AFLA

AFLA 21
The Proceedings of the 21ˢᵗ Meeting of the Austronesian Formal Linguistics Association

edited by
Amber Camp, Yuko Otsuka, Claire Stabile and Nozomi Tanaka

AFLA 21: The Proceedings of the 21st Meeting of the Austronesian Formal Linguistics Association

edited by
Amber Camp, Yuko Otsuka, Claire Stabile and Nozomi Tanaka

The Austronesian Formal Linguistics Association (AFLA) is an organization which promotes the study of Austronesian languages from a formal perspective. The 21st annual meeting of the Austronesian Formal Linguistics Association (AFLA 21) was held at the University of Hawai'i at Mānoa on 23-25 May, 2014. Of the 28 papers presented at the conference, 17 are included in this volume. The contributions include studies in the syntax, semantics, phonetics, phonology, morphology, processing, and acquisition of Austronesian languages.

Asia-Pacific Linguistics

_____

Open Access

AFLA 21:
The Proceedings of the 21st Meeting of the Austronesian Formal Linguistics Association

edited by
Amber Camp, Yuko Otsuka, Claire Stabile and Nozomi Tanaka

A-PL 25

Asia-Pacific Linguistics

―――――――――――――――

Open Access

This collection is dedicated to Professor Robert Blust, University of Hawai'i at Mānoa, both in recognition of his high quality scholarship in Austronesian languages and for his support for AFLA over the years.

Edward L. Keenan
Sandra Chung

Contents

Preface

The 21st annual meeting of the Austronesian Formal Linguistics Association (AFLA 21) was held on 23-25 May, 2014 at the University of Hawaiʻi at Mānoa. The program consisted of 28 presentations, including three plenary talks by Diane Massam, Yuko Otsuka, and Kie Zuraw. This volume includes 17 papers presented at the conference.

Amber Camp, Claire Stabile, Nozomi Tanaka, and Yuko Otsuka

FIJIAN WEAK QUANTIFICATION AS HEAD-INTERNAL RELATIVIZATION[*]

Raúl Aranovich
University of California Davis
raranovich@ucdavis.edu

Fijian weakly-quantified DPs are formed with a determiner preceded by a particle and a word expressing number, quantity, or existence. These anomalous constructions depart from the head-initial phrasal pattern generally found in other nominal expressions. I propose to analyze weakly-quantified DPs as internally-headed (IH) relative clauses. Evidence for this analysis is that these DPs are structurally identical to regular existential sentences. Fijian IH relatives, however, are restricted to existential sentences (Fijian follows a strict N-Rel arrangement in other cases). The use of IH relatives to express weak quantification, then, is a highly restricted construction in Fijian, motivated by the general structure of the Fijian clause, and in the semantics of IH relatives. Following Basilico 1996, I argue that weakly-quantified DPs in Fijian are presuppositional, being outside the scope of the existential quantifier associated with VP (Diesing 1992). This is consistent with an analysis of Fijian as a pronominal argument language (Aranovich 2013), in which only incorporated nouns, pronouns, and proper nouns are internal to the VP. An objection to the IH relative analysis of weakly-quantified DPs in Fijian is that IH relatives are found in OV languages, but Fijian is a VO language. However, exceptions to the correlation between OV and Rel-N orders exist, most of them in the Austronesian family. I speculate that the restricted use of IH relatives in Fijian reflects a general feature of the Austronesian family, preserved as a specialized construction to express weak quantification.

1.   Introduction

Fijian DPs seem well-behaved as head-initial syntactic structures, since they are headed by a determiner (*na*), which is followed by an N. This N has its modifiers to the right, as in (1a), and possessors to the left, between the N and the determiner, as in (1b).[1] But weak quantifiers are anomalous in that they seem

---

[1]   The following abbreviations are used: ACC: accusative, ASP: aspectual, DEM: demonstrative, DET: determiner, DIR: directional, EMPH: emphasis, HAB: habitual, INDEF: indefinite, INT: intensive, LIM: limitation, LK: linker, PART: particle, PERF: perfective, PL: plural, POSS: possessive, PROG: progressive, REL: relativizer, SEQ: sequential, SG: singular, SUB: subordinator, TR: transitive.

to precede the determiner, and are formed with the particle *e* and a word expressing number, quantity, existence, etc. (2a-b) offer some examples.

(1) a. na   no-na   lako
    DET POSS-3.SG journey
    'his journey' (Milner 1956:11)

   b. na   vale   levu
     DET house big
     'the big house' (Milner 1956:11)

(2) a. keirau dau rawata   mai [$_{DP}$ **e**   **levu** na ika]
    1.PL   HAB get.TR   out   PART many DET fish
    'We would catch many fish.'

   b. au   a   qarava   tiko [$_{DP}$ **e**   **dua** na bui-ni-qone]
     1.SG PAST look.after.TR PROG   PART one DET grandmother
     'I take care of an old lady'

I propose to analyze weakly-quantified DPs like the one in (2a-b) as *internally headed relative clauses* (IHRC). Evidence for this analysis is that the DPs in (1a-b) are identical to regular existential sentences.

    I will first give an overview of Fijian syntax, including a more detailed discussion of copulative and existential sentences. After that I will summarize the similarities between existential sentences and weakly-quantified DPs, arguing that they can be accounted by analyzing those DPs as IHRCs. I will then present evidence from the distribution of adverbial particles to support the claim that weakly-quantified DPs are in fact clausal in nature. I will then address a potential objection to the analysis based on the fact that Fijian is an OV language. Before concluding the paper, I will examine the semantics of weakly-quantified DPs, in the context of Basilico's (1996) proposal that IHRCs are inherently quantificational, and my own analysis of Fijian as a pronominal argument language (Aranovich 2013).

2.   An overview of Fijian grammar

Fijian is a VOS language. The verb is obligatorily preceded by a particle expressing agreement with the subject, but an overt DP expressing the subject is not obligatory. In (3a) the particle *e* expresses agreement with a 3rd person singular subject (the proper noun *o Kali*), and in (3b) the 3rd person plural subject is expressed in the particle *eratou* only.[2] These person/number particles can co-occur with an overt subject pronoun, as in (3c). Because this emphatic pronoun is optional, I take Fijian to be a *pro*-drop language.

---

2   The 3.SG particle *e* is often omitted before the emphatic particle *sa*, being used only in emphatic questions. Otherwise, *sa* may precede or follow other personal particles in Fijian. The semantics of *sa*, however, is still poorly understood. Schütz (1985:262) suggests it indicates contrast with a previous action or state.

(3) a. e   tarā   na vale   o   Kali.
      3.SG build.TR DET house DET Kali
      'Kali built a house.'

   b. eratou sa   tarā   oti   na   vale
      3.PL   EMPH build.TR PERF DET house
      'they have already built the house.'

   c. au sa   cakava **ko iau**
      1.SG EMPH do.TR   DET 1.SG
      'I did it myself.' (Milner 1956:99)

A number of aspectual particles expressing tense, aspect, mood, and other categories (including the aspectual/emphatic *sa*, the tense markers *a* PAST, and *na* FUTURE, the sequentials *mani* 'then, accordingly' and *qai* 'then, next', *dui* 'each', dau HABITUAL, *rui* 'excessively', and some others) may separate the person/number particles from the verb.

(4)   Au   **se qai** yadra mai.
        1.SG   ASP SEQ wake   DIR
        'I just woke up.'

Other particles that express a variety of aspectual, temporal, modal, directional, and more adverbial meanings occur in a fixed position after the verb (I will refer to them simply as 'adverbial particles'). These postverbal particles are different from the preverbal particles discussed earlier, even if sometimes they overlap in function.[3] Particles that occur in this position include *tiko* PROGRESSIVE, *tu* INDEFINITE (in time or place), *mai* 'here, hither', *yani* 'away, hence', *rawa* 'possible', *oti* 'finished' or PERFECTIVE, *sara* EMPHATIC, and a few more. There is a fixed order among these postverbal adverbial particles, which may occur alongside each other in a clause. Some examples are provided in (5). In (5b), the verb *tubu* 'grow' is followed by three such particles: *tale* 'also', *tu* 'indefinite', and *ga* 'only'.

(5) a. sa   kani-a **oti**   na   koli na benu
      EMPH eat-TR PERF DET dog DET leftovers
      'the dog finished eating the leftovers.'

   b. e   tubu **tale tu   ga** na   dalo
      3.SG grow also INDEF only DET taro
      'taro grows also.' (Milner 1956:93)

---

[3]   Some particles that appear after the verb can also occur before the verb, with a different but related meaning. *Mai*, for instance, means 'here' or 'hither' in postverbal position, but it behaves as a converb meaning 'come and ...' when it precedes the verb.

3.   Copulative and existential sentences

There is no copula in Fijian. Copulative sentences are formed by juxtaposition of two DPs, or by placing an adjective right after the subject marker.

(6) a. e   lekaleka na   lawa
        3.SG short   DET   net
        'the net is short.' (Milner 1956:13)
    b. na   yaca ni gone o   Tubarua
        DET name of child DET Tubarua
        'the name of the child was T.' (Schütz 1985:79)

Existential sentences can be formed with an expression (usually an adjective) of number or quantity as the head of the predicate. A couple of examples were introduced in (2). Besides *levu* 'many' and *dua* 'one', some of these expressions include other numerals as in (7), and *so* 'a little, a few, some'; *vuqa* 'many'; and *vica* 'a few, how many', as in (8).[4]

(7) a. e   **dua** na   drau-ni-kau ka vakā   na   levu ni   dua na   sede.
        3.SG one DET tree.leaf   REL resembles DET size SUB one DET cent.
        'there is a leaf that is about the size of a penny.'
    b. e   **tolu** na   nodratou waqa na   lewe-ni-koro.
        3.SG three DET POSS.3.PL boat   DET village.crew
        'the villagers have three boats.' (Milner 1956:37)[5]
(8) a. e   **so**   na   vulagi.
        3.SG some DET visitor
        'there were some visitors' (Schütz 1985:329)
    b. e   lewe   **vuqa** (na   tamata)
        3.SG PERSONAL many DET person.
        '(there are) many people' (Miner 1956:36)
    c. e   **vica** na   uvi oqori?
        3.SG few DET yam there
        'how many yams are there?'

4.   Weakly quantified DPs

In general, the same quantificational expressions that occur in existential sentences can also occur in DPs with a weak quantificational meaning.

---

[4]   Schütz includes also *bini* 'plentiful' and *lailai* 'few', but I have no examples of these expressions being used. In fact, he reports that *bini* and *lailai* cannot be used in existential constructions in standard Fijian (Schütz 1985:102). Thus, sentences like **e bini/lailai na vale*. 'there are lots of/few houses' cannot be found in the standard dialect. He also notices that *vuqa* used as a numeral with common nouns is not accepted by Fijian linguists.
[5]   Possessive sentences are expressed in Fijian as existential clauses with a possessed noun.

(9) a. e   **tolu**   na   waqa
       3.SG three DET boat
       'three canoes.' (Milner 1956:14)
   b. e   **vitu**   na   gone lalai
       3.SG seven DET child small
       'seven small children' (Milner 1956: 24)
   c. era yaco mai [e   **so**   na   vulagi]
       3.PL arrive DIR 3.SG some DET   foreigner
       'some visitors arrived (Schütz 1985:329)
   d. erau raica [e   lewe   **vica** (na   yalewa)]
       3.PL see.TR 3.SG PERSONAL few DET   woman
       'they saw a few women.' (Milner 1956:37)

There are several reasons to think that existential clauses and weakly quantified DPs must have a common source, sharing a representation at some level. First, some of the expressions of quantity have identical selectional restrictions. The numerals, *vuqa* 'many', and *vica* 'a few' must be preceded by *lewe* when they refer to people, as in example (9d). Compare it to the use of *vica* in an existential sentence:

(10)   e   lewe   **vica** na tamata?
        3.SG PERSONAL few DET person
        'how many people are there?'

The particle *so* 'some', on the other hand, is exempt from this restriction in existential clauses and in DPs as well. Second, some expressions have the same idiosyncratic uses in existential sentences and in DPs. For instance, *levu* means 'many' in these contexts, but as an adjective it means 'big'. On the other hand, so 'some' cannot be used as an adjective at all.

The close connection between weakly quantified DPs and existential clauses in Fijian has not escaped the attention of dedicated Fijian grammarians. This can be glanced from the following quotes, the first one from Milner's grammar (Milner 1956), the second one from Schütz's (Schütz 1985).

> The Fijian numerals are bases. It is not possible to use *dua* (one) or *rua* (two) for example, as in the English: one house, two men, etc., without using either a sentence, e.g. "the house is one" (see paragraph 35) or a phrase e.g. "the one house." (Milner 1956:23)

> The following examples show how a specified subject and object (respectively), definite because of sentence structure, are made contextually general (...). In each sentence, the NP that has been more general is derived from an existential VP. (Schütz 1985:329)

What these grammarians did not have at their disposal was a natural and elegant way of making the connection between the two categories explicit, making it to follow from more general principles about the structure of human language. I will suggest that an analysis of weakly quantified DPs in Fijian as IHRCs.

5.   Relativization in Fijian and the interpretation of DPs

Example (11) shows an Internally headed Relative Clause in Ancash Quechua (from Cole 1987).

(11)   [nuna **bestya-ta** ranti-shqa-n] alli   bestya-m   ka-rqo-n.
        man   horse-ACC buy-PERF-3   good horse-VALIDATOR be-PAST-3
        'The horse that the man bought was a good horse. '

In this sentence, the head of the main subject is *bestya* 'horse'. But this noun occupies a position inside the relative clause that modifies it. This head noun bears the case suffix appropriate to its function inside the relative clause (i.e. accusative), which is where a relative pronoun would originate from in an externally-headed relative clause.

In Cole's analysis, an IHRC is left-adjoined at S-structure to an empty nominal head, which is coindexed with the internal head of the relative clause (12a). Basilico (1996) offers an alternative analysis in which there is no empty head external to the clause. Rather, the internal head introduces a free variable that is bound by an operator in D (12b). This D takes a sentence (i.e. an IP) as its complement.

(12) a. [$_{NP1}$ [$_{S''}$ nuna bestya-ta$_i$ ranti-shqa-n] [$_{NP2}$ e$_i$]]
    b. [$_{DP}$ [$_{D'}$ [$_{IP}$ nuna bestya-ta ranti-shqa-n] D]]

In this paper I will follow Basilico's analysis. The structure for a relative clause like (9a) is presented in (13). Because Fijian is head-initial, I am placing the head of DP to the left of the IP.

(13)   [$_{DP}$ [$_{D'}$ D [$_{IP}$ e tolu na waqa]]]

The IHRC analysis of weakly-quantified DPs captures the similarity in the use of numerals and expressions of quantity across DPs and existential sentences. It also accounts for the presence of the particle *e* in those DPs, and for the word order facts. But these facts are not enough, in my mind, to justify a clausal analysis of weakly-quantified DPs. What is missing is some other piece of evidence that those DPs do indeed have the internal properties of clausal constructions. Adverbial particles provide that evidence.

Like other predicates in Fijian, existential ones can be modified by adverbial particles: the locative *kina* in (14a), the progressive *tiko* in (14b), *ga* in (14c), and *tu* in (14d).

(14) a. e   levu **kina** na   kubou .
        3.SG big   there DET smoke
        'there is a lot of smoke there.'
     b. e   rua **tiko** na   turaga e   na   gauna koya
        3.SG two PROG DET chief in DET day   those
        'there were two (rival) chiefs in those days' (Milner 1956:29)
     c. e   rua **ga** na   yava-na
        3.SG two only DET foot-3.SG. POSS
        'it has only two feet (i.e. wheels)' (Schütz 1985:92).
     d. sa   rua **tu**   na   tabua.
        EMPH three INDEF DET whale.tooth
        'There are two whale's teeth (no more are expected).'

The prediction made by the IHRC analysis of weakly-quantified DPs is that these particles will also show up between the numeral or quantity expression and the internal DP in these expressions. The examples in (15) show that this is indeed the case.

(15) a. eratou a   rogoca ni sa   yaco mai [e   dua **tale** na   waqa]
        3.PL   PAST hear.TR SUB PART arrive DIR   3.SG one also DET boat
        'they heard that another boat had arrived.' (Milner 1956:49)
     b. [e tolu **ga** na   tamata] eratou kauta   mai na   kedratou kakana.
        3.SG three only DET man   3.PL   bring.TR DIR DET 3.PL.POSS food
        'only three of the men brought their own food.'
     c. e   vuku [e   dua **tale ga** na   qase-ni-vuli]
        3.SG smart 3.SG also only one DET school.teacher
        'There is also another teacher that is smart.'

6.   Fijian IHRC in the context of relativization in Fijian.

A serious objection to the IHRC analysis of weakly-quantified DPs in Fijian is that IHRCs are usually found in OV languages, but Fijian is a VO language. In Cole's (1987) analysis, IH relatives are a sub-type of prenominal relative. He makes the observation that IHRCs are restricted to left-branching languages (OV). To account for this generalization, he proposes the structure in (12a), in which IHRCs are left-adjoined to a null anaphoric head. In this configuration, the anaphoric head commands the antecedent, but it does not precede the antecedent (which is inside the relative clause), making it possible for the antecedent to bind the anaphoric head.

The problem with Cole's analysis is that it is based on a false generalization. IHRCs are found in Mooré, an SVO language of the Niger-Congo family, spoken in Burkina Faso (Lehr et al. 1966, Tellier 1989). Comrie (2006) finds a score of exceptions to the correlation between OV and Rel-N orders, most of them in the Austronesian family (the rest of them in the Sinitic branch of Sino-Tibetan). In fact, Comrie suggests that the combination of OV and Rel-N orders may be a distinctive typological trait of the Austronesian languages. Some languages in this family -- Tukang Besi (Donohue 1999), Seediq, Tagalog (Aldridge 2004)-- display a mix of prenominal, postnominal, and IH relatives, while also being VO. The Tagalog sentences below, for instance, show a head-initial relative (16a) and an internally-headed relative (16b) (Law 2014). These cases weaken the argument in favor of Cole's empty final head analysis, and turn the fact that Fijian has IHRCs in spite of being a VO language into a non-issue.

(16) a. guron-ng   dumating   kahapon
         teacher-LK PERF.arrive yesterday
         'The teacher who arrived yesterday.'
     b. dumating na guro   kahapon
         PERF.arrive LK   teacher yesterday
         'The teacher who arrived yesterday.'

The conclusion I arrive at is that Fijian weakly-quantified DPs are IHRCs. But the construction is not very productive in Fijian. In fact, I have not found it in other types of DP. Fijian relative clauses are normally postnominal, of the externally-headed type. Example (17) shows an object relative.

(17)   au   dau taleitaka [$_{DP}$ na sikoni [$_{CP}$ e   dau bulia   o   koya]]
        1.SG HAB like.TR   DET scone   3.SG HAB make-TR DET she
        'I like the scones that she makes.'

Fijian, then, is a "mixed type" language (like Tagalog) from the point of view of its relativization strategies. The use of IH relatives to express weak quantification, I suggest, is a highly specialized construction in Fijian, which may have been inherited from a more productive construction in an ancestral language. For some reason, however, IHRCs survived as the expression of weak quantification in Fijian. As I will explain in the next section, the reason behind this may be found in the general structure of the Fijian clause, and in the semantics of IH relatives.

7.   The semantic interpretation of IHRCs in Fijian

In a recent paper (Aranovich 2013), I argue that Fijian is a Pronominal Argument (PA) language. Except for incorporated nouns, pronouns, and proper

nouns, all other arguments are external to the VP, and are introduced by a pronominal affix. The PA hypothesis accounts for some observations about transitivity in Fijian. Transitive verbs are distinguished from intransitives by the presence of a verbal suffix, often of a *-Ca* shape. In fact, many two-argument verbs are also used as single-argument verbs when the suffix is absent. Thus, besides *bulu-ta* 'to bury it' and *lako-va* 'to go on/for something' one finds *bulu* 'to be buried' and *lako* 'to go' as the intransitive counterparts. Based on data like these, Schütz (1985) divides Fijian predicates into active (the same argument is subject of the transitive and the intransitive forms, e.g. *lako*) and stative (the subject of the intransitive and the object of the transitive are the same argument, e.g. *bulu*).

When the *-Ca* suffix is used, it is not necessary to specify a complement by means of a DP, as (18) shows.

(18) e   ronqo-ta tiko   na marama.
      3.SG hold-TR PROG DET woman
      'The woman is holding him.'

Pawley (1986) suggests that the *-Ca* ending consists of two suffixes: a transitive extension of the form *-Ci*, and a 3.SG pronominal suffix -a. In a sentence like (18), then, the meaning of the suffix *-a* provides the reference of the internal argument of the verb, saturating the valence of the predicate. When the object DP is overt, on the other hand, the suffix *-a* cross-references the DP. Under the hypothesis that Fijian is a pronominal argument language, object DPs are adjoined to a projection above VP, and are coindexed with the pronominal suffix *-a*.

Like many PA languages, Fijian allows for noun incorporation (NI). In (19), the object is immediately adjacent to the verbal root, without the transitive suffix, and the adverbial particles follow the incorporated noun. Moreover, the object has no article.

(19) keitou tara-vale   tiko.
      1.PL   build-house PROG
      'we are building houses.'

It is worth noting that Fijian has a third construction type in which the object also precedes the adverbial particles, but in which the verb keeps its transitive suffix. This happens when the object is a pronoun or a proper noun.

(20) o   bu-qu   kau-ti   **keitou** rawa
      DET grandmother-1.SG.POSS 3.SG.bring-TR 1.PL   POSSIBLE
      mai ki   vanua.
      DIR to   land
      'my grandmother managed to bring us to land.'

Notice that in this case the transitive suffix has the -Ci form. In Aranovich (2013) I argued that these object DPs are *bona fide* complements of V, and that therefore there can be no pronominal affix *-a* attached after the transitive suffix *-Ci*.

There are some consequences from an analysis of Fijian as a PA language for the semantics of DPs, and the article *na* in particular. In Aranovich (2013) I argue that na-DPs are indefinites. They are interpreted as indefinites in some positions (i.e. when performing a predicative function),[6] and they can also be unselectively bound by a strong adverbial quantifier like *kece*.

(21) a. na   gone levulevu o   Lavinia
        DET child big.big DET Lavinia
        'Lavinia is a chubby baby.'
     b. e   kani-a oti **kece** sara ga **na benu**   na   koli.
        3.SG eat-TR PERF all   INT LIM DET leftovers DET dog
        'the dog did finish eating all the leftovers indeed!'

The adverb *kece* 'all' deserves special mention here. As an adverbial particle, *kece* occurs in a position that is not necessarily adjacent to the DP it modifies, as shown in (21b). It must follow *oti* (and *rawa*), but it must precede the other adverbial particles. The fact that *kece* has a fixed place in the sequence of adverbial particles is evidence for its own adverbial nature.

Following Diesing (1992), I assume that sentences have a semantic representation consisting of an operator, its restriction, and its nuclear scope. Indefinite DPs introduce free variables into the semantic representation of a clause. If an indefinite is in the nuclear scope, the free variable it introduces is bound by an existential quantifier (existential closure), yielding a cardinality reading. But indefinites are ambiguous between this cardinality reading, which is nonpresuppositional, and a presuppositional reading, under which they behave like strongly quantified DPs. Presuppositional indefinites, Diesing suggests, must be outside the nuclear scope, inducing their own restriction. Diesing also notes a syntactic effect on the interpretation of indefinites: only those indefinite DPs that are inside the VP receive a cardinal interpretation, while indefinite DPs that are extracted from the VP (either overtly or at LF) are always

---

[6]   Weakly quantified sentences are not "idiomatic" in that position. A sentence like #*o koya e dua na vu-ni-wai* 'He is a doctor' sounds like a quick translation into Fijian of an English sentence, a typical and frequent error in broadcasting (Schütz 1985:334).

presuppositional. Na-DPs, then, have no quantificational force of their own, and are usually placed outside the VP. In that structural position they can only have a presuppositional reading, not a cardinal one. This presuppositional reading is the one that gives rise to the "definite" interpretation of na-DPs.

A weakly-quantified DP, then, must be presuppositional, since it is outside the scope of the existential quantifier associated with VP. This is precisely the kind of semantic structure that Basilico (1996) argues IH relatives must have. IHRCs are inherently quantificational. They are associated with an operator that binds a variable introduced by the head of the IHRC. I assume that the operator associated with a IHRC creates a generalized quantifier (an expression of type $<<e,t>,t>$). Applying this analysis to a weakly quantified DP like the one below would yield the semantic representation in (22).[7]

(22) a. e rua na tamata
   'two men'
  b. $(\lambda P)(\exists x)(man(x)$ & $two(x)$ & $P(x))$

For the analysis to work, however, it is crucial that the internal head be interpreted as an indefinite, since it needs to introduce a free variable for the existential operator associated with the IHRC to bind.[8] This is not a problem, since I have already provided independent evidence that na-DPs are indefinites. In fact, several languages allow the head of IHRCs to be marked with an indefinite pronoun. This is shown by the Lakhota example in (23).

(23) Mary **owižą wą** kağe ki he ophewathų
    Mary quilt a make the DEM buy.1.SG
    'I bought the quilt that Mary made.' (Williamson 1989:171)

There are, however, some Fijian examples that are still problematic for my approach. Some existential sentences are formed with the help of verbs usually employed to express bodily position (*tiko, tu*), instead of the adjectives denoting quantity or number, which I discussed before. Some examples are presented in (24).

---

[7] Basilico's analysis is an alternative to the "anaphoric head" analysis of Cole. In Cole's analysis, the internal head is anaphorically bound to an empty category which is the head of DP. Regarding the interpretation of the operator, Basilico suggests it is an operator yielding an expression of type e (an individual with a unique denotation), instead of the alternative interpretation as a generalized quantifier.

[8] Moreover, this head must move to a position outside the domain of existential closure (i.e. adjoined to V' or I'). This accounts for the optional movement of the internal head in some languages, discussed in Basilico 1996.

(24) a. e **tiko** e   so   na raisi?
    3.SG stay 3.SG some DET rice
    'is there any rice?' (Schütz 1985:101)

   b. e **tu** e   dua na vu-ni-niu   mai Serua.
     3.SG stay 3.SG one DET coconut.tree in   Serua
     'There is a coconut tree somewhere in Serua.' (Milner 1956:28)

In these existential sentences, the verbs *tu* and *tiko* are followed by weakly-quantified DPs, but these are not in a position where presuppositional indefinites should appear (i.e. in the scope of an existential operator). What I suspect is that the sentences in (24) do not have the semantics of true existential sentences, but this is an area of Fijian grammar that should receive more attention in the future.

8.   Conclusions

I have reviewed the evidence for a common origin of some existential clauses and weakly-quantified DPs. I suggested that this connection can be formalized in an analysis that treats weakly quantified DPs as IHRC's, in which the head is the subject of a relativized existential clause. The occurrence of adverbial particles in weakly quantified DPs is evidence of their clausal status, and provides an argument for the IHRC analysis of those constructions. I have also argued that there is nothing typologically abnormal about a language that has IHRCs and is head-initial, as is the case in Fijian. In fact, other Austronesian languages are like Fijian in having a mixed group of relative clauses (including IHRCs), and being of the VO type.

    From a semantic point of view, The analysis of weakly-quantified DPs as IHRCs is consistent with Basilico's (1996) hypothesis that the internal head must be indefinite, introducing a free variable to be bound by an operator associated with the head of DP. The resulting picture shows a language in which there are no nominal quantifiers to speak of. Strong quantifiers like *kece* are adverbial, while weak quantifiers are predicational. The only true common determiner in Fijian is *na*, which ends up as an indefinite. I have also shown that this system of expressing quantification works quite well for a PA language, an analysis I have suggested for Fijian (Aranovich 2013).

    I would like to speculate that the reason why Fijian preserved the IHRC strategy precisely for existential clauses, and nothing else, is because they allow for the expression of weak quantification without introducing a new category of quantificational determiners, optimally fulfilling an expressive need within the structure of a PA language. Of course, this assumes that Fijian has inherited its IHRCs from its Austronesian ancestry, and that it is not an innovation of the language. Whether this assumption is correct or not I will leave as a question for further research.

References

Aldridge, Edith. 2004. Internally headed relative clauses in Austronesian languages. *Language and Linguistics* 5: 99-129.

Aranovich, Raúl. 2013. Transitivity and polysynthesis in Fijian. *Language* 89: 465-500

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

Cole, Peter. 1987. The structure of internally headed relative clauses. *Natural Language & Linguistic Theory* 5: 277-302 .

Comrie, Bernard. 2008. Prenominal relative clauses in verb-object languages. *Language and Linguistics* 9: 723-733.

Diesing, Molly. 1992. *Indefinites*. Cambridge: MIT Press.

Donohue, Mark. 1999. *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.

Law, Paul. 2014. The syntax of Tagalog relative clauses. *AFLA 21*, University of Hawai'i at Mānoa.

Lehr, Marianne; James E. Redden, and Adama Balima (1966). *Moré basic course*. Washington, D.C.: Foreign Service Institute.

Milner, George Bertram. 1956. *Fijian grammar*. Suva: Government Press.

Pawley, Andrew. 1986. A reanalysis of Fijian transitive constructions. *Te Reo* 29.81–112 .

Schütz, Albert. 1985. *The Fijian language*. Honolulu: University of Hawaii Press.

Tellier, Christine. 1989. Head internal relatives and parasitic gaps in Moore. *Current approaches to African syntax* 6, ed. by Isabelle Haik and Laurice Tuller, 298-318. Dordrecht: Kluwer.

Williamson, Janis S. 1989. An indefiniteness restriction for relative clauses in Lakhota. *The representation of (in)definiteness*, ed. by Eric J. Reuland and Alice G. B. ter Meulen, 168-190. Cambridge: MIT Press.

CONSTITUENT ORDER AND PARSER CONTROL PROCESSES IN CHAMORRO[*]

Manuel F. Borja   Sandra Chung   Matthew Wagers
Inetnun Åmut yan   UC Santa Cruz   UC Santa Cruz
Kutturan Natibu   schung@ucsc.edu   mwagers@ucsc.edu

Based on the comprehension of prenominal versus postnominal relative clauses in Chamorro, we deconstruct the subject gap preference in relative clauses into two parser control processes.

1.   Introduction

Evidence from syntactic typology, on the one hand, and language comprehension, on the other, has revealed a privileged cross-linguistic status for relative clauses in which the relativized element is the subject. On the typological end, Keenan and Comrie's (1977) theory of NP accessibility in relativization includes the universal constraint that every language "must be able to relativize subjects" (1977:67). This constraint has been overwhelmingly confirmed by later typological research (although its status vis-à-vis morphologically ergative languages remains less clear; see e.g. Polinsky et al. 2012). Keenan and Comrie speculated that their Accessibility Hierarchy "*directly reflects the psychological ease of comprehension* [emphasis theirs]", and therefore "subjects are easier to relativize than any other position" on the Hierarchy (1977:88-89). And indeed, psycholinguistic research on various languages has shown that relative clauses with subject gaps are easier to interpret than relative clauses with nonsubject gaps (see e.g. Vasishth et al. 2013). In particular, if the gap in a relative clause is ambiguous, it is usually interpreted as a subject gap—an effect we call the *subject gap preference* (SGP). What interests us here is a strand of research that suggests that the SGP might not hold in certain languages in which relative clauses precede their head NPs (e.g. Yip and Matthews 2007 on Cantonese, Hsiao and Gibson 2003 on Mandarin, and Ekmeçi 1990 on Turkish).[1] Much of this research is controversial

[1] Also relevant is Carreiras et al. 2010 on Basque, but the situation in Basque is complicated by morphological ergativity.

(compare Lin and Bever 2006, Hsu et al. 2009, and Vasishth et al. 2013 on Mandarin; and Özge et al. 2010 on Turkish); all of it focuses on languages that allow *only* prenominal relative clauses.

Is the SGP sensitive to the order of relative clause and head NP, and if so, why should that be? Here we investigate this question by probing the comprehension of relative clauses in Chamorro, an Austronesian language of the Mariana Islands. Like some other Austronesian languages (notably, Tagalog; see Aldridge 2004), Chamorro allows a relative clause to precede or follow the head NP. This pattern is typologically rare: just 31 of the 751 languages whose dominant relative clause order is reported in the *World atlas of language structures* (Dryer and Haspelmath 2013) to be Rel-N and/or N-Rel permit both N-Rel and Rel-N orders.[2] In addition, Chamorro has relative clauses in which the gap can be interpreted as a subject gap or a direct object gap. In an experiment conducted in the U.S. Commonwealth of the Northern Mariana Islands (CNMI) in 2013, we attempted to determine whether the language has a uniform preference for interpreting these gaps, and how such a preference interacts with the order of the relative clause with respect to the head NP. We found that (a) relative clause order does affect where and how comprehenders locate the gap in a relative clause; but (b) whatever the relative clause order, there is also an early preference for subject gaps. Our account of these findings deconstructs the SGP into two parser control processes that are familiar from previous literature: the Active Filler Strategy (Frazier 1987, Frazier and Clifton 1989, Aoshima et al. 2004, Wagers 2014) and the drive to link the verb to its subject. We claim that hypotheses that are readily diagnosed by evidence in the input strongly guide these processes and their interaction.

2. Background: Relative Clauses in Chamorro

Chamorro is a head-initial language that allows a range of null arguments. In clauses, the verb comes first, followed by its arguments and adjuncts, which are ordered flexibly with respect to one another. The default word order is Verb Subject Object. The verb is inflected for subject-verb agreement via a prefix or proclitic that also signals mood and transitivity.[3]

(1) a. Mañaibuk   i palåo'an lemmai.
   AGR.AP.cook.in.coconut.milk the woman breadfruit
   'The woman cooked breadfruit in coconut milk.'

---

[2] We exclude languages that also allow relative clauses that are internally headed, double-headed, or correlative.

[3] The following abbreviations are used: AGR 'agreement', AP 'antipassive', C 'complementizer', L 'linker', LOC 'local case', OBJ 'object', PROG 'progressive', SUBJ 'subject', UNM 'unmarked case', WH 'Wh-Agreement'.

b. Un   gigimin   i   hanum.
   AGR   drink.PROG   the   water
   'You are drinking the water.'

In DP's, the determiner comes first, followed by NP. The point of interest here is that certain modifiers, including relative clauses, can precede or follow NP (Chung 1998, Chung and Ladusaw 2006). When a relative clause is prenominal, NP is inflected with the prenominal form of the linker, *na* (glossed L below). The relative clauses in (2) are enclosed in brackets.

(2) a.   Ågang   atyu   i   [mañaibuk   lemmai]   na
          call   that   the   AGR.AP.cook.in.coconut.milk   breadfruit   L
          palåo'an.
          woman
          'Call that woman who cooked breadfruit in coconut milk.'
    b   Estagui'   i   risuttan   i   CCR...   put   i   [un   gigimin]
         here.is   the   result.L   the   CCR   about   the   AGR   drink.PROG
         na   hånum.
         L   water
         'Here are the results of the CCR...about the water that you drink.'
                              (Commonwealth Utility News, July 2014, p. 8)

When a relative clause is postnominal, it begins with a complementizer that can be viewed as simultaneously spelling out the linker. This complementizer is realized as *i* when the head NP is preceded by a demonstrative plus the linker (see (3a)), and as *ni* otherwise (3b).

(3) a.   Ågang   atyu   na   palåo'an   [i   mañaibuk
          call   that   L   woman   C   AGR.AP.cook.in.coconut.milk
          lemmai].
          breadfruit
          'Call that woman who cooked breadfruit in coconut milk.'
    b.   Impottånti   esti   na   infotmasion   put   i   hanum   [ni   un
          AGR.important   this   L   information   about   the   water   C   AGR
          gigimin].
          drink.PROG
          'This information about the water that you drink is important.'
                              (Commonwealth Utility News, July 2014, p. 8)

The gap in the relative clause is always a DP. When the gap is an argument, the verb of the relative clause registers its grammatical relation via Wh-Agreement, a special agreement that supersedes the normal subject-verb agreement. Wh-Agreement is not overt when the gap is an intransitive subject;

see (2a) and (3a), where the gap is the subject of an antipassive verb. It is realized optionally when the gap is a direct object; see (2b), (3b), and (4).

(4) un guma' [ni finahån-ña / ha fåhan si Juan]
     a house C WH[OBJ].buy-AGR / AGR buy UNM Juan
     'a house that Juan bought'

When the gap is a transitive subject and the relative clause is realis, Wh-Agreement is usually realized as the infix *-um-*, as (5) shows.

(5) a. i [k**um**uentutusi yu' nigap] na palåo'an
       the WH[SUBJ].speak.to.PROG me yesterday L woman
       'the woman who was speaking to me yesterday.'
    b. i lalåhi [ni k**um**akassi i ma'estra]
       the men C WH[SUBJ].tease.PROG the teacher
       'the men who were teasing the teacher.'

But here too, Wh-Agreement is not obligatory. It is possible for a realis relative clause whose gap is a transitive subject to have normal subject-verb agreement instead of Wh-Agreement. Compare (5) with (6).

(6) a. i lalåhi [ni ma kakassi i ma'estra]
       the men C AGR tease.PROG the teacher
       'the men who were teasing the teacher'
    b. i supertyphoon Kim [ni ha danchi ham mågi gi
       the supertyphoon Kim C AGR hit us here LOC
       December 3]
       December 3
       'the supertyphoon Kim that hit us here on December 3' (from a letter)

The overall result is that some relative clauses are systematically ambiguous: their gap can be interpreted as a subject gap or a direct object gap. Importantly, this holds true whether the relative clause precedes or follows the head NP, as can be seen from (7).

(7) a. Hu ågang atyu i [ha kadididak i biha]
       AGR call that the AGR tickle.PROG the old.lady
       na påtgun.
       L child
       'I called that child who was tickling the old woman.' (*subject gap*)
       'I called that child who the old woman was tickling.' (*object gap*)

b. Hu   ågang   atyu   na   påtgun [i   ha   kadididak   i   biha].
   AGR   call   that   L   child   C   AGR   tickle.PROG   the   old.lady
   'I called that child who was tickling the old woman.' (*subject gap*)
   'I called that child who the old woman was tickling.' (*object gap*)

This systematic ambiguity is leveraged in the research reported on below.

3.   Experimental Method, Design, and Participants

We constructed an experiment to probe how speakers interpret the gaps in relative clauses like (7), and how their interpretive preferences interact with relative clause order.

3.1.   Method

The task involved matching an audio stimulus to a picture via touch-tracking on a tablet computer (cf. mouse-tracking; Freeman and Ambady 2010). The experimental software was created in OpenSesame (Mathôt et al. 2012) with an Android back-end and deployed on Google Nexus 10 tablets. Participants heard an audio request to move a small icon (the puck) centered at the bottom of the tablet screen to one of two target pictures, positioned at the left and at the right of the top of the screen. The audio request began with the carrier frame in (8) and continued with a relative clause.

(8)   Chonnik   i   floris   guatu   gi …
       push   the flower   to.there LOC
       'Push the flower over to…'

The target pictures showed the same individuals playing different participant roles in the event named by the verb of the relative clause. For instance, if the verb was *paini* 'comb', participants would have to choose between the pictures in Figure **1** and hear an audio request that continued with the postnominal (and ambiguous) relative clause in (9). Figure **2** schematizes the major trial events.

(9)   atyu   na biha   [i   ha   papaini   i   palåo'an].
       that   L   old.lady C   AGR   comb.PROG   the   woman
       'that old woman who is combing the woman / who the woman is combing'
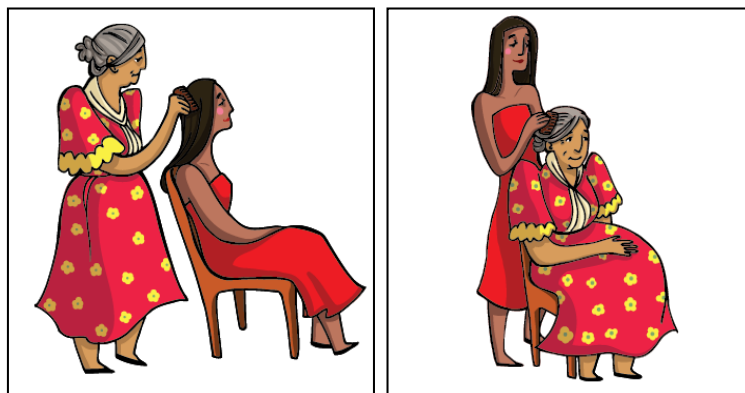
Figure 1.   Picture choices for the verb *paini* 'comb'

An old lady (*biha*) is depicted either as the agent of combing (left) or the patient (right) in this interaction with a younger woman (*palåo'an*).

Participants who moved the puck to the picture on the left interpreted the relative clause in (9) as having a subject gap (as shown in Figure 2D); participants who moved the puck to the picture on the right interpreted the relative clause as having an object gap.

3.2.   Design

Our experiment followed a 3 × 3 design that crossed Relative Clause type with Verb type within the relative clause. The three Relative Clause types were prenominal, postnominal, and headless. In prenominal relative clauses, the gap precedes the head NP (henceforth, the *filler*); in postnominal relative clauses, the gap follows the filler; in headless relative clauses—not discussed below— the filler is unpronounced and so it is unclear whether the gap precedes or follows it. The three Verb types, all realis, were transitive, passive, and Wh-Agreeing. As shown earlier, relative clauses whose transitive verbs have normal subject-verb agreement are ambiguous: they can be interpreted as having a subject gap or a direct object gap. Relative clauses whose verbs are passive or Wh-Agreeing are unambiguous: the gap associated with a (realis) passive verb corresponds to the internal argument, while the gap associated with a Wh-Agreeing verb has the grammatical relation that this special agreement registers.

A. Trial begins with flashing puck.

B. Picture revealed with **audio onset**

C. Response initiated at any time following audio onset.

D. Desired picture selected and confirmed.
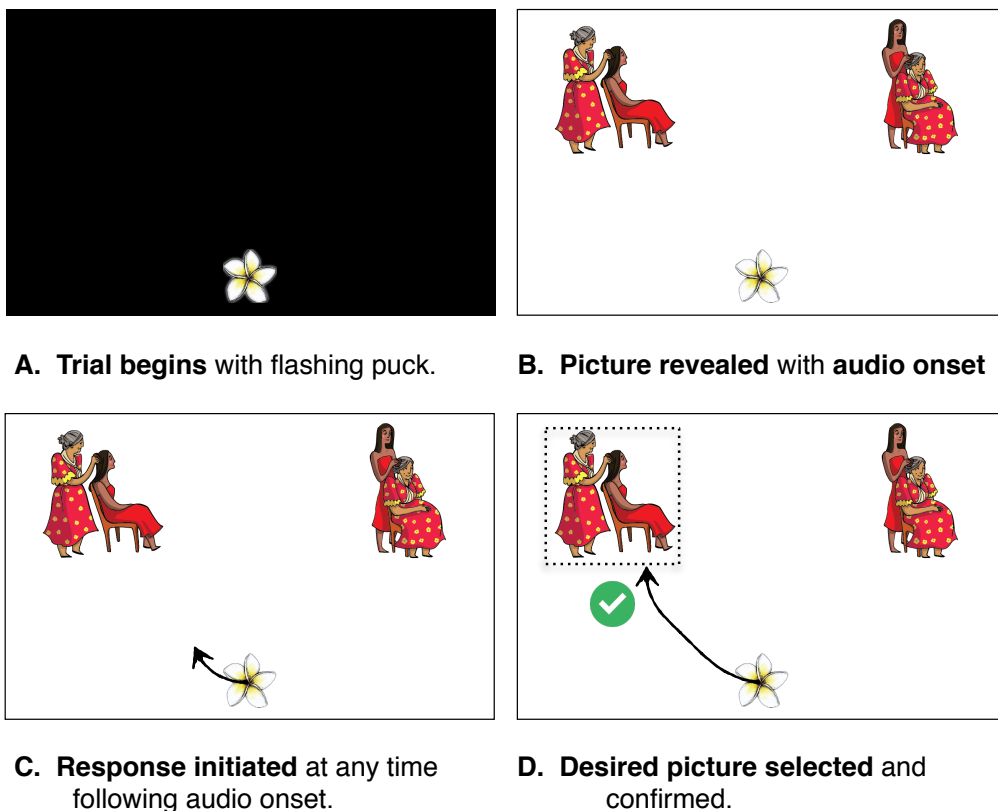
Figure 2. Major trial events

Panel A. Participants are alerted that a new trial is about to start by a black screen and a flashing puck; 2 sec. In this case, the puck is a plumeria blossom (*floris måyu*). *Panel B.* Audio onset is synchronized to presentation of the two pictures. *Panel C.* Participants may begin moving the puck at any point after audio onset. Once participants touch and start to move the puck, we register their trajectory 'initiation time'; trajectory indicated by the crooked arrow. *Panel D.* Once the puck enters the borders of the photo—as defined by the minimum rectangle that contains all the picture's non-transparent pixels—participants may confirm their choice by pressing a 'check' button. They may also change their mind and move the puck to the other picture. At any time after audio offset, participants can press a 'repeat' button (not shown) to hear the sentence again.

Our stimuli (= the audio requests) were drawn from 36 item sets containing relative clauses constructed from 18 reversible transitive verbs (e.g. *dengkut* 'peck', *lasa* 'massage', *galoppi* 'jump over'), combined with 18 pairs of nouns that were equal in humanness and animacy and counterbalanced for position

(e.g. *sihik* 'kingfisher' and *gåyu* 'rooster', or *låhi* 'man' and *palåo'an* 'woman'). These item sets were read by Borja and recorded with a Zoom H4 portable recorder (Zoom Corporation, Tokyo, Japan). Borja also recorded instructions in Chamorro that introduced participants to the task, walked them through some simple examples, and explained how to repeat a stimulus, how to advance to the next stimulus, and when the task had been completed. The entire experiment, including the instructions, took about 15 minutes to complete. Consistent with our past practice, the initial contact with participants, the experiment itself, and the debriefing were conducted in Chamorro (although during the debriefing, some participants switched to English). At the end, each participant was given an information sheet in Chamorro and English about the research.

3.3. Participants

During two weeks in September 2013, 135 Chamorro speakers in the CNMI participated in the experiment: 58 on Rota, 65 on Saipan, and 12 on Tinian. Demographic data were not collected for 2 speakers. The ages of the other speakers ranged from 20 to 70; the median age was 42. Fifty-eight of these speakers were male. Participants received an 8 Gb flashdrive for their participation.

4. Results

4.1. Error Rates

Data from 17 speakers were removed because there were too many errors in the unambiguous conditions. For each speaker, we calculated two measures to assess their ability to comprehend unambiguous relative clauses. *Syntactic subject d-prime* assessed the ability to discriminate syntactic subjects from direct objects in phrase structure, and *thematic agent d-prime*, the ability to discriminate agents from themes in argument structure. For each measure, we estimated the d-prime score by scaling percent correct in one condition against the error rate in a matched condition, as defined in Table 1. We removed participants whose syntactic subject d-prime or semantic agent d-prime was below 0. In other words, a participant had to show *some* discrimination to be included in further analysis. This disjunctive policy identified 9 and 6 participants respectively for exclusion, plus 2 participants who did not meet the criterion in either category. Table 1 (next page) gives further details about this analysis.

We also removed trials with items containing the verbs *hongngang* 'startle' and *tattiyi* 'follow'. These verbs led to error rates above 30% on unambiguous stimuli. In the case of *hongngang,* debriefings from multiple participants indicated confusion over how to interpret the target pictures.

Briefly, for unambiguous relative clauses, we see that participants perform best on extractions marked by subject Wh-Agreement, better than they do either with derived subjects (Passive conditions) or with objects (extractions marked by object Wh-Agreement).

4.2.   Results

The interpretation rates for ambiguous relative clauses in the picture-matching task reveal that the SGP does indeed hold in Chamorro. However, this effect is sensitive to relative clause order. 94% of the ambiguous postnominal relative clauses in our experiment were interpreted as having subject gaps—clear evidence of the SGP. However, just 43% of the ambiguous prenominal relative clauses were interpreted with a subject gap. The size of the contrast between these two interpretation rates suggests that in this language at least, the SGP does not hold absolutely when the gap precedes the filler.

Discrimination	Hits: p_H	False alarms: p_{FA}	
Semantic agent	WH[SUBJ]	Passive	
Syntactic subject	Passive	WH[OBJ]	
D-prime formula	$Z(p_H) - Z(p_{FA})$		*d-prime*

Table 1. Discriminative measures of error on unambiguous conditions

> The 'hits' column lists the conditions from which *correct response* rate was taken for the hit rate. The 'false alarms' column lists the conditions from which the *error rate* was taken for the false alarm rate. To calculate the passive rates, we used only item sets that had the corresponding Wh-Agreement condition. The function Z is the inverse of the cumulative Gaussian. Perfect scores (0 or 1) were corrected with a 0.05 increment. Score distribution for each measure, by participant, is indicated by the boxplots in the right column (maximum = 3.28).

However, inspection of dialectal variation (by island) suggests an interesting asymmetry between subject gap interpretations and object gap interpretations. As Figure **3** shows, speakers from all islands show a strong SGP for postnominal relative clauses—above 90%. However, there is considerably more variation for prenominal relative clauses. On average, prenominal relative clauses show a relatively weaker SGP for all speaker groups. But this consistent relative difference is anchored by different absolute interpretation rates for prenominal relative clauses that vary by island. On Saipan, the interpretation of

prenominal relative clauses is evenly split (53% SGP), whereas on Rota there is an absolute preference for object gap interpretations (35% SGP). Finally, on Tinian speakers show an absolute SGP even for prenominal relative clauses (66% SGP).

Table 2 shows the analysis of this pattern as a mixed-effects logistic regression of subject gap interpretation on RC type and island. There are two crucial statistically discernible effects: in RC Type—prenominal relative clauses lead to fewer subject gap interpretations—and in the interaction between RC Type and Island—for speakers from Rota, prenominal relative clauses lead to *even fewer* subject gap interpretations. Details of the regression are in the table caption.
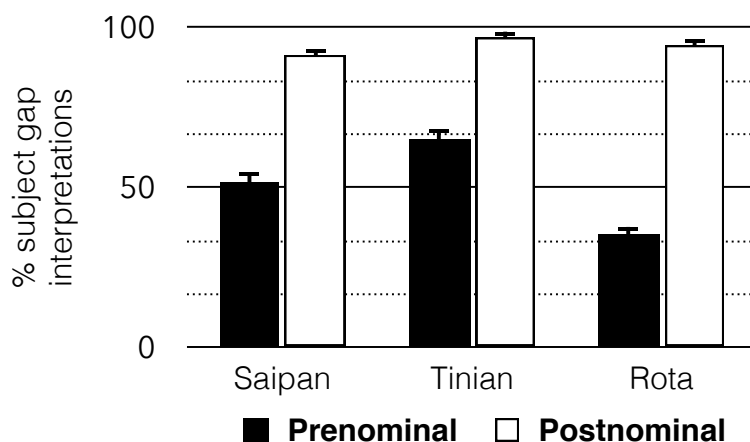
Figure 3. By-island variation in subject gap preference

Average SGP, expressed as a percentage, in each of the three inhabited islands of the CNMI. Filled bars correspond to prenominal relative clauses, open bars correspond to postnominal. Standard error of the mean over participants is shown.

Coefficient	Estimate	Std. err.	z	
Intercept	1.47	.19	7.6	***
RC Type	-3.80	.34	-11.2	***
Island	-0.11	.38	-0.3	
RC Type × Island	-1.94	.60	-3.3	**
Random effects	Item	RC Type × Island		
	Subject	1 [Intercept only]		

Table 2. Logistic regression of subject gap rate on RC Type & Island

Mixed-effects regression with random slopes and intercepts for items, and random intercepts for subjects. Saipan and Tinian speakers were collapsed into one group, since there were only 12 speakers from Tinian. Contrasts were centered and unit scaled: for RC Type, the positive coefficient (0.5) was assigned to prenominal relative clauses; for Island, the positive coefficient was assigned to the Rota group. Coefficient estimates are expressed as logits. Stars are interpreted thus ** *p* < .01; *** *p* < .001. Calculated in R (R Core Team 2013), using the *lme4* package (ver. 1.0-5, Bates et al. 2013).

To sum up the main results: our data show that there is a strong SGP for ambiguous postnominal relative clauses, and a much weaker object gap preference for ambiguous prenominal relative clauses. Moreover, this object gap preference varies considerably depending on the speaker group, while the SGP for postnominal relative clauses varies only minimally. Put another way: the subject gap preference is either really strong, or else it is only somewhat weaker or occasionally reversed.

5. Discussion

5.1 The Proposal

The relatively high, constant SGP in postnominal relative clauses, and the more varying SGP in prenominal relative clauses, suggest to us an explanation in terms of multiple factors. These factors either act in harmony, as in postnominal relative clauses, or in competition, as in prenominal relative clauses. We propose that two familiar factors are operative and that they interact with different strengths and different timecourses. We conceive of these factors as closely reflecting grammatical principles of locality. We assume the parser attempts to optimize the satisfaction of these principles given the information it has. Thus, we adopt an explanation in the spirit of principle-based parsing (Berwick et al. 1991, Merlo and Stevenson 2000).

The first principle is essentially the Active Filler Strategy (AF; Frazier 1987): namely, the parser prioritizes linking fillers to gaps. If the parser recognizes a filler, then it prefers to project a gap in an argument position rather than wait for a pronounced argument. In doing so, it creates a representation that satisfies more grammatical principles. We do not hold, therefore, that the AF needs to literally be a 'strategy'. But a representation that includes a filler but no gap is not legal and the parser will seek to extend the representation to complete the filler-gap dependency.

The second principle we call Subject Satisfaction (SS): the parser prioritizes identifying the subject of a clause. Because subjecthood is correlated with many principles and constraints in grammar, this is really just a cover term.

But for the sake of argument, we assume that the relevant principle in Chamorro is something like: insert into the specifier of T the DP whose agreement features are compatible with the verb's. This principle may be the same across languages or it might differ; it may be reflected in other reinforcing constraints, like 'satisfy the Extended Projection Principle'; or it may compete in languages with other types of morphosyntactic organization (e.g. ergativity; see Carreiras et al. 2010, cf. Polinsky et al. 2012).

Given these two principles, let us first consider what happens in the case of postnominal relative clauses in Chamorro. When the head noun and the relative clause complementizer are encountered, AF requires the parser to postulate a gap. The to-be-encountered relative clause may have an argument in object position, but it will also, of necessity, have a subject. Therefore, the best way of satisfying AF is the most local way: postulate a subject gap. The inflected verb is encountered next. At this point, SS requires that a DP be inserted whose agreement features are compatible with those of the verb. However, given that AF extended the representation with a subject gap, SS is already satisfied, because the filler is linked to a subject gap. The agreement features of the subject are inherited from the filler and can thus be checked against the verb's agreement features. The remainder of the relative clause is consistent with this analysis and activates no countervailing principles. Thus, the subject gap interpretation is arrived at quickly and unanimously. We give an illustration of how this parse is achieved in Figure 4 (on the next page).

Now we consider what happens in the case of prenomial relative clauses. We assume that when the determiner *i* is encountered, immediately followed by a subject-verb agreement morpheme, the comprehender recognizes the relative clause. This is because the sequence *i*-plus-AGR can occur in just two constructions in Chamorro: a prenomial relative clause and a sentential complement (meaning '(the fact) that…'). This ambiguity is resolved in favor of the relative clause given two pieces of evidence in our stimuli: first, the syntactic context requires the larger constituent containing *i* to be a DP in the local case, and, second, the experimental context requires a choice to be made between two pictures. Adopting the relative clause analysis activates AF and thus leads to the postulation of a gap. However, the filler itself has not yet been directly encountered and at this point can only be hypothesized to exist, so we assume that the parser's course of action in postulating a gap may be correspondingly less confident. Concomitantly, because the verb's agreement features have been encountered, this triggers SS. This means that the parser is looking for a subject at roughly the same time as it is postulating a gap. AF and SS can thus be tentatively satisfied by postulating a gap in the specifier of T. This satisfaction is tentative in the sense that the identity of the filler is unknown, and thus it is also unknown whether its agreement features are compatible with the verb's. When the first overt DP inside the relative clause is encountered, that DP offers a better way of satisfying SS because its agreement features are known. This second analysis necessitates linking the gap to the

direct object, an outcome that is also consistent with the verb's transitivity. We give a partial illustration of how this parse is achieved in Figure 5 (on the next page).

In short, prenominal relative clauses trigger a competition between two analyses – or, a rapid succession of analyses – in a way that postnominal relative clauses do not. This is because the location of the relative clause with respect to the head NP in a postnominal relative clause allows the linkage between filler, gap, and the DP in the specifier of T to be checked immediately upon encountering the verb. But in a prenominal relative clause, the filler comes last and thus must initially just be *hypothesized* to exist, usually with no evidence about its contents. Consequently, the linkage between filler, gap, and subject position cannot be checked right away. This is why the subject gap interpretation for prenominal relative clauses arises early in comprehension but will tend to give way to the object gap interpretation. The object gap interpretation accrues greater strength because it links the subject with the overt DP inside the relative clause.
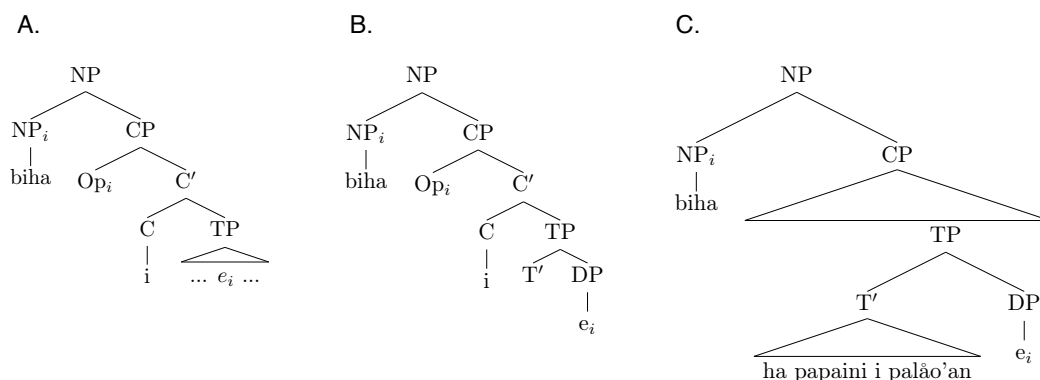
Figure 4. Optimal parse of a postnominal relative clause

String: … *biha i ha papaini i palåo'an* … (see (9))

Tree A. After encountering the noun *biha* and the complementizer *i*, there is enough information to extend the tree with a relative clause, which triggers AF. *Tree B.* The optimal position to postulate an empty category, without further information about the verb's transitivity, is the local one: subject position [specifier shown on the right, to be consistent with Chamorro word order]. *Tree C*. Encountering AGR-plus-V (*ha papaini)* triggers SS. This principle is immediately satisfied because AF has inserted a DP in specifier of T, which inherits features from NP_i. The remaining DP, *i palåo'an,* is consistent with the projected representation.

A. B. C.
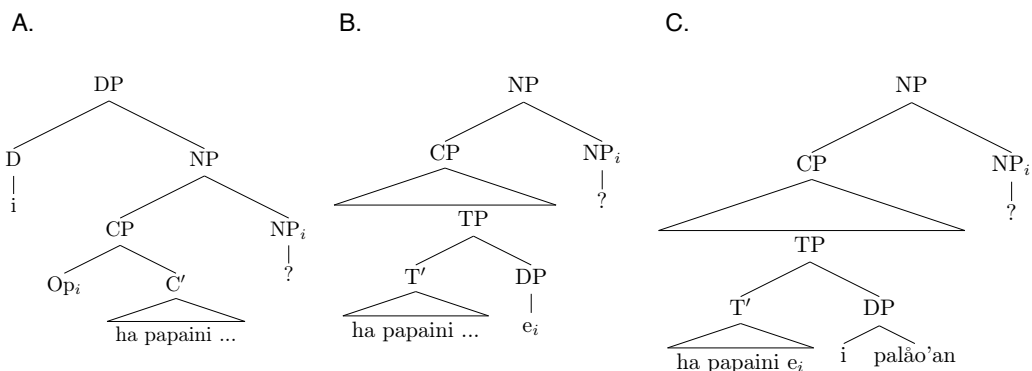
Figure 5. Optimal parse of a prenominal relative clause

String: … *i ha papaini i palåo'an …*

Tree A. Encountering determiner *i* and AGR-plus-V (*ha papaini)* is sufficient to postulate a relative clause. Although NP$_i$ has not yet been heard, AF and SS are operative. *Tree B*. An empty category is postulated in subject position, weakly satisfying both AF and SS. *Tree C*. Encountering the DP *i palåo'an* leads to reanalysis, because SS is better satisfied by a DP whose features are known. The empty category is now in object position.

5.2.   Further Support: Trajectory Dynamics

An important difference between the parsing of the two relative clause types is that, in postnominal relative clauses, the subject gap interpretation can be maintained as the preferred interpretation over time as information accumulates. In prenominal relative clauses, however, there is a strong motivation to reanalyze away from the subject gap interpretation toward an object gap interpretation. Nonetheless, the subject gap interpretation will tend to be arrived at earlier, even if it may also be fleeting.

If the comprehender does *not* reanalyze toward an object gap interpretation, this may be for several reasons. The motivation to reanalyze away from the subject gap interpretation arises because of the near simultaneity of satisfying AF and SS and the higher value placed on checking agreement with a DP whose features are known. Therefore, individual variation at the level of speaker or trial may be found in (i) the weight placed on satisfying AF versus SS, (ii) the perceived informativity of constituents that are encountered versus merely hypothesized, and (iii) the onset of SS and the concomitant linking of agreement features to the subject with respect to encountering the filler. For example, a sample parse may postulate a gap in subject position well in advance of attempting to link agreement to the subject DP. Reanalysis to the object gap interpretation in this instance may not be sufficiently motivated as incoming information is quickly integrated.

In our analysis of comprehenders' touch trajectories, we found evidence that subject gap responses enjoyed a temporal advantage even in prenominal relative clauses. Table **3** shows two measures: the initiation time on the touchscreen (cf. Figure 2C); and the average speed of the response, that is, the speed with which the puck is moved from starting position to the selected picture.[4] In both relative clause types, subject gap interpretations were initiated sooner and executed faster.

Trajectory measure	Prenominal	Postnominal
Difference in initation time (object gap minus subject gap)	162 ms	388 ms
Difference in average speed (subject gap minus object gap)	34 px/s	179 px/s

Table 3. Subject advantage in two measures of trajectory dynamics

The 'subject advantage' was computed for two measures – initiation time and average speed – based on all trajectories initiated between -1000 ms to 2000 ms with respect to audio offset. For both measures, the effect of interpretation (subject v. object) was significant in a linear mixed-effects model with full random effects for subjects and an 'interpretation' intercept for items (which was the maximal, convergent model); the interaction of interpretation and RC Type was not significant.

We further considered just the speaker group from Rota, who showed a stronger object gap preference in prenominal relative clauses. For such relative clauses, they showed no advantage for subject gaps in average speed, but there was a numerically-comparable advantage in initiation times (129 ms, $t = 1.4$, *n.s.*).

In sum, data from touch trajectories reveal that a preference for object gaps in prenominal relative clauses did not translate into a timing advantage – at least in terms of initiating and completing the response. This is consistent with our theory. We predict subject gap interpretations will tend to be computed by the comprehender sooner than object gap interpretations. Further, this pattern in the data suggests that early subject gap interpretations do not always enter into competition with object gap interpretations.

---

[4] Each trajectory was normalized by computing a cubic spline of 15 points on its *x, y* and *t* coordinates. Speed was estimated by calculating the Euclidean distance between successive, normalized (*x,y*) coordinates and dividing it by the corresponding difference in normalized *t*. Average speed was then computed per trajectory.

5.3.   Conclusion

In this study, we asked whether the order of an ambiguous relative clause with respect to the head NP affects how speakers resolve the ambiguity. We found that the speakers' ultimate preference does indeed depend on order. In postnominal relative clauses, there was a strong preference to interpret an ambiguous relative clause as containing a subject gap. In prenominal relative clauses, there was a weak preference for object gaps. The object gap preference varied considerably depending on the speaker group, but the subject gap preference did not. Furthermore, analysis of touch trajectories showed that there was a temporal advantage for subject gap interpretations, regardless of relative clause type. Our account of these facts invoked two parser control processes: Subject Satisfaction, which prioritizes identifying the subject, and the Active Filler Strategy, which prioritizes completing the link between a filler and its gap. The difference between relative clause types was derived as a consequence of the difference in timing and strength of SS and AF. In postnominal relative clauses, AF temporally precedes SS and supplies a known constituent for insertion in the specifier of T. In prenominal relative clauses, AF and SS are cotemporaneous. The fact that the filler has not yet been encountered allows the relative-clause-internal DP to be construed as the subject. It is an important component of our account that encountered constituents act with greater strength than merely hypothesized ones.

What is responsible for the by-island variation in the SGP? Namely, why do speakers from Rota more strongly prefer object gap interpretations for prenominal relative clauses? Recall that transitive relative clauses need not be ambiguous. When overt Wh-Agreement is present, it determines the position of the gap. Although subject Wh-Agreement is optional on all three islands, we conjecture that the loss of this special inflection has gone farther on Saipan and Tinian than on Rota. Thus, for speakers on Rota, the *absence* of subject Wh-Agreement could be taken as an informative cue about what the gap is *not*. Parsing a prenominal relative clause could be particularly sensitive to this cue, since there is greater uncertainty at the onset of the relative clause.

6.   References

Aldridge, Edith. 2004. Internally-headed relative clauses in Austronesian languages. *Language and Linguistics* 5:99-129.

Aoshima, Sachiko, Colin Phillips, and Amy Weinberg. 2004. Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51:23-54.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2013. Lme4: linear mixed-effects models using Eigen and S4. R package version 1.0-5. Online: http://CRAN.R-project.org/package=lme4.

Berwick. Robert, Steven Abney, and Carol Tenny, eds. 1991. *Principle-Based Parsing*. Kluwer Academic Publishers.

Carreiras M., J. Duñabeitia, M. Vergara, I. de la Cruz-Pavía, and I. Laka. 2010. Subject relative clauses are not universally easier to process: Evidence from Basque. *Cognition* 115:79-92.

Chung, Sandra. 1998. *The design of agreement: Evidence from Chamorro*. Chicago: University of Chicago Press.

Chung, Sandra and William A. Ladusaw. 2006. Chamorro evidence for compositional asymmetry. *Natural Language Semantics* 14:325-357.

Dryer, Matthew S. and Martin Haspelmath, eds. 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info, accessed on April 20, 2014.)

Ekmeçi, Ö. 1990. Acquisition of relativization in Turkish. Paper presented at the Fifth International Conference on Turkish Linguistics, SOAS, University of London, England.

Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5:519-560.

Frazier, Lyn and Charles Clifton, Jr. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4:93-126.

Freeman, J.B. and N. Ambady. 2010. MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods* 42: 226-241.

Hsiao, Franny and Edward Gibson. 2003. Processing relative clauses in Chinese. *Cognition* 90:3-27.

Hsu, Chun-Chieh Natalie, Gabriella Hermon, and Andrea Zukowski. 2009. Young children's production of head-final relative clauses: Elicited production data from Chinese children. *Journal of East Asian Linguistics* 18:323-360.

Keenan, Edward L. and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63-99.

Lin, Chien-Jer Charles and Thomas G. Bever 2006. Chinese is no exception: Universal subject preference of relative clause processing. Paper presented at The 19th Annual CUNY Conference on Human Sentence Processing, CUNY Graduate Center, New York, NY.

Mathôt, Sebastiaan, D. Schreij, and J. Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44:314-324.

Merlo, Paola and Suzanne Stevenson. 2000. Lexical syntax and parsing architecture. *Architectures and mechanisms for language processing,* eds. M. Crocker and M. Pickering, 161-188. Cambridge University Press.

Özge, Duygu, Theodoros Marinis, and Deniz Zeyrek. 2010. Comprehension of subject and object relative clauses in monolingual Turkish children.

Proceedings of the Fourteenth International Conference of Turkish Linguistics (ICTL), eds. S. Ay, Ö. Aydın, İ. Ergenç, S. Gökmen, S. İşsever, and D. Peçenek. Wiesbaden: Harrasowitz Verlag.

Polinsky, Maria, Carlos Gomez Gallo, Peter Graff, and Ekaterina Kravtchenko. 2012. Subject preference and ergativity. *Lingua* 122:267-277.

R Core Team. 2013. R: Language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Online: http://www.r-project.org.

Yip, Virginia and Stephen Matthews. 2007. *The bilingual child: Early development and language contact*. Cambridge: Cambridge University Press

Vasishth, Shravan, Zhong Chen, Qiang Li, and Gueilan Guo. 2013. Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS One* 8(10): e77006. doi:10.1371/journal.pone.0077006

Wagers, Matthew W. 2014. Memory mechanisms for *wh*-dependency formation and their implications for islandhood. *Experimental syntax and island effects*, eds. Jon Sprouse and Norbert Hornstein, 161-185. Cambridge University Press.

ILOKANO FREE RELATIVES WITH *MAN**

James N. Collins
Stanford University
jamesnc@stanford.edu

I provide a compositional semantics for a particular kind of free relative con-
struction (analogous to English *wh–ever* phrases) in Ilokano formed with the
particle *man*. I argue that Ilokano free relatives with *man* (*m*FRs) in episodic
sentences are definite, presupposing existence and uniqueness of the *m*FR's
referent. This distinguishes *m*FRs from other kinds of "headless" DPs in
Ilokano that are formed without *wh*-morphology. With a variety of diagnos-
tics, I demonstrate that other kinds of non-*wh*-DPs do not presuppose exis-
tence or uniqueness. I further show that *m*FRs in episodic contexts imply
that the conversational participants are collectively unable to pin down with
certainty the referent of the free relative.

1.   Introduction

This paper describes the semantics of free relatives marked with the morpheme
man (abbreviated as *m*FRs) in Ilokano (Northern Philippine), as in (1).[1]

(1)   *nanglukat ti   tawa   [ti   sinoman   nga   immuna a*
       AP.open   DET window DET who-MAN COMP OP.first COMP
       simmangpet]
       OP.arrive
       'Whoever arrived first opened a window.'

In short, I argue that an *m*FR in an episodic context is a definite (following
Jacobson 1995), presupposing the existence and uniqueness of an individual who
instantiates its descriptive content. In (1), the bracketed *m*FR presupposes that
there is a unique individual who arrived first. Further, I argue that an *m*FR supplies
a not-at-issue meaning component of uncertainty. An utterance of (1) implies
that the conversational participants are mutually unable to uniquely pin down the
identity of the referent of the *m*FR. I show how this condition on the use of an

[1]Abbreviations: AP actor pivot; COMP / C complementiser; DET / D determiner; ERG ergative;
GP goal pivot; NEG negative particle; OP object pivot; PERF perfect; PL plural; PROG progressive;
SG singular; STAT stativity marker; TOP topic marker

*m*FR gives rise to so-called ignorance implications observed in previous work on free relatives with *wh–ever* in English (e.g., Dayal 1997; von Fintel 2000).

I compare *m*FRs to another kind of DP constituent with an embedded clause, namely a headless relative as in (2). Headless relatives lack morphosyntactic features of *m*FRs, namely the *wh*-item, the particle *man* and the overt complementiser *(ng)a*. A bare clause is simply combined with a determiner, *ti* in (2).

(2) *nanglukat ti   tawa   [ti   immuna a   simmangpet]*
     AP.open   DET window DET OP.first COMP OP.arrive
     'A certain one who arrived first opened a window.'

I show that these headless relatives do not impose the same contextual felicity constraints that *m*FRs impose. In particular, they may be used in contexts where the uniqueness and existence of an individual instantiating the descriptive content of the headless relative are not presupposed. I capture this discrepancy by proposing a semantics for *wh*-morphology in *m*FRs which encodes presuppositions of uniqueness and existence. This accounts for the presuppositional semantics of *m*FRs, which demonstrate the requisite *wh*-morphology, and also accounts for the lack of presuppositional semantics in headless relatives, which lack *wh*-morphology.

I also propose that the uncertainty implication of *m*FRs is contributed by the particle *man*. I provide a lexical semantics for *man* which determines that the property denoted by the descriptive content is not held by any one individual across some contextually supplied modal base. Following Lauer (2009), in episodic contexts (to which I devote my attention in this paper), this modal base is the mutual public beliefs of the conversational participants. I show how this semantics for *man* links with its usage in contexts other than in free relatives, in particular its use as a marker of surprise (a mirative), and a marker of politeness in imperatives.

2.   *m*FRs as definites

This section explores the proposal that *m*FRs are semantically definite and compares them in this regard to headless relatives. I show that *m*FRs pass diagnostics suggesting that they are only felicitous in contexts which entail the uniqueness and existence of their referent. I also show that headless relatives are infelicitous in the same contexts.

I use these results to motivate a particular view of the interpretation of *wh*-expressions. Previous accounts of English free relatives (which lack a determiner) suggest that their definite semantics is derived by covert type-shifting or a phonologically null definite determiner. I suggest that the definite semantics is imposed by its *wh*-morphology, thereby alleviating the need for any additional machinery.

