Information structure, prosody and the origins of Formosan relativization

1. Introduction

The purpose of the project was to examine the phenomenon of prenominal relativization attested in Formosan languages (the Austronesian languages spoken in Taiwan), a phenomenon which is unexpected given the verb-initial order of the languages involved and the otherwise fairly robust word order generalizations which state that VO order correlates with the typically head-initial pattern N-REL order, and furthermore that N-REL is dominant as whole, being almost as frequent as REL-N even for OV languages. This generalization is illustrated by data from World Atlas of Language Structures (1).

(1) \[ \text{VO OV} \]

\[ \text{N-Rel} \ 416 \ 113 \]

\[ \text{Rel-N} \ 5 \ 132 \]

Of the five languages which combine VO and REL-N, three are Sinitic (Mandarin, Cantonese and Hakka) one is a close relative of Sinitic (Bai) and one is the Formosan language Amis (with VSO word order). However, Amis is not alone in this respect: REL-N order has been reported in several other verb-initial Formosan languages. The project aims to account for the existence of REL-N in verb-initial languages.

A sample of three languages was chosen, Tgdaya Seediq, Takituduh Bunun and Nanwang Puyuma. These three languages belong to different branches of Austronesian. While Tgdaya Seediq and Takituduh Bunun are fairly close geographically, being spoken within 30 km of one another, Puyuma is spoken in a completely different part of Taiwan, in the south, around Taitung.

All three languages are verb-initial and display at least a certain degree of REL-N order. However, they are maximally different in other ways, as will be seen below. In the
following sections, the relevant background information to the three languages will be reviewed.

2. Surface descriptions
Before departing on descriptions of each of the languages involved, a general overview of some shared properties which deal with voice and relativization must be dealt with.

2.1 Austronesian voice
The Formosan languages, genealogically the highest branches of the Austronesian language family, typically display a voice system sometimes described as the Philippine-type voice system. This is basically a four-voice system, where different marking on the verb reflects whether the grammatical subject prototypically bears the semantic role of Actor, Patient, Location or Instrument, respectively. For the purposes of this description, the “grammatical subject” is defined as an argument which is realized in NOM case, or reflected by a clitic pronoun in NOM case, and which is targeted by certain types of typically subject-oriented mechanisms. It should not be seen as synonymous with the concept of “subject” in e.g., English. In fact, as was noted as early as Schachter (1976), subject properties are split across the semantic agent (or notional subject) and the NOM argument. It is thus not meaningful to discuss which if these corresponds more closely to a “subject” in English. Thus, while the terms Actor and Subject will be used descriptively here to refer to the notional subject and the NOM argument respectively, this is simply a descriptive choice. The four voices are illustrated below in examples from Tgdaya Seediq (2).

2 a) Ga t<m>ekan beras ka Pawan.
    PROG <AV>pound rice NOM Pawan
    ‘Pawan is pounding rice.’

b) Tkan-un=mu ka beras nii.
    pound-PV=1sG NOM rice DEM
    ‘I will pound this rice.’

c) Tkan-an=mu beras ka duhung nii.
    pound-LV=1sG rice NOM mortar DEM
    ‘This mortar is where I pound rice.’

d) S-tekan=mu beras ka seru nii.
    IV-pound=1sG rice NOM pestle DEM
    ‘This pestle is what I pound rice with.’

In the Philippine-type voice system, the major syntactic dichotomy is between Actor Voice (henceforth abbreviated AV) and non-Actor Voice (henceforth abbreviated NAV), where non-Actor subsumes Patient Voice, Locative Voice and Instrument Voice. Thus, while the various non-Actor Voice forms differ in some properties, most notably the semantic role of the grammatical subject, they crucially share the feature that the Actor is realized in an oblique voice, usually Genitive. This dichotomy is increasingly, but not entirely unproblematically, being referred to as a contrast in transitivity, where NAV is identified with a transitive ergative construction, and AV is identified with antipassive
(Aldridge 2004a). Thus, for proponents of this analysis, the Philippine-type Austronesian languages are syntactically ergative. We will not address this debate here, however, since it is not relevant to the project.

One crucial property of Austronesian languages in general, which directly connects the voice system with relativization is the subject-orientation of relativization: relativization can only take place from grammatical subject position. This is the rationale behind the voice system: given that relativization is subject-oriented, a mechanism is required to feed arguments into grammatical subject position, and it is the voice system which fills this role. This is illustrated with examples from Tgdaya Seediq (3).

3 a) gaga inu ka rseno m-n-imah sino?
   LOC where NOM man AF-PST-drink wine
   ‘Where’s the man who drank wine?’

   b) *gaga inu ka sino m-n-imah _ Pawan?
      LOC where NOM wine AF-PST-drink Pawan
      (‘Where’s the wine Pawan drank?’)
      ‘Where’s the wine that drank Pawan?’

   c) gaga inu ka sino n-mah-an na Pawan?
      LOC where NOM wine PST-drink-LF ERG Pawan
      ‘Where’s the wine Pawan drank?’
      (‘Where’s the wine that was drunk by Pawan?’)

With these two major general facts in place, we can proceed with the descriptions of the languages. The surface descriptions given below are the result of sentence elicitation and grammaticality judgments (as well as published sources), and thus can be said to represent the native speaker intuitions of the majority of speakers interviewed. The importance of this caveat will become obvious in later sections.

2.2 Tgdaya Seediq

Seediq word order is robustly subject-final. In AV clauses, ordering is VOS (4a), in NAV clauses, ordering is VAS (4b). Apparent SVO word order often surfaces as well, either as the result of topicalization (4c) or clefting (4d)

4 a) M-n-imah sino macu ka tama=mu.
    AV-PRF-drink wine millet NOM father=3sG
    ‘My father drank millet wine.’

   b) N-mah-an=na tama=mu ka sino kiya.
      PRF-drink-LV=3sG father=3sG NOM wine millet
      ‘My father drank that wine.’

   c) Tama=mu ge, m-imah sino macu (heya).
      father=3sG TOP AV-drink wine millet 3sN
      ‘My father, he drinks millet wine.’

3
Relativization is generally post-nominal (5a). However, some speakers also accept prenominal relativization (5b). In addition, all speakers accept circumnominal relative constructions, where the verb of the relative clause (with clitic pronouns attached) precedes the head noun, while the remaining argument of the relative clause follows it (5c). Interestingly, many, if not most, examples of apparently prenominal relativizations are structurally identical to circumnominal relative clauses, albeit lacking an overt non-subject argument (5d). Since there is no overt relativization morpheme in Tgdaya Seediq, N-Rel and Rel-N are morphologically speaking symmetrical.

5 a) Ni=ku m-enaq sapah s<n>alu=na tama=mu.
   PROG=1sN AV-stay house <PRF.PV>make=3sG father=1sG
   ‘I’m staying in the house my father built.’

b) Ni=ku m-enaq s<n>alu=na tama=mu sapah.
   PROG=1sN AV-stay <PRF.PV>make=3sG father=1sG house
   ‘I’m staying in the house my father built.’

c) Ni=ku m-enaq s<n>alu=na sapah na tama=mu.
   PROG=1sN AV-stay <PRF.PV>make=3sG house GEN father=1sG
   ‘I’m staying in the house my father built.’

d) Ni=ku m-enaq s<n>alu=na sapah.
   PROG=1sN AV-stay <PRF.PV>make=3sG house
   ‘I’m staying in the house he built.’

The empirical facts for prenominal relativization in Seediq are intriguing: while Chang (2000) and Aldridge (2004b) observe that prenominal, postnominal and circumnominal relative clause constructions are grammatically acceptable, this result has only partially been replicated by our own fieldwork: in our sample, two speakers accepted prenominal relative clauses freely, one did not accept them at all, and two accepted them in some cases but not in others. In particular, it turned out that prenominal relative clauses were more readily accepted if the NP was topicalized than if it was realized in an argument position. Thus, (6a) was more generally acceptable than (6b).

6 a) T<n>bugan=na Watan ka huling ka wada=mu bbe-un.
   <PRF>rear=3sG Watan NOM dog NOM PST=1sG beat-PV
   ‘The dog which Watan reared is the one I head.’

b) Wada=mu bbe-un t<n>bug-an=na Watan huling.
   PST=1sG beat-PV <PRF>rear=3sG Watan dog
   ‘I beat the dog which was reared by Watan.’

This mirrors an observation made by Liu (2005) for Jianshi Squiliq Atayal, that while REL-N with a restrictive reading is ungrammatical in an argument position (7a, b), both orders REL-N and N-REL are equally acceptable in a topicalized position (7c, d).
7  a. *?cyux m-ʔuyay qu? [m-n-ihiy yumin ka?] kneril
   AUX AV-hungry NOM AV-PRF-beat Yumin REL woman
   (‘The woman who beat Yumin is hungry.’) (op. cit. p 91.)

   b) cyux m-ʔuyay qu? kneril [ ka? m-n-ihiy yumin ]
      AUX AV-hungry NOM woman [ REL AV-PRF-beat Yumin
      ‘The woman who beat Yumin is hungry.’) (ibid.)

   c) [m-n-ihiy yumin ka?] kneril ga? cyux m-ʔuyay la
      AV-PRF-beat Yumin REL woman TOP AUX AV-hungry PRT
      ‘As for the woman who has beaten Yumin, (she) is hungry.’ (ibid.)

   d) kneril [ka? m-n-ihiy yumin] ga? cyux m-ʔuyay la
      woman REL AV-PRF-beat Yumin TOP AUX AV-hungry PRF
      ‘As for the woman who has beaten Yumin, (she) is hungry.’ (ibid.)

Note that this restriction in Jianshi Squiliq holds for restrictive relativization. We will return to this point in section 4.4, where it will be shown to be of crucial importance. In fact, one of the initial hypotheses underlying the present project is derived from the observation that topicalized position is more liberal when it comes to ordering than argument position. We will use this observation to account for prenominal relativization in Formosan in general. For now, however, suffice it to note that restrictive relativization in an argument position in Jianshi Squiliq Atayal allows N-REL but not REL-N order.

2.3 Takituduh Bunun

Bunun has VSO word order in AV clauses (8a) and VAS word order in NAV clauses (8b). The basic word order generalization in Bunun is therefore V-AGT-PAT, and word order is determined by semantic role rather than surface grammatical role. This fact will prove to be of crucial importance.

8  a) Ma-baliv ca tama-nak is laihli.
     AV-buy NOM father-1sG OBL car
     ‘My father bought a car.’

     b) B<in>aliv is tama ca acu dii.
        <PRF.PV>buy OBL father NOM dog DEM
        ‘This dog was bought by my father.’

However, in our data, we also found a substantial number of VOS examples (9a, b), in particular in cases when the object is indefinite.

9  a) Ma-havu-s qanvang ca binanaauaz.
     AV-drive-OBL buffalo NOM woman
     ‘The woman is driving a buffalo.’
Further, SVO surface order is very frequent, both as the result of topicalization (10a) and clefting (10b).

10 a) Maca binauaz-dii-a, maludaq-i qanvang.
    TOP woman-DEM-TOP AV-beat-OBL buffalo
    ‘This woman, she is beating a buffalo.’

b) Bananaz ca angkuc paan.
    man NOM hold bottle
    ‘It is a man that is holding a bottle.’

Bunun relative clauses are generally described as being predominantly prenominal, at any rate when the desired reading is restrictive (11a), dated cited for Isbukun Bunun from Zeitoun (2000b:100). Non-restrictive relativization is generally postnominal (11b). Note that while REL-N constructions typically bear the overt relativization marker a (corresponding to tu in Isbukun Bunun), this never occurs in N-REL constructions, which are completely unmarked.

11 a) 'adi’ bunun-a hai’ sadu’ kalat-un 'asu’ tu pinaz
    DEM person-NOM TOP see bite-PV dog REL woman
    ‘That person sees the woman who is bitten by a dog.’

b) haiza tatini binanauaz ma-lilikuc-is luqic
    EX single woman AV-hold-OBL stick
    ‘There is a woman holding a stick.’

