysis of word order variation in Yami relative clause construction

is identified by square brackets with the subscript ‘RC,’ and the Head NP is bolded for easy reference.

(1) a. ko ni-ma-cita o [ji yákneng]RC a kanakan.
1.s.gen pa-pf-see nom neg calm lin child
‘I saw the child who cannot hold still.’

b. ko ni-ma-cita o kanakan a [ji yákneng]RC.
1.s.gen pa-pf-see nom child lin neg calm
‘I saw that child, who cannot hold still.’

However, no study has addressed the question whether postnominal or head-initial RCs are the most favored word order in Yami, as in most Philippine-type languages, and what factors best account for the alternation between RCs.

No doubt the most useful way to analyze variable language data and to formulate general principles of linguistic variation is from a variationist approach inseparable from quantification methodology (Milroy, 1987), instead of just focusing on a language specific description or on syntactic construction analysis as previous studies have done. Therefore, in this paper, we present a corpus-based study on word order variation in Yami relative clauses, using Goldvarb X (Sankoff, Tagliamonte, & Smith, 2005), a logistic regression program, to find the most parsimonious model that governs the occurrence of head-initial and head-final relative constructions.

This paper is organized as follows. The introduction is followed by an overview of the Yami speech community and linguistic structures essential to our discussion.

---

2. The abbreviations of the morpheme-by-morpheme glossing are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NF</td>
<td>Nominal affix</td>
</tr>
<tr>
<td>2</td>
<td>NOM</td>
<td>Nominative</td>
</tr>
<tr>
<td>3</td>
<td>OBL</td>
<td>Oblique</td>
</tr>
<tr>
<td>AF</td>
<td>P</td>
<td>Plural</td>
</tr>
<tr>
<td>AUX</td>
<td>PA</td>
<td>Perfective aspect</td>
</tr>
<tr>
<td>CAU</td>
<td>PAR</td>
<td>Particle</td>
</tr>
<tr>
<td>CON</td>
<td>PF</td>
<td>Patient focus</td>
</tr>
<tr>
<td>GEN</td>
<td>PLN</td>
<td>Place name</td>
</tr>
<tr>
<td>H</td>
<td>PN</td>
<td>Personal name</td>
</tr>
<tr>
<td>IF</td>
<td>RED</td>
<td>Reduplication</td>
</tr>
<tr>
<td>INCL</td>
<td>S</td>
<td>Singular</td>
</tr>
<tr>
<td>LOC</td>
<td>SUB</td>
<td>Subjunctive</td>
</tr>
<tr>
<td>LF</td>
<td>SV</td>
<td>Stative verb</td>
</tr>
<tr>
<td>LIN</td>
<td>VF</td>
<td>Verbal affix</td>
</tr>
<tr>
<td>NEG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of relative clauses. In Section 3 the production of RC in discourse is introduced to explain how discourse factors may influence grammatical patterns, followed by a description of grammatical roles in Yami RC construction. The second half of the paper contains the methodology (Section 4), followed by quantitative results, discussion, and conclusion.

2. Yami speech community and its language

In this section, we first describe the genetic position of Yami and provide an overview of Yami clause structure literature.

2.1 An offshore indigenous tribe of Taiwan and its speech community: Yami

The Yami (also known as Tao) are the only Taiwan indigenous group living off Taiwan on Lanyu (Orchid Island), a small offshore island located in the Pacific Ocean 60 kilometers southeast of Taiwan with a population of 3,942 Yami residents (the Council of Indigenous Peoples in Taiwan, 2013). Yami is part of the Philippine language family and is particularly related to Ivatan and Itbayat of Batanes, classified as “Austronesian, Malayo-Polynesian, Philippine, Bashiic, Yami” (Gordon, 2005). Since 1945, Mandarin Chinese has been the national language of Taiwan, ROC, while other languages and dialects were forbidden in public education until 1987. According to Rau (1995), Yami is gradually being replaced by Mandarin Chinese. Iraralay was the only village out of the six on the island where children still used Yami in daily interaction. Comparing the language proficiency, language use and language attitudes among three generations of Yami, Chen (1998) found there was language shift to Mandarin and a decline of Yami language ability among younger speakers. Lin (2007) reexamined language use and language ability among Yami teenagers. The results show that although Yami is still spoken in Iraralay, in the other five villages the use of Yami by teenagers with their parents continued to decline. Moreover, most of the teenagers preferred speaking Mandarin over Yami.

No doubt, Yami is endangered and is rapidly losing child speakers, and will die out in two generations if nothing is done. Since 2005, the second author and her research team have been involved in Yami language documentation and

3. It is well-known that the local people on Orchid Island never called themselves Yami, but identify themselves as pongso no tao ‘people on the island’, and speak circiring no tao ‘human speech.’ In this paper we will use the traditional name Yami, simply because Yami has been commonly used in official documents and academic periodical research.
A corpus-based analysis of word order variation in Yami relative clause construction

conservation, and have published the corpora online. The corpora have served as the basis for writing reference grammars, developing language teaching materials, and building ontology (Chang, Rau, & Dong, 2015a, 2015b, 2015c; Rau & Yang, 2009; Rau, Yang, Chang, & Dong, 2009). The data used in this study are based on the oral corpora of Digital Archiving Yami Language Documentation.

2.2 A brief introduction to Yami clause structure

2.2.1 Yami: A typical Philippine type structure

Yami displays the typical features of a “Philippine-type language” in terms of a unique type of grammatical system known as symmetrical voice (Himmelmann, 2005) or the so-called “focus system”, in which verbal affixes are used to indicate the thematic role of the NP bearing the nominative case in a sentence or the pivot (i.e., topic): Agent Focus (AF) <om>/im-/maN-, Patient Focus (PF) -en, Location Focus (LF) -an, and Instrument Focus (IF) i-, as seen in the following examples:

(2) Focus alternation in Yami

a. Agent Focus (AF; intransitive verb):
   k-om-an so wakay si Salang
   eat Obl sweet.potato Nom PN
   ‘Salang wants to eat a sweet potato.’
   (lit.) The one who wants to eat a sweet potato is Salang’

b. Patient Focus (PF; transitive verb):
   kan-en na ni Salang o wakay
   eat-PF 3.S.GEN GEN PN Nom sweet.potato
   ‘Salang ate the sweet potato.
   (lit.) What Salang ate was the sweet potato’

c. Location Focus (LF; transitive verb):
   ni-akan-an na o mogis ori ni Salang.
   eat-PF 3.S.GEN Nom rice that GEN PN
   ‘Salang ate some rice from there.
   (lit.) What Salang ate a little bit from there was rice’