The data presented here focus on *m*FRs and headless relatives appearing with the determiner *ti*. *ti* signals that the argument is core (as opposed to oblique),

marking the sole argument of an intransitive, and the agent and patient of a transitive. Furthermore, the *m*FRs and headless relatives discussed in this paper are for the most part the "pivot" of their containing clause. Ilokano demonstrates thematic role marking morphology on the verb, typical of Philippine languages. A morpheme on the verb corresponds to a thematic role associated with the event denoted by the verb (e.g., agent, patient, benefactive, instrument, etc.). The pivot is the DP which denotes the participant bearing the thematic role in question. The pivot has a number of particular properties including wide scope with respect to sentential operators such as negation and conditionals. In comparing the semantics of *m*FRs and headless relatives, I will be careful to compare pivots with pivots, and non-pivots with non-pivots, in an effort to keep properties associated with pivothood constant across the compared sentences. A discussion of the compositional semantics of pivots and non-pivots is briefly elaborated on in §2.4.

2.1   Uniqueness

To begin, I will examine whether *m*FRs or headless relatives entail that the descriptive content is uniquely instantiated by an individual or group of individuals. The uniqueness entailment of a noun phrase with descriptive content P, abbreviated as UNIQUENESS throughout, is spelled out informally below.

(3)   UNIQUENESS: If there is an individual x who has property P then at most one individual has property P.

If *m*FRs or headless relatives entail UNIQUENESS as defined in (3), they should allow at most one (plural or singular) individual to instantiate the descriptive content. If UNIQUENESS holds, the predicative content (in our example "opened a window") should apply *exhaustively* to the entire plurality of individuals who arrived first. If UNIQUENESS doesn't hold, it should be possible for the predicative content to apply non-exhaustively to the set of individuals instantiating the descriptive content (some first-arrivers "opened a window" and some didn't).

The example below illustrates a clear contrast: an utterance of an *m*FR is infelicitous in a context in which the predicate does not apply exhaustively to all individuals instantiating the descriptive content of an *m*FR (4a). On the other hand, an utterance of a headless relative is felicitous in the same contexts (4b). The speaker's judgement is slightly complicated by the uninformativity of (4b) in such a context: although the speaker judged (4a) as false and (4b) as true, she cites (4b) as a non-complete description of the scenario.

(4)   [*Context:* A lot of people requested tickets but I only gave
            tickets to some, but not all the people.]

    a.   *#inikkak   ti   ticket [ti   sinoman   nga*
          GP.PERF.give.1SG DET ticket DET who-MAN COMP
          dimmawat]
          AP.PERF.request
          #'I gave a ticket to whoever requested one.'

    b.   *inikkak   ti   ticket [ti   dimmawat]*
          GP.PERF.give.1SG DET ticket DET AP.PERF.request
          'I gave a ticket to a certain one who asked.'[2]

I take this paradigm to be evidence that *m*FRs entail UNIQUENESS, while headless relatives lack this entailment, being judged as true in contexts where the predicate non-exhaustively applies to individuals instantiating their descriptive content.

2.2   Existence

Next, I will examine whether *m*FRs or headless relatives entail that an individual instantiates their descriptive content. In addition I will examine *how* this entailment arises: is it an at-issue entailment, a presupposition, a conventional implicature etc? I abbreviate the entailment as EXISTENCE, spelled out for some descriptive content P below as simple existential quantification. UNIQUENESS entails at most one individual has property P, while EXISTENCE entails at least one individual has property P. The combination of UNIQUENESS and EXISTENCE entails exactly one individual has property P.

(5)   EXISTENCE: There is at least one individual x who has property P.

    Both *m*FRs and headless relatives entail that their descriptive content is instantiated by an individual. For example, neither can be followed with continuations which deny the existence of an instantiator.

(6)   *nanglokat   iti   tawa   [**ti   sinoman** nga   immuna*
        AP.PERF.open DET window DET who-MAN COMP AP.PERF.first
        a   simmangpet] #ngem awan   immuna
        COMP AP.PERF.arrive but   not.exist arrive
        'Whoever arrived first opened a window, (# but no one arrived).'
        ⤳ *There is a person that arrived first*

(7)   *nanglokat   iti   tawa   [ti   immuna   a*
        AP.PERF.open DET window DET AP.PERF.first COMP
        simmangpet] #ngem awan   immuna
        AP.PERF.arrive but   not.exist arrive

---

[2]Speaker comment on (4b): but it's not the whole truth, there's a sea of people who asked for tickets and you didn't give them. It's true but it's not true that whoever asked got a ticket.

'A certain one who arrived first opened a window, (# but no one
arrived).'
⤳ *There is a person that arrived first*

So, at least in positive, episodic contexts, both *m*FRs and headless relatives
convey EXISTENCE. However, if we systematically apply the investigative toolkit
supplied by the literature on not-at-issue meaning (in particular Tonhauser et al.
2013), we arrive at the conclusion that *m*FRs and headless relatives convey EX-
ISTENCE in quite different ways. I will show that these diagnostics point towards
*m*FRs encoding EXISTENCE as a presupposition, while headless relatives encode
EXISTENCE in their asserted content. This puts *m*FRs, but not headless relatives,
in a semantic category with English DPs headed by *the* which are generally taken
to presuppose EXISTENCE and UNIQUENESS (e.g., Strawson 1950; Sharvy 1980;
Barwise and Cooper 1981; Link 1983; Partee 1987; Chierchia 1998).

To establish the nature of these uniqueness and existence implications, I use
diagnostics from Tonhauser et al. (2013). An essential auxiliary definition is the
notion of m-positive and m-neutral contexts (where m is a proposition).

(8) m-POSITIVE AND m-NEUTRAL CONTEXTS: An m-positive context is an
      utterance context that entails or implies m. An m-neutral context is an
      utterance context that entails or implies neither m nor $\neg m$.   (Tonhauser et
      al. 2013:75)

Given the definition of m-positive and m-neutral contexts we can establish
whether or not an entailment of some expression imposes a strong contextual fe-
licity constraint. The use of the term 'strong contextual felicity constraint' follows
Tonhauser et al. 2013, referring to a constraint determining that the expression is
only felicitously uttered if the context of utterance entails a particular proposition.

(9)   STRONG CONTEXTUAL FELICITY: Let S be an atomic sentence that
      contains trigger t of projective content m.

      (i) If uttering S is acceptable in an m-neutral context, then trigger t does
           not impose a strong contextual felicity constraint with respect to m.

      (ii) If uttering S is unacceptable in an m-neutral context and acceptable
           in a minimally different m-positive context, then trigger t imposes a
           strong contextual felicity constraint with respect to m.   (Tonhauser et
           al. 2013:76)

By using these diagnostics, we can determine that an utterance of an *m*FR is
acceptable in an EXISTENCE-positive context, a context entailing EXISTENCE,
(10a), but unacceptable in an EXISTENCE-neutral context (10b).

(10)   a.   [**EXISTENCE-positive context**: Juan and Maria approach a closed room which is not sound-proof. Maria walks inside for a minute. Juan hears singing inside. Maria then comes out and says:]

napintas	*ti*	*boses*	*[ti*	*sinoman*	*nga*
STAT.beautiful	DET	voice	DET	who-MAN	COMP

agkankanta]
AP.PROG.sing

'Whoever is singing has a beautiful voice.'

    b.   [**EXISTENCE-neutral context**: Juan and Maria approach a closed, sound-proof room. Maria walks inside for a minute, then comes out and says:]

#napintas	*ti*	*boses*	*[ti*	*sinoman*	*nga*
STAT.beautiful	DET	voice	DET	who-MAN	COMP

agkankanta]
AP.PROG.sing

'Whoever is singing has a beautiful voice.'

In (10a), both Juan and Maria's belief states entail the existence of an individual who is singing. In (10b), Juan's belief state does not entail the singer's existence (the sound-proof room may or may not contain a singer, as far as Juan believes). As Juan's belief state in (10b) neither entails EXISTENCE nor its negation, it is EXISTENCE-neutral. Thus, the mutual beliefs of Juan and Maria are EXISTENCE-neutral. As the *m*FR is infelicitous in this context, I conclude EXISTENCE is a strong contextual felicity constraint on the use of an *m*FR.

An utterance of a headless relative in an EXISTENCE-neutral utterance context is perfectly felicitous (11). I therefore conclude that EXISTENCE is not a strong contextual felicity constraint on the use of a headless relative.

(11)   [**EXISTENCE-neutral context**: Juan and Maria approach a closed, sound-proof room, Maria walks inside for a minute, then comes out.]

napintas	*ti*	*boses*	*[ti*	*agkankanta]*
STAT.beautiful	DET	voice	DET	AP.PROG.sing

'A certain one who is singing has a beautiful voice.'

These facts fall out of an analysis where the semantics of an *m*FR carries EXISTENCE as a felicity condition on its utterance context: the use of *m*FRs are only felicitous if EXISTENCE holds in the utterance context. My preliminary hypothesis is that EXISTENCE is a presupposition of an *m*FR but not of a headless relative.

This analysis predicts that the existential entailment of *m*FRs should "project": it should scope out of sentential operators such as negation and conditionals.

2.3 Projection of Existence

To diagnose whether or not EXISTENCE is projective when triggered by an *m*FR, I again use diagnostics in Tonhauser et al. 2013. The tests for the projectivity of a proposition p differ based on whether or not p is a strong contextual felicity constraint or not. EXISTENCE is a strong contextual felicity constraint on the use of a *m*FR. We therefore use the following diagnostic:

- If the *m*FR **remains unacceptable** in an EXISTENCE-neutral context even if we negate S or put S as the antecedent of a conditional, then EXISTENCE '**projects**' through negation/conditionals.

- If the *m*FR **becomes acceptable** in an EXISTENCE-neutral context when we negate S or put S as the antecedent of a conditional, then EXISTENCE does not '**project**'.

The data below shows that the former is true: the use of an *m*FR is still unacceptable in a context which is EXISTENCE-neutral, even when the sentence is negated (12) or in the antecedent of a conditional (13).

(12)   a. [**EXISTENCE-positive context**: Juan and Maria are entering a cabin. They know that someone has been there before (the door was unlocked), all the windows are closed. Juan:]

haan	*a*	*nanglokat*	*iti*	*tawa*	*[ti*	*sinoman*	*nga*
NEG	COMP	AP.open	DET	window	DET	who-MAN	COMP

immuna	*a*	*simmangpet]*
AP.first	COMP	AP.open

'It's not the case that whoever arrived first opened a window.'

     b. [**EXISTENCE-neutral context**: Juan and Maria are entering a cabin. They don't know whether anyone has been there before. All the windows are closed. Juan:]

#haan	*a*	*nanglokat*	*ti*	*tawa*	*[ti*	*sinoman*	*nga*
NEG	COMP	AP.open	DET	window	DET	who-MAN	C

immuna	*a*	*simmangpet]*
AP.first	C	AP.open

#'It's not the case that whoever arrived first opened the window.'

(13)   a. [**EXISTENCE-positive context**: Juan and Maria are approaching a cabin. They are very hot, and they know someone has arrived at the house and cooled the house down. Juan:]

no	*nanglokat*	*iti*	*tawa*	*[ti*	*sinoman*	*nga*	*immuna*
if	AP.open	DET	window	DET	who-MAN	COMP	AP.first

a	*simmangpet], namaliis*		*diay*	*balay*
COMP	AP.arrive,	PERF.cold	that	house

'If whoever arrived first opened a window, the house is cool.'

b. [**EXISTENCE-neutral context**: Juan and Maria are approaching a cabin. They are very hot, but they don't know if anyone has arrived at the house yet and cooled the house down. Juan:]

#no	*nanglokat*	*iti*	*tawa*	*[ti*	*sinoman*	*nga*
if	AP.open	DET	window	DET	who-MAN	COMP

immuna	*a*	*simmangpet], namaliis*	*diay*	*balay*
AP.first	COMP	AP.arrive, PERF.cold	that	house

#'If whoever arrived first opened a window, the house is cool.'

Based on these data we can conclude that the existence implication triggered by an *m*FR "projects" through negation and conditionals. Diagnosing the projective behaviour of a headless relative is a little simpler as there is no strong contextual felicity constraint. We merely need to see if EXISTENCE is still implied in negative and conditional sentences. If the implication of EXISTENCE survives, it is projective. The following data demonstrate that if a headless relative is in the pivot position of the sentence, it still entails EXISTENCE even if the sentence is negated or conditionalised.

In (14) and (15), the verb *nanglokat* bears the actor-pivot morpheme *nang-* and therefore the headless relative is the "pivot" (being in the actor thematic role).

(14)   Negation
        [**EXISTENCE-neutral context**: Maria is entering a cabin. She doesn't know whether anyone has been there before. All the windows are closed. Juan is already there, he says:]

haan	*a*	*nanglokat*	*iti*	*tawa*	*[ti*	*immuna*	*a*
NEG	COMP	AP.open	DET	window	DET	AP.first	COMP

simmangpet]
AP.open

'It's not the case that a certain one that arrived first opened a window.'
⤳ *Someone arrived first*

(15)   Conditionals
        [**EXISTENCE-neutral context**: Maria is approaching a cabin. She is very hot, but she doesn't know if anyone has arrived at the house yet and cooled the house down. Juan:]

no nanglokat iti tawa [ti immuna a simmangpet],
if AP.open DET wind. DET AP.first C AP.arrive,

namaliis diay balay
cold   that house

'If a certain one that arrived first opened a window, the house is cool.'
⤳ *Someone arrived first*

We therefore conclude that both *m*FRs and headless relatives project through negation and conditionals, but only *m*FR requires EXISTENCE to be an entailment of the utterance context. As a final diagnostic for the presuppositional status of EXISTENCE for *m*FRs, we can test whether the entailment undergoes filtering, as per Karttunen (1973). As the following filtering sentence cancels the EXISTENCE entailment of the *m*FR, we have evidence that the EXISTENCE entailment behaves much like a presupposition.

(16)   [**Context**: Maria doesn't know whether or not her family has bought flowers today, but she knows they have good taste in flowers]

      no adda sabongda, napintas [ti aniaman daydiay]
      if have flower.3PL, beautiful DET what-MAN there

      'If they have any flowers, then whatever flowers are there are beautiful.' *does not entail* "There are flowers that they bought."

The preliminary hypothesis is therefore that the existence implication is a presupposition triggered by *m*FRs, but not by headless relatives. However, both *m*FRs and headless relatives scope out of negation and conditionals. I therefore suggest that *m*FRs are presuppositional definites in the sense of Strawson 1950, but headless relatives are indefinites whose scope is constrained by Philippine-type verbal morphology.

3.   A semantics for *ti*

The scope of an indefinite is determined by whether or not the indefinite is a pivot or not. As stated earlier, a DP's status as pivot is determined by whether the verb bears morphology matching the thematic role of the DP. If the DP is the pivot, it will necessarily take wide scope with respect to sentential operators like negation and conditionals. If the DP is the logical subject (the intransitive sole argument or the agentive argument of a transitive, regardless of whether it is a pivot), it also necessarily takes wide scope.

If a transitive verb has actor pivot morphology, the non-pivot patient is necessarily a narrow scope indefinite. Obliques which are non-pivots are ambiguously narrow or wide scope. This is summarised in the table below.

Thematic role	If pivot:	If not-pivot
Intransitive sole argument	always wide scope	always wide scope
Transitive actor	always wide scope	always wide scope
Transitive patient	always wide scope	always narrow scope
Oblique	always wide scope	unspecified (both ok)

Non-pivot transitive patients must be non-specific indefinites. Inherently specific pronouns and proper names may not be in the non-pivot transitive patient position. Non-pivot transitive patients are marked by either the determiner *ti* or *iti*, but I have not yet identified a semantic difference associated with this choice. My working hypothesis is that *iti* is composed morphologically of a prepositional case marker *i-* and the determiner *ti*. In (17), the non-pivot transitive patient *iti tawa* is interpreted as an indefinite obligatorily scoping below the negative particle *haan*.

(17) *haan nga   **nang**lokat [i-ti tawa]$_O$ [ti   baket]$_A$*
      not   COMP **AP.**open DET window DET woman
      'A (particular) woman didn't open any windows.'
      $\exists x[\textbf{woman}(x) \land \neg \exists y[\textbf{window}(y) \land \textbf{open}(x, y)]]$

To handle the wide scope indefinite facts, I employ free variable choice functions in the style of Reinhart (1997). The hypothesis is that *ti* takes a property-denoting argument P, and introduces a free variable choice function f. Depending on the value of f, f will take P as its argument and return one individual member of P. The choice function therefore shifts the property to an e-type expression and thus can compose with the rest of sentence.

(18) *ti* $\rightsquigarrow \lambda P. f(P)$, where $f \in D_{\langle \langle e,t \rangle, e \rangle}$

The clausal component of a headless relative is a property type. On composing with the determiner *ti*, the choice function selects an individual who instantiates the property.

(19) *immuna a simmangpet* $\rightsquigarrow$ **arrived-first**
      ti(immuna a simmangpet) $\rightsquigarrow f(\textbf{arrived first})$

The scoping properties of indefinites are handled by existentially closing the free variable choice function. If the DP is a pivot, or an agent, the choice function variable is existentially bound at the assertion level. If the DP is an non-pivot transitive patient, the choice function variable is existentially bound at the VP level, ensuring the indefinite scopes below negation, conditionals and other operators outside of the VP.

4. The modal implication of *m*FRs

Besides EXISTENCE and UNIQUENESS, in episodic contexts, *m*FRs imply UNCERTAINTY. By UNCERTAINTY, I mean that the conversational participants are mutually unable to uniquely identify the referent of the *m*FR. I show this implication in action where the preceding context sets up certainty about the free relative's referent. In such cases the *m*FR is infelicitous.

(20) *#Amok   nga   ni   Carlos ti   nagtakaw ti   alahas ken*
     know-1SG COMP DET Carlos DET AP.steal the jewel and
     timmakas idi Miyerkoles [ti   sinoman   nga   nagtakaw ti
     AP.escape on Wednesday DET who-MAN COMP AP.steal DET
     alahas].
     jewel
     #'I know that Carlos stole the jewels and whoever stole the jewels
     escaped on Wednesday.'[3]

By the same token, we also find that *m*FRs are infelicitous with a *tinnaga ket DP* ('namely DP') parenthetical (cf. Dayal 1997).

(21) *#Timmakas [ti   sinoman   nga   nagtakaw ti   alahas]*
     AP.escape DET who-MAN COMP AP.steal DET jewel
     (tinnaga ket   Carlos) idi Miyerkoles.
     OP.name TOP Carlos   on Wednesday
     #'Whoever stole the jewels, namely Carlos, escaped on Wednesday.'[4]

In (20)-(21), the *m*FR signals ignorance on the part of the speaker as to the referent's identity. However, the use of an *m*FR does not always imply uncertainty on the part of the speaker. In the following "quiz show" context, ignorance can be on the part of the hearer.

(22) [Context: A asks B to guess what kind of animal is inside the box A is
     holding. A drops a banana in the box and listens to the sound of the
     animal happily eating the banana. A gives a clue:]

     kayat [ti   aniaman   nga   adda iti   kahon] ti   saba
     want DET what.MAN COMP exist DET box   DET banana
     'Whatever's in this box likes bananas.'

---

[3]*Speaker comment:* I could say that if I'm trying to piece a puzzle, and I have two pieces of knowledge and I know that Carlos took it, and I know that whoever escaped on Wednesday, therefore I know that Carlos escaped on Wednesday.

[4]*Speaker comment*: Sounds awkward, it's like, hey guys, we know for a fact that it's Carlos who did it, but I'm still using *sinoman*! I can actually say that with my friends and I'm like mocking, I can perhaps say I would say something ... Ok, I'm telling you this, but I'm pretending not to know, but I'm saying "you know who", I drop the bomb at the end and say it's Carlos.

In other uses, the speaker and the hearer may each be certain about the referent of the *m*FR but disagree on the referent's identity, as in the following conversation (adapted from Condoravdi to appear).

(23)   A: *immuna   nga   simmangpet ni   Juan*
     AP.arrive COMP AP.first   DET Juan
     'Juan arrived first.'

    B: *Saan! immunga nga   simmangpet ni   Maria*
     no   AP.arrive COMP AP.first   DET Maria
     'No! Maria arrived first.'

    A: *nanglokat ti   tawa   [ti sinoman   nga immuna   nga*
     AP.open   D   window D   who.MAN   C   AP.arrive C
     simmangpet]
     AP.first
     'Whoever arrived first was the one who opened the window.'

I hypothesise that UNCERTAINTY is a non-at-issue meaning component of *m*FRs. I diagnose this by its insensitivity to sentential operators like negation (24). In contexts which do not support UNCERTAINTY, i.e., contexts where the conversational participants are reasonably assured of the identity of the free relative referent, the *m*FR is infelicitous even in negated and conditionalised sentences.

(24)   *Amok   nga   ni   Carlos ti   nagtakaw ti   alahas*
    know-1SG COMP DET Carlos DET AP.steal   the jewel
    'I know that Carlos was the one who stole the jewels.'

    *#ken **haan nga**   timmakas idi Miyerkoles [ti   **sinoman***
    and   not   COMP AP.escape on   Wednesday DET who-MAN
    nga   nagtakaw ti   alahas].
    COMP AP.steal   DET jewel
    #'...and it's not the case that whoever stole the jewels escaped on Wednesday.'

At this stage I lack the empirical data to conclusively determine whether UNCERTAINTY is better characterised as a presupposition or conventional implicature in the style of Potts 2005. I characterise UNCERTAINTY as a felicity condition on the use of an *m*FR in a given utterance context, though this could be altered to a Potts-style alternate meaning dimension if new data support such an analysis.

5.   Analysis

In this final section I sketch an analysis of Ilokano free relatives and their compositional semantics. The syntactic analysis adapts the Guilfoyle et al. (1992) analysis of Tagalog, with a right branching specifier subject. Following the Paul (2000), Potsdam (2006) analysis of Malagasy, I analyse *wh*-questions as cleft structures: the *wh*-expression in predicate position and the prejacent clause as a sentential subject. (25) is a syntactic analysis of an *m*FR.

(25)

DP

D — ti

IP

PredP — sino man

CP

Op_i [$_{C'}$ nga [$_{IP}$ immuna t_i nga simmangpet]]
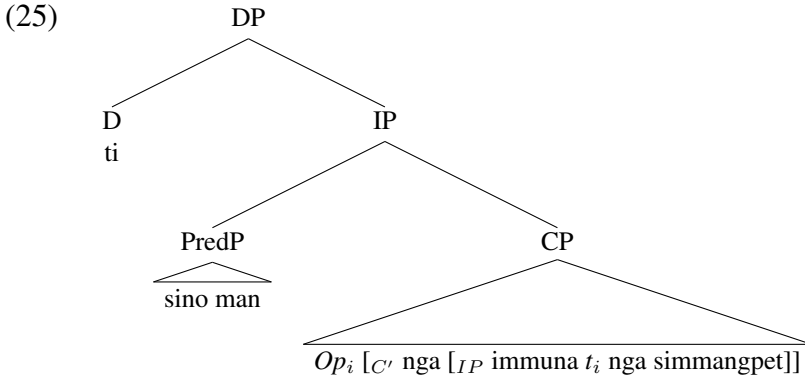
The internal CP constituent forming the "subject" of the cleft structure has an intensional property type.

(26)   *nga immuna nga simmangpet* $\rightsquigarrow \lambda x \lambda w.\textbf{arrived-first}_w(x)$

Recall that *m*FRs, but not headless relatives encode a definiteness presupposition. Both kinds of DPs use the determiner *ti*, but only *m*FRs contain *wh*-morphology. To capture this semantic discrepancy, I encode the definiteness presupposition of an *m*FR on its *wh*-morphology.

The following is a semantics for the *wh*-morphology found in an *m*FR. It must encode a restriction property, e.g., *sino*, 'who', applies only to humans (or at least animates), while *ania*, 'what', applies to inanimates. To capture this, the lexical semantics of *sino* includes an intensional property **human**, and *ania* includes **thing** and so on. The *wh*-item is an expression of type $\langle\langle e, st\rangle, \langle e, st\rangle\rangle$, a function from properties to properties. In (25), the *wh*-item sits in the predicative position, and takes a *wh*-cleft CP as its subject. The *wh*-cleft is a property type and serves as the first argument of the *wh*-item. The *wh*-item *sino* takes the property denoted by the cleft, and returns the property of being the unique maximal human instantiator of that property. For example, the *wh*-item *sino*, takes a property P, and returns the property of being the unique maximal member of P which is human.

(27)   **sino** $\rightsquigarrow \lambda P \lambda x \lambda w.x = \iota y[\textbf{human}_w(y) \wedge P_w(y)]^5$

---

[5] $\iota y[P_w(y)]$ presupposes $\exists x[P_w(x) \wedge \forall z[P_w(z) \rightarrow z \leq x]]$ (i.e., EXISTENCE and UNIQUENESS hold for P). $\iota y[P_w(y)]$ denotes the unique individual y such that y holds P at w.

The semantics of *man* encodes a felicity condition, such that its use is only felicitous if the condition is met. The condition is one of "modal variation". For some input property P, *man* implies that for every individual x in the domain, it is not true that x holds P in every world in a contextually supplied modal base, or equivalently, there is some world in which x does not hold P. It is a *wh*-item-modifier, and therefore is a function which takes a *wh*-item-type expression and returns another *wh*-item-type expression. *man* is an identity function, returning the same *wh*-item as its input, but adds the not-at-issue felicity condition of modal variation.

(28) $man \rightsquigarrow \lambda\mu_{est,est}\lambda P : \forall y[\exists v \in W[\neg P(y)(v)]] . \lambda x\lambda w.\mu(P)(x)(w)$

The felicity condition contains a free variable, modal base W (of type $\langle s, t \rangle$). In episodic contexts, W will anaphorically refer to the conversational common ground (the mutual public beliefs of conversational participants). The end result is that in episodic contexts, *m*FRs entail that for any given individual y, the conversational participants are mutually unable to say with certainty that y holds property P. The following is the result of composing *man* with *sino*, yielding the complex *wh*-item *sinoman*.

(29) $man(sino)$
         $\rightsquigarrow \lambda P : \forall y[\exists v \in W[\neg P(y)(v)]] .$
                                        $\lambda x\lambda w.x = \iota z[\textbf{human}(z)(w) \wedge P(z)(w)]$

(29) states that for any property P, the use of *sinoman* is felicitous iff for all individuals y, there's a world in W in which y doesn't hold P. It's at-issue content is is the property of being the unique human (or animate) instantiator of P. Composing this function with our clausal argument gives the desired free relative semantics.

(30)   *man(sino)(nga immuna a simmangpet)*
         felicity condition: $\forall x[\exists v \in W[\neg\textbf{arrived-first}(x)(v)]]$,
         at-issue content: $\lambda y\lambda w.y = \iota z[\textbf{human}(z)(w) \wedge \textbf{arrived-first}(z)(w)]$

The free relative in an episodic context imposes the felicity condition that for all individuals x, the conversational participants are unable to say with certainty that x arrived first. The at-issue content of the free relative is that it denotes the property of being the unique, maximal individual which is human and arrived first. As *man(sino)(nga immuna a simmangpet)* is a property type, it can combine with (an intensional version of) *ti* which shifts it to an e-type expression.

5.1 Extending the analysis of *man*

The implication of UNCERTAINTY triggered by the use of an *m*FR is captured formally by a felicity condition of modal variation across some contextually supplied modal base W. Following Lauer (2009), in episodic contexts (to which I have devoted my attention in this paper), I take this modal base is the mutual public beliefs of the conversational participants. In episodic contexts, the referent of the *m*FR is inconsistent across the mutual public beliefs of the conversational participants. This amounts to saying that for any individual, the conversational participants are collectively unable to say that the individual instantiates the descriptive content of the FR. This gives rise to the uncertainty implications, similar to those observed in Dayal (1997), von Fintel (2000) among others for English *wh–ever* expressions.

This modal variation analysis of the semantics of *man* makes links with other seemingly disparate uses of *man* in Ilokano besides its use in *m*FRs. For example, *man* can occur as a mirative particle (strictly in main clauses) marking the speaker's surprise about the propositional content of the utterance (31).

(31) *Napudot **man** ita*
       hot   MAN   this
       'It is surprisingly hot today.'

Further, the particle may be used as a marker of politeness in imperatives, as in (32).

(32) *Manang   Biday, ilukatmo   **man** 'ta   bintana.*
       older.sister Biday, OP.open=2SG.ERG MAN that window
       'Older sister Biday, please open that window.'

In Collins (2014), I suggested a unified account of the surprise marker and the politeness marker uses of *man*. In both cases, *man* takes a propositional argument p and returns p again just in case a felicity condition is met: that p is not true in every world across some modal base W. Where in the free relative use, the modal base is the mutual beliefs of the conversational participants, in the mirative and politeness marker uses, the modal base is the set of worlds in which the speaker's expectations are fulfilled.

- *Mirative* in (31):
  at-issue content = *it's hot today*,
  not-at-issue content = *it is not the case that in all the worlds in which my default expectations are true it is hot today*

- *Imperative* in (32):
  at-issue content = *my preferences are that you open the window*,
  not-at-issue content = *it is not the case that in all the worlds in which my default expectations are true you open the window*

The politeness effect of *man* in imperatives arises through a face-saving act of negating the expectation that the listener complies with speakers preferences. The analysis in (30) unites the semantics of the mirative, politeness marker and FRs. *man* encodes a non-at-issue meaning component of ensuring that in some worlds within a modal base, the prejacent is false.

Having set up the semantics of mirative and politeness *man* as imposing a modal variation condition, the link between these uses and the *m*FR use of *man* emerges. The felicity condition imposed by the *m*FR use of *man* in (30) takes the modal variation condition imposed by mirative/politeness-marker *man*, and iterates the condition over every individual in the domain. This way of conceptualizing the felicity condition in (30) is sketched in (33). It determines that for any individual in the domain, there is a possibility given the conversational participants' mutual beliefs that the individual did not arrive first.

(33) $[\![man(who)(arrived\ first)]\!]$ is felicitous iff
     $\exists w \in W[Juan\ didn't\ arrive\ first\ at\ w]$ and
     $\exists w \in W[Maria\ didn't\ arrive\ first\ at\ w]$ and
     $\exists w \in W[Carlos\ didn't\ arrive\ first\ at\ w]$ and
     ...

When the modal variation condition is iterated over every individual in the domain and closed under conjunction, the use of an *m*FR is felicitous just in case there is no individual such that the conversational participants are mutually certain that they arrived first, giving rise to the UNCERTAINTY implication. The proposal is that the basic function of *man* is to introduce a not-at-issue meaning component of modal variation, and this generalizes to its uses as a mirative, politeness marker, and a marker of uncertainty in a free relative.

6.   Conclusion

To summarise the analysis: the existence and uniqueness of a referent must be a common ground belief of conversational participants for a felicitous use of an *m*FR in an episodic context. The existence and uniqueness implications are *not* contributed by the determiner: headless relatives with the same determiner do not show the same contextual felicity constraints, leading us to conclude they are introduced by the *wh*-item. The modal implication of uncertainty is introduced by a felicity condition on the *man* particle, independently motivated by *man*'s use as a mirative/politeness marker.

References

Barwise, Jon and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4:159–219.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339–405.

Collins, James N. 2014. Expectations and the Ilokano mirative. Ms. `http://web.stanford.edu/~jamesnc/collins2014-man.pdf`

Condoravdi, Cleo. to appear. Ignorance, Indifference, and Individuation with *wh-ever*. *Epistemic indefinites*, ed. by L. Alonso-Ovalle and P. Menendez-Benito. Oxford: Oxford University Press.

Dayal, Veneeta. 1997. Free relatives and *ever*: Identity and free choice readings. *Proceedings of SALT VII*, 99–116. Ithaca, NY: CLC Publications, Cornell University.

von Fintel, Kai. 2000. Whatever. *Proceedings of SALT X*, ed. by B. Jackson and T. Matthews, 27–39. Ithaca, NY: Cornell University.

Guilfoyle, Eithne, Henrietta Hung, and Lisa Travis. 1992. SPEC of IP and SPEC of VP: Two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10:375–414.

Jacobson, Pauline. 1995. On the quantificational force of English free relatives. *Quantification in natural languages, vol. II*, ed. by E. Bach et al., 451–487. Kluwer.

Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic Inquiry* 4:167–93.

Krifka, Manfred. 2001. For a structured meaning account of questions and answers. *Audiatur Vox Sapientia: A festschrift for Arnim von Stechow* 52:287–319.

Lauer, Sven. 2011. Free relatives with *-ever*: Meaning and use. Ms., Stanford University.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. *Meaning, use, and interpretation of language*, ed. by R. Bäuerle, C. Schwarze, and A. v. Stechow, 302–323. Berlin: Mouton.

Matthewson, L. 1999. On the interpretation of wide-scope indefinites. *Natural Language Semantics* 7: 79–134.

Partee, Barbara H. 1987. Noun phrase interpretation and type-shifting principles. *Studies in discourse representation theory and the theory of generalized quantifiers*, ed. by J. Groenendijk, D. de Jong, and M. Stokhof, 115–143. Dordrecht: Foris Publications.

Paul, Ileana. 2000. Malagasy Clause Structure. Ph.D. dissertation, McGill University.

Potsdam, Eric. 2006. The cleft structure of Malagasy *wh*-questions. *Clause structure and adjuncts in Austronesian languages*, ed. by H.-M. Gärtner, P. Law, and J. Sabel, 195–232. Berlin: Mouton de Gruyter.

Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.

Reinhart, Tanya. 1997. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20(4):335–397.

Sharvy, Richard. 1980. A more general theory of definite descriptions. *The Philosophical Review* 89:607-624.

Stalnaker, Robert C. 1973. Presuppositions. *Journal of Philosophical Logic* 2:447–457.

Stalnaker, Robert C. 1978. Assertion. *Syntax and Semantics* 9:315–322.

Tonhauser, Judith, David Beaver, Craige Roberts, and Mandy Simons. 2013. Towards a taxonomy of projective content. *Language* 89:66–109.

WHAT MAKES A VOICE SYSTEM?
ON THE RELATIONSHIP BETWEEN VOICE MARKING AND CASE*

Michael Yoshitaka Erlewine
MIT
mitcho@mit.edu

Theodore Levin
MIT
tedlevin@mit.edu

Coppe van Urk
MIT
cvanurk@mit.edu

One of the major questions in Austronesian syntax concerns the relationship between voice marking, extraction, and case. Two common approaches have dominated previous literature. Either voice morphology marks extraction and case, or voice morphology feeds extraction and case. These positions are difficult to distinguish, because of the prominent one-to-one correspondence of voice, case, and extraction. In this paper, we bring new insights to this debate, with original data from Balinese and Dinka, a Nilotic language of South Sudan, which we show exhibits a familiar Austronesian-type voice system. We observe environments in these languages where the correspondence between voice and case and voice and extraction breaks down, in a manner that we argue provides evidence that *voice marks extraction*. Unlike in other extraction-marking languages, however, voice also affects case in Austronesian-type voice systems. We suggest that this is because extraction targets a case position. We account for the changes in case marking in the clause by suggesting that, when the external argument is not extracted, languages must employ alternative strategies to license it.

1. Introduction

In a striking example of syntactic uniformity across genetically and geographically disparate languages, many languages morphologically mark the difference between *non-subject* (a) and *subject* (b) extraction. Consider first the behavior of English *do*-support in (1). English non-subject *wh*-questions require the insertion of *do* to host tense specification, while *do*-support is crucially unavailable in subject *wh*-questions.

(1)   *English T-to-C movement:*
    a. Who *did* Alex see?
    b. Who saw Alex?

Similarly, in French the form of the embedded complementizer varies in long distance *wh*-questions (2). This alternation is conditioned by which element is ex-

tracted. When a non-subject argument is extracted, as in (2a), the complementizer *que* is realized. When the subject is extracted, as in (2b), *qui* appears.

(2)   *French que/qui alternation:*
    a. Qui penses-tu [*que* Marie a   rencontré]?
       who think-you that Marie has met
       'Who do you think Marie has met?'
    b. Qui penses-tu [*qui* a   rencontré Marie]?
       who think-you that has met   Marie
       'Who do you think has met Marie?'

A similar extraction asymmetry is observed in a subset of Mayan languages (see e.g. Stiebels 2006). When a non-subject argument is extracted (3a), the same verb form is used as in a corresponding declarative clause. But when a subject is extracted, the "Agent Focus" form of the verb must be used (3b).

(3)   *Agent Focus in Kaqchikel (Erlewine to appeara):*
    a. Achike xutëj ri a   Juan?   b. Achike xtj-*ö*   ri wäy?
       what   ate the CL Juan   who   ate-AF the tortilla
       'What did Juan eat?'   'Who ate the tortilla?'

Lastly, we observe that in Moro (Niger-Congo; Sudan) non-subject extraction, such as the case of object extraction in (4a), triggers optional *wh*-concord indicated by prefixing all words after the *wh*-word with *nə*. This concord is unavailable in subject extraction examples such as (4b).

(4)   *Moro wh-concord (Rohde 2006; Rose et al. 2014):*
    a. ŋwədʒeki (*nə*).Kuku (*nə*).gətaðoŋo? b. ŋwədʒeki gətaðo   Kuku?
       who   (WH).Kuku (WH).abandon   who   abandon Kuku
       'Who did Kuku abandon?'   'Who abandoned Kuku?'

In contrast to languages that exhibit subject vs non-subject extraction marking, many Austronesian languages appear to display a *more articulated* form of extraction marking. Languages like Atayal (Formosan; Taiwan) have morphology on the verb that not only differentiates subject extractions from non-subject extractions, but also distinguishes between different types of non-subject extractions.[1] Each example in (5) below has one constituent marked with *qu* and in sentence-final position, and the choice of constituent in this position correlates with different *voice* morphol-

---

[1]The "voice" morphology studied here has also been called "focus" and "topic" marking in different corners of the Austronesian literature. Note that this Austronesian voice morphology is distinguished from familiar Indo-European-style active/passive alternations, which are also called "voice."

ogy on the verb. This set of examples illustrates one of the most well-known aspects of Austronesian syntax: *the voice system*.

(5)   *Squliq Atayal voice marking (Liu 2004):*[2]

    a. *M*-aniq qulih qu' Tali'.   *Actor Voice (AV)*
        AV-eat fish   QU Tali
        'Tali eats fish.'

    b. Niq-*un* na'   Tali' qu' qulih qasa.   *Patient Voice (PV)*
        eat-PV GEN Tali   QU fish   that
        'The fish, Tali ate.'

    c. Niq-*an* na'   Tali' qulih qu' ngasal qasa.   *Locative Voice (LV)*
        eat-LV GEN Tali   fish   QU house that
        'In that house, Tali eats fish.'

    d. *S*-qaniq na'   Tali' qulih qu' qway.   *Instrumental Voice (IV)*[4]
        IV-eat   GEN Tali   fish   QU chopsticks
        'With chopsticks, Tali eats fish.'

In the remainder of this paper, we will refer to the argument cross-referenced by voice morphology as *"subject,"* and refer to movement to this subject position as *"extraction."* The terms *"actor"* and *"patient"* will be used to refer to arguments with these thematic roles.

    There are two aspects of the Atayal voice system worth highlighting here, which also hold of many other Austronesian languages. First, in non-actor voice clauses (5b–d), the actor is marked with genitive case. Second, in Ā-constructions such as *wh*-questions, only the "subject" can be Ā-extracted to the left.[5] In Atayal, we thus observe a strict correlation between (a) voice morphology on the verb, (b) the clause-final "subject" constituent, (c) the constituent which can be Ā-extracted, and (d) the pattern of case marking on nominals.

    The differences between plain extraction marking of the type illustrated in (1–4) and Austronesian voice systems have led to two different types of approaches: (i) voice morphology is like extraction marking, but by a different mechanism, such as *wh*-agreement or case agreement (e.g. Chung 1994; Richards 2000; Pearson 2001, 2005); (ii) voice morphology actually encodes argument structure alternations which result in extraction restrictions (Guilfoyle et al. 1992; Aldridge 2004, 2008; Legate 2012, e.g.). Amongst the former group of theories, it is held that voice morphology

---

[2]Translations are modified from Liu (2004), so that the argument cross-referenced by voice morphology is uniformly translated as a topic in English. See Erlewine (to appearb) for evidence that the sentence-final "subject" position in Squliq Atayal tracks the discourse topic, as well as for discussion of *qu*.

[4]There is also a Benefactive Voice with the same prefix *s*- as Instrumental Voice.

[5]This data is not shown here for reasons of space.

is 'cosmetic', like extraction marking. That is, voice morphology does not drive the derivation nor determine case, but simply reflects the results of that derivation. In the latter group of theories, voice plays an important role in determining the course of the derivation. The voice morphology chosen determines which argument can be promoted to "subject" position as well as the case morphology of the external argument. In practice, these two positions are difficult to distinguish, because both proposals can handle the prominent one-to-one correspondence of voice, case, and extraction often found in Austronesian voice systems.

In this paper, we present arguments that voice morphology in Austronesian should be viewed as *extraction marking*, much like the morphological alternations in (1–4). We present two systems in which the one-to-one correspondence between voice, case, and extraction *breaks down*, the Nilotic language Dinka and Indonesian-type languages like Balinese. In Dinka, which exhibits an Austronesian-type voice system, we can dissociate voice and case. In Balinese, we can dissociate voice and extraction. These breakdowns are inconsistent with a view of Austronesian-type voice morphology as *extraction feeding*, because such a view predicts mismatches to be impossible. Crucially unlike in the non-voice languages in (1–4), we propose that voice affects case because *the "subject" position* (the argument referenced by voice morphology) is a *mixed A/Ā-position*. The external argument must move to this position to be case-licensed. If this movement is unavailable, alternative strategies must be utilized to license the external argument. Interestingly, it appears that languages make use of different alternative licensing strategies. We examine three such strategies below.

The remainder of this paper is organized as follows. In section 2, we introduce the Nilotic language Dinka as a language with an Austronesian-type voice system, and we show that voice morphology can be dissociated from the processes governing case marking in Dinka. Section 3 further argues that voice morphology can be dissociated from the extraction restriction, in instances of multiple extraction in Balinese. In section 4, we turn to the question of why voice morphology often triggers changes in case marking throughout the clause. We argue that what distinguishes Austronesian-type voice systems is that extraction targets a case position, so that extraction interferes with the licensing of the actor. We link differences between voice systems to different strategies for licensing the actor in non-actor voices.

2.   Dissociating voice and case in Dinka

If voice morphology is extraction marking, we expect to find *dissociations between voice and case*, since voice would not directly determine case. Rather, in this view, the mechanisms that give rise to case marking should in principle be independent of voice. We find such dissociations in Dinka, a Nilotic language of South Sudan, with a voice system highly reminiscent of Austronesian (Van Urk and Richards 2015).

2.1.   The Dinka voice system

Dinka is a V2 language of the Nilotic family spoken in South Sudan. While not genetically related to Austronesian, it displays a voice system reminiscent of those in the Austronesian family. Dinka has three voices, which reflect the grammatical function of *the clause-initial "subject" position*:

(6) a. Àyén à-*càm*   cuĩ̈n nè̤ pàl.   *Actor Voice (AV)*
      Ayen 3S-*eat*.AV food P knife
      'Ayen is eating food with a knife.'

    b. Cuĩ̈n à-*cέεm*   Áyèn   nè̤ pàl.   *Patient Voice (PV)*
      food 3S-*eat*.PV Ayen.NOM P knife
      'Food, Ayen is eating with a knife.'

    c. Pàl à-*cέεmè̤*   Áyèn   cuĩ̈n.   *Oblique Voice (OblV)*
      knife 3S-*eat*.OBLV Ayen.NOM food
      'With a knife, Ayen is eating food.'

As in Germanic V2 languages, the highest verb or auxiliary occupies second position. Voice morphology appears on this verbal element, marking the grammatical function of the constituent in clause-initial position. In (6), we observe that the verb *câam* 'eat' takes distinct forms which cross-reference the element in first position. If an auxiliary is present, it occupies second position instead of the main verb. Voice morphology then appears on the auxiliary (7a–c).

(7) a. Àyén   à-*cέ*   cuĩ̈n   câam nè̤ pàl.   *AV*
      Ayen.ABS 3S-PRF.AV food.ABS eat.NF P knife.ABS
      'Ayen has eaten food with a knife.'

    b. Cuĩ̈n   à-*cíi*   Áyèn   câam nè̤ pàl.   *PV*
      food.ABS 3S-PRF.PV Ayen.NOM eat.NF P knife.ABS
      'Food, Ayen has eaten with a knife.'

    c. Pàl   à-*cénnè̤*   Áyèn   cuĩ̈n   câam.   *OblV*
      knife.ABS 3S-PRF.OBLV Ayen.NOM food.ABS eat.NF
      'With a knife, Ayen has eaten food.'

Regardless of where voice morphology appears, the clause-initial XP always appears in the unmarked case, usually called "absolutive" in the Nilotic literature (e.g. Dimmendaal 1985; Andersen 1991, 2002).

    As in many Austronesian languages, voice marking restricts overt Ā-extraction. The constituent undergoing Ā-extraction must be the argument cross-referenced by voice morphology:

(8) a. Yeŋà *cé̤* cui̤n câam nè̤ pàl? *AV*
    who PRF.AV food.ABS eat.NF P knife.ABS
    'Who has eaten food with a knife?'

  b. Yeŋṳ́ *cí̤i* Áyèn câam nè̤ pàl. *PV*
    what PRF.PV Ayen.NOM eat.NF P knife.ABS
    'What has Ayen eaten with a knife?'

  c. Yeŋṳ́ *cé̤nnè̤* Áyèn cui̤n câam. *OblV*
    what PRF.OBLV Ayen.NOM food.ABS eat.NF
    'What has Ayen eaten food with?'

It is important to note that non-initial actors appear in a dedicated case, the "marked nominative", while non-initial patients are unmarked (i.e. absolutive). Case alternations are not realized using affixes as is common cross-linguistically, but rather by alternations in tone. Observe that the external argument, *Àyén*, in clause initial position (6a,7a) bears a distinct tonal pattern from the same argument in non-initial position: *Áyèn*, in the (b–c) examples above. In this respect as well, Dinka behaves like Austronesian languages, which display dichotomies between subject and non-subject actors. We will return to this point below.

2.2.   Voice is independent of case

V2 in Dinka, as in many other V2 languages, is limited to certain types of clauses. In non-V2 clauses, no argument is extracted to the front of the clause, resulting in verb-initial order. We will use these environments to see whether voice morphology patterns with case or with extraction. As we will see, voice patterns with extraction, and only default voice morphology appears in non-V2 environments.

    Matrix *yes-no* questions are verb-initial in Dinka, with no constituent fronted to the clause-initial position (9).

(9)   *Verb-initial yes-no question with AV and marked nominative actor:*
    Nhiár Máyèn Àdít?
    love.AV Mayen.NOM Adit.ABS
    'Does Mayen love Adit?'

As (9) shows, a novel pattern emerges in such a clause. Elements following the highest verb or auxiliary appear just as they do when they are not in clause-initial position in V2 clauses. Thus, word order is strict (the actor must come before the patient) and the actor and patient are case-marked just as they are when not extracted. The actor appears in the "marked nominative", just as in Patient Voice or Oblique Voice (e.g. 7b–c). The patient is in the absolutive, just as in the Agent Voice (7a) or Oblique Voice (7c). Voice morphology, however, is necessarily Agent Voice in *yes-no* questions. We treat this as an instance of *default* marking, since Agent Voice

otherwise does not appear with "marked nominative" case on the subject (as in 7a).

There are a number of other verb-initial environments which make the same point, that alternations in voice are not necessary to drive case marking on non-"subject" nominals. Following the complementizer *yè*, verb-initial order is found, with (default) AV on the highest verb/auxiliary, but "marked nominative" on the postverbal subject (10).

(10) *Verb-initial order under* yè *complementizer:*
      À-yừkkừ   luêel, [yè *nhiár*   Máyèn   wôɔk].
      3S-PRF.1PL say.NF C *love*.AV Mayen.NOM 1PL.ABS
      'We say that Mayen loves us.'

Another environment with verb-initial order is in *bɛ́*-clauses. These clauses are found with a set of verbs that usually function as control verbs in other languages. In Dinka, these verbs select for a verb-initial clause always headed by the future auxiliary *bɛ́* (11a). V2 is ungrammatical (11b).

(11) *Verb-initial order in* bɛ́-*clauses:*
   a. Bòl   à-cɛ́   Àyén   lɔ̂ɔŋ [*bɛ́*   Máyèn   rɛ̀ɛr].
      Bol.ABS 3S-PRF.AV Ayen.ABS beg.NF FUT.AV Mayen.NOM stay.NF
      'Bol has begged Ayen for Mayen to stay.'
   b.*Bòl   à-cɛ́   Àyén   lɔ̂ɔŋ [Màyén   (à-)*bɛ́*   rɛ̀ɛr].
      Bol.ABS 3S-PRF.AV A.ABS beg.NF Mayen.ABS (3S-)FUT.AV stay.NF
      'Bol has begged Ayen for Mayen to stay.'

In a number of environments, then, case marking and voice morphology diverge in Dinka.

It is not the case, however, that these verb-initial clauses have completely fixed AV morphology. For example, when an argument is long-distance extracted out of the *bɛ́*-clause still drives changes in voice morphology. This supports the view that these verb-initial clauses simply do not front any constituent to initial position and AV morphology is the default realization, rather than a view that these clauses are somehow voice-deficient. Note also that the pattern of case-marking in the AV embedding in (11b), with no marked nominative argument, clearly contrasts from the AV clause in (11) above, with a marked nominative actor.

(12) *Long-distance extraction triggers voice alternations in* bɛ́-*clauses:*
   a. Yeŋɔ́ lɛ̀ŋ-kừ Àyén   [*bíi*   Máyèn   gɔ̂ɔr]?
      what beg-1PL Ayen.ABS FUT.PV Mayen.NOM write.NF
      'What are we begging Ayen for Mayen to write?'

b. Yeŋà lɛ̀ŋ-kừ   Àyén   [bɛ́   àkɛ́kôol gɔ̂ɔr]?
   who beg-1PL Ayen.ABS FUT.AV story.ABS write.NF
   'Who are we begging Ayen for *t* to write a story?'

These facts clearly show that voice morphology tracks extraction and does not correlate with case marking on any particular nominal.This follows under a view in which voice morphology functions as extraction marking. In contrast, under an extraction feeding view of voice morphology, even if extraction was independently blocked in verb-initial environments, we would expect voice and case to correlate. Actor voice morphology on the verb should trigger absolutive case on the external argument, while marked nominative case should be limited to non-actor voices.

In this section, we saw that voice and case can be dissociated in Dinka. In verb-initial environments, verbs and auxiliaries are marked with AV morphology, but the actor argument does not occupy clause-initial positon. Rather the actor surfaces in its base position and bears "marked nominative" case. Crucially, this dissociation is surprising if voice morphology is an argument structure alternation that affects the pattern of case assignment and feeds extraction to subject position. Adopting instead the view the that voice morphology *marks* extraction, we posit that AV is a default form which arises when no argument occupies the subject position.

3. Voice and multiple extraction in Balinese

In this section, we present another breakdown of the common one-to-one correspondence of voice, case, and extraction. Specifically, we observe a dissociation between voice morphology and extraction in Indonesian-type languages, such as Bahasa Indonesia (Chung 1976; Cole and Hermon 2005), Jambi Malay (Yanti 2010), and Balinese. Here we focus on Balinese.

We concentrate here on the Actor Voice and Patient Voice in Balinese:[6]

(13) a. *Actor Voice (AV):*   b. *Patient Voice (PV):*

   Polisi *ng*-ejuk   Nyoman.   Nyoman ∅-ejuk   polisi.
   police AV-arrest Nyoman   Nyoman PV-arrest police
   'A policeman arrested Nyoman.'   'A policeman arrested Nyoman.'

In this impoverished voice system, any argument that is promoted to the pre-verbal subject position other than the actor is cross-referenced with Patient Voice morphology. As we have seen above for Atayal and Dinka, voice morphology imposes an extraction restriction. When the actor argument appears in subject position, AV morphology must appear on the verb (14a). When the patient argument appears in subject

---

[6]The Actor Voice and Patient Voice are described as the two "active" voices in Indonesian-type languages, which also have other "passive" voices. See e.g. Arka (2003); Aldridge (2008).

position, PV morphology must appear on the verb (14b).

(14) a. *Actor extraction ⇒ AV:*

   Nyen *ng/*∅*-alih   ci   ditu   ibi?
   who   AV/*PV-search   you   there   yesterday
   'Who looked for you there yesterday?'

  b. *Patient extraction ⇒ PV:*

   Apa   *ng/∅-alih   ci   ditu   ibi?
   what   *AV/PV-search   you   there   yesterday
   'What did you search for there yesterday?'

   However, in Balinese, extraction of multiple arguments to pre-verbal position is also possible. Multiple extraction occurs when the actor is fronted to be in immediate preverbal position and subsequently another argument undergoes *wh*-movement.[7] In such cases, the verb is PV:

(15)   Buku   cen   Nyoman *ng/∅-paca?
   book   which   Nyoman *AV/PV-read
   'Which book did Nyoman read?'

We can explain these facts if we view voice morphology as extraction marking. If PV surfaces whenever a patient is extracted and AV is the default realization, we expect to see PV if multiple extraction is ever possible. Under an extraction marking view, voice morphology is logically independent of the extraction restriction. In contrast, under a view in which voice morphology drives argument structure alternations, voice morphology is the mechanism by which the extraction restriction is created. As a result, we do not expect to find such dissociations.

   Under this kind of proposal, the difference between a voice system like Balinese, in which multiple extraction is possible, and Dinka, in which it is not, must be independent of the mechanisms behind voice. Rather, we might imagine that the difference between Dinka and Balinese is much like the difference between English, in which multiple instances of Ā-movement—such as topicalization and *wh*-movement—cannot co-occur, and Italian, in which multiple instances of Ā-extraction can target the same left periphery.

---

[7]This configuration is reminiscent of the famed "bodyguard" construction in Malagasy (Keenan 1976). Due to space limitations, we will not describe our analysis for these multiple extraction constructions here.

4.   The relationship between voice and case

We now turn to the relationship of voice systems and case/nominal licensing. We have argued so far that Austronesian-type voice morphology is extraction marking. However, unlike extraction marking in non-voice system languages (English, French, Kaqchikel, Moro in §1), voice often has *repercussions for case throughout the clause*. We propose that what unifies voice systems is that the position in the clause periphery targeted by Ā-extraction, the "subject" position, is a *case position*. Here we will call this subject position Spec,CP.

In AV clauses, the actor is licensed in this case position. In non-AV (NAV) clauses, when a XP that is not the actor undergoes Ā-movement, it deprives the actor of its licenser, since extraction targets its case position. Thus, the actor must be licensed using an alternative strategy. We suggest that differences between voice systems arise in part because voice languages handle the problem of licensing the external argument actor in non-actor voices differently. In particular, we will demonstrate *three strategies* for licensing the actor in NAV:
   1. Ergative/genitive marking in Formosan/Philippine-type voice systems;
   2. Oblique (prepositional) case in the Dinka voice system;
   3. Pseudo-noun incorporation in the Balinese voice system.
Even genetically close languages may use different strategies, while some genetically distant languages (e.g. Formosan or Philippine-type and Dinka) use conceptually very similar strategies.

4.1.   Strategy 1: ergativity

The first strategy we identify is to license the actor with ergative/genitive case. This is the strategy observed in Atayal. Here we will consider the following AV and PV examples:

(16)   *Actor Voice (AV):*   (17)   *Patient Voice (PV):*
       M-aniq sehuy (qu) Yuraw.   Niq-*un na*   Yuraw (qu) sehuy.
       AV-eat   taro   QU   Yuraw   eat-PV GEN Yuraw QU   taro
       'Yuraw eats taro.'   'Yuraw eats taro.'

Consider first the derivation of the AV clause in (16). As mentioned previously, we propose that the "subject" position is Spec,CP. In the AV clause, the external argument *Yuraw* cannot be licensed in its base position, Spec,*v*P (18). It moves to Spec,CP and is licensed there (19).

(18)

```
              vP
          ___/  \___
     ☹DP        v    VP
      |            __/  \__
    Yuraw         V       ☺DP
                  |        |
                maniq    sehuy
```

(19)

```
                 CP
            ___/    \___
       ☺DP            C    TP
        |                 __/  \__
      Yuraw              T       vP
                         |      /  \
                       maniq   t    tV sehuy
```
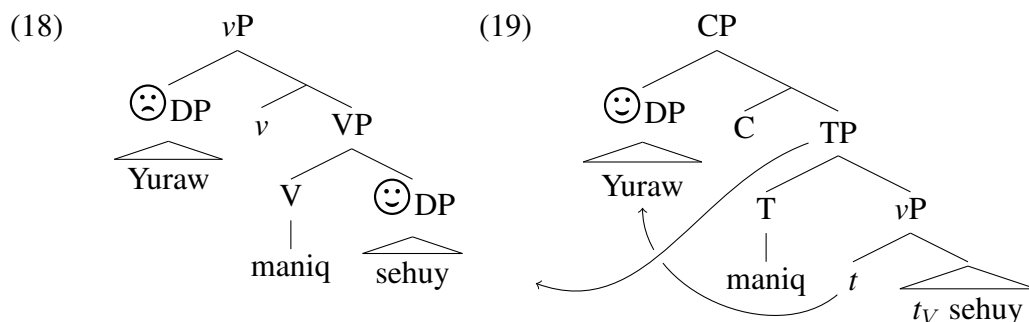
As in a number of Austronesian languages, the "subject" position is clause-final. TP-fronting yields the observed word order (Aldridge 2004). *Qu* marks the DP in the subject position, and is not a case marker (Erlewine to appearb).

In NAV clauses, the actor is *genitive marked*. This genitive-marked actor has been analyzed as an *ergative* argument in some previous work (Huang 1994; Starosta 1999; Aldridge 2004). Consider the derivation of the PV clause in (17). We assume that, in non-actor voices, the "subject"—in this case the patient *sehuy*—occupies the case position that licenses the actor in AV clauses. As a result, the actor is deprived of licensing (20).

(20)

```
                    CP
               ___/    \___
          ☺DP            C    TP
           |                 __/  \__
         sehuy              T       vP
                            |      /  \
                          niqun  ☹DP   tV  t
                                  |
                                Yuraw
```
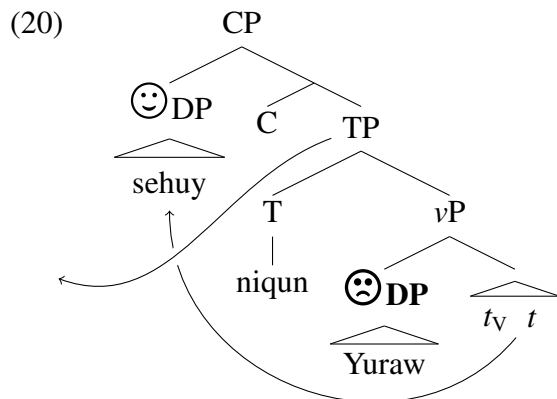
This is precisely the configuration where the actor is given ergative/genitive case, using a particular choice of *v* (Aldridge 2004; Woolford 2006; Legate 2008). This strategy allows the actor to be licensed in non-actor voices, when the "subject" position is unavailable.[8]

---

[8]Under our proposal here, voice is extraction marking, not a *v* head, but moving a non-actor to "subject" position necessarily correlates with the choice of an ergative/genitive-assigning *v*. There are important questions here regarding derivational look-ahead. We will leave these issues for future work. We note that ergative/genitive-assignment could be conceived of as an application of a last-resort or default rule, as in Imanishi (2014), which avoids (or reframes) this issue. See also footnote 9 below.

4.2.   Strategy 2: oblique case

We find a different strategy in Dinka. In Dinka NAV clauses, actors appear in a dedicated case, *"marked nominative"* (Kãűnig 2006; Van Urk and Richards 2015), which is tonally marked:

(21)   *Actor Voice (AV):*
        Ayén   à-cé̤   cṳ̂ìn   câam.
        *Ayen.*ABS   3S-PRF.AV   food.ABS   eat.NF
        'Ayen has eaten food.'

(22)   *Patient Voice (PV):*
        Cṳ̂ìn   a-cí̤i   *Áyèn*   câam.
        food.ABS   3S-PRF.PV   *Ayen.*NOM   eat.NF
        'Food, Ayen has eaten.'

"Marked nominative" is unlike familiar subject cases. "Marked nominative" does not pattern like ergative case. It is not linked to transitivity or semantic properties of the verb and can be found with unergatives and unaccusatives in environments that suppress V2:

(23) a. *Adít*   à-nìn.   (24) a. *Galàm*   à-cé̤   dhuòŋ.
        *Adit.*ABS   3S-sleep.AV   *pen.*ABS   3S-PRF.AV   break.NF
        'Adit is sleeping.'   'The pen broke.'

     b. Nìn   *Ádìt*?   b. Cé̤   *galám*   dhuòŋ?
        sleep.AV   *Adit.*NOM   PRF.AV   *pen.*NOM   break.NF
        'Is Adit sleeping?'   'Did the pen break?'

"Marked nominative" also does not behave like nominative, however, because it is the marked case. The absolute appears in all default contexts, as in citation form and on nominal predicates (Andersen 1991, 2002).

(25)   Adít   e-*dupióoc*.
        Adit.ABS   COP-*teacher.*ABS
        'Adit is a teacher.'

In addition, "marked nominative" is also assigned by some prepositions (Andersen 2002):

(26) a. Yï̤n nhiàr   ɣò̤n   è̤ *Máyèn*.
        you love.AV   house.LNK   P *Mayen.*NOM
        'You love Mayen's house.'

b. Cuĩ̵n   à-cĩ̵i   câam nè̤ pàl   nè̤ *Áyèn.*
   food.ABS 3S-PRF.PASS eat.NF P   knife.ABS P   *Ayen.*NOM
   'The food has been eaten with a knife by Ayen.'

Instead, we propose that "marked nominative" is an oblique case, assigned by a null preposition, inserted as a repair to license a caseless nominal. We draw on the notion of repair in Rezac (2012), who suggests that prepositional material may be added at the end of a phase to license certain DPs that fail to acquire case. For similar proposals regarding prepositions for nominals which would otherwise be unlicensed, see Stowell (1981) on English *of*-Insertion and Halpert (2012) on Bantu augment nominals.[9]

The derivation for Dinka NAV clauses is illustrated by the following trees. As in Atayal, we posit that the "subject" moves to Spec,CP and deprives the actor of its usual licensing position (27). To license the actor, a silent preposition is inserted, which assigns case to the actor (28).

(27)

(28)
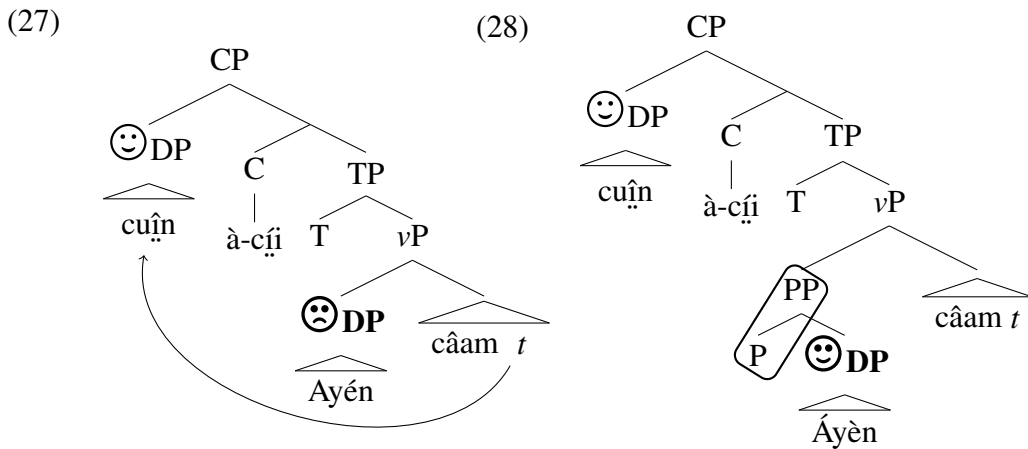
In this view, "marked nominative" is actually a *prepositional case*, and absolutive is the only real structural case (which is why it behaves like the unmarked case). There is then no structural licensor for the subject outside of the voice system. As a result, "marked nominative" emerges as a necessary repair in non-subject voices and in structures in which the voice system is not available.

---

[9]Strategies 1 and 2 could be seen as very similar, if we analyzed ergative/genitive case in Formosan and Philippine languages (Atayal above) as a last resort repair. See Imanishi (2014) for such a proposal for ergativity in Mayan.

4.3.   Strategy 3: pseudo-noun incorporation

A different strategy is found in Balinese. In Balinese, the actor in non-actor voices forms a single "phonological word" with the verb (Clynes 1995). In Balinese NAV clauses, post-verbal actors undergo *Pseudo-Noun Incorporation* (PNI), by means of Morphological Merger (Levin 2014). Such actors display *strict head-head adjacency* with the verb.

Evidence for this adjacency requirement is presented here. First, pre-nominal adjectives are banned on post-verbal actors. Adjectives that can normally appear pre- and post-nominally must appear post-nominally when modifying a post-verbal actor:

(29) a. *Actor Voice (AV):*

[(Liu)   cicing (liu)]   ngugut Nyoman.
(many) dog   (many) AV.bite Nyoman
'Many dogs bit Nyoman.'

    b. *Patient Voice (PV):*

Nyoman gugut   [(*liu)   cicing (liu)].
Nyoman PV.bite (*many) dog   (many)
'Many dogs bit Nyoman.'

In addition, the post-verbal actor shows a *definiteness effect*. The definite suffix *-e* and overt determiners like *ento* 'that' are illicit (Wechsler and Arka 1998, 441):

(30) a. I   Wayan gugut   cicing.   b.*I   Wayan gugut   cicing-*e* (*ento*).
ART Wayan PV.bite dog   ART Wayan PV.bite dog-DEF (*that*)
'A dog bit Wayan.'   'The dog bit Wayan.'

We propose that this is because the presence of an NP blocks PNI of D. In support of this, we see that pronouns and proper names can undergo PNI.

(31) a. Be-e   daar   *ida*.   b. Be-e   daar   *Nyoman*.
fish-DEF PV.eat 3SG   fish-DEF PV.eat *Nyoman*
'(S)he ate the fish.'   'Nyoman ate the fish.'

We can derive this if pronouns and proper names occupy D^0 (e.g. Postal 1966; Longobardi 1994; Elbourne 2001) and lack an NP, satisfying head-head adjacency.

The behavior of non-subject actors in Balinese is inconsistent with either of the two alternative licensing stagies above. If non-subject actors were case marked either by lexical case or P-insertion, we would not expect to find the head-head adjacency requirement. However, if head-head adjacency is necessary to create a structure to which Morphological-Merger can apply, then the behavior of Balinese can be captured.

5. Conclusion

In this paper, we presented two examples where the one-to-one correspondence of voice, case, and extraction can break down in Austronesian-type voice system languages. We argue that voice morphology is a form of *extraction marking*, which tracks the argument moved to the "subject" position (Chung 1994; Richards 2000; Pearson 2001, 2005). By connecting this position to the licensing of the actor in AV clauses, we arrive at a unified explanation for *the quirky behavior of actors in NAV clauses*:

1. Ergative/genitive marking in Formosan and Philippine languages;
2. Oblique case marking in Dinka (Nilotic);
3. Pseudo-noun incorporation in Balinese.

Specifically, the idea is that (a) the external argument actor lacks structural licensing in its Spec,*v*P position, (b) the actor is licensed in the subject position in AV, and (c) another strategy is necessary for licensing the subject in NAV clauses.

This view of voice morphology leaves open a number of questions. First of all, we have left open the question of why and how voice languages come to show more articulated extraction marking, as we saw with in Atayal examples in (5). One appealing answer is that such non-PV non-actor voices reflect argument structure alternations that are necessary to turn PP arguments into nominals that can occupy the "subject" position, as in Rackowski's (2002) treatment of Tagalog and Van Urk's (in preparation) analysis of the Dinka oblique voice. A second question is what mechanism ultimately yields non-subject extraction marking (e.g. Chung 1994; Pesetsky and Torrego 2001; Rizzi and Shlonsky 2007; Erlewine to appeara). This issue is especially important, since some theories of extraction marking crucially link non-subject extraction morphology to case, such as Pesetsky and Torrego's (2001) proposal.

References

Aldridge, Edith. 2004. Ergativity and word order in Austronesian languages. Doctoral Dissertation, Cornell University.

Aldridge, Edith. 2008. Phase-based account of extraction in Indonesian. *Lingua* 1118:1440–1469.

Andersen, Torben. 1991. Subject and topic in Dinka. *Studies in Language* 15:265–294.

Andersen, Torben. 2002. Case inflection and nominal head marking in Dinka. *Journal of African Languages and Linguistics* 23:1–30.

Arka, Wayan. 2003. *Balinese morphosyntax: a lexical-functional approach*. Canberra: Pacific Linguistics.

Chung, Sandra. 1976. An object-creating rule in Bahasa Indonesia. *Linguistic In-*

quiry 7:41–87.

Chung, Sandra. 1994. *Wh*-agreement and "referentiality" in Chamorro. *Linguistic Inquiry* 25:1–44.

Clynes, Adrian. 1995. Topics in the phonology and morphosyntax of Balinese—based on the dialect of Singaraja, North Bali. Doctoral Dissertation, Australia National University.

Cole, Peter, and Gabriella Hermon. 2005. Subject and non-subject relativization in Indonesian. *Journal of East Asian Linguistics* 14:59–88.

Dimmendaal, Gerrit. 1985. Prominence hierarchies and Turkana syntax. In *Proceedings of ACAL 14*, 127–148.

Elbourne, Paul. 2001. E-type anaphora as NP-deletion. *Natural Language Semantics* 9:241–288.

Erlewine, Michael Yoshitaka. to appeara. Anti-locality and optimality in Kaqchikel Agent Focus. *Natural Language & Linguistic Theory* URL `http://ling.auf.net/lingbuzz/001841/current.pdf`.

Erlewine, Michael Yoshitaka. to appearb. Subject marking on non-subjects in Squliq Atayal. In *Proceedings of the 20th meeting of the Austronesian Formal Linguistics Association (AFLA 20)*.

Guilfoyle, Eithne, Henrietta Hung, and Lisa Travis. 1992. Spec of IP and Spec of VP: Two subjects in Austronesian languages. *Natural Language & Linguistic Theory* 10:375–414.

Halpert, Claire. 2012. Argument licensing and agreement in Zulu. Doctoral Dissertation, Massachusetts Institute of Technology.

Huang, Lillian Meijin. 1994. Ergativity in Atayal. *Oceanic Linguistics* 33:129–143.

Imanishi, Yusuke. 2014. Default ergative. Doctoral Dissertation, Massachusetts Institute of Technology.

Keenan, Edward. 1976. Remarkable subjects in Malagasy. In *Subject and topic*, ed. Charles N. Li and Sandra A. Thompson. Academic Press.

KÃűnig, Christa. 2006. *Case in Africa*. Oxford University Press.

Legate, Julie Anne. 2008. Morphological and abstract case. *Linguistic Inquiry* 39:55–101.

Legate, Julie Anne. 2012. Subjects in Acehnese and the nature of the passive. *Language* 88:495–525.

Levin, Theodore. 2014. Pseudo noun incorporation is M-Merger: evidence from Balinese. Handout from LSA 2014.

Liu, Adlay Kun-long. 2004. On relativization in Squliq Atayal. Master's thesis, National Tsing Hua University.

Longobardi, Giuseppe. 1994. Reference and proper names: a theory of N movement in syntax and logical form. *Linguistic Inquiry* 25:609–665.

Pearson, Matthew. 2001. The clause structure of Malagasy: A minimalist approach. Doctoral Dissertation, University of California at Los Angeles.

Pearson, Matthew. 2005. The Malagasy subject/topic as an A′-element. *Natural*

Language & Linguistic Theory 23:381–457.

Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*. MIT Press.

Postal, Paul M. 1966. On so-called "pronouns" in English. In *19th Monograph on Languages and Linguistics*. Georgetown University Press.

Rackowski, Andrea. 2002. The structure of Tagalog: Specificity, voice, and the distribution of arguments. Doctoral Dissertation, Massachusetts Institute of Technology.

Richards, Norvin Waldemar, III. 2000. Another look at Tagalog subjects. In *Formal issues in Austronesian linguistics*, ed. Ileana Paul, Vivianne Phillips, and Lisa Travis, 105–116. Springer.

Rizzi, Luigi, and Ur Shlonsky. 2007. Strategies of subject extraction. In *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*, ed. Uli Sauerland and Hans-Martin GÃd'rtner, number 89 in Studies in Generative Grammar, 115–160. Mouton de Gruyter.

Rohde, Hannah. 2006. The syntax of questions in Moro. Handout from ACAL 37.

Rose, Sharon, Farrell Ackerman, George Gibbard, Peter Jenks, Laura Kertz, and Hannah Rohde. 2014. In-situ and ex-situ *wh*-question constructions in Moro. *Journal of African Languages and Linguistics* 35:91–125.