To summarize, restrictivity appears to be an important factor for relativization ordering in Bunun. Importantly, it appears that the Bunun connection between restrictivity and REL-N and non-restrictivity and N-REL is the reverse of the pattern we found for Jianshi Squliq and (some speakers of) Tgdaya Seediq.

2.4 Nanwang Puyuma

Puyuma is traditionally described as alternating VOS/VSO: both orders are grammatical (12a, 12b). The grammaticality of both orders was confirmed by our own fieldwork. Nevertheless, in elicitation tasks, VOS dominates more or less to the exclusion of VSO. Further SVO is extremely frequent, both as the result of topicalization (12c) and clefting (12d). In fact, sometimes SVO occurs in cases where it is not immediately linearly obvious whether we are dealing with clefting or topicalization (12e). Prosodically and information-structurally, however, (12e) can be shown to be equivalent to a cleft. It is felicitous as a reply to a subject wh-question and it bears a typical intonational contour with a H tone on the subject, cf. section 7.

12 a) s<em>arem dra bunga i nama=li
    <AV>grow OBJ sweet.potato DET.FAM father=1sG
    ‘My father grows sweet potatoes.’
b) s<em>arem i nama-li dra bunga
  &lt;AV&gt;grow DET.ANIM father=1sG OBJ sweet.potato
  ‘My father grows sweet potatoes.’

c) na ma’inayan i, ma-’ekan dra belbel
  DET man TOP AV-eat OBJ banana
  ‘The man, he is drinking wine.’

d) amau i nama-li na auka i balrangaw
  COP DET.FAM father=1sG DET go to Taitung
  ‘It’s my father who is going to Taitung.’

e) na balaka tr<em>ima dra palriding
  DET foreigner <AV>buy OBJ vehicle
  ‘The foreigner bought a bike.’

Puyuma relativization allows both N-REL (13a) and REL-N (13b), as well as circumnominal relativization constructions (13c, d). The circumnominal relativization constructions are described as *headless relative clauses* by Huang (2000c:179) and analysed as *internally headed relative clauses* by Aldridge (2004b). Teng (2008) does not mention this construction as a possibility, and in fact almost explicitly excludes them from her description, stating that “Puyuma has postnominal external and prenominal externals RCs”. We will discuss the status of these clauses in section 5.7.

13  a) malriay idru na balaka na tr<em>ekelr dra eraw
  drunk DEM DET foreigner DET &lt;AV&gt;drink OBJ wine
  ‘The foreigner that drank wine got drunk.’

b) malriay idru na tremekelr dra eraw na balaka
  drunk DEM DET &lt;AV&gt;drink OBJ wine DET foreigner
  ‘The foreigner that drank wine got drunk.’

c) me-nau=ku dra penuakpuk dra traw dra walak
  AV-see=1sN OBJ hit OBJ man OBJ child
  ‘I saw a person who hit a child.’ (Huang 2000c:179)

d) matia=ku dra me-kan=ku
  AV-dream=1sN OBJ AV-eat=1sN
  &lt;PV.PRF&gt;cook tasty OBJ corn DEM aunt Muya.
  ‘I dreamed I ate the tasty corn cooked by aunt Muya.’ (Huang 2000c:179)

The contrast between prenominal and postnominal relativization constructions can be described in terms of restrictivity: thus, while restrictive relativization constructions allow both REL-N and N-REL, non-restrictive relativization constructions only allow N-REL. This fits the same pattern as Bunun and Tsou (and, as we shall see, reflects a common pattern across Formosan languages, with the exception of Atayalic languages).
2.4 Background summary

The background information serving to formulate the problem is based primarily on grammaticality judgment tasks. Thus, it described potentially grammatical constructions in the languages involved, but does not address the issue of when and if these different constructions are actually used. As we shall see from what follows, there are rather large discrepancies in all three languages between these two viewpoints.

3. Previous accounts

3.1 Rel-N is not a calque

Before outlining hypotheses which have been presented to date, our first step must be to eliminate the possibly most obvious candidate: external influence. During the last few hundred years, the majority population of Taiwan has been Taiwan Hokkien (except in some areas, e.g. near Miaoli and Hsinchu, with a substantial Hakka population). Both Hokkien and Hakka share the typical Sinitic REL-N order also found in Mandarin Chinese. Between 1895 and 1945 Taiwan was part of the Japanese Empire and it was during this period that most Formosan populations became bilingual, as the result of universal schooling with Japanese as the medium of education. Given that Japanese is a REL-N language, this might also be viewed as a potential source of influence. However, there are several reasons to discount this as a possible explanation.

Firstly, we have no evidence that Formosan languages have been heavily influenced by Sinitic or by Japanese (other than modern loanwords). If Formosan languages had been influenced by either Sinitic or Japanese, the most obvious feature to borrow would have been SVO order, which is typologically less marked than the extant VOS or VSO, and which is present as an alternative topicalization order in most of these languages anyway. Preserving verb-initial order but borrowing the highly marked REL-N order is a very unlikely scenario.

Secondly, if REL-N were a calque from Sinitic or Japanese, we would expect to find it especially robustly in those few Formosan languages which have developed SVO order, presumably under Sinitic influence. However, in one of these languages, Thao, both N-REL (14a) and REL-N (14b) appear to cooccur. While not mentioned explicitly in either ooof the sources, the examples suggest that the relevant difference is one of restrictivity (cf. section 3.2). However, the distribution in Thao suggests that this language, despite its SVO clause-level word order, is no different from other Formosan languages like Bunun when it comes to relativization order.

14 a) yaku’ myaran m-angqtuqtu
    1s often AV-think
    nak a’ azazak i tuLi m-acupiS pataSan
    1sG LIG child LOC Kaohsiung AV-study book
‘I often think of my child who is studying in Kaohsiung. (Huang 2000a: 132)

b) sa pinacay ihu a shput
    TOP hit 2s REL person
‘the person who hit you’ (Blust 2003:227)

Thirdly, if it were a calque from either Sinitic or Japanese, we would expect it to have developed fairly recently. While the Formosan peoples have shared Taiwan with Sinitic populations for the past 300 years, there was scarcely any widespread contact between
most aboriginal tribes and outside society until after the advent of the Japanese in 1895. Nevertheless, REL-N constructions are found in the very earliest sources, presumably before bilingualism could have had any noticeable effect on language structure (15a, b).

15 a) mastaan madain to bunun \textit{Bunun} \\
very big REL person \\
’a very big person’ (Ogawa & Asai 1935: 647)

b) o maJa-foʔis ?a majakakai \textit{Amis} \\
? INCH-star REL sibling-NMLZ \\
‘The siblings who became stars.’ (op.cit: 552)

Fourthly, REL-N / N-REL alternation is found also in, e.g. Tagalog, which is outside the influence of Japanese and Sinitic (16a, b).

16 a) masarap ang pagkai-ng [ni-luto mo] \textit{Tagalog} \\
tasty NOM food-LIG PV-cook 2sG \\
‘The food you cooked is tasty.’ (Schachter 1972:124)

b) masarap ang [ni-luto mo-ng] pagkain \\
tasty NOM PV-cook 2sG-LIG food \\
‘The food you cooked is tasty.’ (ibid.)

Thus, REL-N as a possible order is attested throughout the highest branches of Austronesian, and can hardly be analysed as anything other than an original property of the earliest stages of the language family. Thus, the issue must be investigated from the internal point of view of Austronesian. How, then, can we reconcile verb-initial order with REL-N?

3.2 The flat structure account

Tang (2008) observes that in Austronesian relativization constructions, the relative clause on its own can serve as an independent noun phrase (17a), on a par with the head noun (17b), and that an overt relativization construction is indistinguishable from an apposition of these two elements (17c). The facts are illustrated here with Seediq data. Thus, it is perhaps more apt to refer to this construction as clausal nominalization rather than relativization.

17 a) m-n-ekan=ku b-n-rig-an=su \\
AF-PST-eat=1sN <PST>-buy-LF=2sE \\
‘I ate what you bought.’

b) m-n-ekan=ku bunga \\
AF-PST-eat=1sN sweet.potato \\
‘I ate sweet potatoes.’

c) m-n-ekan=ku bunga b-n-rig-an=su \\
AF-PST-eat=1sN sweet.potato <PST>-buy-LF=2sE \\
‘I ate the sweet potatoes that you bought.’
This view is developed for Puyuma in Teng (2008:80ff), who proposes the following structure for a relativization construction, where an NP is built up of several elements of the form np, which can be a noun, an adjective, or a demonstrative (18).

Independent evidence for this analysis comes from the fact that each element within the complex NP is case-marked independently (19a), and could act as an NP on its own. Further, not only are ADJ and N freely ordered with respect to each other, but two adjectives within the same NP can actually be realized on either side of the N (19b).

In Holmer & Karlsson (2012) we argue that Teng’s structure should be revised slightly, in that DEM does not appear to be an np element of the same category as ADJ, N or REL. For one thing, DEM is always initial within an NP (cf. Teng 2008:82). Furthermore, doubling DEM actually has yet another effect, in that it permits Puyuma NPs to override the restrictivity constraint on ordering (20).

Thus, while restrictive relative clauses can be either N-REL och REL-N, non-restrictive relative clauses can only have the order N-REL. Relativization with a proper noun, which is non-restrictive by definition, is ungrammatical with REL-N order (21a). However, this ordering is grammatical if both elements are marked with DEM (21b, c).

---

18) 

\[ NP \]

\[ np \]

Dem \[ nadrunu \]

those

Nmk\[r\]

XP

Nmk\[r\]

XP

\[ malruwadi \]

sibling

\[ ma’inayan \]

male

19) a) \[ qa \] \[ qa ]\[ suan \] \[ qa] saygu \[ me-lulu’ \]

OBJ other OBJ dog OBJ can

ITR-hunt

‘other dogs that can hunt’ (Tang 2008:938)

b) \[ idri \] \[ na\]

matrina \[ na \]

kaagungan \[ na \]

utreutrem i

DEM DET big DET buffalo DET black TOP

‘As for the big black buffalo...’

20. Restrictive: N-REL, REL-N

Non-restrictive: N-REL, *REL-N

21) a) \[ *malriay \] \[ na \]

tr<em>ekeir \]

drunk DET <AV>drink OBJ wine DET.FAM Sigimulri

(‘Sigimulri, who drank wine, got drunk’)

b) \[ idru \] \[ na \]

saigu s<em>enay idri \]

Samingad i, kaadru i Puyuma that DET able <AV>sing this Samingad TOP live LOC Nanwang

‘Samingad, who is good at singing (lives in Nanwang)’
c) idru na sager dra eraw idri Miduay i, s<em>a</em>-senay la
that DET like OBJ wine this Miduay TOP <AV>-RED-sing PRF
‘Miduay who likes wine, started singing’

Thus, it is rather the presence of DEM which makes an NP independent enough to be able to occur in apposition constructions entirely free of ordering restrictions (in this respect, DEM serves the same function as the determiner in e.g., the REL-N language Basque (22a), which allows a relative clause to occur in postnominal position (22b) or to occur as an argument on its own (22c).

22 a) erosì ditut Itziar-rek idatzi dituen liburu-ak
buy AUX.3pA.1sE Itziar-ERG write AUX.3pA.3sE.REL book-PL.ABS
‘I bought the books that Itziar wrote.’

b) erosì ditut liburuak, Itziarrek idatzi dituen-ak
buy AUX.3pA.1sE book-PL.ABS Itziar-ERG write AUX.3pA.3sE.REL-PL.ABS
‘I bought the books, those that Itziar wrote.’

c) erosì ditut Itziar-rek idatzi dituen-ak
buy AUX.3pA.1sE Itziar-ERG written AUX.3pA.3sE.REL-PL.ABS
‘I bought those that Itziar wrote.’