As (2) shows, the root \textit{kan} ‘eat’ is affixed in four different ways to reflect the semantic role of the Nominal cases (marked with gray boxes): \textit{k-om-an}, \textit{kan-en}, \textit{akan-an}, and \textit{i-akan}. Being an Austronesian language, Yami morphological case marking of NPs include Nominative, Genitive, Locative, and Oblique case markers. A definite NP is usually coded with a nominative or genitive case marker. Yami is morphosyntactically ergative (Rau & Dong, 2006) in that the subject (= pivot) of an intransitive clause (such as \textit{si Salang} in 2a) is marked the same as the object of a transitive clause (= pivot) (such as \textit{o wakay}, \textit{o magis}, and \textit{o among ya} in 2b–d) with the nominative case\(^5\) (\textit{si} for personal names and kinship terms and \textit{o} for common nouns), while the subject of the transitive clause is marked differently, with the genitive case (\textit{ni} for personal names and kinship terms and \textit{na} for common nouns).

With regard to Yami syntax, based on Liao’s (2002) categorization of grammatical roles for Austronesian languages, four grammatical roles can be distinguished to express Yami clause structure with its core arguments:\(^6\) A (transitive subject), O (transitive object), S (intransitive subject), and E (extension to core), with E referring to the second argument of a dyadic intransitive verb, as \textit{so alibang-bang} ‘flying fish’ in (4). The designations of the Core arguments (AOSE) as the required syntactic and semantic functions follows Dixon (1979, 1994).

Clausal construction in Philippine-type languages is typically right branching (predicate-initial): the predicate occurs first, followed by its modifiers. In Yami, verbal clauses are divided into two types: transitive and intransitive. The intransitive constructions contain only one complement (actor) (e.g., (3) with VS structure) marked with the Nominative case, and include stative verbs, involuntary verbs and dynamic verbs with \textit{af} (e.g., \textit{k-om-an} in 2a) or contain double complements consisting of Nominative case and Oblique or Locative cases (see Example 4 with VSE structure). On the other hand, transitive constructions usually have two nominal complements: Agent (actor) marked with the Genitive case and Patient

\(^5\) Following the practices by Liao (2002) and Reid and Liao (2004), we use nominative to refer to the absolutive case in an ergative-absolutive system.

\(^6\) Throughout this paper, A, O, S, and E are used to refer to these four arguments, mentioned here for easy reference.
(undergoer) marked with the Nominative case (see Example 5 with VAS structure). Transitive verbs include PF, LF, IF verbs, as kan-en, akan-an, and i-akan in (2b-2d), potential ma-verbs and involuntary ka-...-an verbs with an expressed actor.

(3) **om-oli** ko simararaw.
    AF-go.home 1.S.NOM noon
    Vintrans S
    ‘I will go home at noon.’

(4) **man-zaneg ka so alibangbang.**
    AF-cook 2.S.NOM OBL flying.fish
    Vintrans S E
    ‘You will cook flying fish. (lit. You are the one who will cook flying fish).’

(5) **ni-i-ka-m’ying no mehakay o mavakes a.**
    PA-IF-VF-laugh GEN man NOM woman PAR
    Vtrans A S
    ‘The man laughed at the woman.’

2.2.2 *Yami RC structure*

Rau and Dong’s two Yami reference grammars (2006, pp. 124–126; 2016) contain detailed descriptions of Yami RC structure. According to Rau and Dong (2006), the basic word order of RCs is postnominal (head-final) RC. Yami RCs are connected to the head nouns by the linker/ligature *a*. The position of Head NP can be either initial as in (6) or final as in (7), and there is a zero pronominal trace (ø) in the RC that refers to the Head NP, as *tazokok* ‘tazokok bird’ in (6), and *wakay* ‘sweet potato’ in (7).

(6) **aro a tazokok a [om-oli ø do ili].**
    Many LIN bird.name LIN AF-go.home LOC village
    ‘(There are) many tazokok birds that went back to the village.’
    (Rau & Dong, 2006, p. 124)

(7) **[ko ni-pangay ø do vanga] a wakay.**
    1.S.GEN PA.PF-put LOC pot LIN sweet.potato
    ‘The sweet potato that I put in the pot.’
    (Rau & Dong, 2006, p. 125)

Rau & Dong (2016) propose that a RC appearing before the modified Head NP (head-final RC) can be seen as a restrictive RC that provides essential old information about the Head NP, as in Example (8), in which the Head NP *meakay isyo* ‘the man’ is the subject of the AF RC (*nimai do jia nokakyab* ‘came here yesterday’). On the other hand, an RC occurring after the modified Head NP (head-initial RC), such as (9), functions like a non-restrictive RC and is used as a complementation strategy...
because *rarakeh ori* ‘that old man’ is modified by a following RC to provide extra information about the man who did the action of *nitomolok sia* ‘poked him’ before.

(8) ko i-ka-kza o [ni-m-ai do jia nokakyab]$_{RC}$ a meakay 1.s.gen if-sv-like NOM PA-AF-come LOC this.LOC yesterday LIN man isyo.
   that(an expression of reminder)
   ‘I like the man who came here yesterday.’  
   (Rau & Dong, 2016)

In Philippine-type languages, the primary strategy for structuring RC is to relativize the nominative noun phrase and to replace it with a gap in the RC; that is, there is a linker/ligature between the head noun in the main clause and the relative clause. Note that the head of the RC is a verbal form (Reid & Liao, 2004). Similar to other Philippine-type languages, Yami relative clauses are connected to the head nouns by the linker/ligature *a*, which introduces the subordinate construction. Recall Examples (1a) and (1b), repeated here as Examples (10) and (11), to illustrate RCs in Yami.

(10) ko ni-ma-cita o [ø ji yákneng]$_{RC}$ a kanakan. 1.s.gen PA-PF-see NOM NEG calm LIN child
   ‘I saw the child who cannot hold still.’

(11) ko ni-ma-cita o kanakan a [ø ji yákneng]$_{RC}$. 1.s.gen PA-PF-see NOM child LIN NEG calm
   ‘I saw that child, who cannot hold still.’