Starosta, Stanley. 1999. Transitivity, ergativity, and the best analysis of Atayal case marking. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*.

Stiebels, Barbara. 2006. Agent Focus in Mayan languages. *Natural Language & Linguistic Theory* 24:501–570.

Stowell, Timothy Agnus. 1981. Origins of phrase structure. Doctoral Dissertation, Massachusetts Institute of Technology.

Van Urk, Coppe. in preparation. The syntax of phrasal movement: A Dinka case study. Doctoral Dissertation, Massachusetts Institute of Technology.

Van Urk, Coppe, and Norvin Waldemar Richards, III. 2015. Two components of long-distance extraction: Successive cyclicity in Dinka. *Linguistic Inquiry* 46:113–155.

Wechsler, Stephen, and Wayan Arka. 1998. Syntactic ergativity in Balinese: An argument structure based theory. *Natural Language & Linguistic Theory* 16:387–441.

Woolford, Ellen. 2006. Lexical case, inherent case, and argument structure. *Linguistic Inquiry* 37:111–130.

Yanti. 2010. A reference grammar of Jambi Malay. Doctoral Dissertation, University of Delaware.

ALTERNATIVE SEMANTICS FOR FOCUS AND QUESTIONS: EVIDENCE FROM SĀMOAN[*]

Vera Hohaus

University of Tübingen

vera.hohaus@uni-tuebingen.de

Anna Howell

University of Tübingen

anna.howell@uni-tuebingen.de

This paper investigates the semantics of the particle *'o* which marks noun phrases under a variety of conditions in Sāmoan, a Polynesian language. We argue that *'o* marks those noun phrases for which an alternative-semantic value needs to be calculated. Evidence for such an analysis comes from distribution, the absence of certain locality constraints, and intervention effects. Accordingly, the *'o*-marking of *wh*-phrases in questions provides evidence for an alternative-semantics approach to interrogatives and against a quantificational one.

1. Introduction

In Sāmoan, a Polynesian language with approximately 300,000 speakers, noun phrases are marked with the particle *'o* in a number of configurations. Among them, focused and topicalized noun phrases, *wh*-phrases as well as noun phrases associated with the exclusive particle *na'o* ('only') or with the free-choice item *so'o* ('any') receive *'o*-marking. The syntactic and information-structural function of this particle and its cognate *ko* in other Polynesian languages has been a subject of debate: For Sāmoan, the particle has been previously analyzed as nominative case marking by e.g. Downs (1949) or as serving some information-structural purpose (Pawley 1966; R. Clark 1969; Chapin 1970; Mosel and Hovdhaugen 1992). In other Polynesian languages, *ko* has received an analysis "... as a preposition, a copular preposition, a focus or topic complementizer, a pred morpheme, and a tense morpheme", to quote the overview in Massam, Lee, and Rolle (2006: 3). However, the semantic implications of any of these analyses have never been spelled out in detail. This paper offers a semantic perspective on the function of *'o* in Sāmoan. We take as a starting point the observation that the distribution of the particle *'o* correlates with the main constructions for which a semantics relying on a second tier of semantic interpretation, referred to as the alternative-

or focus-semantic value, has been proposed. We suggest that *'o* marks those noun phrases for which an alternative-semantic value (i.e. a semantic value different from the ordinary semantic value) needs to be calculated. As a consequence, we argue that the *'o*-marking found on *wh*-phrases indicates a semantics for interrogatives that operates on alternatives rather than one which treats *wh*-phrases as quantifiers.

The plot of the paper is as follows: In section 2 we discuss the data pertaining to the distribution of *'o*-marking of focus and topics as well as with exclusive particles and free-choice items. Section 3 then spells out the semantic analysis in detail. In section 4, we look at constituent questions in Sāmoan, and extend our analysis to explain the occurrence of *'o* with *wh*-phrases. Section 5 looks at further predictions of the proposed analysis, including predictions relating to island (in)sensitivity and intervention effects. Section 6 concludes and provides an outlook on extensions of the analysis in Sāmoan and other Polynesian languages.

2. Data

A primary motivation for the semantic account of the particle *'o* which we propose in this paper comes from its distribution: As we will show in this section, *'o*-marking occurs in the core constructions for which a second tier of semantic interpretation, the alternative- or focus-semantic value (Rooth 1985, 1992), has been suggested to play a role:

The particle *'o* in Sāmoan occurs with noun phrases that are the focus, as in (1b), or topic, as in (3), of a sentence.[1] This includes both new information focus, in (1b), as well as contrastive focus, in (2b), and extends also to cases of multiple focus, as indicated by data from the literature in (4).[2]

(1)   **New Information Focus**

    a.   *'O   ā   mea'ai na   'aumai e   Pita?*
         ALT. what food   TAM(past) bring   ERG. Peter
         'What food did Peter bring?'

---

[1] Unless otherwise indicated all data come from work with Sāmoan native speakers conducted in Germany, Hawai'i and Sāmoa. The material was designed following Matthewson (2004) and Matthewson (2011). The original orthography of the examples (especially with respect to diacritics) has been preserved. Abbreviations used in glosses are ABS. = absolutive case marker, ALT. = alternative marker, DEM. = demonstrative, DET. = determiner (specific, singular), DIR. = directional particle, EMPH. = emphatic particle, ERG. = ergative case marker, fut. = future, INDEF. = determiner (indefinite), NEG. = negation, PART. = particle, pl. = plural, PREP. = preposition, PRN. = pronoun, prog. = progressive, sg. = singular, and TAM = tense-aspect marker.
[2] We were however unable to replicate the finding of Chapin (1970)'s fieldwork that multiple *'o*-marked constituents are acceptable in Sāmoan.

b. **['O** *le* *talo*] *na* *aumai e* *Pita.*
   ALT. DET. taro TAM(past) bring ERG. Peter
   'Peter brought the TARO.'

(2) **Contrastive Focus**

a. *Na* *alu i* *Apia lou* *tinā?*
   TAM(past) go PREP. Apia your mother
   'Did your mother go to Apia?'

b. *Leai,* **['o** *lo'u tamā*] *na* *alu i* *Apia, 'a* **['o** *lo'u*
   no   ALT. my father TAM(past) go PREP. Apia but ALT. my

   tinā] *na* *nofo i* *le* *fale.*
   mother TAM(past) stay PREP. DET. house
   'No, my FATHER went to Apia but my MOTHER stayed at home.'
   (Mosel and So'o 1997: 52)

(3) **Topic**

   'Afai **[o** *mea'ai*]*, e* *fiafia Luka i* *panikeke.*
   if   ALT. food   TAM like Luke PREP. pancakes
   'As far as food is concerned, Luke likes pancakes.'

(4) **Multiple Foci**

   ['O *le* *tama*] **['o** *le* *teine*] *sa* *alofa i* *ai.*
   ALT. DET. boy   ALT. DET. girl TAM(past) love PREP. PRN.
   'The BOY loved the GIRL.'
   (Chapin 1970: 375)

Sāmoan *'o* also obligatorily co-occurs in the equivalent of English *only*, the exclusive particle *na'o*, as in (5), and in the equivalent of English *any*, the free-choice item *so'o*. Both items employ alternatives in their semantics. (See Rooth (1985, 1992) as well as Beaver and B. Z. Clark (2008) for analyses of the semantics of English *only*, and Krifka (1995), Aloni (2007), and Chierchia (2013) for alternative-semantics analyses of English *any*.)

(5) **Exclusive Particle**

   [*Na* *('o) *Luka*] *'o lo'o* *ia te* *a'u.*
   only   ALT. Luka TAM(prog.) be.with PRN.(1 sg.)
   'Only LUKE is currently here with me.'

(6) **Free-Choice Item**

   E *mafai* *ona tupu le* *ositioafaraiti i* **[*so**('o)**
   TAM possible that grow DET. arthritis   PREP. any+ALT.

   se *tagata*]*.*
   INDEF.(sg.) person
   'Anybody can get arthritis.'

In the examples above, *'o*-marking always co-occurs with fronting of the focus or topic, but there are examples of the *in-situ* use of *'o* as a means to mark focus in the literature, such as (7) and (8). However, we were unable to reproduce these examples in our own fieldwork. *'O*-marked noun phrases are only acceptable to the native speakers we have consulted when fronted. The exception to this generalization are noun phrases associated with the exclusive particle *na'o* ('only') or with the free-choice item *so'o* ('any'), both of which can occur either fronted, as in (5), or *in situ*, as in (9). One possible explanation for this pattern is that in Sāmoan the initial phrase is prosodically maximally prominent and speakers aim to align the different structural representations. (See also Calhoun (2013).)

(7) *Ai   lava se   mauoloa* [*o   lou tama*]*?*
     perhaps EMPH INDEF.(sg.) rich   ALT. your father
     'Perhaps your father is a rich person.'
     (Mosel and Hovdhaugen 1992: p.264, no. (6.71))

(8) . . . *'ae o'o a'e ia*   [*'o   le   'autaunonofo*],. . .
        but reach DIR. EMPH.   ALT. DET. harem
     '. . . but the wives went up to her.'
     (Mosel and Hovdhaugen 1992: p. 273, no. (6.141))

(9) *I   le   1960 na   pau ai   le   fuainumera o tagata*
     PREP. DET. 1960 TAM(past) limit PRN. DET. number   of people
     mamai i   le   3,000 ma i   le   1979 e   tusa   [*na*
     sick   PREP. DET. 3,000 and PREP. DET. 1979 TAM be.equal.to   only
     ('o*) le   10*].
     ALT. DET. 10

     'In 1960, it restricted the number of people that got sick to 3,000
      and in 1979, it was equal to only TEN.'

Intonational focus marking is also available in Sāmoan (cf. Calhoun 2013). An interesting question, which we have not pursued in this paper, is the interaction between these two focus-marking strategies. For example, we have noted above that both focus and topics receive *'o*-marking. It might be the case that Sāmoan differentiates sentences with *'o*-marked noun phrases as focus and sentences with *'o*-marked noun phrases as topic intonationally, like in Māori (cf. Bauer 1991). This remains to be tested empirically. Further investigation may also reveal evidence of competition between these two strategies, and in fact, Sasha Calhoun (p.c.) speculates that focus marking strategies might currently be in flux and an area of linguistic change in the language.

3. Alternative Semantics for Focus

In the previous section, we observed that *'o*-marking in Sāmoan occurs in exactly those constructions whose semantics makes use of alternatives. To explain this observation, we propose in this section that the function of the particle *'o* is to mark those noun phrases for which alternatives need to be calculated. To do so, we first need to spell out our assumptions about the way focus-alternatives are generated and manipulated by the grammar.

The basic idea under this type of account is that focus gives rise to the generation of alternatives. For example, in the sentence in (5), the *'o*-marking on the noun phrase *Luka* will give rise to the generation of the alternative set in (10). Focus-sensitive operators work with these alternatives. In the case of an operator like *only*, for example, the other alternatives are excluded as false.

(5)   [***Na 'o***   *Luka*] *'o lo'o*   *ia te*   *a'u.*
         only ALT. Luka TAM(prog.) be.with PRN.(1 sg.)
         'Only LUKE is currently here with me.'

(10)   {that John is here with me, that Mark is here with me,
         that Matthew is here with me, that … }

There are several ways of technically implementing an analysis of this kind for focus, namely structured propositions (cf. e.g. Krifka 1992), a two-tier semantics with focus- and ordinary-semantic values (cf. esp. Rooth 1985, 1992), and distinguished variables (cf. e.g. Beck 2006, to appear). Here, we will use distinguished variables, a framework in which, at the level of Logical Form, focus-marking is realized as a distinguished variable that is bound by an alternative-evaluating operator higher up in the structure. (In this case: the squiggle operator $\sim$.) Variables as we know them receive a value assignment via the function g. Distinguished variables receive their value assignment via a second assignment function h. Essentially, the distinguished variable tells us where to compute alternatives and of what type these alternatives are. Thus, (5) has the Logical Form in (11). When evaluated with respect to the assignment function h, the focused constituent, *Luka*, is replaced by a distinguished variable, *ii*, as in (12b).

(11)
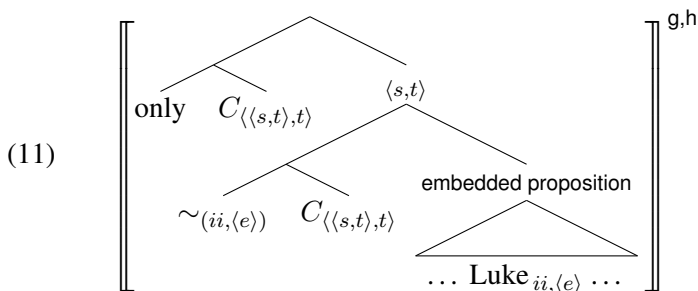
(12) a. ordinary semantic interpretation:
$[\![$ -embedded proposition- $]\!]^g = \lambda w.\, \text{be.here.with.me}_w(\text{Luke})$

   b. focus-semantic interpretation:
$[\![$ -embedded proposition- $]\!]^h = \lambda w.\, \text{be.here.with.me}_w(h(ii, \langle e \rangle))$

The squiggle operator $\sim$, in (13), unselectively binds distinguished variables and adds a presupposition regarding the value assignment to a free variable C of type $\langle \langle s, t \rangle, t \rangle$ it introduces. Namely, it adds the presupposition that the value assigned to this free variable is a subset of the set of alternatives generated by replacing the distinguished variable with all possible instantiations of it. Operators sensitive to alternatives such as *only* and its Sāmoan equivalent in (14) also come with a contextual variable C, which receives the same value as the contextual variable that comes with the squiggle operator. That is, they work with the alternatives the squiggle operator generated with the help of the distinguished variable. For the example from (5), this yields the desired truth conditions in (15), which state that all alternative propositions to the proposition that Luke is with me are false.

(13)   If $\alpha = [[\sim C]\,\beta]$, then for any g, h:
$[\![\alpha]\!]^g$ is only defined if $g(C) \subseteq \{p : \exists h\,[p = [\![\beta]\!]^h]\}$.
Then $[\![\alpha]\!]^g = [\![\beta]\!]^g$ and $[\![\alpha]\!]^h = [\![\beta]\!]^\emptyset$.

(14) $[\![\,na\,]\!] = [\![\,only\,]\!] = \lambda C_{\langle\langle s,t\rangle,t\rangle}.\,\lambda p_{\langle s,t\rangle}.\,\lambda w_{\langle s\rangle}.\,\forall q\,[C(q)\;\&\;q(w) \to q = p]$

(15) $\lambda w.\,\forall q\,[C(q)\;\&\;q(w) \to q = [\lambda w'.\,\text{Luke is with me in } w']]$
presupposition: $g(C) \subseteq \{p : \exists x \in D_{\langle e\rangle}\,[p = [\lambda w.\,x \text{ is with me in } w]]\}$
with e.g. $g(C) = \{\lambda w.\,\text{Luka is with me in } w,\; \lambda w.\,\text{Paulo is with me in } w,\\ \lambda w.\,\text{Ioane is with me in } w, \ldots\}$

In summary, this system of focus interpretation has three main ingredients: (i) distinguished variables to generate alternatives, (ii) alternative-evaluating operators such as $\sim$ to bind distinguished variables and introduce a set of alternatives into the semantics via the presupposition, and (iii) operators sensitive to alternatives such as *only*. This system is used for focus interpretation but not limited to focus. There are a number of other alternative-evaluating operators and operators sensitive to alternatives (cf. also Beck 2006), including questions, free-choice and negative polarity items, disjunction and certain quantifiers. Thus, a prediction of this analysis of Sāmoan *'o* is that the particle show up in other constructions that have been analyzed to involve alternative-evaluating and alternative-sensitive operators. This prediction is borne out. As we have seen in section 2, the particle *'o* not only marks focus but also topics, and it co-occurs in the free-choice item *so'o* ('any'). *Wh*-phrases, which we will discuss in detail in section 4, are also obligatorily *'o*-marked as is disjunction, which we will discuss in section 5.

3.1.   Excursus on the Syntax of ‘*o*-Constructions

Let us point out that the semantic analysis we propose has syntactic implications as well. Polynesian *ko*-constructions such as the Sāmoan example in (16) can potentially be analyzed either as clefts, as in (17), as pseudoclefts, as in (18), or as dislocation structures, as in (19). (See especially Potsdam (2009) as well as Potsdam and Polinsky (2011) for discussion.) Under the analysis presented above, interpretation proceeds from structures that are not in any way bi-clausal. Rather, the ‘*o*-marked constituent has undergone movement, as in (19). (This is also the analysis adopted in Pizzini (1971).)

(16)   ‘*O   Ioane e   umi.*
          ALT. John   TAM long
          ‘John is tall.’

(17)   **Cleft**:
          It is John [RelCl *who is tall*].

(18)   **Pseudo-Cleft**:
          [RelCl *Who is tall*] *is John*.

(19)   **Movement of Noun Phrase
          and Verb Raising in Derivation of VSO**
          (cf. e.g. Collins, to appear):
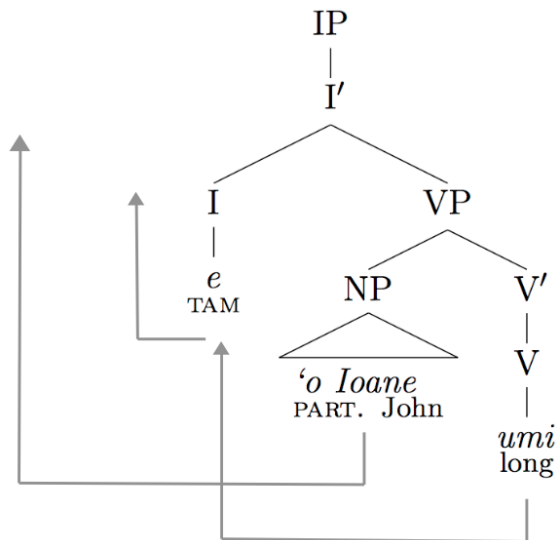
Because of the lack of an overt copula and of overt expletives in Sāmoan, arguments in favor of such a dislocation analysis are not straightforward (cf. also Potsdam & Polinsky 2011). However, evidence in favor of the structure in (19) above comes from the following: First, tense-aspect markers are ungrammati-

cal in '*o*-marked constituents, as in (20), contrary to expectations for a bi-clausal structure. (Mosel and Hovdhaugen (1992: 500) also observed that the '*o*-marked constituent in Sāmoan is "...unmarked for tense-aspect or mood and does not accept any verb or negative particle.")

(20)   **[(*Na)   'o   le   meleni]** [*na   'ai e   Sioane analeilā*].
          TAM(past) ALT. DET. melon   TAM(past) eat ERG. John   yesterday
        'It was the melon that John ate yesterday.'

Second, the possibility of multiple fronted '*o*-constituents, as in (4), is unexpected under both, a cleft- and a pseudocleft-analysis. Third, as far as they are accepted by native speakers, the occurrence of '*o*-marked constituents *in situ*, as in (7) to (8) provides additional evidence against a bi-clausal structure. Finally, the unavailability of headless relative clauses as arguments elsewhere in Sāmoan, as discussed in Mosel and Hovdhaugen (1992: 633), also weakens the empirical support for a cleft-analysis under which fronted '*o* marked constituents are headless relative clauses. We conclude from this brief discussion that for Sāmoan, a dislocation analysis is preferable.

4.   Interrogatives

As mentioned in the previous section, a prediction made by this analysis is that '*o*-marking should be found in other environments where distinguished variables are used in the semantic calculation. *Wh*-questions provide an interesting opportunity to test this prediction, because certain approaches derive the meaning of *wh*-questions from a set of alternatives generated by a distinguished variable in the *wh*-phrase. Interestingly, Sāmoan *wh*-phrases are also marked by the particle '*o*. Based on the occurrence of '*o*-marking in interrogatives, we argue in this section that questions in Sāmoan should receive an alternative-semantics analysis rather than a quantificational one.

4.1.   Data

Constituent questions in Sāmoan also obligatorily employ '*o*, as illustrated in (21) below.[3] (See also the example in (1a) from section 2.)

(21)   **[*('O) ai]** *na   sau ma le   talo?*
          ALT. who TAM(past) come with DET. talo
        'Who came with the taro?'

---

[3] In order to avoid unnecessary complexity, we focus only on argument questions here.

In embedded questions, like the one in (22), the question particle *pē*, is additionally present in the complementizer of the embedded clause.[4] Examples from the literature such as the one in (23) indicate that the question particle *pē* may have also been realized overtly in matrix questions in Sāmoan in the past, though matrix sentences marked with the question particle were rejected by our consultants.

(22) *'Ou te iloa [**po** ['**o** ai] e alofa iai Sina],...*
      I TAM know Q ALT. who TAM love PREP.+PRN. Sina
      'I know who Sina loves...'

(23) ***Po** [o le tusi a ai] sa e fa'aaogā?*
      Q ALT. DET. book of who TAM you use
      'Whose book did you use?'
      (Downs 1949: 44)

4.2. Approaches to the Semantics of Interrogatives

Questions denote the set of possible answers to them, an idea going back to Hamblin (1973). (Cf. also Krifka (2011) for an introduction to the semantics of questions.) There are multiple ways to derive these sets compositionally.

Quantificational Analyses of Interrogatives. The most prominent approach (e.g. Karttunen (1977) as well as Groenendijk and Stokhof (1984)), developed primarily for English, derives this set of propositions by analyzing *wh*-words as existential quantifiers, which obligatorily undergo movement to produce an interpretable structure, and by the covert question operator Q, in (25b). Thus, the question in (24) is assigned the Logical Form in (26), where *wh*-pronouns have a lexical entry as in (25a), yielding the set of propositions in (27).

(24) *What did Mary eat?*

(25) a. $[\![\,what\,]\!] = \lambda P_{\langle e,\langle\langle s,t\rangle,t\rangle\rangle}.\,\lambda p_{\langle s,t\rangle}.\,\exists x\,[(P(x))(p)]$
     b. $[\![\,Q\,]\!] = \lambda p_{\langle s,t\rangle}.\,\lambda q_{\langle s,t\rangle}.\,[p = q]$

(26) $[_{\langle\langle s,t\rangle,t\rangle}\;what\;[_{\langle e,\langle\langle s,t\rangle,t\rangle\rangle}\;1\;[_{\langle\langle s,t\rangle,t\rangle}\;Q\;[_{\langle s,t\rangle}\;Mary\,[\,eat\;t_{1,\langle e\rangle}\,]]\,]]\,]$

(27) $[\![\,what\,]\!](\lambda x_{\langle e\rangle}.\,\lambda q_{\langle s,t\rangle}.\,[q = [\lambda w.\,\mathsf{eat}(x)(\mathsf{Mary})(w)]])$
     $\Leftrightarrow \lambda p_{\langle s,t\rangle}.\,\exists x\,[p = \lambda w.\,\mathsf{eat}(x)(\mathsf{Mary})(w)]$
     e.g. {that Mary ate taro in w, that Mary ate some papaya in w}

In the interpretation in (27), the question operator Q combines with a proposition and turns it into a set of propositions. The existential quantifier introduced by the *wh*-word subsequently binds the trace left by its movement, yielding the set of possible answers to the question in (24) of the shape that Mary ate some thing.

---

[4] The question particle *pē* is realized as *pō* when followed by an *'o*.

Alternative Semantics for Questions. A second, more recent approach (e.g. Kratzer and Shimoyama (2002) as well Beck (2006)) capitalizes on the similarity between interrogative semantics and focus semantics, and analyzes *wh*-words on par with focus-marked constituents as elements which introduce distinguished variables, as shown in (28), and which thus generate alternatives. The question operator Q, in (29), is another alternative-evaluating operator, just like the squiggle operator, in (13).

(28)   a. $[\![\,what_{ii,\langle e\rangle}\,]\!]^g$ UNDEFINED
         b. $[\![\,what_{ii,\langle e\rangle}\,]\!]^h = h(ii,\langle e\rangle)$

(29)   If $\alpha = [Q_{i,\langle\tau\rangle}\ \beta]$, then for any g, h:
         $[\![\alpha]\!]^g = \{p : p = \exists x \in D_{\langle\tau\rangle} : [\![\beta]\!]^{h[x/i]}\}$.

Under this account, movement is not required at Logical Form. Rather, the *wh*-pronoun remains *in situ* and its distinguished variable is bound by the Q-operator, as illustrated in (30). Like the squiggle operator $\sim$, Q generates a set of propositions. It does so by taking the alternative semantic value of the embedded proposition and replacing the distinguished variable introduced by the *wh*-pronoun with all of its possible instantiations. Thus, only the alternative semantic value of the embedded proposition, in (31), is used in the generation of the question meaning. Its ordinary semantic value is undefined.

(30)
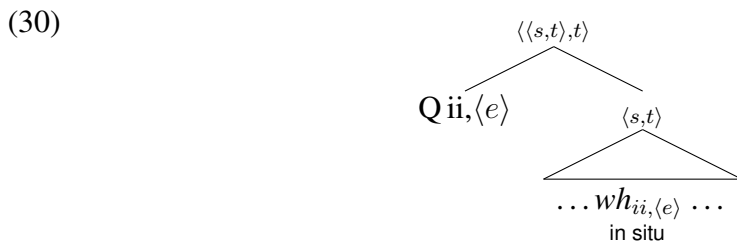
(31)   a.   ordinary semantic value:
               $[\![\text{ -embedded proposition- }]\!]^g$ UNDEFINED
         b.   focus-semantic value:
               $[\![\text{ -embedded proposition- }]\!]^h = \ldots (h(ii,\langle e\rangle))$

The obligatory *in-situ* interpretation of *wh*-pronouns, while somewhat unintuitive in *wh*-fronting languages, is desirable in *wh-in-situ* languages like Japanese, in which *wh*-phrases do not undergo covert movement (cf. Shimoyama 2006). However, an alternative-semantics analysis of questions has also been argued for in some *wh*-fronting languages, in particular Tinglit, a Na-Dené language spoken in Southeast Alaska and Western Canada (Cable 2010).

4.3.   Applying the Analysis to Sāmoan

The perspective from Sāmoan suggests a unified treatment of interrogatives and focus, and thus is another language that provides support for an alternative-based analysis of interrogatives. We suggest that *wh*-words in Sāmoan questions such as (21) are interpreted *in situ*, yielding the Logical Form in (32). The denotation, in (33), is the set of propositions that are possible answers to the question. This is the same set of propositions which a quantificational analysis would generate, but they have been calculated using the focus-alternatives of the proposition embedded under the focus-sensitive operator Q.

(21)   **['O   ai]** *na   sau ma le   talo?*
        ALT. who TAM(past) come with DET. taro
    'Who came with the taro?'

(32)
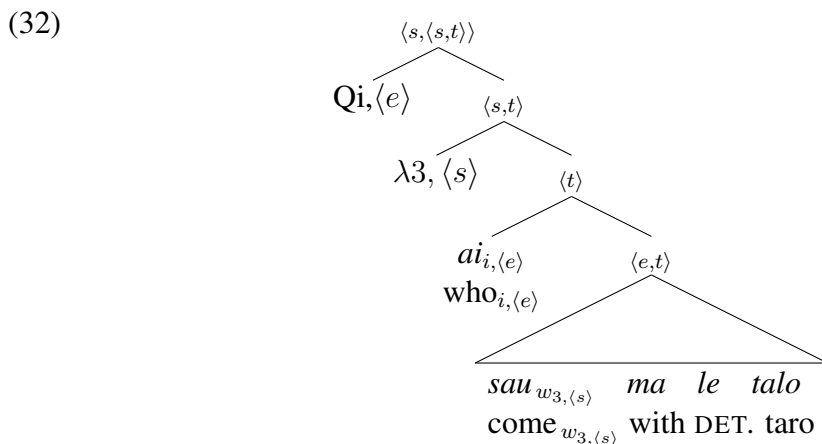
(33) $\{p : \exists x \in D_{\langle e \rangle} : p = [\![\,[\lambda 3, \langle s \rangle\, [[ai_{\,i,\langle e \rangle}]\,[sau_{\,w_3,\langle s \rangle}\, ma\ le\ talo]]]\,]\!]^{h[x/i]}\} \Leftrightarrow$
    $\{p : \exists x \in D_{\langle e \rangle} : p = \lambda w.\ x$ came with the taro in $w\}$
    e.g. $\{$that Malia came with the taro in w, that Eseta came with the taro in w, that Alofa came with the taro in w,... $\}$

This analysis for Sāmoan makes some predictions about sensitivity to syntactic islands and the presence of intervention effects in constructions with *'o*, which will be explored in the next section.

Interim Summary.   The analysis proposed thus far for *'o*-marking in Sāmoan can be summed up as follows: Constituents containing distinguished variables, including constituents containing focus and *wh*-pronouns, are marked with the particle *'o*. Two covert operators, the squiggle operator $\sim$ for focus and Q for questions bind these distinguished variables to generate sets of alternatives, which yield question meanings, or are used by focus-sensitive particles like *na'o* ('only').

We hypothesize that the inventory of alternative-sensitive operators in Sāmoan (in Table 1) additionally includes the free-choice item *so'o* ('any') and a covert ASSERT-operator for topics.

alternative-sensitive operators	alternative-evaluating operators	distinguished variables
na 'o (= only)	covert ~	
so 'o (= any)	*pē* or covert Q	marked by *'o*
covert ASSERT		

Table 1: Alternatives in Interpretation: The View from Sāmoan

5. Predictions: Island (In)sensitivity and Intervention Effects

Aside from arguments based on the distribution of the particle *'o*, diagnostics used in the semantic literature to test for an alternative semantics, such as island insensitivity (Rooth 1985; Shimoyama 2006) and the presence of intervention effects (Beck 2006; Erlewine and Kotek, to appear) provide additional evidence for the analysis proposed in the previous sections.

5.1. Island Insensitivity of Focus Association

A primary motivation for introducing alternatives into the interpretation component of the grammar comes from the fact that association with focus is not subject to the same locality constraints as movement or Quantifier Raising are. Whereas Quantifier Raising is blocked in English by a relative clause boundary, no such locality restrictions affect association with focus, as illustrated by the contrast in (34) below. Crucially, the exclusive particle *only* can associate with a focused constituent across a syntactic island.

(34)   a.   *I know the person* [$_{RelCl}$ *that* [*every girl*] *loves*].
             = 'The person that I know is such that every girl loves them.'
             ≠ 'For every girl, I know the person that is such that she loves them.'

       b.   *I only know the person* [$_{RelCl}$ *that MARY loves*].
             = 'Only for Mary do I know the person that she loves.'

Similarly in Sāmoan, although relative clauses are otherwise islands for movement, as illustrated by (35), we find that *na'o* ('only') can associate with a focused constituent across a relative clause: In (37) in the context of (36), association is with the focused noun phrase *Malia* across the relative clause boundary.

(35) *[*'O ai*] *'o lo'o*   Malia i   le   [*tama*
      ALT. who TAM(prog.) Mary PREP. DET. boy

      [RelCl *e*   *alofa i*   *ai*   _]]*?*
        TAM love PREP. PRN.

      (Lit.) 'Who is Mary talking to the boy that _ loves?'

(36) <u>Context</u>: Sina is very well informed. She is always the first to know who
      has asked whom on a date, and who is in love with whom. That's why
      shortly after three girls move to town, some of the boys in the village ask
      Sina whether she has any information about the new girls. She answers:

(37) [***Na 'o*** [*le*   *tagata lava* [RelCl *e*   *alofa i*   *ai*
      only ALT. DET. person EMPH   TAM love PREP. PRN.

      Malia]]]   *ou te*   *iloa.*
      Mary]   I   TAM know

      'I only know the very person who MARY loves.'

The example suggests that an adequate analysis of the particle *na'o* should not
involve movement out of the relative clause but rather employs association with
focus alternatives.

5.2. Intervention Effects

A second prediction made by our analysis of *'o*-marking in Sāmoan is that we
should find certain interactions between alternative-evaluating operators. Beck
(2006) proposes that the presence of the squiggle operator $\sim$, an unselective
binder of distinguished variables, at Logical Form, separating a distinguished vari-
able from the alternative-evaluating operator that is intended to bind it is the cause
of so-called intervention effects. One such configuration that gives rise to an in-
tervention effect is sketched in (38). Intervention effects have been observed in
a variety of languages when a certain class of operators including exclusive par-
ticles, certain quantifiers and negation intervene between a *wh*-pronoun and its
evaluating *Q*-operator, as in the Korean example in (39).

(38) *[Qii ... [$\sim$ ... [*wh*-phrase$_{ii}$ ... α_i]]]

(39) *Minsu-man nuku-lûl   po-ass-ni?*
      Mina-only who-ACC. invite-past-Q
      (intended) 'Who did only Mina invite?'
      (Beck 1996: p. 28, ex. (55-a))

The ungrammaticality of (39) indeed follows from the semantics of the $\sim$-operator:
It unselectively binds all distinguished variables in its scope, preventing the *Q*-
operator from binding the distinguished variable of the *wh*-pronoun. (The reader

is referred to Beck (2006) for details of the proposal.) Under this analysis, the presence of intervention effects can be used as a diagnostic for constructions which employ semantic alternatives.[5] (See also Beck, to appear, as well as Erlewine and Kotek, to appear.) If the insertion of an intervener results in ungrammaticality, we have further evidence for an alternative-semantics analysis. Thus, our proposal for the semantics of Sāmoan constituent questions, taken together with Beck (2006)'s account, makes the prediction that intervention effects should arise in questions.

Unfortunately, it is well know that movement obviates intervention effects (Beck 1996; Pesetsky 2000), so in order to test this prediction for questions, we require *in-situ wh*-phrases. In Sāmoan, however, *wh*-phrases cannot remain *in situ*. Moreover, multiple questions, where intervention effects have been observed in languages like English and German, are ungrammatical, as is true more generally across Polynesian languages. However, in addition to *wh*-questions, intervention effects have been observed in another type of question, in alternative questions. The term alternative question here refers to a particular reading of an apparent *yes/no*-question containing a disjunction under which possible answers to the question are the two disjuncts rather than *yes* and *no*. For example, English (40a) has both, an alternative question and a *yes/no*-question reading, as we can see from the possible answers in (40b) and the paraphrases in (41). In the presence of an intervener such as negation, in (42), however, the alternative-question reading is unavailable. (For discussion see e.g. Beck and Kim (2007).)

(40)   a.   <u>Question</u>:
               Did Sally teach Syntax or Semantics?

        b.   <u>Possible Answers</u>:
               Yes./ No.
               Syntax./ Semantics.

(41)   a.   <u>Alternative Question Reading</u>:
               'Which of Syntax and Semantics did Sally teach?'

        b.   <u>Yes/No-Question Reading</u>:
               'Did Sally teach Syntax or Semantics, or not?'

(42)   a.   <u>Question</u>:
               Didn't Sally teach Syntax or Semantics?

        b.   <u>Possible Answers</u>:
               Yes./ No.
               # *Syntax./ Semantics.*

---

[5] Other analyses (Beck 1996; Tomioka 1997; Haida 2007; Mayr 2014) attribute intervention effects to other grammatical properties of interveners and make no predictions regarding the interaction of multiple alternative-evaluating operators.

As this type of question is also available in Sāmoan, it can be used as a diagnostic for alternative semantics. An example of a Sāmoan alternative question is given in (43). Possible answers to (43) are, for instance, *'O le māketi.* ('To the market.') as well as *Leai.* ('No.').

(43) *E   ō   le fanau   a Tavita i   le   māketi*
      TAM go(pl.) the children of David PREP. DET. market
      po 'o   *le   falesā?*
      or ALT. the church
      'Are David's kids going to the market or the church?'

Just like in English, negation causes an intervention effect in this type of questions. In the Sāmoan example in (44), *'O le māketi.* ('The market.') is not a possible answer to the question, whereas *Leai.* ('No.') is. Note that the disjunction *po'o* ('or') also contains *'o*, as our analysis predicts if disjunctions employ alternatives.

(44) *E   le'i   ō   le   fanau   a Tavita i   le   māketi*
      TAM NEG. go(pl.) DET. children of David PREP. DET. market
      po 'o   *le   falesā?*
      or ALT. DET. church
      'Aren't David's kids going to the market or the church?'

Thus, the presence of intervention effects in Sāmoan alternative questions provides further evidence for our alternative-semantics analysis of *'o*.

6.   Summary and Outlook

The paper offered a perspective from semantic theory on the function of the particle *'o* in Sāmoan. The particle marks those noun phrases for which alternatives need to be calculated during interpretation. Evidence for this analysis comes from the distribution of *'o* as well as from intervention effects and the absence of locality constraints in association with focus: The particle marks topic, focus, and *wh*-phrases, and obligatorily co-occurs with the exclusive particle *na'o* ('only'), the free-choice item *so'o* ('any'), and the disjunction *po'o* ('or'). Although constituent questions in Sāmoan involve movement, the data provide evidence in favor of an alternative-based approach to the semantics of interrogatives and against a quantificational analysis. At this point, we can identify two pathways for extensions of this analysis, first, within the language, and second, across Polynesian languages:

In Sāmoan, the particle *'o* also occurs in identity statements, as in (45), and in appositives such as (46). An alternative-semantics approach to these con-

structions might provide a new perspective on their syntax and semantics.

(45) **[*O*   *puaa fanua*]   [*o*   *puaa popoto ia*].
       PART. pig   cultivated PART. pig   clever EMPH.
       'The domestic pigs are very clever pigs.'
       (Mosel and Hovdhaugen 1992: p. 503, no. (11.17))

(46) . . . *ma lana tama,* [*'o*   *Sina*].
        and her   child   PART. Sina
       '. . . and her child, Sina.'
       (Mosel and Hovdhaugen 1992: p. 500, no. (11.4))

Another pathway for further investigation is the extension of the analysis to *ko* in other Polynesian languages, which seems to share the distribution of its Sāmoan counterpart. An example is the interrogative from Niuean in (47) below.

(47) **[*Ko*   *hai*]** *ne*   *lalaga e*   *kato* *ē?*
       PART. who TAM(non-fut.) wove ABS. basket DEM.(sg.)
       'Who wove this basket?'
       (Massam, Lee, and Rolle 2006: p. 15, no. (25))

This shared distribution, however, needs to be investigated against the syntactic micro-variation in the relevant constructions already observable in the literature (cf. e.g. Bauer 1991; Cook 1999; Pearce 1999; Massam, Lee, and Rolle 2006). A diachronic perspective on this micro-variation might in the end help us better understand the division of labor at the interface of syntax, semantics, and phonology when it comes to constructions that invoke alternatives.

References

Aloni, Maria. 2007. Free choice, modals, and imperatives. *Natural Language Semantics* 15:65–94.

Bauer, Winifred. 1991. Māori *ko* again. *Te Reo* 34:3–14.

Beaver, David I., and Brady Z. Clark. 2008. *Sense and sensitivity: How focus determines meaning.* Malden: Wiley-Blackwell.

Beck, Sigrid. 1996. *Wh*-constructions and transparent Logical Form. Doctoral dissertation, Eberhar Karls Universität Tübingen.

Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14:1–56.

Beck, Sigrid. To appear. Focus-sensitive operators. *Handbook of information structure*. Oxford: Oxford University Press, to appear.

Beck, Sigrid, and Shin-Sook Kim. 2007. Intervention effects in alternative questions. *The Journal of Comparative Germanic Linguistics* 9:165–208.

Cable, Seth. 2010. *The grammar of Q: Q-particles,* wh-*movement, and pied-piping*. Oxford: Oxford University Press.

Calhoun, Sasha. 2013. The interaction of prosody and syntax in Samoan focus marking. Ms., Victoria University of Wellington, Wellington, Nz.

Chapin, Paul G. 1970. Samoan pronominalization. *Language* 46:366–78.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention.* Oxford: Oxford University Press.

Clark, Ross. 1969. Some verbless sentences in Samoan. *Oceanic Linguistics* 8:105–19.

Collins, James N. To appear. Diagnosing predicate fronting in Samoan. *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL)* 32.

Cook, Kenneth W. 1999. Hawaiian *he*, *'o*, and *i*: Copular verbs, prepositions, or determiners? *Oceanic Linguistics* 38:43–65.

Downs, E. A. [2]1949. *Everyday Samoan: Samoan grammar.* Papakura: Southern Reprints.

Erlewine, Michael, and Hadas Kotek. To appear. Covert pied-piping in English multiple *wh*-questions. *Linguistic Inquiry*.

Groenendijk, Jeroen, and Martin Stokhof. 1984. Studies in the semantics of questions and the pragmatics of answers. PhD thesis, Universiteit van Amsterdam.

Haida, Andreas. 2007. The indefiniteness and focusing of *wh*-words. Doctoral dissertation, Humboldt-Universität zu Berlin.

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10:41–53.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1:1–44.

Kratzer, Angelika, and Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. *Proceedings of the Tokyo Conference on Psycholinguistics (TCP)* 3, 1–25.

Krifka, Manfred. 1992. A compositional semantics for multiple focus constructions. *Informationsstruktur und Grammatik*, ed. Joachim Jacobs, 17-53. Opladen: Westdeutscher Verlag.

Krifka, Manfred. 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25:209–57.

Krifka, Manfred. 2011. Questions. *Semantics: An International Handbook of Natural Language Meaning*, eds. Claudia Maienborn, Klaus von Heusinger and Paul Portner, 1742-85, Berlin: De Gruyter.

Massam, Diane, Josephine Lee, and Nicholas Rolle. 2006. Still a preposition: The category of *ko*. *Te Reo* 49:3–37.

Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70:369–451.

Matthewson, Lisa. 2011. Methods in crosslinguistic formal semantics. *Semantics: An international handbook of natural language meaning*, eds. Claudia Maienborn, Klaus von Heusinger, and Paul Portner, 268-84, Berlin: De Gruyter.

Mayr, Clemens. 2014. Intervention effects and additivity. *Journal of Semantics* 31:513-554.

Mosel, Ulrike, and Even Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Scandinavian University Press.

Mosel, Ulrike, and Ainslie So'o. 1997. *Say it in Samoan.* Canberra: Pacific Linguistics.

Pawley, Andrew. 1966. Samoan phrase structure: The morphology-syntax of a Western Polynesian language. *Anthropological Linguistics* 8:1–63.

Pearce, Elizabeth. 1999. Topic and focus in a head-initial language: Māori. *Proceedings of the Annual Meeting of the Austronesian Formal Linguistics Association (AFLA)* 6, 249–63.

Pesetsky, David. 2000. *Phrasal movement and its kin.* Cambridge: The MIT Press.

Pizzini, Quentin A. 1971. An analysis of predicate complement constructions in Samoan. *Oceanic Linguistics* 10:38–62.

Potsdam, Eric. 2009. Austronesian verb-initial languages and *wh*-question strategies. *Natural Language and Linguistic Theory* 27:737–71.

Potsdam, Eric, and Maria Polinsky. 2011. Questions and word order in Polynesian. *Topics in Oceanic Morphosyntax*, eds. Claire Moyse-Faurie, and Joachim Sabel, 83–109, Berlin: De Gruyter.

Rooth, Mats. 1985. Association with focus. Doctoral dissertation, University of Massachusetts at Amherst.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1:75–116.

Shimoyama, Junko. 2006. Indeterminate phrase quantification in Japanese. *Natural Language Semantics* 14:139–73.

Tomioka, Satoshi. 1997. Pragmatics of LF intervention effects: Japanese and Korean *wh*-interrogatives. *Journal of Pragmatics* 39:1570–90.

PERSON-BASED ORDERING OF PRONOMINAL CLITICS IN RIKAVUNG PUYUMA: AN INVERSE ANALYSIS[*]

Haowen Jiang   Loren Billings
Rice University   Hankuk University of Foreign Studies
haowen@rice.edu   sgnillib@gmail.com

This study argues that in Rikavung Puyuma pronominal clitics' variations in form are best accounted for by means of their linear position with respect to either other clausal elements or each other. Not only do these pronouns exhibit no case distinctions, and thus constitute just a single paradigm, they also demonstrate ordering relative to each other based only on grammatical person. Given that a first-person clitic pronoun *always* precedes a second-person one, and that the clitic /ta(w)/ is added specifically in front of both clitic pronouns to indicate second person acting upon first person, we also argue that /ta(w)/ is an inverse marker and that there are thus no overt third-person clitic pronouns.

Many Austronesian languages of Taiwan and the Philippines attest two sets of pronominal clausal clitics. If these co-occur (i.e., in a transitive clause), then they can cluster in several patterns.[1] The most common is for the Actor (usually bearing ERG case) to be first (Lee and Billings 2008:195–197; Yen and Billings 2014). Far less frequent is the opposite ordering, at least in those languages where the transitive Actor is never the subject (Quakenbush and Ruch 2008). In accusatively aligned Mantauran Rukai (southern Taiwan) both active and passive clauses require the subject to be initial within the clitic-pronoun cluster (Yen and Billings 2011). Seediq (northwestern Taiwan) also orders its clitic pronouns with the subject first (as Holmer and Billings 2014 argue). As another type, most Central Philippine languages (Bloomfield 1917; Kaufmann 1916; Lee and Billings 2008) also order a monosyllabic pronoun first in the clitic cluster. These languages thus utilize prosodic weight as their primary pronoun-ordering criterion. As yet another type, the current study discusses cluster-internal ordering based only on grammatical person: previously found in most Danao and Manobo languages, both subgroups spoken primarily in the southern

[1] It is also possible for the pronominal clitics *not* to form a cluster. That other type is found elsewhere in Puyuma as well as in neighboring Paiwan, where the Actor immediately precedes the lexical verb and the Undergoer (if overt) follows right after it (Huang et al. 1999:186, 188).

Philippines, and one dialect of Atayal, spoken in north-central Taiwan (Chang 2012; Holmer and Billings 2014:122–123; Hung and Billings 2009/2010; Kaufman 2010; Peng and Billings 2008). In each of the aforementioned languages with person-based ordering, through various means, the two pronouns in the morphosyntactic-clitic cluster are unambiguously differentiated as to their forms so that the ordering does not obscure the two pronouns' respective (semantic) roles or (syntactic) relations. In all the known languages of the area that order two speech-act participants (hereafter abbreviated as *SAP*s) using grammatical person as the primary ordering criterion, the first-person pronoun precedes the second-person form. In addition, in all such languages an SAP pronoun precedes any third-person form. Languages can also use a mixture of ordering criteria. For instance, all known Atayalic languages order an SAP before a third-person pronoun but two dialects (of Atayal proper) order clusters of SAP clitic pronouns based only on prosodic weight (Billings and Kaufman 2004:17; Li 1995:40/2004:403, citing "Mei (p.c.)"; Liao 2004:285–296, 2005).

　　　　Rikavung, a subdialect of Puyuma (spoken in southeastern Taiwan), also utilizes grammatical person in ordering its first- and second-person clitic pronouns relative to each other. (In our analysis, there are no overt third-person clitic pronouns.) Despite extensive variation in Rikavung's various clitic-pronominal forms, we argue that none of these forms encodes a case distinction. Rather, we demonstrate in this paper that the variation is positionally defined allomorphy, along the lines of the analysis of Mantauran Rukai in Zeitoun (2007) and developed further in Yen & Billings (2011). Each of the variant forms in (1)—whether between parentheses, indicating optionality, as in (1a–b, e), or not, as in (1d), where each 2SG form is required in at least certain environments—can be used for both semantic roles/syntactic relations.[2]

(1)　　　Clitic-pronoun inventory
　　　　a. ku (~ ka)
　　　　　　1SG
　　　　b. mi (~ niam)
　　　　　　EXCL1PL
　　　　c. ta
　　　　　　INCL1PL
　　　　d. u ~ nu
　　　　　　2SG
　　　　e. mu (~ nmu)
　　　　　　2PL

---

[2] Compare (1) with the pronominal inventory in Tamalakaw Puyuma (Tsuchida 1980:196, 1992:744, 1995:795). That neighboring subdialect is discussed further in section 2 below. To our knowledge, the only publication to date reporting sentential Rikavung data is by Suenari (1969), who spells the name as *Rikavong*. All sentential data in this study from Rikavung, shown in the International Phonetic Association's transcription, come from our own field notes.

More specifically, the clitic pronouns' *forms* display neither ergative nor accusative alignment. Elsewhere in the grammar of Rikavung (but not in the clitic-pronominal system), there is clear ergative alignment.[3] We demonstrate that syntactic relations do not play a role in any of the variations in (1a–b, d–e).

       This study consists of just two body sections. The first of these lays out the argument alignment in Rikavung based on the pronouns' positioning and forms, demonstrating that there is just a single, case-neutral paradigm of clitic pronouns. The latter section then presents and justifies our inverse analysis of Rikavung, where we show that it is grammatical person that determines the order of two pronouns relative to each other within the clitic cluster.

1.   Argument Alignment and Variation in Pronominal Clitics

This section argues that there is only one paradigm of clitic pronouns in Rikavung. As such, none of the variation found in four out of the five pronominal forms listed in (1a–e) above is used to encode morphological case. Each instance of variation is positionally determined allomorphy. As such, the pronominal *forms* do not contribute to argument alignment. However, their *positioning* does reveal a relatively marked kind of alignment.

1.1.   Pronominal Positioning within the Rest of the Clause

Here we address the ordering of bound pronouns relative to free elements. This then establishes that these morphemes are clitics rather than affixes. The clitic pronouns' *positioning* exhibits neither accusative nor ergative but rather horizontal alignment, where both of a transitive verb's arguments behave differently from an intransitive verb's only argument (Comrie 2013:29).

       The sole pronoun of an intransitive clause can be analyzed as having Wackernagel positioning, after some clause-initial element (and thus it is clearly a clitic). For example, compare the positioning of the INCL1PL pronoun after the verb, in (2a), with its positioning between Neg and the verb, as (2b) shows.

---

[3] Only SG-personal *nouns* show clear ergative alignment, as to their case-markers' forms:

(i)	ta=	sukun-anaj **i**	**misak**	(iii)	ta(w)=	**ta=**	sukun-anaj <u>ni</u> <u>misak</u>
	INCL1PL=	push-TR *ABS*	(name)		INV=	INCL1PL=	push-TR *ERG* (name)
	'Let's push Misak.'				'Misak pushed us.'		
(ii)	s<əm>ukun	i	misak	(iv)	s<əm>ukun	=ta	kani misak
	<INTR>push	*ABS*	(name)		<INTR>push	=INCL1PL *OBL*	(name)
	'Misak pushed {someone/something}.'				'We pushed Misak.'		

In (i) the INCL1PL Actor adds a cohortative reading. With both PL-personal and definite-common nouns, the ABS and ERG cases are neutralized (distinct from OBL case), whereas with indefinite-common nouns, the ERG and OBL cases are neutralized (distinct from ABS case). See Jiang (2013) re these markers in Rikavung and Teng (2009:827) for those in three other varieties.

(2) a. s<əm>ənaj   =ta
        <INTR>sing   =INCL1PL
        'Did we sing?'

    b. ħa(z�̪i)   =ta   s<əm>ənaj
        NEG   =INCL1PL   <INTR>sing
        'Didn't we sing?'

Because in both of (2a–b) the clitic is adjacent to the verb, albeit not phonologically affiliated with it in (2b), we cannot eliminate the possibility that the pronoun is of the head-adjacent type rather than merely after the first element of the clause. The important point here, however, is that the pronoun is a clitic (rather than an affix), not exactly which specific type of clitic it is.

    By contrast, both of a transitive verb's pronouns immediately precede the lexical verb regardless of whether Neg is present. The pair in (3a–b) shows transitive clauses with only the Actor realized as a pronoun (and the Undergoer unexpressed); the data in (4a–b) show both arguments as pronominal clitics.[4]

(3) a. <u>ta</u>=   sukun-anaj
        INCL1PL=   push-TR
        'Did we push {her/him/them}?'

    b. ħa(z̪i)   =<u>ta</u>   sukun-an
        NEG   =INCL1PL   push-TR.DEP
        'Didn't we push {her/him/them}?'

(4) a. {<u>ku</u>/<u>ka</u>}=   {*u/nu}=   sukun-anaj
        1SG=   2SG=   push-TR
        'I pushed you.'

    b. ħa(z̪i)   ={<u>ku</u>/<u>ka</u>}   ={*u/nu}   sukun-an
        NEG   =1SG   =2SG   push-TR.DEP
        'I didn't push you.'

---

[4] Transitive data in which only the Undergoer clitic pronoun is overt are listed in (14) through (18), in §2 below. Transitive imperative (including prohibitive) clauses, though paradigmatically attesting transitive verbal morphology, nonetheless position a lone overt first-person Undergoer clitic pronoun *after* the lexical verb (with the second-person Actor unexpressed): for example, /sukun-anaj={**ku**/*ka}/ (push-TR=1SG) 'Push me.' In this syntagmatic sense, imperative clauses behave intransitively. For the current purposes, the variation in only the 1SG pronoun here and in (4), discussed further below in §1, can be ignored. In addition, all of our data's transitive verbs are from one of the three transitive subtypes (not labeled as such in the glosses). Our glossing of the two transitive affixes as '-TR' and '-TR.DEP(endent)' is in the spirit of a recent analysis of Nanwang Puyuma (Ross 2009:299); cf. also *connegative* verbs in Seediq (Holmer and Billings 2014:111 fn. 1, and the sources cited there). In our transitive examples, Actors are <u>underlined</u>; Undergoers, in **bold** type. For pragmatic reasons (that are for the most part irrelevant to this study), where it is odd to tell one's addressee(s) what {she/he/they} did, data involving [+you] pronouns—if transitive, with a [+you] Actor—were often elicited as yes/no interrogatives.

The pronouns in a transitive clause do not alternate as to their positioning and thus are less obviously recognizable as clitics. We therefore utilize two other tests to show that these pronouns too are clitics. To begin, short /ħa/ 'NEG' is allowed only if at least one clitic follows right after it, as in (2b), (3b), and (4b) above. Otherwise, long /ħaz̪i/ 'NEG' is required.[5] We have not detected any semantic difference between these two Neg forms. For the current purposes, we show only that /ħa/ is phonologically deficient, requiring at least one more syllable in its prosodic word. Additional evidence comes from an accentual test, where the lexical pitch of Neg is shifted to the last of any following clitics. Thus, the only pitch in the first prosodic word in (2a–b) and (3b) is on the only clitic pronoun, and in (4b) it appears on the latter of two clitic pronouns. In each of (20b) and (21b) below there are three clitics; the high pitch is realized on the third clitic in each.[6] Thus, preverbal pronouns are also clitics. We preliminarily analyze this paradigm of pronominal forms as arguments: merged and moved in the syntax. For instance, these clitic pronouns can be cross-referenced to free SAP pronouns that are in apparent A-bar positions (not exemplified here).

This subsection has demonstrated that the pronominal forms in Rikavung Puyuma are clitics. In an intransitive clause, the lone pronominal clitic follows the initial affirmative verb, whereas in a negated clause it follows Neg and precedes the verb. In both of these environments, the pronoun is phonologically enclitic. In a transitive clause, up to two clitic pronouns precede the lexical verb.

---

[5] For instance, (ii) is the negated counterpart of (i), itself repeated from a preceding footnote:

(i)	s<əm>ukun	i	misak	(ii)	{ħaz̪i/*ħa}	s<əm>ukun	i	misak
	<INTR>push	ABS	(name)		NEG	<INTR>push	ABS	(name)
	'Misak pushed {someone/something}.'				'Misak didn't push {anyone/anything}.'			

[6] The situation with negation in Rikavung is strikingly similar to the properties of *ne* 'NEG' in the Slavic language Bulgarian, as described in Rudin et al. (1999:553–566). Namely, *ne* is accented but doesn't itself bear stress. If no clitic immediately follows *ne*, then it is unstressed, as in (i).

(i)	ne=	vaLI		(iii)	ne	=*MI*	*se*=	STRUva	[…]
	NEG=	rain;PRS.3SG			NEG	=1SG.IO	REFL=	seem;PRS.3SG	
	'It isn't raining.'				'It doesn't seem to me […]'				
(ii)	ne	=*ME*	boLI	(iv)	ne	=*SÂM*	*ti*=	*go*=	DAla
	NEG	=1SG.DO	hurt;PRS.3SG		NEG	=be;PRS.1SG	2SG.IO	M/N.3SG.DO=	given;F.SG
	'It doesn't hurt me.'				'I haven't given it to you.'				

If at least one clitic follows immediately, as in (ii) through (iv), then only the first clitic after *ne* is stressed. These data are all from Rudin et al. (1999:562), where transliterated Bulgarian spelling is used—with any clitics italicized, stressed syllables in majuscule, and the glosses modified slightly. We have inserted the equals signs showing the directions of phonological affiliation, though Rudin et al. (1999:566) actually remain noncommittal as to the proclitic status of the reflexive clitic in (iii) and of the two object-agreement clitics in (iv). One difference between the languages is that whereas Rikavung attests a disyllabic Neg option (i.e., /ħaz̪i/), Bulgarian has available only monosyllabic *ne*. Thus, if no clitic follows Neg, as in (i), then the accent of Neg is not realized as stress on any syllable. Our own alternative view of (i) is that if no clitics follow it, *ne* cannot form a prosodic word of its own and must procliticize to the verb.

Returning to the variation in form found in most of the clitic pronouns, (4) above already shows an environment that permits variation in one of the pronouns (i.e., 1SG) but not in the other pronoun (2SG). Two quite distinct kinds of variation in form are found, as signified by the parentheses above only in (1a–b, e) but not in (1d). In the former type, dubbed *facultative* variation, one of the variant forms is allowed in all environments but the other variant is allowed only in a subset thereof. In the latter type, called *obligatory* variation, certain environments require one form to the exclusion of the other. We start with the latter type, in (1d), where each variant is required at least in some environments.

1.2.   Obligatory Variation: 2SG

As (4) above already shows, in a 1SG>2SG transitive clause only the /nu/ 2SG variant is allowed.[7] The 2SG pronoun in (4) would be defined as ABS and ACC under ergative and accusative alignments, respectively. Next, we show an intransitive clause, where negation affects the 2SG variation in two distinct ways. In an affirmative intransitive clause, /u/ is required, as (5a) shows.

(5) a. s<əm>ənaj   ={u/*nu}
         <INTR>sing =2SG
         'Did you sing?'
     b. ħa   ={*u/nu}   s<əm>ənaj
         NEG   =2SG   <INTR>sing
         'Didn't you sing?'
     c. ħaz̧i   ={u/*nu}   s<əm>ənaj
         NEG   =2SG   <INTR>sing
         'Didn't you sing?'

The negated counterparts of (5a) are shown in (5b–c). As mentioned above, there are two Neg variants. Short Neg, in (5b), requires /nu/; long Neg, in (5c), allows only /u/. Under ergative alignment, in all three examples the 2SG pronoun would by definition be ABS case; under accusative alignment, all of these 2SG pronouns would be labeled as NOM. Nonetheless, in (5a, c) the form must be /u/, whereas in (5b) it can only be /nu/. The distribution of 2SG pronouns in an intransitive clause therefore correlates with neither purported alignment type.

Even at this point, it is already clear that the distinction between /u/ and /nu/ is not one of morphological cases. Corroborating evidence comes from 2SG>3 clauses. In an affirmative context only /nu/ is allowed, as shown in (6a). The negated counterparts of (6a) are shown in (6b–c). As in (5b–c) above, short and long Neg in (6b–c) require only /nu/ and /u/, respectively. Under ergative alignment, the 2SG pronouns in (6a–c) would all be predicted to be ERG; under accusative alignment, the same pronouns would all be defined as NOM.

---

[7] The symbol > abbreviates the preceding person/number features 'acting upon' the latter ones.

(6) a. {*<u>u</u>/<u>nu</u>}=   sukun-anaj
       2SG=   push-TR
       'Did you push {her/him/them}?'
   b. ħa   ={*<u>u</u>/<u>nu</u>}   sukun-an
       NEG =2SG   push-TR.DEP
       'Didn't you push {her/him/them}?'
   c. ħaẕi   ={<u>u</u>/*<u>nu</u>}   sukun-an
       NEG =2SG   push-TR.DEP
       'Didn't you push {her/him/them}?'

Table 1 summarizes how the 2SG forms correlate with the cases under each purported alignment type. For completeness, it also covers (17), (20), and (22d, g) in section 2 below. The table's middle row shows that /nu/ is consistent with neither alignment, whereas in the table's upper row /u/ is inconsistent with ergative but at least consistent with accusative alignment. However, accusative alignment overall would fail to account for how, in the NOM column of table 1, /u/ is required in (5a, c) and (6c) but /nu/ is obligatory in (5b) and (6a–b). As the bottom row of table 1 also shows, there are two environments in which both 2SG forms are allowed, also posing problems for ergative alignment. Instead of the two 2SG variants encoding distinct case forms (as has been proposed in the literature on the Nanwang dialect of Puyuma—quite plausibly, in our view—by Ross and Teng 2005:756, e.g.), the 2SG variation in Rikavung is allomorphic, based only on positioning, not on semantic roles, syntactic relations, or morphological cases. Namely, immediately following either a(n intransitive) verb or long Neg, /u/ is obligatory; clause-initially or right after either short Neg or another pronoun, /nu/ is required. (And as we show in section 2 below, there is a choice between the 2SG allomorphs right after a nonpronominal clitic.)

Table 1: Summary of Variation in 2SG Pronominal Clitics' Forms

Align-ment / Form	Ergative		Accusative	
	ERG	ABS	NOM	ACC
/u/ required	(6c)	(5a, c)	(5a, c), (6c)	
/nu/ required	(6a–b), (20a–b), (22g)	(4a–b), (5b)	(5b), (6a–b), (20a–b), (22g)	(4a–b)
Variation		(17a–b), (22d)		(17a–b), (22d)

1.3.   Facultative Variation: 2PL, EXCL1PL, and 1SG

The preceding discussion of the 2SG pronominal forms' variation has shown that each variant is required to the exclusion of the other in at least one environment. In the remaining three person/number combinations (in which pronominal variation is also observed), one of the pronominal variants is always possible, whereas the other variant form is restricted to specified environments.

    Moving first to the 2PL variation, listed in (1e) above, a pattern similar to that of the 2SG is found. Starting with intransitive clauses, both of (7a–b) show that the /mu/ pronominal variant is required. Unlike the 2SG data above, however, here short and long Neg do not select distinct 2PL variants.

(7) a. s<əm>ənaj   ={mu/*nmu}
         <INTR>sing =2PL
         'Did you sing?'
     b. ħa(ʑi)   ={mu/*nmu}   s<əm>ənaj
         NEG   =2PL   <INTR>sing
         'Didn't you sing?'

Next, with 2PL>3 clauses, /mu/ is also required—once again, regardless of polarity (or, if negated, the choice of the short or long form of Neg):

(8) a. {mu/*nmu}=   sukun-anaj
         2PL=   push-TR
         'Did you push {her/him/them}?'
     b. ħa(ʑi)   ={mu/*nmu}   sukun-an
         NEG   =2PL   push-TR.DEP
         'Didn't you push {her/him/them}?'

All remaining environments—including (18), (21), and (22e, h)—allow either 2PL variant.[8] For example, in (9a–b), encoding 1SG>2PL, there is no restriction.

(9) a. {ku/ka}= {mu/nmu}=   sukun-anaj
         1SG=   2PL=   push-TR
         'I pushed you.'
     b. ħa(ʑi)   ={ku/ka}   ={mu/nmu}   sukun-an
         NEG   =1SG   =2PL   push-TR.DEP
         'I didn't push you.'

    Table 2 sums up the data where /mu/ is required (and /nmu/ is disallowed), on the upper row, and where both are allowed, on the lower one.

---

[8] The choice of non-1SG clitic-pronominal allomorphs is free right after nonpronominal clitics.

Table 2: Summary of Variation in 2PL Pronominal Clitics' Forms

Align-ment / Form	Ergative		Accusative	
	ERG	ABS	NOM	ACC
/mu/ required	(8a–b)	(7a–b)	(7a–b), (8a–b)	
/nmu/ optional	(21a–b), (22h)	(9a–b), (18a–b), (22e)	(21a–b), (22h)	(9a–b), (18a–b), (22e)

As the bottom row of this table shows, the environments where both forms are allowed are inconsistent with either purported alignment. As the upper row demonstrates, ergative alignment is untenable for the /mu/ variant. However, the pattern is at least consistent with accusative alignment. As in the NOM column of table 1 above, under an accusative alignment in table 2 there would still be no way to predict, on the basis of morphological cases, the environments in which /nmu/ is optional. Therefore, we also analyze the 2PL variation as positional allomorphy: /nmu/ is allowed only immediately after another clitic.[9] As (18a–b) and (22e) below show, the clitic right before /nmu/ need not be pronominal.

We turn next to the EXCL1PL variation, shown in (1b) above. Here the data are considerably more limited even than those of the 2PL (primarily because the EXCL1PL pronoun does not co-occur with any other clitic pronoun, a phenomenon we discuss separately below). First, in an intransitive clause, only /mi/ is found (regardless of polarity or whether Neg is short or long):

(10) a. s<əm>ənaj ={mi/*niam}
     <INTR>sing =EXCL1PL
     'We sang.'

   b. ħa(ʐɨ) ={mi/*niam} s<əm>ənaj
     NEG =EXCL1PL <INTR>sing
     'We didn't sing.'

---

[9] Both variations in second-person clitic pronouns in Rikavung involve one /n/-initial variant (with their counterparts consisting only of every segment after this /n/). From comparing the Puyuma dialects, we've argued (separately, as Jiang and Billings 2015) that this boundary *n in Proto Puyuma appeared at boundaries between various morphophonologically bound elements. Billings (1996) discusses a similar phenomenon in Russian involving a boundary *n (between a preposition and either a possessive or a personal pronoun). This prehistoric situation continues to be relevant in Rikavung in the second-person clitic pronouns to differing degrees: 2SG /nu/ required only clause-initially or after either another clitic pronoun (invariably 1SG) or short Neg, but 2PL /nmu/ permitted only immediately after any other clitic (not clause-initially or after short Neg). As (17), (18), and (22d–e) below also show, /nu/ and /nmu/ are in free variation (with /u/ and /mu/, respectively) immediately after the nonpronominal clitics /ta(w)/ 'INV' and /a/ 'IRR'.

In an affirmative EXCL1PL>3 clause, there is an option. However, in its negated counterpart only /mi/ is allowed (once again, regardless of the length of Neg):

(11) a. {<u>mi</u>/<u>niam</u>}=   sukun-anaj
        EXCL1PL=   push-TR
        'We pushed {her/him/them}.'
    b. ħa(z̩i)   ={<u>mi</u>/*<u>niam</u>}   sukun-an
        NEG   =EXCL1PL   push-TR.DEP
        'We didn't push {her/him/them}.'

As with the 2PL situation above, *no* environment prohibits one of the EXCL1PL variants: /mi/. Rather, those listed in (10a–b) and (11b) merely prohibit the /niam/ variant. The distribution of only /mi/, shown in the upper row of table 3, is incompatible with ergative alignment but is consistent (probably due merely to the dearth of data) with accusative alignment. As the same table's lower row of data shows, also including (15) and (22b) below, the environments where both variants are possible cannot be explained by either purported alignment type.[10] As in the preceding two tables, there would be no way to account for the combined distribution of NOM forms under a purported accusative alignment.

Table 3: Summary of Variation in EXCL1PL Pronominal Clitics' Forms

Alignment / Form	Ergative		Accusative	
	ERG	ABS	NOM	ACC
/mi/ required	(11b)	(10a–b)	(10a–b), (11b)	
/niam/ optional	(11a)	(15a–b), (22b)	(11a)	(15a–b), (22b)

As alluded to above, neither of the EXCL1PL forms co-occurs with any other clitic pronoun. (At this point we offer no explanation for this fact.) In order to encode either EXCL1PL>2SG or EXCL1PL>2PL, a work-around strategy is employed: initial /iniam mu/ (where /iniam/ is a case-neutral EXCL1PL nonclitic pronoun and /mu/ 'TOP' is homophonous with but distinct from the 2PL clitic pronoun), followed by a pause, then by the rest of the sentence, as in (4) and (9) above, in which the 1SG clitic-pronoun form stands in for any EXCL1PL pronoun.

    The final instance of variation in form among the clitic pronouns is with the 1SG, in (1a) above. Like the preceding 2PL and EXCL1PL situations, the /ku/ variant is possible in all environments, whereas the /ka/ form is allowed only in certain pronominal combinations. We also show that the distribution of /ka/ is inconsistent with both ergative and accusative alignments. To begin, (4) and (9) above both exemplify the 1SG pronoun in 1SG>2SG and 1SG>2PL transitive

---

[10] As a preceding footnote mentions, right after a nonpronominal clitic there is a choice in forms.

clauses, respectively, in both of which 1SG variation does occur. By contrast, there are environments that prohibit /ka/. For example, intransitive clauses do not allow /ka/ (regardless of polarity or the form of Neg used):

(12) a. s<əm>ənaj ={ku/*ka}
          <INTR>sing =1SG
          'I sang.'
     b. ħa(z̲i) ={ku/*ka}   s<əm>ənaj
          NEG   =1SG   <INTR>sing
          'I didn't sing.'

Another environment prohibiting the /ka/ variant is 1SG>3 transitive clauses:

(13) a. {<u>ku</u>/*<u>ka</u>}= sukun-anaj
          1SG=   push-TR
          'I pushed {her/him/them}.'
     b. ħa(z̲i) ={<u>ku</u>/*<u>ka</u>}   sukun-an
          NEG   =1SG   push-TR.DEP
          'I didn't push {her/him/them}.'

Table 4: Summary of Variation in 1SG Pronominal Clitics' Forms

Align-ment / Form	Ergative		Accusative	
	ERG	ABS	NOM	ACC
/ku/ required	(13a–b)	(12a–b), (14a–b), (22a)	(12a–b), (13a–b)	(14a–b), (22a)
/ka/ optional	(4a–b), (9a–b)	(20a–b), (21a–b), (22g–h)	(4a–b), (9a–b)	(20a–b), (21a–b), (22g–h)

As this final table shows (so long as additional examples from section 2 are also considered), the environments where the /ku/ variant is required are compatible with neither ergative nor accusative alignment. The same goes for the environments where /ka/ is permitted. The only tenable observation is that /ka/ is a 1SG variant found only immediately before another clitic pronoun. (And this following clitic pronoun invariably encodes only either 2SG or 2PL.)

This concludes our discussion of pronouns' variation as to their *form*. Only one of these does not vary in this way: INCL1PL /ta/, in (1c). Still, this form varies in its *function*. It encodes INCL1PL, as in footnote 3, (2), and (3) above, as well as (16) below.[11] It is also possible to use /ta/, instead of the 1SG forms

---

[11] In the spirit of the formal person/number features proposed by McKaughan (1959), 1SG and EXCL1PL are each [+me, –you], INCL1PL is [+me, +you], 2SG and 2PL are each [–me, +you],

mentioned so far, /ku/ or /ka/, to encode 1SG (possibly even EXCL1PL) under restricted conditions. Before discussing that variation in function, however, it is necessary to present the facts about this paper's subtitle: *An inverse analysis*.

The variations in form above in (1a–b, d–e) have now all been accounted for positionally. The 2SG pronoun is realized as /u/ immediately following the (intransitive) verb or long Neg but as /nu/ clause-initially or after either short Neg or another clitic pronoun. The 2PL pronoun allows /mu/ in all environments but /nmu/ only after another clitic. Similarly, the EXCL1PL pronoun must be realized as /mi/ after the (intransitive) verb or Neg but is allowed to be /niam/ only elsewhere. Finally, the /ka/ 1SG variant is found only right before another clitic pronoun, whereas its /ku/ counterpart is allowed in all environments.

2.   The Remaining Combinations: Inverse

If none of the forms in (1a–e) above encodes case, then in Rikavung there is but a single paradigm of clitic pronouns. How, then, do these pronouns convey their syntactic relations, especially if they co-occur (i.e., in a transitive clause)? Various examples above have shown that in an SAP>3 clause only the SAP pronoun is overt: (3), (6), (8), (11), and (13). If both pronouns are SAPs, as so far (4) and (9) have shown, the 1SG clitic pronoun precedes the second-person form.[12] Still not discussed are the eight clitic-pronominal combinations in which the Undergoer is at least as person-prominent as the Actor—where the *lower* the ordinal numeral of the grammatical person, the *greater* its person-prominence.

In 3>SAP environments, an overt SAP pronoun immediately follows the inverse (INV) clitic, as each of (14) through (18) show. (The INV clitic can be pronounced without its final labiovelar glide but see further discussion below.)

(14) a. ta(w)=   {**ku**/*****ka**}=   sukun-anaj
         INV=   1SG=   push-TR
         '{She/He/They} pushed me.'
     b. ħa(z̩i)   =ta(w)   ={**ku**/*****ka**}   sukun-an
         NEG   =INV   =1SG   push-TR.DEP
         '{She/He/They} didn't push me.'

Because the 1SG pronoun does not precede any clitic pronoun in (14), /ka/ is not acceptable. The other 3>SAP examples, where the SAP clitic pronouns do show

and 3 is [–me, –you]. Two pronouns sharing either [+me] or [+you] cannot co-occur (or at least not as a clitic cluster). This restriction entails that /ta/ functioning as an INCL1PL pronoun cannot co-occur with any other SAP clitic pronoun. As such, the only transitive combinations involving a functionally INCL1PL pronoun are INCL1PL>3, above in (3), and 3>INCL1PL, below in (16).