While the detailed issue of which level is relevant for apposition is still open to debate, it still seems clear, in particular in view of the fact that headless relativizations are structurally identical to apparently headed relativizations, that a flat structure is the best approximation of what is actually going on in an Austronesian relative clause construction. At the same time, it raises important questions:

a) while it does allow the issue of ordering to be separated from headedness, in that V-initial order need not strictly predict N-REL ordering, it does not really explain why REL-N actually occurs;

b) if a flat structure can account for REL-N in the Formosan languages, it is surprising that other head-initial languages are strictly N-REL; is a flat structure only an available option in Austronesian?

Tang (2008:953ff.) offers three potential explanations a) initial case-marking, b) external influence, and c) restrictivity. These three accounts will be addressed in turn.

In a flat structure, there can still be certain word order preferences, e.g. Case-first, and if a case-marked demonstrative is merged as the obligatory first NP1 of a complex flat NP structure, the resulting collocation obligatorily has the form DEM-N (as is the case in, e.g. Amis, Rukai, Puyuma and Paiwan). If the placement of other modifiers is generalized to that of DEM, this would lead to a more general pattern Modifier-Head, which would include ADJ-N and REL-N as possible instances.

This argument creates some serious problems, however. Firstly, it is not clear why DEM-N should generalize to attract REL leftwards (there are numerous examples, such as English, Spanish, German and Russian, in which this does not happen). Secondly, it is not clear how REL-N order could develop in an N-DEM language such as Takituduh Bunun (23a–d). Thus, this account is not a very convincing explanation.
23 a) b-in-aliv is tama ca acu dii
   <PV.PRF>buy OBL father NOM dog DEM
   ‘This dog was bought by my father.’

   b) madangqac ca puwaq dii
       red NOM flower DEM
   ‘This flower is red.’

   c) panak-a-s bunun aiza ca tama-k-inak
       hit-LV-OBL person that NOM father-LNK-1sG
   ‘My father was hit by that person.’

   d) Cima ma-ludaq-is uva’az-un?
       who AV-hit-OBL child-DEM
   ‘Who hit the child?’

The second possible explanation presented by Tang capitalizes on the fact that some Formosan languages have developed SVO word order, presumably under influence from Sinitic. Examples of such languages are Saisiat and Thao. Given that these languages have shifted from verb-initial to subject-initial, they are less likely to display typical verb-initial properties.

However, there are two problems with this approach. Firstly, from a typological point of view, SVO also implies N-REL, everywhere except in Sinitic. Thus, a shift to SVO word order could hardly have caused the development of REL-N by means of language internal pressure, although both REL-N and SVO together might be the result of massive external influence. Secondly, as we have noted above, the SVO language Thao allows N-REL word order (24), which is not predicted by this account, although it is typologically expected otherwise.

24. yaku’ myaran m-angqtuqtu
    1s often AV-think
    nak a’ azazak i tuLi m-acupiS pataSan
    1sG LIG child LOC Kaohsiung AF-study book
    ‘I often think of my child who is studying in Kaohsiung.’ (Huang 2000a: 132)

The most promising account builds on the observation that relativization order is closely aligned with restrictivity in several Formosan languages, as well as the geographically close but non-Formosan language Tau /Yami on Orchid Island (25a, b).

25 a) ko ni-ma-cita o [ji á-kneng] a kanakan
    1s PST-able.PF-see NOM NEG SUB-still REL child
    ‘I saw the child who cannot hold still.’ (Rau & Dong 2006:125)

   b) ko ni-ma-cita o kanakan a [ji á-kneng]
       1s PST-able.PF-see NOM child REL NEG SUB-still
       ‘I saw that child, who cannot hold still.’ (op. cit.126)
The restrictive vs. non-restrictive readings become clearer when the relevant examples are taken in context: examples (26a) and (26b) are consecutive in a Tsou narrative.

26 a) Ø-o-si Ø-li-a na ino no cmoi, UF-PST-3 UF-meet-UF DET mother GEN bear 
ci m-o ea-oko-a REL AF-PST have-child-UF 
“He met with a bear mother, who had a cub.”

26 b) aṃi na’no pak’i really very angry 
na m-o ea-oko-a ci cmoi DET AF-PST have-child-UF REL bear 
’Bears who have cubs are really very irritable.’

If the head noun of the construction is a proper name, the only possible reading is non-restrictive, making personal names the safest test case. The restrictive / non-restrictive contrast is illustrated below in two minimal pairs from Amis (27a, b) and Tsou (27c, d).

27 a) maulah ci aki tura like NOM Aki DEM 
[[maro’ay i fiyaw nira] a fafahiyan] live-NML LNK next.door 3sg REL woman 
‘Aki likes the woman who lives next door.’ (Wu 2000:151)

27 b) maulah ci aki tura ci-panay-an like NOM Aki DEM OBJ-Panay-OBJ 
satu [[mao’ay i fiyaw nira] DEM live-NML LNK next.door 3sg 
‘Aki likes Panay, who lives next door.’ (op.cit. 155)

27 c) o’-u-cu aiti PF-1s-PRF see.PV 
[[t’o o-si tposi to pasuya ci] tposi] NOM PF-3s write.PF by Pasuya REL book 
‘I saw the book that Pasuya wrote.’ (Chang 1997, cited from Zeitoun 2000a:135)

27 d) o’-u-cu aiti PF-1s-PRF see.PV 
[tposi ci o o-si tposi to pasuya] book REL NOM PF-3s write.PF by Pasuya 
‘I saw that book, it was written by Pasuya.’ (op.cit.:136)

This fact holds not only for Formosan languages, but appears to be a more general principle which is found in e.g. Tibetan (28a, b) and Mandarin (28c, d).

28 a) [tshong-khang nang las-ka byed-mkhan phrugu] cig yod-pa red store in work do-NMLZ guy a exist PRF 
‘There was a guy who worked in the store.’ (De Lancey 1999:244)
b) mgokhrid de sngon-la bod-la rgyami ma-yong gong-la
   chief DET formerly-LOC Tibet-LOC Chinese NEG-come before-LOC
   sprangpo [slong-kag ‘gro-mkan] de’dra yin-pa red
   beggar beg-PURP go-NMLZ thus be-PRF
   ‘The chief, before the Chinese came, had been a beggar,
   one who went around to beg.’ (ibid.)

c) [ he-jiu de ] neige ren shi wo baba
   drink-wine MOD DEM person COP 1s Dad
   ‘The person who is drinking wine is my Dad.’

d) zai yangtai you ren [ zai he-jiu ]
   LOC terrace EX person PROG drink-wine
   ‘On the terrace there is someone drinking wine.’

In Holmer (2011) the relevance of restrictivity is summarized as follows, on the basis of published descriptions of the languages involved: while there is no one-to-one mapping between restrictivity and ordering, there is a hierarchical gradient whereby restrictive relativization is preferably prenominal, and non-restrictive preferably postnominal.

29) PRENOM.REL - N - POSTNOM.REL Source

Tsou  RESTR ------>| Tang 2008:956
Amis  ------|<------ NON.RESTR Tang 2008:956
Paiwan RESTR ------------------------>| Tang 2008:956
Puyuma |<------ NON.RESTR Tang 2008:956
Bunun  RESTR ------>| Zeitoun 2000b:100
      |<--------------------- NON.RESTR

Rukai  RESTR ------------------------>| Zeitoun 2000c:123
      |<--------------------- NON.RESTR

Atayal  RESTR ------------------------>| Tang 2008:956
    |<--------------------- NON.RESTR

The generalization seems to hold for all the languages on the list, although the correlation appears to be trivial in Atayal, given that both orders are available for both readings. In fact, the situation in Atayal is even more complex, because a more detailed examination of the Jianshi dialect of Sqliq Atayal suggests that the restrictivity pattern for relativization may the reverse of that found in other Formosan languages (Liu 2005). Thus, for arguments which are realized in argument positions, as opposed to topicalized position, while non-restrictive relative clauses can appear in either order (30a, b), restrictive relative clauses can only surface in N-REL order (30c, d).
Thus, as far as restrictivity is concerned, Jianshi Squliq displays the reverse pattern of that found in, e.g., Paiwan. We will address the issue of Jianshi Squliq in section 4.4. For now, we will attempt an account of the majority pattern in Formosan.

Assuming that there is a shared underlying relativization position with the NP (in itself not an unproblematic assumption), the restrictive - non-restrictive contrast could conceivably be derived by some kind of movement: either leftward movement of a stressed or focused (and therefore restrictive) element, or rightward movement of a non-restrictive element. Given that there is more cross-linguistic evidence for leftward movement in general, and given that some syntacticians, e.g., Kayne (1994), even exclude the theoretical possibility of rightward movement, we will first examine the leftward movement scenario.

If we assume leftward movement of a stressed constituent, the next question is what it is that moves leftwards. Here, Tgdaya Seediq opens up for us an interesting avenue to explore: the circumnominal relativization construction (henceforth CNR). The structure of CNR constructions in both Seediq and Puyuma is fixed: preceding the head noun is the verb of the relative clause (predominantly, and in Seediq, exclusively in Patient Voice), and following the head noun is the remainder or the relative clause (usually only the ergative agent). The surface structure is as in (31a) and two examples from Tgdaya Seediq are given in (31b, c).

31  a) [ V – ..... ]       N       [ .... AGT ]

  b) Wada=mu gguy-un ka [l<n>amu=na] nasi [na baki].
     PST=1sE steal-PV NOM <PV.PRF>-pick=3sE pear ERG old.man
     ‘I stole the pears picked by the old man.’

  c) Wada=mu sbet-un ka [t-n-bug-an=na] huling [na Watan].
     AUX=1s beat-PV NOM <PST>rear-LV=3s dog GEN Watan
     ‘I beat the dog which Watan bred.’

The CNR construction, which is accepted by all speakers, including those for whom REL-N is substandard, can be pretheoretically derived by leftward movement of the verb past the head noun, stranding the remainder of the relative clause in postnominal position.
We are not concerned here with the exact mechanism by which this movement takes place, and are aware that if the relative clause is part of an NP of which the head noun is the structural head, syntactic movement may constitute a breach of Travis’ (1984) Head Movement Constraint.

The interesting point comes when we look at reduced forms of the above construction, which simply involve the leftward movement of the verb and its clitic, without the overt realization of the agent (32a). Given that these constructions are more frequent in Seediq narrative than actual circumnominal constructions with overt agents, it follows that it is not entirely clear what (32a) is a reduction of: is it a reduction of a circumnominal (32b) or of a prenominal (32c)?

32 a. Wada=mu g guy-un ka [l\<n\>amu=na] nasi. PST=1SE steal-PV NOM <PV.PRF>pick=3SG pear ‘I stole the pears which he picked.’

b. Wada=mu g guy-un ka [[[l\<n\>amu=na] nasi [na baki]]. PST=1SE steal-PV NOM <PV.PRF>pick=3SG pear GEN old.man ‘I stole the pears that the old man picked.’

c. Wada=mu g guy-un ka [[[l\<n\>amu=na baki] nasi]. PST=1SE steal-PV NOM <PV.PRF>pick=3SG old.man pear ‘I stole the pears that the old man picked.’

A situation like this could easily lead to the reanalysis of CNR constructions as prenominal relative clauses. If this reanalysis has taken place, it would follow that prenominals are simply the result of pied-piping of the agent with the verb when it is moved past the noun. This could possibly explain the occurrence of REL-N in Seediq, and fits well with the fact that CNR constructions are generally better than REL-N.

There are two problems with this analysis, however. Firstly, it does not carry over to most other Formosan languages, where CNR is not attested. In fact, it is even doubtful whether it carries over Puyuma, where CNR does occur, but are significantly less frequent than REL-N, and where, as we shall see in section 5.7, their distribution and internal structure can be derived directly from other language-internal facts, namely main clause word order patterns.

Secondly, if it is the verb that moves past the noun, this type of movement should be open to a verb in either Actor Voice (AV) or Non-Actor Voice (NAV). However, in Seediq, CNR constructions do not obtain in AV (33a). Huang (2000c) does, indeed, present an example of a CNR in Actor Voice in Puyuma (33b). They are however, less frequent, and in our data we do not have a single example, suggesting that NAV is at any rate preferred.