Following the description of Rau and Dong (2006), there is a zero pronominal trace ø in both (10) and (11) RCs that refer to its head noun *kanakan* ‘child.’ The head of the RC is a verbal form *ji yákneng* ‘cannot hold still.’ Comparing these two constructions, the two head nouns in the main clauses are placed in different positions: head-final in (10) and head-initial in (11). Thus, we group these two different orders according to the position of the head noun in the main clause: the head-final group, where the head noun follows the RC (i.e., [RC] + *a* + Head NP) and the head-initial group, in which the head noun precedes the RC (i.e., Head NP + *a* + [RC]). By means of a quantitative analysis we hope to discover functional explanations of word order variation between head-initial and head-final RCs in Yami.
3. Proposed factors determining the variation of Yami RC

To find motivations for the word order variation, this study sought factors from studies in sociolinguistics and functional/typological approaches to language. In studies of English RCs a well-known issue is the investigation of patterns of variation in constraining variant choice – the variation of that, WH and zero relative markers. Some studies have found internal factors such as humanity of antecedents to be a key factor in constraining the variant choice of who in subject position (Guy & Bayley, 1995; Levey, 2006; Quirk, 1957; Tagliamonte, 2002; Tagliamonte, Smith, & Lawrence, 2005). Some also found a relationship between existential sentences and who, although the results have been contradictory. Tagliamonte (2002) found who is disfavored in existentials, while Levey (2006) found who is slightly favored in existentials. Social factors were also considered as constraining factors. D’Arcy and Tagliamonte (2008) found that younger speakers disfavor who. A few studies also reported that gender is at least slightly associated in conditioning variation (Levey, 2006; Tottie & Rey, 1997).

Fox and Thompson (1990, p. 297) identified information flow, which denotes “the interactionally determined choices that speakers make which determine intonational, grammatical, and lexical choices,” as a cluster of key factors to explain the perception and production of English RCs. Information flow, including information status, humanness, definiteness, and grammatical roles, is both cognitive and interactional. Following a functional/typological approach to language, many unexplained grammatical phenomena can be easily captured by certain discourse-level explanations, especially by the notion of information flow (Chafe, 1976, 1987; Du Bois, 1987; Givón, 1983; Prince, 1981). In fact, many cognitive linguists have investigated the complexity of syntactic processing of RCs (e.g., Gibson, 1998; King & Just, 1991; MacWhinney & Pléh, 1988) and proposed animacy relations (Traxler, Morris, & Seely, 2002), the referential status of the nouns in the RC (Warren & Gibson, 2002), and RC structure (Fox & Thompson, 1990; MacWhinney & Pléh, 1988) as cognitive strategies/factors to ease RC processing. As pointed out by Haviland and Clark (1974), the speaker will apply a given/new strategy when understanding an RC sentence in its discourse context to establish coherence, while the listener will try to extract the new information and integrate it with old information already in memory.

To sum up, based on the factors identified in previous literature on sociolinguistics and psycholinguistics, we propose that the factors determining the variation of Yami RC are: information status of Head NP, humanness of Head NP, definiteness of Head NP, grammatical role of Head NP, and grammatical role of NP$_{rel}$ (relativized NP). In the next section, we discuss these factors and illustrate them with Yami examples. The following explanations of the operational definitions of
all the factors on which the coding of our quantitative data was based will pave the way for our research design.

3.1 Information status, definiteness, humanness

The first factor is the information status of the NP in the relative clause. Chafe (1976, 1987) distinguishes the notion of given and new information in terms of consciousness. Given information is “knowledge which the speaker assumes to be in the consciousness of the addressee”, while new information is “what the speaker assumes he is introducing into the addressee’s consciousness by what he says”. Following Chafe (1987) and Du Bois (1980), Fox and Thompson (1990) describe information status as follows: a “given” referent is presumed to be in the focal consciousness of the hearer; a “new” referent is not in the consciousness of the hearer. In other words, a new referent is something mentioned the first time in the discourse, while a given referent has been previously mentioned in the discourse.

In addition to given/new information, an NP can be coded as definite versus indefinite and human versus non-human. A definite Head NP may be modified with a deictic or a possessive pronoun when the speaker assumes that the hearer will be able to identify the referent from a possible range of referents (Chafe, 1976, p. 39).

Applying the above-mentioned definitions, the head-final NP Mikowkow in Example (12) is coded as an old, definite, and human referent because the Head NP is mentioned in the previous discourse.

(12) a ka-tenng-an na si Mikowkow a [ni-mi-ayob ø so ayob PAR VF-know-VF 3.s.gen nom pn LIN PA-AF-wear OBL clothes no longtsang a ka no abtan no longtsang]RC GEN farm LIN CON GEN pants GEN farm

‘He saw Mikowkow, who was wearing farm clothing (i.e., prison inmate clothing).’

In Example (13), the Head NP kanakan ‘child’ is definite because it is followed by a deictic ya. By contrast, the head noun of rarakeh ‘old person’ in Example (14) is indefinite because it is introduced by an oblique case marker so.

(13) no maka-cita sira so asi-asi no kayo-kayo am, ori o when.PA AF-see 3.p.nom OBL RED-fruit GEN RED-tree PAR that nom i-ka-bsoy no [ma-ni-sibo ø ]RC a kanakan ya am, IF-VF-full GEN AF-RED-go_to_the_mountains LIN child this PAR

‘When they saw fruit, those kids who were going into the mountains would pick some to fill their stomachs.’

7. A zero pronominal trace ø indicates the co-referent inside the RC (NPrel).
‘Every time I saw an old man who was carrying a heavy load of food, I would go and help.’

3.2 Grammatical roles in Yami RC

From the perspective of universal typology, Keenan and Comrie (1977) proposed that there is an accessibility hierarchy of RCs, named the Noun Phrase Accessibility Hierarchy: Subject > Direct Object > Indirect Object > Oblique case > Genitive > Object of Comparison (“>” means “is more accessible than”), which stresses the grammatical function of the Head NP and predicts that subject Head NPs are easier to relativize than object Head NPs. Moreover, as Himmelmann (2005) points out, various RC structures are governed by the grammatical role of the NP within the RC; consequently, we predict that the grammatical roles of all four core arguments in Yami play a role in determining the word order between prenominal and postnominal RCs.