[12] As has been mentioned above, neither the EXCL1PL nor the INCL1PL pronoun co-occurs with any overt clitic pronoun. Namely, INCL1PL>2 is semantically anomalous (footnote 11) and (as has been mentioned in §1.3) EXCL1PL>2 is prohibited. As such, 1SG>2SG (4) and 1SG>2PL (9) are the only overt pronominal combinations encoding first person acting upon second person.

a variation in form—namely, only (15), (17), and (18) but not (16)—allow either variant of the SAP clitic pronoun immediately after the INV clitic.

(15) a. ta(w)= **{mi/niam}**= sukun-anaj
    INV=   EXCL1PL=   push-TR
    '{She/He/They} pushed us.'
   b. ħa(ʐi) =ta(w) ={**mi/niam**} sukun-an
    NEG   =INV   =EXCL1PL   push-TR.DEP
    '{She/He/They} didn't push us.'

(16) a. ta(w)= **ta**=   sukun-anaj
    INV=   INCL1PL= push-TR
    '{She/He/They} pushed us.'
   b. ħa(ʐi) =ta(w) =**ta**   sukun-an
    NEG   =INV   =INCL1PL   push-TR.DEP
    '{She/He/They} didn't push us.'

(17) a. ta(w)= **{u/nu}**= sukun-anaj
    INV=   2SG=   push-TR
    '{She/He/They} pushed you.'
   b. ħa(ʐi) =ta(w) ={**u/nu**} sukun-an
    NEG   =INV   =2SG   push-TR.DEP
    '{She/He/They} didn't push you.'

(18) a. ta(w)= **{mu/nmu}**= sukun-anaj
    INV=   2PL=   push-TR
    '{She/He/They} pushed you.'
   b. ħa(ʐi) =ta(w) ={**mu/nmu**} sukun-an
    NEG   =INV   =2PL   push-TR.DEP
    '{She/He/They} didn't push you.'

Moreover, in (19a–b), encoding 3>3, there is no overt clitic pronoun at all.[13]

(19) a. taw= sukun-anaj
    INV= push-TR
    '{She/He/They}$_i$ pushed {her/him/them}$_j$.'
   b. ħa(ʐi) =taw   sukun-an
    NEG   =INV   push-TR.DEP
    '{She/He/They}$_i$ didn't push {her/him/them}$_j$.'

---

[13] Somewhat surprisingly to us, optionally pronouncing (16a–b) without the labiovelar glide of the INV clitic results in an *acceptable* sequence of homophonous [ta] syllables (of the INV and INCL1PL clitics). However, in (19a–b) the same glide is obligatory. Without it, the only interpretation deemed acceptable by our consultant speakers is INCL1PL>3, as in (3a–b) above.

From the data so far, in (14) through (19), /ta(w)/ would appear to encode a third-person Actor. Indeed, this is Tsuchida's analysis of /taw/ in Tamalakaw (1980:196), and Teng's analysis of it in Tamalakaw, Ulivelivek, and Katripul (2009:824, 826, 834, 2015:410, 419), three closely related Puyuma varieties.

The crucial data in (20a–b) and (21a–b) don't involve any third-person entity. The invariant semantics contributed by /ta(w)/ in all of (14) through (21) is that the Undergoer is at least as person-prominent as the Actor and that the first clitic pronoun immediately after /ta(w)/ encodes the Undergoer.[14]

(20) a. ta(w)= {**ku/ka**}= {*u̱/ṉu}= sukun-anaj
      INV= 1SG= 2SG= push-TR
      'Did you push me?'

   b. ħa(z̧i) =ta(w) ={**ku/ka**} ={*u̱/ṉu} sukun-an
      NEG =INV =1SG =2SG push-TR.DEP
      'Didn't you push me?'

(21) a. ta(w)= {**ku/ka**}= {m̱u/ṉmu}= sukun-anaj
      INV= 1SG= 2PL= push-TR
      'Did you push me?'

   b. ħa(z̧i) =ta(w) ={**ku/ka**} ={m̱u/ṉmu} sukun-an
      NEG =INV =1SG =2PL push-TR.DEP
      'Didn't you push me?'

In fact, Tsuchida reports a sentential 2SG>1SG example (1980:199) from Tamalakaw identical in the relevant respects to our Rikavung example in (20a), as well as diagrammatic data (1980:200) of only the clitic sequences similar to our (20) and (21). In order to maintain his analysis of /taw/ in Tamalakaw as a marker of only a third-person Actor, Tsuchida resorts to calling these sequences the result of morphological "fusion" (1980:199). Such fused forms are also known in the literature as *portmanteau* forms. Still in the spirit of Tsuchida (1980), Teng (2015:422–423) proposes a reduction to the list of portmanteaux to just the SAP>SAP pairs, /kunu/ ~ /kanu/ '1SG>2SG', /kanmu/ '1SG>2PL', /takunu/ ~ /takanu/ '2SG>1SG', and /takanmu/ '2PL>1SG'—i.e., Tamalakaw's counterparts to the clitic sequences in our Rikavung (4), (9), (20), and (21), respectively.[15] We argue that there are no portmanteau pronouns in Rikavung.

---

[14] Here we address issues about Rikavung in Teng (2008), the existence of which we learned about since our AFLA-21 talk. These portmanteaux are reported: /takunu/ '2SG>1SG' and /taku(n)mu/ '2PL>1SG', consistent with our (20) and (21), respectively. The addition of /ta-/ encodes "a reverse of roles between the two speech act participants." Hence, for example, /kunu/ is used for 1SG>2SG, as in our (4), whereas /takunu/ is used, as in (20), "when the actor changes from first person to second person": the essence of what is known in the linguistic literature as a morphosyntactic inverse. Teng (2008) uses /taw/ for 3.ERG but /ta-/ as part of the portmanteaux.
[15] Whereas Tsuchida (1980:199–200) lists only /takanu/ as the 2SG>1SG portmanteau, Teng (2015:422) also lists a variant, /takunu/. In addition, whereas Tsuchida (1980:200) lists

Paradoxically, our primary evidence against Tsuchida's and Teng's portmanteau approaches is inspired by data elsewhere in their studies. Namely, Tsuchida (1980:200) reports that in Tamalakaw between /taw/ and an SAP pronoun—namely: 1SG /ku/, EXCL1PL /mi/, INCL1PL /ta/, 2SG /u/, or 2PL /mu/; cf. our Rikavung examples (14) through (18) above, respectively—it is possible to insert /a/, which is glossed there as a FUT marker. In addition, Tsuchida (1980:199) lists a sentential example in Tamalakaw with the sequence /taw a ku/; cf. our Rikavung example, identical in the relevant respects, in (22a) below. Similarly, Teng (2015:424) very effectively harnesses the insertion of PFV /=la/ to demonstrate that sequences of clitics in Katripul (another Puyuma dialect) are *not* portmanteaux. Using those two linguists' ideas, we then verified that in Rikavung the clitic /a/, which we gloss as IRR, can immediately follow /ta(w)/ 'INV', as (22f) below shows. If there is at least one clitic pronoun after the INV clitic, then IRR /a/ can appear between the INV clitic and the pronoun, as (22a–e, g–h) also show. The fact that the IRR clitic can go between /ta(w)/ and the SAP clitic pronoun specifically as in (22a–e) demonstrates that the clitic sequences in (14) through (18) above must not be analyzed as portmanteaux. Rather, each of these sequences consists of two independent morphemes: INV and a pronoun. Of even more relevance to the current discussion, the fact that the IRR clitic can follow /taw/ but precede *both* clitic pronouns in (22g–h) demonstrates that the acceptable clitic sequences in (20) and (21) must also not be portmanteaux.[16]

(22) a. taw= a=   {**ku/*ka**}= sukun-anaj   [cf. (14a) above]
       INV= IRR= 1SG=   push-TR
       '{She/He/They} would like to push me.'

   b. taw= a=   {**mi/niam**}= sukun-anaj   [cf. (15a) above]
       INV= IRR= EXCL1PL=   push-TR
       '{She/He/They} would like to push us.'

   c. taw= a=   **ta**=   sukun-anaj   [cf. (16a) above]
       INV= IRR= INCL1PL= push-TR
       '{She/He/They} would like to push us.'

   d. taw= a=   {**u/nu**}= sukun-anaj   [cf. (17a) above]
       INV= IRR= 2SG=   push-TR
       '{She/He/They} would like to push you.'

   e. taw= a=   {**mu/nmu**}= sukun-anaj   [cf. (18a) above]
       INV= IRR= 2PL=   push-TR
       '{She/He/They} would like to push you.'

   f. taw= a=   sukun-anaj   [cf. (19a) above]
       INV= IRR= push-TR
       '{She/He/They}$_i$ would like to push {her/him/them}$_j$.'

/takanmu/ for the 2PL>1SG portmanteau, Teng (2015:422) shows the optional epenthetic vowel between the nasals (and reports that this form also allows a PL-Undergoer interpretation).
[16] For phonological reasons, the INV clitic's final glide in each of (22a–h) is all but obligatory.

g. taw=   a=   {**ku/ka**}=   {*<u>u</u>/<u>nu</u>}=   sukun-anaj   [cf. (20a) above]
   INV=   IRR=   1SG=   2SG=   push-TR
   'Would you push me.'

h. taw=   a=   {**ku/ka**}=   {<u>mu</u>/<u>nmu</u>}=   sukun-anaj   [cf. (21a) above]
   INV=   IRR=   1SG=   2PL=   push-TR
   'Would you push me.'

Another reason to favor our INV analysis of /ta(w)/ in Rikavung comes from the same two aforementioned studies. Tsuchida (1980:199) observes that in Tamalakaw whereas /a/ (which he glosses as FUT) can appear between /taw/ (considered to be a third-person pronoun there) and 1SG /ku/, as in our Rikavung example (22a) above, /a/ cannot interrupt a sequence of 1SG /ka/ and 2SG /nu/. The order in Tamalakaw, Tsuchida reports, must be /ka nu a/ followed by the verb. The same example also refutes Teng's claim that in Tamalakaw /kanu/ is inseparable from the verb (2015:422). We have verified that in Rikavung (4a) and (9a) above, IRR /a/ also can be inserted between the 2SG or 2PL clitic pronoun and the verb. Teng (2015:422–424) also draws a distinction (in all of the Nanwang, Katripul, and Tamalakaw dialects) between various combinations consisting of only SAP pronouns (which she considers to be portmanteaux) and 3>SAP clitic clusters, where only in the latter can the two clitics be separated. Now, under both authors' accounts of Tamalakaw, where what we consider to be the INV clitic is analyzed instead as a third-person pronoun, the inseparability of two pronouns only if they both encode SAPs would be a coincidence. Under our INV analysis of Rikavung, /ta(w)/ is not a pronoun. Therefore, the relevant restriction is that IRR /a/ cannot interrupt any cluster of *pronominal* clitics.

Even more support for our analysis comes from the fact that in all of (15) through (18), (20), and (21), under these two authors' assumptions, the order would be third person *before* SAP. In every known clitic-ordering system based primarily on person in the Philippine-Formosan area, any SAP invariably *precedes* a third-person pronoun (Holmer and Billings 2014:122–124).

To be sure, there is quite a bit of additional allomorphy in Rikavung. For instance, there exists, in addition to /ta(w)/, another INV variant: /tu/. (It is unclear to us whether this form is a borrowing from other Puyuma dialects such as Nanwang.) Additionally, a 1SG Undergoer within a pronominal cluster can also be encoded using the INCL1PL clitic pronoun, /ta/, after the INV clitic and before a second-person pronoun. Namely, /ta/ can replace /ku/ or /ka/ in (20) and (21) without affecting the free translations there. This variation in the function of /ta/ resembles a mechanism also widely observed in Philippine languages. However, we find it peculiar that this use of /ta/ is apparently limited in Rikavung to 2>1 INV clauses. We have also verified that this special use of pronominal /ta/ is not part of a portmanteau (again, using the IRR-insertion test, not exemplified here).

To summarize section 2, we have shown that /ta(w)/ is not a pronoun. Rather, this clitic is added to indicate that the clitic pronoun right after it en-

codes the Undergoer. If two such pronouns follow /ta(w)/, then the latter one encodes the Actor. We have also shown that no pronominal combinations are portmanteaux. Finally, Rikavung attests no overt third-person clitic pronouns.

3.   Conclusion

This paper has demonstrated that in Rikavung Puyuma there is a single paradigm of clitic pronouns. Despite plenty of variation, no case distinction can be made. Furthermore, if two clitic pronouns co-occur, their relative order is determined only by grammatical person. Finally, an INV clitic is employed to indicate that the Undergoer is at least as person-prominent as the Actor.

References

Billings, Loren A. 1996. Sandhi phenomena and language change. *Interfaces in phonology*, ed. Ursula Kleinhenz, 60–82 (Studia grammatica 41). Berlin: Akademie Verlag.

Billings, Loren, and Daniel Kaufman. 2004. Towards a typology of Austronesian pronominal clisis. *Proceedings of AFLA 11* [the 11th annual meeting of the Austronesian Formal Linguistics Association], ed. Paul Law, 15–29 (ZAS Papers in Linguistics 34). Berlin: Zentrum für Allgemeine Sprachwissenschaft.

Bloomfield, Leonard. 1917. Tagalog texts and grammatical analysis. *University of Illinois Studies in Language and Literature* 3(2–4):1–408 [157–564].

Chang, Yiying Ann. 2012. Pronominal ordering in Plngawan Atayal. *Proceedings of ConSOLE XIX*, eds. Enrico Boone, Kathrin Linke, and Maartje Schulpen, 177–190. Leiden: Student Organization of Linguistics in Europe.

Comrie, Bernard. 2013. Ergativity: Some recurrent themes. *The acquisition of ergativity*, eds. Edith L. Bavin and Sabine Stoll, 15–34 (Trends in Language Acquisition Research 9). Amsterdam: Benjamins.

Holmer, Arthur, and Loren Billings. 2014. Clitic pronouns in Seediq. *Papers from 12-ICAL* [the 12th International Conference on Austronesian Linguistics], *vol. 2: Argument realisations and related constructions in Austronesian languages*, eds. I Wayan Arka and N. L. K. Mas Indrawati, 111–139 (Asia-Pacific Linguistics 013; Studies on Austronesian Languages 002). Canberra: College of Asia and the Pacific, Australian National University.

Huang, Lillian M., Elizabeth Zeitoun, Marie M. Yeh, Anna H. Chang, and Joy J. Wu. 1999. A typological overview of pronominal systems of some Formosan languages. *Selected papers from the fifth International Conference on Chinese Linguistics*, eds. Hsu Wang, Feng-fu Tsao, and Chin-fa Lien, 165–198. Taipei: Crane.

Hung, Silvia Yu-ju, and Loren Billings. 2009/2010. Topicality and pronominal ordering in two Manobo languages. Presented at the panel on Pronoun-ordering typology, 11th International Conference on Austronesian Linguistics (Aussois, France; June 2009). Uploaded to the Rutgers Optimality Archive <roa.rutgers.edu> (June 2010).

Jiang, Haowen. 2013. The possession-modification scale and a reappraisal of "genitive" in Paiwan, Rukai, and Puyuma. Presented at the second National Taiwan University Workshop on Discourse and Grammar in Formosan Languages, Taipei <www.academia.edu/3693931>.

Jiang, Haowen, and Loren Billings. 2015. Evidence from Puyuma for only a single clitic-pronoun paradigm in Proto Austronesian. Panel on Topics in the morphosyntax of higher-order Austronesian subgrouping: Considering the Nuclear Austronesian hypothesis, 13th International Conference on Austronesian Linguistics (Academia Sinica, Taiwan).

Kaufman, Daniel. 2010. The grammar of clitics in Maranao. *Piakandatu ami Dr. Howard P. McKaughan*, eds. Loren Billings and Nelleke Goudswaard, 179–204. Manila: Linguistic Society of the Philippines and SIL Philippines.

Kaufmann, John. 1916 (©1939). *Principles of Visayan grammar*. Oroquieta, Manila: Catholic Trade School.

Lee, Celeste, and Loren Billings. 2008. Clitic-pronoun clusters in Central Philippine. *SEALS XIV volume 1: Papers from the 14th annual meeting of the Southeast Asian Linguistics Society 2004*, eds. Wilaiwan Khanittanan and Paul Sidwell, 193–203 (Pacific Linguistics E-5). Canberra: Research School of Pacific and Asian Studies, Australian National University.

Li, Paul Jen-kuei. 1995. The case-marking system in Mayrinax, Atayal. *Bulletin of the Institute of History and Philology* [Academia Sinica] 66(1):23–52 (re-issued in Li 2004:387–414).

Li, Paul Jen-kuei. 2004. *Selected papers on Formosan languages*, vol. 1 (Language and Linguistics Monograph Series C3). Taipei: Institute of Linguistics, Academia Sinica.

Liao, Hsiu-chuan. 2004. Transitivity and ergativity in Formosan and Philippine languages. Doctoral dissertation, University of Hawai'i at Mānoa.

Liao, Hsiu-chuan. 2005. Another look at the order of clitic pronouns in Wulai Atayal. *Concentric: Studies in Linguistics* 31(1) [Special issue: Austronesian linguistics]:47–63.

McKaughan, Howard P. 1959. Semantic components of pronoun systems: Maranao. *Word* 15(1):101–102.

Peng, Adam, and Loren Billings. 2008. Binukid pronominal clisis. *Studies in Philippine Languages and Cultures* 17 [Special issue: 10-ICAL pronoun papers, ed. Sue McQuay]:179–212.

Quakenbush, J. Stephen, and Edward Ruch. 2008. Pronoun ordering and marking in Kalamianic. *Studies in Philippine Languages and Cultures* 17 [Special issue: 10-ICAL pronoun papers, ed. Sue McQuay]:213–234.

Ross, Malcolm. 2009. Proto Austronesian verbal morphology: A reappraisal. *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, eds. Alexander Adelaar and Andrew Pawley, 295–326 (Pacific Linguistics 601). Canberra: Research School of Pacific and Asian Studies, Australian National University.

Ross, Malcolm, and Stacy Fang-ching Teng. 2005. Formosan languages and linguistic typology. *Language and Linguistics* [Academia Sinica] 6(4):739–781.

Rudin, Catherine, Christina Kramer, Loren Billings and Matthew Baerman. 1999. Macedonian and Bulgarian *li* questions: Beyond syntax. *Natural Language and Linguistic Theory* 17(3):541–586.

Suenari, Michio. 1969. A preliminary report on Puyuma language (Rikavong dialect). *Bulletin of the Institute of Ethnology* [Academia Sinica] 27:141–164.

Teng, Stacy Fang-ching. 2008. The internal relationships among Puyuma dialects: Evidence from morphosyntax. Unpublished slides from talk presented at the second Workshop on Internal Relationships of Formosan languages, Academia Sinica, Taipei (cited with permission).

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

Teng, Stacy Fang-ching [Stacy F. Teng]. 2015. A comparative study of the personal pronominal systems in three Puyuma dialects. *New advances in Formosan linguistics*, eds. Elizabeth Zeitoun, Stacy F. Teng, and Joy J. Wu, 407–429 (Asia-Pacific Linguistics 017; Studies on Austronesian Languages 003). Canberra: College of Asia and the Pacific, Australian National University.

Tsuchida, Shigeru. 1980. Puyuma-go (Tamarakao hōgen) goi: Gohō gaisetsu oyobi tekisuto tsuki [Puyuma (Tamalakaw dialect) vocabulary: With grammatical notes and texts]. *Kuroshio no minzoku, bunka, gengo* [Ethnology, cultures, and languages along the Black Current], eds. Kuroshio Bunka no Kai [Black Current Cultures Committee], 183–307. Tokyo: Kadokawa Shoten.

Tsuchida, Shigeru. 1992. Puyuma-go [The Puyuma language]. *Gengogaku daijiten* [The Sanseido encyclopedia of linguistics], eds. Takashi Kamei, Rokurō Kōno, and Eiichi Chino, vol. 3, 742–753. Tokyo: Sanseido.

Tsuchida, Shigeru. 1995. Alienable and inalienable distinction in Puyuma. *Austronesian studies relating to Taiwan*, eds. Paul Jen-kuei Li, Cheng-we Tsang, Dah-an Ho, Ying-kuei Huang, and Chiu-yu Tseng, 793–804 (Symposium Series of the Institute of History and Philology 3). Taipei: Institute of History and Philology, Academia Sinica.

Yen, Doris Ching-jung, and Loren Billings. 2011. Sequences of pronominal clitics in Mantauran Rukai: V-deletion and suppletion. *Proceedings of the eighteenth meeting of the Austronesian Formal Linguistics Association (AFLA), Harvard University, Cambridge (MA), March 4–6, 2011*, eds. Lauren Eby Clemens, Gregory Scontras, and Maria Polinsky, 168–182. London, Ontario, Canada: Linguistics Program, University of Western Ontario <westernlinguistics.ca/afla/proceedings/afla18/>.

Yen, Doris Ching-jung, and Loren Billings. 2014. The cluster-internal ordering of clitics in Kavalan (East Formosan, Austronesian). *Proceedings of the thirty-eighth annual meeting of the Berkeley Linguistics Society: General session and thematic session on language contact*, eds. Kayla Carpenter, Oana David, Florian Lionnet, Christine Sheil, Tammy Stark, and Vivian Wauters, 523–537. Berkeley: Berkeley Linguistics Society.

Zeitoun, Elizabeth. 2007. *A grammar of Mantauran (Rukai)* (Language and Linguistics Monograph Series A4-2). Taipei: Institute of Linguistics, Academia Sinica.

CORRELATIONAL COMPARATIVES (CCS) IN MALAGASY[*]

Edward L. Keenan
University of California at Los Angeles
edward.keenan1@gmail.com

Baholisoa Ralalaoherivony
Université d'Antananarivo
rabaholy@gmail.com

We study CCs (*The more you work the more you earn*) in Malagasy, which present a novel use of *no*-Focus structures. We provide some new data patterns and advocate a Binary Operator analysis, briefly comparing Malagasy CCs with ones from other languages. They differ from but share properties with purely paratactic constructions and correlatives. We consider their relation to pseudoclefts as in Paul 2001, 2008, Potsdam 2006 and Kalin 2009.

1. Introduction

CCs[1] assert a proportional dependency between two scalars. A scalar is a function whose values are linearly ordered; for F,G scalars G *is proportional to* F iff $F(x) < F(y)$ implies $G(x) < G(y)$, all x,y. F is the *independent* scalar, G the *dependent* one. (1) is a Malagasy CC, scalars bracketed.

(1) Arakaraka   [ind ny ∅-iasana]   no   [dep ∅-ahazoana   vola   ]
     proportionate.to   DET PRES-work.CT   FOC   PRES-receive.CT money
     [ind The more one works][dep the more one earns (receives money)]

So CCs are true/false claims, cognitively similar to conditionals (McCawley 1988, Beck 1997, Hsiao 2003, Brasoveanu 2008, Smith 2010) since proportionality is defined in conditional terms. But CCs are not syntactically conditionals, which are of the form *Raha* ϕ *dia* ψ (*If* ϕ *then* ψ). Proportional dependency may express cause-effect (Working more causes earning more) but may just indicate correlation (Beck, Brasoveanu, Smith): *The taller you are the taller your siblings*.

2. Distribution

CCs in Malagasy occur as verbal complements like other assertions. They also negate, take propositional modals, form tags, host S-level questions, can be

---

[1] Also called *Comparative Correlatives* (den Dikken 2005) and *Comparative Conditonals* (McCawley 1988, Beck 1997). CCs are contentious: Culicover and Jackendoff 1999 argue that CCs in English are paratactic but unrepresentable in (then) current syntactic theories. Den Dikken 2005 counterargues that they are correlatives with a minimalist representation. Smith 2010 provides a categorially oriented subordinate –main clause analysis.

topicalized into, and admit *Why?* questions, (2). Finally CCs cannot be extracted from, (3), as only subjects extract (Keenan 1972) and CCs are not of Predicate+Subject form.

(2) a. Heveriko/Mihevitra   aho   fa   arakaraka ny iasana no ahazoana vola
    think.TT.1S.GEN/AT.think 1S.NOM that [(1)   ]
    'I think that the more one works the more one earns.'
  b. Tsy fantatro   na   [arakaraka ny iasana no ahazoana vola]
    not know.TT.1S.GEN whether [(1)   ]
    na   tsia
    or   no
    'I don't know whether the more you work the more you earn or not.'
  c. Tsy   arakaraka ny iasana no ahazoana vola
    not   [(1)   ]
    'It is not so that the more you work the more you earn.'
  d. Toa   arakaraka ny iasana no ahazoana vola
    seem [(1)   ]
    'It seems that the more you work the more you earn.'
  e. Arakaraka ny iasana no ahazoana vola,   sa tsy izany?
    [(1)   ],   or not that
    'The more one works the more one earns, not so?'
  f. Moa   ve   arakaraka ny iasana no ahazoana vola?
    PRT   Q   [(1)   ]?
    'Is it the case that the more you work the more you earn?'
  g. Raha Rabe$_i$ aloha   dia arakaraka ny   ihinana.ny$_i$ no ahazoa.ny$_i$
     if   Rabe adverb TOP prop.to   DET eat.CT.3GEN NO receive.CT.3GEN
    aina.
    life
    'As for Rabe, the more he eats the more he improves.'
  h. Nahoana   no   arakaraka ny iasana no ahazoana vola?
    why   NO   [(1)   ]?
    'Why is it that the more one works the more one earns?'

(3) *ny vola   (izay) arakaraka ny   iasanao no ahazoana
    the money (that) prop.to   DET work.CT NO receive.CT
    'the money that you receive the more you work'

3. Internal Structure

First, we present three empirical observations. One, *arakaraka* 'proportionate to' in (1) is a preposition, the lexical reduplication of *araka* 'according to/pertaining to', which has a wider usage than *arakaraka* but occasionally substitutes for it in CCs. PPs may be predicate and nominal modifiers, (4a-b),

and given the absence of a copula, may head P1s (tensed one place predicate phrases), (4c). (5) and (6) show the predicate modifier and P1 head use of *arakaraka*.

(4) a. Miteny ami.ko   izy.
      speak   with.my 3
      'He speaks with me.'
    b. trano anaty ala
      house in   forest
      'house in (the) forest'
    c. [tandrify ny paositra]   [ny trano.ko].
      opposite   the post.office   the house.my
      'My house is opposite the post office.'

(5) [P1 Tokony hatao   arakaraka   ny   fidiran'ny   vola] [ny
      should FUT.TT.do prop.to   the   entering'the money the
      fandalinana   azy]
      using.up   3.ACC]
      'Spending should be done proportionate to revenue.'   [Raki]

(6) a. [P1 Arakaraka   ny   halavan'ny olona] [ny havesara.ny].
          prop.to   the height'the   people   the weight.their
          'People's weight is proportional to their height.'
    b. [P1 Arakaraka ny isan'ny   mponina   ao   an-tanana   iray][ny
          prop.to   the number'the inhabitant   LOC in-village   one   the
          isan'ny   trano   ao].
          number'the   house   LOC
          'The number of houses in a village is proportionate to the number of
          inhabitants there.'

Given its meaning *arakaraka* selects a scalar nominal: ones denoting height, weight, length, number, amount, salary, etc. Both its argument and that of the P1 it heads must be scalar for a proportionality comparison to make sense. Further *arakaraka* can impose a scalar interpretation on DPs like 'the washed clothes' in (7) which are scalable but not inherently scalar: [entity ⇒ scalar] is a common semantic type shift (Schwarzchild 2006). So (7b) paraphrases (7a):

(7) a. [Arakaraka   ny   lamba sasana] [ny rano   lany]
          prop.to   the clothes washed the water used
          'The water used is proportionate to the clothes washed.'
    b. [Arakaraka ny   habetsaky ny lamba sasana] [ny   habetsaky ny
          prop.to   the quantity   the clothes washed the   quantity   the

rano   lany]
water used
'The quantity of water used is proportionate to the amount of clothes washed.'

Two, the object of *arakaraka* in (1) is *ny iasana*, in which the DET *ny* combines with a P1 in the circumstantial (CT) voice, the morphology of verbs whose DP sisters ("subjects") denote a circumstance of the action or state expressed by the verb, (8c): time, location, instrument, cause, cost, manner,.... [RR] 1971:112–17 lists 13 such plus three other CT selecting constructions.

(8) a. [P1 n.an.enjika an'i   Hery tamin'io   fiara io] Rabe
         PST.AT.chase ACC'ART Hery PST.PREP'that car   that Rabe
         'Rabe chased Hery with that car.'
    b. [P1 n.enjeh.in-dRabe   tamin'io   fiara io] i   Hery
         PST.chase.TT-Rabe.GEN PST.PREP'that car   that ART Hery
         'Hery was chased in that car by Rabe.'
    c. [P1 n.an.enjeh.an-dRabe   an'i   Hery] io fiara   io
         PST.[AT.chase].CT-Rabe.GEN ACC'art Hery that car   that
         'That car was used by Rabe to chase Hery.'

AT verbs are built by prefixing *m.i-*, *m.aN-*, or *m.a-* to roots or by prefixing already derived AT verbs with causative or reciprocal morphemes. TT verbs are either zero affixed roots or built by prefixing *a-*, *voa-* or *tafa-* (rarely, infixing *-in-*) to roots or suffixing *-ina* to roots or causatives. CT verbs are built by suffixing *-ana* to AT verbs.

Worth emphasizing is that P1s (any voice) are not clauses (TPs/IPs). They need an argument to form clauses of the True/False sort. So (up to isomorphism) they denote functions from argument denotations to truth values (or, intensionally, propositions).

(9a) illustrates the widely used *ny*+P1 argument, P1 in any voice (Ntelitheos 2012, Potsdam and Polinsky 2014). (9b) and (10a-b) show that roots which lack AT and hence CT morphology use AT causative forms (Rajaona 1972:206–210; Dez 1980:118) as circumstantials.

(9) a. Faly be aho t.amin'ny   n.andeh.anan.tsika t.any   Antsirabe
        happy big 1.NOM PST.PREP'the PST.go.CT.1P.INCL PST.there Antsirabe
        'I was very happy when we went to Antsirabe.'
    b. Faly be aho t.amin'ny   n.aha-t.eto   ahy   t.aloha
        happy big 1S.NOM PST.PREP'the PST.CAUSE-PST.here 1S.ACC PST.before
        'I was very happy when I was here before.'

(10) a. Arakaraka ny n.iterah.a.ny   maro no   n.aha.tanora
       prop.to   DET PST.give-bear.CT.her   many NO   PST.CAUSE.young
       an-dRasoa
       ACC-Rasoa
       'The more Rasoa gives birth the younger she is (appears).'
    b. Arakaraka ny maha-adala   no   ihomehezana
       prop.to   DET CAUSE-crazy   NO   laugh.CT
       'The crazier you are the more you laugh.'

Three, the verb in the post-*no* scalar in a CC must, like the pre-*no* one, be in CT voice:

(11) a. *Arakaraka ny miasa   (ianao) no mahazo   vola   (ianao)
          prop.to   DET AT.work (2S.NOM) NO AT.receive money (2S.NOM)
     b. *Arakaraka ny atao   (ny) asa   no azo   (ny)   vola
          prop.to   DET TT.do (DET)work NO receive.Ø (DET) money
     c. *Arakaraka ny sasa.na   (ny)   lamba no taom.ina (ny)   rano
          prop.to   DET wash.TT   (DET) clothes NO bring.TT (DET) water

So far then CCs have the form [[_PP_ *arakaraka* [*ny* P1_CT]] *no* [P1_CT]]. Both P1s are tensed, usually the same but not always, (12a). Both take PP complements, accusative objects, (12b), and agent phrases (possessors of the verb), usually the same, though (12c) is a found example with different agents. Absence of agent phrases yields an "arb" 'one' interpretation, as in (1), not an existentially quantified one. Finally the scalars may be internally coordinate, (12d).

(12) a. Arakaraka ny n.anomboh.a.ny   ihany no h.iafarany
        prop.to   DET PST.begin.CT.3GEN   thus   NO FUT.end.CT.3GEN
        'As its beginning (of the story) built up so its ending fades out.'

                                                                    [Etsy 231]

    b. Arakaraka ny if.anerasera.ko amin'i   Soa no tsy ahafantarako
       prop.to   DET REC.socialize.1S.GEN prep'ART Soa NO not know.1S.GEN
       azy
       3ACC
       'The more I socialize with Soa the less I understand her.'
    c. Arakaraka ny h.androsoan'ilay rangahy   no h.ihemoan'ny
       prop.to   DET FUT.go.CT'that   gentleman NO FUT.retreat.CT'the
       miaramila
       soldier
       'The more the old man would advance the more the soldiers would back
       up.'   [Etsy 161]

d. Arakaraka ny   iasana   no [ahazoana vola   sy itomboana
prop.to   DET work.CT NO receive.CT money and expand.CT
fahasalamana]
health
'The more you work the more you earn money and expand your health.'

4. Semantic Interpretation

Write [w]$_s$ for the interpretation of an expression w in a situation s. For φ = [[$_{PP}$ *arakaraka* p] *no* [q]] a CC, interpret φ as True in s iff [p]$_s$ is proportional to [q]$_s$. This analysis assumes A1–A3 below, and raises the issue of uniform compositional interpretation of expressions of the form [X *no* Y].

A1: Both pre-*no* [*arakaraka* p] and post-*no* q denote scalar functions though they differ (slightly) in category. This arises as follows. ARAKARAKA selects or imposes scalarity on its argument, as in (7a). NO serves to pass this scalar interpretation on to the next argument, q, facilitated by the fact that q is a P1 and so is already interpreted as a function, perhaps even one with ordered values, like 'the number of houses' in (6b). This view is independently supported by our analysis of the conjunction *sady...no...* below. So *no* behaves like an internal conjunction (Zhang 2009) as in e.g. French coordinations [x A x B], where x = *et* 'and', *ou* 'or', or *ni* 'nor', as in *et Jean et Marie* 'both John and Mary', etc. The second conjunction places no independent selection requirements on its argument(s), it just maintains those imposed on the first.

A2: [*arakaraka,no*] is assigned a semantic interpretation. This may be done compositionally using a structure for CCs like <[arakaraka,no],[DP$_{CT}$][P1$_{CT}$]>. A "minor" rule would move the designated word *no* between its arguments. (See Larson's 1985 treatment of [*Either/Whether, or*]). This rule is explanatory. Moving *no* serves to identify the second argument of the proportionality function. Maintaining the voice of the first argument is similarly motivated.

A3: The pre-*no* [*ny*+P1$_{CT}$] determines a scalar, like the post-*no* P1$_{CT}$, which says that nominalizing the P1 preserves semantic properties. Randriamasimanana (2007) and Potsdam and Polinsky (2014) provide extensive independent support for this. Their examples use active P1s, as in (13a). But (13b) shows that circumstantial P1s behave comparably. In both (13a-b) the complement of 'buy' has its future tense selected by the governing verb 'intend' regardless of whether the Det *ny* is present or not. Equally in both cases the agent of 'buy' is the same as that of 'intend'. So nominalization preserves agenthood and temporal dependence, properties of P1s.

(13) a. n.i.kasa   (ny)   h.i.vidy   akanjo ho an'i   Soa   aho
PST.AT.intend (DET) FUT.AT.buy clothes for'ART   Soa   1NOM
'I intend to buy clothes for Soa.'

b. no.kasa.i.ko   (ny)   h.ividi.anana akanjo   i   Soa
   PST.intend.PASS.1GEN (DET)   FUT.buy.CT   clothes ART Soa
   'Soa was intended to be bought clothes for by me.'
   '(= I intended to buy clothes for Soa).'

In addition, while *ny* is often translated as definite, that interpretation is plausibly a default. *Ny* cooccurs naturally with weak (indefinite) as well as strong quantifiers (Keenan 2006, 2008):

(14) a. Nitsangana   [ny mpianatra rehetra] / [ny ankabeazan'ny mpianatra]
        stood.up   DET student   all   /   DET majority.of'the student
        'All the students stood up.   / Most of the students stood up.'
     b. Nitsangana ny mpianatra maro / maromaro /   sasany / vitsivitsy
        stood.up   DET student   many / somewhat many   / some   / few
        / roa / iray
        / two / one
        'Many / several / some / few / two / one student(s) stood up.'

Nor is it formally problematic to access the P1 interpretation from that of [*ny*+P1]. It suffices that *ny* denote a one to one function from P1 denotations to DP denotations. Quite generally nominalizing functions are one to one: if P1s p,q have different meanings so do nom(p) and nom(q). E.g. *walk* and *talk* have different meanings and so do *walking* and *talking*, *walkers* and *talkers*, *a/the walk* and *a/the talk*. So given a nominalization meaning we uniquely retrieve the P1 meaning (not an expression which denotes it) it is a function of. We turn now to some merits of our binary operator analysis and then one issue it raises.

5. Motivating Binary Operators

Binary operators are functions taking two arguments. Functions *select* their arguments: to define them we must state what their domain is, that is, what objects (pairs of objects, etc.) they apply to. The defining conditions are the *selectional restrictions* of function denoting expressions. Those functions we call *operators* are grammatical invariants (Keenan & Stabler 2003). Binary operators are often expressed by pairs of expressions one of which is largely predictable from the other, often identical to it. The two parts identify the two arguments of the function, which motivates its discontinuous expression. We will see a variety of plausible candidates for binary operators in Malagasy. Some candidates in English are: *neither...nor...*, *more...than...*, *else...else...*, *different...different...*, *for...to...*, *poss...ing...*[2], as in:

---

[2] Thanks to Ed Stabler for drawing my attention to the For-To and Poss-Ing cases.

(15) *Neither* John *nor* Mary left / *Ni* Jean *ni* Marie ne sont partis
  More students *than* teachers signed the petition   (Keenan 1987)
  Ben recognized Sue but no one *else* recognized anyone *else* (Keenan 1987)
  Different people like *different* things   (Keenan 1992)
  For Warren *to* object at that point was unexpected
  Zelda*'s* leav*ing* early annoyed me

5.1.   Binary Operators in Malagasy

5.1.1.   *Sady...no...* 'as well as'

Sady...no... 'as well as' coordinates predicate phrases of any arity (P0 = TP, P1 = tensed VP, P2 = Tensed Transitive Verb Phrase). *Sady* also occurs between Pn's without *no* (16).

(16) ... [nanda   sady   nikifikifika] izy ...
   ... refused SADY   shook.head she ...
   '... she refused and shook her head ...'

(17) a. [[Sady nikapoka   no nandaka]   bibilava roa] Rabe   (P2)
        SADY PST.AT.beat NO PST.AT.kick   snake   two Rabe
        'Rabe was beating and kicking two snakes'
         (≠ was beating two... and kicking two...)
    b. ... ny vehivavy [[sady   nibaby   no nitrotro]   ny zana.ny]
       ... the women   SADY   PST.back.carry NO chest.carry the child.3GEN
        '... the women who carried their children on their back and on their chest'
                              [IKM.52]

(18) a. [Sady matanjaka no   kinga   saina] Rabe   [P1]
        SADY strong   NO agile   mind   Rabe
        'Rabe is strong as well as quick-witted.'
    b. [Sady mahazo   masoandro tsara   no mahazo rivotra   madio]
        SADY get   sun   good NO get   air   clean
        ny toerana   manerinerina
        DET place   high
         'A high location gets good sun as well as clean air.'   [T- 4]
    c. olona [sady tsara fanahy no kinga saina] no   n.anampy ahy
       person [SADY nice spirit   NO agile mind   FOC PST.help   me
       h.anolo   ny   kodiarana
       FUT.change   the tire
        'People who were nice and smart helped me change the tire.'

(19) ... satria sady maty ny zavany   no very ny volany   ...   namidiny
    ... because SADY gone the thing.his NO lost   the money.his ... he paid.for
    azy   (P0)
    them
    '... because his things were gone and the money he paid for them lost' [Ang]

Importantly, *sady...no...* requires its two arguments to be *semantically comparable*. Speakers balk at conjoining stage level and individual level predicates, (20), though both can be stage level, (17a), and both individual level, (18a-b). So *sady* alone cannot select one over the other. Rather, *sady…no…* selects a pair of semantically comparable expressions.

(20) *?[Sady   tsara tarehy no mikapoka   bibilava roa] Rasoa.
         SADY   good face   NO PRES.AT.beat snake   two Rasoa
    '*? Rasoa is pretty as well as is beating two snakes.'

Further, *sady...no...* selects predicates, but not DPs, (21), which do coordinate freely with *sy* 'and'. Category restrictions on coordinators are fairly widespread: Zhang (2009:45). Also, not too unsurprisingly, the *no* in *sady...no...* occasionally iterates, (22).

(21) *Efa   nivoaka   sady Rabe no Ranaivo.
    already gone.out   SADY Rabe NO Ranaivo
    'Rabe as well as Ranaivo have gone out.'

(22) [Sady mitsara no mikapoka, no manambitamby ary masiaka] no fomba
    SADY judges NO beats   NO cajoles   and severe   NO manner
    fitondra zatovo
    raise   young
    'Judging, beating, cajoling, and being severe are manners of raising the young.'

5.1.2.   Other Candidates for Binary Operators

(23) *a.leo*.ko   maty miaraka amin-dRazay tokoa, *toy izay* hamela   azy
    TT.prefer.my die   together with-Razay   emph, than   FUT.leave her
    eo am-pelatànan'ny ratsy
    LOC at-palm'the   bad
    'I prefer to die with Razamalala than to leave her in the hands of the bad
    guys.'   [IKM 24]

(24) a. Tsy avy any *na* ny rainao *na* ny reninao
       not come there NA the father.your NA the mother.your
       'Whether it's your father or your mother, that person didn't come from
       there.'   [Bleu 60]

   b. Nijininika ra ny ratrany *hatrany* am-piasana *ka*
      flowed blood the wound.his from LOC-NOM.work CONJ
      hatrany amin'ny trano
      to prep'the house
      'His wound bled from his (place of) work to his house.'

   c. *Hatramin*'ny oviana *ka* *hatramin*'ny oviana no h.ijanon.a.nao
      from'the when CONJ up.to'DET when NO FUT.stay.CT.2S.GEN
      any?
      there
      'From when to when will you stay there?'

We also find discontinuous unary operators whose two parts identify edges of
constituents. Demonstratives are a prominent case. They may self-embed, (25).

(25) a. Hatramin'*izao* daty anoratanay *ity* lahasoratra
       since'this date write.CT.1PL.EXCL.GEN this article
       ity *izao*, dia ...
       this this, and so...
       'Since the day of writing this article...' [Newspaper 1994]

   b. ...an'ireo biby nomeny anjara eto amin'*ity* anganon'*izao* andro *izao*
      ACC'those animals given role here in'this tale'this day this
      ity?
      this
      '...those animals given a role in this tale of our time' [IS]
   c. Ny nahatongavany *teo* amin'*io* toeran'*io* *teo* hono dia
      the PST.arrive.CT.his PST.there at'that place'that there they.say TOP
      nobaben'ny reniny
      carried'the mother.his
      'His arrival there at that place was by his mother carrying him on her
      back.'

Lastly, cross linguistically the expression of CCs often involves discontinuous
expressions with the second the same as, (26a-e), or constrained by, the first,
(27).

(26) a. *Yuè* tián   de píngguǒ *yuè* haǒchī   Mandarin
    more sweet REL apple   more delicious   Hsiao 2003
    'The sweeter an apple the more delicious it is.'

  b. Ni *yuht*   gong, keoi *yuht*   ganzoeng   Cantonese
    you surpass talk, 3S   surpass nervous   Leung 2005
    'The more you talk the more nervous s/he is.'

  c. Khun *ying*   phuut, khaw *ying* dandeng   Thai
    you increasingly talk,   3S   increasingly nervous Leung 2005
    'The more you talk the more nervous he gets.'

  d. Ali *makin*   besar   *makin*   sombong   Indonesian
    Ali increasingly old   increasingly arrogant   Leung 2005
    'The older Ali gets the more arrogant he becomes.'

  e. *The* bigg*er* they are *the* hard*er* they fall   English

(27) a. *Kolkoto*   četeš, *tolkova*   maučavaš   Bulgarian
    how.much read.2S, that.much learn.2S   Bhatt 2008
    'As much as you read, that much you learn.'

  b. [IP[*Jiitnaa*   suuraj chamk-aa][*utnii*   ThanD baRh-ii]]   Hindi
    How-much.MSG sun.M shine-PF   that-much.F cold.F increase-PF
    'The more the sun shone, the colder it got.'   Den Dikken 2005

5.3.   Malagasy CCs: Paratactic? Correlative?

CCs are, "semantically paratactic" by definition: to be CCs they must present two scalar expressions. In (26) and (27) the two scalars look structurally identical but this is less so in Malagasy: one scalar is a DP, the other a P1, neither a full clause. Still they are both circumstantial and CT P1s may include all the subcategorized arguments of the active verb. On plausible structures for CCs, the two scalar constituents are on the same level: in <[arakaraka,no],[DP$_{CT}$][P1$_{CT}$]> each c-commands the other, in <[[arakaraka [DP$_{CT}$]][no [P1$_{CT}$]]> neither DP$_{CT}$ nor P1$_{CT}$ c-commands the other. Pronominalization patterns are consistent with c-command neutrality. Both antecedent-pronoun orders in (28a-b) are acceptable:

(28) a. Arakaraka ny ifaneraserako   ami.n'i Soa$_j$ no tsy ahafantarako
    prop.to   DET socialize.CT.my PREP.'ART Soa NO not understand.CT.my
    azy$_j$
    3ACC
    'The more I socialize with Soa the less I understand her.'

b. Arakaraka ny ifaneraserako ami.ny$_j$ no tsy ahafantarako
    prop.to DET socialize.CT.my PREP.3GEN NO not understand.CT.my
    an'i Soa$_j$
    ACC'ART Soa
    'The more I socialize with her the less I understand Soa.'

In a context in which *-ny* 'his/her/their' does not have a plausible antecedent in the preceding discourse the binding in (28b) is acceptable, more so than in English.

    Equally Malagasy CCs resemble correlatives in that the second half contains an item, *no*, syntactically and semantically dependent on one, *arakaraka*, in the first half. But in distinction to classical correlatives, as (27), the dependent element is not referential, e.g. a pronoun or demonstrative, and lacks nominal features (number, gender, case). And the item it is dependent on is not a relative or interrogative word, it is just a preposition. Still the cooccurrence restriction between the two is a grammatical relation over and above mere branching structure.

    Lastly, in some languages (English, German) CCs present the morpho-syntax of basic comparatives (x *is bigger than* y). But not so in Malagasy (nor, it appears, in Mandarin or Cantonese). Malagasy has two types of basic comparatives:

(29) a. Lehibe (kokoa) noho ianao izy
      big COMP prep 2S.NOM 3.NOM
      'He is bigger than you.'
   b. Lehibe m.i.hoatra anao izy
      big PRES.AT.surpass 2S.ACC 3.NOM
      'He is bigger than you.'

But their distinctive comparative elements, *noho* and *mihoatra*, are not part of CC structure, and contra German according to Beck 1997, comparatives can independently occur in CCs:

(30) Arakaraka ny ahazoanao vola noho Rabe no maha.tezitra azy
    prop.to DET receive.CT.2S.GEN money PREP Rabe NO cause.angry him
    'The more your salary increases compared to Rabe's the more that angers him.'

6. Problem: a unified syntax for [X no Y]?

The core use of *no* is in Clefts, (32b), in which the post-*no* string is a P1, any voice, and the pre-*no* one, the focused constituent, is its subject:

(31) a. N.i.homehy   Rabe   b. Rabe no n.i.homehy
      PST.AT.laugh   Rabe   Rabe NO PST.AT.laugh
      'Rabe laughed.'   'It was Rabe who laughed.'

Paul 2001, 2008, Potsdam 2006, Law 2007 and Kalin 2009 treat the focused constituent–*Rabe* in (31b), as a predicate and (except for Law) the *no* phrase as its definite DP subject, roughly, (*the one*) *who laughed*. Law, Pearson 2009, and Gärtner 2009 note some serious problems with the *no*+P1 = DP part of this analysis, but this pseudocleft approach is supported by two syntactic facts and one, more complex, semantic one, which an adequate analysis of *no* must account for.

       Syntactically, in general negative polarity items (npi's) and the polar question particle occur at the right edge of P1s, (32a-b), and this placement treats focused constituents as P1s, (32c-d):

(32) a. Tsy nihomehy   akory Rabe   b. Nihomehy   ve Rabe?
       not PST.AT.laughat all Rabe   PST.AT.laugh   Q Rabe?
       'Rabe didn't laught at all.'   'Did Rabe laugh?'

    c. Tsy Rabe akory no n.i.homehy   d. Rabe ve no n.i.homehy?
      not Rabe at.all NO PST.AT.laugh   Rabe Q   NO PST.AT.laugh
      'It wasn't Rabe at all who laughed'   'Was it R who laughed?'

Semantically we infer from (32c-d) that someone laughed – we just deny it was Rabe in (32c) and query whether it was him in (32d). The Cleft analysis predicts these facts, which provide substantive support that CCs and Clefts share structure. Both have the gross form *X no Y* and npi's and Q distribute identically in the two cases:

(33) a. Arakaraka ny iasana   ve no ahazoana   vola?
       prop.to   DET work.CT Q NO receive.CT money
       'Is it so that the more you work the more you earn?'
    b. Tsy arakaraka ny iasana   akory no ahazoana vola
       not prop.to   DET work.CT at.all NO receive.CT money
       'It is not at all so that the more you work the more you earn.'
    c. Tsy arakaraka ny ifaneraserako   amin'i   Soa no tsy
       not prop.to   DET REC.frequent.CT.my with'ART Soa NO not
       ahafantarako   azy
       understand.CT.my 3ACC
       'It is not so that the more I socialize with Soa the less I understand her.'

But semantically CCs differ from core clefts in not presupposing existential quantification over the post-*no* predicate. (33b) just denies that, and (33a) just queries whether, the proportionality relation holds between receiving money and working hard. Neither implies that receiving money correlates with any other activity. This semantic judgment is supported by the natural reading of (33c), the negation of (28b), which does not even suggest that there is an activity aside from my frequenting Soa which my understanding of her decreases with.

This absence of presupposition supports that CCs are structurally different from core clefts (identical structures assumed to be interpreted identically). But data in Gärtner 2009 give us pause. From Raharirinina-Rabaovololona 1991, he notes (34a), which is not obligatorily presuppositional but can be used as a general statement about the function of knives, in contrast to (34b). Gärtner draws a parallel, (35), with the French *c'est...qui/que* construction which translates core clefts. In (36a-d) we note some further cases of not clearly presuppositional uses (all present tense, note):

(34) a. (Ny) antsy no ∅-andidi.ana mofo
      The   knife NO PRES-cut.CT   bread
      'Knives serve to cut bread.'
    b. Ity antsy ity   no ∅-andidi.ana mofo
      this knife this   NO PRES-cut.CT bread
      'It is this knife which is for cutting bread.'

(35) a. Que-ce qui se passe?
      'What's going on?'
    b. C'est maman qui me bat
      'Mum's just beating me (that's all).'

(36) a. Amin'izao fotoana izao mantsy   dia ny fiaraha-monina mihitsy   no
      PREP'now time   now to.be.sure TOP the society   itself   NO
      marary
      sick
      'At the current time society itself is sick.'   [JR 88]
    b. Ny fifehezana ny vidin-javatra no lalana iray hahazoa-mitsimbina   ny
      the controlling the price-thing NO way   one FUT.able.CT-look.after the
      fahefa-mividy
      power-to.buy
      'Controlling prices is one way of being able to protect purchasing
      power...'   [JR 126]
    c. – Iza   no te-hiala?   – Tsy misy
      who NO wants-FUT.leave   not exist
      – 'Who wants to leave?'   – 'No one.'   (Eric Potsdam, pc)

6.1. Why are clefts presuppositional?

The pseudocleft approach offers a syntactic answer: the *no* phrase in clefts is a definite DP (or free relative) independently presupposed referential. But we noted that several objections to this analysis have been raised. A further one is that *no* has additional uses not really compatible with a definite DP or free relative analysis. The coordinator *sady...no...* is one such. A second, little noted, is *no* with strong speech act markers: imperatives (Koopman 2005), hortatives, and prohibitives:

(37) a. Vonjeo   aho!
     save.TT.IMP   1S.NOM
     'Save me!'
   b. Io zaza io$_i$ no vonjeo   fa nianjera e$_i$!
     that child that NO save.TT.IMP for fell   he
     'Save that child (for he has fallen)!'

(38) a.b. Ny bokiko   ihany no aoka ho entinareo   fa ny azy no
         the book.my only   NO HORT FUT carry.TT.2PL.GEN but the 3ACC NO
         aza   entina!
         PROH carry.TT
         'Let us take my books only! but his, don't take!'

Third, Gärtner 2009 and Pearson 2009 discuss *no* used to conjoin two full clauses (no gaps) with a temporal overlap or a causative/enabling interpretation. For reasons of space we just add one example in which the enabling clause is a question, a natural structure type in Malagasy. Note that what we question in (39c) is the reason the following clause (a cleft!) holds.

(39) a. Moa lany   ve ny rano   no tsy misasa   ianao?
       PRT used.up Q the   water NO not AT.wash.INTRANS 2S.NOM
       'Is it because the water is used up that you don't wash?'
   b. Nifandidy antsy tamin'iza   hianao no feno ra   toy io ny
     PST.REC.cut knife PST.with'who? 2S.NOM NO full blood like that the
     lobakao?
     shirt.2S.GEN
      'With whom did you mutually knife cut whence your shirt is full of
     blood like that?'   [Tak 314]
   c. Nahoana   no tsy izaho   no nalain'i   Damo ho vady?
     why   NO not 1S.strong NO PST.take.TT'ART Damo FUT spouse
     'Why was it that it wasn't I who was taken by Damo as (his) spouse?'

Our response to the question in 6.1 is more semantic. The focused (pre-*no*) constituent in a cleft, core or non-core, identifies an *inferable concomitant* (property, circumstance) of the event or state determined by the predicate. Being a participant – Agent, Theme, Goal, ... are such concomitants, but so are Time, Place, Means, Manner, Cause, Purpose, ... as appropriate. (We infer an Instrument if you open a can but not if you greet a visitor; we infer Place and Time in both.) So negating or questioning a focused constituent just challenges the identification of the concomitant, not whether there is one, which is independently assumed.

In crucial contrast are CCs. The pre-*no* clause does not identify an inferable concomitant of the verb denotation. It compares it to something *extraneous* – hence the two independent predicates. Negating and questioning just naturally challenge whether the comparison holds.

These remarks highlight the semantic difference between clefts and CCs, but why should npi's and Q distribute similarly in the two cases? We suggest that CCs are built from a common type of non-core cleft but differ in interpretation in virtue of the lexical meaning of *arakaraka* 'proportionate to', which is inherently relational. Moreover this difference in interpretation induces a correlative type restructuring. The relevant non-core clefts are in (40c) and (41c).

(40) a. Manapaka bozaka amin'ny antsinay izy
        PRES.AT.cut grass with'the knife.1PL.EXCL.GEN 3.NOM
        'He cuts grass with our knife.'

  b. Anapahany bozaka (*amin')ny antsinay
        cut.CT.3GEN grass with'the knife.our
        'The knife is used by him to cut grass.'

  c. (Amin')ny antsinay no anapahany bozaka
        with'the knife.our NO cut.CT.3GEN grass
        'It is with our knife that he cuts grass.'

The verb in (40a) is active; in (40b) circumstantial with the subject Instrumental. PPs do not occur in (clause final) subject position. But when focused the preposition (here) remains. For some other obliques, even with the same preposition, (40b) is not acceptable and the preposition in (40c) is obligatory (See [RR]:112). And in this pattern npi's and Q behave as in (41a-b), like those in CCs, (33a-b), repeated as (43a-b) here:

(41) a. Miarahaba anao amin-kafaliana be izahay
        PRES.AT.greet 2s.ACC with-joy great we.EXCL.NOM
        'We greet you with great joy.'

    b.*Iarahabanay   anao ny hafaliana be
    c. Amin-kafaliana be   no   iarahabanay   anao
       with-joy   great NO greet.CT.1PL.EXCL   2S.ACC
       'It is with great joy that we greet you.'

(42) a. Tsy amin-kafaliana be   akory no iarahabanay   anao
        not with-joy   great at.all NO greet.CT.1PL.EXCL 2S.ACC
        'It isn't at all with great joy that we greet you.'
    b. Amin-kafaliana be   ve   no   niarahaban-dRabe   anao?
       with-joy   great Q   NO PST.greet.CT-Rabe.GEN 2S.ACC
       'Was it with great joy that Rabe greeted you?'

(43) a. Tsy arakaraka ny   iasana   akory no ahazoana vola
        not prop.to   DET work.CT at.all NO receive.CT money
         'It is not at all so that the more you work the more you earn.'
    b. Arakaraka ny iasana   ve no   ahazoana   vola?
       prop.to   DET work.CT Q NO   receive.CT money
       'Is it so that the more you work the more you earn?'

CCs however do not behave exactly like focused PPs from circumstantial clauses, as they do not accept the Bodyguard construction. Compare (44a) with (44b).

(44) a. Amin-kafaliana be   izahay   no miarahaba   anao (+Bodyguard)
       with-joy   great 1PL.EXCL.NOM NO PRES.AT.greet 2S.ACC
       'It is with great joy that we greet you.'

    b.*Arakaraka ny iasanao   Rabe no mahazo   vola (+Bodyguard)
       prop.to   DET work.CT.2S.GEN Rabe NO AT.receive   money
       'The more you work the more Rabe receives money.'

We should note that the Bodyguard construction might seem unnatural – it looks as though the active subject and an oblique are simultaneously fronted. But it is in fact quite widely used. The following (partial) paradigm is boringly natural:

(45) a. Mipetraka   aiza izy?   b. Aiza   no mipetraka izy?
       PRES.AT.live where 3NOM   where NO AT.live   3NOM
       'Where does he live?'   'Where does he live?'

    c. Aiza izy   no mipetraka?   d. Aiza   no ipetrahany?
       where 3NOM NO PRES.AT.live   where NO PRES.live.CT.3GEN
       'Where does he live?'   'Where does he live?'

So if the initial scalar [*arakaraka* [*ny* P1$_{CT}$]] in a CC is a mere PP (44b) should be grammatical, parallel to (44a). But it isn't. We conclude then that CCs are not simply clefts with a focused PP. Rather *arakaraka* and *no* form a syntactic and semantic unit, which retains properties of the preposition and focus particle it is formed from.

Sources

[Ang]   *Contes des Aieux Malgaches* (*Anganon'ny Ntaolo* "Tales of the Ancestors") 1992 (1904). Collected and edited by Lars Dahle; then John Sims. Bilingual edition translated by D. Dorian and L. Molet. Institut des Langues et Civilisations Orientales, Paris.

[JR]   Jedele, Thomas and Lucien E. Randrianarivelo. 1998. *Malagasy Newspaper Reader*. Dunwoody Press. Kensington, Maryland. October 1992-November 1995.

[Bleu] Ravololomanga, Bodo. 1996. *Le Lac Bleu et autres contes de Madagascar*. L'Harmattan.

[Etsy] Razanadrakoto, Lucien. 1988. *Etsy Ho Ahy Ny Tananao* A.C.E. Tananarivo.

[IKM]   Rajohanesa, P. 1963. *Ilay Kintana Mamirapiratra*. Imprimerie Luthérienne. Tananarive.

[IS]   *Isambilo* T-4. 1990. Rajemisa-Raolison, Régis and Lala Rakotoson-Raolisoa. Ambozontany,

[Raki] *Rakibolana*. 1985. Rajemisa-Raolison, Régis. Ambozontany, Fianarantsoa

[RR]   Rajemisa Raolison, Régis. 1971. *Grammaire Malgache* Ambozontany, Fianarantsoa.

[Tak]   Siméon Rajaona (ed). 1972. *Takelaka Notsongaina II: Lahatsoratra Tsotra* (Selected Pages II: Basic Writings). Ambozontany, Fianarantsoa.

[T-4]   1989. *Fanazarana Hiteny sy Fanabeazam-boho* T-4. Ranaivomanana, Raymond. FOFIPA Antananarivo.

References

Beck, Sigrid. 1997. On the semantics of comparative conditionals. *Linguistics and Philosophy* 20:229- 271.

Bhatt, Rajesh. 2008. Comments on 'Comparative Correlatives in Greek: The syntax of *oso*'. The Proceedings of the 2007 MIT Greek Syntax-Semantics Workshop.

Brasoveanu, Adrian. 2008. Uniqueness effects in correlatives. A. Grønn (ed.), *Proceedings of Sinn und Bedeutung* 12:47–65. ILOS, Oslo.

Culicover, Peter and Ray Jackendoff. 1999. The view from the periphery: The English comparative correlative. *Linguistic Inquiry* 30.4:543–571.

Den Dikken, Marcel. 2005. Comparative correlatives comparatively. *Linguistic Inquiry* 36.4:497–532.

Dez, Jacques. 1980. *Structures de la Langue Malgache*. POF 2 rue de Lille, 75007 Paris.

Gärtner, Hans-Martin. 2009. On the prospects of a clause combining approach to "focusing" *no*-constructions in Malagasy. *Natural Language and Linguistic Theory* 27:789–816.

Hsiao, Su-Ying. 2003. On proportional correlative constructions in Chinese and Mongolian. *Journal of Taiwanese Languages and Literature* 1.1:243–272.

Kalin, Laura. 2009. News about the no-phrase: Specificational pseudoclefts in Malagasy. *Proceedings of the 16th Annual Meeting of the Austronesian Formal Linguistics Association (AFLA 16)*, *eds*. Sandra Chung, Daniel Finer, Ileana Paul and Eric Potsdam, 61-75.

Keenan, Edward L. 1972. Relative clause formation in Malagasy (and some related and some not so related languages). *The Chicago Witch Hunt*. 169–189.

Keenan, Edward L. 1976. Remarkable subjects in Malagasy. *Subject and Topic*, ed. Charles N. Li, 247-301. New York: Academic Press.

Keenan, Edward L. 1987. "Multiply-Headed NPs *Linguistic Inquiry* 18.3:481-90.

Keenan, Edward L. 1992. Beyond the Frege Boundary. *Linguistics and Philosophy* 15:199-221.

Keenan, Edward L. 2006. Quantification in Malagasy. Quantification: A Cross-Linguistic Perspective, ed. Lisa Matthewson, 319-352. Bingley, UK: Emerald.

Keenan, Edward L. 2008. The definiteness of subjects and objects in Malagasy. *Case and Grammatical Relations*, eds. Greville G. Corbett and Michael Noonan, 241-261. Amsterdam: John Benjamins.

Keenan, Edward L. and Edward P. Stabler. 2003. *Bare grammar: Lectures on linguistic invariants*. CSLI, Stanford Ca.

Koopman, Hilda. 2005. Malagasy imperatives. *Proceedings of the 12th Annual Meeting of the Austronesian Formal Linguistics Association (AFLA XII)* (UCLA WPL): 141–160.

Larson, Richard. 1985. On the syntax of disjunction scope. *Natural Language and Linguistic Theory* 3:217–264.

Law, Paul. 2007. The syntactic structure of the cleft construction in Malagasy. *Natural Language and Linguistic Theory* 25:765–823.

Leung, Tommi Tsz-Cheung. 2005. Typology and universals of comparative correlatives. Paper presented at the Association for Linguistic Typology 6th Biennial Meeting (ALT VI). Pedang, Indonesia.

McCawley, James. 1988. The comparative conditional construction in English, German and Chinese. *Berkeley Linguistics Society 14,* 176-189. Berkeley Lingusitics Society, Berkeley, Ca.

Ntelitheos, Dimitrios. 2012. *Deriving nominals: A syntactic account of Malagasy nominalizations*. Leiden: Brill.

Paul, Ileana. 2001. Concealed pseudo-clefts. *Lingua* 111:707-727.

Paul, Ileana. 2008. On the topic of pseudo-clefts. *Syntax* 11.1:91-124.

Pearson, Matthew. 2009. Another look at *no*: Pseudo-clefts and temporal clauses in Malagasy. *Proceedings of the 16th Annual Meeting of the Austronesian Formal Linguistics Association (AFLA 16)*, *eds*. Sandra Chung, Daniel Finer, Ileana Paul and Eric Potsdam, 165-179.

Potsdam, Eric. 2006. More concealed pseudoclefts in Malagasy and the Clausal Typing Hypothesis. *Lingua* 116:2154-2182.

Potsdam, Eric & Maria Polinsky. 2014. Control into nominalized clauses in Malagasy. Paper presented at the 21st Annual Meeting of the Austronesian Formal Lingusitics Association (AFLA 21). Honolulu, HI.

Raharinirina-Rabaovolona, Lucie. 1991. *Lexique-Grammaire des Composés du Malgache: Les Adverbes de Temps*. Thèse de doctorat. Université Paris 7.

Rajaona, Siméon. 1972. *Structure du Malgache* (*Etude des formes prédicatives*) Ambozontany, Fianarantsoa. 206-210.

Randriamasimanana, Charles. 2007. Malagasy control structures. *Concentric: Studies in Linguistics* 33.2:91-120.

Schwarzchild, Roger. 2006. The role of dimensions in the syntax of noun phrases. *Syntax* 9:67–110.

Smith, E. Allyn. 2010. *Correlational comparison in English*. Doctoral dissertation, Ohio State University, Columbus, OH.

Zhang, Niina. 2009. *Coordination in syntax*. Cambridge: Cambridge University Press.

THE ARGUMENT STRUCTURE OF AMIS EXPERIENCER VERBS: ANOTHER LOOK AT THE *MA-* MORPHOLOGY[*]

Jonathan Kuo
University of Hawai'i at Mānoa / Academia Sinica
kuocc@hawaii.edu

This paper investigates the argument structure of Amis Experiencer verbs and identifies an additional function of *ma-*, which has not been discussed in the literature. Adopting the predicate decomposition approach, I categorize Amis Experiencer verbs into two types: a) internally caused psych verbs, which employ *ma-* as v_{Be} and *mi-* as v_{Do}, and b) externally caused psych verbs, which involve *mi-* as v_{Cause} and *ma-* as "anticausative". By proposing the distinctness of VoiceP and vP functional projections, I further show that the absolutive case assignment of these experiencer verbs is not at all arbitrary. The discovery of the anticausative function of *ma-* in Amis may serve as a crucial link between the stative function and the eventive function. In all these different type of verbs, *ma-* invariably licenses the *v*P-internal arguments. This suggests the possibility of the development of *ma-* from stative to eventive, with the transition of anticausative.

1.   Introduction

The Philippine-type voice morphology has recently been treated as markers of intransitivity/transitivity (ITR/TR): Actor Voice (AV) indicates (syntactic) intransitivity whereas Undergoer Voice (UV) marks transitivity (Aldridge 2004, Ross & Teng 2005, Chang 2011, inter alia). This is shown in the Amis example (1), where AV/ITR *mi-* licenses the Actor 'person' and UV/TR *-en* licenses the Undergoer 'pig' as the absolutive argument.[1]

(1) a. **mi**-adup   **k-u-ra**   **tamdaw** t-u   fafuy
      ITR-hunt   ABS-CN-that   person   OBL-CN   pig
      'That person hunts pigs.'

---

[*] I would like to thank the native speaker consultants, Ofad, Lisin, and Talud. Special thanks to Yuko Otsuka, Shinichiro Fukuda, Elizabeth Zeitoun, and Stacy Teng for their valuable comments.
[1] List of abbreviations: 1/2/3SG: 1/2/3 person singular, 1/2/3PE/I: 1/2/3 person plural exclusive/inclusive, ABS: Absolutive, ACAUS: anticausative, CAUS: causative, CN: common noun marker, ERG: ergative, GEN: genitive, IRR: irrealis, ITR: intransitive, OBL: oblique, TR: transitive, PPN: Personal proper noun.

    b.   adup-**en**   n-u-ra   tamdaw   **k-u**   **fafuy**
         hunt-TR   ERG-CN-that person   ABS-CN   pig
         'That person hunted the pig.'

Unlike *mi-* or *-en*, which has a particular voice function, *ma-* in Amis has been found to have different voice "possibilities", depending on the root it attaches to:

(2) a. **ma**-fanaq   **ci**   **sawmah** t-u-ra   wawa[2]
      ITR-know   PPN   sawmah OBL-CN-that child
      'Sawmah knows that child.'

    b. **ma**-palu   ni sawmah **k-u-ra**   **wawa**
       TR-beat   ERG sawmah ABS-CN-that   child
       'Sawmah beat that child.'

Instead of identifying two homophonous *ma-*'s, Wu (2007a) takes a polysemy view and analyzes *ma-* as AV and UV, respectively. This paper identifies an additional grammatical function of *ma-*, namely **anticausative**, found in a particular type of Amis Experiencer verbs (e.g. 'annoy'). For the sake of convenience, the rationale of the analysis is provided here, to be justified carefully in later sections.

(3)   The causative alternation (modified based on Schäfer 2008:9)
    a. **agent/causer**$_{SUB}$ V_{tr} undergoer/theme e.g. *John broke the window*.
    b. **undergoer**$_{SUB}$ V_{intr} e.g. *The window broke*.

(4)   Amis *mi-/ma-* argument alternations
    a. **mi**-'esam   **k-u-ra**   **lalangaw**(t-u   tamdaw)
      CAUS-annoy ABS-CN-that fly   OBL-CN   person
      'The fly is annoying (people).'

    b. **ma**-'esam   **k-u-ra**   **tamdaw** (t-u-ra   lalangaw)
      ACAUS-annoy   ABS-CN-that person   OBL-CN-that fly
      'That person is annoyed (with that fly).'

(3) demonstrates how verbs expressing a change of state can participate in the so-called causative alternation, used either as a transitive/causative verb or an intransitive/anticausative verb. In the transitive variant, the initiator of the event

---

[2] Unlike common nouns, personal proper nouns in Amis do not carry overt absolutive marker such as *k-*. The PPN alone implies the absolutive marking (e.g. *ci sawmah*), while the addition of the suffix *-an* indicates the oblique marking (e.g. *ci sawmah-an*).

(i.e. agent/causer) surfaces as the subject. In the intransitive variant, this participant is missing, and the undergoer surfaces as the subject. (4) provides a parallel case of Amis, in which *mi-* licenses the initiator of an emotion, and *ma-* affixation licenses the undergoer of the emotion by "detransitivizing" the two-place pysch(ological) predicate 'annoy'. The anticausative analysis provided in (4) is based on a careful investigation on the derivational properties of Amis voice markers *mi-* and *ma-*, to be revealed in Section 4.

The paper is organized as follows. Section 2 introduces the dichotomy of Amis Experiencer verbs established in Wu (2007b), and discusses the challenges these verbs bring to the analysis of voice markers, especially *ma-*. Section 3 reviews a decomposition approach to Experiencer verbs based on the contrast of "internally and externally caused eventualities" (Levin & Rappaport Hovav 1995). In Section 4, I provide a minimalist analysis of Amis Experiencer verbs, thereby identifying the functions of *mi-* and *ma-* in terms of the flavor of *v* (Harley 2009). Adopting Alexiadou et al. (2006), I argue for the case of anticausative *ma-* in Amis Undergoer-Experiencer class. Section 5 provides a discussion regarding how the presence of anticausative *ma-* may serve as an intermediate stage between its original "AV" and innovative UV usage.[3] Section 6 concludes.

2. Actor$_{Exp}$ vs. Undergoer$_{Exp}$ verbs in Amis: An unsolved puzzle

The study of psych(ological) verbs has proven insightful to the theory of argument structure. One of the significant findings is the distinction between Subject Exp(eriencer) verbs and Object Exp verbs, as shown in the following English examples.

(5)   Subj$_{Exp}$ vs. Obj$_{Exp}$ verbs in English
    a. **[Indiana Jones]$_{Exp}$** feared [the snakes]$_{Target}$.
    b. [The snakes]$_{Causer}$ frightened **[Indiana Jones]$_{Exp}$**.

Various approaches have been applied to the dichotomy of Exp verbs (e.g. Belletti & Rizzi 1988, Dowty 1991, Grimshaw 1990, Landau 2010, Pesetsky 1995). Among others, Pesetsky (1995) carefully examines the truth conditions of sentence pairs with their Subj$_{Exp}$/Obj$_{Exp}$ verbs denoting the same emotion (e.g. *fear* in 5). He concludes that these two types of verbs have distinct argument structures and involves different thematic roles (see also Levin & Grafmiller 2013).

Wu (2007b) proposes a similar dichotomy of Experiencer verbs in Amis based on their argument structures, semantic interpretations, and derivation

---

[3] Later in this study, I will show that "AV" is not an adequate term to characterize the properties of *ma-*.

constraints, as briefly summarized below.