33 a) *Ni=ku beebu [m-n-eeguy] laqi [hlama]. PROG=1sN hit.AV AF-PST-steal child candy ‘I am beating the child that stole the candy.’

b) me-nau=ku dra [penuakpuk] dra trau [dra walak] AV-see=1sN OBJ hit OBJ person OBJ child ‘I saw a person who hit a child.’ (Huang 2000c:179)

This asymmetry could be explained by referring to the Agent / Object asymmetry
observed by Aldridge (2004b): Agents, but not objects, can be preposed (34a, b).

34 a) Pawan-ni wada=na bbe-un ka dangi=na.
   Pawan-DEF PST=3sE hit-PV NOM friend=3sG
   ‘Pawan hit his friend.’ (Aldridge 2004b:109)

   b) *Patis m-n-atis ka seediq kiya.
      book AV-PST-write NOM person that
      (‘That person wrote a book.’) (ibid.)

Similarly, Agents can also be postposed (35a), and in some cases constructions with postposed Agents are better than constructions where the Agent in situ (35a, b), presumably for parsing reasons (in 35b, the distance between the verb with its voice marking and the subject argument the voice refers to crosses two full NPs). Regardless of the reasons for Agent postposing, the interesting fact remains that we have no attested instances of postposed objects in our data.

35 a) S-sebuc=na ricah ka btakan Pawan.
   IV-strike=3sG plum NOM bamboo Pawan
   ‘Pawan strikes down plums from the tree with a bamboo.’

   b) ?S-sebuc=na ricah na Pawan ka btakan.
      IV-strike=3sG plum by Pawan NOM bamboo
      (‘Pawan strikes down plums from the tree with a bamboo.’)

Capitalizing on these Agent / Object asymmetry, we can explain the NAV restriction on CNR constructions if we follow Aldridge (2004b) in deriving them, not by head movement of the verb, but by AGT extraction, followed by remnant movement of the VP (i.e. the remaining V) past the head noun. However, this implies that CNR is a sub-type of REL-N, combined with AGT extraction. If this is the case, we cannot use CNR as a way to account for REL-N, but rather vice versa: given REL-N, CNR can be accounted for straightforwardly. This begs the question of why circumnominal relativization is preferred to REL-N among speakers of Tgdaya Seediq, but the explanation might simply be heavy shift, i.e. a preference for heavier elements to be postposed, cf. the split relative clause in German perfect constructions (36). Otherwise, deriving CNR from REL-N is actually an advantage, in that most other Formosan languages do not display CNR, but do display REL-N.

   I have DET book bought REL 2s always have wanted.2s
   ‘I have bought the book that you always wanted to have.’

Thus, promising as it might appear from the point of view of Tgdaya Seediq, CNR does not appear to be a valid avenue for deriving REL-N. Thus, we must either assume movement of the whole relative clause to derive restrictivity, or we can propose an alternative mechanism which can derive non-restrictivity.

One important clue is the following: it is striking that the connection between prenominal placement and restrictivity which was referred to above appears to be typical of head-final structure: it is found in, e.g. the SOV language Tibetan, and for the
Panoan SOV language Shipibo, spoken in Peru, as well for Mandarin, which is SVO but has clear head-final characteristics otherwise. Similar facts could be found in Basque, likewise an SOV language. In the light of this, it is highly unexpected that this generalization should hold in VSO/VOS languages. For this generalization to be able to explain the word order variation in Formosan languages, we are indirectly equating them with head-final structures a priori. Thus, this cannot in itself explain the existence of REL-N in Formosan.

One rationale for the placement contrast between restrictive and non-restrictive relative clauses could be that non-restrictive N-REL is really an afterthought, i.e. an assertion presented as additional information, and not a real relative clause. It would then follow that REL-N must be the basic pattern, and N-REL derived from it. The restrictivity pattern can only be explained if we have already derived REL-N. Therefore, the next section will directly address the issue of how REL-N can be derived.

4. Rel-N as a matrix construction

Hitherto we have tried to account for REL-N in Austronesian by generalizing patterns which are attested in other languages or language types. However, there is an alternative which we have not addressed yet: the possibility that REL-N is the result of specifically the Austronesian context, notably the combination of two patterns recurring across the entire family: subject-final word order and subject-oriented relativization.

4.1 Subject-final word order

Austronesian languages are typically verb-initial. Some are described as VSO, others as VOS, and it might therefore appear that to describe Austronesian languages as subject-final is an exaggeration. However, Kroeger (1993) and Billings (2005) outline two important principles which affect clause-level word order: EARLY ACTOR (i.e. the Actor should be the first post-verbal argument) and LATE SUBJECT (i.e. NOM should be the last argument). Both of these generalizations serve as strong tendencies rather than as absolutes.

In Actor Voice (henceforth AV), where the Actor and the NOM subject are the same argument, these two principles are competing, EARLY ACTOR pulling towards VSO and LATE SUBJECT pulling towards VOS. Depending on which if these principles is stronger in a given language, a language will realize surface word order VSO (37a) or VOS (37b) respectively. In some cases, such as Tagalog (37c, d), both orders are equally acceptable. In fact, our data shows that Bunun also has VOS order as a possible option (37e).

37 a) Ma-baliv ca tama-nak is davuc.
   AV-buy NOM father-1sG OBL wine
   'My father bought wine.'

   b) M-n-ari sino ka tama=mu.
      AV-PRF-buy wine NOM father=1sG
      'My father bought wine.'

   c) s<um>ulat ng-liham si=Juan
      <AV>write OBL=letter NOM=Juan
      (Billings 2005:307)
      'Juan wrote a letter.'

In Non-Actor Voice (henceforth NAV), in contrast, both principles conspire to generate the order VAS (38a, b), and VSA is highly marked (38c) or ungrammatical (38d).

This implies word order may be underlyingly subject-final even in cases where the surface order in Actor Voice appears to be VSO. This is the case across most of the highest branches of Austronesian, throughout Formosan and the Philippine languages.

### 4.2 Subject-oriented relativization

We have already touched on the subject-orientation of Austronesian relativization in section 2.1. Across most Austronesian languages, (gapped) relativization is obligatorily subject-oriented (although gapless relativization may target other roles, which is not an issue here). Thus, to be able to relativize on an argument, it must bear the subject role corresponding to the voice of the verb. This is illustrated with data from Tagalog (39a, b) and Takituduh Bunun (39c, d).

39 a) *isda-ng nagbigay ang lalake sa bata
   fish-LIG AV.give NOM man OBL child
   ('the fish that the man gave to the child')

b) isda-ng ibinigay ng lalake sa bata
   fish-LIG IV.give GEN.man OBL.child
   'the fish that the man gave to the child'
   (lit. 'the fish that was given to the child by the man')
To summarize, in Formosan and Philippine languages, the relativized head of a relativization construction is always NOM subject, and the NOM subject of a main clause is always final. It follows that a REL-N construction in a Formosan language differs only minimally from a main clause construction: while they differ in whether the morphosyntactic marker is a nominative case-marker or a relativizer, the content elements in both constructions are identical in both form and order. This is illustrated below for Bunun (40a, b) and Tsou (40c, d).

40  
a) [b<in> aliv is tama-k-inak ca acu dii]  
   <PST.PV>buy OBL father-LNK-1sG NOM dog this  
   ‘This dog was bought by my father.’

b) [b<in> aliv is tama-k-inak a] acu
   <PST.PV>buy OBL father-LNK-1sG REL dog
   ‘this dog that was bought by my father.’

c) i-si ana 'o tacumü
   AUX.PV-3s eat.PV NOM banana
   ‘S/he ate the banana.’  (adapted from Zeitoun 2000a:67)

d) mo okosi 'o i-si ana ci tacumü
   AUX.AV small NOM AUX.PV-3s eat REL banana
   ‘The banana s/he ate is small.’  (Zeitoun 2000a:135)

In Paiwan, since the relativizer and NOM marker are syncretic, subject-final main clauses and Rel-N relative clauses are identical in every respect (41a, b).

41  
a) v-in-eļi ni kai a kun
   <PV.PRF>buy GEN Kai NOM skirt
   ‘The skirt was bought by Kai.’  (Tang 2008:918)

b. k-in-asengseng ni kui a kun
   <PV.PRF>make GEN Kui REL skirt
   ‘skirts that were made by Kui’ (ibid.)

In most other Formosan languages, the morphological realization of the relativization marker is distinct from that of the nominative morpheme, so a relative clause can be distinguished by virtue of the relative marker, even if it is linearly identical to a main clause (42a). Nevertheless, the relative marker is often syncretic with other functional elements in the clause, e.g. topic markers in Takitudu Bunun (42a, b), complementizers in Isbukun Bunun (42c, d).
42 a) ma’dikla kaun-un ca b<in>aliv=inak a hutan
bad eat-PV NOM <PF.PRF>-buy=1s REL sweet potato
‘The sweet potato I bought isn’t good to eat.’

b) maca muca’an ak Taipei a, nanu na minbuqbuq
if go-1s Taipei TOP certainly FUT get.lost
‘If I go to Taipei, I certainly get lost.’

c) kalat-un ’asu’ tu pinaz hai manuaz
bite-PV dog REL woman TOP beautiful
‘The woman whom the dog bit is beautiful.’ (Zeitoun 2000b: 100)

d) haiyap saikin tu kusian-as taihuku’
know 1s COMP go-2s Taipei
‘I know you go to Taipei.’ (op.cit. 97)

Even in languages like Takituduh Bunun, where REL is formally distinguished from NOM, we often find examples where the REL appears to have been replaced by a NOM marker (43a, b). Further, there are also other constructions which display unexpected uses of the NOM markers, e.g. in existential constructions (43c, d).

43 a) Muqna siza-u-s [ma-havun qanvang ca bananaz-un]...
again carry-PV-OBL AV-drive.off buffalo NOM man-DEM
‘Again, it is carried by that man who is driving off a buffalo, he is wearing a hat.’

b) kusbai ca tamuhung panacaan [laincun-un=cia ca pangka]
fly NOM hat located push-PV=3sG NOM table
‘The hat has flown, and is located on the table he is pushing.’

c) haiza [tas’a ca tulkuk-nak]
EX 1 NOM chicken-1sG
‘I have a chicken.’ (lit. ‘There exists [my chicken is one].’)

d) haiza dia katundaq-is paan ca bananaz-un-a, tamuhung-an
EX still kick-OBL bottle NOM man-DEM-TOP hat-LV
‘There is that man kicking a bottle, wearing a hat.’

Interestingly, this kind of construction is not restricted to Formosan languages. In Tukang Besi (Austronesian: Celebic; Indonesia), matrix clause constructions with NOM marking on the head noun can also serve in lieu of an argument (44). In Tukang Besi, this takes place under certain restricted circumstances, which have to do with the grammatical function of the relativized noun, both within the relative construction and within the superordinate main clause.

44) No-wila-mo [ku-’ita’-e na mia].
3R-go-PF 1S-see-3O NOM person
‘The person I saw has left.’ (Donohue 1999:386)
The Tukang Besi data is particularly striking: here it is clear from the placement of the case marker that the head noun is acting as the surface subject of the relative clause. At the same time, there is no other potential argument for the matrix verb, so it is equally clear that the same argument is semantically the subject of the matrix clause.

The construction in Tukang Besi is clearly an instance of an *internally headed relativization clause* construction (henceforth IHRC), of the type otherwise attested in, e.g. Quechua, Korean and Japanese (45).

John-TOP Mary-NOM three-CL-GEN apple-ACC peeled-NOMLZR-ACC ate  
'John ate three apples, which Mary had peeled.' (Hoshi 1995:131)

In IHRC constructions in these languages, the head is linearly within the relative clause, since it corresponds to the object of the relative clause and the relevant languages have SOV order. However, the structurally most crucial feature of an internally headed relative clause is not that the head is linearly enclosed by the relative clause, but that it is located in situ in the structural position to which it corresponds within the relative clause (viz. in clause-internal object position in Japanese).