As mentioned earlier, we adopted Liao’s (2002) categorization of grammatical roles of the Yami clause, a revised version of Dixon’s Basic Linguistic Theory (Dixon, 1979, 1994; Dixon & Aikhenvald, 2000). As illustrated in Examples (3)–(5), four grammatical roles are distinguished to express Yami RC structure with its arguments: A (transitive subject), O (transitive object), S (intransitive subject), and E (extension to core). Thus seven combinations of Yami RC constructions are distinguished, by which the grammatical role of the Head NP within the main clause, and NP$_{rel}$ (relativized NP), that is, the co-referent inside the RC, were coded. The seven types, as illustrated in (15-21), include A-S, S-S, O-S, E-S, S-O, O-O, and E-O.

The term ‘X-relative’ refers to the role of the NP$_{rel}$. For instance, Subject-relative refers to a RC in which the NP$_{rel}$ is the subject of the RC. ‘A-S’ refers to a type of the RC, in which the Head NP has the grammatical role of A and the NP$_{rel}$ has the grammatical role S. The relative construction in (15) for instance, illustrates A-S construction: the Head NP of the RC in the main clause is a Transitive Subject, while the NP$_{rel}$ in the RC is an Intransitive Subject. The S-S RC construction in (16) indicates the Head NP of the RC in the main clause is an Intransitive Subject, while the NP$_{rel}$ in the RC is also an Intransitive Subject. As mentioned earlier, in this study, examples of RC are given in brackets, the Head NP is in bold,
and NP_{rel} is indicated by a zero pronominal trace \( \emptyset \). The following are examples of Yami relative clauses and their syntactic representations.

(15) **A-S:** Transitive subject Head NP modified by intransitive subject RC

\[
\text{ikongodo} \quad \text{ji} \quad \text{na} \quad \text{ni-ma-ziman} \quad \text{no} \quad [\text{mi-gowgaw} \ \emptyset]_{\text{RC}} \quad \text{a} \quad \text{kanakan} \\
\text{because} \quad \text{NEG} \quad 3.\text{S.GEN} \quad \text{PA-PF-alert} \quad \text{GEN} \quad \text{AF-probe_in_a_hole} \quad \text{LIN} \quad \text{child} \quad \text{ori} \ldots \\
\text{that} \ldots \\
\text{‘The child who was probing the hole didn’t even notice.’}^8
\]

(16) **S-S:** Intransitive subject Head NP modified by intransitive subject RC

\[
\text{ma-niring} \quad \text{o} \quad \text{asa} \quad \text{ka} \quad \text{ra-rakeh} \quad \text{a} \quad [\text{ma-masil} \ \emptyset]_{\text{RC}} \quad \text{ori} \quad \text{am}, \\
\text{AF-speak} \quad \text{NOM} \quad \text{one} \quad \text{CON} \quad \text{RED-old} \quad \text{LIN} \quad \text{AF-fish} \quad \text{that} \quad \text{PAR} \\
\text{‘An old man, who was fishing, said’}^9.
\]

(17) **O-S:** Object Head NP modified by intransitive subject RC

\[
\text{to} \quad \text{ko} \quad \text{a-cita} \quad \text{si} \quad \text{mina} \quad \text{apen} \quad \text{Kalalanet} \quad \text{ito} \quad \text{a} \quad [\text{mi-saboay} \ \emptyset] \\
\text{Aux} \quad 1.\text{S.GEN} \quad \text{PF-see} \quad \text{NOM} \quad \text{late} \quad \text{grandfather} \quad \text{PN} \quad \text{that} \quad \text{LIN} \quad \text{AF-carry} \quad \text{so} \quad \text{cinengeh} \quad \text{na} \quad [\text{RC} \quad \text{am} \\
\text{OBL} \quad \text{construction_material} \quad 3.\text{S.GEN} \quad \text{PAR} \\
\text{‘I saw the late grandfather Kalalanet, who was coming back from cutting wood with a pile of construction material on his back.’}
\]

(18) **E-S:** Extension to core Head NP modified by intransitive subject RC

\[
\text{am-ian} \quad \text{so} \quad \text{ta-tlo} \quad \text{a} \quad \text{pa-pat} \quad \text{a} \quad \text{ka} \quad \text{kanakan} \quad \text{a} \\
\text{AF-have} \quad \text{OBL} \quad \text{RED-three} \quad \text{LIN} \quad \text{RED-four} \quad \text{LIN} \quad \text{CON} \quad \text{child} \quad \text{LIN} \\
[\text{ni-man-mo-mosi} \ \emptyset]_{\text{RC}} \quad \text{dang} \quad \text{am}. \\
\text{PA-AF-RED-catch_bugs} \quad \text{then} \quad \text{PAR} \\
\text{‘There were also three or four children, who were catching bugs.’}
\]

(19) **S-O:** Intransitive subject Head NP modified by object RC

\[
\text{ma-ságpaw} \quad \text{o} \quad \text{ra-rakeh} \quad \text{ori} \quad \text{a} \quad [\text{kaod-en} \quad \text{da} \quad \emptyset], \\
\text{SV-heavy} \quad \text{NOM} \quad \text{RED-old_person} \quad \text{that} \quad \text{LIN} \quad \text{ROW-PF} \quad 3.\text{P.GEN} \\
\text{‘The old man, whom they carried in a rowing boat, was really heavy.’}
\]

---

8. Yami deictics *ya* ‘this (close to me),’ *ori* ‘that (close to you),’ and *ito* ‘far (close to him/her/them)’ are Nominative bound forms occurring after the noun head (Rau & Dong, 2006).

9. In Yami, the AF man- is added to the root *ciring* ‘speak, say’ to form the intransitive verb *maniring* ‘someone said.’ However, like *say* in English, *koan* ‘say, think’ in Yami is usually considered a transitive verb (e.g., *koan na si ina am, “ka mararaken” koan na. “You are very stingy,” he told my mother.’).
A corpus-based analysis of word order variation in Yami relative clause construction

4. Methodology

The goal of this study is to find out what factors account for the alternation between head-initial and head-final RCs in Yami. We mainly intend to address what factors best account for the alternation between head-initial and head-final RCs in terms of information flow. However, we also did an extra analysis to find out whether there is a relationship between the RC constructions and the variation of Yami RC. A multivariate analysis was designed to address the main research question using VARBRUL analysis, and Chi-square tests were conducted to examine the relationship between the grammatical roles of Head NP and the word order variation of RC.