Table 1. Psych verbs in Amis

	ma-√ examples	The reading of *mi-*√ counterparts	*pa-ka-*√	√-*en*
Actor$_{Exp}$ class	*ma-ulah* 'like', *ma-talaw* 'fear', *ma-inal* 'envy', *ma-ngudu* 'respect'…	purposive/motional	Yes	Yes
Undergoer$_{Exp}$ class	*ma-esam* 'annoyed', *ma-lanang* 'annoyed (by sound)'	Causative	No	No

Based on the Role and Reference Grammar (RRG), Wu (2007b) uses the terms **Actor**-Experiencer and **Undergoer**-Experiencer verbs for the dichotomy in Amis. She discovers that while all emotion-denoting roots (e.g. *talaw* 'fear', *'esam* 'annoy') allow *mi-* and *ma-* voice morphology, the interpretation of *mi-* marked verbs differs.[4]

(6)   Amis Actor$_{Exp}$ vs Undergoer$_{Exp}$ verbs: semantic motivation
   a. **mi-ulah** ci   sawmah t-u-ra   wawa
      ITR-like PPN sawmah OBL-CN-that   child
      'Sawmah (will) like that child purposefully.'
      ("purposive/motional" reading)

   b. **mi-'esam**   k-u-ra   lalangaw (t-u   tamdaw)
      ITR-annoy   ABS-CN-that   fly   OBL-CN   person
      'That fly is annoying (people).'
      ("causative" reading)

Supporting evidence for this dichotomy also comes from the observation that only one type of roots allows certain affixations (e.g. *pa-ka-* causative and -*en* TR). Here I only address the semantic motivation. As shown in (6), Wu observes that with the same *mi-* voice marking, only certain Exp verbs (e.g. *mi-ulah*) describe emotions carried out by the experiencer purposefully. For other *mi-* marked Exp verbs (e.g. *mi-'esam*), the experiencer is not intentional—there is an overt causer responsible for the emotion.

---

[4] In her RRG analysis, Wu chooses AV/UV for voice markers and NOM/GEN/DAT for the case of arguments (and adjuncts). For the sake of consistency, the glossing of the Amis examples cited from other works (e.g. Wu 2006, 2007a, 2007b and Tsukida 2008) will be adjusted accordingly.

Wu's observation on Amis Actor$_{Exp}$/Undegoer$_{Exp}$ verbs is compatible with the findings about Subj$_{Exp}$/Obj$_{Exp}$ verbs in the literature. This is especially clear when Pesetsky's (1995) thematic labels are adopted, as shown below.

(7)   Amis Actor$_{Exp}$ vs Undergoer$_{Exp}$ verbs: argument structure
   a. *mi*-marked Actor$_{Exp}$ verbs: <**Exp**, Target>   (e.g. 6a)
   b. *mi*-marked Undergoer$_{Exp}$ verbs: <Causer, **Exp**>   (e.g. 6b)

(7) justifies Wu's terminology for the dichotomy: a) Actor$_{Exp}$ verbs, which select Exp as the external argument, and b) Undergoer$_{Exp}$ verbs, which select Exp as the internal argument. Wu further argues that Actor$_{Exp}$ class denotes internally motivated psych states and Undergoer$_{Exp}$ class denotes externally triggered psych states.

   While Wu's arguments for the Actor$_{Exp}$/Unergoerfor$_{Exp}$ division are valid, information about the argument structure of *ma-* marked Exp verbs is incomplete in her study. Here, I continue this line of research by presenting the following sentences and demonstrating a puzzle regarding the function of *ma-* voice morphology.

(8)   The function of *ma-* in Amis Actor$_{Exp}$/Undergoer$_{Exp}$ verbs
   a. **ma**-ulah **ci**   **sawmah** t-u-ra   wawa
      ITR-like PPN sawmah OBL-CN-that   child
      'Sawmah likes that child.'

   b. **ma**-'esam   **k-u-ra**   **tamdaw** (t-u-ra   lalangaw)
      ITR-annoy   ABS-CN-that   person   OBL-CN-that   fly
      'That person is annoyed (with that fly).'

The argument structure of *ma-* marked Exp verbs is interesting. At first glance, it provides a false impression that there is no distinction between the previously established Actor$_{Exp}$/Unergoerfor$_{Exp}$ dichotomy, as both verbs in (8a-b) choose the experiencer as the absolutive argument. With respect to syntactic transitivity, *ma-* in both sentences should be glossed as ITR, as suggested by the case marking pattern (i.e. no ergative). However, *ma-* in (8b) is questionable with respect to its thematic (or macrorole) features. As discussed previously, Undergoer$_{Exp}$ class denotes externally triggered psych states and selects the experiencer as the internal argument. In (8b), *ma-* cannot be analyzed as AV, because it licenses the undergoer experiencer as the absolutive argument; it cannot be treated as a typical UV (or TR) either, because there is no ergative argument (cf. 1b). The argument structure of Amis Exp verbs such as (8b) thus suggests a third function of *ma-* other than AV and UV. In the following sections, I make the case for this particular *ma-* as the anticausative marker.

3. Internally-caused vs. externally-caused eventualities

This section introduces a predicate decomposition approach to the Subj$_{Exp}$/Obj$_{Exp}$ dichotomy. The semantic motivation for the dichotomy is mainly based on the distinction of internally caused and externally caused psych states. For the sake of presentation, let me begin with the semantics of causative/inchoative alternation of dynamic verbs. Consider the following examples.

(9)   The causative/inchoative alternation in English: *broke, open, melt…*
　　a. transitive/causative: e.g. *John **broke** the window*.
　　b. intransitive/inchoative: e.g. *The window **broke***.

(10)   Verbs without causative/inchoative alternation in English: *laugh, play, speak…*
　　a. transitive: e.g. *The teacher laughed the child
　　　　　　　　　　　　(intended for 'The teacher made the child laugh.')
　　b. intransitive: e.g. *The child **laughed***.

The examples above demonstrate two verb classes in terms of their participation in the causative/inchoative alternation. Levin & Rappaport Hovav (1995) argue that non-alternating intransitive verbs (e.g. *laugh*) describe "internally-caused" eventualities, as "some property inherent to the argument of the verb is "responsible" for bringing about the eventuality" (p. 91). Externally caused verbs (e.g. *break*), on the other hand, "imply the existence of an "external cause" with immediate control over bringing about the eventuality described by the verb" (p. 92). The difference between non-alternating and alternating verbs can be captured by means of predicate decomposition, as shown below.

(11)   Lexical semantic representation of non-alternating verb *laugh* (e.g. 10)
　　laugh: [x ACT <*laugh*>]   (intransitive)

(12)   Lexical semantic representation of alternating verb *break* (e.g. 9)
　　a. break: [x CAUSE [BECOME [y <*broken*>]]]   (transitive/causative)
　　b. break: [BECOME [y <*broken*>]]   (intransitive/inchoative)

The event structure templates above are based on Rappaport Hovav & Levin's (1998) framework. The *primitive* predicates such as ACT, CAUSE, or BECOME, correspond to the generally acknowledged event types. The *constant* (e.g. <*laugh*>, <*broken*>) hosts the idiosyncratic meaning of a verb. The x and y *variables* represent arguments linked to the primitives. Consider the difference between the primitives associated with the initiator (i.e. x) of internally/externally-caused eventualities, i.e. ACT and CAUSE. The label ACT

indicates that x itself is responsible for bringing about the eventuality. In (12a), CAUSE is used to show that x and y participate in different subevents: the resultant state of y (i.e. BECOME) is triggered/caused by an external component x.

The predicate decomposition approach is useful in many facets. The theoretical constructs such as primitives, constants, and variables help establish verb classes in a systematic fashion. In addition, the semantic templates as shown in (12) represent distinct argument structures, and thus disambiguate the verbs especially in languages using the same form for both variants (e.g. *break* (tr.) and *break* (intr.)). In the next section, I show how this predicate decomposition approach can be applied to morphologically complex Amis verbs.

This section ends with extending the predicate decomposition approach to psych verbs. It has been argued that FEAR verbs and FRIGHTEN verbs contrast in a similar way as (11) and (12) do (DiDesidero 1999). They are summarized below.[5]

(13)   Lexical semantic representation of $Subj_{Exp}$ verbs: *fear, hate, dread…*
         fear: [x <*fear*> y][6]

(14)   Lexical semantic representation of Obj_{Exp} verbs: *frighten, amaze, confuse…*
         a. *frighten*: [x CAUSE [BECOME [y <*frightened*>]]] (transitive/causative)
         b. *frightened*: [BECOME [y <*frightened*>]]   (intransitive/inchoative)

The lexical representations of $Subj_{Exp}$/Obj_{Exp} verbs as shown in (13) and (14) are well-motivated: the former denote self-initiated psych states (e.g. *fear*) and the latter denote externally-triggered psych states (e.g. *frighten*). The difference between $Subj_{Exp}$ and Obj_{Exp} verbs in terms of their event structures is also reflected by their participation in middle construction. The following examples are taken from Hale & Keyser (2002:37–38).

---

[5] DiDesidero (1999) in fact proposes subcategories within Sub_{Exp} and Obj_{Exp}. For the purpose of this study, I introduce only one of the semantic representations for each class of Exp verbs.
[6] Here, I address some differences between (11) and (13) which do not influence the analogy, including the presence of ACT in (11), and the involvement of a second participant in (13). In (11), ACT is required to indicate the dynamicity of the event. For stative verbs such as 'fear', it is thus not required. In (13), the presence of y suggests the bivalent nature of 'fear'. Note that this is not always the case for psych predicates (e.g. *angry*, *happy*). Furthermore, it is observed that all Sub_{Exp} verbs, regardless of their valency, do not allow inchoative/causative alternation. Thus the semantic templates for non-alternating/alternating dynamic verbs can still be extended to Experiencer verbs.

(15)   The middle voice of Subj$_{Exp}$ verbs: ungrammatical
    a. *<u>*John's talent*</u> **envies** *easily*. (cf. *Everyone envies <u>John's talent</u>*.)
    b. *<u>*French films*</u> **love** *easily*.   (cf. *My kids love <u>French films</u>*.)

(16)   The middle voice of Obj$_{Exp}$ verbs: grammatical
    a. *<u>Politicians</u> **anger** easily*.   (cf. *The truth angers <u>politicians</u>*.)
    b. *<u>I</u> **worry** easily*.   (cf. *Economic downturns worry <u>me</u>*.)

As shown in (16), middle voice "detransitivizes" the verb, making the internal argument the subject of the sentence, and the agent (or causer) implicit. The ability for Obj$_{Exp}$ verb class to undergo such a syntactic process is crucial to this study. In the next section, I demonstrate how Amis Undergoer$_{Exp}$ verbs get detransitivized by means of the anticausative *ma-* marker.

4.   The argument structure of Amis Exp verbs: a minimalist analysis

This section explores the argument structure of Amis Experiencer verbs. I adopt the minimalist framework, as it assumes a delicate division of labor between the components of verbal predicates, and proposes structural correlations between thematic roles and their hierarchical positions in relation to the syntactic heads (Harley 2010, 2012). Section 4.1 begins with some theoretical assumptions adopted for the basic morphosyntax of Amis. The function of *mi-* and *ma-* in verbs that are not emotion-denoting will be identified first in terms of the flavors of v (Harley 2009). The investigation of Actor$_{Exp}$/Undergoer$_{Exp}$ verbs is conducted in Section 4.2. It will be shown that the so-called voice markers (e.g. *mi-/ma-*) provide different verbalizing functions based on the semantic characteristics of the root (i.e. internally caused/externally caused). Inspired by Alexiadou et al. (2006), I argue that *ma-* in Undergoer$_{Exp}$ verbs is best analyzed as anticausative, or more precisely, "anticausative-II" in terms of Schäfer's (2008) typology.

4.1   Theoretical Assumptions

The idea that VP consists of at least two projections have been commonly accepted by Chomskyan linguists. Building on the work on Marantz (1984, 1997) and Kratzer (1996), the traditional "VP" came to be understood as a cover term for a functional projection, vP (or VoiceP), where external arguments are introduced, and a lexical VP or √P, which introduces selected internal arguments (Harley 2013a). Inspired by Pylkkänen (2002:122–25), some scholars propose a further division of this functional projection, as summarized below.

(17) The distinctness of Voice and v (based on Harley 2013b)

Functions of Voice:
→ External argument introduction
→ Check a structural case on
   an internal argument

Functions of v:
→ Verbalize √
→ Causing/initiating subevent
   introduction
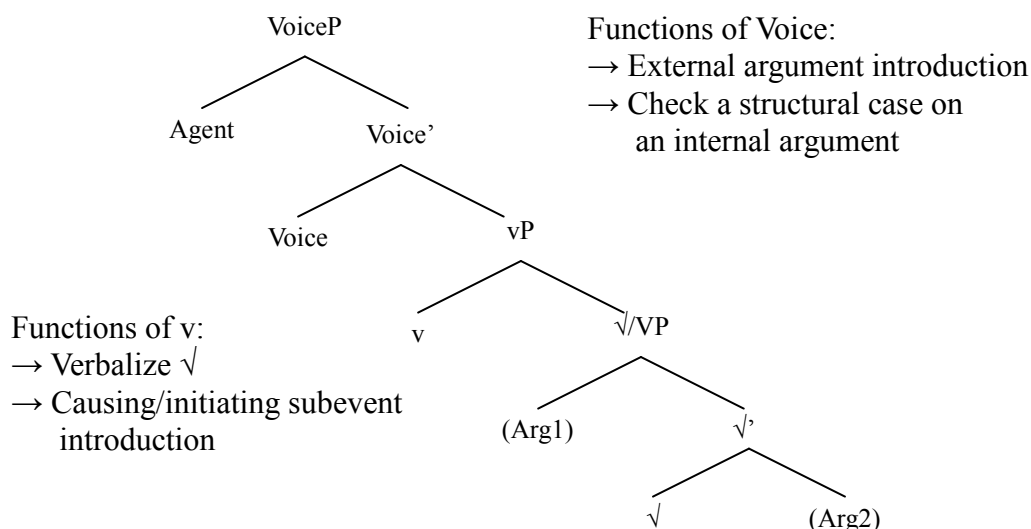
Pylkkänen (2002, 2008) proposes a "bundling parameter" to account for the fact that some languages unify all the functions above in a single projection, and others project VoiceP and vP independently. This study embraces the distinctness of Voice and v in Amis for theoretical and empirical reasons. Theoretically, the separation of Voice and v is motivated from Schäfer's (2008) typology of anticausatives. Empirically speaking, Voice and v represent the inflectional and derivational properties of Philippine-type voice markers, respectively.

For the transitive and intransitive clauses in Amis, I follow Aldridge's (2004, 2012) ergative approach, with slight modification according to the previously mentioned tripartite structure: VoiceP-vP-√. The following demonstrates the structure of an intransitive sentence such as (1a).

(18)   The structure of Amis intransitives
    a. mi-adup   k-u   tamdaw   t-u   fafuy   ((1a))
       ITR-hunt   ABS-CN   person   OBL-CN   pig
       'The person hunts pigs.'

b.
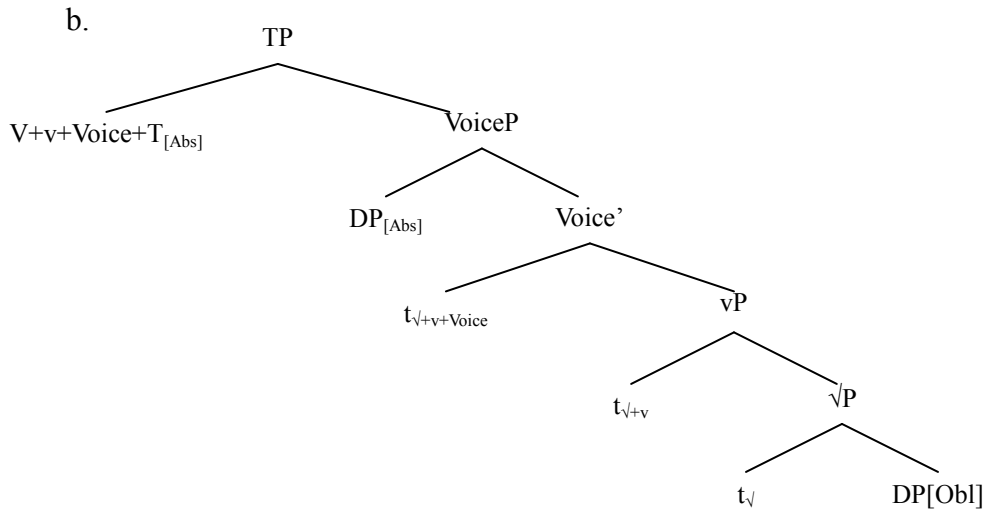
For the interest of this study, I focus on how absolute case feature is checked in intransitive clauses (including antipassives). Aldridge (2012) argues that in *v*-Type ergative languages such as Tagalog, T_{Fin} optionally carries an absolute case feature. The feature will only be checked when there is a DP with an unvalued case feature in its c-command domain (p. 6). In (18), the intransitive Voice does not have a case feature, and T must value absolutive case. Since the external argument is the first DP in T's c-command domain, this is the DP which will receive absolutive case.

In previous examples, I gloss Amis voice markers as TR/ITR to highlight the puzzle regarding the inflectional properties of *ma-*. Here, with the assumed three-layered predicate structure, I focus on the derivational (verbalizing) properties of voice marking. Harley (2009) characterizes different types of *v* in terms of feature clusters like [±dynamic], [±change of state], and [±cause], as shown below.

(19)   The flavors of v: (Harley 2009)
    a. v_{Cause}: [+dynamic], [+change of state], [+cause]
    b. v_{Become}: [+dynamic], [+change of state], [-cause]
    c. v_{Do}: [+dynamic], [-change of state], [-cause]
    d. v_{Be}: [-dynamic], [-change of state], [-cause]

In Section 3, I refer to Rappaport Hovav & Levin's lexical semantic templates for the distinction between internally and externally caused eventualities. The idea of primitive predicates is well-preserved in Harley's model (i.e. v_{Cause} = CAUSE, v_{Become} = BECOME, v_{Do} = ACT, and v_{Be} = STATE). The derivational properties of Amis voice markers have been discussed in Wu's (2007a) RRG analysis. The following is a summary of the verbalizing functions of *mi-* and *ma-*, with alternative glossing (cf. TR/ITR in prevous examples) to highlight their distinct functions.

(20)   The functions of *mi-* verbalizer
    a. mi-adup   k-u   tamdaw   t-u   fafuy
       DO-hunt   ABS-CN   person   OBL-CN   pig
       'The person hunts pigs.'

    b. mi-tuniq[7]   k-u   kuwaq   t-u   ti'ti'
       CAUS-soft   ABS-CN   papaya   OBL-CN   meat
       'The papaya will tenderize meat.' (Wu 2006:171)

(21)   The functions of *ma-* verbalizer
    a. ma-adup   n-u   tamdaw   k-u   fafuy
       DO-hunt   ERG-CN   person   ABS-CN   pig
       'The person hunted the pig.'

    b. ma-su'su'   k-u-ra   wawa
       BE-fat   ABS-CN-that   child
       'That child is fat.'

    c. ma-patay   k-u   wacu   nira
       BECOME-dead   ABS-CN   dog   3SG.GEN
       'His dog is/becomes dead.'

(20) and (21) suggest that the functions of *mi-* or *ma-* differ according to the

---

[7] The identification of *mi-* as 'CAUSE' may seem less convincing to those possessing the knowledge about the causative marker *pa-* (or its equivalent) in the Austronesian language family. While *pa-* is commonly observed in Amis causative constructions, I argue that the difference between *mi-* and *pa-* is lexically and syntactically driven. The usage of *mi-* as CAUSE, to my knowledge, is only restricted to state-denoting roots (e.g. 'soft' in (20), 'annoy' in Section 4.2.2). The general causative marker *pa-*, on the other hand, applies to both the root level and the verb level. The following examples show the ability of *pa-* to add another external argument (i.e. causer) to the event denoted by the root (*pi-* and *ka-* can be considered as a phonological output of the features of *v* combined with [-infinitive]).

(i) The productivity of Amis causative marker *pa-*
    a. pa-nanum   cingra   ci   aki-an   t-u-ra   sayta   (cf. *nanum* 'water')
       CAUS-water   3SG.ABS   PPN   Aki-OBL   OBL-CN-that   soda
       'He gave Aki that soda (to drink).' (Wu 2006:301)
    b. pa-pi-nanum   ci   ina   ci   mama-an   (cf. *mi-nanum* 'drink (water).')
       CAUS-PI-water   PPN   mother   PPN   father-OBL
       'Mother asked Father to drink water.' (Wu 2006:264)
    c. pa-ka-roray   ci   aki   kitanan   (cf. *ma-roray* 'get tired')
       CAUS-KA-tired   PPN   Aki   1PI.OBL
       'Aki made us tired.' (Wu 2006:268)

root it attaches to. This motivates a careful identification of the root types. For example, *mi-* is used for a self-initiating 'hunting' event (e.g. DO in 20a), or it can also derive a change-of-state verb when attached to a state root (e.g. CAUSE in 20b). The usage of *ma-* is wider: it can occur in a self-initiating event (e.g. DO in 21a), in stative verb (e.g. BE in 21b), or in a change-of-state verb (e.g. BECOME in 21c).

The discussion of *mi-* and *ma-* here excludes the case of psych verbs deliberately. In the following subsection, I show how these verbalizers contribute to the formation of Amis Exp verbs. Furthermore, I explain why *ma-* in externally caused psych verbs is best analyzed as anticausative, based on the case pattern and argument structure.

4.2 The argument structure of Amis Exp verbs

This section presents the syntactic structures of two types of Amis Exp verbs. Hereafter the terminology "Actor$_{Exp}$/Undergoer$_{Exp}$" verbs will be replaced with internally caused/externally caused psych verbs. This is based on the concern that the so-called experiencer of the former class may have two possible structural positions under the tripartite analysis (see Section 4.2.1); thus the "Actor" label is not ideal, as it does not fully address the structural correlations between thematic roles and their hierarchical positions in internally caused psych verbs.

4.2.1 Internally caused psych verbs

In Section 4.1, I have addressed the importance of identifying root categories, as they correlate with the functions of the verbalizers (i.e. voice markers). Along this line, I argue that the semantic distinction between internally caused and externally caused eventualities in Amis exists at the root level. For internally caused psych states such as 'like', the emotion is motivated by the experiencer. Consider below for a typical usage of 'like' and its syntactic representation.

(22) *ma-* marked internally caused psych verbs (e.g. 'like')
  a. ma-ulah **ci sawmah** t-u-ra wawa ((8a))
    BE-like PPN Sawmah OBL-CN-that child
    'Sawmah likes that child.'

b.

```
                        TP
                      /    \
                    T        vP
                 ma-ulah   /    \
                     DP[Abs]     v'
                     ci sawmah  /   \
                             vBe     √P
                            (ma-)   /   \
                                   √     DP[Obl]]
                                 (ulah)  tura wawa
```
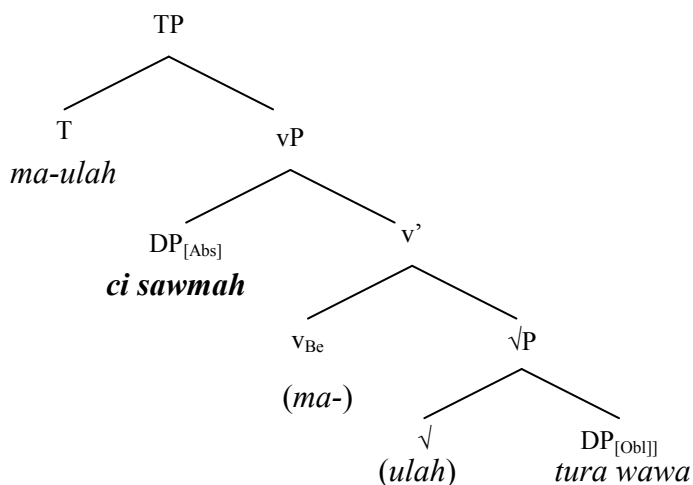
As mentioned earlier, the experiencer of self-motivated emotions such as 'like' serves as the "external argument". I argue that this experiencer is located in [Spec, v_{Be}] as the state-holder (Ramchand 2008, Rothmayr 2009). The Voice projection is absent in (22), as there is no involvement of the agent or causer. The experiencer is assigned absolute case as it is the highest DP within T's c-command domain.

As far as frequency and interpretation are concerned, *ma-* affixed internally caused psych verbs are "unmarked", as opposed to their *mi-* counterparts—native speakers only acknowledge the usage of these *mi-* verbs in special contexts. Following Didesidero (1999:168–69), I argue that some state-denoting roots also allow for the activity event structure (e.g. [x <LOVE> y]) vs. [x $ACT_{<LOVE>}$ y]). (23b) shows the syntactic representation of *mi-ulah*, with its "motional/purposive" interpretation implied by v_{Do}.

(23)   *mi-* marked internally psych verbs (e.g. 'like (purposefully)')
   a. mi-ulah   **ci   sawmah** t-u-ra   wawa   ((6a))
      DO-like   PPN Sawmah OBL-CN-that   child
      'Sawmah (will) like that child purposefully.'

b.

```
                          TP
                     ┌─────┴──────┐
                     T          VoiceP
                  mi-ulah     ┌───┴────┐
                           DP[Abs]    Voice'
                          ci sawmah ┌───┴────┐
                                  Voice      vP
                                   Ø      ┌───┴───┐
                                        v_Do      √P
                                        (mi-)  ┌───┴────┐
                                               √       DP[Obl]
                                            (ulah)   tura wawa
```
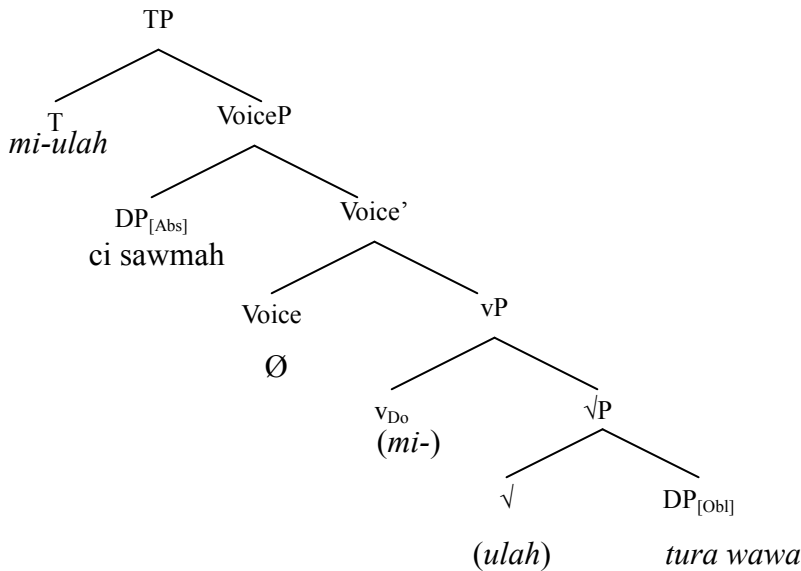
In (23b), the attachment of v_{Do} turns a state of emotion into an event. According to the distinctness of v and Voice (e.g. 17), v is responsible for the semantics of the event, but not for the introduction of its external argument. The Voice projection is thus required to license the Actor/Agent of 'liking' in its specifier position.[8] This external argument as the highest DP gets the absolutive case.

4.2.2 Externally caused psych verbs

As mentioned in Section 3, externally caused eventualities presuppose the existence of an external causer and a theme undergoing a change of state. I therefore propose the structure of *mi-'esam* 'annoy' as follows. In (24b), the change-of-state root *'esam* merges with v_{Cause} *mi-*, denoting the externally caused psych state 'annoy'. The experiencer undergoing the change of state is introduced as the complement of √P. T's absolutive case feature is checked with the causer in [Spec, Voice].

(24)   *mi-* marked externally caused psych verbs (e.g. 'annoy')
  a. mi-'esam   **k-u-ra**   **lalangaw** (t-u   tamdaw)
     CAUS-annoy   ABS-CN-that   fly   OBL-CN   person
     'That fly is annoying (people).'

---

[8] Some may question our treatment of *mi-* and *ma-* as pure verbalizers and the Voice head as zero. Alternatively, voice markers can represent the phonological realizations of the combined features of v and Voice (i.e. lexical insertion). For the sake of exposition, I enforce the separation of v and Voice in this study to pinpoint the function and hierarchical position of *ma-* in externally caused psych verbs.

b.

```
                    TP
          ┌──────────┴──────────┐
          T                   VoiceP
       mi-'esam          ┌──────┴──────────┐
                    DP[Abs]              Voice'
                   kura lalangaw    ┌──────┴──────────┐
                                  Voice              vP
                                    Ø         ┌───────┴───────┐
                                          V_Cause           √P
                                           (mi-)       ┌─────┴─────┐
                                                       √         DP[Obl]
                                                    ('esam)     tu tamdaw
```
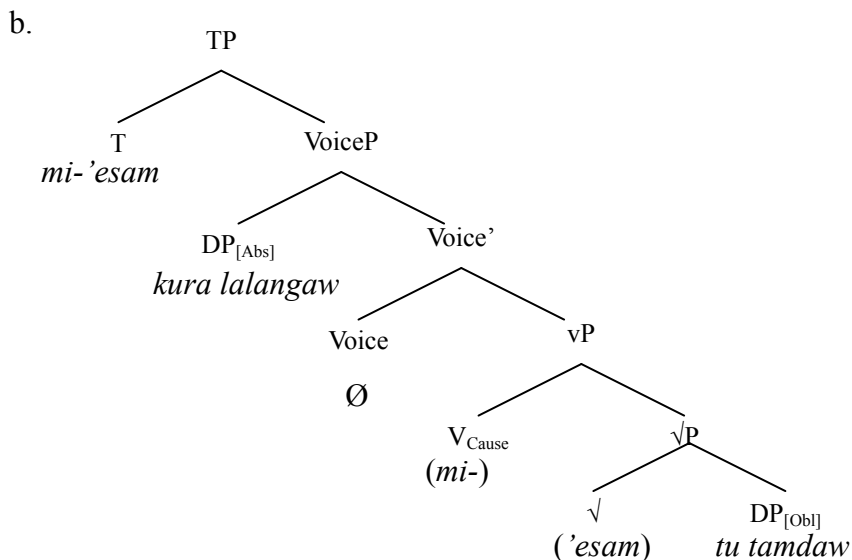
Finally I deal with *ma-* marked externally caused psych verbs, where absolutive case is found on the experiencer internal argument. Inspired by Alexiadou et al. (2006), I propose the anticausative analysis, as shown in (25b).

(25)   *ma-* marked externally caused psych verbs (e.g. 'annoyed')
   a. ma-'esam   **k-u-ra**   **tamdaw**   (t-u-ra   lalangaw)
      ACAUS-annoy   ABS-CN-that   person   OBL-CN-that   fly
      'That person is annoyed (with that fly).'
   b.

```
                    TP
          ┌──────────┴──────────┐
          T                   VoiceP
       ma-'esam         ┌──────┴──────────┐
                     Voice                vP
                     (ma-)        ┌────────┴────────┐
                               vP                   PP
                        ┌───────┴───────┐     ┌──────┴──────┐
                    V_Cause           √P     P           DP[Obl]
                     (mi-)      ┌──────┴──┐   Ø        tura lalangaw
                               √        DP[Abs]
                            ('esam)    kura tamdaw
```
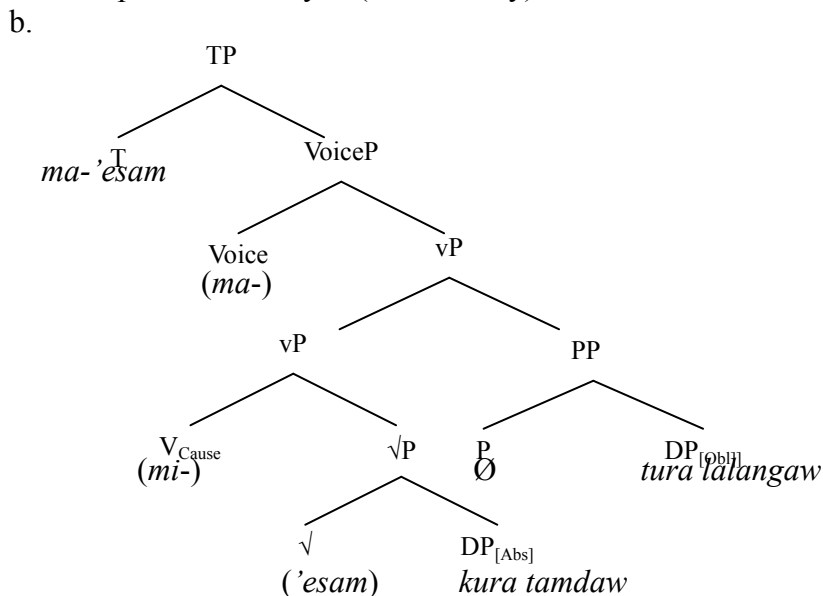
Embracing the semantics of externally caused eventualities, I argue that v_Cause

always play a role in the formation of verbs of this type. Therefore, *ma-'esam* is formed by attachment of *mi-* to the root in the first place. In addition, I propose a Voice head *ma-* with neither D-feature nor thematic feature; consequently, the causer cannot be projected in its specifier position (Schäfer 2008:176). Alternatively, the causer PP, being thematically licensed by v_{Cause}, can be made explicit by means of adjunction. The null P is responsible for assigning oblique case to its complement.[9] As a result, the experiencer internal argument is the remaining DP to receive absolutive case.

One challenge for this proposal is the surface form of the predicate, which clearly shows no trace of v_{Cause} *mi-* (compare 25a and 25b). There are some possible accounts for the lack of *mi-* in the real utterance. Theoretically, when fusion occurs by means of head-to-head movement, some languages choose to only realize one vocabulary item (Halle and Marantz 1993). This is arguably the case in Amis. Most importantly, the <u>ma</u>-<u>mi</u>-√ template, as a result of Ca reduplication, has become a strategy applied across *mi-*√ verbs for the irrealis mood. This makes the same template impossible for anticausative case (e.g. 25). Consider the sentences below.

(26)   Irrealis mood by means of Ca reduplication in Amis
    a. ma-mi-nanum   kaku,   mi-tapadang   kisu   (Wu 2006:126)
       IRR-DO-water   1SG.ABS   DO-call   2SG.ABS
       'When I was about to drink water, you called me. (So I didn't drink)

    b. ma-mi-'esam   k-u-ra   tamdaw   takuwanan,
       IRR-CAUS-annoy   ABS-CN-that   person   1SG.OBL
       mi-laliw   kaku
       DO-leave   1SG.ABS
       'That person is about to annoy me, (so) I am leaving.'
       (not 'That person is annoyed with me, (so) I am leaving.')

Before ending this section, I introduce an alternative analysis involving a simpler structure and explain why it should not be accepted. In his cross-linguistic examination, Schäfer (2008) identifies three possible structures for anticausatives, as summarized below.

(27)   The structures of anticausatives across languages (Schäfer 2008:176)
    a. anticausative-I: [Expl. [Voice$_{\{D, \emptyset\}}$ [v [Root]]]]
    b. anticausative-II: [Voice$_{\{\emptyset\}}$ [v [Root]]]
    c. anticausative-III: [v [Root]]

---

[9] The treatment of null P is based on the observation that all adjuncts in Amis, except location-like participants, are introduced to the sentences without overt prepositions. Location-like adjuncts are introduced by the preposition *i* instead.

The simplest structure possible for anticausatives is the lack of Voice projection for the predicate (e.g. 27c). The other two structures involve Voice with different features. Previously, I have propose the presence of Voice for Amis *ma-* marked externally caused psych verbs. As there is no expletive in Amis, these verbs naturally belong to anticausative-II (e.g. 27b). However, it is tempting to analyze the structure of *ma-'esam* as a case of anticausative-III, given the previously identified verbalization functions of *ma-*, v_{Become} in particular. Consider the following representations (with the experiencer included).

(28) Two possible analyses for *ma-'esam* 'annoyed'
a. anticausative-II: [Voice$_{\{Ø\}}$ (=*ma-*) [v_{Cause} (=*mi-*) [Root(=*'esam*) DP$_{exp}$]]]
b. anticausative-III: [v_{Become} (=*ma-*) [Root (=*'esam*) DP$_{exp}$]

In light of Wu's (2007b) findings on the Exp verb dichotomy, I argue that anticausative-II is a more convincing analysis than anticausative-III. As briefly stated in Section 2, one class of psych roots allows *pa-ka-* and *-en* affixations, whereas the other does not. This is demonstrated below.

(29) Amis internally/externally caused psych verbs: derivation constraints
a. pa-ka-**inal**   k-u   tafulod   aku   t-u   tao
   CAUS-KA-envious   ABS-CN   bag   1SG.GEN   OBL-CN   others
   'My bad made other people feel envious.' (Wu 2006:219)

a'. *pa-ka-**'esam**   k-u-ra   lalangaw t-u-ra   tamdaw
    CAUS-KA-annoy ABS-CN   fly   OBL-CN-that   person
    'That fly made that person feel annoyed.'

b. **ngudu**-en   aku   k-u   matu'asay
   respect-DO   1SG.ERG   ABS-CN   old.man
   'I will show respect to the old people.' (Wu 2006:223)

b'. *'**esam**-en aku   k-u   matu'asay
     annoy-DO 1SG.ERG   ABS-CN   old.man
     'I will annoy the old people.'

The representation of anticausative-II provides nice structural motivations for the derivation constraints possessed by externally caused psych roots/verbs. Given that the causer is already thematically licensed by v_{Cause}, this predicate does not allow other verbalizers (e.g. *pa-* or *-en*), which repeatedly and redundantly license the initiator (e.g. causer/agent) of the psych state.

5.   Discussion

This section discusses the implications of the establishment of the anticausative *ma-* in Amis. In the introduction section, I have presented an interesting finding about this marker's ability to license either the external argument or the internal argument, depending on the verb/root types: the former is typically found in stative verbs whereas the latter is mostly associated with dynamic verbs. This has been exemplified in *ma-fanaq* 'BE-know' and *ma-palu* 'DO-hit' in (2).[10] A reasonable question is whether one should view this as a case of homophony or polysemy. The identification of anticausative, I argue, supports the polysemy (or multifunction) of *ma-*, by which I mean that the verbalizer *ma-*, originally serving as v_{Be}, gradually developed other functions such as anticausative and UV in Amis.

The stative function of *ma-* can be traced back to the Proto-Austronesian (PAn) level (Blust 2009). Supporting evidence comes from the observation that most (if not all) Formosan languages rely on the reflex of PAn *ma- to form stative verbs. The UV function (or transitive usage) of *ma-*, on the other hand, is rarely attested in other Formosan languages (except Kavalan; see Huang & Sung 2008), but identified in some other Western-Malayo-Polynesian languages outside Taiwan (e.g. Tagalog). This suggests that the UV function of *ma-* is innovative.

To defend the polysemy view, one needs to show that the development from stative *ma-* to an undergoer-licensing *ma-* is motivated rather than arbitrary. Prior to the discovery of anticausative, such an argument is difficult to maintain. After all, the ability for the same morpheme to license an external argument with some roots and an internal argument with others creates the burden of acquisition (De Guzman 1992). However, upon scrutiny, certain similarity exists in these *ma-* verbs in terms of the hierarchical position of the absolutive argument. Based on the tripartite structure of verb phrases, it is observed that it is always the *v*P-internal arguments that are assigned absolutive case. (30a-c) are repeated from (21), and (30d) is repeated from (25a).

(30)   The "P-marking" characteristic of Amis *ma-* verbs
    a. ABS = complement of √P
        ma-adup   n-u   tamdaw   k-u   fafuy
        DO-hunt   ERG-CN person   ABS-CN pig
        'The person hunted the pig.'

---

[10] Here I focus on the derivational properties of *ma-* (e.g. v_{Be} and v_{Do}), instead of its inflectional properties (ITR/TR).

b. ABS = specifier of v

ma-su'su'	k-u-ra	wawa
BE-fat	ABS-CN-that	child

'That child is fat.'

c. ABS = specifier of v

ma-patay	k-u	wacu	nira
BECOME-dead	ABS-CN	dog	3SG.GEN

'His dog is/becomes dead.'

d. ABS = complement of √P

ma-'esam	k-u-ra	tamdaw	(t-u-ra	lalangaw)
ACAUS-annoy	ABS-CN-that	person	OBL-CN-that	fly

'That person is annoyed (with that fly).'

In light of the typology of semantic alignment (Donohue & Wichmann 2008), Tsukida (2008) demonstrates that Amis *ma-* verbs show features similar to that of 'P-marked' verbs. (p.292) (see also Jiang 2011 for a related discussion). Along this line, *ma-* stative verbs can be related with innovative *ma-* dynamic verbs: both license *v*P-internal arguments as absolutive. The idea about "stative to eventive/UV" becomes more appealing, if one acknowledges the innovative "anticausative" *ma-* as the intermediate stage of the development. This is likely when we recall the syntactic/semantic representation of anticausative *ma-* verbs. It contains a "change of state" and a "causing event", and thus shares properties from two extremes of *ma-* verbs (i.e. stative and eventive). Furthermore, anticausative verbs preserve the original P-marking characteristic for selecting the internal argument as the absolutive argument.

So far, the development of *ma-* from stative to anticausative to eventive/UV is merely my speculation, motivated by the shared grammatical properties of these *ma-* verbs. More research should be done to deal with the questions regarding how this two-step development may have been triggered. Before ending this section, I introduce my hypothesis for the development of *ma-* from anticausative to eventive/UV—which may be skeptical at first glance, considering the fact that the former is (syntactically) intransitive and the latter transitive. I argue that in the case of anticausatives, it is likely for the speakers to introduce an ergative argument to emphasize the volitionality of the causer in change-of-state contexts. For example, consider the following two case frames for the verb *ma-patay*.

(31)   Two possible case frames for *ma-* verbs

a.
ma-patay	k-u	oner	t-u	sapaiyo	n-u	'edu
MA-dead	ABS-CN	snake	OBL-CN	medicine	GEN-CN	mouse

'Snakes may die from the poison for killing mice.' (Wu 2006:347)

    b. ma-patay   (nira)   kina   tamdaw   (Fata'an Amis)
       MA-dead   3SG.ERG this.ABS person
       'This person was killed (by him/her).'
       'This person is dead (because of him/her).' (Tsukida 2008:284)

ma-patay has the intransitive case frame in (30a). It is thus an anticausative verb with the absolute theme and the oblique causer. In (30b), the "transitive" case frame is found for the "same verb", and the reading is ambiguous: the initiator of the causing event can be an agent (i.e. volitional) or a causer. I hypothesize that sentences like (30b) are the origin of *ma-*'s eventive/UV function. The anticausative *ma-* first applied to change-of-state roots, and was reanalyzed as UV for this innovated case frame. The innovative function was later extended to event-denoting roots such as 'hunt'. This hypothesis, along with other concerns for the development of *ma-*, awaits further study.

6.   Conclusion

In this paper, I have identified the anticausative function of Amis verbal morphology *ma-*, based on a careful examination of the argument structure of Experiencer verbs. A predicate decomposition approach is adopted to distinguish internally caused psych verbs and externally caused pysch verbs. Particularly, the former involves *mi-* as v_{Do} and *ma-* as v_{Be}, whereas the latter involves *mi-* as v_{Cause} and *ma-* as anticausative. By proposing the distinctness of Voice and v in verbs phrases, I have also analyzed *ma-* marked verbs as having the structure of anticausative-II in terms of Schafer's (2008) typology. The discovery of the anticausative function serves as a crucial link between the stative function and the eventive/UV function; this suggests a possible development path of *ma-*, which separates Amis from other Formosan languages.

References

Aldridge, Edith, 2004. Ergativity and verb-initial word order in Austronesian languages. Doctoral dissertation, Cornell University.

Aldridge, Edith. 2012. Antipassive and ergativity in Tagalog. *Lingua* 122:192–203.

Alexiadou, Artemis, Elena Anagnostopoulou, and Florian Schäfer. 2006. The properties of anticausatives crosslinguistically. *Phases of Interpretation,* ed. Mara Frascarelli, 187–211. Berlin: Mouton.

Belletti, Adriana, and Luigi Rizzi. 1988. Psych verbs and θ-theory. *Natural Language and Linguistic Theory* 6:291–352.

Blust, Robert. 2009. *The Austronesian languages*. Canberra: Pacific Linguistics.

Chang, Henry Yung-li. 2011. Transitivity, ergativity, and status of O in Tsou. *Language and Cognition: Festschrift in Honor of James H-Y. Tai on His 70th Birthday*, ed. Jung-hsing Chang & Jenney Y.-C. Kuo, 277–308. Taipei: Crane.

de Guzman, Videa. 1992. Acquisition of voice affixes in psychological verbs in Tagalog. Paper presented at the 3rd International Symposium on Language and Linguistics, Pan-Asiatic Linguistics, Bankok.

DiDesidero, Linda Barbara. 1999. Psych verbs: Acquisition, lexical Semantics, and event Structure. Doctoral dissertation, Northwestern University.

Donohue, Mark, and Sren Wichmann. 2008. *The typology of semantic alignment*. New York: Oxford University Press.

Dowty, David. 1991. Thematic proto roles and argument selection. *Language* 67:547–619.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge: MIT Press.

Hale, Ken, and Samuel Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge: MIT Press.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. *The view from building 20: Essays in Linguistics in honor of Sylvain Bromberger*, ed. Ken Hale and Samuel Jay Keyser, 111–76. Cambridge: MIT Press.

Harley, Heidi. 2009. The morphology of nominalizations and the syntax of vP. *Quantification, definiteness and nominalization*, eds. Monika Rathert and Anastasia Giannadikou, 129–44. Oxford: Oxford University Press.

Harley, Heidi. 2010. A minimalist approach to argument structure. *The Oxford handbook of linguistic minimalism*, ed. Cedric Boeckx, 426–47. Oxford: Oxford University Press.

Harley, Heidi. 2013a. External arguments and the mirror principle: On the distinctness of Voice and v. *Lingua* 125:34–57.

Harley, Heidi. 2013b. The 'bundling' hypothesis and the disparate functions of little v. Paper presented at Little v Workshop, Leiden University.

Huang, Shuping, and Li-May Sung. 2008. The undergoer focus *ma-* in Kavalan. *Oceanic Linguistics* 47:159–84.

Jiang, Hao-wen. 2011. Middle voice in Formosan languages. Ms., Rice University.

Kratzer, Angelika. 1996. Severing the external argument from its verb. *Phrase structure and the lexicon*, eds. Johan Rooryck and Laurie Zaring, 109–37. Dordrecht: Springer.

Landau, Idan. 2010. *The locative syntax of experiencers*. Cambridge: MIT Press.

Levin, Beth, and Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge: MIT Press.

Levin, Beth, and Jason Grafmiller. 2013. Do you always fear what frightens you? *From quirky case to representing space: Papers in honor of Annie Zaenen*, eds. Tracy Holloway King and Valeria de Paiva, 21–32.

Stanford: CSLI On-line Publications.

Marantz, Alec. 1984. *On the nature of grammatical relations*. Cambridge: MIT Press

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4: 201–25.

Pesetsky, David. 1995. *Zero syntax: Experiencers and cascades*. Cambridge: MIT Press.

Pylkkänen, Liina. 2002. Introducing arguments. doctoral dissertation, MIT.

Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge: MIT Press.

Ramchand, Gillian. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge: Cambridge University Press.

Rappaport Hovav, Malka, and Beth Levin. 1998. Building verb meanings. *The projection of arguments*, eds. Miriam Butt and Wilhelm Geuder, 97–134. Stanford: CSLI Publications.

Ross, Malcolm, and Stacy Fang-ching Teng. 2005. Formosan languages and linguistic typology. *Language and Linguistics* 5:433–90.

Rothmayr, Antonia. 2009. *The structure of stative verbs*. Amsterdam: John Benjamins.

Schäfer, Florian. 2008. *The syntax of (anti-)causatives: External arguments in change-of-state contexts*. Amsterdam: John Benjamins.

Tsukida, Naomi. 2008. Verb classification in Amis. *The typology of semantic alignment*, ed. Mark Donohue and Søren Wichmann, 277–296. New York: Oxford University Press.

Wu, Joy Jin-lan. 2006. Verb classification, case marking, and grammatical relations in Amis. Doctoral dissertation, State University of New York at Buffalo.

Wu, Joy Jin-lan. 2007a. "Voice" markers in Amis: A role and reference grammar analysis. *Language and Linguistics* 8:95–142.

Wu, Joy Jin-lan. 2007b. Experiencer types and psych predicates in Amis. Paper presented at The 14th Annual Meeting of Austronesian Formal Linguistics Association. Montreal: McGill University.

NASAL ASSIMILATION IN JAKARTA INDONESIAN[*]

Ferdinan Okki Kurniawan
Cornell University
fok3@cornell.edu

This paper examines nasal assimilation in Jakarta Indonesian (JI), a colloquial variety of Indonesian spoken in Jakarta, Indonesia. The paper makes a twofold contribution. First, rather than relying on impressionistic observation, data in this study are drawn from a naturalistic spoken corpus (Gil and Tadmor 2014) and a production task. Secondly, using these two sources of data, this investigation finds that the speakers of JI produce variation in the nasal assimilation at prefix-root boundary. Zuraw (2010) proposed that the variation in Tagalog nasal substitution is an evidence of lexical variation. In lexical variation, each word is considered to have just one form and it applies across lexicon.This study demonstrates that the nasal assimilation in JI exhibits different types of variation from Tagalog. The patterns of variation in nasal assimilation in JI does not show any evidence of lexical variation. Rather, the variation in JI is conditioned by inter- and intra-speaker differences.

1. Introduction

Indonesian, a western Austronesian language of Indonesia which in its standard form serves as the national language of Indonesia, has a widely used verbal prefix /məN-/ which alternates in its shape at prefix-root boundary. The nasal in coda position of the prefix assimilates to a root-initial obstruent. This phonological process is commonly known as nasal assimilation. The current literature (Pater 1999, 2001) investigates the formal driving force of nasal assimilation in Indonesian within the Optimality Theory framework (OT; Prince and Smolensky 1993, 2004). This pattern was first described systematically by Lapoliwa (1981) within a generative framework. All previous studies were devoted to the standard variety of Indonesian spoken in formal contexts (standard Indonesian (SI)). Less attention, however, has been given to the pattern of nasal assimilation in a more colloquial variety of Indonesian as a casual everyday language spoken in Jakarta (Jakarta Indonesian (JI)) with the cognate prefix /N-/.

Interestingly, nasal assimilation in JI exhibits variation beyond the phonological conditioning environment. The variation found in JI occurs when

the nasal prefix (henceforth N-prefix) patterns with root-initial voiced obstruents [b-, d-, g-, and d͡ʒ-]. The variation is exemplified in (1):

(1)   /N+bəli/   ŋəbəli ~ mbəli   'to buy' (JI)

This study supports the impressionistic observation that the speakers of JI produce these two different representations in (1).

   In order to describe the patterns of variation of nasal assimilation in JI, this study examines the variation which occurs in a naturalistic spoken corpus of JI (Gil and Tadmor 2014). A production task was also conducted with JI native speakers to observe whether or not the results from this task mirror the results from the spoken corpus. The corpus shows variability in the utterances produced by the speakers. This study investigates the sources of this variability, whether the variability is due to lexical variation, following Zuraw's (2010) analysis on Tagalog, a related Austronesian language—or as a consequence of inter- or intra-speaker differences.

   In the rest of this introduction, I provide a brief overview of the relatonship between SI, JI and Betawi Malay before presenting the pattern of N-prefixation in JI. I then provide a brief comparison with Tagalog before summarizing the goals of this paper.

1.1.   An overview of Standard Indonesian, Jakarta Indonesian and Betawi Malay

As mentioned above, SI has served as the formal and national language of Indonesia. One increasing and influencing colloquial variety in and around Jakarta is JI. This also serves as the basis of the colloquial variety in urban areas across Indonesia. Most of the previous studies of nasal assimilation in Indonesian focus on the standard variety of Indonesian. SI is not spoken on daily basis and only spoken in formal situations. SI is taught at formal school and children have limited or no exposure to this standard variety until they start to go to school (Sneddon 2006). As vernacular spoken at home, JI is acquired by children in Jakarta as their first language.

   As a vernacular, JI emerged since World War II (Wallace 1976), followed by huge influx of migrants into the capital city. The second generation of these migrant families have been forming a new linguistic variety called Jakarta Indonesian. Before the emergence of JI, there was an indigineous variety of Malay, namely Betawi Malay (henceforth BM) which emerged in around the 17th-18th centuries in Jakarta. The lack of understanding of JI and BM often causes scholars to consider them under the same variety. BM should not be confused with JI.[1] Nowadays, BM is spoken by a small minority group in

---

[1] Wallace (1976) termed these JI speakers as Modern Jakarta Malay speakers, while the BM speakers in this study were categorized under his term as Traditional Jakarta Malay speakers.

Jakarta while JI is spoken widely by more educated speakers associated with higher socio-economic status in Jakarta (Sneddon 2006). Sneddon reports that JI is the variety spoken by educated people in Jakarta in informal situation, while BM is the vernacular spoken by inhabitants of 'old kampungs' in Jakarta, limited to Betawi communities.

Both BM and JI form a dialect continuum with SI. BM, JI and SI are different in some important respects in terms of their phonology but mutually intelligible. Differences in the patterns of nasal assimilation show this evidence. This paper, however, does not aim to address the socio-phonological differences between JI and BM. Rather, this investigation uses these two socio-ethnic groups to map the variations of the nasal assimilation. Let us now turn to description of N-prefix which is conditioned by root-initial consonants in JI.

1.2.   The N-prefix in Jakarta Indonesian

The N-prefix in JI is cognate with N-prefix *məN-* in SI.[2] The description below is based on my observations and coincides closely with the patterns described by Ikranagara (1980), Muhadjir (1981) and Sneddon (2006). I use the placeless nasal symbolized with *N-* as the underlying representation of the N-prefix. The pattern of the N-prefix in JI is presented in this section. Firstly, the roots that begin with liquids, glides and vowels in JI are presented in (2).

(2)   Root-initial sonorant consonants and glides:

a.	/N+lamar/	ŋəlamar	'to propose'
b.	/N+rusak/	ŋərusak	'to destroy'
c.	/N+jakin+in/	ŋəjakinin[3]	'to believe'
d.	/N+wabah/	ŋəwabah	'to be epidemic'
e.	/N+harus+in/	ŋəharusin	'to require'
f.	/N+aŋkat/	ŋaŋkat	'to lift'
g.	/N+obat+in/	ŋobatin	'to medicate'
h.	/N+ekor/	ŋekɔr	'to follow'
i.	/N+isi/	ŋisi	'to fill'

Schwa epenthesis occurs with root-initial liquids, glides and glottal fricative as in (2a-e) while a velar nasal is realized with root-initial vowels as in (2f-i). The next set of data in (3) illustrates the N-prefix which is conditioned by root-initial voiceless consonants.

(3)   Root-initial voiceless consonants:

---

[2] For further description of SI, see Lapoliwa (1981).
[3] [j] is orthographically written as <y>.

a. /N+pilih/   milih   'to choose, to vote'
b. /N+tulis/   nulis   'to write'
c. /N+kasih/   ŋasih   'to give'
d. /N+sapu/   ɲapu   'to sweep'
e. /N+t͡ɕari/   ɲari   'to seek'

The pattern in (3) where by an initial voiceless stop is replaced by a homorganic nasal is often referred to as nasal substitution.[4]

(4)   Root-initial nasals:

a. /N+makan/   makan   'to eat'
b. /N+nilai/   nilai   'to grade'
c. /N+ɲaɲi/   ɲaɲi   'to sing'
d. /N+ŋaŋgur/   ŋaŋgur   'to do nothing'

The root-initial nasals in (4a-d) show similar process as (3). This could be intrepreted as nasal substitutions or deletion.

       For root-initial voiced obstruents there is a pattern of variation that is displayed in (5a-d) below.

(5)   Root-initial voiced obstruents:

a.   /N+bəli/:   i. mbəli   'to buy'
                  ii. ŋəbəli
b. /N+dapət/:   i. ndapət   'to get'
                  ii. ŋədapət
c. /N+d͡ʑawab/:   i. ɲd͡ʑawap   'to answer'
                  ii. ŋəd͡ʑawap
d. /N+guntiŋ/:   i. ŋguntiŋ   'to cut with scissors'
                  ii. ŋəguntiŋ

       The form in (5a-d. i) are what would be expected and are used by BM and some JI speakers. However, the form in (5a-d. ii) with schwa epenthesis‒ similar to the forms seen for liquids and glides‒are also observed for some speakers.[5] One of the key questions to be addressed is what conditions this variation and why.

---

[4] Following serial rule ordering approach, Ikranagara (1980) proposed that the underlying nasal assimilates to the root-initial voiceless consonants which then forms homorganic cluster. This process is then followed by deletion of the initial consonant. In the more recent studies, Pater (2001) and Zuraw (2010) termed these two phonological processes as nasal substitution.

[5] Ikranagara (1980) did not mention such variation in BM while Muhadjir (1981) mentioned such variation.

It should be noted that there are also bare verb forms. These and N-prefix forms may occur in the same and different syntactic position in a sentence, see Tjung (2006) and Chung (2008) for further details. Since this variation is due to morphosyntactic conditioning, this study does not include the bare verb as a variant of N-prefix forms.

Before looking more closely at the patterns of variation, it is useful to look at a similar pattern observed in Tagalog, a western Austronesian language, as discussed by Zuraw (2010).

1.3. Nasal Substitution in Tagalog

In her study, Zuraw proposed a model of lexical variation to account for observed patterns in Tagalog nasal substitution. She takes lexical variation to be different from free variation. In free variation, the same speaker can produce two different forms of pronunciation. Free variation is a results from stylistic variation affected by social factors such as degree of formality, dialect, etc. In lexical variation each word is considered to have just one form and it applies across the lexicon, crucially the choice of form does not follow from frequency so must be part of each lexical entry. For example, */paŋ + poʔók/* in Tagalog is realized as *pam-poʔók 'local'* and */paŋ-RED[6] + pighatiʔ/* is realized as *pa-mi-mighatiʔ 'being in grief'*.[7] *p* in *poʔók* is resistant to nasal substitution. Instead, it forms homorganic clusters as in *pam-poʔók*, while in another lexical item such as in *pighatíʔ*, the coda nasal is substituted becoming *pa-mi-mighatíʔ*.

Based on dictionary and corpus data of written language and an experiment, Zuraw argued that the pattern of nasal substitution in Tagalog above is accounted for as lexical variation which its case is variable on a word-by-word basis.

The corpus and dictionary investigation conducted by Zuraw show interesting evidence on the lexical variation in nasal substitution in Tagalog. The data show that there is a general lexical trend in the behaviour of the nasal coda in the N-prefix.

As reported in (5), JI also shows variation in the shape of the N-prefix with root-initial voiced obstruents. This current study focuses the investigation in these patterns of variation. This study closely examines whether or not the variation observed in JI show similar pattern as the variation in Tagalog.

1.4. Goals of the present study

Currently, JI is becoming the most prestigious informal variety, not only in Jakarta but also throughout Indonesia. However, compared to SI, JI is in fact

---

[6] RED: reduplication.
[7] The pattern is generated to *t, k, s, ʔ, d,* and *g-* root initials.

barely studied at all. Therefore, this paper aims to shed a light about a phonological aspect of JI, namely the variation pattern of nasal assimilation. Rather than relying on impressionistic observation, data in this study are drawn from naturalistic spoken corpus in informal settings (Gil and Tadmor 2014). It is important to see how this variation is actually produced spontaneously by the native speakers in naturalisitic data. To the best of my knowledge, none of the prior studies about nasal assimilation in Indonesian used a large data set from naturalistic conversation. The use of corpus could help us to verify impressionistic observation and understand more about the variation pattern of nasal assimilation. The investigation in JI corpus aims to seek evidence whether the two different forms occur in (5) are due to inter-speaker variation, i.e. dialect or due to intra-speaker variation. This investigation is also intended to see whether or not the variation pattern in JI exhibits a similar pattern as Tagalog lexical variation studied by Zuraw (2010). The results from JI corpus is presented in the next section.

2.   Results from JI corpus

In this section, the results from the corpus are presented in two main parts: the results that show inter-speaker variation and the results that shows intra-speaker variation. Let us now begin with a brief description about the corpus.

2.1.   About the corpus

The corpus used in this study is Betawi-Jakarta corpus (Gil and Tadmor 2014) with data collected from 2004-2012. There are a total of 47,802 utterances which were produced by adult native speakers that have been coded in computerized database. The analysis includes only the speakers who produced relatively high tokens of N-prefix forms and could clearly be identified as BM or JI speakers in the corpus metadata.

Although the corpus is large in size, the results that show N-prefix forms are quite limited. This is most probably because the use of bare verb and passive form indicated with *di-* prefix are more productive than the use of active voice indicated with N-prefix. However, although the N-prefix forms are quite limited, they still show us interesting distributional pattern of variation.

2.2.   Inter-speaker (dialect) differences

As mentioned in 2.1., there are two main groups of speakers in this study. The first one is the speakers of Betawi ethnicity. The second one is those who have ethnic background other than Betawi but were born and grew up in Jakarta and their parents are not of Betawi ethnicity. These speakers are considered as JI native speakers.

The investigation will be limited to the distribution of variation that are only spoken by these two groups of speakers. This section is intended to see whether or not the variation is due to sociolinguistic background of the speakers (inter-speaker differences) especially as determined by dialect background. I conducted a corpus search based on the tokens of both forms. This is done to observe whether one group of speakers prefers to have the homorganic clusters [mb-, nd-, ŋg-, ɲd͡ʑ-] or prefer to have schwa epenthesis forms [ŋəb-, ŋəg-, ŋəd-, ŋəd͡ʑ-].

BM speakers that were included in (6) are marked as 'Betawi' ethnicity in the Betawi-Jakarta corpus metadata. There are a total of 43 speakers in the corpus that are marked with Betawi ethnicity. The search was done on these 43 speakers. The results of the retrieval, however, do not come from all of these 43 speakers. The results retrieved are only from the speakers who produced the relevant tokens. The results in (6) below present the total tokens produced by BM speakers by place of articulation.

(6) Total tokens by BM speakers

	mb-	ŋəb-	nd-	ŋəd-	ɲd͡ʑ-	ŋəd͡ʑ-	ŋg-	ŋəg-
total tokens	39	15	4	11	21	17	21	20
number of speakers	15	6	4	5	14	8	11	12
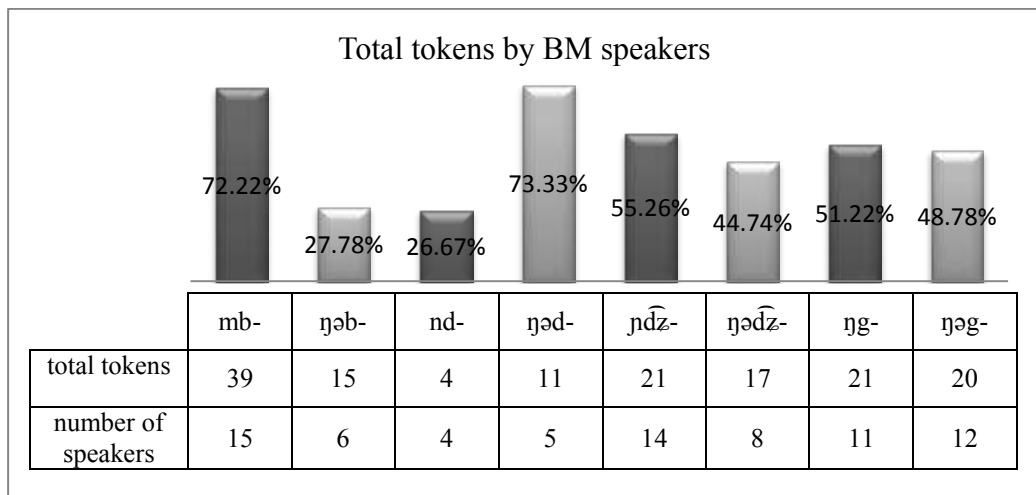
The total number of tokens in (6) for each segments are counted from the tokens produced by BM speakers. The variation of N-prefix and the root-initial consonants are indicated in the first row of the table. The percentage in the bars tells what percentage of total tokens produced in the corpus for each variation. The black bar indicates percentage of the homorganic clusters and the grey bar indicates the percentage of the epenthetic schwa form (totalling 100%). The total number of tokens are indicated in the second row. The third row table indicates total number of speakers who produce the total number of tokens. For example, the segment [ŋəb-] is produced 15 times in the corpus by 6 speakers while the segment [mb-] is produced 39 times by 15 speakers. One speaker may produce the form more than one time. From this data, we can see that the

homorganic consonant clusters have higher tokens than the epenthesis schwa forms for [mb-, ŋg- and ɲd͡ʑ-], but for alveolars [nd-] is lower than [ŋəd-].

Let us know consider the results from JI speakers shown in (7). There are 6 speakers found in the corpus that can definitely be identified as JI speakers. As mentioned before, JI speakers are those who have ethnic background other than Betawi but were born and grew up in Jakarta and their parents are not of Betawi ethnicity. The results of the retrieval, however, do not come from all of these 6 speakers. The results retrieved are only from the speakers who produced relevant tokens.

The next results in (7) show the total tokens produced by JI speakers.

(7) Total tokens by JI speakers

	mb-	ŋəb-	nd-	ŋəd-	ɲd͡ʑ-	ŋəd͡ʑ-	ŋg-	ŋəg-
total tokens	3	19	2	5	2	2	2	4
number of speakers	3	4	2	2	1	2	2	2
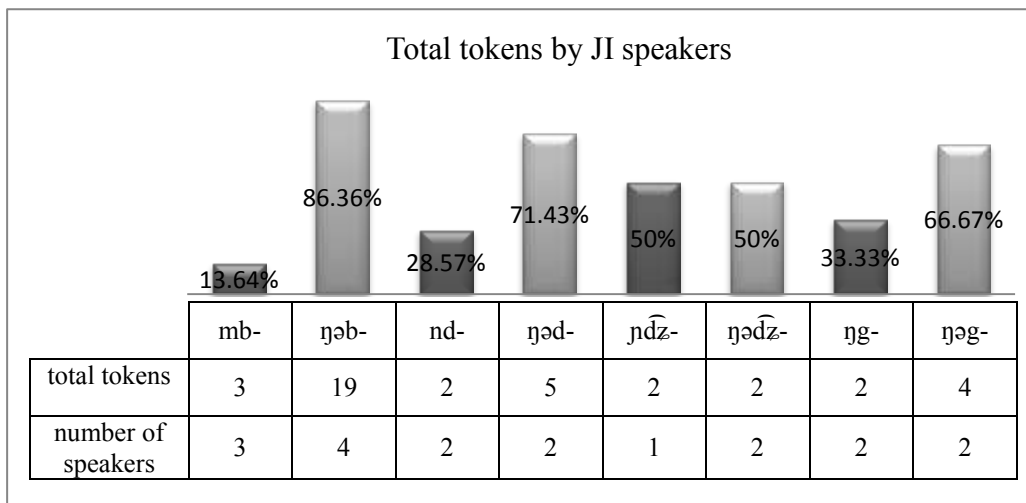
JI data in (7) show that the epenthetic schwa in [ŋəb-, ŋəd-], and [ŋəg-] result in higher tokens than homorganic clusters [mb-, nd-, ŋg-], while [ŋəd͡ʑ-] and [ɲd͡ʑ-] show the same number of tokens. The table below in (8) shows the percentage of the tokens of homorganic clusters. The table is organized based on place of articulation. The percentage that are highlighted in grey are from the tokens which have less than 10 tokens.

(8) Percentage of nasal assimilation (homorganic clusters) across place of articulation

Labial		Alveolar		Palatal		Velar	
BM	JI	BM	JI	BM	JI	BM	JI
72,22%	13,64%	26,67%	28,57%	55,26%	50%	51,22%	33,33%

From the description in the table above, we can see a big difference of percentage in the labials and velars. There is no significant difference in

alveolars and palatals. Thus, we can see that BM and JI pattern differently in the labial and velar sounds. The dialectal background of the speakers indeed plays an important role for the occurrence of variation especially for labial and velar sounds. JI speakers show preference to utter form with schwa epenthesis over the homorganic clusters, while BM speakers produce the homorganic clusters more than the schwa epenthesis, especially for [b]-initial root.

Unlike BM, data for JI are limited in the corpus. Except for epenthetic schwa form in *ŋəb-,* the data for both homorganic clusters and epenthetic schwa forms show less than 5 tokens. This might be caused by low number of JI speakers that were involved in the project compared to BM speakers. The next subsection will demonstrate the results from intra-speaker differences.

2.3.   Intra-speaker differences

The corpus exhibits distribution of N-prefix variation within the same lexical items. One lexical item might be produced in two forms by the same individual speaker. The results in (9) below show the intra-speaker variation does occur in the corpus.

(9) Distribution of N-prefix within the same lexical items

BM Speakers	Underlying representation	Homorganic clusters	Epenthetic schwa
DADBTW	/N+gajəm/ 'to eat'	ŋgajəm (1)	ŋəgajəm (1)
AFRBTJ	/N+daptar/ 'to register'	ndaptar (1)	ŋədaptar (1)
SALBTW	/N+gəletak/ 'to lie down'	ŋgəletak (1)	ŋəgəletak (1)
MLYBTJ	/N+batɕa/ 'to read'	mbatɕa (2)	ŋəbatɕa (2)
SIRBTJ	/N+ d͡ʑual/ 'to sell'	ɲd͡ʑual (2)	ŋəd͡ʑual (1)
JI Speakers	Underlying representation	Homorganic clusters	Epenthetic schwa
EXPOKK	/N+dəŋər+IN/ 'to listen'	ndəŋərin (1)	ŋədəŋərin (1)

We can see from the results in (9) above that indeed the same speaker sometimes produce both forms even for the same lexical item. The number in the parenthesis indicates the numbers of tokens. This results are different from what Zuraw (2010) found based on her study in the written and dictionary corpus of Tagalog. She considered that the general lexical trend from the Tagalog written corpus is that the lexical items are already lexicalized or listed as a single lexical item in the Tagalog speakers' mind and while the same initial sound might show variation, there is no variation between speakers or within speaker for particular lexical items. Here, intra-speaker variation occurs in Betawi-Jakarta corpus where both forms can be uttered in spontaneous speech by the same speaker.

It should also be noted that the intra-speaker variation in the data above does not seem to be caused by different situations. All the speakers uttered them when they interacted with other speakers in informal settings. In summary so far, the variation of nasal assimilation is not only observed between dialects, but also occurs as intra-speaker variation.

To look more systematically at both intra- and inter-speaker differences with more reliable data, I conducted production task that will be presented in section 3.

3.   Evidence from production task in JI: preliminary results

This production task was conducted to test whether the results from the production task mirror the results from the naturalistic corpus. The proposed hypothesis is JI speakers produced epenthetic schwa forms more frequent than the homorganic cluster forms. We can formulate this into H_1: ^{n}epenthetic schwa > ^{n}homorganic clusters. The null hypothesis for this study would be H_1 : ^{n}epenthetic schwa = ^{n}homorganic clusters. This section will present the test items, participants and location, and the production task results.

The production task was conducted with only JI speakers and did not examine the inter-speaker (dialectal) variation since the data collection was limited to a more homogenous group of Indonesian speakers in Ithaca, New York, USA. At the present time, there is no BM speakers reside in Ithaca. As a further goal, this production task could be extended to examine the inter-speaker (dialectal) variation more closely.

3.1.   Test Sentences and Participants

I will briefly explain the methodology used in this production task. This includes test sentences, participants and location.

3.1.1. Test Items

The test items in the production task are designed as follows. 48 words which begin with [b-, d-, d͡z-], and [g-] are chosen for the test items. If these words are prefixed with N-prefix, they are predicted to surface as either homorganic cluster forms [mb-, nd-, ɲd͡z-] and [ŋg-], or epenthetic schwa forms [ŋəb-, ŋəd-, ŋəd͡z-] and [ŋəg-].

Each word is embedded in two different sentences. Thus, there are a total of 96 test sentences recorded by male speakers. The participants listen to the test sentences in passive voice construction and afterward they produce active sentences item by item. In Indonesian, passive voice is indicated by *di*-verbal prefix as we can see in (10). They are asked to produce the active voice which is indicated by the N-prefix. The example below is one of the test sentences in the production task:

(10)   Subject hear:

uang   itu   udah   *dibalikin*   Toni
money that   already   Pass-return-Ben/Caus[8]   Toni
'that money was returned by Toni'

Expected response:

Toni udah *mbalikin*/*ngebalikin* uang itu.
'Toni has returned the money.'

     The sentences and their topics are composed in a colloquial style and everyday situations to avoid participants producing SI form. Since the participants are not expected to respond in SI, then the JI benefactive/causative marker *–in* is used. SI has causative/benefactive marker *–i* and *–kan*.

     The order of the test sentences is randomized. Distracters which consist of words that begin with [p- , t-] and [k-] are placed in every 6 test items to avoid subjects produce bias response with only to one variant..

3.1.2. Participants

There are a total of 8 subjects, 3 male and 5 female participants. They are refrerred to in the results as S1, S2, S3, S4, S5, S6, S7 and S8. They are Cornell University graduate students, their spouses or faculty. The data collection was done at their home or office in Ithaca, New York. Their voices are recorded using Edirol by Roland type R-09HR, 24 bit 96KHZ Wave/MP3 recorder. The age range of participants is between 25-45 years old. All of them are educated speakers of Indonesian.

3.2.   Results

Similar to the results from corpus, the results from this production task are reported in two parts: inter-speaker differences and intra-speaker differences.