It is the linear ordering which prompts Aldridge (2004b) to analyse circumnominal relative clauses in Seediq as internally headed relativization. However, the head in a CNR relativization construction in Seediq is never in in situ position, since relativization is always subject-oriented (as is Austronesian relativization in general) and grammatical subjects are always clause-final in the relevant corresponding matrix constructions in Seediq (we will see an exception in Puyuma in section 5.7). A NOM argument in sentence-medial position (46a) simply does not occur in a main clause in Seediq (46b, c), as opposed to Puyuma.

46) a) Wada=mu sbet-un ka [t<n>bug-an=na] huling [na Watan].  
AUX=1sG beat-PV NOM <PRF>rear-LV=3sG dog GEN Watan  
'I beat the dog which Watan bred.'

b) T<n>bug-an=na Watan ka huling.  
<PRF>rear-LV=3sG Watan NOM dog  
The dog was bred by Watan.'

c) #T<n>bug-an ka huling na Watan  
<PRF>rear-LV=3sG NOM dog GEN Watan  
'Watan's dog was bred.'  
(# The dog was bred by Watan.)

Rather, from a structural point of view, in the Austronesian context of subject-final structure and subject-oriented relativization, the best candidate for an IHRC is precisely an example like (44) from Tukang Besi; this is incidentally also the analysis proposed by Donohue (1999:385ff). Given the surface similarities, we further suggest that Austronesian Rel-N is simply a linear reanalysis of an IHRC, i.e. of a subject-final matrix construction.
4.3 Potential objections to an IHRC analysis
There are two possible objections which might be raised to the analysis of Austronesian Rel-N as an internally headed construction. The issue of word order and the issue of definiteness.

Firstly, IHRCs are usually only reported from SOV languages. In fact, Cole (1987) goes as far as to theoretically exclude IHRC as a possibility in languages which are not SOV. Dryer (2013) lists one other verb-initial language which has clear (linearly recognizable) internally headed relativization: Kutenai, an isolate spoke in Idaho. Kutenai has VOS word order (47a) and IHRCs (47b).

```
47a) qa·k waqayq-ni ƙaq’anxuʔna-s niʔ tiŋamu
   thus roll.up-IND door-OBV DET old.woman
   ‘The old woman rolled up the tepee door.’

47b) nʔ=ip-s-i [niʔ-s k=wu-κat paŋki-s misaʔ]
   IND=die-OBV-IND DET-OBV SUBORD=see woman-OBV Mike]
   ‘The woman that Mike saw died.’
```

There is no a priori reason to assume that Kutenai should be exceptional in this case. In fact, for a verb-initial language, REL-N itself is just as unexpected, so there is no reason to prefer externally headed REL-N to IHRC as an analysis of a VOS language. Finally, the VOS languages Seediq and Puyuma do display constructions which are linearly analogous to IHRCs (although we argue here that they are not structurally speaking true IHRCs), namely circumnominal relative clauses (44) discussed in section 3.2. Thus, it is not a question of whether or not IHRCs occur in Seediq (or Puyuma), but rather a question of which of the existing constructions in these languages best corresponds to our definition of an IHRC.

```
   AUX=1s beat-PV NOM <PST>rear-LV=3sG dog GEN Watan
   ‘I beat the dog which Watan bred.’
```

In fact, assuming subject-oriented and subject-final constructions, we would necessarily expect a head which is located in situ within the relative clause, i.e. at the right margin. In such as case, a typical IHRC structure is less visible, since the clause-internal position of the head is a question of bracketing rather than linearity. There may be many more IHRCs in verb-initial languages, but they may be harder to detect, unless they are overtly case-marked, as is the case in Tukang Besi, cf ex. (44) above.

The second possible objection is that an internal head in an IHRC is normally excluded from displaying overt definiteness marking. In one sense, this is not a problem, since the Formosan languages lack overt definiteness. However, the NOM subject is usually described as being prototypically definite, to the extent that voice morphology is often harnessed to ensure that the definite argument is realized in NOM subject position (49a, b).

```
49a) M-n-ekan wawa ka huling.
   AV-PRF-eat meat NOM dog
   ‘The dog ate meat.’
```
b) P<n>uq-an  huling ka wawa.
<PRF>eat-LV  dog NOM meat

‘A dog ate the meat.’

This might be considered as an argument against an IHRC analysis. However, even though NOM is prototypically definite, this is just one of its functions, and nothing absolute. Thus, subject-oriented relativization can force an argument to NOM subject for other reasons, regardless of definiteness (50a), or even referentiality (50b).

50  a) niqan kingan [seedaq  [h<m>anguc kretun neepah]]
   EX  1  person  <AV>cook  kretun field
   ‘There was one person who cooked wild kretun mushrooms.’

   b) Uka hari [laqi  [q<m>erac plale]]  di,
   NEX  a.bit  child  <AV>catch  butterfly PRF
   ado uka naq [seedaq  [m-ari plale]]  di.
   because NEX just person AV-buy butterfly PRF
   ‘Not many children nowadays catch butterflies,
   because nobody buys butterflies.’

Therefore, we must conclude that potential objections to an IHRC analysis are invalid. Thus, we propose that REL-N in Austronesian is diachronically a matrix construction, which either replaces the argument it refers to (as in a true IHRC), or is construed in apposition to the clause as a whole, e.g. in topicalized position, a position which is open not only to arguments, but also to stage-setting clauses. The purpose of this project is twofold:

a) to find independent evidence (e.g. prosodic, information structural etc.) for or against an IHRC analysis of REL-N; and

b) assuming that REL-N represents the IHRC origin of REL-N, to account for the distribution of N-REL in the languages involved.

4.4 Assertion and restrictivity
A matrix clause represents an assertion, and restrictivity is not an issue. Whereas a relative clause can be either restrictive or non-restrictive, an assertion is non-restrictive by definition (51).

51  a) Paul bought a book, I read it.  (no restrictivity)

             b) I read the book that Paul bought.  (potentially restrictive)

If prenominal relativization is a matrix clause, we would expect it to be non-restrictive, which contradicts what is found in e.g. Tsou and Amis, but interestingly enough fits well with what has been described for Jianshi Squiliq Atayal, where REL-N in argument positions can only have a non-restrictive reading (52a, b).
52 a) cyux m-ʔuyay qu? kneril [ ka? m-n-ihiy yummin ]
   AUX AV-hungry NOM woman [ REL AV-PRF-beat Yumin
   'The woman who beat Yumin is hungry.' (Liu 2005:91)

   b) *cyux m-ʔuyay qu? [m-n-ihiy yummin ka?] kneril
   AUX AV-hungry NOM AV-PRF-beat Yumin REL woman
   ('The woman who beat Yumin is hungry.') (ibid.)

Similarly, those few examples of REL-N we found in recorded narratives in Tgdaya Seediq appear to occur with quantification constructions, which are typically non-restrictive contexts (53).

53) hbaro [m-n-enaq Hori hini ka tanah tunux cbeyo]
   many AV-PST-stay Puli here NOM red head long.ago
   'Many Japanese used to stay here in Puli long ago.'
   (lit. 'It was many that the Japanese stayed here in Puli long ago.')

If a matrix clause is used to establish the reference of an argument, this can work fairly straightforwardly in topicalized position, but much less so in argument position, as can be seen quite clearly from English paraphrases (54a–c).

54 a) The dog came in, // I fed it.
   b) I fed the [dog [that came in]].
   c) *I fed [the dog came in].

Interestingly enough, we find cases where topicalized REL-N is preferable to REL-N in argument positions. Thus, Huang (2000b:155) mentions that REL-N is only found in topicalized position in Mayrinax Atayal (55). In previous fieldwork similar facts have been found for Seediq, where topicalized REL-N is more generally acceptable than REL-N in argument position.

55. t<um-in>uting 'i' kuing cu' hisa'
   <AF-PRF>hit ACC 1s ADP yesterday
   ka' kanairil ga' awuni la
   REL woman TOP arrive PRF
   'As for the woman who hit me yesterday, she has arrived.' (Huang 2000b:154)

This being the case, we can see that there are reasons to identify REL-N with non-restrictivity, at least in Atayalic, at the same time as there are other reasons to identify REL-N with restrictivity in other Formosan groups. Assuming that restrictive relativization is the prototypical function of relative clause constructions, it appears that there are two different strategies towards non-restrictivity. One, that found in Atayalic, appears to be preserving what may be an archaic matrix construction, while the other, found in, e.g. Tsou and Amis, appears to be a question of adding a reduced subject-less clause as an afterthought after the head noun. Given that these two strategies for non-restrictivity occur, with concomitant alternating REL-N and N-REL order, respectively, the crucial issue for this project is how investigate how restrictive relative clauses are formed, and if they harness one or other of the two strategies, or if they instead diverge from the relevant non-restrictive for disambiguation.
5. The project
The main task of the project has thus been to investigate the actual distribution of NP-internal ordering patterns in three Formosan languages. The methods used were primarily various kinds of elicitation tasks (director-matcher tasks, image description tasks, question-answer tasks based on images illustrating contrasting actions and contrasting referents), designed to elicit restrictive relativization constructions, constructions where the noun bears contrastive focus, and double-adjective constructions. Some film stimuli to elicit continuous narrative were also included for control purposes.

5.1 Tgdaya Seediq results
In the Tgdaya Seediq data, there were virtually no unambiguous examples of REL-N recorded. Regardless of the fact that several of the speakers accept REL-N structure, they did not use them in any of the tasks involved. It even proved difficult to identify REL-N in performance errors in the tasks. Some examples were found in free narrative, and these occurred with quantification constructions (56a, cf. also 53 above), which is exactly the same kind of construction as REL-N examples mentioned by Chang (2000), cf. (56 b). In these examples, the presence of the NOM ka marker preceding the head noun suggests that the relative clause (e.g., [musa lmanguy ka qbhenil]) is not only linearly identical to a main clause, but in fact structurally identical.

56 a) egu riyung ka [musa l<m>anguy ka qbheni]  
much very NOM go <AV>swim NOM bird  
‘Many birds went to swim there.’  
(lit. ‘It was very many / much that birds went to swim.’)

b) egu riyung wada m-ekan ido ka laqi  
many very PST AV-eat rice NOM child  
‘the children who ate rice were many’ (Chang 2000:135)

Further, at least one example of reduced REL-N (or reduced circumnominal relative constructions) was also found in narrative, but not in the experimental data. It is not clear whether this is a reduction of a circumnominal or a prenominal relative construction (57).

57. m-os a l<m>amu p<n>huma=na heyi qhuni  
AV-go <AV>pick <PV.PRF>plant=3sG fruit tree  
‘goes to pick the fruit he has planted’

Neither of the types above are prenominal relativization in a sense directly comparable to that found in e.g. Bunun och Puyuma, and since none of them occurred in the experimental data, the function and variation could not be analysed and compared with the other two languages. Therefore, the next step was to investigate the possibility of ADJ-N order. Here the two youngest speakers more or less consistently produced ADJ-N, while the older speakers consistently produced N-ADJ. There was no systematic variation across different contexts or elicitation tasks, rather, the distribution appeared to be completely idiolectal. The reasons why prenominal adjectives are preferable to
preponominal relative clauses presumably has to do with heaviness: heavy elements are more likely to be postposed.

There were no further systematic patterns found within Seediq. However, the general pattern found in Seediq can be related to patterns of language-internal variation in Bunun, which will be demonstrated in the following section.

5.2 Takituduh Bunun results

Given that the elicitation tools were designed to elicit restrictive relative clauses, we were expecting to find exclusively REL-N order. In actual fact, the results also contained many existential clauses, most of which had subjects with internal N-REL order (58a) although some also had REL-N order (58b).

58 a) Haiza ta’a binauaz angkuc lukic.
   EX 1 woman hold stick
   ‘There is a woman holding a stick.’

   b) Haiza ca kavila mali binanauaz
   EX NOM play ball woman
   ‘There is a woman who is playing with a ball.’

For the purpose of statistics, existentials were excluded from the survey. Nevertheless, it is interesting to note that some apparent REL-N constructions displayed the NOM marker ca instead of the REL marker a, indicating that the existential in these cases had a sentential subject rather than an NP subject containing a relative clause (59). Examples like these provide support for an analysis of REL-N as being derived from or even sometimes identical to, an internally headed relative clause.