Based on the previous research on RC discussed in Section 3, we hypothesize the following five factor groups can account for the variation of Yami RC: information status of noun phrase (NP), humanness, definiteness, grammatical role of Head NP, and grammatical role of the NP_{ref}. With the help of Goldvarb X (Sankoff et al., 2005), this quantitative study aims to analyze the systematic patterns of word order variation (head-initial and head-final) in Yami relative constructions.
The data comprise 71 texts from the Yami corpora: 20 texts from Rau and Dong (2006), 27 texts from Yami Language Documentation,\(^\text{10}\) and 24 texts from the Yami online Dictionary.\(^\text{11}\)

4.1 Data coding of relative clauses

We follow Dixon’s (2010) four criteria to define the canonical relative construction:

1. The construction involves a main clause (MC) and an RC, making up one sentence which consists of a single unit of intonation.
2. There is one common (shared) argument (CA) of these two clauses.
3. The RC functions as syntactic and semantic modifier of the CA of the MC.
4. The RC must have the basic structure of a clause with at least a predicate required as the core argument.

Based on the descriptions of Yami RCs (Rau & Dong, 2006; 2016), a relative construction in Yami involves two clauses, the relative clause (RC) and the main clause. The common argument (Head NP) in the main clause is connected with the RC by a linker \(a\). The RC modifies the Head NP of the main clause. However, not all constituents connected with the linker \(a\) meet the criteria of RC, so some were excluded from coding. For example, in (22), the LINKER \(a\) connects the auxiliary verb \(oyod\) ‘truly, real’ with the following main verb. In (23), the LINKER \(a\) connects two verbs \(layii\) ‘cry’ and \(omgonagonay\) ‘move’ to form a serial verb construction with the shared Agent \(na\) ‘he’. In addition, in (24), the head noun ‘crab shell’ is not expressed and thus the grammatical role of the Head NP in the main clause cannot be determined, even though its construction can be guessed as (S-O).

(22) \(oyod\ a\ ji\ ko\ a-viay\ ya?\)
real LIN NEG 1.S.NOM AF-alive this
‘Oh, will I really not live?’

(23) \(ori\ to\ na\ laví-i\ a\ om-gona-gonay\ so\ lima\ na\ ori\ am,\)
therefore AUX 3.S.GEN CRY-LF. LIN <AF>RED-MOVE OBL hand 3.S.GEN that PAR
‘He kept crying and tried to pull it out.’

(24) \(aro\ a\ [ni-ni-aka-kan-an\ da]_{RC}\)
many LIN RED-PA-RED-eat-LF 3.P.GEN
‘(There are) many (the crab shells) from the eaten (crab).’

\(^{10}\) http://yamiproject.cs.pu.edu.tw/yami_ch/corpus2.htm

\(^{11}\) http://yamibow.cs.pu.edu.tw/corpus_ch.htm
Notice that Yami RCs with a zero head are very rare, indicating that the head of Yami RC must be overt. However, the data of these narratives, either story-telling or description of speakers’ own life experiences, show that Yami clauses allow their referent to be implicit, such as the unexpressed subject of the intransitive verb (the crab) in the following example:

(25) mi-ratateng am, aráko rana ji rana maka-ngay do vahay na
Af-later_on PAR big already NEG already AF-contain LOC home 3.SGEN
am,
PAR
‘After a while, (the crab) grew too big for its home.’

In total, 193 RCs out of 71 texts were coded as eligible tokens for quantitative analysis of word order variation between head-initial and head-final RCs in Yami. The detailed coding scheme can be found in Appendix.

4.2 Coding reliability

Thirty percent of the texts (i.e., the 20 texts of Rau & Dong, 2006) were coded by both authors. Inter-coder reliability was performed by running Cohen’s Kappa statistics (Cohen, 1960). The results of the coding reliability of the five independent variables are given in Table 1. All Kappa values are higher than 0.80 ($p < 0.001$). According to Landis and Koch (1977), values of Kappa higher than 0.80 (i.e., 0.81–1.00) are considered “almost perfect agreement.” After inter-coder reliability was established, the first author was responsible for the rest of the coding.

Table 1. Kappa statistics for inter-coder reliability on data coding

<table>
<thead>
<tr>
<th>Factor</th>
<th>Value of Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information status of NP</td>
<td>0.94</td>
</tr>
<tr>
<td>Humanness</td>
<td>1.00</td>
</tr>
<tr>
<td>Definiteness</td>
<td>0.82</td>
</tr>
<tr>
<td>Grammatical role of Head NP</td>
<td>0.97</td>
</tr>
<tr>
<td>Grammatical role of NPrel</td>
<td>1.00</td>
</tr>
</tbody>
</table>

5. Results and discussion

The presentation of the results is divided into two parts. As VARBRUL is an exploratory tool based on a logistic regression for variationist linguistics (i.e., working for categorical dependent variables, such as head-final RC vs. head-initial RC),
we show the two stages of VARBRUL runs in Table 2 and Table 3 to seek the most parsimonious model to account for the variation between head-initial and head-final RC variation. In the following sections, we provide answers to the research questions mentioned above by first presenting the distribution of head-initial/final NP in Yami RC, followed by the results of the inferential statistics.

5.1 Information flow factors and variation of Yami RCs

First we come to the question of what factors in terms of information flow best account for the alternation between head-initial and head-final RCs and test if information flow can explain the word order variation.

5.1.1 Occurrence of variation of Yami RC

The raw numbers and percentages of head-initial/final Yami RC constructions indicates that Yami head-final RCs outnumbered head-initial RCs (Table 2).