3.2.1. Inter-speaker differences

The followings are the production results across 8 speakers:

---

[8] Pass: passive, Ben: benefactive, Caus: causative.

(11)   Production Task in JI

Production Task Result across 8 JI speakers
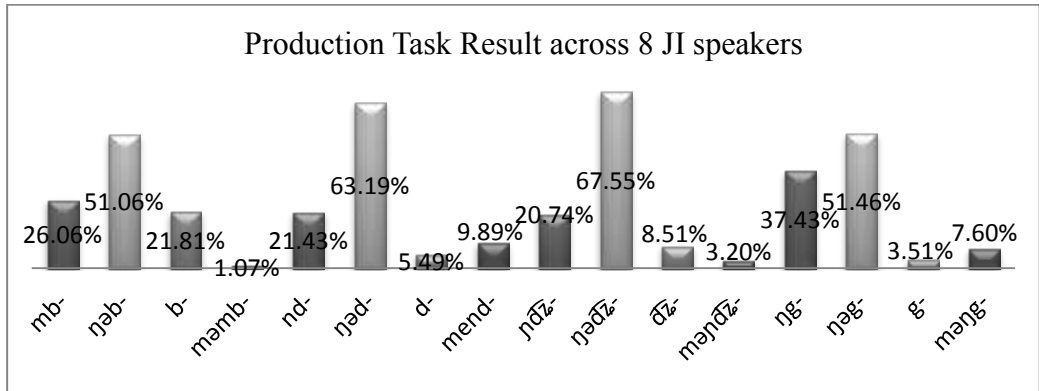
Right away, we can see that the epenthetic schwa forms are produced in greater numbers than the homorganic clusters, corresponding to the naturalistic data presented in the previous section but *not to the exclusion* of the homorganic forms. The percentage in (10) are obtained from the production of 96 test items for each of stop-initial consonant. For example, there are 96 test items (100%) for b-initial root. 26.06 % of the total test items are produced with homorganic clusters form [mb-], 51.06% are produced with epenthetic schwa form [ŋəb-], 21,81% are produced with bare verb form [b-], and 1.07% are produced with SI form [məmb-]. Since the task to produce either homorganic clusters or epenthetic schwa forms are controlled in the instruction, pre-training, training and test sentences, the bare verb and SI forms *məN-* are unexpected. Most of the results of the bare verb and SI forms are below 10%, unless for bare verb forms [b-] that reach 21.81%. This high percentage of bare verb production is produced by one speaker (S1). S1 has the same sociolinguistic background as the other speakers involved in this production task. However, it seems that he has different pattern of distribution from other 7 speakers in bilabial roots. It is still unclear what caused this difference. Therefore, I leave S1 out from the analysis of the results.

Now we turn to the core results. The distribution of the homorganic cluster, epenthetic schwa, bare verb and SI forms. Since together the total is 100%, I only present percentage of homorganic cluster forms in (12) and (13). Thus, other percentage that are not shown in the figures are the percentage from the other forms: epenthetic schwa, bare verb and SI form *məN-*. Let us now see the homorganic cluster forms across place of articulation by JI speakers in (12).

(12)   Results by speakers (pooled across place of articulation)

N-prefix forms

A   100%   S7

B   36.17%   S3   24.71%   S2   16.67%   S6

C   7.61%   S4   4.21%   S8   3.19%   S5
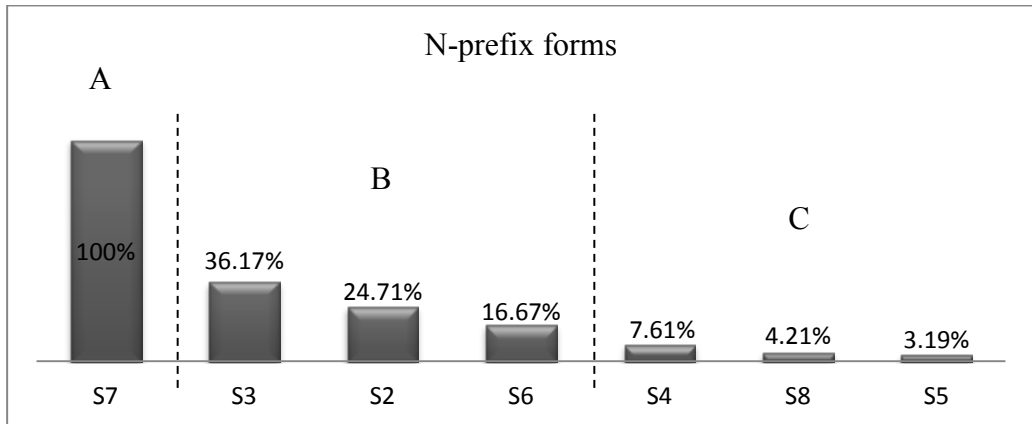
The figure (12) presents the percentage of total homorganic clusters *mb-*, *nd-*, *nd͡ʑ-*, and *ŋg-* produced by each speaker. There are three types of speakers that we can observe in (11). The first type (type A) consists of the speaker who consistently produced 100% homorganic cluster forms without any other variants. S1 falls into this category. The second type (type B) consists of the speakers who produced more variations. S3, S2 and S6 are under this category. They produced homorganic cluster forms between 16-37%. The third type of speakers (type C) consists of the speakers who produced very limited distribution of homorganic cluster forms. S4, S8 and S5 can be categorized into this type. They produced produced homorganic clusters forms between 3-8%. These categories are divided by the dash lines in (12) above.

Type A, that has only 1 speaker, is actually what would be predicted for BM speakers. Type B and type C, that together have 6 speakers, produced much more epenthetic schwa than homorganic cluster forms. This results consistent with the proposed hypothesis for this production task. In the hypothesis, it is predicted that the epenthetic schwa forms have greater distribution than homorganic cluster forms. This results also corresponds to the result from JI corpus data in (7) where epenthetic schwa has greater distribution than homorganic cluster forms.

Among the three types, type B shows the highest degree of variability. It is now important to look closely at what conditions this variability. The results in (13) below present the details.

(13)   Distribution of place of articulation by type B speakers
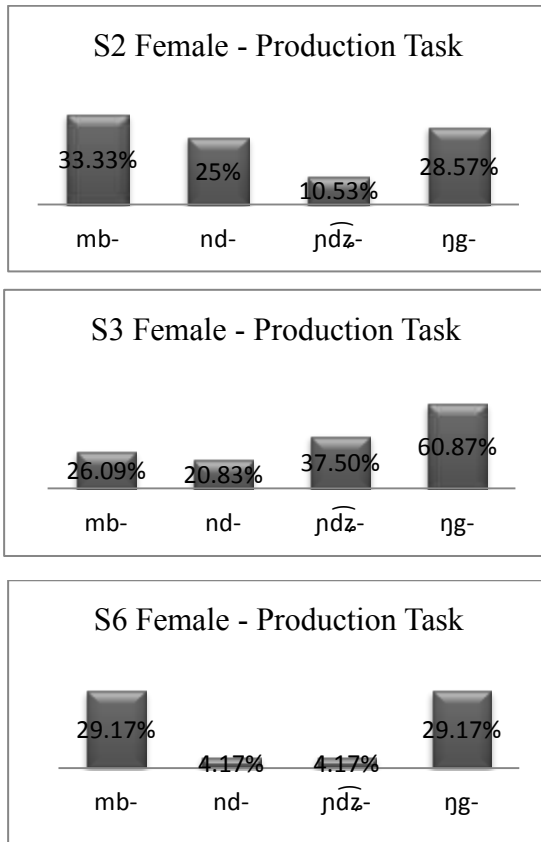
The speakers who produce more varations (S3, S2, and S6) show more variability in the figures (13). These figures exhibit interesting pattern. For labials and velars, all 3 speakers produced more than 25%. For alveolars and palatals, most of them produced not more than 25%, except for palatals that are produced by S3.

Based on this evidence, it turns out that place of articulation is another source of variation. A general pattern that we can draw is that the alveolar and palatal sounds produced in the homorganic cluster forms are less productive than the bilabial and velar sounds. Let us now turn to the results that show the intra-speaker differences.

3.2.2. Intra-speaker differences

The following table is the intra-speaker variation that are produced by 8 speakers. Almost all speakers produced intra-speaker variation. There is only one speaker (S7) that consistently avoids epenthetic schwa.

(14)   Intra-speaker variation in the production task

Speakers' code	Gender	Variation within place of articulation	Percentage
S3	F	21	43.75%
S2	F	13	27.08%
S6	F	13	27.08%
S5	F	7	14.58%
S4	M	5	10.42%
S8	F	1	2.08%
S7	M	0	0%

The number in the third column tells us about the number of test items that are produced with variation within place of articulation. The variation occurs across lexical items. For example, S3 produced variation twice in two test items. She produced [ndəket] 'to get closer' in the first test sentence and [ŋədəket] 'to get closer' in the second test sentence. In the next test item, she produced [ŋəd͡ʐəmput] 'to pick up' in the first test sentence and [ɲd͡ʐəmput] 'to pick up' in the second test sentence.

Interestingly, the types of speakers in (12) match with the order of distribution in (14). In (14), we can see that type B (S3, S2 and S6) produced the highest percentage of intra-speaker variation, which is between 27-44%. Type C (S5, S4 and S8) produced intra-speaker variation between 2-15%. Type A (S7) did not produce intra-speaker variation at all.

The table in (14) shows an interesting fact that in production task the same speakers produce the variation that mirror the results from the naturalistic spoken corpus. There is no lexical variation found so far in JI production task.

4.   Conclusion and orientation for further research

Based on the corpus results, Betawi Malay and Jakarta Indonesian speakers produced two different patterns of variations of the N-prefix: inter- and intra-speaker variations. For inter-speaker variation, BM speakers produced high numbers of the homorganic cluster forms [mb-, ɲd͡ʐ-] and [ŋg-] except for [nd-]. JI speakers produce higher numbers of the schwa epenthesic forms [ŋəb-, ŋəd-] and [ŋəg-] unless for [ŋəd͡ʐ-]. Across place of articulation, even though the numbers of tokens are low, significant difference of the tokens of the homorganic cluster forms between BM and JI speakers can be seen in the labial and velar forms. That is, the forms [mb-] and [ŋg-] are produced in a greater numbers by BM speakers. Thus, dialectal background of the speakers plays important role for the inter-speaker variation. For intra-speaker variation, the

corpus shows that the same individual speaker produced homorganic clusters and epenthetic schwa forms for the same lexical item.

The results from production task that was conducted on JI speakers mirorr the results from the corpus. Overall, as predicted in the hypothesis, JI speakers produced high numbers of epenthetic schwa forms than the homorganic cluster forms. The distribution across place of articulation shows that the bilabial and velar sounds in the homorganic clusters forms are robustly produced by JI speakers while the alveolar and palatal forms are produced in smaller numbers. Intra-speaker variation also occured in the production task. 6 out of 7 speakers produced both homorganic clusters and epenthetic schwa forms for the same place of articulation. Thus, intra-speaker variation in both corpus and production task for JI demonstrate that the process of lexicalization has not occured. The results show that there is no evidence of lexical variation observed. This is different from what Zuraw's (2010) finding based on her study in the written and dictionary corpus of Tagalog. She considered that the lexical items are already lexicalized or listed as a single lexical item in the Tagalog speakers' mind. Such case does not occur in JI.

The results from this study identify multiple sources of variation in terms of inter- and intra-speaker variation. Another source of variation found in this study is place of articulation. For further research, since most of recent studies examined nasal assimilation and substitution within formal account, it is therefore important to inquire what are the formal driving forces behind nasal assimilation process in JI. Some models of OT account have been developed to account for variation. It is important to find out how these models handle such variation observed above. It is also important to look more closely at what motivates the low distribution of alveolar and palatal sounds in homorganic clusters forms. More data collection on JI and BM is needed in the production task to fully understand factors conditioning the variation.

References

Chung, Sandra. 2008. Indonesian clause structure from an Austronesian perspective. *Lingua* 118:1554-82.
Gil, David and Uri Tadmor. 2014. The MPI-EVA Betawi-Jakarta database. A joint project of the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology and the Center for Language and Culture Studies, Atma Jaya Catholic University.
Ikranagara, Kay. 1980. Melayu Betawi grammar. *NUSA Linguistic Studies in Indonesian and Languages in Indonesia*, Vol. 9, Atma Jaya University, Jakarta.

Lapoliwa, Hans. 1981. *A generative approach to the phonology of Bahasa Indonesia*. Canberra: Department of Linguistics, Research School of Pacific Studies, Australia National University.

Muhadjir. 1981. Morphology of Jakarta dialect: Affixation and reduplication. *NUSA Linguistic Studies in Indonesian and Languages in Indonesia*, Vol. 11. Jakarta: Atma Jaya University.

Pater, Joe. 2001. Austronesian nasal substitution revisited. *Segmental phonology in Optimality Theory: Constraints and Representations*, ed. L. Lombardi, 159-182. Cambridge University Press.

Pater, Joe. 1999. Austronesian nasal substitution and Other NC effects. *The prosody morphology interface*, eds. Rene Kager, Harry van der Hulst, and Wim Zonneveld, 310-343. Cambridge University Press.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden: Blackwell. Originally circulated. 1993 as Technical Report TR-2. Rutgers Center for Cognitive Science/Technical Report CU-CS-696-93. University of Colorado at Boulder Department of Computer Science.

Sneddon, James. 2006. *Colloquial Jakartan Indonesian*. Canberra, ACT: Pacific Linguistics.

Tjung, Yassir. 2006. The formation of relative clauses in Jakarta Indonesian: A subject-object asymmetry. Doctoral dissertation, University of Delaware, Newark, DE.

Wallace, Stephen. 1976. Linguistics and social dimensions of phonological variation in Jakarta Malay. Doctoral dissertation, Cornell University, Ithaca, NY.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28:417-72.

THE SYNTAX OF TAGALOG RELATIVE CLAUSES[*]

Paul Law
City University of Hong Kong
paul.law@cityu.edu.hk

This paper provides a unified analysis of relative clauses in Tagalog according to which the head is raised from the TP complement of C leaving behind a copy. The superficial differences among the relative clauses are suggested to be due to the pronunciation of different copies. There are therefore just two types of headed relative clauses, one is externally (initially) headed and the other is internally (medially or finally) headed. Headless relative clauses are the result of deleting both copies. The account explains why they are all subject to the same constraint on movement and why the position of the clause-internal head of the RC necessarily coincides with the absolutive argument in the declarative. It is argued that the linker *na/-ng* in relative clauses is not always in C; it may also occur in the same position as the marker *ang* for the absolutive argument in the declarative.

1.   Introduction

In this paper, I show that superficially four different types of relative clauses (RCs) in Tagalog (head-initial, head-medial, head-final and headless) can be derived in essentially the same way. In particular, I suggest that all four have the same underlying D CP structure, and the head of the RC is raised from the TP complement of C to SpecCP (Vergnaud 1974, Kayne 1994). The different positions of the head noun are the results of different copies being deleted (Chomsky 1995) (the overt head of the RC is in bold and underlined; the strike-out represents deletion):[1]

(1) a. $[_{DP}$ D $[_{CP}$ $[_{TP}$ ... V ... NP ...]]]   (underlying structure)
    b. $[_{DP}$ D $[_{CP}$ **NP** $[_{TP}$ ... V ... ~~NP~~ ...]]]   (head-initial RC)
    c. $[_{DP}$ D $[_{CP}$ ~~NP~~ $[_{TP}$ ... V ... **NP** ...]]]   (head-medial RC)

---

[1] If copies of a moved phrase (or head) constitute the same phrase (or head), then the reason why only one copy can be kept can be attributed to Kayne's (1994) Linear Correspondence Axiom, according to which c-command relation maps to linear precedence. As movement is to a c-commanding position, if both copies of the moved phrase are kept, then the phrase would precede itself, a logical impossibility. This is why only one of the copies of the raised head in the RC can be kept. I thank Ed Keenan for raising this issue.

d.   [$_{DP}$ D [$_{CP}$ ~~NP~~ [$_{TP}$ ... V ... ~~NP~~ ...]]]   (headless RC)

As the head of the RC is external to TP in (1b), and internal to TP in (1c), the former may be called the externally headed RC and the latter the internally headed RC. I shall argue that for some languages the head-final RC is in fact a special case of an internally headed one, just as the head-medial RC is. Descriptively, then, there are just two types of overtly headed RCs, one is externally headed and the one is internally headed.

Particular Tagalog examples instantiating the schemata in (1) are given in (2):[2]

(2) a. **guro**-ng   d<um>ating   kahapon.   (head-initial RC)
        teacher-LK   <AP.PERF>arrive   yesterday
        'Teacher who arrived yesterday.'

   b. d<um>ating   na **guro**   kahapon.   (head-medial RC)
        <AP.PERF>arrive   LK teacher   yesterday
        'Teacher who arrived yesterday.'

   c. d<um>ating   kahapon   na **guro**.   (head-final RC)
        <AP.PERF>arrive   yesterday   LK   teacher
        'Teacher who arrived yesterday.'

   d. d<um>ating   kahapon.   (headless RC)
        <AP.PERF>arrive   yesterday
        'One who arrived yesterday.'

The account explains in straightforward way why all four types of RCs are subject to the same constraint on extraction, and why the position of the postverbal head noun coincides with that of postverbal absolutive argument in the declarative. This last point makes it possible to assimilate head-final RCs to head-medial ones. The head of the RC in the end-position, just like that in a medial position, is the same as that of the absolutive argument in the declarative.

The paper is organized as follows. I first briefly review the arguments for the head-raising analysis of RCs (Vergnaud 1974) and show that most of the evidence can be replicated for Tagalog. I then consider the constraints to which RCs in Tagalog are subject, motivating a unified account for them. I next demonstrate that the postverbal positions of the head of the RC coincide with those of the absolutive argument in the declarative. This justifies head-final RCs being a special case of internally headed ones. I argue that Aldridge's (2003, 2004) remnant TP fronting analysis of head-final RCs is empirically and conceptually inadequate for Tagalog. Lastly, I consider the morpho-syntax of the linker *na/-ng* arguing that it cannot always be in the C position as is commonly thought, but can also be in the same position as the marker *ang* for

---

[2] Abbreviations: ABS=absolutive, AP=anti-passive, APP=applicative, AY=the morpheme *ay*, ERG=ergative, LK=linker, OBL=oblique, P=preposition, PERF=perfective, PL=plural, S=singular.

the absolutive argument. I conclude the paper with some brief remarks on the cross-linguistic variation with respect to the type of RC present or absent in specific languages.

2. The head-raising analysis of relative clauses

A familiar argument that RCs should be analyzed in terms of movement is that they resemble other cases of overt movement, e.g., *wh*-movement in questions. As shown in (3), a *wh*-phrase can be moved out of a complement clause in questions (*wh*-phrases with a strike-out are the position from which the *wh*-phrases move), and the head of a RC, too, can be related to a position in a complement clause, as shown in (4):

(3) a.   Who did they say that that Mary saw ~~who~~?
    b. ?Who did they wonder whether Mary saw ~~who~~?
    c. *Who did they leave before Mary saw ~~who~~?

(4) a.   The man that they said that Mary saw.
    b. ?The man that they wondered whether Mary saw.
    c. *The man that they left before Mary saw.

Thus, if *wh*-questions involve movement, then so should RCs. But the similar facts in (3) and (4) do not say much about what moves in RCs.

There are essentially two views about movement in RCs. One is that what moves is a phonetically null counterpart of the *wh*-phrase in questions, known as the empty operator (Chomsky 1986, Browning 1987), as in (5a), and the other is that the head of the RC moves to SpecCP (Kayne 1994), as in (5b):

(5) a.   The man [$_{CP}$ O_i [that [$_{TP}$ they said [$_{CP}$ t_i [that [$_{TP}$ Mary saw t_i]]]]]]
    b.   The [$_{CP}$ **man** [that [$_{TP}$ they said [$_{CP}$ ~~man~~ [that [$_{TP}$ Mary saw ~~man~~]]]]]]

Arguments favoring the head-raising analysis are based on the reconstruction effect. A reflexive in the head of the RC can be bound by an antecedent that does not c-command it, and part of an idiom expression can be discontinuous with the rest of it:

(6) a.   D [$_{CP}$ [that [$_{TP}$ John saw pictures of himself]]]
    b.   D [$_{CP}$ **pictures of himself**$_i$ [that [$_{TP}$ John$_i$ saw ~~pictures of himself~~$_i$]]

(7) a.   The [$_{CP}$ **advantage** [that [$_{TP}$ they took ~~advantage~~ of John]]]
    b.   The [$_{CP}$ **tabs** [that [$_{TP}$ the NSA kept ~~tabs~~ on everybody]]]

If the head of the RC actually originates in the RC, then the binding of the reflexive as well as the idiomatic interpretations of discontinuous idioms can be

accounted for straightforwardly. What is bound in (6b) is the reflexive in the copy the raised head leaves behind in the RC. There, it is c-commanded and bound by its antecedent. Similarly, the idiomatic interpretations of the idioms in (7) are possible, as the part of the idiom that is moved leaves behind a copy in the clause where it is interpreted. The assumption here is that idiomatic interpretation requires that parts of an idiom be at least in the VP, if not contiguous. The head-raising analysis thus provides a very straightforward account for binding and idiomatic interpretations in RCs.

Largely the same facts obtain in Tagalog. As we will look at movement in some detail in the next section, I give here an example of binding in which a reflexive in the head of the RC is bound by an antecedent that does not c-command it:

(8) a. nakita ko   ang larawan   ng kanya-ng sarili na b<in>ili   ni John.
         saw   ERG ABS picture   LK 3S-LK   self LK <PERF>buy ERG
         'Maria saw pictures of himself that John bought.'
    b. D [CP **larawan ng kanya-ng sarili** [[TP binili ni John ~~larawan ng kanya-ng sarili~~]]]

As in (6b), what is bound is the reflexive in the copy of the raised head left behind in the TP.

I cannot replicate the idiom argument for Tagalog, however, for I could not elicit many examples of idiom. To the extent I can get any, they behave much like English rigid idioms of the type *kick the bucket*. Parts of this type of idiom cannot be displaced. Thus, expressions like *the bucket that John kicked* or *how many buckets did they kick?* do not have idiomatic readings, in contrast with idioms like *take advantage of*. Like that in (7b), the example *how much advantage did they take of John?* retains the idiomatic interpretation of *take advantage of*. Nevertheless, as we will see, there are other facts that are more easily accommodated in the head-raising analysis than in the null operator account.

3.   Constraint on movement and relative clauses in Tagalog

As is well-known, movement in Tagalog and other Austronesian languages is subject to the constraint that only the absolutive (or subject) argument can be extracted, and the clause out of which movement takes place must also be the absolutive argument of the matrix predicate (Schachter 1976, Keenan 1976, Dell 1981, Kroeger 1993).[3] For descriptive convenience, I will call this constraint the absolutive-only constraint on extraction (for a theoretical account, see Nakamura 1994, Richards 2000, Rackowski and Richards 2005):

---

[3] As my account would later be compared with Aldridge's (2003, 2004) remnant TP fronting analysis, I assume with her that Tagalog is an ergative language, rather than a nominative language (Kroeger 1993). This choice has no bearing on the analysis of RCs.

(9) **The absolutive-only constraint on extraction**
   a.   NP ... V ... ~~NP~~ ...   (only if NP is absolutive/subject)
   b.   NP ... V ... [CP ... V ... ~~NP~~ ...]   (only if NP and CP are absolutive/subject)

PPs and adverbs are not subject to absolutive-only constraint in (9a) and may freely extract. Long-distance extraction of these is nevertheless restricted by the constraint in (9b).

The effect of (9a) can be seen in the *ay*-inversion construction, examples of which are given in (10)-(12) (cf. Schachter and Otanes 1972):

(10) a.   ang   guro   ay   b<um>ili   ~~ang   guro~~   ng   libro.
           ABS   teacher   AY   <AP.PERF>buy   ABS   teacher   OBL   book
           'It's the teacher who bought the book.'
   b. *ng   libro   ay   b<um>ili   ang   guro   ~~ng   libro~~.
           OBL   book   AY   <AP.PERF>buy   ABS   teacher   OBL   book
           'It's the book that the teacher bought.'

(11) a.   ang   libro   ay   b<in>ili   ng   guro   ~~ang   libro~~.
           ABS   book   AY   <PERF>buy   ERG   teacher   ABS   book
           'It's the book that the teacher bought.'
   b. *ng   guro   ay   b<in>ili   ~~ng   guro~~   ang   libro.
           ERG   teacher   AY   <PERF>buy   ERG   teacher   ABS   book
           'It's the teacher that bought the book.'

(12) a.   ang   bata   ay   b<in>igy-an   ng   babae   ~~ang   bata~~   ng   kendi.
           ABS   child   AY   <PERF>give-APP   ERG   woman   ABS   child   OBL   candy
           'It's the child that a woman gave candy to.'
   b. *ng   kendi   ay   b<in>igy-an   ng   babae   ang   bata   ~~ng   kendi~~.
           OBL   candy   AY   <PERF>give-APP   ERG   woman   ABS   child   OBL   candy
           'It's the candy woman gave to the child.'
   c. *ng   babae   ay   b<in>igy-an   ng   babae   ang   bata   ~~ng   kendi~~.
           ERG   woman   AY   <PERF>give-APP   ERG   woman   ABS   child   OBL   candy
           'It's the candy woman gave to the child.'

The effect of (9b) can be seen in (13), even though it is not obvious that the clausal complement is the absolutive argument of the matrix predicate:

(13) a.   ang   libro   ay   s<in>abi   ni   Fred [na   b<in>ili   ni Maria
            ABS   book   AY   <PERF>say   ERG   LK   <PERF>buy   ERG
            ~~ang   libro~~]
            ABS   book
            'It is the book that Fred said that Maria bought.'

b. *ang libro ay nag-sabi   si Fred [na b<ini>li   ni Maria
    ABS   book   AY AP.PERF-say   ABS   LK <PERF>buy   ERG
    ~~ang   libro~~]
    ABS   book
    'It is the book that Fred said that Maria bought.'

Contrary to absolutive DP argument, the clausal complement is never marked with *ang*. However, as the matrix argument in (13a) is ergative, it is plausible that the clausal complement of the matrix predicate is absolutive. This is the same in other cases with two DP arguments, e.g., (10a) and (11a). If one of the arguments is ergative, then the other is absolutive.

Another way to look at the constraint in (9b) is that the matrix clause with an ergative argument can host the absolutive argument of the embedded clause, but that with an absolutive argument cannot. (13a) is good but (13b) is bad, precisely because the matrix clause in the former has an ergative argument, whereas that in the latter has an absolutive argument.

Turning now to RCs, we can see that they are subject to the same absolutive-only constraint on extraction. Examples in (14)-(16) clearly show the effect of (9a):

(14) a. kilala ko ang **guro**-ng [b<um>ili   ~~guro~~   ng   libro]
         know   1S ABS teacher-LK <AP.PERF>buy teacher OBL book
         'I know the teacher who bought the book.'

   b. kilala ko ang ~~guro~~ [b<um>ili   na   **guro**   ng   libro]
       know   1S ABS teacher <AP.PERF>buy LK   teacher OBL book
       'I know the teacher who bought the book.'

   c. kilala ko ang ~~guro~~ [b<um>ili   ng   libro na **guro**]
       know   1S ABS teacher <AP.PERF>buy OBL book   LK teacher
       'I know the teacher who bought the book.'

   d. kilala ko ang ~~guro~~ [b<um>ili   ~~guro~~   ng   libro]
       know   1S ABS teacher <AP.PERF>buy teacher OBL book
       'I know the one who bought the book.'

(15) a. *kilala ko ang **libro**-ng [b<um>ili   ang guro   ~~libro~~]
         know   1S ABS book-LK   <AP.PERF>buy ABS teacher book
         'I know the teacher who bought the book.'

   b. *kilala ko ang ~~libro~~ [b<um>ili   ang guro   na **libro**]
       know   1S ABS book   <AP.PERF>buy ABS teacher LK book
       'I know the teacher who bought the book.'

   c. *kilala ko ang ~~libro~~ [b<um>ili   na   **libro**   ang guro]
       know   1S ABS book   <AP.PERF>buy LK   book   ABS teacher
       'I know the teacher who bought the book.'

   d.   *kilala ko ang ~~libro~~ [b<um>ili   ang guro   ~~libro~~]
             know 1S ABS book   <AP.PERF>buy ABS teacher book
             'I know the one who bought the book.'

(16) a. gusto ko ang **libro**-ng [b<in>ili   ng   guro   ~~libro~~]
             like   1S ABS book-LK   <PERF>buy ERG   teacher   book
             'I like the book that the teacher bought.'

      b. gusto ko ang ~~libro~~ [b<in>ili   ng   guro   na   **libro**]
             like   1S ABS book   <PERF>buy ERG teacher LK   book
             'I like the book that the teacher bought.'

      c. ??gusto ko ang ~~libro~~ [b<in>ili   na   **libro** ng   guro   ]
             like   1S ABS book   <PERF>buy LK   book   ERG   teacher
             'I like the book that the teacher bought.'

      d. gusto ko ang ~~libro~~ [b<in>ili   ng   guro   ~~libro~~]
             like   1S ABS book   <PERF>buy ERG teacher book
             'I like the book that the teacher bought.'

The examples in (14) and (16) are good precisely because the extracted argument is absolutive. This is regardless of the position of the head or whether the head is overt. In (14) and (16) the upper copy is pronounced in the a-examples, and the clause-internal copy is pronounced in the c-and d-examples (on the marginal status of (16c), see the discussion of (25) and (26) below).

      The ungrammaticality of the examples in (15b-c) is most striking. The head of the RC apparently remains in the clause, but these examples are nevertheless as ungrammatical as that in (15a) where the head occurs outside the clause. They are thus evidence that some movement takes place, even though the head is still in the clause. It is not obvious how the null operator account of RCs would explain why (15b-c) are excluded. As the head is in the clause, it is hard to see where the null operator can be said to move from. For the head-raising account in terms of head-raising and copy-and-deletion, these cases are fairly straightforward. The head of the RC in (15b-c) is raised to SpecCP much like that in (15a). They are therefore all subject to the same absolutive-only constraint. Regardless of which copy of the raised head is pronounced, the result is the same.

      The effect of (9b) in RCs can be seen in the examples in (17)-(18). An embedded absolutive argument can be extracted to the matrix clause only if the matrix clause does not already host an absolutive argument:

(17) a. gusto ko ang **libro**-ng s<in>abi   ni Fred [na b<in>ili
             like   1S ABS book-LK <PERF>say ERG   LK <PERF>buy
             ni   Maria ~~libro~~]
             ERG   book
             'I like the book that Fred said that Maria bought.'

b. gusto ko ang ~~libro~~ s<in>abi ni Fred [na b<in>ili ni
like 1S ABS book <PERF>say ERG LK <PERF>buy ERG
Maria na **libro**]
LK book
'I like the book that Fred said that Maria bought.'

c. gusto ko ang ~~libro~~ s<in>abi ni Fred [na b<in>ili na
like 1S ABS book <PERF>say ERG LK <PERF>buy LK
libro ni Maria]
book ERG
'I like the book that Fred said that Maria bought.'

d. gusto ko ang ~~libro~~ s<in>abi ni Fred [na b<in>ili ni
like 1S ABS book <PERF>say ERG LK <PERF>buy ERG
Maria ~~libro~~]
book
'I like the one that Fred said that Maria bought.'

(18) a. *gusto ko ang **libro**-ng nag-sabi si Fred [na b<in>ili
like 1S ABS book-LK AP.PERF-say ABS LK <PERF>buy
ni Maria ~~libro~~]
ERG book
'I like the book that Fred said that Maria bought.'

b. *gusto ko ang ~~libro~~ nag-sabi si Fred [na b<in>ili ni
like 1S ABS book AP.PERF-say ABS LK <PERF>buy ERG
Maria na **libro**]
LK book
'I like the book that Fred said that Maria bought.'

c. *gusto ko ang ~~libro~~ nag-sabi si Fred [na b<in>ili na
like 1S ABS book AP.PERF-say ABS LK <PERF>buy LK
libro ni Maria]
book ERG
'I like the book that Fred said that Maria bought.'

d. *gusto ko ang ~~libro~~ nag-sabi si Fred [na b<in>ili ni
like 1S ABS book AP.PERF-say ABS LK <PERF>buy ERG
Maria ~~libro~~]
book
'I like the book that Fred said that Maria bought.'

Again, the contrast between (17b-c) and (18b-c) is most interesting. In both cases, the head has apparently not been moved out of the clause; yet, they show the effect of (9b). This can be explained if the head in all these cases has in fact moved, the appearance of the head in a clause-internal position is simply due to the pronunciation of the lower copy.

Also relevant to the analysis of RCs is the lack of preposition stranding in Tagalog. The examples in (19c) show that the complement of a preposition may not be moved in the *ay*-inversion construction:

(19) a. i-b<in>igay   ng babae   [PP sa bata] ang mangga.
     APP-<PERF>give   ERG woman   P child ABS mango
     'A woman gave the mango to the child.'

   b. ang   mangga ay i-b<in>igay   ng babae   [PP sa bata]
     ABS   mango AY APP-<PERF>give ERG woman   P   child
     ~~ang   mangga~~.
     ABS   mango
     'It is the mango that a woman gave to the child.'

   c. *bata   ay i-b<in>igay   ng   babae [PP sa ~~bata~~] ang
     child AY APP-<PERF>give ERG   woman   P   child   ABS
     mangga.
     mango
     'A woman gave the mango to the child.'

   d. * (ng) babae   ay i-b<in>igay   ~~ng babae~~ [PP sa bata] ang
      ERG woman   AY APP-<PERF>give ERG woman P child ABS
     mangga.
     mango
     'A woman gave the mango to the child.'

The lack of preposition stranding in Tagalog may not seem surprising, given that preposition stranding is rare cross-linguistically. As we will see shortly when we consider RCs, there is reason to believe that the ungrammaticality of the (19c) is related to the absolutive-only constraint on extraction, rather than to an independent constraint barring preposition stranding. It is ruled out for the same reason as those in (19c-d) are. The extracted phrase is not the absolutive argument.

The examples in (20) shows that it is not possible for the complement of a preposition to be relativized, regardless of the position of the head or whether the head is overt:

(20) a. *si   Pedro ang **bata**-ng   i-b<in>igay   ng babae   sa ang
     ABS   ABS child-LK APP-<PERF>give ERG woman   P ABS
     mangga.
     mango
     'Pedro is the child who a woman gave the mango to.'   (cf. (19c))

   b. [CP child [TP gave Maria to ~~child~~ mango]]

(21) a. *si   Pedro ang i-b<in>igay   ng   babae [PP sa **bata**] ang
         ABS   ABS   APP-<PERF>give   ERG   woman   P   child   ABS
         mangga.
         mango
         'Pedro is the child who a woman gave the mango to.'   (cf. (19d))

   b. [CP ~~child~~ [TP gave Maria [PP to child] book]]

(22) a. *si   Pedro ang i-b<in>igay   ng   babae sa ang mangga.
         ABS   ABS APP-<PERF>give   ERG   woman P   ABS mango
         'Pedro is the one who a woman gave the mango to.'

   b. [CP ~~child~~ [TP gave Maria [PP to ~~child~~] book]]

While (20a) and (22a) may be excluded because the preposition is stranded, the explanation does not extend to (21a). The preposition is clearly not stranded, the head of the RC remaining in the PP. In the head-raising and copy-and-deletion analysis, (21a) is straightforwardly excluded. The head in (21a) is raised just as it is in (20a) and (22a). All these cases are ruled out, since the extracted argument is not the absolute argument.

4.   The postverbal positioning of the head of the relative clause

There is reason to suppose that head-final RCs are a special case of internally headed ones. The positioning of the head in the clause-final position is just the same as that of the absolutive argument in the declarative.
        Postverbal order of arguments in the declarative is relatively free (Schachter and Otanes 1972). Thus, the word-orders in (23) are three of the six possible word-orders for a three-place predicate like *bigyan* 'give':

(23) a. b-in-igy-an   ni   Fred ang bata   ng   lapis   kahapon.
         <PERF>give-APP   ERG   ABS child OBL pencil yesterday
         'Fred gave a pencil to the child yesterday.'

   b. b-in-igy-an   ni   Fred ng   lapis ang bata   kahapon.
         <PERF>give-APP   ERG   OBL pencil ABS child   yesterday
         'Fred gave a pencil to the child yesterday.'

   c. b-in-igy-an   ni   Fred ng   lapis kahapon   ang bata.
         <PERF>give-APP   ERG   OBL pencil yesterday ABS child
         'Fred gave a pencil to the child yesterday.'

Now, given that the head of an internally headed RC remains in the clause, it should come as no surprise that it may occur in the same positions as those of the absolutive argument in the declarative. The examples in (24) show that the positioning of the head of the RC coincides exactly with that of the absolutive argument in the declarative in (23):

(24) a. b-in-igy-an ni Fred na **bata** ng lapis kahapon.
   <PERF>give-APP ERG   LK child OBL pencil yesterday
   'Child who Fred gave a pencil to yesterday.'

  b. b-in-igy-an ni Fred ng lapis na **bata** kahapon.
   <PERF>give-APP ERG   OBL pencil LK child yesterday
   'Child who Fred gave a pencil to yesterday.'

  c. b-in-igy-an ni Fred ng lapis kahapon na **bata**.
   <PERF>give-APP ERG   OBL pencil yesterday LK child
   'Child who Fred gave a pencil to yesterday.'

As Aldridge (2004) pointed out, speakers prefer the ergative agent argument to precede the absolutive theme argument. There is detectable contrast between (25a) and (25b):

(25) a. b<in>ili ng guro ang libro.
   <PERF>buy ERG teacher ABS book
   'The teacher bought the book.'

  b. ??b<in>ili ang libro ng guro.
   <PERF>buy ABS book ERG teacher
   'The teacher bought the book.'

They detect the same contrast in the internally RCs in (26):

(26) a. gusto ko ang ~~libro~~ [b<in>ili ng guro na **libro**]
   (=(16b))
   like 1S ABS book <PERF>buy ERG teacher LK book
   'I like the book that the teacher bought.'

  b. ??gusto ko ang ~~libro~~ [b<in>ili na **libro** ng guro]
   (=(16c))
   like 1S ABS book <PERF>buy LK book ERG teacher
   'I like the book that the teacher bought.'

The similarity between (25) and (26) is hardly surprising. As it remains in the clause, the head of the RC should be able to occur in the same positions as those of the absolutive argument in the declarative.

5. Head-final relative clauses and the remnant TP fronting analysis

Aldridge (2003, 2004) suggested that head-final RCs be distinguished from internally headed RCs in that the former are derived by fronting of the remnant TP after the head is raised to SpecCP, while the latter are derived by raising the head not to SpecCP but to the Spec of a functional projection FP;

the head is related to a base-generated null operator in SpecCP:[4]

(27) a. b<in>ili   ni   Maria-ng **libro**.   (head-final RC)
         <PERF>buy ERG   -LK   book
         'The book Maria bought'
     b. [$_{DP}$ [$_{TP}$ bought t_i Maria]$_j$ [$_{CP}$ **book** t_j]]

(28) a. i-b<in>igay   na **kendi** ng babae   sa bata. (internally headed RC)
         APP-<PERF>give LK candy ERG woman P child
         'The candy the woman gave to the child.'
     b. [$_{DP}$ [$_{CP}$ Op$_i$ [$_{TP}$ give [$_{FP}$ **candy**$_i$ [$_{vP}$ woman [$_{VP}$ ... t_{NP} ...]]]]]]

     The reason for the distinction, she argued, is that the head of a head-final RC may be followed by a PP, but not a DP argument. In her terms, a PP may, but a DP may not, be stranded after the head of a head-final RC. The grammatical contrast in (29) is due to the different derivations in (30):

(29) a. i-b<in>igay   ng   babae-ng **kendi** sa bata.
         APP-<PERF>give ERG woman-LK candy P child
         'Candy the woman gave to the child.'
     b. *b<in>igy-an   ng   babae-ng **bata** ng kendi.
         <PERF>give-APP ERG woman-LK child OBL candy
         'Child to whom the woman gave candy.'

(30) a. [$_{DP}$ [$_{TP}$ gave woman t_{NP} t_{PP}] [$_{CP}$ [$_{NP}$ **candy**] [$_{XP}$ [$_{PP}$ to child] t_{TP}]]]
     b. *[$_{DP}$ [$_{TP}$ gave woman t_{NP} t_{DP}] [$_{CP}$ [$_{NP}$ **child**] [$_{XP}$ [$_{DP}$ candy] t_{TP}]]]

The examples in (29) are head-final RCs, even though the head is not at the end, the head being separated from the predicate by an ergative argument (see note 4). In the remnant TP fronting analysis, the head-final RC is derived by remnant TP fronting so that the predicate would end up in front of the raised head. Thus, the PP or DP that follows the head of the head-final RC must first be moved out of TP, and the remnant TP is subsequently fronted. The first step is possible with PP but not with non-absolutive DP, given the absolutive-only constraint on extraction (see the derivations in (30)). The contrast in (29) is thus explained.

     If the head-final RC is a special case of an internally headed RC and the clause-internal head of the RC may be in any position in which the absolutive argument in the declarative may appear, then the example in (29b) is problematic. On this view, (29b) should be grammatical. As it turns out,

---

[4] As we will see, what is taken by Aldridge to be a head-final RC with the derivation in (27b) does not always have the head at the end of the RC. One way to recognize a head-final RC, in Aldridge's analysis, is to observe that the head is not adjacent to the predicate (see the discussion of (29) below).

example (29b) is indeed possible if the form of the linker is *na* preceded by a small pause indicated as *#*, as shown in (31a):

(31) a. b-in-igy-an   ng   babae # na **bata** ng kendi.   (cf. (29b))
        <PERF>give-APP ERG woman   LK child OBL candy
        'Child who the woman gave candy to.'
    b b-in-igy-an   ni   Fred   na **bata** ng   kendi.
        <PERF>give-APP   ERG   LK child OBL candy
        'Child who Fred gave candy to.'

Likewise, a similar example with the linker *na* (with or without a pause) is also possible if the argument preceding the linker *na* ends in a consonant, as shown in (31b).
       In fact, other examples comparable to (29b) with the linker *-ng* are also possible. The example in (32a) differs minimally from that in (29b) in that the ergative argument separating the head of the RC and the predicate is a (modified) pronoun:

(32) a. b<in>igy-an   ko-ng **bata**   ng   kendi.
        <PERF>give-APP   1S-LK   child   OBL candy
        'Child who I gave candy to.'
    b. b<in>igy-an   **nami-ng (mga) bata**   ng   kendi.
        <PERF>give-APP   1PL-LK   PL   child   OBL candy
        'One who we children gave candy to.'
        NOT 'the child who we gave candy to.'

The examples in (31) and (32a) are clearly problematic for the remnant TP fronting analysis. There is apparently no reason why the form of the linker or the ergative argument being a pronoun should make a difference with respect to stranding of a DP after the head of the RC. Example (32b) might be taken to be an internally headed RC, not a head-final one, for the head is not separated from the predicate by an ergative argument (see note 4). We will see shortly that it is very similar to example (29b), and should be given the same structure of a head-final RC.
       It seems that the reason why (29b) is judged not to have the indicated reading is due to an independent reason having to do with processing. According to my consultants, (29b) is grammatical with *babaeng bata* being understood as 'female who is a child, i.e., girl', a reading that is possible in general. In other words, (29b) is a headless RC with the reading 'someone who a girl gave candy to'. This point is re-enforced by the examples in (32). The string *kong bata* in (32a) cannot be understood as 'I who is a child', for only a plural pronoun can be modified (Schachter and Otanes 1972). Consequently, *ko* 'I' is interpreted as the ergative argument and *bata* 'child' as the head of the RC.

(32b) is unproblematic. The plural pronoun *nami* 'we' can be modified, and the string *naming bata* (*-ng* is the linker) is accordingly interpreted as 'we who are children, we children'. The interpretation of *naming bata* as the ergative argument is just the same as that of *babaeng bata* in (29b).

It is therefore clear that the examples in (29) do not support the remnant TP fronting analysis.[5] In fact, the remnant TP fronting account is empirically problematic insofar as it does not explain why the positioning of the head of the RC should co-incide with that of the absolutive argument in the declarative.

6.   The morpho-syntax of the linker *na/-ng*

The morpho-syntax of the linker *na/-ng* in RCs has not been systematically studied. It has a direct bearing on the structure of RC, and hence deserves a close look.

It is commonly assumed that the linker *na/-ng* is in the C position of the RC. This may be correct for the externally headed (head-initial) RC, but not obviously so for the internally headed (head-medial and head-final) ones:

(33) a. **guro**-ng   d<um>ating   kahapon.   (externally headed RC)
            teacher-LK   <AP.PERF>arrive yesterday
            'Teacher who arrived yesterday.'

      b. d<um>ating   na **guro**   kahapon.   (internally headed RC)
            <AP.PERF>arrive LK   teacher yesterday
            'Teacher who arrived yesterday.'

(33a) differs from (33b) in that the linker *na/-ng* is to the right of the head of the RC in the former and is to the left of it in the latter. If it is always in C, then it must be that the head of the RC lands in SpecCP in (33a) but in a lower position in (33b). It is not clear where that lower position can be and why it should be the case.

In addition, in order to account for the position of the predicate in (33b) it must be assumed that remnant movement of some sort takes place to the effect that the predicate ends up before the raised head of the RC. Apart from the fact that remnant movement analysis cannot explain why the position of the head of the RC should be the same as that of the absolutive argument in the declarative, certain facts indicate that the linker *na/-ng* cannot always be in C.

If the linker *na/-ng* is always in C, then it is predicted that what appears before it is a syntactic constituent in SpecCP. The prediction is not entirely borne out. While what appears before the linker *na* in (34a) may be a syntactic constituent, but it clearly cannot be in (34b), the oblique argument and the adverb being separated from the predicate they are related to:

---

[5] By contrast, the head-final RC in Japanese can only be derived by remnant TP fronting (Kayne 1994). This is because Japanese does not allow an argument to occur after the predicate.

(34) a. b-in-igy-an   ni Fred ng lapis na **bata** kahapon.
       <PERF>give-APP ERG   OBL pencil LK child yesterday
       'Child who Fred gave pencil to yesterday.'

    b. b-in-igy-an   ni Fred na **bata** ng lapis kahapon.
       <PERF>give-APP ERG   LK child OBL pencil yesterday
       'Child who Fred gave pencil to yesterday.'

    c. gusto ko ang ~~libro~~ s<in>abi ni Fred [na b<in>ili
       like 1S ABS book <PERF>say ERG   LK <PERF>buy
       na **libro** ni Maria]
       LK   book ERG
       'I like the book that Fred said that Maria bought.'

The problem is even worse in (34c). What appears before the linker *na* before the head of the RC cannot possibly a syntactic constituent, for it comprises the matrix clause and part of the embedded clause.

      In light of these facts, I would like to suggest that whereas the linker appearing to the right of the head of the head-initial RC is in C, that occurring to the left of the head of the internally headed RC is in the same position as the marker *ang* for the absolutive argument in the declarative (see also Otsuka 2014 for the view that it is an allomorph of *ang*). Evidence for this claim comes from the distribution of the linker *na/-ng* in RCs being the same as that of the absolutive marker *ang* in the declarative.

      In the declarative, the absolutive marker *ang* may be separated from the noun by a PP, by the plural morpheme *mga* and an adjective, or by a pronominal possessor:

(35) a. binasa   ni Max ang nasa mesang libro.
       <PERF>read ERG   LK on table book
       'Max read the book on the table.'

    b. binili   ni Fred ang mga bagong <u>libro</u>.
       <PERF>buy ERG   ABS PL new   <u>book</u>
       'Fred bought the new books.'

    c. gusto   ni Pat ang aking guro.
       like   ERG   LK my   teacher
       'Pat likes my teachers.'

In the internally headed RC, too, the linker *na/-ng* and the noun of the head of the RC may be separated by a PP, by the plural morpheme *mga* and an adjective, or by a pronominal possessor:

(36) a. binasa   ni Max na nasa mesang **libro**.
       <PERF>read ERG   LK on table book
       'book on the table that Max read'

b. binili   ni Fred   na mga bagong **libro**.
   <PERF>buy ERG   LK PL   new   book
   'new books that Fred bought'

c. gusto   ni Pat na aking **guro**.
   like   ERG   LK my   teacher
   'Teachers of mine that Pat likes'

In the declarative, the absolutive marker *ang* may not be separated from the noun by a manner or sentential adverb:

(37) a. binasa   ni Max (malimit) ang (*malimit) libro.
        <PERF>buy ERG   often   ABS   often   book
        'Max (often) read books.'

   b. binili   ni Fred (kahapon) ang (*kahapon)   libro.
      <PERF>buy ERG   yesterday LK   yesterday   book
      'Fred bought a book (yesterday).'

Nor may the linker *na/-ng* in the internally RC be so separated from the head noun:

(38) a. binasa   ni Max (malimit)   na (*malimit) **libro**.
        <PERF>read ERG   often   LK   often   book
        'Books that Max (often) read.

   b. binili   ni Fred (kahapon) na (*kahapon)   **libro**.
      <PERF>buy ERG   yesterday LK yesterday   book
      'Book Fred bought (yesterday)'

The identical distribution of the linker *na/-ng* in the internally headed RC and the absolutive marker *ang* in the declarative is too striking to be a pure co-incident. It would be explained if they in fact occur in the same position.

7.   Conclusion

If the different types of RC are derived essentially in the same way by head-raising, with the differences being due to the pronunciation of different copies of the raised head, then an issue that arises is how cross-linguistic variation with respect to the type of RCs present or absent in particular languages can be accounted for. Specifically, why should it be that Tagalog has the option of pronouncing the lower copy of the raised head resulting in an internally headed RC, but English does not, English having no internally headed RCs?

The answer to this question obviously must be that English does not have the option of pronouncing the lower copy. But why should that be? When we look at other cases of movement, it becomes clear that in English the copy in the final landing site quite generally must be pronounced. This is most self-

evident in questions, raising to subject and object (also known as exceptional Case-marking). In these cases, the lower copy of the raised phrase is necessarily deleted and the upper copy in the landing site is pronounced. Therefore, the lack of internally headed RCs in English is just the same lack of option of pronouncing the lower copy. By contrast, in Tagalog any copy of the raised phrase may be pronounced and the other copies are deleted (see Law 2014 for examples). As well, the relatively free postverbal word-order of arguments can be taken to be the result of pronouncing different copies arising from movement. In this light, it is unsurprising that Tagalog has both externally and internally headed RCs.

It is worth pointing out that there might be independent factors requiring that a particular copy be pronounced. For instance, in Tagalog the upper copy of the phrase preceding *ay* in the *ay*-inversion construction must be pronounced, even though it need not be in other cases of movement. This is due to the independent fact that the position before *ay* is necessarily stressed. The stress cannot be realized if the phrase undergoing *ay*-inversion is not pronounced.

To what extent the account proposed here can be extended to other languages is an issue that can only be adequately addressed by detailed and systematic investigations of RCs in particular languages.

References

Aldridge, Edith. 2003. Remnant movement in Tagalog relative clause formation. *Linguistic Inquiry* 34: 631-640.

Aldridge, Edith. 2004. Internally headed relative clauses in Austronesian languages. *Language and Linguistics* 34: 631-640.

Browning, M. A. 1987. Null operator constructions. Doctoral dissertation, MIT.

Chomsky, Noam. 1995. *The Minimalist program*. Cambridge, Ma. MIT Press.

Chomsky, Noam. 1986. *Barriers*. Cambridge, Ma. MIT Press.

Dell, Francois. 1981. On certain sentence complements in Tagalog. *Cahiers de linguistique Asie-Orientale* 10: 19-41.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, Ma. MIT Press.

Keenan, Edward. 1976. Remarkable subjects in Malagasy. In *Subject and topic*, ed. Charles Li, 247-301. New York. Academic Press.

Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford, Ca. CSLI Publications.

Law, Paul. 2014. A unified theory of relative clauses. In *Proceedings of the 16th Seoul International Conference On Generative Grammar*, ed. by Jong-un Park and Il-Jae Lee, 203-221. The Korean Generative Grammar Circle, Seoul, Korea.

Nakamura, Masanori. 1994. An economy account of *wh*-extraction in Tagalog. In *Proceedings of the 12th West Coast Conference on Formal Linguistics*, 405-42. Stanford, Ca. CSLI Publications.

Otsuka, Yuko. 2014. Relative clauses in Polynesian and the role of D. Paper presented at the 21st Austronesian Formal Linguistics Association, University of Hawai'i, Mānoa.

Rackowski, Andrea and Norvin Richards. 2005. Phase edge and extraction: A Tagalog case study, *Linguistic Inquiry* 36: 565–599.

Richards, Norvin. 2000. Another look at Tagalog subjects. In *Formal Issues in Austronesian Linguistics*, ed. Ileana Paul, Vivian Phillips, Lisa Travis, 105-116. Dordrecht. Kluwer Academic Publishers.

Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. In *Subject and topic*, ed. Charles Li, 491-518. New York. Academic Press.

Schachter, Paul and Fe Otanes. 1972. *Tagalog reference grammar*. Berkeley and Los Angeles, Ca. University of California Press.

Vergnaud, Jean-Roger. 1974. French relative clauses. Doctoral dissertation, MIT.

APPLICATIVES AS SECONDARY PREDICATES[*]

Diane Massam
University of Toronto
diane.massam@utoronto.ca

This paper re-examines Niuean *aki* constructions, which are usually considered to be instrumental applicatives. It is argued that when we consider the full range of *aki* constructions, they do not conform to the diagnostics for applicatives. Instead, I propose that we consider *aki* to be one in the class of secondary predicates found in Niuean (cf. Ball 2005, 2008).

1. Introduction

1.1. Overview

Niuean is considered to have an instrumental applicative construction, as shown in (1b). This applicative is symmetrical, in that both Instrument & Theme act as objects for extraction, pro-drop, raising, etc. (Chung 1978, Seiter 1980).[1]

(1) a. Kua hele tuai   e   ia   <u>e   falaoa</u>   <u>aki e   titipi</u>
     Perf cut Perf   Erg.P 3.Sg Abs.C bread   with Abs.C   knife
     'He has cut the bread with the knife.' (FN)

   b. Kua hele aki tuai   e   ia   <u>e   titipi</u> <u>e   falaoa</u>
     Perf cut use Perf   Erg.P 3.Sg Abs.C knife   Abs.C bread
     'He has cut the bread with the knife.' (FN)[2]

The following properties, in (2), are generally considered to hold of applicative constructions (Marantz 1993, Pylkkänen 2002, 2008, Cuervo 2003, McGinnis

[1] Symmetry is a hallmark of Niuean arguments: all of subjects, objects and applied arguments are eligible for extraction etc. Because this is not a property unique to these applicative-like constructions, I do not address this issue in this paper (Seiter 1980, Massam 1985, Larson, Longenbaugh & Polinsky 2015). I do not discuss prepositional cases (1a), see Ball (2005, 2008).
[2] Abbreviations are: Abs absolutive; C common; Erg ergative; Gen genitive; Imp imperative; Instr instrument; Lnk linker; Neg negative; Nfut nonfuture; P proper; Perf perfect; Pl plural, Q question particle; Sg singular.

2008, Kim 2011). However, Niuean applicatives do not fully conform to this view, as indicated in (2) and discussed below.

(2)   Generally, an applicative or Appl head:

	Niuean:
a. introduces an argument in its specifier position,	NO
b. assigns this argument its theta role,	PARTLY TRUE
c. case-licenses it as a core internal argument,	USUALLY
d. is high, relating its argument to an event, or low, relating two internal arguments in a possession relation.	NEITHER

In this paper I will argue that the Niuean applicative head *aki* is a relatively high (above v, below Voice) secondary predicate, with a meaning akin to 'use/bring about by means of'. Essentially this has been previously argued for by Ball (2005, 2008) in a lexically-driven analysis (HPSG). Here, I use new data and different arguments to posit, within Minimalist Theory, that *aki* is best analyzed as a secondary predicate. Like some other secondary predicates, it shares two arguments with the primary verb: an affecting external argument (User) and a raised or object-shifted internal argument (Means), which it also case-licenses (cf. Georgala, Paul, and Whitman 2008, Georgala and Whitman 2009, Georgala 2012, Carrier 2014, for related views). Because *aki-* assigns secondary roles to arguments, this analysis has the theoretical consequence that full thematic structure is composed derivationally. In an alternative use, *aki* can also be a clitic (Roberge and Troberg 2010), resuming a relativized instrument or means participant in an otherwise gapless relative clause (Massam 1998, cf. Massam and Roberge 1997), but I will not discuss this use in any detail in this paper.

1.2. Background on Niuean

Niuean is a Polynesian language of the Tongic sub-group (Pawley 1966, Otsuka 2006). It has [V S O PP] word order, ergative case, and isolating morphology, as illustrated in the following examples. As indicated in the glosses, the case markers inflect for the proper or common value of the following nominal.

(3) a.   Ne   nakai kai   [he   tama]   [e   apala]
          Past   not   eat   Erg.C child   Abs.C apple
          'The child did not eat the apple.'   (FN)

    b.   Kua   fehola [e   tau   fānau]
          Perf   flee   Abs.C Pl   children
          'The children fled.' (Sp:123)

My assumptions regarding clause structure for this paper are as follows, illustrated in (4) which shows a transitive sentence. I assume that transitive

agents are merged in the specifier of a peripheral applicative head (Kim 2011), which assigns an Affecting Agent/Causer theta role, along with ergative case. I assume the object is merged as sister to the verb, and that it moves to the specifier of a null light v, which assigns absolutive case. I assume the verb undergoes head movement through v and Appl to INFL to achieve VSO word order. (This is a simplified version of my fuller analyses of Niuean clause structure.[3])

(4)   Basic structure for a transitive clause in this paper (*Sione cut bread*)
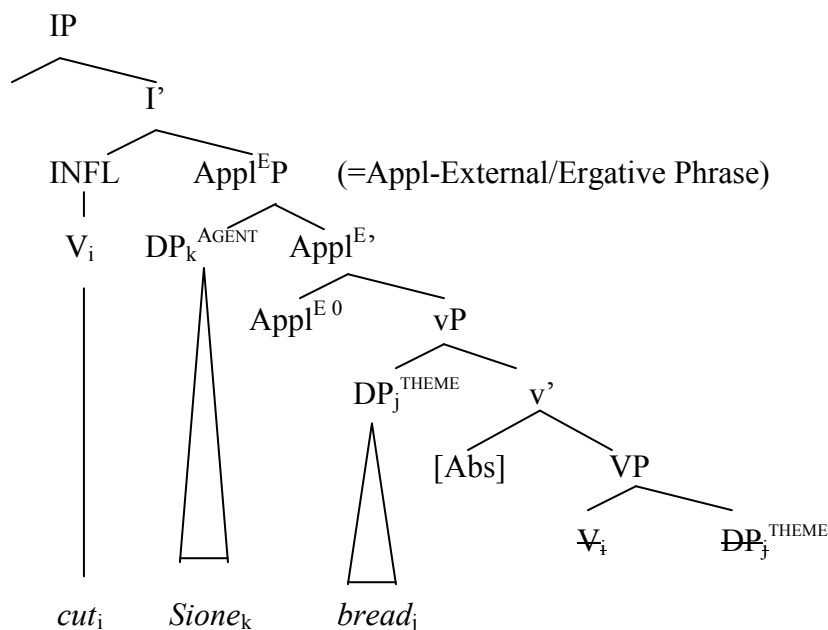
As for unaccusatives, I consider that they are the same, without the peripheral Appl^{E}P. For unergatives (including noun incorporation (NI) sentences) I assume the external arguments are merged in a lower position than transitive agents, in specifier of VP (or a low Vc above VP, but I use Spec VP here for simplicity), receiving a theta role of 'Doer' rather than Affecting Agent, and raising to the specifier of the light v to receive absolutive case (Massam 2009, cf. Tollan 2015). This is shown below. I assume a null cognate object as sister to V in unergatives (Hale and Keyser 1993).

---

[3] In particular, I am assuming head movement rather than vP predicate fronting, and object shift rather than predicate-externally merged arguments, and NI not PNI for this paper, contrary to my usual assumptions, due to space limitations (cf. Massam 2001, 2010).

(5) Basic structure for an unergative/NI clause in this paper
     (*Sione jumped/bread-cut*)

(CO – cognate object, IN = incorporated noun)
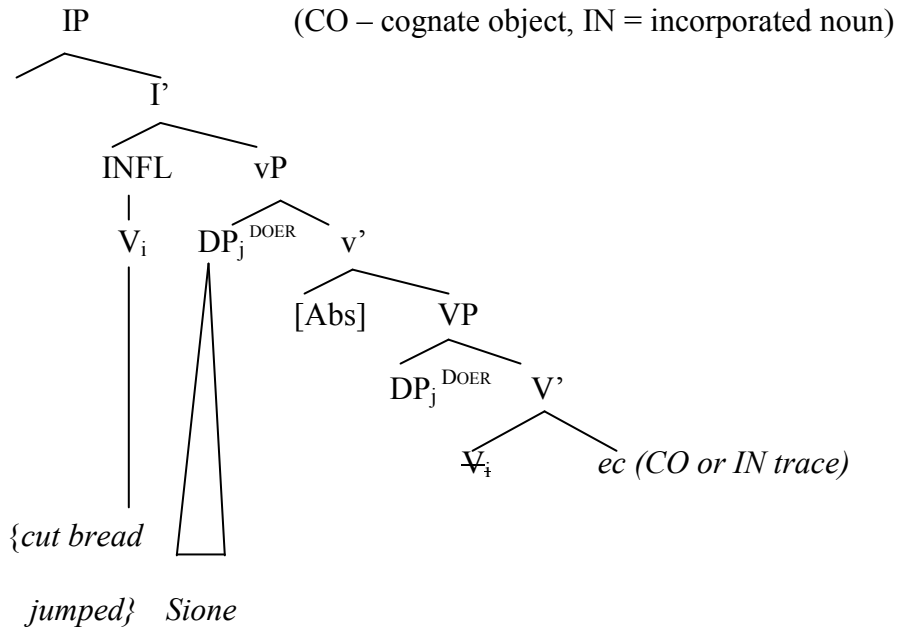
2. Questions Raised by Niuean *aki* Constructions

2.1. Does *aki* introduce a new instrument argument in its specifier?

Let us now address the question whether *aki* introduces a new instrument argument in its specifier, which is considered to be the normal function of an applicative head. At first, the answer would seem to be yes, given (1b), where there is a seemingly new argument bearing an instrument role. Here, it looks like a fairly standard high applicative. However, the applicative marker *aki* has another function, as in (6) (Gould et al 2009).

(6) Kua fakatotō   aki e   au   e   tau tupe   a   ia
     Perf cause-carry Instr Erg.P 1.Sg Abs.C Pl   money Abs.P   3.Sg
     'I gave him some money.'/ 'I made him carry some money.' (Sp:324)

In such cases, *aki* is used in causative constructions, to license a third argument. Here, the answer to the question above would seem to be no, because in such sentences, all of the arguments are thematically linked to the verb (or in case of the Causer, to the causative prefix *faka-* 'CAUS'). The argument *I* is the Causer, the argument *him* is the Doer of *carry* and the argument *money* is the Theme of *carry*. So here, the argument related to *aki* bears a theme role, not an instrument role, and it would appear to get this role from the root verb.

188

In fact, in such sentences, *aki* looks somewhat like a low applicative, in that the two internal arguments are related by possession (Pyllkanen 2002, 2008), and overall the sentence is like a double object construction, with the translation 'give'. However, morpheme order suggests that in (6), the verb is below *aki*, not above it, as a low Appl analysis would suggest. (See Georgala, Paul, and Whitman 2008, Georgala and Whitman 2009, Georgala 2012, and Carrier 2014 for similar arguments in other languages.) A low Appl structure would be as in (7a), but the morpheme order of (6) would suggest a structure as in (7b).

(7) a. Low Appl Structure
        (Pyllkänen 2002)
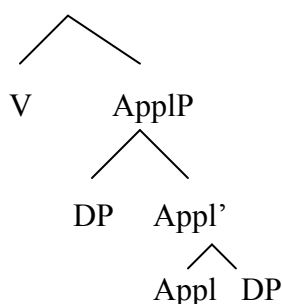
b. Niuean structure
        (left adjunction of head)

```
         /\                              /\
        /  \                            /  \
       V   ApplP                    Vᵢ+aki  VP
           /\                              /\
          /  \                            /  \
        DP   Appl'                      DP   V'
             /\                              /\
            /  \                            /  \
         Appl  DP                         V̶ᵢ   DP
```
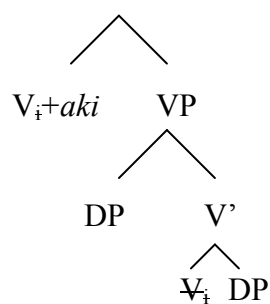
Furthermore, in other causatives with *faka-* the relation between the two internal arguments is not one of possession unlike low applicatives according to Pyllkänen (2002) (as also discussed for Inuktitut by Carrier 2014).

(8)   Fakagagau   aki   <u>e   uga</u>   <u>e   haana   a   lima</u>.
       Cause-bite   with   Abs.C crab   Abs.C 3.Sg.Gen Lnk hand
       'Make the coconut crab bite his hand.' (Sp:98)

Here the relation is one of biting not of possession. In Niuean texts and dictionaries, there are many examples of causative constructions with *aki*. Some examples are given (translated into English, but with Niuean word order) in (9).

(9)   Other examples with causative *faka-*:
       make-keep-awake-with song us, make-block-with truck road, make-hit-with stick head, make-beautify-with flowers house, make-cheer-with his tricks us, make-prettify-with girl flower, make-bite-with crab hand, make-grip-with pliers lid, make squirt-with water garden, make-cool-with ice drink, make-smell-with dog crab, make-plop-with hand water

       In all these cases, the root verb provides the core relation between the two internal arguments; hence it does not appear that *aki* is playing a role in terms of thematic relations. I argue that we can provide a unified analysis of *aki*,

by extending to regular instrumentals the view that the internal arguments are both arguments of the root verb, that is, the higher argument within VP is, in both causatives and instrumentals, a non-volitional subject (whether a causee or an instrument), or inanimate 'agent' of the root verb, thus forming part of an internally transitive VP. In case of causatives, examples include *crab bite hand, pliers grip lid,* and *dog smell crab*, and, in case of non-causatives, examples include *knife cut bread, club slay hero* etc. (cf. Fillmore 1968, Jackendoff 1987 for supporting views about instruments)**.** This argues for a structure like (7b) with V lower than *aki*.

Further support for this structure is found in that the theme of the root verb can undergo noun incorporation, whereas the other argument of the root verb does not (as noted by Carrier 2014 for Inutktitut). If NI targets the V sister (Baker 1988, Massam 2001), this argues that the theme is sister to the verb at merge. I return to the absence of *aki* in (10) later.

(10) a. Volu   niu   oti nakai e   koe   <u>e   tau matā</u>.
         Grate coconut   all Q   Erg.P 2.Sg   Abs.C Pl scraper
         'Did you grate the coconut with all the scrapers?' (S:269)

     b. Ua   fakatakitaki kato   e   koe   <u>haku   tama</u>.
         Neg.Imp cause-hold   basket   Erg.P 2.Sg   1.Sg.Gen child
         'Don't you make my child hold your bags.' (Sp:292)

The structure of the *aki* phrase is thus posited to be along the lines of (11).

(11)   Preliminary proposal for the structure of an *aki* phrase

NVS=Non-volitional subject   (instrument/causee)

aki   VP

          NVS   V'

               V   Theme
use crab   bite   hand
use scraper grate coconut
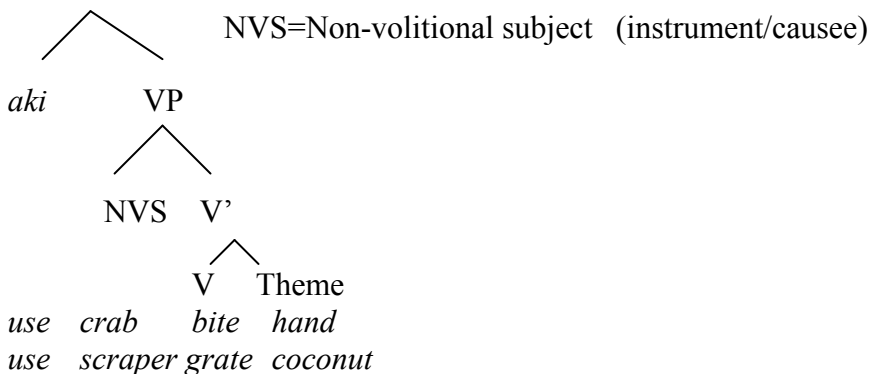
I next argue, following Georgala (2012), that the applicative head *aki* is a raising applicative, which triggers the movement to its specifier of one of the internal arguments of the root verb as in (12).

(12) a. [__ *aki* [NVS V THEME]

 b. [{NVS/THEME} *aki* [NVS V THEME]

As schematized in (12b), this raising analysis allows for an account of the word order variation in Niuean shown in (13): the *aki*-licensed argument is sometimes the NVS and sometimes the Theme. Since, as noted in Footnote 1, Niuean raising to subject and raising to object can freely target the external or the internal argument (Seiter 1980, Massam 1985, Larson, Longenbaugh, & Polinsky 2015), I put aside the apparent locality violation here, simply noting that as the choice of argument has implications for focus, perhaps there is no locality violation, as the closest argument with the relevant focus feature is the one that is targeted for movement. Below, (13a, d) show raising of the NVS, and (13b, c) show raising of the theme.

(13) a. Ne haha aki   e   Sione <u>e   akau</u>   <u>e   kulī</u>.
     Pst hit with Erg.P Sione Abs.C stick   Abs.C dog
     'John hit the dog with a stick.' (FN)

 b. Ne haha aki   e   Sione <u>e   kulī</u>   <u>e   akau</u>.
     Pst hit with Erg.P Sione Abs.C dog   Abs.C stick
     'John hit the dog with a stick.' (FN)

 c. Kua fakaepo   aki   he   matua <u>e   asekulimi</u> <u>e   muke</u>.
     Perf cause-taste with Erg.C mother Abs.C ice-cream Abs.C baby
     'The mother let the baby taste the ice cream.' (Sp:56)

 d. Fakagagau aki <u>e   uga</u>   <u>e   haana   a   lima</u>.
     cause-bite with Abs.C crab   Abs.C 3.Sg.Gen Lnk hand
     'Make the coconut crab bite his hand.' (Sp:98)   (repeated from (8))

I consider that the other argument is licensed by absolutive from v, which is between *aki* and the main verb (not shown in (7) and (11)).

(14) a. [THEME *aki* [NVS v [~~NVS~~ ~~V~~ ~~THEME~~]

 b. [NVS *aki* [THEME v [~~NVS~~ ~~V~~ ~~THEME~~]

Aside from focusing the *aki*-licensed argument, another interpretive effect of the constructions seems to be that the argument licensed by v acts as the true affected argument (sentential theme) of the entire complex predicate so the readings for (13) are more or less like *The dog got hit with the stick, The stick got used in hitting the dog, She fed the baby with the ice-cream,* and *His hand got bitten by the crab.* This sentential theme role is determined derivationally. If

the other movement pattern had occurred, the readings would be the reverse for each sentence.

In summary, *aki* is a raising element. Georgala (2012) considers such applicatives to be expletives, thematically vacuous elements that serve only to case-license an argument that is already theta-licensed. Is this the case for *aki*?

2.2.   Is *aki* always and only a case-licenser?

The sentence below demonstrates the answer to this question is no, because *aki* can be present even if not necessary for case, with only one internal argument.

(15) a. Ne   hopo   aki   e   ia   e   kave   toua
        Pst   jump   with   Erg.P 3.Sg   Abs.C cord   rope
        'She jumped with a rope.' (FN)

     b. Ne   hele   falaoa aki   e   ia   e   sisipi.
        Past   cut   bread with   Erg.P 3.Sg   Abs.C   knife
        'He cut bread with a knife.'   (FN)

If *aki* is not there in such sentences, the instrument appears to be interpreted as a Theme affected by the action, as in (16) (and (10a)), even though it is still an instrument as well.

(16) Kua tā   fakatino   he   tama   e   malala
     Perf draw   picture   Erg.C child   Abs.P   charcoal
     'The child has been drawing pictures with the charcoal.' (S: 51b)
     (and the charcoal was affected, e.g. all used up)

I hypothesize that the causative construction necessarily requires a sentential Theme so if the internal argument theme is incorporated or unexpressed, the NVS must become the sentential theme and so must be licensed by v, so *aki* does not appear in case of NI or unergatives in causatives, as in (10b) and (17). In regular non-causative NI or unergatives, though, as in (15a, b), *aki* will normally still appear, and the means or instrument argument is not interpreted as a sentential theme, as it is not licensed by v.

(17) Kua   fakalele   e   ia   e   manulele.
     Perf cause-fly Erg.P 3.Sg   Abs.C   bird
     'He made the bird fly.' (Sp:180)

In addition, *aki* does not always assign case, as shown below, where the instrument has been extracted.