59) haiza ca katundaq-is paan ca bananaz-un-a, tamuhung-an
   EX NOM kick-OBL bottle NOM man-DEM-TOP hat-LV
   ‘There is a man kicking a bottle, wearing a hat.’

As far as non-existential examples are concerned, where we have no reason to expect non-restrictive relativization, we expected uniform REL-N order. However, in actual fact, of the total of 386 non-restrictive relative clause constructions elicited in the data, only 63% displayed REL-N order (60a), while 37% instead displayed N-REL order (60b). Single examples also surfaced with circumnominal order (60c), but these are too few and unsystematic to be analysed, and may represent performance errors.

60 a) Icaan [ma-laincu-s pangka a] bananaz ca tamuhung.
   LOC AV-push-OBL table REL man NOM hat
   ‘The hat is on the man pushing a table.’

   b) Maca tamuhung a, munhaan-in binauaz [ma-laincun pangka daing i].
   TOP hat TOP go-PRF woman AV-push table big DEM
   ‘The hat, it has gone to the woman pushing a big table.’

   c) Icaan [laincun-u-s] pangka-[c]is [bananaz] ca tamuhung
   LOC push-PV-OBL table-OBL man NOM hat
   ‘The hat is on the tables pushed by the man.’
We discount the circumnominal examples and focus on the contrast between REL-N and N-REL. The most salient factor determining the order of REL and N appears to be voice form of the verb within the relative clause. Theoretically, all four combinations are possible: AV, REL-N (61a); AV, N-REL (61b); NAV, REL-N (61c); and NAV, N-REL (61d).

61 a) Icaan ma-ludaq-is qanvang a bananaz ca tamuhung.  
   LOC AV-beat-OBL buffalo REL man NOM hat  
   'The hat is on the man who is hitting the buffalo.'

   b) Maca tamuhung a,  
      TOP hat TOP  
      munhaan-in binauz dii ma-lalaincun pangka daing dii.  
      go-PRF woman DEM AV-push table big DEM  
      'The hat, it has gone to the woman who is pushing the big table.'

   c) Haiza ta’a ca tulkuk ihaan tandangaus-is kalat-u-s acu a bananaz.  
      EX 1 NOM chicken LOC front-OBL bite-PV-OBL dog REL man  
      'There is a chicken in front of the man who is being bitten by a dog.'

   d) Maca tamuhung a, munhaan-in mali na katundaq-un bananaz dii.  
      TOP hat TOP go-PRF ball FUT kick-PV man DEM  
      'The hat, it has gone to the ball that i being kicked by the man.'

However, although all combinations do occur, some are much more frequent than others. The results are presented below:

62) AV   NAV  
   REL-N  210  35  
   N-REL  71   70

As we can see, there is a strong correlation between voice and ordering. As many as 75% of the relative clauses with an AV verb appeared in REL-N constructions, whereas only 33% of relative clauses with a NAV verb did. This strong pattern requires to be explained.

One possible explanation for the Bunun pattern might stem from the fact that voice alternation is usually only an issue for 2-place verbs. Thus, while a transitive verb like ma-laincun / laincun-un ‘push AV / PV’ can meaningfully display a voice alternation, depending on which of the two arguments of the verb is realized as NOM subject, the same can not be said of maluduldul ‘stand’, which has a single argument. Thus, in practice, NAV verbs usually occur with 2-place verbs and are therefore far more likely to represent transitive constructions than are AV verbs.

63) _______ stand _______ push  
   AV:    ma-luduldul    ma-laincun

   NAV:    -      laincun-un
This implies that some relative clauses in the data consist simply of a verb, while others consist of both a verb and a non-subject argument (Agent or Object, depending on the voice of the verb). The vast majority of NAV relative clauses would for the reasons outlined above belong to the latter group, although the data does contain a handful of NAV verbs where the Agent is omitted (64a). Thus, the ordering generalization might not primarily be based on voice, but on transitivity, which in turn would stem from a preference for heavier elements to occur further rightwards on the clause (cf. Heavy Shift). There are a couple of striking examples in the data, with double AV relative clauses, one intransitive and one transitive. These invariably place the intransitive relative clause prenominally and the transitive relative clause postverbally (64b).

64 a) Icaan bungu-s ludaq-un a qanvang ca tamuhung.
   LOC head-OBL hit-PV REL buffalo NOM hat
   'The hat is on the head of the buffalo that is being beaten.'

   b) Maca tamuhung a m-u’ulusqu haan
      TOP hat TOP AV-move LOC
      ma-luduldul a binanauaz ma-laincun pangka.
      AV-stand REL woman AV-push table
      'As for the hat, it has moved to the woman who is standing pushing a table.'

To test the transitivity connection, we therefore grouped the examples according to transitivity, which was defined in a strictly linear fashion. Any relative clause which contained an overt argument of any kind, be it Agent (for NAV verbs), Patient (for AV-verbs) or Location (for some verbs of movement), was treated as transitive for the purposes of this study, while relative clauses which only contained the verb itself were treated as intransitive (including NAV verbs where the Agent was omitted). This approximation is motivated since it is the structural heaviness of the relative clause which is the relevant factor. The results are given below.

65) Intransitive  Transitive
   REL-N  123           122
   N-REL   14           127

The hypothesis that ordering depends on heaviness was partially confirmed: intransitive relative clauses turned out to be overwhelmingly prenominal (90%), while transitive relative clauses are more or less equally likely to be REL-N or N-REL. This implies that heaviness clearly seems to be part of the equation. However, it can not be the whole answer, since we still have no account for the ordering distribution for transitive relative clauses. Here, voice still seems to be the better approximation.

The next step was therefore to look at intransitive and transitive relative clauses separately, and to check these against the voice of the verb. Taking intransitive relative clauses first, as a control group, it is clear that heaviness is still by far the strongest factor (66). Impressionistically, the REL-N dominance seems to be slightly weaker for NAV than for AV, but this is not significant.
66) **Intransitive relative clauses**

\[
\begin{array}{cc}
AV & NAV \\
REL-N & 108 & 15 \\
N-REL & 10 & 4 \\
\end{array}
\]

For transitive relative clauses, the picture is completely different. Here, there is a strong correlation between AV and REL-N: 63% of AV relative clauses have REL-N order, but only 23% of NAV relative clauses have this order.

67) **Transitive relative clauses**

\[
\begin{array}{cc}
AV & NAV \\
REL-N & 102 & 20 \\
N-REL & 61 & 66 \\
\end{array}
\]

Thus, there are two factors which independently of one another trigger N-REL order: a) heaviness on the part of the relative clause, and b) NAV marking on the verb in the relative clause. Heaviness requires no further comment, but the correlation between word order and voice is extremely interesting and merits further study.

### 5.3 Voice and relativization order in Bunun

We have noted in section 2.3 that clause-level word order appears to be semantically determined: basic word order in Bunun is V-AGT-PAT, irrespective of which of these happens to be the grammatical NOM subject. Thus, in AV, the order is VSO (Xa) and in NAV, the order is VAS (68b).

68  a) Ma-ludaq ca bunun is acu.  
    AV-hit NOM person OBL dog  
    'The person hit a dog.'

   b) Ludaq-u-s bunun ca acu.  
    hit-PV-OBL person NOM dog  
    'The dog was hit by a person.'

Thus, it is only in NAV that a matrix clause and a REL-N relative construction are potentially ambiguous, if we discount the distinction between the REL marker a and the NOM marker ca (cf. 69 with 68b).

69) ludaq-u-s bunun a acu  
    hit-PV-OBL person NOM dog  
    'the dog which was hit by a person'

It is therefore perhaps not surprising that it is exactly the type of structure in (69), NAV in REL-N, which is the least frequent combination in our data. In Holmer (2014) it is argued that the rationale underlying the voice correlation is disambiguation: other factors being equal, N-REL is harnessed to ensure that the relative clause is maximally distinguishable from a main clause. If this is true, it would follow that REL-N is the
underlying basic structure in relativization, and that N-REL is derived. We will return to this conclusion in section 5.6.

5.4 Seediq topicalization revisited
In Bunun, we have shown that disambiguation from matrix clauses is an important factor in relative clause ordering, in that it links the correlation between voice and word order in relative clauses with the correlation between voice and word order in main clauses. This is visible in Bunun, where main clause word order is dependent on voice. In Seediq, which is uniformly subject-final (VOS, VAS), this pattern of covariation can not be seen. But interestingly enough, the correlation between main clause word order and relative clause word order seems to follow the same principle. Given that word order is uniformly subject-final, REL-N should always be potentially ambiguous, and therefore preferably avoided. And in fact, this is what occurs: even speakers who generally accept REL-N tended not to use this order in the elicitation experiments. Thus, the pattern accounting for word order in Bunun can be extended naturally to Seediq.

In fact, the analogy possibly goes one step forward: there is one context where disambiguation between main clause and relative construction has a low functional load, and that is in stage-setting topicalization constructions. This can be illustrated straightforwardly in English (70a, b).

70  a) Bill bought a book, I read it.

b) The book that Bill bought, I read it.

Here it does not really matter whether the stage-setter is a single NP or a whole clause, the net result is more or less the same. This should imply that topicalization constructions are a position where relative constructions should not be subject to the same disambiguation pressure. And that is indeed what we find. One of the explicanda of the current project was the fact that fronted position appears to be more tolerant of REL-N than other contexts, both in Seediq and Jianshi Squiliq Atayal (cf. the discussion in section 2.2).

Here, Bunun gives us further unexpected support: over and above the existential contexts outlined in section 5.2 above, one context where we can find unambiguous matrix clauses, with an overt NOM marker ca instead of the REL marker a, in positions where we would otherwise expect an argument, is in topicalized position (71).

71) A ma-ludulul ca uva'az a, aiza puwaq. TOP? AV-stand NOM child TOP, EX flower

’Where the child is standing, there is a flower.’
(lit. ’The child is standing, there is a flower.’)

5.5 Voice pattern in Puyuma
In Puyuma, given that both orders are expected by speakers, we were expecting a more or less equal distribution of N-REL and REL-N. However, there was a slight preference in favour of N-REL, which was the order found in 89 of the 134 relative clause examples in the data (i.e. 66%). Given the ordering variation, one of the first factors to examine is voice. Interestingly enough, while there is a significant correlation between voice and
word order, it is the diametrical opposite of the situation in Bunun. In Puyuma, AV correlates with N-REL, not with REL-N. The relevant data is given below.

<table>
<thead>
<tr>
<th></th>
<th>N-Rel</th>
<th>Rel-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV</td>
<td>61</td>
<td>20</td>
</tr>
<tr>
<td>NAV</td>
<td>28</td>
<td>25</td>
</tr>
</tbody>
</table>

Thus, while for NAV, N-REL and REL-N are more or less equally frequent, for AV, N-REL outnumber REL-N by 3 to 1. This is surprising in the light of subject-finality. However, it is interesting to note that it does correlate with main clause word order in the same way as in Bunun. In Puyuma, main clause word order is fairly free. Thus, both VOS and VSO are accepted by speakers as grammatically correct in AV (73a, b), and likewise, both VAS and VSA are accepted in NAV (73c, d).

5.6 Bracketing ambiguity in Puyuma

However, there are other factors at play in Puyuma. One of these has to do with the fact that there is no relative marker, and instead both the head noun and the relative clause are marked by the same kind of determiner. In other words, there is an even more severe kind of ambiguity to avoid: bracketing ambiguity. This is illustrated by a hypothetical example (74).
74) ? ulaya kandri kana walak na trematrekelr dra enay na terekuk be DEM DET child DET drink OBJ water DET chicken

This example might theoretically be acceptable, since the case-marking of the arguments is correct and the ordering of each element is in theory possible. However, we found no examples whatsoever of this kind, and the reason is presumably ambiguity. We can recognize that na trematrekelr dra enay 'that is drinking water' is a relative clause, but given that we have no REL marker linking the relative clause with the head noun, we have no way of linearly identifying whether it is a postnominal relative clause referring to the child (75a), or a prenominal relative clause referring to the chicken (75b), although the two structures could in fact be distinguished prosodically (N-REL and REL-N have two different prosodic structures, cf. section 6).