Table 2. Numbers and percentages of head-initial/final Yami RC

<table>
<thead>
<tr>
<th>Variation of Yami RC</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Head-initial</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Information status of Head NP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New</td>
<td>31</td>
<td>41.9</td>
</tr>
<tr>
<td>Given</td>
<td>35</td>
<td>29.4</td>
</tr>
<tr>
<td>Humanness of Head NP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>22</td>
<td>26.5</td>
</tr>
<tr>
<td>Non-human</td>
<td>44</td>
<td>40.0</td>
</tr>
<tr>
<td>Definiteness of Head NP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definite</td>
<td>36</td>
<td>35.3</td>
</tr>
<tr>
<td>Non-definite</td>
<td>30</td>
<td>33.0</td>
</tr>
<tr>
<td>Grammatical role of Head NP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitive subject (A)</td>
<td>5</td>
<td>26.3</td>
</tr>
<tr>
<td>Intransitive subject (S)</td>
<td>19</td>
<td>25.3</td>
</tr>
<tr>
<td>Transitive object (O)</td>
<td>15</td>
<td>38.5</td>
</tr>
<tr>
<td>Extension to core (E)</td>
<td>27</td>
<td>45.0</td>
</tr>
<tr>
<td>Grammatical role of NP_{rel}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intransitive subject (S)</td>
<td>52</td>
<td>31.1</td>
</tr>
<tr>
<td>Transitive object (O)</td>
<td>14</td>
<td>53.8</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>34.2</td>
</tr>
</tbody>
</table>
Although the results in Table 2 cannot determine which factor is significant to the variation, it can indicate how often each factor occurred with head-initial RCs, or with head-final RCs. As Table 2 shows, the proportion of each factor among five groups of all RCs (Total column) indicates these Yami RCs occur most frequently with given Head NP (61.7%), human Head NP (57%), definite Head NP (52.8%), intransitive subject of Head NP (38.9%), and intransitive subject of NP_rel (86.5%). The factors which the head-final RCs occur more frequently with are almost the same as for all RCs. Comparing across the factors among the five groups, the head-initial RCs can be seen to occur more frequently with new Head NP (41.9%), non-human Head NP (40%), definite Head NP (35.3%), extension to core of Head NP (45%), and transitive object of NP_rel (53.8%).

Furthermore, head-initial tokens constitute 34.2% of the total 193 tokens, while head-final RCs constitute 65.8%, which shows that head-final RCs are the favored position, confirming the claim by Rau & Dong (2006, p. 125) that Yami RC preceding the modified Head NP (i.e., head-final RCs) is the basic (unmarked) construction. Therefore, after the initial VARBRUL run, we decided to treat head-initial constructions as the application (marked) value in VARBRUL analysis to seek the most parsimonious model to account for the occurrence of the head-initial RCs.

### 5.1.2 Factors accounting for head initial RCs

After a binomial up-and-down run for head-initial word order of Yami RC application, the factor groups of definiteness, humanness, and grammatical role of Head NP were eliminated as they did not contribute significantly to the variation. The final inferential statistics from the VARBRUL analysis with the significant factor groups are presented in Table 3. The model presented in Table 3 is considered reliable because the total Chi-square has a value of 7.61, less than the critical value 9.21 (\(df = 2, p < 0.01\)). This indicates the remaining two factor groups are independent of each another. Therefore, we can interpret VARBRUL weights\(^{12}\) (probabilities) in Table 3 to find out the influence of the factors.

The grammatical roles of NP_rel (range = 32) and the information status of Head NP (range = 18) were chosen as the best combination of factor groups to account for word order variation in Yami RC. The fact that the grammatical roles

---

\(^{12}\) There is a standard formula to interpret the VARBRUL weights. For each factor, there is a value/weight (\(P_i\)) ranging between 0 and 1; a value of greater than 0.5 indicates that the factor being considered has a favoring effect on the occurrence of the variant, while a value of less than 0.5 indicates a disfavoring effect. A factor weight of 0.5 is equivalent to no effect. In addition, the effect for the significant factor groups is indicated by the range, which is the difference between the highest and lowest factor weight in the group.
of NP$_{rel}$ is significant confirms the claim by Himmelmann (2005) that different RC structures are governed by the grammatical role of the NP within the RC.

Table 3. Significant factors accounting for production of head-initial RC

<table>
<thead>
<tr>
<th>Factor group</th>
<th>VARBRUL weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information status of Head NP</td>
<td></td>
</tr>
<tr>
<td>New</td>
<td>0.60</td>
</tr>
<tr>
<td>Given</td>
<td>0.42</td>
</tr>
<tr>
<td>Range</td>
<td>18</td>
</tr>
<tr>
<td>Grammatical role of NP$_{rel}$</td>
<td></td>
</tr>
<tr>
<td>Transitive object</td>
<td>0.71</td>
</tr>
<tr>
<td>Intransitive subject</td>
<td>0.39</td>
</tr>
<tr>
<td>Range</td>
<td>32</td>
</tr>
<tr>
<td>Input probability = 0.33</td>
<td></td>
</tr>
<tr>
<td>Total Chi-square = 7.61</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 indicates that a transitive object of the NP$_{rel}$ favors the occurrence of head-initial RCs (0.71). In other words, an object-relative construction is more likely to co-occur with a head-initial RC.

The second most important factor group is information status of the Head NP. A Head NP with new information is found to favor head-initial RCs (0.60). In other words, Yami speakers tend to use head-initial RCs to make the referent of the new Head NP salient, as it has not been mentioned before and is new to the hearer. In addition, the grammatical role of the NP$_{rel}$ also favors Object, as mentioned in the previous paragraph, due to the salience of the head referent: new head-initial referents are not grounded in a previous structure (i.e., old information) but new, as shown in Example (26), when they are uttered in the main clause. Example (26) serves to illustrate a head-initial RC construction. The Head NP *kavakavatanen ko* ‘my own story’ is new to the discourse in the main clause and functions as the transitive object (O) of the RC.

(26) ori rana ya na kavos-an no kava-kavatanen ko a [ya that already aux 3.s.gen end-lf gen red-story 1.s.gen lin aux ko on-onong-an ø ]$_{RC}$
1.s.gen red-state-lf

‘This is the end of my own story, which I stated.’
5.2 Analysis of unmarked Yami RCs: Head-final RC

The results in Table 3 also indicate that intransitive subject of the NP_{rel} and a Head NP with given information favor the occurrence of head-final RCs. In other words, a subject-relative construction is likely to use a restrictive RC (i.e., a head-final RC) to identify the referent of the head noun with given information in Yami. The two RCs in Example (27) serve to demonstrate a head-final Yami RC construction.

In Example (27), the given Head NPs _kanakan_ ‘child’ and _rarakeh_ ‘old person’ are introduced (in 27a) by the previous NPs: _síno ori_ ‘who is that’ and _asa ka tao a ra-rakeh ori_ ‘an elder,’ respectively. In (27b) the given Head NP _kanakan_ ‘child’ in the main clause is an transitive subject (A) ‘The child didn’t even notice the elder,’ and is modified by the RC with a zero trace of an intransitive subject (S) RC [migow-gaw ø] ‘who was probing the hole.’ Similarly, the given Head NP _rarakeh_ ‘old person’ is the transitive object in the main clause (O) ‘The child who was probing the hole didn’t even notice the elder’ in (27c) and is modified by an RC with a zero trace of an intransitive subject (S) [amian ø do likod na ori] ‘who was behind him.’