(18) e   kave toua [ne fã.e   hopo aki   <u>a   ia</u>]
      AbsC   cord rope [Nfut Prog   jump with   Abs.P 3.Sg]
      'the rope that she is jumping with' (Massam 1998)

Usually, A-bar extraction does not affect case in Niuean: if an object is extracted, the subject remains ergative. The A-bar bound trace is thus case marked and the sentences is construed as transitive, as in (19).

(19) e   kofe   [ne   taute   e   au]
      Abs.C coffee   lNfut make   Erg.P 1.Sg]
      'the coffee that I made' (FN)

In cases such as (18), it appears that *aki* serves as a resumptive clitic encoding instrumental means (cf. Roberge and Troberg 2010, who argue that a Romance applicative head position can be realized as a clitic, also Massam 1998). The relative clause in (18) is thus really gapless, and the sentence is construed as an intransitive, with an absolutive subject. *Aki* can be compared to locative *ai* in this use (Chapin 1974, Massam & Roberge 1997).

     In summary, *aki* assigns (or encodes, as in (18)) a secondary focused role of 'by means of' to the element to which it assigns case, either the NVS or Theme (in contrast to v, which secondarily marks the sentential affected theme). Thus *aki* contributes semantically to the sentence and is not an expletive case-licenser. (20) presents the full structure for *aki* sentences, augmented from (11). One argument raises to specifier of vP and the other to specifier of *aki*P, as determined either by focus features or randomly, by equidistance, and each receives its acquired role in that position, in addition to its core role.

(20) The structure of the *aki-v-V* phrase

       *aki*P
       /\
  _ *aki*   *v*P   acquired thematic roles (Means/AffectedTheme)
         /\
      _   *v'*
        /\
     [abs] VP   core thematic roles (NVS/Theme)
         /\
     NVS   V'
         /\
       V   Theme

__	*use*	__	*crab*	*bite*	*hand*
__	*use*	__	*flower*	*beautify*	*girl*
__	*use*	__	*pliers*	*grip*	*lid*
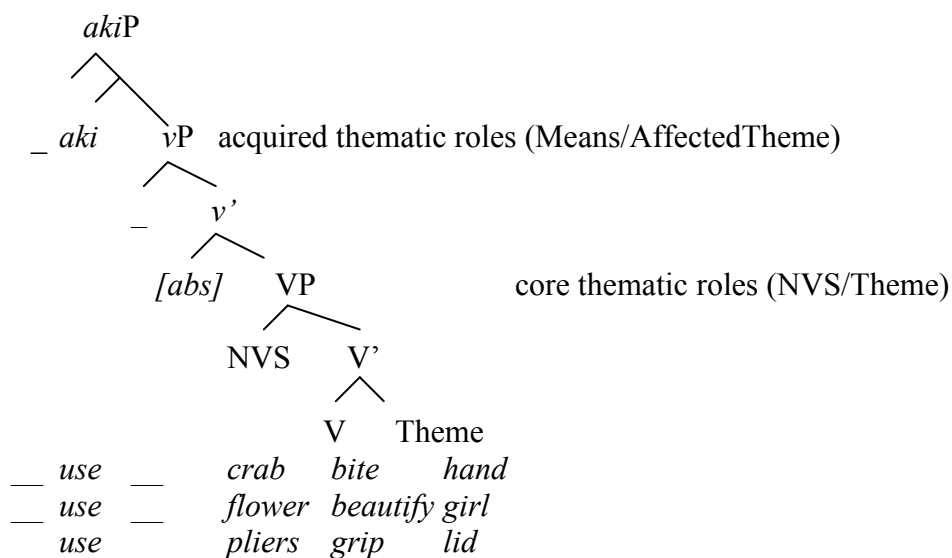
2.3.   The highs and lows of *aki*

Pylkkänen (2002, 2008) argues that low applicatives cannot appear with unergatives, as they require an internal argument because they create a possession relation between the applied argument and the internal argument. Niuean *aki* can appear with unergative verbs, making the verb transitive, as in (21), hence it is not a low applicative. (Note that *tohitohi* "write" is, prior to *aki*, an obligatorily unergative verb, with *tohi* as its transitive counterpart.)

(21)   Ne tohitohi   aki   e   Sione   e   pene.
        Pst writing   with   Erg.P Sione   Abs.C   pen.
        'Sione is writing with the pen.' (Massam 1998, cf. Ball 2008)

However, *aki* does require, in a certain sense, a VP with two arguments, although one can be a null cognate object, as in (22), the assumed structure for unergatives (Hale and Keyser 1993). In the analysis proposed here, however, *aki* does not create a relation between the two arguments in the VP, rather, the Verb already does this.

(22)   Unergative structure
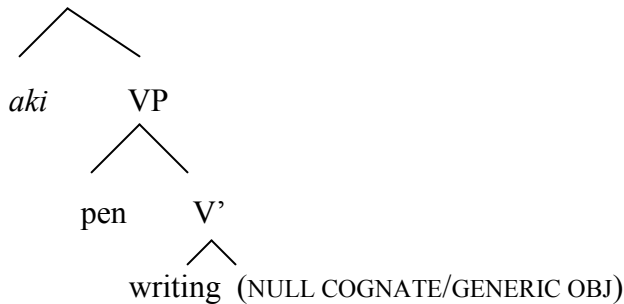
Thus, *aki* is not a low applicative. However, given its use in causative sentences such as (6) and (8), it is clear that it is not a high applicative either, as it does not relate an argument to an event, rather it licenses and augments an argument that is already involved in the event. I have extended this view to include the non-causative sentences such as (1b) as well, considering the instrument in such cases to be a NVS, just like the causee in (6) and (8).

        What *aki* does seem to do, in fact, is create a new relation between the external argument and one internal argument, a relation of 'use' or 'by means of". There are potential counterexamples to this claim, in which an external argument appears not to be required for *aki*. Double unaccusatives, with an NVS do appear with *aki*. An example is in (23) and others (in the form of English translations) are in (23b).

(23) a. Ne fakakofu   aki   <u>e</u>   vaka   <u>e</u>   tau laukou
       Pst cause-cover with   Abs.C canoe Abs.C Pl leaves
       'The canoe is covered with leaves.' (FN, Massam 1998)

     b. Volcanic rock makes-secure the island / These actions make-unpopular the leader / The sickness makes-suffer the people

However, it is notable that all such sentences include the causative prefix *faka-*. From Pyllkanen (2002) we know that causative heads such as *faka-* can merge in an Appl^E head without taking a causer argument as specifer. We can thus maintain our generalization that *aki* relates an external argument (either the CAUS head itself or its specifier) to an internal argument via a 'by means of' relation, that is, it does require a peripheral Appl^E, but this Appl^E can simply express CAUS as in (23), rather than introduce an Agent or Causer. This is shown in (24).

(24)   CAUS may or may not have an argument in its specifier (e.g. (23))

Appl^E/Caus   *aki*P

                 (← vP with Abs is left out of this tree, see (20))

      aki   V'

          NVS   V'

             V   Theme
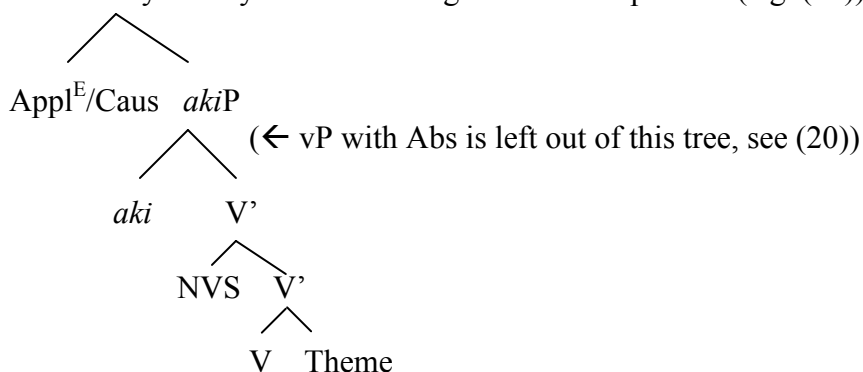
In summary, *aki* does not really show the behaviour of a high or a low applicative, but instead seems to act like a secondary predicate that shares and relates two arguments, by assigning each a secondary role, one (User) to an external argument whether explicit or implicit (agent, causer or CAUS) and the other (Means) to an internal argument (NVS or Theme), via object shift or raising.

3. Context: Niuean has a rich array of secondary predicates

I will briefly put this idea into context. Niuean, like other Oceanic languages, has a rich array of secondary predicates (Bril and Ozanne-Rivierre 2004, Massam 2013) and the claim here is that *aki* is one such secondary predicate (cf. Ball 2008). The full range of secondary predicates is provided in (25) and a hypothetical full predicate is suggested in (26).

(25) Secondary predicates in merge/scope order[4]
    a) <u>Pre-verb</u> + VERB   (Modals, evidentials)
    b) <u>Light verb</u> + VERB   (agentive, stative)
    c) VERB + Aspectual Adverb   (always, immediately)
    d) VERB + Locative/Temporal   (*ai* resumptive clitic)
    e) VERB + <u>Universal Quantifier/Completion</u>   (*oti*)
    f) VERB + Instrumental   (*aki*)
    g) VERB + <u>Secondary Predicate</u>   (some can co-occur)
                                   a. Resultatives
                                   b. Modifiers
                                   c. Depictives
    h) VERB + NP   (Noun Incorporation)
    i) VERB + VERB   (Compound Verbs)

(26)   [want do hit bash dog dead use all there always just Perf Q]
       Did (he) always just want to really hit dogs dead with all (the sticks)
       there?)

Due to successive movements of the verb, which I will not outline here, most of these predicates end up in inverse order, as shown in (27). See Massam (2010, 2013) for details.

(27)   *Surface positions of elements within the predicate*:
       a   b   i   h   g   f   e   d   c
       preV + LtV + VB(+V1) + NP + <u>V2*</u> + *aki* + Quant + Loc + Adv

The main point to note here is that *aki* is adjacent on each side to another secondary predicate in the predicate complex, and that most of these secondary predicates also share one or two arguments with the root verb. It thus is plausible that it too is a secondary predicate. Unlike the others, though, *aki* can case-license an argument, and it does not appear as an independent verb (but see Ball 2005, 2008, who argues that it is verbal even in its prepositional use).

4. Conclusion

In conclusion, in this brief paper I have provided an outline of a unified analysis of *aki* in its causative and instrumental uses, and I have argued that it is neither a high nor a low applicative, as it does not display the usual properties of applicatives, and it does not create a relation between two internal arguments, nor does it relate an argument to an event. Instead, I have argued that it is a

---

[4] Note *aki* and *oti* appear in both orders, due, I think, to the fact that *oti* can be a completion resultative (over the event), or a quantifier over an argument. There are other internal-order questions or variations, but I won't deal with them here.

transitive secondary predicate (Ball 2008), which assigns a secondary role (User) to the external argument (explicit or implicit), and a secondary role (Means) via object shift and case-licensing, to one internal argument in an internally transitive VP clause (cf. Georgala 2012, Carrier 2014). In addition, I noted that *aki* can also directly encode Means itself, as a resumptive clitic in a relative clause (Chapin 1974, Massam and Roberge 1997, Massam 1998)**.**

References

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing.* Chicago: University of Chicago Press.

Ball, Douglas. 2005. A unified analysis of Niuean *aki*. *Austronesian and Theoretical Linguistics*, eds. Raphael Mercado, Joey Sabbagh, and Lisa Travis, 249–270. Amsterdam: Benjamins.

Ball, Douglas. 2008. *Clause structure and argument realization in Tongan.* Ph.D. dissertation, Stanford University.

Bril, Isabelle, and Françoise Ozanne-Rivierre (eds.). 2004. *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundedness.* Berlin: Mouton de Gruyter.

Carrier, Julien. 2014. Redefining the low applicative: Evidence in Inuktitut. Ms., University of Toronto, Toronto, On.

Chapin, Paul. 1974. Proto-Polynesian *ai. *Journal of the Polynesian Society* 83:259–307.

Chung, Sandra. 1978. *Case marking and grammatical relations in Polynesian.* Austin: University of Texas Press.

Cuervo, Maria-Cristina. 2003. *Datives at large.* Doctoral dissertation, MIT.

Fillmore, Charles. 1968. The case for case. *Universals in Linguistic Theory*, eds. Emmon Bach, and Robert Harms, 21–294. New York: Holt, Rinehart and Winston.

Georgala, Effi, Waltraud Paul, and John Whitman. 2008. Expletive and thematic applicatives. *Proceedings of West Coast Conference on Formal Linguistics 26*, eds. Charles B. Chang, and Hannah J. Haynie, 181–189. Somerville: Cascadilla Press.

Georgala, Effi, and John Whitman. 2009. Ditransitives and applicative structure in Greek. In *Chicago Linguistic Society 43:1*, eds. N. Adams, A. Cooper, F. Parrill, and T. Wier, 77-91. Chicago: Chicago Linguistics Society.

Georgala, Effi. 2012. *Applicatives in their structural and thematic function: A Minimalist account of multitransitivity.* Doctoral dissertation, Cornell University.

Gould, Isaac, Diane Massam, and Philip Patchin. 2009. Faka-Niue: Understanding cause in Niuean. *Toronto Working Papers in Linguistics* 31, University of Toronto, Toronto, On.

Hale, Ken, and Samuel Jay Keyser. 1994. On Argument structure and the lexical

expression of syntactic relations, eds. Ken Hale, and Samuel Jay Keyser, 53–109. *The View from Building 20*. Cambridge: MIT Press.

Jackendoff, Ray. 1987. *Consciousness and the computational mind*. Cambridge: MIT Press.

Kim, Kyumin. 2011. *External argument introducers*. Doctoral dissertation, University of Toronto.

Larson, Bradley, Nicholas Longenbaugh, and Maria Polinksy. 2015. Subject-object parity in Niuean and the labeling algorithm. Paper presented at West Coast Conference on Formal Linguistics. Simon Fraser University.

Marantz, Alec. 1993. Implications of asymmetries in double object construction. *Theoretical Aspects of Bantu Grammar*, ed. Samuel Mchombo, 113–150. Stanford, CA: CSLI Publications.

Massam, Diane. 1985. *Case theory and the projection principle*. Doctoral dissertation, MIT.

Massam, Diane. 1998. *Aki* and the nature of Niuean transitivity. *Oceanic Linguistics* 37:12–28.

Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language and Linguistics Theory* 19:153–197.

Massam, Diane. 2009. The structure of unergatives. *Proceedings of the Austronesian Formal Linguistics Association 16*, eds. Sandy Chung, Daniel Finer, Ileana Paul, and Eric Potsdam, 125–135. London: University of Western Ontario.

Massam, Diane. 2010. Deriving inverse order. *Austronesian and Theoretical Linguistics*, eds. Raphael Mercado, Joey Sabbagh, and Lisa Travis, 271–296. Amsterdam: John Benjamins Publishing Company.

Massam, Diane. 2013. Nuclear complex predicates in Niuean. *Lingua* 135:56–80.

Massam, Diane, and Yves Roberge. 1997. Operator-bound clitics and Niuean *ai clitics, pronouns and movement,* eds. James Black, and Virginia Motapanyane, 273–300. Amsterdam: John Benjamins Publishing Company.

McGinnis, Martha. 2008. Applicatives. *Language and Linguistic Compass* 2:6, 1225–1245.

Niue: A History of the island. 1982. Alofi, Niue: Institute of Pacific Studies of the University of the South Pacific and the Government of Niue.

Otsuka, Yuko. 2006. Niuean and Eastern Polynesian: A view from syntax. *Oceanic Linguistics* 45:429–56.

Pawley, Andrew. 1966. Polynesian languages: A subgrouping based on shared innovations in morphology. *Journal of the Polynesian Society* 75:39–64.

Pylkkänen, Liina. 2002. *Introducing arguments*. Doctoral dissertation, Cambridge: MIT.

Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge: MIT Press.

Roberge, Yves, and Michelle Troberg. 2010. The high applicative syntax of the

dativus commodi/incommodi in Romance. *Probus* 21:249–289.

Seiter, William. 1980. *Studies in Niuean syntax*. New York: Garland.

Sperlich, Wolfgang. 1997. *Tohi Vagahau Niue: Niue language dictionary*. Government of Niue and the University of Hawaiʻi Press.

Tollan, Rebecca. 2015. Unifying transitivity and (un)ergativity: A view from Samoan. Ms., University of Toronto, Toronto, On.

AN ACOUSTIC INVESTIGATION OF JAVANESE STOP CONSONANT CLUSTERS

Marc Matthews
University of Florida
mrmarc@ufl.edu

Previous research of Javanese stop consonants has shown that the phonemic contrast is realized phonetically with stiff- and slack voice on the following vowel in CV syllables with simple onsets. Little is known about the phonetic realization of stiff- and slack voice on vowels in complex CCV syllables in which the vowel is separated from the stop by an intervening segment. In the present study, simple and complex stiff- and slack voiced syllables ([pV, plV] and [bV, blV], respectively) containing the vowels [a], [i], and [u] are analyzed in terms of F0, F1 and F2 frequency, H1-H2 and H1-F2 amplitudes, the cues most commonly identified in the literature as signally stiff- and slack voice. Vowels following complex slack- and stiff-voice stop onsets show the same types of acoustic markers as their simple CV counterparts. Slack voiced vowels show characteristically lower F0 and F1 frequencies than stiff voiced vowels, which is taken to be an indicator of larynx lowering. Differences in H1-H2 and H1-F2 amplitudes are inconsistent for [a], but follow predicted patterns for the high vowels. In addition to vowels, lateral approximants following stiff- and slack voice stops also show the acoustic characteristics of stiff- and slack voice, respectively. While the differences between stiff- and slack voice appear distinct in Javanese, modal voiced lateral approximants do not occupy a consistently intermediate position relative to the two nonmodal voice qualities. The lack of a phonemic contrast between modal and nonmodal segments in Javanese is proposed to account for this observed phenomenon.

1.   Introduction

Javanese stop consonants are phonetically unique among the Austronesian languages. Unlike more familiar voiced and voiceless consonants, Javanese stop consonants do not contrast in terms of aspiration or Voice Onset Time (VOT), e.g. the timing of the onset of vocal fold vibration relative to the release of the stop closure. Rather, Javanese stop consonants differ in terms of laryngeal tension, also known as *phonation type* or *voice quality*. Stops represented orthographically with <p,t,th,c,k> are produced with *stiff-voice*; those written as <b,d,dh,j,g> are produced with *slack-voice*. Earlier studies of Javanese stop consonants used terms such as 'light' and 'heavy', 'tense' and 'lax', and 'clear' and 'breathy' to refer to stiff-voice and slack-voice, respectively. However, following the descriptions of voice quality and laryngeal settings of Ladefoged and Maddieson (1996), the terms 'stiff' and 'slack' are thought to be the most appropriate. Furthermore, Thurgood (2004) demonstrated that true breathy voice

is used in Javanese as a form of emphasis, whereas stiff voice and slack voice are used to contrast stop consonants.

Research beginning with Fagan (1988) has shown that the phonemic contrasts of Javanese stiff- and slack voiced stops are realized phonetically on the following vowel. Previous studies have shown that the most reliable acoustic cues of stiff and slack voice are relative pitch and formant frequencies, as well as differences in harmonic amplitudes. The data used in these studies have typically come from simple CV syllables in which the vowel carrying the acoustic cues for phonation type immediately follows the stop segment. The present study asks whether and to what extent these same acoustic cues can be seen in vowels following stop consonant cluster onsets.

2.   Nonmodal Phonation Types

There is a continuum of laryngeal tension within which phonation is possible. Ladefoged and Maddieson (1996) describe this continuum as containing five distinct laryngeal settings ranging from creaky voice to breathy voice. When the vocal folds vibrate at their tensest setting, it is perceived as creaky (or laryngealized) voice; when the vocal folds vibrate at their loosest setting, it is perceived as breathy voice. Modal (or clear) voice is the default setting for voicing, wherein the vocal folds are neither especially tense nor lax (Laver 1980). Stiff voice and slack voice are intermediate settings between creaky, modal, and breathy voice, respectively. Voicelessness lies on either side of the continuum. When the vocal folds are completely taut and closed, as in a glottal stop, voicing does not occur. Similarly, voicing does not occur when the vocal folds are completely open, as in a voiceless glottal fricative.

2.2.   Acoustic Correlates of Stiff and Slack Voice

Javanese stop (and affricate) consonants are produced with either stiff- or slack voice. The two nonmodal phonation types do not contrast phonemically with modal voice in Javanese. That is, there are no modal voiced stops in Javanese, nor are there stiff- or slack voiced fricatives, nasals or approximants. The physical mechanisms and thus the acoustic cues that signal a particular phonation type are varied, and often involve a combination of cues (Blankenship 2002, Wayland and Jongman 2003). I summarize Laver's (1980) and Ladefoged and Maddieson's (1996) descriptions of stiff- and slack voice as follows. In stiff voice, the vocal folds are slightly tenser and held closer together than in modal voice. Additionally, the larynx is raised in the throat. The perceptual consequence of this phonation type is that stiff voiced segments sound slightly laryngealized and higher in pitch than modal segments. Conversely, slack voiced segments are produced with the vocal folds held slightly looser and more open than in modal voice, and with the larynx lowered in the throat. As a result, slack

voiced segments are perceived to be slightly breathy and lower in pitch than modal segments.

Several acoustic effects result from the differing mechanisms involved in stiff- and slack voice phonation types. The predicted acoustic effect of larynx lowering is the lowering of the fundamental (F0) and formant frequencies, whereas larynx raising has the opposite effect (Laver 1980). The findings of Fagan (1988) and Thurgood (2004) support this prediction in part. Both studies show that F0 and first formant (F1) frequencies are lower in vowels following slack voiced stops compared to stiff voiced stops. Surprisingly, the second formant (F2) frequency of slack voiced vowels was often found to be *higher* than those in stiff voiced vowels. A higher F2 in slack voiced segments is unexpected and cannot be explained by the effect of larynx lowering.

As laryngeal tension decreases, the vocal folds remain open for larger portions of a single open-closed cycle, which boosts the spectral energy of the first harmonic (Wayland and Jongman 2003). The open quotient is the ratio of the open phase of the vocal folds to a complete cycle (Blankenship 2002), and is thought to be larger for slack and breathy voice, and smaller for stiff and creaky voice. As the open quotient increases (e.g. in slack voice compared with stiff voice), the spectral energy in the first harmonic increases and energy in the higher harmonics decreases. Therefore, the difference in amplitudes between the first and second harmonics (H1-H2) is expected to be higher for slack voiced segments compared to stiff voiced segments.

Spectral slope—a measure of the relative concentration of acoustic energy in a waveform—is also often used as a cue to phonation type (Blankenship 2002). Due to the boost in H1 and dampening of the higher harmonics in slack voice, the difference in amplitudes of the first harmonic and the second formant (H1-F2) will be higher for slack voiced segments compared to stiff voiced segments. That is, slack voiced segments are expected to have a steeper spectral slope than stiff voiced segments.

In summary, the perception of nonmodal phonation types such as stiff- and slack voice likely arise from a combination of acoustic cues, rather than from a single cue (Blankenship 2002). The most commonly used measures of nonmodal phonation are fundamental frequency (F0), the first and second formant frequencies (F1, F2), as well as the difference in amplitudes between the first and second harmonics (H1-H2) and the first harmonic and second formant (H1-F2). Of these, F1 frequency and H1-F2 appear to be the most robust acoustic cues across studies, with slack voiced exhibiting lower F1 and higher H1-F2 compared with stiff voiced segments (Fagan 1988, Ladefoged and Maddieson 1996, Laver 1980, Thurgood 2004, Wayland and Jongman 2003). This information is summarized in Table 1 below. Modal voiced segments are expected to be intermediate to stiff- and slack voice in all of these measurements, as modal voice represents the default setting of vocal fold vibration, neither tense nor lax.

Table 1. Expected characteristics of stiff- and slack-voice
(adapted from Ladefoged and Maddieson 1996, Blankeship 2002)

	Stiff voice <p>	Slack voice <b>
A. Articulatory characteristics:		
	- raised larynx	- lowered larynx
	- vocal folds slightly tense and close	-vocal folds slightly lax and open
	- lower open quotient	- higher open quotient
B. Auditory characteristics:		
	- slightly laryngealized	- slightly breathy
	- higher pitch	- lower pitch
C. Acoustic characteristics:		
a. F0 and Formant frequencies:		
	- higher	- lower
b. H1-H2 (open quotient) and H1-F2 (spectral slope):		
	- lower	- higher

3.   Research Questions

The analyses and conclusions regarding the acoustic correlates of stiff- and slack voice in Javanese stops cited in section 2 were done on vowels following simple onsets, e.g. CV syllables. This study represents the first step in studying the acoustic correlates of slack- and stiff voice in Javanese syllables with complex onsets. Javanese phonotactics allow prenasalized stop clusters, as well as stop-approximant clusters as onsets. This study looks specifically at complex onsets consisting of a bilabial stop followed by a lateral approximant, and compares those to simple onsets of stiff- and slack voiced bilabial stops and modal voiced lateral approximants. I attempt to answer the following two research questions. First, to what extent can the effects of stiff- and slack voiced stops be measured in vowels of ClV syllables, in which a lateral approximant separates the stop and vowel? Second, can the acoustic effects of stiff- and slack voiced stops be observed in the intervening lateral approximant in ClV syllables?

Bilabial stops were chosen over lingual stops because of the possible confounding effects that a lingual gesture might have on the actions of larynx raising or lowering. Labial gestures are not expected to effect the tension surrounding the larynx or the ability to raise or lower it. Lateral approximants were chosen as the intermediate segment in target syllables for two reasons. First, lateral approximants have a rather well defined harmonic structure (Reetz and Jongman 2009), in which it may be possible to measure the effects of stiff- and slack voice. Second, lateral approximants do not generally block the spreading of articulatory gestures or features. In light of these observations, it is predicted that the same acoustic cues for stiff- and slack voice will be present to similar degrees in vowels following simple and complex onsets of the same

type. Furthermore, with regard to the second research question, if stiff- and slack voice phonation types act to raise or lower frequencies or amplitudes in vowels, then that effect should also be observed in an intervening lateral approximant. Although lateral approximants are produced with modal voice in Javanese, given the right environment (e.g. following a stiff- or slack voiced stop) I predict that they will also be produced with stiff- or slack voice.

Blankenship 2002 raises the question of whether the phonemic status of a nonmodal segment influences how it is realized phonetically. In Mazatec, which contrasts creaky, modal and breathy segments phonemically, all three phonation types occupy distinct phonetic spaces. However, in Tagalog, in which the nonmodal phonation types occur as allophones, no such clear distinction was found. Javanese presents an interesting case because the nonmodal phonation types, stiff voice and slack voice, contrast phonemically with each other within the stop series, but do not contrast with any modal segment. That is, Javanese has stiff voiced stops and slack voiced stops, but not modal voiced stops. Therefore, of additional interest to the present study is the relationship of modal voice to stiff- and slack voice in Javanese lateral approximants. Based on Ladefoged and Maddieson's (1996) continuum of laryngeal tension, it is predicted that modal voiced laterals (e.g. word initially) will exhibit characteristics intermediate to lateral approximants following stiff- and slack voiced stops.

4. Methodology

The data reported in this study were elicited from one adult female speaker of the Central Javanese dialect. A word list (Appendix A) containing 45 target words was used to elicit the data. Target words began with one of the onset sequences [bl], [pl], [p], [b], [l], followed by one of the point vowels [a], [i], [u]. The list contained three target words for each onset and vowel combination. The target word list was randomized and divided into three pages. Two distractor words were added to the beginning and end of each page (12 words total) in order to reduce the presence of page-turning noise, starting- or ending-intonation, or other undesirable prosodic and acoustic effects. The consultant confirmed that all of the words on the list were commonly used words. She sat in a quiet room and said each word twice inside the carrier phrase *aku kondo ___ ping loro* 'I say ___ two times'. There were six tokens for each consonant (cluster) and vowel combination, except for [plu]. Due to an oversight, one of the [plu] words elicited (*plung* 'something small falling into liquid') was not produced with the target [u] because of a phonological laxing rule on closed syllables. As a result, there were only four tokens in the [plu] set.

Target words were extracted and analyzed using the Praat software. Measurements of F0, F1, F2, H1-H2 and H1-F2 were taken over a 25ms interval starting at the onset of each vowel and lateral approximant. Following Fagan's

(1988) and Thurgood's (2004) criteria, a vowel onset was considered to be the onset of voicing and identifiable first and second formant bands following a stop release. These criteria were also used to identify the onset of a lateral approximant in a stop-lateral onset cluster. In the case of vowels that followed lateral approximants, the vowel onset was identified as the steady state portion of voicing following the transition from the lateral. For consistency, all onset measurements were taken from the nearest zero point of the waveform at the onset. Frequencies (Hz) were measured using wide-band spectrograms. Harmonic amplitudes (dB) were measured using 25.6ms (Hamming) FFTs.

Results of the data analysis are presented in the next section. It was determined that inferential statistical measures would be inappropriate for the current data set, due to its relatively small size and in light of the fact that it was obtained from a single speaker. Therefore, the data presented in the following section summarize the observed trends in general terms.

5.   Stiff- and Slack Voice in Vowels

The first question investigated in the present study is whether and to what extent the acoustic correlates of stiff- and slack voice are observed in vowels following complex stop cluster onsets. To that end, acoustic measurements described in section 2.2 and section 4 were compared for vowels following simple and complex stiff voiced onsets (e.g. [pV] and [plV]) and following simple and complex slack voiced onsets (e.g. [bV] and [blV]).

Figures 1 and 2 summarize the mean vowel frequencies for slack- and stiff voiced vowels, respectively. Based on the effects of larynx lowering associated with the production of slack voice, vowel frequencies of slack voiced vowels are predicted to be lower than stiff voiced vowels, regardless of onset type. Within each group of stiff voiced onsets and slack voiced onsets, slight differences were observed between simple and complex onsets. However no consistent pattern emerged within that variation, e.g. vowels following complex slack voiced onsets were not consistently higher or lower in frequency than simple slack voiced onsets for any of the three frequency measures. Furthermore, stiff voiced onsets and slack voiced onsets occupy distinct, non-overlapping regions in terms of F0 and F1 frequency. This trend is observed in the mean vowel frequencies of all three vowels collapsed together (figures 1a, 2a) and for each vowel individually (figures 1b, 2b), but is most apparent when comparing the mean vowel frequencies of slack voiced [a] vowels with those of stiff voiced [a] vowels. Mean vowel F0 and F1 are lower following slack voiced onsets compared to those following stiff voiced onsets. Mean F2 frequencies for slack- and stiff voiced vowels are not as clear-cut. Consistent with previous findings (Fagan 1988, Thurgood 2004), F2 of back vowels are generally higher following slack voiced onsets than stiff voiced onsets. Looking at the F2 frequencies for each of the three vowels, there is considerable overlap in the

range of F2 frequencies between slack- and stiff voiced vowels, particularly in [a], and F2 is nearly identical for [i] following both slack- and stiff voiced onsets.

Figure 1.
a. Mean vowel frequencies (Hz): Slack voice

	[a]	[i]	[u]
■ Simple [bV]	-0.13	13.47	7.72
◪ Complex [blV]	-1.30	9.98	4.27
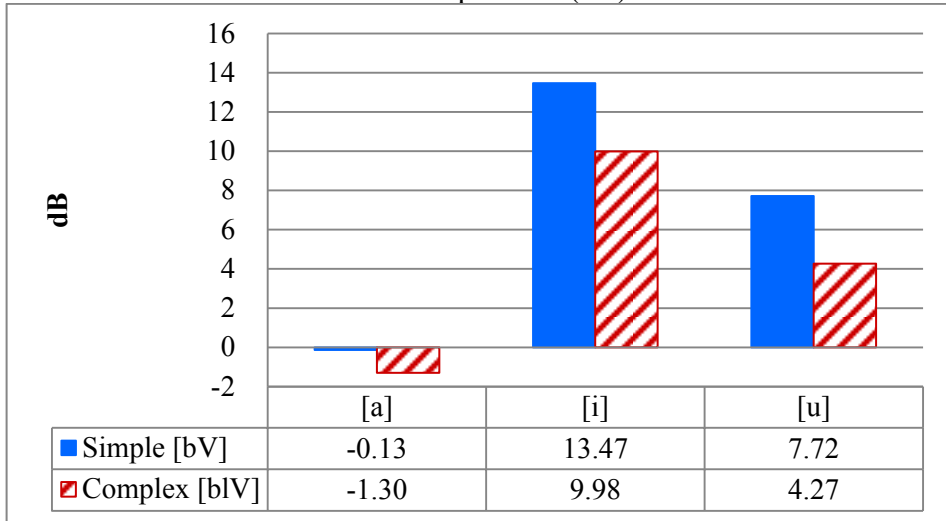
b. Slack voice vowel frequencies (Hz): Breakdown by vowel and onset type

Vowel	Onset type	F0	F1	F2
[a]	Simple [bV]	207.66	614.18	1489.51
	Complex [blV]	198.42	706.17	1899.68
[i]	Simple [bV]	219.45	291.49	2479.16
	Complex [blV]	222.21	340.84	2476.46
[u]	Simple [bV]	220.97	353.53	1691.49
	Complex [blV]	215.71	365.93	1419.17

Figure 2.
a. Mean vowel frequencies (Hz): Stiff voice

	[a]	[i]	[u]
■ Simple [pV]	2.98	19.10	19.58
▨ Complex [plV]	2.70	21.82	20.43
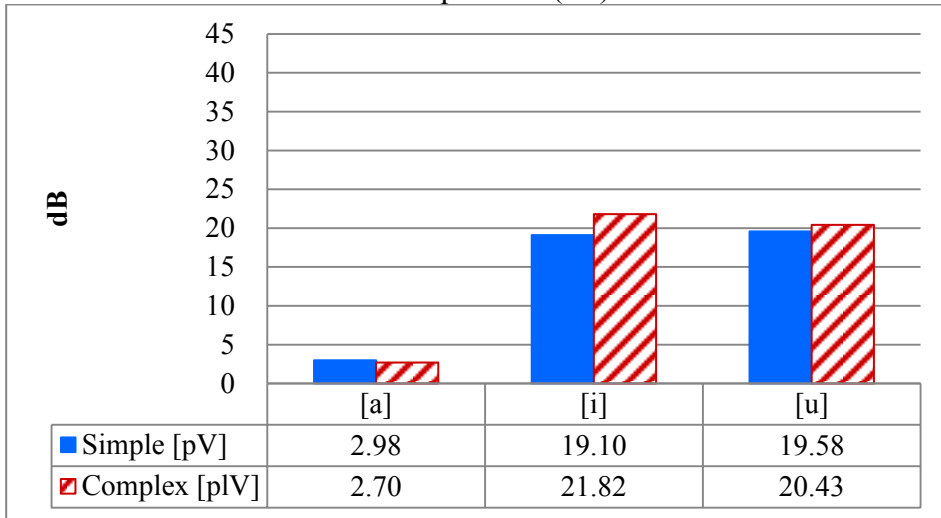
b. Stiff voice vowel frequencies (Hz): Breakdown by vowel and onset type

Vowel	Onset type	F0	F1	F2
[a]	Simple [pV]	237.06	884.50	1606.40
	Complex [plV]	228.08	903.39	1648.01
[i]	Simple [pV]	239.95	366.71	2674.00
	Complex [plV]	239.45	387.18	2453.40
[u]	Simple [pV]	238.41	457.53	823.37
	Complex [plV]	253.78	439.35	1284.51

Differences in H1-H2 amplitudes for slack- and stiff voiced vowels are summarized in figure 3. It was predicted that H1-H2 would be greater (e.g. a higher number) for slack voiced vowels than stiff voiced vowels, regardless of onset type, based on the prominence of spectral energy in H1 which results from the higher open quotient in slack voice (Blankenship 2002, Wayland and Jongman 2003). This prediction is supported by the mean vowel H1-H2 for all three vowels together. There is some variation in the range of H1-H2 for simple and complex onsets for each phonation type, however two distinct clusters emerge for averaged slack voiced onsets and stiff voiced onsets. Looking at each vowel individually, the predicted pattern (e.g. slack > stiff) is most readily observed for [u]. H1-H2 amplitudes for [i] form distinct groups to a lesser extent. The predicted pattern is not observed for [a], in which H1-H2 of slack voiced [a] is lower than stiff voiced [a].

Figure 3.
Vowel H1-H2 (dB): Comparison of slack and stiff voice

| Vowel | Onset Type | Phonation type | |
		Slack [b-]	Stiff [p-]
mean	Simple	10.53	3.79
	Complex	6.48	3.15
[a]	Simple	-0.13	-0.77
	Complex	-1.30	0.05
[i]	Simple	13.47	8.92
	Complex	9.98	4.70
[u]	Simple	7.72	-0.57
	Complex	4.27	1.55

The final measure of phonation type considered in this study is that of spectral slope (H1-F2), summarized in figure 4. Similar to the predictions for open quotient H1-H2 above, spectral slope was predicted to be higher for slack voiced vowels than in stiff voiced vowels, based on the prominence of H1 and dampening of spectral energy in the higher frequencies occurring in slack voice. Again, this prediction is supported by the mean H1-F2 for all three vowels together, and for the high vowels [i] and [u]. Here we can observe two distinct groups for slack voiced vowels and stiff voiced vowels. There is some variation between H1-F2 amplitudes for simple and complex onsets of the same type, yet despite this range the two groups do not overlap for the high vowels, and there is no consistent pattern in the variation between onset types. The predicted pattern is again not observed for [a], where the mean H1-F2 for complex slack voiced onsets (e.g. [bla]) falls within the range of simple and complex stiff voiced onsets (e.g. within the range for [pa] and [pla]).

Figure 4.
Vowel H1-F2 (dB): Comparison of slack- and stiff voice

| Vowel | Onset Type | Phonation type | |
		Slack [b-]	Stiff [p-]
mean	Simple	39.55	20.83
	Complex	30.45	22.47
[a]	Simple	9.57	2.98
	Complex	2.73	2.70
[i]	Simple	28.70	19.10
	Complex	29.05	21.82
[u]	Simple	40.83	19.58
	Complex	29.12	20.43

In summary, vowel measurements were taken for stiff- and slack voiced vowels in terms of fundamental frequency, formant frequencies, and harmonic amplitudes. As predicted, slack voiced vowels had lower F0 and F1 values than stiff voiced vowels. This was particularly true for [a]. F2 values were higher on average for slack voiced values, consistent with previous research yet inexplicable by the assumed effect of larynx lowering. Differences in harmonic amplitudes (H1-H2 and H1-F2) also supported our predictions, such that the average H1-H2 and H1-F2 for all three vowels together was higher for slack voice than stiff voice. However, the prediction was not borne out for [a], in which H1-H2 was lower for slack voice than stiff voice, and H1-F2 overlapped for the two phonation types.

5.2.   Stiff- and Slack Voice in Lateral Approximants

The second aim of the present study is to investigate whether and to what extent stiff- and slack voice can be observed in lateral approximants that occur after stiff- and slack voiced stops, respectively. To that end, the acoustic measurements described above were compared for lateral approximants in stiff-, slack- and modal voiced onsets (e.g. the [l] in [plV], [blV] and [lV] syllables). Based on the observations that laterals contain a relatively well-defined formant structure and do not typically block feature spreading, it was predicted that the same acoustic cues observed in stiff- and slack voiced vowels (e.g. relative changes in F0, F1, F2, H1-H2 and H1-F2) would also be observed in so-called stiff- and slack voiced laterals. Furthermore, based on the proposed continuum of laryngeal settings (Ladefoged and Maddieson 1996), it was predicted that modal [l] would exhibit intermediate values to stiff- and slack voiced [l].

The mean F0, F1 and F2 frequencies of lateral approximants for the three phonation types are shown in figure 5a for all three following vowels together. As was predicted and observed for vowels (section 5.1), F0 and F1 of slack voiced [l] is lower than F0 and F1 of stiff voiced [l]. Also in line with the current findings for vowels, F2 of slack voiced [l] is higher than F2 of stiff voiced [l]. Modal [l] does show intermediate average values to stiff- and slack voiced [l], however modal [l] does not appear to occupy its own distinct space in any of these three measures. In terms of F0, modal [l] and slack voiced [l] are nearly identical. Similarly, modal [l] and stiff voiced [l] are nearly identical in terms of F1 and F2. This trend becomes more apparent when looking at the identity of the following vowel (figure 5b). We see that when [l] precedes [a], the F0, F1 and F2 frequency values follow the predicted pattern (e.g. slack [l] is lower than modal [l], which is lower than stiff [l]). However, when preceding [i], F1 of modal [l] is higher and F2 is lower than both stiff- and slack voiced [l]. Similarly, when [l] precedes [u], F0 and F2 of modal [l] are lower and F1 is higher than both stiff- and slack voiced [l].

Figure 5.
a. Mean lateral approximant frequencies (Hz)

	F0	F1	F2
■ Slack [blV]	216.76	352.45	1902.31
◪ Modal [lV]	217.03	463.24	1722.89
◩ Stiff [plV]	240.08	471.95	1717.27
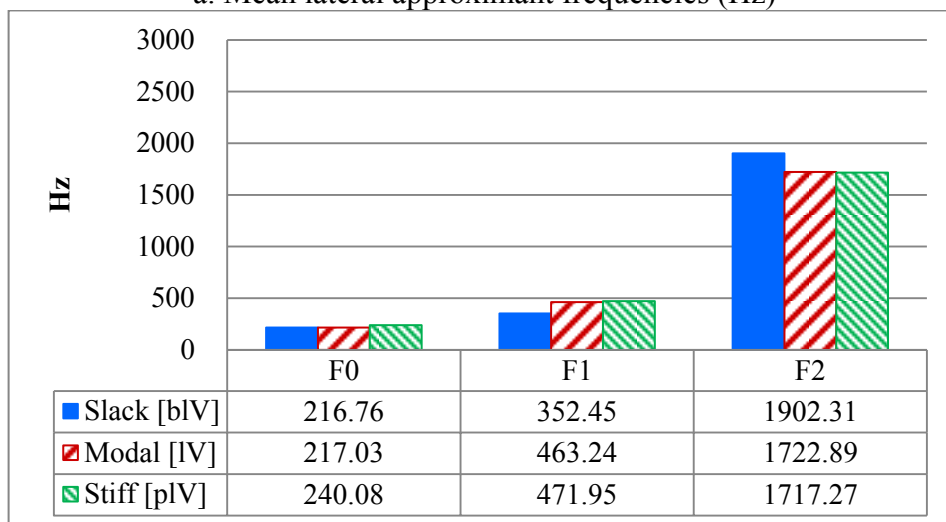
b. Lateral approximant frequencies (Hz): Breakdown by phonation type and following vowel

Following Vowel	Phonation Type	F0	F1	F2
[a]	Slack	211.50	379.99	1903.34
	Modal	223.67	496.37	1685.73
	Stiff	234.32	545.27	1527.58
[i]	Slack	223.98	334.67	1964.19
	Modal	225.71	473.47	1733.48
	Stiff	238.08	459.71	1822.95
[u]	Slack	214.81	342.68	1839.40
	Modal	201.73	419.89	1749.47
	Stiff	247.84	410.87	1801.29

We can see similar trends in the differences in harmonic amplitudes of stiff- and slack voiced laterals as those obtained for vowels (figures 3, 4). Figure 6 shows the H1-H2 amplitude difference for lateral approximants. As was observed for vowels in section 5.1, stiff- and slack voiced laterals follow the predicted pattern such that H1-H2 is greater for slack voiced laterals than stiff voiced laterals. This is true for all following vowels collapsed together, and for each following vowel individually. As was observed in the fundamental and formant frequencies of modal [l] (figure 5), H1-H2 amplitude of modal [l] fails to occupy an intermediate position to stiff- and slack voiced [l]. On average and especially when preceding [u], H1-H2 of modal [l] is lower than stiff voiced [l], which was predicted to be the lowest of the three phonation types. When [l]

precedes [a] and [i], H1-H2 of modal [l] is intermediate yet nearly identical to stiff voiced [l].

Figure 6.
H1-H2 of lateral approximants (dB)

	mean	[a]	[i]	[u]
Slack [blV]	4.52	-0.78	7.52	6.82
Modal [lV]	-6.06	-5.50	-4.68	-7.98
Stiff [plV]	-2.52	-5.83	-5.27	3.55
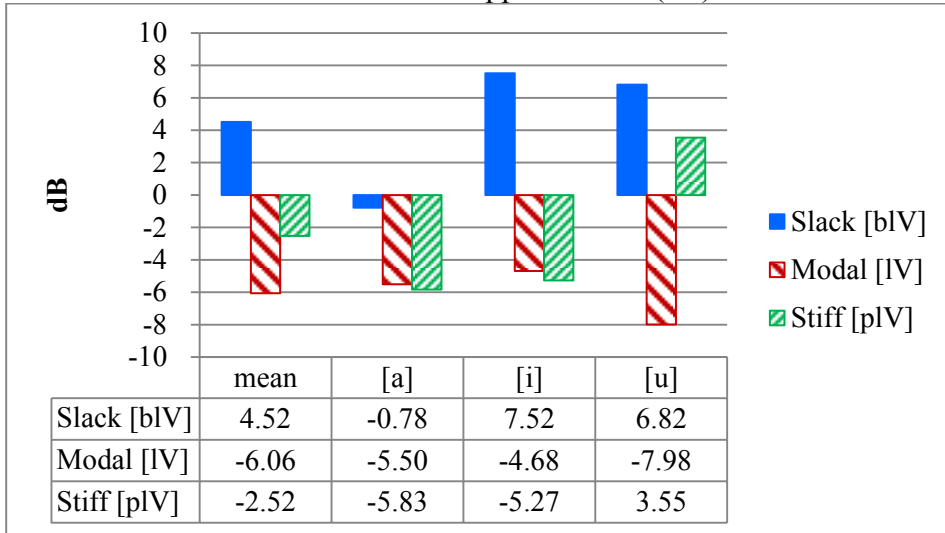
Figure 7 shows the differences in spectral slope (H1-F2) of lateral approximants. Again, we see that H1-F2 of slack voiced laterals is higher than stiff voiced laterals, except when preceding [u]. However, H1-F2 of modal [l] does not appear to occupy a consistent intermediate position to that of stiff- and slack voiced [l]. Looking at each following vowel, we see that when preceding [i], H1-F2 of modal [l] is intermediate to stiff- and slack voiced [l], although it is very close to the average value for stiff voiced [l]. Surprisingly, H1-F2 of modal [l] is much lower than both stiff- and slack voiced [l] when preceding [a] and [u].

Figure 7.
H1-F2 of lateral approximants (dB)

	mean	[_a]	[_i]	[_u]
Slack [blV]	22.12	18.88	24.73	22.75
Modal [lV]	8.61	5.67	10.52	9.63
Stiff [plV]	17.26	15.25	9.77	26.75
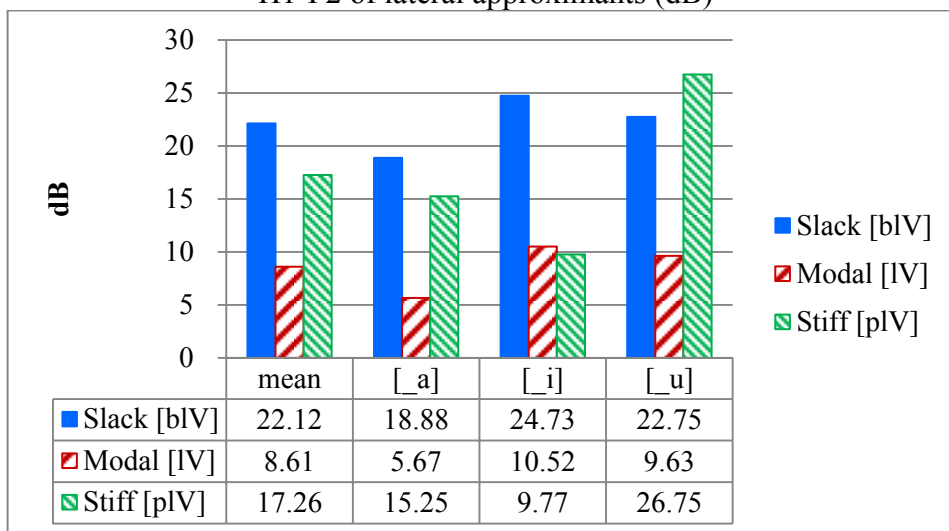
In summary, lateral approximants that were preceded by stiff- and slack voiced stops showed the acoustic cues associated with their respective phonation type. Similar to the findings for stiff- and slack voiced vowels (section 5.1), slack voiced laterals exhibited lower F0 and F1, higher F2, and higher H1-H2 and H1-F2 than their stiff voiced counterparts. Unexpectedly, modal [l] did not occupy a distinct intermediate space in terms of these acoustic measurements, particularly when modal [l] preceded the high vowels [i] and [u].

6.   Discussion

The data presented in the previous section suggest that the entire voiced portion of speech following a stiff- or slack voiced stop show acoustic cues for stiff- and slack voice, respectively, whether that voiced portion is a single vowel or a lateral approximant followed by a vowel. Furthermore, the identity of the vowel appears to have an effect on the particular cues that are used and on their relative prominence. F0 and F1 frequencies were shown to be lower for slack voiced laterals and vowels than stiff voiced laterals and vowels, particularly for [a]. Fagan (1988) and others have interpreted a lower F1 in slack voiced vowels as indicating larynx lowering. However, amplitude differences (H1-H2 and H1-F2) for [a] and laterals preceding [a] did not exhibit the predicted behavior, whereas they were much more apparent for the high vowels [i] and [u]. These two measures of open quotient and spectral slope, respectively, indicate the relative degree of glottal tension via the prominence of H1 relative to the upper harmonics associated with stiff- and slack voice (Blankenship 2002). If we consider the articulatory actions involved in the production of each of the three vowels in the present study, the tongue is least constricted in the production of

[a]. Due to the greater degree of tongue advancement and constriction in high compared to low vowels, the speaker may be less able to achieve larynx-lowering in [i] and [u], and instead relies on a similar action of relaxing glottal constriction, resulting in more apparent differences in open quotient and spectral slope for high vowels. This may explain why F1 differences are less pronounced for the high vowels and more pronounced for the low vowel [a], and conversely why the predicted amplitude differences observed for the high vowels and not for [a]. In order to validate this proposed explanation, however, it will be necessary to first determine experimentally whether the speaker actually uses larynx lowering in the production of slack voice, as is assumed in the literature, and the extent to which the action of larynx lowering is affected by vowel height.

In addition to assessing the relationship between stiff- and slack voice in simple and complex onsets, this study also sought to determine the relationship among stiff-, slack- and modal voice. Ladefoged and Maddieson's (1996) distinction of five laryngeal settings predicts that modal voice will be intermediate to stiff- and slack voice, however Blankenship's (2002) findings suggest that the phonemic status of a nonmodal sound can influence whether it is realized as phonetically distinct from modal sounds. As described in section 2.2, the phonemic distinction between stiff- and slack voice is only relevant within the stop series (including affricates as stops). Stiff- and slack voice do not contrast phonemically with any modal segment in Javanese. The so-called stiff- and slack voiced laterals and vowels described in the preceding sections arise as contextual variants only when preceded by a stiff- or slack voiced stop, respectively. Although the phonemically contrasting stiff- and slack voiced [l] occupied consistently distinct position in terms of frequency and amplitude, non-contrasting modal voiced [l] did not. In many cases, frequency and amplitude values for modal [l] overlapped with those for either stiff- or slack [l]. In other cases, values for modal [l] were more extreme than those for either stiff- or slack [l]. These findings lend further support to Blankeship's (2002) notion that the phonology of a language influences the ways in which particular sounds are realized, such that an idealized continuum (e.g. Ladefoged and Maddieson 1996) may only be phonetically implemented when a phonemic contrast is present. It appears that, for the Javanese speaker in the present study, contrasting phonation types occupy distinct regions of the laryngeal continuum. However, as noted previously, there are no modal sounds that contrast with stiff- or slack voiced sounds in Javanese. Therefore, the phonetic implementation of modal voice need not be as careful as the implementation of nonmodal voice settings. In more concrete terms, when a Javanese speaker produces a [la] syllable, there is no chance that it will be confused with a stiff voiced [la], and so it is permissible to produce that syllable with slightly stiff voiced features or slightly slack voiced features. However, it is very important for the speaker that slack

voiced [ba] not be confused with stiff voiced [pa]. Thus, their production must be more distinct and closer to the hypothesized continuum.

7.   Conclusion

In this preliminary study of nonmodal phonation in Javanese, it was demonstrated that stiff- and slack voice are expressed on the vowel following a stiff- or slack voiced stop, respectively, regardless of the presence of a lateral approximant between the stop and vowel. Furthermore, it was shown that the same acoustic cues for nonmodal phonation present in the vowel are also observed in the intervening lateral approximant. The most prominent acoustic cues for stiff- and slack voice appear to be influenced by the identity of the vowel. The lowering of F0 and F1 following slack voiced stops was most readily observed for the low back vowel [a], and to lesser extents for the high vowels [i] and [u]. Clearer differences in H1-H2 and H1-F2 for stiff- and slack voice were more readily observed for the high vowels than for [a], which showed greater overlap between phonation types. To explain the difference in acoustic cues relative to vowel identity, it was proposed that larynx lowering is easier to achieve for low and back vowels compared to high and front vowels, and that modulations of glottal tension are used to greater extents in order to produce stiff- and slack voice for the high vowels. However, whether this proposed explanation actually accounts for the observed phenomena must be borne out by future research.

Stiff- and slack voice in Javanese appear to occupy distinct position on the continuum of laryngeal tension in terms of fundamental and formant frequencies, and harmonic amplitude differences. This difference can be attributed to the contrasting phonemic status of stiff- and slack voiced stops in Javanese. Similar measurements of word-initial modal voiced lateral approximants did not distinguish them from the so-called stiff- and slack voiced lateral approximants which resulted from the expression of stiff- and slack voice in [plV] and [blV] syllables, respectively. Following the work of Blankenship (2002), we proposed that the lack of a clear boundary between modal voice on the one hand and stiff- and slack voice on the other hand may be due to the lack of a phonemic contrast between modal and nonmodal segments in Javanese.

The present findings are limited in their generalizability, as they are draw from a relatively small data sample elicited from a single speaker. Future studies of nonmodal phonation in Javanese stops should therefore collect a larger data set from a larger sample of speakers, allowing a more thorough analysis using inferential statistics. The future research should also elicit stiff- and slack voice stop clusters at additional places of articulation, and involving different approximants. It was shown that the lateral approximant [l] does not block the spreading of nonmodal phonation to the following vowel in [plV] and [blV] clusters, and that the approximant takes on those features of nonmodal

phonation. It remains to be seen whether this is true of the other approximants in Javanese, particularly the trilled approximant [r], which may be less amicable to the spreading of phonatory features than [l], [j] or [w]. Finally, the role that the various acoustic cues for stiff- and slack voice play in the categorical perception of Javanese stop consonants and consonant clusters must also be considered in future research.

Appendix A.   Word list

Target words are listed in alphabetical order with English gloss.
1. babad   'history'
2. bajag   'pirate'
3. bapak   'father'
4. bibar   'after'
5. biji   'ore'
6. bikak   'open (karma register)'
7. blabag   'blackboard; one who is responsible for X'
8. blabar   'strong cord used as a boundary marker'
9. blacan   'wild cat'
10. bligo   'wax-gourd'
11. blimbing   'star fruit'
12. blirik   'spotted hen'
13. bludag   'to overflow, spill out/over; to be boastful; to be generous
                     to others'
14. bluluk   'coconut in its first stage of development on the tree'
15. blumbang   'pond, pool'
16. budhal   'departure'
17. buka   'to go'
18. bungah   'excited'
19. lagu   'song'
20. larang   'expensive'
21. layang   'note'
22. lima   'five'
23. lipet   'multiplied'
24. lisah   'oil'
25. lucu   'cute'
26. lukisan   'picture'
27. luweh   'the optimum
28. pager   'fence'
29. pangan   'to eat'
30. papat   'four'
31. pindah   'move'
32. pitu   'seven'

33. piye   'how'
34. pladen   'one who serves or waits on people'
35. plak-plek   'to keep slapping'
36. plana   'saddle'
37. pliket   'sticky'
38. plilik   'to stare wide-eyed'
39. plisir   'handbreadth'
40. plucut   'something slippery falling out of one's hand'
41. plung   'sound of something falling into liquid'
42. pluntur   'cord used for hanging various gamelan instruments'
43. pulisi   'police'
44. putih   'to white'
45. puter   'turn'

References

Blankenship, Barbara. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30:163–191.

Fagan, Joel L. 1988. Javanese intervocalic stop phonemes: The light/heavy distinction. *Studies in Austronesian Linguistics. Ohio University Center for International Studies*.

Ladefoged, Peter, and Maddieson, I. 1996. *Sounds of the world's languages.* Cambridge: Blackwell.

Laver, John. 1980. *The phonetic description of voice quality.* Cambridge: Cambridge University Press.

Reetz, Henning, and Jongman, Allard. 2009. *Phonetics: Transcription, production, acoustics, and perception.* Malden: John Wiley & Sons.

Thurgood, Ela. 2004. Phonation types in Javanese. *Oceanic linguistics* 43:277–295.

Wayland, Ratree, and Jongman, Allard. 2003. Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics* 31:181–201.

DECOMPOSING MALAY ANAPHORIC EXPRESSIONS[*]

Hiroki Nomoto
Tokyo University of Foreign Studies
nomoto@tufs.ac.jp

Previous studies on anaphoric expressions in Malay centre on two forms, i.e. '*diri* + pronoun' and '*diri* + pronoun + *sendiri*', and analyse them non-compositionally. This paper shows that a compositional analysis of Malay anaphoric expressions is not only possible but is empirically more desirable than a non-compositional analysis, as it can account for a wider range of anaphoric expressions in a systematic manner.

1.   Introduction

Previous studies on anaphoric expressions in Standard Malaysian/Singapore Malay (e.g. Cole and Hermon 1998, 2005; Nomoto 2011) centre on two forms: (i) '*diri* + pronoun' and (ii) '*diri* + pronoun + *sendiri*'. They are non-compositional in that they treat these multimorphemic forms as single lexical items/words on a par with English reflexives such as *himself*. Consequently, the anaphoric properties of the relevant forms are ascribed to the whole expression.

This study proposes an alternative analysis. The proposed analysis is compositional, and hence the anaphoric properties of a multimorphemic form are ascribed to its constituent parts. It is demonstrated that the proposed analysis can deal with a wider range of anaphoric expressions in Malay, which include (i) '*diri* + pronoun', (ii) '*diri* + pronoun + *sendiri*', (iii) '*diri* + non-pronoun', (iv) '*diri* + non-pronoun + *sendiri*', (v) *diri*, (vi) *diri sendiri*, (vii) '(non-)pronoun + *sendiri*', and (viii) *sendiri*. The proposed analysis offers more insight into the study of anaphoric expressions in related dialects/languages (e.g. Gil 2001; Paul 2004; Davies 2008; Kartono 2013), as these dialects/languages have been reported to employ similar multimorphemic anaphoric expressions.

This paper is organized as follows. Section 2 demonstrates that Malay indeed possesses the various anaphoric expressions listed above by presenting examples of them. Section 3 reviews Cole and Hermon's (2005) analysis of patterns (i) and

(ii) as a representative of non-compositional analyses. After pointing out problems of their analysis, I propose an alternative compositional analysis in section 4 and show how it accounts for the anaphoric properties of the various types of anaphoric expressions in Malay in section 5. Section 6 concludes the paper.

2.   The Inventory of Malay Anaphoric Expressions

Besides the two types that have attracted previous researchers' attention ((i) and (ii)), Malay has many other patterns of anaphoric expressions ((iii)–(viii)). All eight patterns are easily found in naturally occurring texts. They are all made up of one or more elements from *diri*, *sendiri* and noun phrases. This simple fact already suggests the possibility of a compositional treatment. I will show in section 4 that it is possible to reduce the eight patterns to just three. Examples of the eight patterns are given below.

(i) '*diri* + pronoun'   It is this pattern that has intrigued previous researchers the most, as it shows a hybrid property of reflexives and pronominals. At first, one may regard it as reflexive because, in elicitation sessions, native Malay speakers will normally use this pattern to translate English expressions with reflexives such as *himself* and *herself*. In support of this initial hypothesis, the pattern allows a local c-commanding antecedent, as indicated by index j in (1)–(2) below. However, a closer inspection reveals that the pattern also exhibits pronominal behaviours. Thus, *dirinya* also allows non-local and non-c-commanding antecedents, as indicated by index i in (1) and (2) respectively. The referent indicated by index k is a discourse referent that is salient in the current discourse but is not referred to by any DP in the sentence.

(1)   Ali$_i$ kata [Siti$_j$ mengambil gambar *diri-nya$_{i/j/k}$*].
         Ali say Siti take   picture DIRI-3
         'Ali said Siti took a picture of *him/herself/her*.'

(2)   [Bapak Siti$_i$]$_j$ tidak suka *diri-nya$_{i/j/k}$*.
          father Siti   not   like DIRI-3
         'Siti's father does not like *her/himself/him*.' (Cole and Hermon 2005:631)

(ii) '*diri* + pronoun + *sendiri*'   Unlike '*diri* + pronoun', this pattern shows regular reflexive behaviours. It is hence subject to Condition A of the canonical binding theory; *dirinya sendiri* must be locally bound in (3) below.[1]

(3)   a.   Ali$_i$ kata Siti$_j$ mengambil gambar *diri-nya sendiri$_{*i/j/*k}$*.
                Ali say Siti take   picture DIRI-3 own

---

[1] Note that examples like (i) involve a combination of pattern (i) '*diri* + pronoun' and the adverbial *sendiri* meaning 'alone, by oneself', and should not be confused with pattern (ii).

'Ali said Siti took a picture of *herself*.'

b. [Ibu   Siti$_i$]$_j$ mengambil gambar *diri-nya sendiri*$_{*i/j/*k}$.
    mother Siti   take   picture DIRI-3   own
    'Siti's mother took a picture of *herself*.'

For patterns (iii)–(viii), I only present examples (all taken from the DBP (Dewan Bahasa dan Pustaka) Corpus).[2] The anaphoric properties of these patterns will be described in section 5, where I show that they can be accounted for by the compositional analysis to be proposed in section 4.

(4) **(iii) '*diri* + non-pronoun'**[3]

Segala kejadian biologikal dan fizikal   ke atas *diri remaja   itu*
all   incident biological and physical to top DIRI adolescent that
[...].
'All the biological and physical incidents that happen to *the adolescent* [...].'

(5) **(iv) '*diri* + non-pronoun + *sendiri*'**

Apabila tersalah memilih teman, ia akan memberi kesan yang besar
when   mistake choose friend it will give   effect REL big
terhadap *diri remaja   itu sendiri*.
towards DIRI adolescent that own
'If friends are chosen wrongly, that will have a big influence on *the adolescent*.'

(6) **(v) *diri***

Itu penting, sebab   dalam hidup ini keyakinan *diri* adalah 50%
that important because in   life this confidence DIRI be   50%
daripada kemenangan.
from   victory
'That is important because in life *self*-confidence is 50% of the success.'

(7) **(vi) *diri sendiri***

Sampai bila   adik   harus membohongi *diri sendiri* dan
until   when younger.sibling should deceive   DIRI own   and

(i)   John fikir (yang) diri-nya sendiri akan pergi ke KL besok.
       John think that   DIRI-3 alone   will go   to KL tomorrow
       'John thinks that he himself will go to KL (= Kuala Lumpur) tomorrow.'

(Cole and Hermon 2005:634)

[2] See my AFLA handout (http://www.tufs.ac.jp/ts/personal/nomoto/ handout_afla21.pdf) for more examples.

[3] Note that non-pronouns are not used as pronoun substitutes in this pattern, unlike *adik* 'younger sibling' in (7), which is used as a substitute for the second person pronoun.

diri-nya?
DIRI-3
'When can you stop deceiving *yourself* and him?'

(8) **(vii) '(non-)pronoun +** *sendiri*'

Bangsa Melayu di-katakan tidak mampu berfikir dalam bahasa   *mereka*
ethnic Malay PASS-say not able   think in   language their
sendiri.
own
'Ethnic Malays are said to be unable to think in *their own* language.'

(9) **(viii)** *sendiri*

Di samping itu, responden tidak pasti   sama ada mereka memiliki
at side   that respondent not   certain whether they   have
sikap   bangga terhadap bahasa   *sendiri*.
attitude proud   towards language own
'Moreover, the respondents are not certain whether they take pride in *their own* language.'

3.   Non-compositional Analysis: Cole and Hermon (2005)

In this section, I briefly review Cole and Hermon's (2005) study as a representative of non-compositional analyses of Malay anaphoric expressions. As noted at the outset of this paper, Cole and Hermon focus on two patterns, i.e. (i) '*diri* + pronoun' and (ii) '*diri* + pronoun + *sendiri*'. They do so presumably for typological considerations rather than descriptive considerations concerning the anaphoric system of a specific language. In particular, pattern (i) in Malay could pose a serious problem to some rather solid typological generalizations about long-distance reflexives if it were actually a reflexive (Cole and Hermon 1998, 2005). They conclude that pattern (i) is in fact not a reflexive, and hence does not affect the relevant typological generalizations. This conclusion is based on the following analysis of pattern (i).

In Cole and Hermon's analysis, "*dirinya* is not treated by the syntax as a complex DP [...], but rather as a unitary lexical entry" (643). They assume the existence of a paradigm of '*diri* + pronoun' shown in Table 1, which is reminiscent of the English reflexive pronoun paradigm.
Cole and Hermon claim that "*diri* + pronoun is unspecified in the lexicon with regard to the features [αanaphor] and [αpronominal]" (631). In other words, pattern (i) is neither a reflexive nor a pronominal; it has properties of both. This unspecified feature analysis is able to account for the hybrid property of '*diri* + pronoun' that we saw in the last section (cf. (1)).

Given this analysis, a similar analysis has to be assumed for true reflexives '*diri* + pronoun + *sendiri*', though Cole and Hermon are not explicit about them.

Table 1: The full paradigm of '*diri* + pronoun' (Cole and Hermon 2005:629)

Person	Singular	Plural
First Person	diri saya/diri-ku	diri kami/diri kita
Second Person	diri kamu/diri-mu	diri kamu/diri-mu
Third Person	diri-nya	diri mereka/diri-nya

That is to say, the lexicon of Malay should have '*diri* + pronoun + *sendiri*' as a unitary lexical entry with a paradigm similar to Table 1. Each lexical entry must be specified in the lexicon as [+anaphor, −pronominal] to capture its canonical reflexive behaviours.

While Cole and Hermon's study has bolstered the general theory of long-distance reflexives by explaining away the typologically unusual behaviours of reflexive-looking forms in Malay, i.e. '*diri* + pronoun', it is problematic when taken as an analysis of a phenomenon in a specific language. To begin with, the various other anaphoric expressions presented in section 2 (iii)–(viii) appear to result from combinations of the constituent parts of '*diri* + pronoun (+ *sendiri*)' (i)–(ii). Yet, under Cole and Hermon's non-compositional analysis (or any non-compositional analysis for that matter), the relation between (iii)–(viii) and (i)–(ii) remains unclear. Furthermore, the pronoun slot in '*diri* + pronoun (+ *sendiri*)' is in fact not restricted to pronouns, but also available for other DPs, as is the case with (iii) '*diri* + non-pronoun' and (iv) '*diri* + non-pronoun + *sendiri*'. It is implausible to think that these patterns with non-pronouns are also unitary lexical entries, for the number of non-pronouns is infinite, unlike that of pronouns.

Given these problems, it is more desirable if a compositional analysis is possible that can account for the anaphoric properties of '*diri* + pronoun (+ *sendiri*)' as well as the other anaphoric expressions. I will propose one such analysis in the next section.

4.   Compositional Analysis

The proposed compositional analysis has three main ingredients. The first ingredient is the syntax and semantics of *diri*. A proper understanding of *diri* enables us to see how *diri* is involved in anaphoric expressions and why. It also leads us to the second main ingredient, i.e. the null unspecified possessive pronoun *pro*, which plays an important role in reducing the numerous different patterns into just three and thereby greatly simplifying the description and analysis of the anaphoric expressions in Malay. The last ingredient is the semantics of the intensifier *sendiri* 'alone, own'. I argue that the semantics of *sendiri* involves a kind of focus semantics and that the local binding property characteristic of anaphors results from it.

4.1. *Diri*

I make the following two claims about *diri*. First, *diri* is an NP that takes a posses-
sor argument. Second, it denotes a function from an individual to that individual's
physical self:

(10) $[\![\text{diri}]\!] = \lambda x.x$'s physical self

To put this informally, *diri* means 'someone's body'. Given that one's body,
constituting his/her physical self, is the entire whole inalienably possessed by
him/her, one could analyse *diri* as a pseudo-identity function.[4] In other words, the
formula in (10) can be approximated as in (11), which roughly says that a person's
body is that person himself/herself.

(11) $[\![\text{diri}]\!] = \lambda x.x$'s physical self $\approx \lambda x.x$'s $x \approx \lambda x.x$

Two possibilities exist regarding the approximations in (11). The first possibil-
ity is that they are hypothesized diachronic processes in the grammaticalization
of *diri*. That is, there are two *diri* morphemes synchronically, i.e. (10) and the
identity function ($\lambda x.x$ in (11)). Only the former but not the latter preserves the
meaning of 'physical self'. The other possibility is that there is only one *diri*
morpheme and the approximations in (11) take place in the speaker's mind/brain
as synchronic subconscious inference processes when *diri* is used in anaphoric
expressions. In what follows, I will assume the first possibility for simplicity's
sake, and treat *diri* as an identity function, unless otherwise noted. Further study
is needed to determine which possibility is actually the case.

There is evidence for the 'physical self' meaning of *diri*. First, *diri* can be
used in contrast with words such as *jiwa* 'soul', *hati* 'heart' and *batin* 'inner self',
words representing the mental self of an individual:

(12)   a.   Saya terpaksa "meremajakan semula" *diri   dan jiwa* saya.
                 I   have.to   rejuvenate   DIRI and soul my
                 'I had to rejuvenate my body and soul.'
           b.   [...] cuba-lah   fahami   lagi siapa Ade ini, fahami
                        try-PARTICLE understand more who   Ade this understand
                 diri   dan hati Ade.
                 DIRI and heart Ade
                 '[...] try to understand more who I am, to understand my physical
                 self and heart.'   (DBP Corpus)

Second, if only the physical aspect of an individual changes, it is a change in *diri*:

---

[4] Kartono (2013) glosses *diri* in Indonesian, Palembangese and Jambi as 'body'. This is not
a precise description of *diri* in Malay (and probably even in these languages). It is *badan* and
tubuh that refer to 'body'. *Diri* is not interchangeable with these words, as it is an abstract notion
reflecting the Malay perception of the self.

(13)   Ular itu telah kembali kepada *diri* asal-nya, se-orang putera raja.
       snake that PRF return   to   DIRI original-3 one-CLF prince king
       'The snake has transformed itself back into the original self, a prince.'
                                                        (*Si Bongsu dengan Kak Nam*)

These examples also lend support to the claim that *diri* is an NP. It is smaller than DP because it can be followed by a possessor DP and other modifiers. Moreover, it can be coordinated with another non-DP noun phrase.

4.2.   The Null Unspecified Possessive Pronoun *pro*

The second ingredient for the compositional analysis of various anaphoric expressions in Malay is the null unspecified possessive pronoun *pro*. It is comparable to *one's* in English, and is a part of the pronominal paradigm of the language.

Where anaphoric expressions are concerned, *pro* occurs in the following two contexts: as a possessor argument of *diri* and with the intensifier *sendiri*. With the introduction of *pro*, it becomes possible to reduce the eight different patterns into just three types, as shown in (14).

(14)   (i)   '*diri* + pronoun'   ⎫
       (iii)   '*diri* + non-pronoun'   ⎬ '*diri* + DP'
       (v)   *diri* = *diri* + *pro*   ⎭

       (ii)   '*diri* + pronoun + *sendiri*'   ⎫
       (iv)   '*diri* + non-pronoun + *sendiri*'   ⎬ '*diri* + DP + *sendiri*'
       (vi)   *diri sendiri* = '*diri* + *pro* + *sendiri*' ⎭

       (vii)   '(non-)pronoun + *sendiri*'   ⎫ 'DP + *sendiri*'
       (viii) *sendiri* = '*pro* + *sendiri*'   ⎭

When *diri* appears to occur with no noun phrase, as in (v) and (vi), it actually contains the phonologically null DP *pro* as its complement. Likewise, when *sendiri* appears to occur with no noun phrase, as in (vi) and (viii), it actually modifies *pro*. Notice that (vi) can be parsed in two ways: [*diri* + *pro*] + *sendiri* and *diri* + [*pro* + *sendiri*]. Also notice that '*diri* + DP + *sendiri*' is a combination of '*diri* + DP' and 'DP + *sendiri*'.

Now, the various anaphoric expressions in Malay can be summarized as in Table 2. Emphatic forms are so called because their denotations are virtually identical to those of the corresponding non-emphatic forms (cf. (11)), but have more expression.[5] The forms containing *sendiri* are referred to as focussed forms because the semantics of *sendiri* involves focus semantics, as we shall now see.