75 a) ? ulaya kandri [kana walak na trematrekelr dra enay] na terekuk ex DEM DET child DET drink OBJ water DET chicken ('The chicken is by the child that is drinking water.')

b) ? ulaya kandri kana walak [na trematrekelr dra enay na terekuk] ex DEM DET child DET drink OBJ water DET chicken ('The chicken that is drinking water is by the child.')

This fact seems to account for the surface observation made in Holmer & Karlsson (2012) that while both orders obtain freely if the relevant NP is alone in postverbal position, e.g. if the subject has been topicalized (76a, b), only Rel-N is found in a non-subject argument position in a VOS clause (76c).

76. na kabung i... as for the hat...

a) ... tu=pi-kabung-aw kandru kana ma'inayan na ma'kan dra belbel 3s=CAUS-hat-PV DEM DET man DET eat OBJ banana '

b) ...ulaya kana ma'kan kana belbel na mainayan ex DET eat DET banana DET man '

... it is worn by the man eating a banana.'

b) ...ulaya kana ma'kan kana belbel na mainayan ex DET eat DET banana DET man '... it is on the man eating a banana.'

c) Ulaya kana matengadaw kana babayan na kabung.

be DET sitting DET woman DET hat 'The hat is on the woman who is sitting.'

The generalization seems to be as follows: if there are two NPs linearly adjacent to each other, the only natural placement for the relative clause which avoids bracketing ambiguity is if it does not intervene linearly between the two nouns. Thus, in a VOS structure, O would normally only display the internal order REL-N and S would normally only display the internal order N-REL. This is illustrated graphically below (77).

(77) V _ O _ S
     [Rel-N] [N-Rel]
This hypothesis can be partially confirmed by our data: we have found no examples which contradict this generalization, i.e. where the relative clause intervenes between two otherwise linearly adjacent NPs. However, this is only partial confirmation, since the test situation (2 linearly adjacent NPs where one contains a relative clause) is rather infrequent. Where this situation occurs, the structure of the example exactly mirrors (76c). In most of the examples, it is either the case that one of the arguments is topicalized, placing the arguments neatly on opposite sides of the verb (78a), or, in verb-initial clauses, one of the arguments is realized by a clitic pronoun, which can not be modified by a relative clause (78b). Given the potential bracketing ambiguity which would have to be solved by judicious placement of the relative clause, it is not surprising that alternative strategies are favoured.

78 a) i na kabung i, ulaya tu=pi-kabung-aw
   TOP DET hat TOP EX 3sG-CAUS-hat-PV
   kandru kana babayan na s<em>asulit dra papadaran
   DET DET woman DET <AV>push obj table
   ‘The hat, it is worn by the woman who is pushing the table.’

   b) tu=salpitay na suwan na k<em>akarat dra traw
      3sG=hit-LV DET dog DET <AV>bite OBJ person
      ‘The dog that bit the person was beaten by him....

The two aforementioned factors (voice correlation and bracketing ambiguity) are fairly good predictors of word order. Other factors being equal however, there seems to be an information structural account as well: thus, it appears that N-REL is a way of emphasizing the contrastive function of the content of the relative clause. Thus, giving the two examples (79a, b), one speaker specified that (79b) with N-REL order emphasizes that the sweet potatoes were grown by the speaker's father.

79 a) sagar=ku m-ekan dra
    like=1sN AV-eat OBJ
    tu-s<in>arem-an kan nama=li na bunga
    3sG<-PV.PRF>grow DET father=1sG DET sweet.potato
    ‘I like to eat the sweet potatoes grown by my father.’

   b) sagar=ku m-ekan dra
      like=1sN AV-eat OBJ
      bunga tu-s<in>arem kan nama=li
      sweet.potato 3sG<-PV.PRF>grow DET father=1sG
      ‘I like to eat the sweet potatoes grown by my father.’

Note that this implies that it is postposing which indicates emphasis, not preposing. This suggests two things: firstly, that the information structural connection between REL-N and restrictivity is more complex than previously assumed; and secondly, that, that it is N-REL which appears to be the derived order, while REL-N is in some sense basic. Thus, the information structural contrast can account for the attested variation, but hardly for REL-N order as a whole. We will return to this second point presently.
5.7 Information structure and Puyuma

Given that the mechanisms of relative clause placement are affected by parsing considerations, potentially including ambiguity with respect to main clauses, and affected by the voice of the verb, we tested the information structural behaviour of adjectives in double adjective constructions. The initial hypothesis was that double adjective constructions would make movement processes more transparent, and specifically, that constructions which respected Sproat and Shih’s (1991) Adjective Order Restrictions would be more likely to be basic, while constructions which violated these restrictions would be more likely to be derived by movement. To test adjective order restrictions in a controlled fashion we designed a stimulus set which contained referents which differ in both colour and size, i.e. which could elicit examples like the big black buffalo, where either the colour or the size could be made contrastive by the succession of images.

The results, presented in Holmer & Karlsson (2012) show that there ordering patterns differed systematically depending on whether the NP occurred in topicalized position or in postverbal argument position (i.e. as the subject of an existential).

In topic position, Adjective Ordering Restrictions were not violated, but this result was not significant, since there was substantial word order variation (e.g. N-COLOUR-SIZE, SIZE-COLOUR-N, SIZE-COLOUR-N). The surprising pattern that appeared to emerge for arguments in topic position was that a contrast in COLOUR was expressed by the ordering SIZE-N-COLOUR, while a contrast in SIZE was expressed by the ordering SIZE-COLOUR-N. In the first case, contrastiveness is expressed by rightward movement, in the second case either by leftward movement or no movement at all (assuming that SIZE-COLOUR-N is the underlying pattern).

In existential subjects, it instead surfaced that contrastiveness on the ADJ was expressed by prenominal placement, regardless of the semantic domain of the adjective. Thus (80a, b) represent two consecutive utterances in the elicitation session.

80 a) ...ulaya a kaagungan a makiteng a utreutrem
   EX DET buffalo DET small DET black
   ‘there is a small black buffalo’

   b) ...ulaya matrina a kaagungan a utreutrem
   EX big DET buffalo DET black
   ‘there is a big black buffalo’

Thus, we can see that in existentials and topics behave differently when it comes to how contrastive information is expressed. In one case it appears that movement takes place rightwards, in the other case leftwards. This is unexpected under an assumption that movement serves to place an argument in a given functional position.

A partial solution is offered by prosody. In Puyuma, the collocation ADJ-N is normally marked by a L pitch contour on the adjective, which links ADJ to the N, while N is always marked by a H boundary tone, showing that the information package is finished (which presumably has to do with the fact that the N is the central part of the NP). This illustrated schematically in (81).

81) \[ \text{ADJ}_L \quad \text{N}_H \quad // \]
It follows that right-movement can serve to place an argument in a position where it can be marked as belonging to a separate prosodic domain, thus marking it contrastively. An alternative strategy would appear to be left movement, but it appears that this always coincides with special prosodic marking of the emphasized adjective, in other words, it is not primarily the movement which bears the functional load, but the prosodic marking.

The contrast between SIZE and COLOUR can also be solved by prosody, assuming an underlying order SIZE-COLOUR-N and the prosodic boundary marking on N. Emphasis on SIZE does not require movement, since it is always possible to mark the SIZE adjective prosodically, and single it out from the remainder of the NP. However, if the same mechanism is ised to mark COLOUR prosodically, it does not mark COLOUR specifically, instead, it marks both SIZE and COLOUR simultaneously, since it is a process of boundary marking rather than stress.

Thus we can see that word order can be used to place adjectives in positions where the fixed prosodic pattern of the NP can focus them. When this occurs, it appears that the underlying structure is ADJ-N, and that derivations from this are harnessed in various way for emphasis, unless the prosodic pattern allows emphasis in situ. In its own way, prosody therefore also supports the proposal from section 4.

The same tests were also conducted for Bunun in Seediq. However, the results were subject to quite a lot of individual variation and were less reliable. In Seediq, two speakers more or less consistently N-OLD-NEW, which fits well with the generalization that rightward movement indicates focus. For other speakers, the data was too unsystematic to see a clear pattern. In Bunun, two speakers were quite consistent in producing COLOUR-N-SIZE, regardless of the information structural content, which suggests that they were simply self-mirroring.

As far as we can see, therefore, information structure is sometimes a relevant factor in the placement of adjectives within an NP, which would suggest that it behaves similarly for relative clauses, although more subtly, since these are subject to other length-related and voice-related restrictions as well. And as we have noted, this seems to indicate that the underlying point of departure is a structure where modifiers precede the noun.

5.7 Puyuma circumnominals

In our Puyuma data, we also found some circumnominal relative clauses. There were just a handful, and all were, interestingly enough, in NAV. With a single exception, they were all topicalized. Example (82) is a typical example.

82. idru tu=sinasulut na papadaran kana ma'înayan i, 
   DEM 3sG-PV.PRF>push DET table DET man TOP
   ulaya kabung isatr
   EX hat above
   ‘The table that the man is pushing, there is a hat on it.’

The fact that circumnominal constructions preferentially surface with NAV verbs is highly unexpected, given that it is exactly in NAV that the NOM subject of a main clause often occurs in clause-medial position (i.e. VSA order). Thus, it is exactly in NAV that a circumnominal relative clause would be potentially ambiguous with respect to a main clause. In fact, in (82) above, the only indicator that we are dealing with an argument rather than with a full main clause is the demonstrative idru. If we isolate the string
following *idru* (i.e., *tu=sinasulut na papadaran kana ma’inayan*), this is a fully grammatical main clause, ‘The table is pushed by the man’.

One partial solution might stem from the fact that these structures are usually found in topicalizations, where, as we noted above, main clause ambiguity is least a problem (since whether the discourse stage is set by a full clause or by a single argument is functionally irrelevant). Nevertheless, it is still intriguing why there should be a preference for NAV, given the potential ambiguity. After all, all things being equal, while ambiguity might not be a serious problem, structurally unambiguous constructions should in theory be at least as acceptable (and therefore as frequent) as structurally ambiguous constructions. The acceptability of NAV circumnominals is not a problem, but the preference is.

However, the solution is as simple as the problem appears to be complex. Let us capitalize briefly on the similarity with matrix clause and the placement in topicalized position. If we assume that these constructions are not a specially derived construction class, but rather internally headed relative clauses, i.e. matrix clauses which represent arguments, then the facts fall out straightforwardly.

Matrix clauses in AV normally display VOS word order, as in all our experimental data. VSO is accepted as grammatical, but we have no spontaneous examples in our data. Matrix clauses in NAV display either VAS or VSA, and both are equally frequent. If follows that that the only commonly occurring matrix construction where the subject is clause-internal is VSA in NAV. If we view circumnominal relative constructions as matrix clauses, we should be looking more closely at attested matrix clause types rather than at all logically possible combinations of ordering and voice.

And under this approach, we find that topicalized internally headed relative constructions occur with all the attested matrix orders: VOS (83a), VAS (83b) and VSA (83c). However, it is only the latter which linearly forms a circumnominal relative clause. Both VOS and VAS are, linearly speaking, surface instances of REL-N, and therefore were simply not classified as circumnominals.

83 a) *tu-tr<in>ekel-an kan nama-li na eraw, saima*  
        3s-<PRF>drink-LV OBJ.DET father-1s DET wine little  
        ‘The wine was drunk by my father, it is very little.’  
        (= [The wine drunk by my father], it is little.)

b) *na k<em>akarat dra traw na suwan i,*  
        DET <AV>-bite OBJ person DET dog TOP  
        *tu=sa<em>lit-ay la kana walak*  
        3s<em>=beat-LV PRF DET child  
        ‘The dog that bit a person, it was beaten by the child.’  
        (= [The dog bit a person], it was beaten by a child.)

c) *idru tu=sinasulut na papadaran kana ma’inayan i,*  
        DEM 3s<em>-<PV.PRF>push DET table DET man TOP  
        *ulaya kabung isatcream*  
        EX hat above  
        ‘The table that the man is pushing, there is a hat on it.’