(27) a. ma-cíta na no [m-angay ø do keysakan]_{RC} a _asa ka_ tao a _ra-rakeh_ ori am, “a _síno ori o ya_ person LIN RED-old_person that PAR PAR who that NOM AUX mi-pi-patotog do Kasngenan a ya mi-gi-gowgaw AF-VF-upside_down LOC PLN LIN AUX AF-RED-probe_in_a_hole ori a, _síno ori?”

b. do ji na ni-ma-ziman no [mi-gowgaw ø]_{RC} a because NEG 3.S.GEN PA-PF-alert GEN AF-probe_in_a_hole LIN kanakan child

c. ori o [am-ian ø do likod na ori]_{RC} a _ra-rakeh_ ori that NOM AF-exist_at LOC back 3.S.GEN that LIN RED-old_person that a, PAR ‘When _an elder_ who happened to walk toward the sea saw (the child), he thought,

“Who is that? Why is he upside-down at Kasngenan with his hand in the hole?”

The child who was probing the hole didn’t even notice _the elder_ who was behind him.’

In other words, Yami RCs favor intransitive Subject RCs to modify a given head-final NP, known to both speaker and listener in a discourse; that is, a given NP
is grounded in the previously mentioned discourse (Givón, 1993), and intransitive Subject RCs are favored to express features of their Subjects what Fox and Thompson call a pivot (Fox & Thompson, 1990, p. 307).

To sum up, these VARBUL results have a plausible functional explanation. According to Givón (1993), a new Head NP RC serves to make the new referent salient (i.e., at the center of attention in the discourse) and cataphorically grounds it to the following discourse, while a given Head NP RC is used to connect the Head NP anaphorically with a prior mental structure. In Yami, the head-final NP is definitely more salient and more topical than the head-initial NP because a given referent has already been mentioned in previous discourse and is well established in the listener’s mental schema. Therefore, Yami speakers tend to use head-final RC to restrict the given referent to the previous discourse, as shown in Example (27) above. It is not surprising for head-final Yami RCs to favor given Head NP, as Subject heads (=pivot) in RCs have a tendency to convey given information (Du Bois, 1987; Fox & Thompson, 1990).

5.3 Grammatical roles of RCs and the variation of Yami RC

To explore whether there is a relationship between grammatical roles of Head NP and the word order variation of RC, we examine the occurrence of grammatical roles of the Head NP and NP_{rel} in Yami. Table 4 shows that the constructions of Extension to core RC (E) and transitive subject RC (A) as NP_{rel} do not occur in the data. This is not surprising as Keenan (1985, p. 156) found that only subjects of

<table>
<thead>
<tr>
<th>Head NP</th>
<th>NP_{rel}</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>O</td>
</tr>
<tr>
<td>A</td>
<td>19</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>9.8%</td>
<td>9.8%</td>
</tr>
<tr>
<td>S</td>
<td>64</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>33.2%</td>
<td>5.7%</td>
</tr>
<tr>
<td>O</td>
<td>36</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>18.7%</td>
<td>1.6%</td>
</tr>
<tr>
<td>E</td>
<td>48</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>24.9%</td>
<td>6.2%</td>
</tr>
<tr>
<td>Total</td>
<td>167</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>86.5%</td>
<td>13.5%</td>
</tr>
</tbody>
</table>
main clauses can be relativized in the Philippine languages, that is, NP_{rel} is always understood to function as the Pivot. As Yami is an ergative language, S and O are marked the same as the pivot of the clause.

In addition, the S-S construction (64 tokens, 33.2%) is produced more frequently than any other Yami RC construction. Among the 64 S-S RCs, 15 of them are existential RCs, as in Examples (28) and (29), constructed with the verb *abo* ‘not exist, absent’ to be Yami negative existentials. On the other hand, Yami positive existential clauses are formed with the verb *mian–amian* ‘exist; have,’ illustrated by (18) above. As existential clauses are used to introduce new information (Givón, 2011) or a sentence-focus structure with new information (Lambrecht, 1994, 2002), this explains why they are found with a relatively higher frequency in narratives. In Yami, an existential clause introducing new information tends to occur with Subject (= pivot) RCs, which makes the new referent salient and grounds a new Head NP (Givón, 1993).

(28) *abo no_exist [ji ani-ma-niahey ø]RC a tao-tao do ka-pongso-an no_exist nom NEG PA-AF-fear LIN RED-people LOC NF-island-NF ta sia ya, 1.p.nom.incl 3.s.obl this ‘The people living on our island were all shivering with fear.’ (lit. ‘there were no people on our island who did not fear it.’)

(29) *ta because if *no mina abo o ra-rakeh* ori a because if PA absent nom RED-old_person that LIN [ni-t-om-olok ø sia]RC am,… PA<AF>poke_bottom 3.s.obl PAR ‘Because if not for the old man who had poked him …’

To test whether there is a relationship between the position of Head NP and Yami RC constructions, the data were analyzed using a Chi-square test. We report the Fisher-Freeman-Halton\(^\text{13}\) \(p\) value \((FI = 17.89; p = 0.004 < 0.05)\), indicating that a significant relationship exists so we can interpret the cell differences by examining

---

\(^{13}\) The Chi-square test assumes that the expected values are greater than one, or no more than 20% of the cells have expected frequencies below five in order to avoid an unreliable Chi-square approximation. When the assumption is not met in the data, Fisher’s Exact (Fisher-Freeman-Halton) Test is recommended instead of the Pearson chi-square to see if there is a relationship between two categorical variables.
the standardized residuals.\textsuperscript{14} In Table 5, one cell has a standardized residual that exceeds the criterion: S-S RCs occurred less than expected in head-initial RCs (standardized residual = −2.1).\textsuperscript{15} It is also observed that the S-S combination occurred frequently in head-final RCs (\(n = 52; 26.9\%\)), although the standardized residual (1.5) did not reach the threshold of 2.0. To conclude, the most frequent S-S relative clause construction favors the head-final position.