It follows that circumnominal relative clauses in AV do not occur in the spontaneous data for the simple reason that the corresponding matrix clause, VSO, does not occur in
the spontaneous data. Thus, identifying REL-N with a matrix construction helps us to explain the distribution, not only of N-REL and REL-N, but also of circumnominal constructions, as an epiphenomenal consequence of the interaction of matrix clause order and head-internal relativization. Over and above this, the major mechanism driving variation is disambiguation.

Note that it follows that circumnominal relative constructions in Puyuma and in Seediq are two fairly different kinds of constructions, although they are similar on the surface. The similarities appear to be coincidental, since each can be explained satisfactorily on the basis on language-internal facts, while a common analysis which takes language-internal facts into account does not seem to work, given that the structure otherwise found in each language are too different for the same analysis to hold. One way of subsuming both under the same view would be to argue that circumnominal relative constructions in Seediq are also internally headed constructions with rightwards agent extraction, a mechanism which can be shown independently to occur in Seediq. However, we do not address this further here since circumnominals are fairly infrequent in Seediq and do not appear in the experimental material.

5.8 Summary of structure
As far as the structural facts are concerned, we can conclude that the major factor underlying the variation between REL-N and N-REL order in the three languages investigated is structural disambiguation: in all three languages, patterns which distinguish relative constructions from the corresponding matrix construction are favoured. Given that the matrix orders in the three languages differ, the resulting relativization patterns also differ in an entirely predictable manner, which also covers the behaviour of circumnominal relative constructions in Puyuma. In Puyuma, where relative clauses bear the same case-marked articles as nouns, another potential source of ambiguity, bracketing ambiguity, forces another pattern of disambiguation.

While disambiguation turned out to be the most salient factor, we found that information structure could motivate ordering in cases where disambiguation was irrelevant. This was clearest in Puyuma, where N-REL implies emphasis on the content of Rel. Insofar as ADJ placement directly reflects relativization order, we can see the effects of information structure more clearly.

One important point to note here is that both disambiguation and information structure serve to motivate the appearance of N-REL in certain well-defined contexts (in Seediq, this single context covers basically the whole language, but can be still be defined as disambiguation from a uniform subject-final structure). It follows that the results do not per se explain the existence of REL-N in head-initial languages. Instead, they account for the occurrence of N-REL under the assumption that REL-N is the underlying structure in all three languages. Recall that a plausible solution for the existence of REL-N in Austronesian is proposed in section 4, namely that REL-N derives historically (and even synchronically in some cases) from subject-final matrix constructions.

6. The view from prosody
One of the hypotheses underlying the research project was the assumption that if there are structural parallels or structural identity between a matrix clause and a corresponding relativization construction with the same linear order, this might be reflected in prosodic similarities. This hypothesis was partially confirmed.
One striking parallel between matrix structures and relative clause structures is the fact that the prosodic grouping distribution seems to be very similar. Thus, for matrix clauses, two surface word order patterns surface, generally speaking: VOS (often, but not exclusively, realized instead as VSO in Bunun; in our data we actually found more spontaneous VOS than VSO) and SVO (whether topicalized or cleft). For relativization constructions, two surface word order structures also surface: REL-N and N-REL.

While the VOS pattern and the REL-N patterns are fairly homogeneous, both in information structural function and prosodic (although function and prosodic structure differ between languages), the SVO pattern and the N-REL pattern can each be teased apart into two separate constructions, with corresponding information-structural and prosodic properties.

Thus, a surface SVO construction can either serve as a cleft construction (i.e. where the new contributed information is the identity of SUBJ) or as a topicalization construction (where the new contributed information is the content of VO). These diametrically opposite readings are distinguished prosodically, with the cleft construction including a prosodic break between S and VO, while the topicalization construction being a single prosodic unit. In analogy, N-REL is also heterogeneous, in that it can either represent a restrictive construction, in which case it is a single prosodic unit, or a non-restrictive construction, in which case there is a prosodic break between N and REL.

Therefore, prosody mirrors information structure in disambiguating two linearly identical constructions at NP level and two linearly identical constructions at clause-level. In this parallelism, cleft constructions correspond prosodically to non-restrictive relative clauses, while topicalization constructions correspond to prosodically to restrictive postnominal relative clauses.

Thus, we found that intonation in all three languages was a cue to syntactic structure. However, the primary function of intonation in these languages is to serve as a cue for information structure (in term of Givenness and Newness) and discourse structure (in terms of Topic and Comment). We found prosodic differences between relative clauses and matrix clauses, and these can be explained by differences in information/discourse status instead of a direct reflection of the syntax.

Intonation in Puyuma serves primarily to mark the status of phrases as informatively complete or incomplete (Karlsson & Holmer 2011). In Puyuma prosodic boundaries are of two types, realized as High and Low boundary tones (H and L) respectively. We find evidence that the L tone serves as a connector, implying that phrases ending with and L boundary are informatively incomplete. This also explains the H final boundary in declaratives (cuing completeness of information) as contrasting to L final boundary in interrogatives, as a result of request of information. In matrix clauses, S is always phrased as an own phrase. We get then the following phrasing patterns: VO\text{h}\text{h} S\text{l} (VO is the information requested), S\text{h} VO\text{h} (S is the information requested), S\text{i} VO\text{h} (VO is the information requested) with tropicalized S. Both word order and intonation are thus cuing information structure, although word order SVO is linearly ambiguous. REL-N is phrased as one prosodic phrase, and N is not separated by a prosodic boundary as S would be in a matrix clause. The L boundary on V is enhanced, having a connecting function and marking [VOS] clause as one prosodic unit.

In Bunun, intonation primarily serves to mark the division of the utterance in Topic and Comment, by always marking the initial boundary of Comment. These differences in use of intonation between Puyuma and Bunun yield differences in prosodic contours of relative clauses and matrix clauses. In Puyuma VOS and Rel-N are phrased differently as
the reflection of differences in information status, with VO in VOS being phrased as a complete unit and REL (i.e. in practice VO or VA) in REL-N being phrased as an incomplete unit). In Bunun, on the other hand, VS/VOS/VAS and Rel-N are phrased similarly as a reflection of same Comment-Topic Structure, V being start of Comment in both cases.

The analysis of the relation between discourse structure (in terms of Topic and Comment), information status (in terms of NEW and GIVEN) and intonation gives more evidence in favor of restrictivity as a possible factor for the choice of REL-N or N-REL order. We found that Bunun is prosodically a phrase language, i.e. it uses intonation primarily for marking boundaries. Our analysis shows that boundaries coincide with syntactic grouping but their distribution differs as a reflection of division into Topic and Comment. Thus, Comment (information that is given about Topic) is marked by an initial boundary. Relative clauses are intonationally integrated or separated from the main clause depending on whether they are part of Topic or not. REL-N clauses form tonally one prosodic unit and can only have restrictive a reading while N-REL clauses have an internal prosodic boundary, with V separated and marked as a separate prosodic unit. This gives a non-restrictive reading to the N-REL clause. However, when the same N-REL clause appears as a description of the next image of the picture stimuli, it becomes restrictive and is prosodically integrated as one unit.

7. Summary

The crucial result of the project is that evidence suggests that REL-N is Formosan is not a calque from Sinitic or Japonic, but derived from within the language family itself on the basis of two inherent properties of Austronesian: subject-final structure and subject-oriented relativization. The project has not addressed the issue of derivation as such, but has found substantial evidence showing that for each instance of variation between N-REL and REL-N in the languages examined, the functional motivation for derivation takes REL-N as its starting point and derives N-REL.

Thus, N-REL can appear as the result of a strategy disambiguating relative clauses from main clauses, it can appear for information structural reasons, e.g. to harness a prosodic boundary on N and thus emphasize the contrastive function of the relative clause. The disambiguation function is less important in topicalized position, and therefore, it comes as no surprise that REL-N is more readily preserved in topicalized position cross-linguistically. The disambiguation pattern also covaries predictably from language to language depending on the surface matrix order across the two major voice types in the languages involved, and we also found that Puyuma circumnominal relativization patterns exactly mirror existing matrix orders.

Furthermore, we have seen cases in both Seediq and Bunun where a Rel-N construction surfaces with a NOM marker instead of a relative marker (in Puyuma, where relative clauses are marked by a determiner, this is not an issue, since the relative clause in in practice always accompanied by the NOM marker referring to its subject). This never occurs with an N-REL construction in our data, further supporting the identification of Rel-N with matrix VOS/VAS. The parallelism between REL-N and VOS/VAS is further corroborated by prosodic data, lending further support to the proposed analysis.

Thus, for the languages analysed above, as far as we can see, by extension for other Formosan languages, the correct analysis seems to be the following:
a) relativization constructions are derived from subject-final matrix constructions, i.e. underlyingly internally headed relativization constructions;
b) these matrix constructions are still evident on the surface in many cases, and survive most clearly in topicalized environments;
c) disambiguation from matrix clauses pushes relativization constructions towards N-REL, and the frequency of N-REL is directly proportional to the probability of ambiguity in REL-N;
d) if ambiguity is not an issue, information structure may also push towards N-REL, to emphasize the content of REL;
e) N-REL is always an option for non-restrictive relativization constructions, where the apparent relative clause can plausibly be analysed as a subject-less main clause placed in direct apposition to the apparent matrix clause, following it;

Therefore, REL-N is not only not surprising, but in fact predictable in the Austronesian context, despite the head-initial structure of the languages. We suggest that it is predictable precisely because of the conspiracy of two salient facts in Austronesian languages: robustly SUBj-final structure and SUBj-oriented relativization. In any language or language family where this combination obtains, we predict that REL-N should at least be a diachronic stage along the development of relativization (at least until subordination strategies generalize to a head-initial pattern otherwise found in a verb-initial language, or until disambiguation needs force a shift to N-REL). Under the analysis proposed, REL-N in Formosan is an archaism surviving in the highest branches of the Austronesian phylum, from the Formosan languages across languages in the Philippines, but becoming successively rarer as we progress downward throughout the genealogical tree.

8. Dissemination
This report represents the full summary of the results attained by the project. Partial results of the project have been presented at the following conferences, workshops and invited talks.


Over and above this project report, partial results of the project are also published in the following.


9. Spinoff
The issue of information structure turned out to have higher relevance for main clause phenomena than for relativization, and was easier to identify in discourse. The research issues have therefore branched out into a follow-up project, *Integrating the Structures of Information and Discourse: a cross-linguistic approach* (INSID), P.I. Anastasia Karlsson, funded by VR / The Swedish Research Council. The current Formosan relativization project has further generated a vast amount of recorded, transcribed and tagged data from three Formosan languages, specially targeting information structure and word order, which is also being used for the project INSID and which will serve for future research projects. Finally, the project has generated new elicitation tools which have been used for the elicitation of word order data in other languages, and which will be harnessed and developed in future projects.

Certain questions remain. Given that the results of the project point to a diachronic development of word order which has little to do with cross-categorial harmony or headedness but which still tends to make the languages typologically harmonic, one
followup question is to investigate other cases where we can see traces of word order changes taking place, to see to what extent the results are generalizable to other situations. How powerful is disambiguation as a mechanism to drive language change? Finally, the fact that Seediq generalizes N-REL out of a derivation pattern which is supported by voice data in both Bunun and Puyuma begs the question of what happens in Tsou, which is at least descriptively REL-N despite its VOS word order. Replicating the same experiments in Tsou is a promising future avenue of research.

References
Basilico, David. 1996. Head position and internally headed relative clauses. Language 72: 498-532


Ogawa, Hisayoshi & Asai, Erin. 1935. Taiwan Takasagokuzu densetsu shi [Myths and traditions of the Formosan native tribes]. Taipei: Taihoku Imperial University


Lund, March 26th, 2015