### Table 5. Cross-tabulation of word order variation by Yami RC construction

<table>
<thead>
<tr>
<th>Position of head noun</th>
<th>Relative construction</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A-S</td>
<td>S-S</td>
</tr>
<tr>
<td>Head-final clause</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>14</td>
<td>52</td>
</tr>
<tr>
<td>% of total</td>
<td>7.3%</td>
<td>26.9%</td>
</tr>
<tr>
<td>Std. residual</td>
<td>0.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Head-initial clause</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>% of total</td>
<td>2.6%</td>
<td>6.2%</td>
</tr>
<tr>
<td>Std. residual</td>
<td>−0.6</td>
<td>−2.1</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>64</td>
</tr>
<tr>
<td>% of total</td>
<td>9.8%</td>
<td>33.2%</td>
</tr>
</tbody>
</table>

### 6. Conclusion

This paper has examined the variation between head-initial and head-final relative clause constructions in Yami quantitatively, based on a functional approach to word order variation in Austronesian languages (Chang & Rau, 2011; Rau, 2000). After performing a variable rule analysis on 193 RCs extracted from 71 Yami texts, we found that head-final RCs (65.8%) ([RC] + a + Head NP) are produced more frequently than head-initial RCs (34.2%) (Head NP + a + [RC]). The result that head-final RCs are more predominant in Yami is contrary to Keenan’s (1985, p. 144) claim that most of RC formation strategies in Philippine languages are head-initial. This raises an important question for determining the canonical word order in Yami. Perhaps the variation order in Yami is undergoing change in canonical word order, but confirmation of this awaits future investigation.

\textsuperscript{14} Standardized residuals with a positive value mean that the corresponding observed frequency is greater than expected, while negative residuals indicate that the corresponding observed frequency is less than expected. Absolute values of standardized residuals less than 1 are not significant, while those greater than 2 are significant.

\textsuperscript{15} Standardized residuals that meet or exceed the criterion of +/- 2 are indicated in bold.
After examining the occurrence of Yami RCs, we also found that: (1) in Yami RC, subject head nouns outnumber other grammatical roles of Head NP; (2) the S-S construction is produced more frequently than other Yami RC constructions (3) two other types of RC never occur in our data: transitive subject (A) and extension to core (E) relatives. Moreover, an existential clause introducing new information tends to occur with Subject in Yami RCs. To sum up, the results indicate that the primary function of Yami RCs is to modify the topic of the discourse, provide information status (new or given) and function as a semantic distinction between restrictive information and additional information with the use of word order.

One remaining problem we encountered was how to interpret constructions with the linker a. Verbs can be connected with the following main verbs with the linker a to form a Serial Verb Construction (SVC). Thus, in some circumstances, it would be difficult to distinguish a relative clause from a V-a-V construction. Take (30) for example.

(30) kateymárahett no kanakan ya a [mang-alalas ø]RC,
                             CON-very-AF-bad GEN child    this LIN AF-cheat
‘Therefore it is really bad for the child to trick (elders).’ ‘Therefore the child is really bad, who tricks (elders).’

The linker a can be interpreted either as connecting two verbs (kateymárahett and mangalalas), sharing the same Genitive Agent (kanakan ya), or as connecting the modified head noun (kanakan ya) with the RC ([mang-alalas ø]); the Patient (‘elders’) is understood from the context and thus is not expressed. Although polysemy seldom causes problems for communication among language users, it might create problems in analyzing data in cross-linguistic contexts (Ravin & Leacock, 2000). Both SVC and RC constructions can encode the complement clause function (Dixon, 2006). A further detailed study needs to be conducted on the distribution and functions of the morphosyntax of Yami linker a to clearly distinguish a serial verb construction from RC construction. We speculate that SVCs and RCs probably form a continuum, in which an SVC serves as a Verb-like construction, while an RC construction functions as a Nominal construction, as proposed in Rau and Dong (2016).

All in all, this study has provided a systematic approach to examine Yami RCs from a discourse-pragmatic perspective. Although the small number of tokens of RCs extracted from our endangered language documentation corpus was not ideal for a logistic regression, it is the best corpus data available for this language. For future studies, an analysis of the relationship between external factors (e.g., gender, text, and age) and word order variation of Yami RC will be necessary. Thus, we hope this study will pave the way for future study on syntactic variation in other Austronesian languages, the RCs of which may display similar variation.
Acknowledgements

This study was partially supported by an NSC grant (NSC 100-2420-H-194-011-MY3), provided to the second author. A previous version of this paper was presented at the NWAV AP 2012 in Japan. We would like to thank Prof. Laurie Reid for valuable comments.

References


Meyerhoff, Miriam, & Evans, Nicholas (2016). All the way up, and the way down: Bridging micro-variation and macro-variation. Plenary workshop presented at NWAV AP 4, Chiayi, Taiwan: National Chung Cheng University.


**Appendix. Coding sheet for Yami relative constructions**

Position of Head NP

- **i** = head initial (Head NP + a + [RC])
- **f** = head final ([RC] + a + Head NP)

Information status

- **n** = new
- **g** = given

Humanness

- **h** = human
- **n** = nonhuman

Definiteness

- **d** = definite
- **n** = non-definite
Grammatical role of Head NP
A = transitive subject  
S = intransitive subject  
O = transitive object  
E = extension to core

Grammatical role of NP$_{red}$
S = intransitive subject  
O = transitive object

Abstract (Chinese)

達悟語的關係子句 (relative clause; RC) 可以置於所修飾的名詞 (中心語; head noun) 之前 ([關係子句] + 繫詞 + 中心語)，如例句 ko nimacita o [ji yákneng] a kanakan ‘我看見了那個靜不下來的小孩’ 中的關係子句「靜不下來的」前置修飾中心語 kanakan ‘小孩’，作為限定用法 (restrictive)；亦可將關係子句置於中心語之後(中心語 + 繫詞 + [關係子句])，如例句 ko nimacita o kanakan a [ji yákneng] ‘我看見了那個小孩靜不下來’，作為補充修飾中心語用法 (non-restrictive)。根據 VARBRUL 變異分析顯示，中心語後置 (head-final) 是達悟語關係子句最常用的詞序結構，說話者選擇此詞序結構是將已知指涉對象 (given referent) 與先前所提訊息內容作連結。結果也顯示相較於其他語法角色 (grammatical roles)，主語 (Subject) 為最常見的中心語語法角色，而且主語為中心語被主事焦點關係子句修飾是最常出現的結構，這表示達悟語關係子句最主要的功能是用來修飾主語以達到主題連続性 (topic continuity)。

關鍵字: 達悟語, 關係子句結構, 詞序, 變異, 訊息流

Address for correspondence
Hui-Huan Chang
National Chung Cheng University
Chiayi County, Taiwan
ann.ronald@msa.hinet.net