---

[5] In this connection, Kartono (2013:50–53) points out that emphatic forms in Indonesian (more specifically '*diri* + pronoun' in a subject position) are used to express respect and empathy. This description seems also valid in Malay.

Table 2: The composition of Malay anaphoric expressions

	Non-focussed	Focussed (*sendiri*)
Non-emphatic	**(A) DP** [$_A$ DP] *pro*/-nya/Ali 'one's/his/her/Ali's'	**(C) DP + sendiri** [$_C$ [$_A$ DP] sendiri] *pro*/-nya/Ali sendiri 'one's/his/her/Ali's own'
Emphatic (*diri*)	**(B) diri + DP** [$_B$ diri [$_A$ DP]] diri *pro*/-nya/Ali 'one's/his/her/Ali's physical self'	**(D) diri + DP + sendiri** [$_D$ [$_B$ diri DP] sendiri]/ [$_D$ diri [$_C$ DP sendiri]] diri *pro*/-nya/Ali sendiri 'one's/his/her/Ali's own physical self'

4.3. *Sendiri*

According to Alsagoff (1992), *sendiri* 'alone, own' (in a non-subject position) requires a local antecedent when used by itself, as in (15).

(15)   Mariam$_i$ kata bahawa Ali$_j$ menjual kereta *sendiri*$_{*i/j}$.
       Mariam say that   Ali sell   car   own
       'Mariam said that Ali sold *her/his car.'   (Alsagoff 1992:41)

Alsagoff's other example (16) and the additional example in (17) show that the antecedent of *sendiri* must c-command the NP modified by *sendiri* too. These properties seem inherent to *sendiri*.[6]

(16)   Ali$_i$ mengusik doktor itu$_j$ di rumah *sendiri*$_{i/*j}$.
       Ali tease   doctor that in house own
       'Ali teased the doctor in his own house.'   (Alsagoff 1992:40)

(17)   [Ibu   Ali$_i$]$_j$ menjual kereta *sendiri*$_{*i/j}$.
        mother Ali   sell   car   own
       'Ali's mother sold her/*his own car.'

   In the present analysis, *sendiri* in the examples above is in fact *pro sendiri*. Given this analysis, the role of *sendiri* is to restrict the otherwise unspecified referent of *pro* to that of its antecedent. Crucially, *sendiri* restricts possible interpretations to a reflexive one.

---

[6] One might wonder whether they can be attributed to parts of *sendiri*, as *sendiri* contains *diri*. However, synchronically, *sendiri* should not be analysed into smaller parts. The etymology of *sendiri* is not clear. Zaharani Ahmad (p.c.) suggested to me that it emerged from *se-orang diri* [one-CLF DIRI] 'alone', which is highly probable, given the fact that the related language Minangkabau employs *surang*, the equivalent of *seorang* in Malay, in contexts where Malay uses *sendiri* such as *dirinya sendiri* (Yusrita Yanti, p.c.).

Let us look at the function of *sendiri* (in a non-subject position) in more detail. *Sendiri* plays a dual role, syntactic and semantic. Syntactically, it searches the sentence for the antecedent of the expression it combines with. As we have seen above, the antecedent must locally c-command *sendiri*. Semantically, *sendiri* involves a kind of focus semantics, as Gil (2001) proposes for *sendiri* in Riau Indonesian.[7] This focus semantics has to do with its meaning translated into English as 'own'. *Sendiri* induces a set consisting of the potential referents of the expression it combines with, and excludes from this set all members but the referent of the antecedent.

Given the expression in (18a), where DP_a (a for "antecedent") denotes individual a ($[\![DP_a]\!] = a$), 'DP_f *sendiri*' induces a set of the potential referents of DP_f (f for "focus"), as in (18b). Call this set F. F must include a, otherwise the expression is ungrammatical. *Sendiri* entails that among the members of F, V applies only to a (18c).

(18)   a. DP_a V [$_{OBJ}$ NP DP_f sendiri]
       b. $F = \{a, b, c, d, \dots\}$
       c. $\forall x[\text{V}(x) \rightarrow x = a]$

If DP_f is a pronoun whose ϕ-features are compatible with DP_a, then the referent of DP_f becomes identical to that of DP_a ($[\![DP_a]\!] = [\![DP_f]\!] = a$). A reflexive interpretation is obtained in this case. For example, in (15) with the expression 'Ali V [$_{OBJ}$ NP *pro* sendiri]', the ϕ-features of *Ali* and *pro* are compatible with each other, as the latter is unspecified in terms of ϕ-features. Thus, the otherwise unrestricted referent of *pro* is fixed to Ali, giving rise to a reflexive interpretation. By contrast, if *pro* is replaced by another pronoun whose ϕ-features are incompatible with those of *Ali* as in (19), the sentence becomes ungrammatical. This is because Ali is not in F, the set of potential referents of *saya* 'my'.[8]

(19)   *Ali$_a$ menjual [$_{OBJ}$ kereta saya$_f$ sendiri].
        Ali sell   car   my   own
        ('*Ali sold my own car.')

In the examples above, the DP that *sendiri* combines with is a pronoun. However, when *sendiri* combines with a non-pronoun DP as in *Siti sendiri* 'Siti herself', it works slightly differently. Specifically, the DP itself becomes the antecedent of *sendiri*. Moreover, the set F consists not of the potential referents of the DP, but of the DP's actual referent and its alternatives, because the referent of

---

[7] Gil refers to the function of *sendiri* as 'conjunctive operator'. Moreover, the specific formulation of it to be presented below differs from Gil's.

[8] One might wonder whether the sentence is grammatical if *Ali* is used as a substitute for the first person pronoun. Such a sentence is unacceptable or very unnatural at best for an independent reason: when a proper name is used as a substitute for a pronoun, the substitution must occur throughout the sentence consistently.

a non-pronoun DP is already fixed to a particular individual. As a consequence of the first difference, *sendiri* does not affect the referent of the DP. Thus, I will not discuss cases involving non-pronouns below, i.e. patterns (iii) and (iv) as well as pattern (viii) with non-pronouns.

5.   Accounting for Each Pattern

This section discusses how the compositional analysis proposed in the last section accounts for the various anaphoric expressions. Since one of the patterns in focussed forms (C), i.e. (viii) *sendiri*, has been already explained in the last section, I begin with focussed forms (C). I will then turn to emphatic focussed forms (D), as *sendiri* plays a crucial role here too. Emphatic forms (B), which unlike the other two forms do not involve *sendiri*, are discussed at the end.

5.1.   DP + *Sendiri* (C. Focussed Forms)

As seen in section 4.3, *sendiri* restricts possible interpretations to a reflexive one, by limiting the referent of the expression it combines with to that of its antecedent, which must locally c-command it. To repeat the account of pattern (viii), while the referent of *pro* in (20a) is unconstrained, allowing both reflexive and non-reflexive interpretations, that of *pro* in (20b) is restricted to that of *sendiri*'s antecedent, i.e. Ali. As a result, only a reflexive interpretation is available in (20b).

(20)   a.   Ali$_i$ menjual kereta $pro_{i/j}$.[9]
               Ali sell   car   *pro*
               'Ali sold his/her/their/etc. car.' (reflexive and non-reflexive)
        b.   Ali$_i$ menjual kereta $pro_{i/*j}$ sendiri.
               Ali sell   car   *pro*   own
               'Ali sold his/*her/*their/*etc. own car.' (reflexive only)

Another pattern of focussed forms, i.e. (vii) 'pronoun *sendiri*', differs from (viii) only in that an overt pronoun is used instead of *pro*. Hence, no additional mechanism is necessary to account for (vii). In (21a) without *sendiri*, the referent of the third person pronoun -*nya* can be either that of the local subject *Ali* or that of the matrix subject *Mariam*, or even a discourse referent that does not appear in the sentence. A reflexive interpretation is obtained in the first case whereas a non-reflexive interpretation is obtained in the second and third cases. However, the addition of *sendiri* rules out the latter interpretative possibility and forces a reflexive interpretation, as in (21b).

---

[9] The same surface string has another structure without the possessor pronoun *pro*. This alternative structure is irrelevant here; it is associated with an interpretation in which the possessor is unimportant (e.g. 'Ali sold a car.').

(21)   a.   Mariam$_i$ kata bahawa Ali$_j$ menjual kereta-*nya*$_{i/j/k}$.
             Mariam say that   Ali sell   car-3
             'Mariam said that Ali sold her/his car.' (reflexive and non-reflexive)
       b.   Mariam$_i$ kata bahawa Ali$_j$ menjual kereta-*nya*$_{*i/j/*k}$ *sendiri*.
             Mariam say that   Ali sell   car-3   own
             'Mariam said that Ali sold *her/his own car.' (reflexive only)

Notice that in my analysis (the relevant use of) *sendiri* always modifies the possessor. This is obvious when the possessor is overt. Where there does not seem to be any possessor, as in pattern (viii), my analysis assumes a null DP possessor *pro*. My analysis thus accords with the following observation by Gil (2001:112): "in Standard Malay/Indonesian, the actual form *sendiri* appears to function as a reflexive only in possessive constructions." It is worth noting here that while Malay and Modern English allow both reflexive and non-reflexive interpretations only in possessive constructions, some languages allow both interpretive possibilities for non-possessive pronominals as well (cf. Table 3). These languages include Old English (e.g. van Gelderen 2000; König and Siemund 2000; Keenan 2002), Madurese (Davies 2008), Jambi (Mudung Darat and Tanjung Raden dialects) (Cole et al. 2010), and Palembangese (Kartono 2013). Conversely, other languages distinguish between two kinds of possessive pronouns for reflexive and non-reflexive interpretations. Russian is one such language (Kazuhiro Kojima, p.c.). I am not aware of a language that has distinct reflexive and non-reflexive forms only in possessive constructions.

Table 3: Cross-linguistic variation in the reflexive vs. non-reflexive formal distinction

	Non-possessive	Possessive
Old English, Madurese, etc.	−	−
Modern English, Malay	+	−
Russian	+	+

5.2.   *Diri* + DP + *Sendiri* (D. Emphatic Focussed Forms)

As seen in section 2, (ii) '*diri* + pronoun + *sendiri*' shows typical reflexive behaviours. It must be bound locally, as in (22b). This is not a property of the expression as a whole but one of *sendiri* alone, as is the case with focussed forms discussed above. (22a) without *sendiri* has both reflexive and non-reflexive interpretations, putting aside the reason why it does for the moment (see section 5.3). With the addition of *sendiri*, the interpretation is restricted to a reflexive one, as in (22b). This restriction can be achieved in two ways, as *diri-nya sendiri* can be parsed in two ways: (a) *[[diri -nya] sendiri]* and (b) *[diri [-nya sendiri]]*.

Since *diri* is an identity function (cf. section 4.1), these two parses can be rewritten as (a) *[[-nya] sendiri]* and (b) *[[-nya sendiri]]* respectively, and end up being denotationally identical to pattern (vii) 'DP + *sendiri*' of focussed forms (C).

(22)   a.   Ali$_i$ kata [Siti$_j$ mengambil gambar *diri-nya$_{i/j/k}$*]. (= (1))
              Ali say Siti   take   picture DIRI-3
              'Ali said Siti took a picture of him/herself/her.' (reflexive and non-reflexive)

       b.   Ali$_i$ kata Siti$_j$ mengambil gambar *diri-nya sendiri$_{*i/j/*k}$*. (= (3a))
              Ali say Siti take   picture DIRI-3 own
              'Ali said Siti took a picture of *him/herself/*her.' (reflexive only)

It is expected that local binding is also required for (vi) *diri sendiri*, because it is in fact '*diri pro sendiri*' in my analysis and hence differs from (ii) '*diri + pronoun + sendiri*' only in that the former involves a null pronoun. This prediction is borne out, as in (23).

(23)   Ali$_i$ kata Siti$_j$ mengambil gambar *diri pro$_{*i/j/*k}$ sendiri*.
           Ali say Siti take   picture DIRI   own
           'Ali said Siti took a picture of *him/herself/*her.'

5.3.   *Diri* + DP (B. Emphatic Forms)

There are two emphatic form patterns to consider: (i) '*diri* + pronoun' and (v) *diri*. The first pattern has been the main concern in previous studies. The second pattern is in fact *diri pro*, and differs from the first pattern only with regard to the type of pronoun used, overt or null. I discuss these two patterns in separate subsections below, because the choice of the pronoun affects the resultant interpretation considerably.

5.3.1. Pattern (i): *Diri* + Pronoun

Recall that *diri* is an NP that takes a possessor argument (cf. section 4.1). Thus, in this pattern, *diri*'s possessor argument is saturated by the pronoun, resulting in an emphatic pronoun. English does not have a distinct form corresponding to this Malay form, but a similar meaning can be conveyed by prosodic emphasis.

Being an emphatic pronoun, '*diri* + pronoun' behaves pronominally. For instance, *dirinya* in (24) can take as its antecedent the matrix subject *Ali* and a discourse referent indicated by index k. What is surprising, however, is that local binding is also possible, as indicated by index j.

(24)   Ali$_i$ kata Siti$_j$ selalu memuji [*diri-nya*]$_{i/j/k}$.
           Ali say Siti always praise   DIRI-3
           'Ali says Siti always praises him/her/herself.'

I adopt an account for this local binding property suggested (but rejected) by Cole and Hermon (2005). Under this analysis, the interpretation of '*diri* + pronoun' is determined by the antecedent of the possessor pronoun. This is because *diri* denotes an identity function (cf. (11)), and hence the indices of the possessor and the entire phrase become identical, as shown in (25).

(25)   dirinya: y_1

        diri: $\lambda x.x$ -nya: y_1

Now, a direct object possessor can be coreferential with the local subject, as in (26). Replacing *anak lelaki* in (26) by *diri*, (27) is obtained, where local binding holds between the embedded subject *Siti* and *dirinya*, making *dirinya* look like a reflexive pronoun.

(26)   Ali kata Siti$_j$ selalu   memuji [anak lelaki-*nya*$_j$].
       Ali say Siti always praise   child male-3
       'Ali says Siti always praises her son.'

(27)   Ali kata Siti$_j$ selalu   memuji [*diri-nya*$_j$]$_j$.
       Ali say Siti always praise   DIRI-3
       'Ali says Siti always praises herself.'

Of course, a direct object possessor can be coreferential with other DPs in the sentence and discourse referents, in which case a non-reflexive reading is obtained.

Cole and Hermon (2005) reject the analysis above. They expect sentences (28) to pattern with (29) rather than (30) in terms of Condition C, presumably confusing the index of the possessor *-nya* in (29) with that of the whole DP.[10] What the analysis actually predicts, however, is (28) patterning with (30) rather than (29) based on the correct indexing shown in the parentheses after the Malay sentences.

(28)   a. *Diri-nya$_i$ mencium John$_i$. ([diri-nya$_i$]$_\mathbf{i}$)
           DIRI-3   kiss   John
           'He kissed John.'
     b. *Diri-nya$_i$ di-cium   (oleh) John$_i$. ([diri-nya$_i$]$_\mathbf{i}$)
           DIRI-3   PASS-kiss by   John
           'He was kissed by John.'

(29)   a. Anjing-nya$_i$ menggigit John$_i$. ([anjing-nya$_i$]$_\mathbf{j}$)
           dog-3   bite   John
           'His dog bit John.'
     b. Anjing-nya$_i$ di-pukul (oleh) John$_i$. ([anjing-nya$_i$]$_\mathbf{j}$)
           dog-3   PASS-hit by   John

---

[10] Paul (2004) makes a similar mistake in her discussion of Malagasy *ny tenany* [DET self.3(GEN)].

'His dog was hit by John.'

(30)   a.   *Dia$_i$ mencium John$_i$. (dia$_\mathbf{i}$)
           3   kiss   John
           'He kissed John.'
    b.   *Dia$_i$ di-cium   (oleh) John$_i$. (dia$_\mathbf{i}$)
           3   PASS-kiss by   John
           'He was kissed by John.'

5.3.2. Pattern (v): *Diri*

When *diri* appears to be used by itself, the meaning is generic, and not one of simple reflexivity. Recall that *diri* in such cases is in fact accompanied by the null unspecified pronoun *pro* (i.e. *diri pro*). I argue that the generic meaning is due to the unspecified nature of *pro* accompanying *diri*; *pro* means 'one's' rather than 'my/your/his/her/etc.'. Furthermore, the meaning of *diri* involved in this pattern, in most instances, is not an identity function but 'physical self' (10). This is because most instances of this pattern involve a reflexive meaning, which cannot be captured by the composition of an identity function and *pro*. Arguably because of the use of the 'physical self' *diri* instead of the identity function *diri*, (v) is much less common than the other patterns discussed above. Its distribution seems to be lexically determined, even though the semantic contribution of *diri* is clear. In this respect, (v) resembles *self-* in English (e.g. *self-care*, *self-ignite*).[11]

    Diri (= *diri pro*) can occur as either a noun modifier or a verbal complement. In (31), *diri* modifies an NP. An NP modified by *diri* is interpreted generically as in (31a), and cannot be associated with a particular individual as in (31b).[12] *Gambar diri* could mean 'selfie', but like nouns with *self-* in English, the use must be established in the speaker community in order to become fully acceptable.

(31)   a.   Siti tidak lalai   menjaga [kecantikan/maruah *diri*].
           Siti not   careless keep   beauty/dignity   DIRI
           'Siti does not fail to take care of her beauty/dignity.' (cf. self-beauty, self-dignity)

---

[11] Chung (1976) analyses *diri* in Indonesian as a clitic attaching to a verb, as it cannot be separated from the verb by a PP, as in (i). Gil (2001) reaches a similar conclusion for *diri* in Riau Indonesian. However, under the present analysis, where *diri* is actually a phrase containing the null possessor DP *pro*, the same data needs a different explanation, e.g., the adjacency requirement on Accusative Case assignment (Ramli 1995).

(i)   a.   Apakah dia membunuh *diri* [di kebun]?
           Q   he kill   DIRI at garden
           'Did he kill himself in the garden?'
    b.   *Apakah dia membunuh [di kebun] *diri*?
           Q   he kill   at garden DIRI   (Chung 1976:44)

[12] (31b) is acceptable with the irrelevant reading 'Siti took a standing picture'.

b. *Siti mengambil [*gambar diri*].[13]
    Siti take   picture DIRI
    For: 'Siti took a picture of herself.' (cf. self-picture)

When *diri* is used as a verbal complement, the VP containing it describes an event that is generally self-directed rather than one that can be other-directed as well as self-directed. For example, *memukul diri* 'to hit oneself' in (32) describes a self-hitting action conducted for specific purposes (e.g. religious rituals), but not just any kind of self-hitting.[14]

(32) Ali_i kata $Siti_j$ memukul $diri_{*i/j/*k}$.
        Ali say Siti hit   DIRI
        'Ali said Siti hit herself.' (cf. self-hit)

(33) shows examples of VPs with *diri* that are commonly used.

(33)   a. membunuh diri   c. melibatkan diri
          kill   DIRI   involve   DIRI
          'to commit suicide'   'to get involved'
    b. melarikan/melepaskan   d. menyerahkan diri
          make.run/release   yield   DIRI
          diri   'to surrender'
          DIRI
          'to run away'

These phrases suggest that the generic meaning 'V one's physical self $\approx$ V oneself' makes a transitive verb semantically intransitive-like.[15] This quasi-intransitivization explains the inherently self-directed meaning of VPs with *diri*. It also provides an account for what appears to be local binding in (32), because an intransitivized transitive verb has reflexivity encapsulated in the verb meaning. Hence, one can treat the transitive VP *memukul diri* 'x hits one's physical self' in (32) as if it means 'x hits x' or, to put it more intransitively, 'x x-hits' ($\lambda x.x$ hits x).

6.   Conclusion

This paper has shown that a compositional analysis of Malay anaphoric expressions is not only possible but is empirically more desirable than a non-com-

---

[13] Examples of *gambar diri* were actually found in the Internet, though they were very rare. All attested examples were in Indonesian, and *gambar diri* had a specialized meaning, namely 'selfie', and hence was not associated with a particular individual.

[14] Not all speakers accept (32). It is acceptable only for those who know or can imagine some inherently self-directed hitting action. Nomoto (2011) thus reports a sentence with the same phrase, i.e. *memukul diri*, as ungrammatical.

[15] It is possible that a third *diri* morpheme exists that technically intransitivizes a transitive verb.

positional analysis, as it can account for a wider range of anaphoric expressions in a systematic manner. The complex anaphoric expressions comprising 'body' plus pronoun in some related languages are known to behave similarly to '*diri* + pronoun' in Malay. These languages include Malagasy (Paul 2004), Madurese (Davies 2008), Jambi (Cole et al. 2010; Kartono 2013), Javanese and Palembangese (Kartono 2013). The proposed analysis extends to these languages straightforwardly. By contrast, in non-compositional analyses, it remains unclear why the relevant binding properties are associated with 'body' + pronoun, but not with some other forms, and why the association is consistent across languages. In fact, Paul (2004) suggests a compositional analysis for *ny tenany* in Malagasy. This paper has developed her basic idea into a more complete and plausible hypothesis.

I would like to end this paper with an important empirical finding recently put forward by Kroeger (2014). He points out cases of '*diri* + pronoun + *sendiri*' (pattern (ii)) occurring in a subject position, where it cannot be bound locally.

(34)   *Diri-nya sendiri* selalu di-utamakan-nya.
         DIRI-3   own   always PASS-prioritize-3
         'Himself is always prioritized by him.' (i.e. 'He always gives priority to
         himself.')   (Kroeger 2014:18)

Examples like this can be real counterexamples to the previous generalization that pattern (ii) does not occur in subject positions (Cole and Hermon 2005). (34) is not an apparent counterexample of the type mentioned in footnote 1, where *dirinya* (pattern (i)) happens to be immediately followed by the adverbial *sendiri* meaning 'alone, by oneself', because *sendiri* in (34) cannot be moved to a clause-final adverbial position without changing the meaning. The status of pattern (ii) is extremely important in Malay syntax, as it is one of the few diagnostics for determining the precise phrase structure of a construction (see Kroeger 2014 for a relevant discussion, and Kartini and Nomoto 2012 for an example of an argument for a particular syntactic structure based on pattern (ii)). While a non-compositional analysis like Cole and Hermon's (2005) cannot handle examples like (34), the present analysis should be able to handle them. The key lies in understanding the function of *sendiri* in the subject position, which is something that is glossed over in this study and needs to be explored in future research.

References

Alsagoff, Lubna. 1992. Topic in Malay: The other subject. Doctoral dissertation, Stanford University.

Chung, Sandra. 1976. An object-creating rule in Bahasa Indonesia. *Linguistic Inquiry* 7:41–87.

Cole, Peter, and Gabriella Hermon. 1998. Long distance reflexives in Singapore Malay: An apparent typological anomaly. *Linguistic Typology* 2:57–77.

Cole, Peter, and Gabriella Hermon. 2005. The typology of Malay reflexives. *Lingua* 115:627–644.

Cole, Peter, Gabriella Hermon, and Yanti. 2010. Anaphora in traditional Jambi Malay. In *Austronesian and theoretical linguistics*, ed. Raphael Mercado, Eric Potsdam, and Lisa deMena Travis, 327–344. Amsterdam: John Benjamins.

Davies, William D. 2008. Madurese reflexives with reference to Malay/Indonesian. *Lingua* 118:1603–1616.

van Gelderen, Elly. 2000. *A history of English reflexive pronouns: Person,* self, *and interpretability*. Amsterdam: John Benjamins.

Gil, David. 2001. Reflexive anaphor or conjunctive operator: Riau Indonesian *sendiri*. In *Long-distance reflexives*, ed. Peter Cole, Gabriella Hermon, and C.-T. James Huang, volume 33 of *Syntax and Semantics*, 83–118. San Diego: Academic Press.

Homer, Vincent. 2009. Backward control in Samoan. In *Proceedings of the Sixteenth Meeting of the Austronesian Formal Linguistics Association (AFLA 16)*, ed. Sandy Chung, Daniel Finer, Ileana Paul, and Eric Potsdam, 45–59.

Kartini Abd. Wahab, and Hiroki Nomoto. 2012. Passives without 'by' in Malay. Paper presented at the 19th Meeting of the Austronesian Formal Linguistics Association (AFLA 19).

Kartono, Bambang. 2013. A puzzle in binding: Half reflexives and locally bound pronouns. A comparative study of anaphoric systems in Indonesian, Javanese, Palembangnese, City Jambi and Village Jambi. Master's thesis, Utrecht University.

Keenan, Edward L. 2002. Explaining the creation of reflexive pronouns in English. In *Studies in the history of English: A millennial perspective*, ed. Donka Minkova and Robert P. Stockwell, volume 1, 325–355. Berlin: Mouton de Gruyter.

König, Ekkehard, and Peter Siemund. 2000. Intensifiers and reflexives. In *Reflexives: Forms and functions*, ed. Zygmunt Frajzyngier and Traci S. Curl, 41–74. Amsterdam: John Benjamins.

Kroeger, Paul. 2014. Passive agents in Malay: The binding properties and discourse functions of agentive =*nya*. In *Current trends in Malay linguistics*, 5–29 ed. Siaw-Fong Chung and Hiroki Nomoto, volume 57 of *NUSA*. Jakarta and Tokyo: Universitas Katolik Indonesia Atma Jaya and Tokyo University of Foreign Studies.

Nomoto, Hiroki. 2011. Teori tambatan dalam bahasa Melayu. *Toukyou Gaidai Tounan'ajiagaku* 16:1–16.

Paul, Ileana. 2004. NP versus DP reflexives: Evidence from Malagasy. *Oceanic Linguistics* 43:32–48.

Ramli Haji Salleh. 1995. *Sintaksis bahasa Melayu: Penerapan teori Kuasaan dan Tambatan.* Kuala Lumpur: Dewan Bahasa dan Pustaka.

RELATIVE CLAUSES IN POLYNESIAN
AND THE ROLE OF D[*]

Yuko Otsuka
University of Hawai'i
yotsuka@hawaii.edu

Using a modified DP movement approach to relativization, this paper examines various types of relative clauses in Polynesian: (1) headed relatives; (2) free relatives morphologically identical to headed relatives; (3) free relatives seemingly headed by a determiner; and (4) Tahitian headed relatives containing a determiner and an optional predicate marker *'o* between the head noun and the relative clause. It is argued that the presence or absence of a determiner in free relatives is determined by whether the free relative pronoun D is morphologically realized or receives null spell-out. Tahitian *'o*-relatives are analyzed as involving relativization of a nominal, instead of verbal, construction. The proposed analysis of Tahitian *'o* relatives sheds new light on the morphological difference between headed and free relatives in Tagalog, in which the former, but not the latter, contain a "linker" between the head noun and the relative clause.

1.   Introduction

This paper examines the morphological differences between headed and free relatives in some Polynesian (PN) languages. Free relatives in Polynesian typically occur in argument *wh*-questions as part of a pseudo-cleft construction. In some PN languages such as Hawaiian and Tahitian, free relatives differ from headed relatives with respect to the item occurring in the clause-initial position, as illustrated with Tahitian examples in (1). Imperfective aspect is expressed by *e* in headed relative clauses, but by *tē* in free relative clause. [1]

[1] Abbreviations: ABS= absolute, ACC = accusative, DEF = definite, DEM = demonstrative, DET= determiner, DIR = directional, ERG = ergative, EXCL = exclusive, FUT = future, IPFV = imperfective, LNK = linker, PFV = perfective, PL = plural, POSS = possessive, PRED = predicate, PRS = present, PST= past, REF = referential, REL = relative, SG = singular, TR = transitive, 1 = first person, 2 = second person, 3 = third person.

(1) TAHITIAN
  a. te ta'ata [*e*   hōpoi mai i   te fāraoa]
    DEF[2] man IPFV bring DIR ACC DEF bread
    'the man who will bring the bread' (Coppenrath and Prevost 1975:277)

  b. 'o   vau   [*tē*   taora 'i te   'ofa'i]
    PRED 1.SG IPFV throw ACC DEF stone
    'It is I who will throw stones.' (lit. the one who will throw stones is I.]
                   (Tryon 1970: 87)

    Not all PN languages show such a morphological difference between headed and free relative clauses. In Tongan, Niuean, and Māori, for example, free relatives are morphologically identical to the headed relative clause, as seen in the Tongan examples in (2).

(2) TONGAN
  a. mo   e   tohi [na'e lau 'e   Sione __]
    with REF book PST   read ERG John
    'with the book John read'

  b. ko   e   hā   [na'e lau 'e   Sione __]?
    PRED REF what PST   read ERG John
    'What did John read?' (lit. what John read is what?)

    The objective of this paper is twofold. First, using a version of DP movement approach to relativization proposed by Tonoike (2008), I propose a uniform analysis of all types of relative clauses in PN: (a) headed relatives; (b) free relatives with a determiner; and (c) free relatives without a determiner. Second, I put forward a historical account for the typological variation across PN languages regarding the morphosyntactic difference between headed and free relatives.

    One phenomenon that requires special treatment is an idiosyncratic behavior of headed relatives in Tahitian, in which the regular tense/aspect marker can be replaced with the corresponding form occurring in free relatives (e.g., *tē* vs. *e*) and the predicate marker *'o* may optionally appear between the head noun and the relative clause, as illustrated in (3).

---

[2] Whether the so called "definite" articles in PN indicate definiteness, referentiality, or specificity is a matter of debate. In this paper, I use the label DEF, following the traditional grammarians, but remain agnostic about the actual semantic analysis, except for Tongan, for which I use the label REF, as definiteness is indicated by other means (Anderson and Otsuka 2006).

(3) TAHITIAN

te	taʻata	(ʻo)	[tē/*e	hōpoi	mai	i	te	fāraoa]?
DEF	man	PRED	IPFV	bring	DIR	ACC	DEF	bread

'the man who will bring the bread'

I propose that these relative clauses are derived from a nominal construction, in which the predicate nominal is a free relative (Section 6). The proposed analysis of Tahitian headed relatives with *ʻo* sheds new light on the morphological difference between headed and free relatives in Tagalog, in which the former, but not the latter, contain a "linker" *ng/na* between the head noun and the relative clause. I propose that headed relatives in Tagalog are also derived from nominal constructions (rather than verbal ones) and that the so-called linker is an allomorph of *ang*, which is used to mark the predicate nominal in an embedded clause. The proposed analysis provides a unified account for both head-initial and head-final relatives in Tagalog without postulating any unusual (and unmotivated) operation such as remnant TP movement to derive head-final relatives (cf. Aldridge 2003).

2.   Basic Morphosyntax of PN languages

PN languages are predicate initial. In verbal constructions, the clause-initial position is occupied by a tense/aspect/mood marker (TAM). Noun phrases are always marked by a determiner (except in Samoan, in which definite plural nouns are zero-marked), which is preceded by a case marker. Western PN languages such as Tongan and Samoan show an ergative pattern of case marking. In Tongan, for example, ERG is marked by *ʻe* and ABS is marked by *ʻa*. The ABS marker is obligatory with proper names, but may be dropped when it is followed by a pronoun (personal as well as demonstrative) or a referential article *e* (which has an allomorph *he* used with ERG).

(4) TONGAN

a.
Naʻe	kata	ʻa	Sione.
PST	laugh	ABS	John

'John laughed.'

b.
Naʻe	lau	ʻe	Sione	(ʻa)	e	tohi.
PST	read	ERG	John	ABS	REF	book

'John read a book.'

In Eastern PN languages such as Hawaiian and Tahitian, canonical transitive construction shows a nominative-accusative pattern of case marking:

NOM is unmarked, while ACC is marked by a particle *i*. This is illustrated in (5) with Hawaiian examples.

(5) HAWAIIAN
  a. Ua   hele   au.
      PFV   go   1.SG
      'I went.'

  b. Ua   'ai ke   kanaka   i   ka   poi.
      PFV   eat DEF   man   ACC   DEF   poi
      'The man ate the poi.' (Elbert and Pukui 1979:39)

    In nominal constructions, the predicate nominal precedes the subject noun phrase, as shown in (6). The former is marked by a predicate marker, a reflex of Proto-Polynesian (PPn) *ko, and the latter appears in ABS/NOM.

(6) TONGAN
  Ko   e   faiako   'a   Sione.
  PRED   REF   teacher   ABS   John
  'John is a teacher.'

3.   Headed vs. Free Relatives in Polynesian

In this section, some morphological differences between the three types of relative clauses in PN are described: (a) headed relatives, (b) free relatives identical to headed relatives, and (c) free relatives different from headed relatives. The latter two types are in complementary distribution.

3.1. Headed Relative Clauses

Relative clauses in Polynesian are head-initial, that is, the relative clause follows the NP it modifies. Like simple declarative clauses, relative clauses contain a TAM in the clause-initial position. Relative clauses do not contain what can be analyzed as a relative pronoun. In Western PN languages, typically, the relative clauses are identical to simple declarative clauses except that the former contains a gap in lieu of the relativized NP.

(7) TONGAN
  a. Na'e   lau   'e   Sione 'a   e   tohi.
      PST   read   ERG   John ABS   REF   book
      'John read a book.'

b. mo   e   tohi [na'e   lau   'e   Sione __]
   with   REF   book   PST   read   ERG   John
   'with the book John read'

In Eastern PN as well as Niuean, a distinct set of TAMs (hereafter "dependent" TAMs) are used in relative clauses, as shown in (8) below. Morphological differences between the dependent TAMs and those used in matrix clauses are illustrated in Table 1 with Hawaiian examples.

(8) Hawaiian

   a. *Ua*   hae   ka 'īlio.
     PFV   bark   DEF dog
     'The dog barked.'

   b. ka 'īlio   [*i*   hae __]
     DEF dog   PFV   bark
     'the dog that barked' (Hawkins 1982:109)

		REGULAR	DEPENDENT
PERFECTIVE		ua	i
PRESENT		ke … nei	e … nei
IMPERFECTIVE	+CONTINUOUS	e … ana	e … ana
	−CONTINUOUS	e	e

TABLE 1. HAWAIIAN TAMS

3.2. Free Relatives Identical to Headed Relatives

Free relatives typically occur in argument *wh*-questions in PN languages, as the subject of a pseudo-cleft construction (cf. Potsdam and Polinsky 2011). In Tongan, Niuean, and Māori, free relatives are identical to headed relatives except for the absence of the head noun, as illustrated in (9)-(11).

(9) TONGAN

   ko   e   hā   [na'e   lau   'e   Sione __]? (cf. 7b)
   PRED   REF   what   PST   read   ERG   John
   'What did John read?' (lit. what John read is what?)

(10)   MĀORI

   a. te wahine   [*e*   tuu   mai   ra __].
     DEF woman   IPFV   stand   DIR.1   DEM.3
     ' the woman who is standing there.' (Biggs 1998:159)

b. Ko   wai   [*e*   haere __]?
   PRED who IPFV go
   'Who is to go?' (lit. the one that will go is who) (Harlow 2001:193)

(11)   NIUEAN

    a. ke he   tagata [*ka*   nofo __   i   kaina]
      to DEF   person FUT   stay   at home
      'to the person who's going to stay at home' (Seiter 1980:93)

    b. Ko   e   heigoa [*ka* tā   e   lautolu __]?
      PRED ABS   what FUT build ERG 3.PL
      'What are they going to build?' (Seiter 1980:110)

3.3. Free Relatives Different from Headed Relatives

In Hawaiian and Tahitian, free relatives contain an item different from the expected dependent TAMs. This is illustrated in (12) and (13) below, where the items in question, Hawaiian *kai* and Tahitian *tē*, are glossed with a question mark.

(12)   HAWAIIAN

    a. ka 'īlio [*i*   hae __]
      DEF dog   PFV   bark
      'the dog that barked' (Hawkins 1982:109)

    b. 'O   wau [*kai*   noho me   lākou].
      PRED 1.SG PFV? stay with 3.PL
      'I am the one who stayed with them.' (Creeland 1994:351)

(13)   TAHITIAN

    a. te   ta'ata [*e*   hōpoi mai i   te   fāraoa]
      DEF   man   IPFV bring DIR.1 ACC DEF bread
      'that man who will bring the bread'

                               (Coppenrath and Prevost 1975:277)

b. 'o   vau   [*tē*   taora   'i   te   'ofa'i][3]
   PRED   1.SG   IPFV?   throw   ACC   DEF   stone
   'It is I who will throw stones.' (Tryon 1970:87)
   [lit. the one who will throw stones is I]

At first glance, it appears that free relatives and headed relatives use different kinds of TAMs, e.g., *i* vs. *kai* in Hawaiian. However, what appears to be a special TAM can be analyzed as consisting of a definite article *ka*/*te* and a dependent TAM that is used in the headed relative clause: *te* + *e* → *tē* in Tahitian and *ka* + *i* → *kai* in Hawaiian. Thus, we may conjecture that at least historically, these items consist of two separate morphemes, although they are synchronically spelt as a single item.

Samoan permits both types of free relatives, as shown in (14). Samoan is a little different from Hawaiian and Tahitian, however. In the second type, the relative clause includes a regular TAM in addition to the item in question. In (14c), for example, the additional item *lē* is followed by a regular TAM, *na*. Thus, *lē* could be analyzed as a complementizer or a free relative pronoun. Samoan data will be discussed in section 5.3.

(14)   SAMOAN
   a. le   tamāloa   [na maitau-ina   tupe]
     DEF   man   PST count-TR   money
     'the man who counted the money' (Chung 1978:86)

   b. 'o   ai   [na   fasia __ le   maile]?
     PRED   who   PST   hit   DEF   dog
     'Who hit the dog?' (lit. the one that hit the dog is who?)
                   (Mosel and Hovdhaugen 1992:633)

   c. 'o   ai   [*lē*   na   fasia   a'u]
     PRED   who   REL?   PST   hit   1.SG
     'Who is the one who hit me?' (Mosel and Hovdhaugen 1992:433)

The above observation leads to two questions. First, assuming that these items historically consisted of two separate items, how should they be analyzed synchronically? Are they (a) special (third) type of TAMs, (b) complementizers, or (c) still comprised of two morphemes, article and TAM? Second, how can we account for the typological variation within PN with respect to the derivation of free relatives?

---

[3] Tryon (1970) uses *te* with a short vowel to represent this item. I use *tē* with a long vowel, following Coppenrath and Prevost (1975).

4.   Theoretical Assumptions

Two approaches to relativization have been proposed and widely accepted in the literature: the operator movement approach (Chomsky 1981, inter alia) and the raising approach (Kayne 1994). In the following analysis of PN relative clauses, I adopt a third approach, which is proposed by Tonoike (2008).

4.1. Operator Movement Approach

In the Operator (OP) movement approach, a relative clause is generated as an adjunct to an NP and contains an operator in an argument position in the base structure. This operator moves to [Spec, C] (*wh*-movement) and is realized as a relative pronoun if pronounced in this position. The gap in the relative clause is therefore a trace of the moved operator. While widely accepted, this approach has a technical problem in view of the Minimalist Program (Chomsky 1995 and subsequent work); namely, it requires a special operation/device (such as coindexation or predication) to establish the coreference between the head noun and the extraction site.

(15) $[_{DP}$ DP$_i$ $[_{CP}$ OP$_i$ $[_{C'}$ C $[_{TP}$ … <OP> …]]]]

```
                      DP
                    /    \
                 DP_i      CP
               the book   /  \
                      OP_i     C'
                      which   /  \
                            C      TP
                                  /  \
                               … <OP> …
                              I read <which>
```
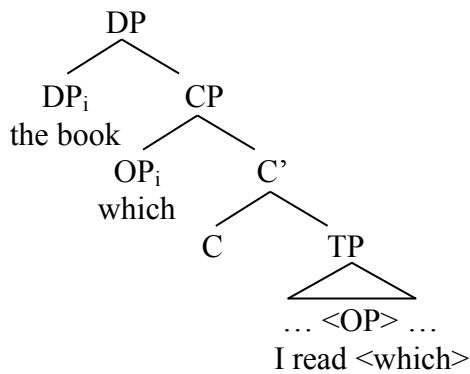
4.2. Kaynean Raising Approach

In contrast, Kayne (1994) proposes an analysis in which a relative clause is treated as a complement of D. In this approach, the "head noun" is base generated within the relative clause. For relative clauses without a relative pronoun, it is assumed that the base generated item is an NP and undergoes movement to [Spec, C], as schematically illustrated in (16a). For those with a relative pronoun, it is proposed that the item undergoing movement is a DP headed by a relative pronoun and that after raising to [Spec, C], the complement NP moves to [Spec, D], DP-internally. This derivation is shown in (16b).

(16) a. [$_{DP}$ D [$_{CP}$ NP [$_{C'}$ C [$_{TP}$ … <NP>…]]]]

```
              DP
            /    \
           D      CP
          the    /  \
              NP      C'
             book    /  \
                    C      TP
                  that    /\
                      … <NP> …
                     I read <book>
```

(16) b. [$_{DP1}$ D [$_{CP}$ [**DP2** NP [$_{D'}$ D <NP>] [$_{C'}$ C [$_{TP}$ … <DP2>…]]]]]

```
                DP1
              /     \
             D       CP
            the     /   \
                 DP2      C'
                /  \     /  \
              NP    D'  C     TP
             book  /  \      /\
                  D   <NP>  … <DP2> …
                which <book> I read <which book>
```
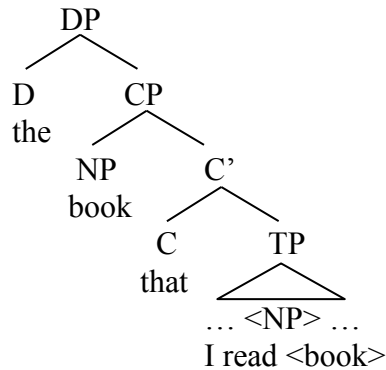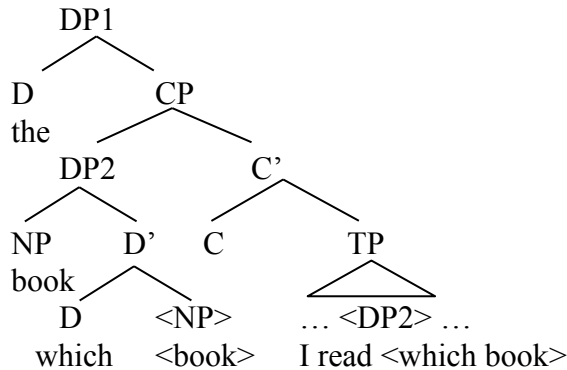
A significant advantage of this approach is that coreference is naturally explained in terms of movement. On the other hand, it comes with some disadvantages (Borsley 1997; also see Bianchi 2000 for counter-arguments). First, in analyzing the relative clause without a relative pronoun, the relevant argument is argued to be an NP rather than DP. This is in conflict with a widely accepted assumption that semantically, NP is a predicate and requires a DP layer to become an argument (Stowell 1991, Longobardi 1994). Also, the proposed derivation allows D to take a CP complement rather than an NP complement as standardly assumed. (See Donati and Cechetto 2011, which addresses this issue and proposes some modification to Kayne's model.) Second, in order to account for relative pronouns, it postulates a highly unusual operation of DP-internal NP movement that raises the complement NP to [Spec, D]. Third, syntactically, relative clauses behave like adjuncts: they are optional as adjuncts; they show the binding contrast observed by Van Riemsdijk and Williams (1981) and Freidin (1986), as illustrated in (17). Coindexation of the embedded subject *John* with the pronominal subject of the matrix clause is ungrammatical when

the embedded clause is a complement (17a), but permissible if the embedded clause is an adjunct (17b).

(17) a. Which claim [that John$_i$ was guilty] was he$_{*i/j}$ willing to discuss?
      b. Which claim [that John$_i$ made] was he$_{i/j}$ willing to discuss?

4.3. Modified DP Movement Approach

A third approach, which I call modified DP movement approach, has been proposed by Tonoike (2008). Similar to Kayne's analysis, this approach assumes that the head noun is base generated as DP in the argument position within the relative clause and undergoes movement to [Spec, C]. This DP subsequently undergoes sideward movement to create a new syntactic object (Nunes 2001), to which the relative CP adjoins. These three steps are indicated in the tree diagram below as (I) relativization, (II) sideward movement, and (III) adjunction.

(18)

It is assumed that each movement of DP leaves a copy of D alone. Thus, strictly speaking <DP> should be <D> in the diagram in (18). If the copy in [Spec, C] is pronounced, it is realized as a relative pronoun. In free relatives, the moved DP is headed by a free relative pronoun such as *what* in English and lacks NP complement. In this structure, only the highest copy is pronounced. The derivations of headed and free relatives are schematically shown in (19) and (20), respectively.

(19) [$_{DP}$ DP [$_{CP}$ <D> [$_{C'}$ C [$_{TP}$ … <D> …]]]]
      [$_{DP}$ the book][$_{CP}$ <the> [$_{C'}$ C [$_{TP}$ … <the> …]]]
                 ↓ Spell-Out
              which

(20)   [DP DP [CP <D> [C' C [TP … <D> …]]]]
       [DP what] [CP <what> [C' C [TP … <what> …]]]

This approach addresses the disadvantages of the previous approaches that have been just discussed. First, unlike the OP movement approach, coreference is naturally accounted for in terms of movement. Second, unlike Kayne's approach, relative clauses are analyzed as adjuncts. Third, relative pronoun is regarded as a spell-out of a copy of D and hence, there is no need to postulate unusual movement.

Additionally, this approach provides a natural explanation for resumptive pronouns. Consider Tongan as an example. Relativization of an ergative argument requires a resumptive pronoun, as shown in (21).

(21)   a. TONGAN: ERG ungrammatical gap
        *ki   he   fefine$_i$ [na'e   kai __$_i$ 'a   e   ika]
        to   REF   woman   PST   eat   ABS   REF   fish
        'to a woman who ate a fish'

(21)   b. TONGAN: ERG resumptive pronoun
        ki   he   fefine$_i$ [na'a   *ne*$_i$ kai 'a   e   ika]
        to   REF   woman   PST   3.SG   eat ABS   REF   fish
        'to a woman who (she) ate a fish.'

Since pronouns are D-heads, the present analysis with an assumption that movement leaves a copy of D instead of DP gives a natural account for why resumption is always pronominal. It is a spell-out of the lowest copy of D (i.e., the copy in the position in which its Case feature is valued).[4,5] In the operator movement analysis and Kayne's approach, the resumptive pronoun presumably occurs in lieu of the copy of OP and a copy of NP, respectively. Either way, it is not clear how such an item can and must be morphologically realized as a pronoun (i.e., D-head).

---

[4] I assume that Tongan has a language specific constraint that the lowest copy of D with a Case feature [ERG] must be pronounced. See Otsuka 2006 for discussion on how such constraint can be captured in terms of C's features.

[5] With respect to free relatives, however, a reasonable question arises as to how a copy of a phonetically null D (i.e., free relative pronoun) can possibly be pronounced. I assume that the free relative pronoun D has phi-feature with a default value of [3.SG] and a Case feature (valued in the course of derivation).

5.   Analysis of Relative Clauses in PN

In this section, the three types of relative clauses in PN languages are analyzed, using the modified DP movement approach.

5.1. Headed Relatives

Headed relatives behave the same way in all five languages under examination and can be analyzed as follows. A DP headed by a definite/referential article is base generated in the argument position and undergoes cyclic movement of relativization. The relevant derivation is illustrated in the bracketed structure in (22) with a Hawaiian example.[6]

(22)   HAWAIIAN [$_{DP}$ ka NP][$_{CP}$ <D> C [$_{TP}$ … <D> …]]
               [$_{DP}$ ka 'īlio][$_{CP}$ <ka> i [$_{TP}$ hae <ka>]]

5.2. Free Relatives

Free relatives come in two types. One preceded by a determiner (section 3.3) and the other, without (section 3.2). For the former, I propose that in these languages, the definite article may also function as a free relative pronoun, taking no NP complement. This is illustrated with a Hawaiian example in (23). In other words, I argue that what appears to be a special TAM in free relatives such as Hawaiian *kai* and Tahitian *tē* consists of two separate items in separate structural position, a determiner *ka*/*te* and a TAM.

(23)   HAWAIIAN [$_{DP}$ ka][$_{CP}$ <ka> C [$_{TP}$ … <ka> …]]
               [$_{DP}$ *ka*][$_{CP}$ <ka> *i* [$_{TP}$ noho <ka> me Kalei]]

        As for the latter, I propose the same derivation, with the only difference that the free relative pronoun is phonetically null. This is illustrated in (24) with a Tongan example. I assume V-to-T-to-C movement in Tongan (Otsuka 2000, 2011).

(24)   TONGAN   [$_{DP}$ D][$_{CP}$ <D> C   [$_{TP}$ … <D> …]]
               [$_{DP}$ Ø][$_{CP}$ <Ø> na'e lau [$_{TP}$ <na'e lau> 'e Sione <Ø>]]

---

[6] I remain agnostic about the position of the verb in Hawaiian, as it requires detailed examination to conclude whether the verb-initial order in Hawaiian is derived as a result of V-raising (as proposed by Otsuka (2000, 2011) for Tongan) or VP-raising (as proposed by Massam (2000, 2001) for Niuean).

5.3. Historical Account of the Typological Variations

The present analysis essentially proposes that the difference between the free relatives with a determiner and those without is not structural, but only morphological, namely, whether the free relative pronoun has an overt form (syncretic with the definite article) or phonetically null. In this section, I propose a diachronic account of how such a difference came about. My argument is that the difference resulted from two different kinds of reanalysis of the original construction.

As a point of departure, I argue that historically, the source of present day free relatives was headed relatives containing *me'a* 'thing' as a dummy head noun. Such a construction is synchronically available, as shown in (25).

(25) TONGAN
      Ko   e   hā   ('a)   e   me'a [na'e fakatau 'e Sione __]?
      PRED   REF what   (ABS) REF thing PST buy   ERG John
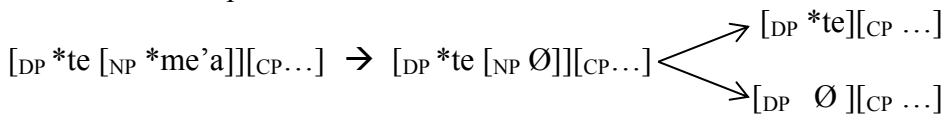   'What is the thing that John bought?' (lit. the thing John bought is what)

Superficially, free relatives are derived from this construction by omitting this dummy head noun *me'a*. In fact, this situation is referred to as "contraction" in Hawaiian within the pedagogical framework (Creeland 1994:350-351). In (26), (a)-sentences represent the "full" relative clause with the dummy head *mea* 'thing'; (b)-sentences are the contracted version, where this dummy head *mea* is omitted and the definite article *ka* and the TAM *e* (imperfective) or *i* (perfective) are fused into a single item so to speak, or at least orthographically represented as such.

(26) HAWAIIAN
      a. 'O   wau   [ka   mea   e   wala'au   me   lākou]
          PRED   1.SG   DEF   thing IPFV   talk   with   2.PL
          'I am the one who should talk with them.'

      b. 'O   wau   [ke   wala'au   me   lākou]
          PRED   1.SG   DEF.thing.IPFV talk   with   2PL

(27) HAWAIIAN
      a. 'O   Kanoa [ka mea i   noho   me   Kalei]
          PRED   Kanoa DEF thing PFV   stay   with   Kalei
          'Kanoa is the one who stayed with Kalei.'

      b. 'O   Kanoa [kai   noho   me   Kalei]
          PRED   Kanoa DEF.thing.PFV   stay   with   Kalei

Formally, this can be captured as replacement of the dummy N with a phonetically null N. Simple replacement would yield a construction in which a regular definite D is followed by a phonetically null N. I propose that this situation led to two different innovations. One is to allow the definite D to take no complement, whereby the definite D gained a new function as a free relative pronoun. This took place in Tahitian and Hawaiian. The other is to add a phonetically null free relative pronoun to the language's inventory of D's, while maintaining the constraint that an overt D must take an NP complement with phonological content. This option is taken by Tongan, Niuean, and Māori. The two scenarios are illustrated in (28) (reconstruction of PPn forms are based on Clark 1976).

(28)   Historical development of free relatives in PN

$$[_{DP} *te [_{NP} *me'a]][_{CP}\ldots] \rightarrow [_{DP} *te [_{NP} \emptyset]][_{CP}\ldots] \begin{cases} [_{DP} *te][_{CP} \ldots] \\ [_{DP}\ \emptyset\ ][_{CP} \ldots] \end{cases}$$

Samoan facts do not exactly fit this historical model. Recall that Samoan permits two kinds of free relatives, one with *lē* and one without. The formal similarity of this item *lē* and the definite article *le* gives us reasonable suspicion that Samoan *lē* free relatives are just another instance of free relatives with a definite article. Unlike free relatives in Hawaiian and Tahitian, however, Samoan *lē* free relatives additionally contain a regular TAM. The relevant examples are repeated here in (29).

(29)   SAMOAN
     a. 'o   ai   [na   fasia __   le maile]?(= 14b)
         PRED who   PST   hit   DEF dog
         'Who hit the dog?' (lit. the one that hit the dog is who?)
                                        (Mosel and Hovdhaugen 1992:633)

     b. 'o   ai   [*lē*   na fasia   a'u]   (= 14c)
         PRED who DET.REL   PST hit   1.SG
         'Who is the one who hit me?' (Mosel and Hovdhaugen 1992: 433)

In Mosel and Hovdhaugen 1992, this item is analyzed as bimorphemic, *l=ē*, where *ē* is called "relative preform". It turns out that this clause-initial element inflects for definiteness/specificity and number: *lē*, *ē*, *sē*, corresponding to a set of articles *le* [+DEF, SG], Ø [+DEF, PL], and *se* [–DEF]. This fact clearly indicates that the initial consonant in these items is D. The inventory of TAMs in Samoan does include *e*, which is synchronically analyzed as "generic"

and does not indicate any specific tense/aspect, but expresses "general truth" or "habitual action" (Mosel and Hovdhaugen 1992:366). It is plausible that Samoan went through a similar path as Hawaiian and Tahitian following the deletion of *meʻa: the definite article *le* acquired a new function as a free relative pronoun. When the TAM in the relative clause was *e*, the sequence *le* …*e* arose as a result. It seems, however, that Samoan took a step further. This sequence has been reanalyzed as a single unit, not only phonologically, but also morphosyntactically, resulting in a new type of D, a free relative pronoun *lē*. With this reanalysis, the relative clause in free relatives lacks a TAM, since the original TAM *e* no longer exists. This has led to the use of a regular TAM in free relatives. Once *lē* has been added to the inventory of determiners, gradually other forms such as *ē* for plural and *sē* for indefinite were added so as to fit the three-way paradigm of the determiners. If this speculation is on the right track, similar innovations may take place in future in other languages that use the definite determiner as a free relative pronoun.

6.   Tahitian *ʻo* in Headed Relatives

It should be noted at this point that the facts about headed relatives in Tahitian are a little more complicated than what has been described so far. Headed relatives may also contain *tei*/*tē* instead of the dependent TAMs *i*/*e*. Furthermore, and even more problematically, when headed relative clauses contain *tei*/*tē*, they may optionally be preceded by a predicate marker *ʻo*. See (30) below. The distribution of the two kinds of headed relatives seems to be governed by aspect. While Coppenrath and Prevost (C&P) (1975:277-278) state that it is "always possible to paraphrase the relative clause containing a 'relative pronoun' [*tei*/*tē*] as a relative clause without it", they also note that the omission of *te* is strongly preferred in continuous aspect. Tryon (1970:87-88) makes a stronger statement, namely, relative clauses in Tahitian contain just a dependent TAM in continuous aspect, but they contain *tē* or *tei* and is preceded by *ʻo* in non-continuous aspect.

(30)   TAHITIAN
    a. te   taʼata [*e*   hōpoi mai   i   te   fāraoa]
         DEF   man   IPVF   bring DIR.1 ACC   DEF   bread
         'thet man who will bring the bread' (C&P 1975: 277)

    b. te   taʻata (*ʻo*) [*tē*   hōpoi mai   i   te   faraoa]?
         DEF   man   PRED   DEF.IPFV   bring DIR.1 ACC   DEF bread
         'the man who will bring the bread' (C&P 1975: 268)

      Relative clauses like the one in (30b) raise two questions: (a) why *ʻo* is present between the head noun and the relative clause; and why the definite

article *te* (in the form *tē*) is used in the relative clause instead of a simple dependent TAM (*e* or *i*). I propose that these questions can be readily answered if we assume Tahitian headed relatives involve nominal constructions rather than verbal constructions. The relevant nominal construction is illustrated in (31).

(31)   TAHITIAN
[PRED ʻo   Tahiti] [SUBJ   tō   mātou   ʻāiʻa]
        PRED Tahiti   POSS   1.PL.EXCL country
'Our country is Tahiti.' (C&P 1975: 35)

Headed relatives with *ʻo te/tei* can be derived by relativizing the subject of the nominal construction in which the predicate nominal contains a free relative, as illustrated in (32).

(32)   a.   TAHITIAN: Nominal construction
              [PRED ʻo te [e   hōpoi mai   i   te   faraoa]]
                    PRED DEF IPFV bring DIR.1   ACC   DEF   bread
              [SUBJ   te   taʻata].[7]
                      DEF   man
              'the man is the one who will bring the bread.'

       b.   TAHITIAN: Relativization
              [DP te taʻata][CP <te> [ʻo te [e hōpoi mai i te faraoa]]] [SUBJ <te>]
              'the man who is the one who will bring the bread'

The presence of *ʻo* and *te* in (32b) has nothing to do with the operation of relativization per se, but is simply due to the fact that the relevant clause contains a nominal predicate which itself is a free relative. The optionality of *ʻo* may be understood as indication of two kinds of reanalysis: (a) reanalysis of the relevant construction as involving relativization of a verbal construction; and (b) reanalysis of *tē* as a general relative pronoun in a way similar to how Samoan *lē* came to function as a free relative pronoun (section 5.3).

       The present analysis of *ʻo* in Tahitian headed relatives has an interesting implication for the analysis of the reflexes of PPn *ko. In addition to marking predicate nominals, *ko is used in "appositional" constructions, as illustrated in (33).

---

[7] Though the reverse order may be preferred for this particular example, in general the subject DP and the predicate DP are interchangeable (C&P 1975:36).

(33)   a.   TAHITIAN
           Te   'oire   ra,   'o   pape'ete
           DEF   village   DEM   PRED   Papeete
           'the village (of) Papeete' (C&P 1975:39)

       b.   TONGAN
           (ki he) tamasi'i   ko   Sione
            to REF boy   PRED   John
           'to a boy (called) John'

       c.   HAWAIIAN
           (i) ka   luahine,   'o   Pele
           to DEF   old.woman   PRED   Pele
           'to the old woman, Pele' (Hawkins 1982: 56)

One may speculate that at least historically, such constructions were derived through relativization of the subject of the nominal construction. The process in question is illustrated in (34) with a Tongan example.

(34)   TONGAN
       a.   [PRED   Ko   Sione] [SUBJ   'a   e   tamasi'í].[8]
                   PRED   John   ABS   REF   boy.DEF
           'The boy is John.'

       b.   ('a) e   tamasi'i [CP <e> [C'   ko   Sione <e>]]
           ABS REF   boy   PRED   John
           'the boy, who is John'

8.   Tagalog Relative Clauses

The present analysis of Tahitian *'o*-relative constructions sheds new light on the analysis of relative clauses in Tagalog. Tagalog is also predicate initial. The syntactically prominent argument of a sentence is marked by a pre-nominal

---

[8] Synchronically, a more natural expression would be (i) topic construction or (ii), where the subject is more individuated.

(i)   ko   e   tamasi'i eni,   ko Sione.
      PRED REF boy   this PRED John
      'This boy here, (it is) John'

(ii) Ko   Sione 'a   e   tamasi'i na'e   'alú.
      PRED   John   ABS REF boy   PST   go.DEF
      'The boy who went was John.'

marker *ang*, which I analyze as a definite determiner here. Headed and free relatives show morphological differences in Tagalog, too, as illustrated in (35). Argument *wh*-questions contain a free relative, which is unmarked in a way similar to Tongan. In (35a), the free relative is preceded by *ang*, as it is the subject of the nominal construction. In headed relatives, a linker *ng*/*na* (phonologically conditioned allomorphs) is obligatory between the head noun and the relative clause, as in (35b). Note also that the linear positions of the head noun and the relative clause can be switched, as in (35c).

(35) a. [DP PRED ano] [DP SUBJ ang [CP b<in>ili ____ ni Maria]]
              what   DEF buy.PFV   DET Maria
          'What did Maria buy?' (lit. the thing Maria bought is what?)

      b. (ang) libro=*ng*   [b<in>ili ni Maria]
           DEF   book-LNK   buy.PERF DET Maria
          'the book Maria bought'

      c. (ang) [b<in>ili ni Maria]=*ng* libro
           DEF PFV -buy DET Maria-LNK   book
          'the book Maria bought'

Otsuka (to appear) proposes the following analysis of these facts about Tagalog relatives, using the modified DP movement approach. Tagalog is similar to Tahitian in that the free relative pronoun must be pronounced. Since in Tagalog only *ang*-marked DPs may undergo *wh*-extraction, I assume that the free relative pronoun in Tagalog is *ang* and hence the copy of D inside the relative clause is represented as *<ang>* in (36).[9]

(36) [DP ang] [CP <ang> [C' b<in>ili <ang> ni Maria]]
           DET   buy.PFV   DET Maria
      'the (thing which) Maria bought

Analysis of the headed relatives is not that straightforward, however. First, there is the obligatory linker, which curiously cannot occur in free relatives. Second, a successful analysis of Tagalog relative clauses must account for the free alternation between head-initial and head-final linear orders. I propose that unlike free relatives, headed relatives in Tagalog at least historically involve relativization of nominal constructions in which the

---

[9] Though it is represented as *ang* in the final position in this and following examples, the morphological realization of the D in the final position is not always *ang*. Similar to Samoan, its morphological form is determined by larger morphosyntactic context, showing inflection for Case (Schachter and Otanes 1975).

predicate DP is a free relative. Consider the nominal construction in (37a), whose structure is represented in (37b).

(37) a. ang libro ang b<in>ili=ko
     DEF book DEF buy.PFV=1.SG
     'The one I bought was the book.'

   b. [DP PRED ang libro] [DP SUBJ [D ang][CP <ang> [C' b<in>ili=ko <ang>]]]
     DEF book   DEF   buy.PFV=1.SG

Relativization of the subject in (37b) would result in the structure in (38) (the internal structure of the predicate DP is omitted).

(38) [DP [**DP** (ang) [CP binili=ko]] [CP <ang> [TP [DP_PRED ang libro] <ang>]]]

     'the one I bought which was the book'

The linear sequence of this awkward double relative construction is identical to a verbal construction: *binili-ko ang libro* 'I bought the book'. I suggest that the need to disambiguate these two constructions (relative clause vs. verbal construction) led to the replacement of *ang* in (38) with *ng*. When this replacement took place historically, this gave rise to an allomorph of *ang* which marked the predicate DP in a relative clause. Synchronically, this allomorphy may be formally understood as realization of C's clause-type feature on D: when its value is [Rel(ative)], it is spelled out as *ng*. Head-initial relatives can be derived in a similar fashion; only, the subject and predicate DPs are switched, as shown in (39). The replacement of *ang* with *ng* is also necessary here, as the linear sequence of the resulting head-initial relative is identical to a nominal construction: *ang libro ang binili=ko* 'what I bought was the book'.

(39) a. [DP_PRED   ang b<in>ili=ko] [DP_SUBJ   ang libro]
             DET buy.PFV=1.SG   DEF book
     'The book is the one I bought.'

   b. [DP [DP_SUBJ (ang) libro] [CP <ang> [TP [DP_PRED ang b<in>ili=ko]
     <ang>]]]

   c. [DP [DP_SUBJ (ang) libro] [CP <ang> [TP [DP_PRED *ng* b<in>ili=ko]
     <ang>]]]
     'the book which is the one I bought'

The lack of linker in free relatives is also accounted for in this analysis. Since free relatives involve verbal constructions, they do not contain a predicate DP marked by *ang*. Therefore, it is neither necessary nor possible for *ng* to arise as a replacement of *ang*.

(40)   [$_{DP}$ [$_D$ ang] [$_{CP}$ <ang> [$_{TP}$ b<in>ili=ko   <ang>]]]
                 DEF   buy.PERF=1.SG
         'the (thing which) I bought'

An apparent problem for this analysis is that the structure in (39a) is synchronically considered ungrammatical (Schachter & Otanes 1976:530). Nevertheless, I argue that it is reasonable to assume that nominal constructions like (39a) historically existed (Starosta, Pawley, and Reid 1982; Kaufman 2009). According to Starosta, Pawley, and Reid (1982), such constructions were reanalyzed as verbal constructions, giving rise to the synchronic *binili-ko ang libro* 'I bought *the book*', in which 'the book' is in focus. This essentially is semantically identical to 'the book is what I bought', the intended meaning of (39a). It is thus reasonable to conjecture that (39a) is synchronically generable, but that it is blocked because of the corresponding verbal construction.

   It should be noted that the linker *ng/na* appears to mark noun modifiers of any type, including adjectives and numerals, and that these modifiers can also occur before or after the noun that is modified: *dalawa-ng libro* vs. *libro-ng dalawa* for 'two books', *mabuti-ng tao* vs. *tao-ng mabuti* for 'good person'. In the present analysis, these modifier-noun constructions are also treated as relative clauses: e.g., 'the books that are two', 'the person who is the good one', in which the predicate DP is marked by an allomorph *ng/na* of the definite determiner *ang*.

   Aldridge (2003) proposes an alternative analysis of Tagalog headed relative clauses using Kayne's model. Head-initial relatives are derived by means of NP movement to [Spec, CP] (41a). As for head-final relative clauses, Aldridge proposes subsequent remnant TP movement to [Spec, DP] to derive (41b).

(41)   a.   [$_{DP}$ [$_{CP}$ book [$_{TP}$ Maria bought t$_{book}$]]]
        b.   [$_{DP}$ [$_{TP}$ Maria bought t$_{book}$] [$_{D'}$ [$_{CP}$ book [$_{C'}$ t$_{TP}$]]]]

In addition to the lack of clear theoretical motivation for the optional remnant TP movement, Aldridge's analysis has two problems. First, it does not explain the categorial status as well as the structural position of the linker. In the structure proposed in (41a), the linker *ng* occurs below [Spec, C] and above TP in head-initial relatives. Therefore, one may hypothesize that it is a complementizer. (Incidentally, sentential complements of verbs such as *think*

are introduced by a homophonous item, *ng/na*, even showing the same allomorphy). However, this hypothesis fails to account for the position of the linker in head-final relatives such as (35c), which precedes the head noun in [Spec, C]. If the linker were C in the structure proposed in (41b), the expected order would be *binili ni Maria libro-ng* instead. Note also that it fails to account for the lack of linker in free relatives. The second problem is the position of the determiner in head-final relatives. If the structure in (41b) is correct, this D is located below the TP which has undergone movement to [Spec, D]. Thus, the predicted linear order would be *[binili ni Maria] ang libro* (ignoring the linker) instead of the actual *ang [binili ni Maria]-ng libro*.[10]

9. Conclusion

I have examined two phenomena concerning relativization in PN languages: (a) morphological differences between headed and free relatives in languages like Hawaiian and Tahitian; (b) crosslinguistic variation within PN regarding this morphological difference between headed and free relatives. Using the modified DP movement approach (Tonoike 2008), I have argued that the items occurring in free relatives in Hawaiian (*kai* and *ke*) and Tahitian (*tei* and *tē*) in lieu of a regular TAM should be regarded as consisting of two independent items, a definite determiner and a dependent TAM. I have also proposed a historical account for the observed typological variation. Namely, the difference between two types of free relatives, those containing the definite article and those which do not, arose due to two innovations that took place separately in relevant languages: (a) addition of a new function (free relative pronoun) to the definite article *te; and (b) addition of a phonetically null free relative pronoun to the inventory of D's.

Intriguingly, the distribution of the two types of free relatives does not coincide with any subgrouping or typological classification (e.g., case alignment and presence or absence of dependent TAMs), as shown in Table 2. The present diachronic analysis of PN free relatives leads to an interesting observation about syntactic innovations in general: syntactic innovations are arbitrary in terms of motivation as well as direction and therefore, cannot be used as evidence for subgrouping. This very point is supported by another observation made in this study: Tagalog and Tahitian utilize nominal constructions as a major source of relativization, but no other PN languages show evidence of such a strategy.

---

[10] See Law (this volume) for another alternative approach. He argues (a) headed relatives in Tagalog all have the same structure, the only difference being which copy is pronounced; (b) linker is a "category-less morpheme filling any empty head position." As such, it occurs as C in head-initial relatives and as D in head-final relatives). Law's basis for the latter claim is that neither *ang* nor *ng* may appear to mark the head N of the head-final relatives.

	Tongan	Niuean	Samoan	Māori	Hawaiian	Tahitian
D in free realtives	−	−	±	−	+	+
Subgroup	Tongic	Tongic	Ellicean	EP	EP	EP
Case	ERG	ERG	ERG	ACC	ACC	ACC
Dependent TAMs	−	+	−	+	+	+

TABLE 2. DISTRIBUTION OF TWO TYPES OF FREE RELATIVES IN PN

References

Aldridge, Edith. 2003. Remnant movement in Tagalog relative clause formation. *Linguistic Inquiry* 34:631-640.

Anderson, Victoria and Yuko Otsuka. 2006. The phonetics and phonology of "definitive accent" in Tongan. *Oceanic Linguistics* 45:33-53.

Bianchi, Valentina. 2000. The raising analysis of relative clauses: A reply to Borsley. *Linguistic Inquiry* 31:123-140.

Borsley, Robert. 1997. Relative clauses and the theory of phrase structure. *Linguistic Inquiry* 28:629–647.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Ma: MIT Press.

Chung, Sandra. 1978. Case marking and grammatical relations in Polynesian. Austin, Tx: University of Texas Press.

Clark, Ross. 1976. *Aspects of Proto-Polynesian syntax*. Auckland: Linguistic Society of New Zealand.

Coppenrath, Hubert and Paul Prevost. 1975. *Grammaire approfondie de la langue tahitienne: Ancienne et moderne*. Papeete: Librairie Pureora.

Creeland, Hōkūlani. 1994. *'Ōlelo 'ōiwi*. Honolulu: 'Aha Pūnana Leo.

Donati, Caterina and Carlo Cechetto. 2011. Relabeling heads: A unified account for relativization structures. *Linguistic Inquiry* 42:519-560.

Elbert, Samuel H. and Mary Kawena Pukui. 1979. *Hawaiian grammar*. Honolulu: University of Hawai'i Press.

Freidin, Robert. 1986. Fundamental issues in the theory of binding. *Studies in the acquisition of anaphora*, ed. Barbara Lust, 151-188. Dordrecht: Reidel.

Harlow, Ray. 2001. *A Māori reference grammar*. Auckland: Longman.

Hawkins, Emily. 1982. *Pedagogical grammar of Hawaiian: Recurrent problems*. Honolulu: Hawaiian Studies Program, University of Hawai'i at Mānoa.

Kaufman, Daniel. 2009. Austronesian nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35:1-49.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, Ma: MIT Press.

Longobardi, Giuseppe. 1994. Reference and proper names. *Linguistic Inquiry* 25:609-665.

Massam, Diane. 2000. VSO and VOS: Aspects of Niuean word order. *The syntax of verb initial languages,* eds. Andrew Carnie and Eithne Guilfoyle, 97-116. Oxford: Oxford University Press.

Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language and Linguistic Theory* 19:153-197.

Mosel, Ulrike and Evan Hovdhaugen.1992. *Samoan reference grammar*. Oslo, Norway: Scandinavian University Press.

Nunes, Jairo. 2001. Sideward movement. *Linguistic Inquiry* 31:303-344.

Otsuka, Yuko. 2000. *Ergativity in Tongan*. Doctoral thesis, Oxford University.

Otsuka, Yuko. 2006. Syntactic ergativity in Tongan: Resumptive pronouns revisited. *Ergativity: Emerging issues*, eds. Alana Johns, Diane Massam, and Juvenal Ndayiragije, 79-107. Dordrecht: Springer.

Otsuka, Yuko. 2011. PRO vs. null SE: Case, tense, and empty categories in Tongan. *Syntax* 14:265-296.

Otsuka, Yuko. To appear. On Tagalog relative clauses. *In untiring pursuit of better alternatives*, eds. Hiroki Egashira, Hisatsugu Kitahara, Kazuo Nakazawa, Tadao Nomura, Masayuki Oishi, Akira Saizen, Motoko Suzuki. Tokyo: Kaitakusha..

Potsdam, Eric and Maria Polinsky. 2011. Questions and word order in Polynesian. *Topics in Oceanic morphosyntax*, eds. Claire Moyse-Faurie and Joachim Sabel, 83-109. Berlin: Mouton de Gruyter.

van Riemsdijk, Henk and Edwin Williams. 1981. NP-Structure. *The Linguistic Review* 1:171-217.

Schachter, Paul and Fe Otanes. 1972. *Tagalog reference grammar*. Berkley: University of California Press.

Starosta, Stanley, Andrew Pawley, and Lawrence Reid. 1982. The evolution of focus in Austronesian. *Papers from the Third International Conference on Austronesian Linguistics 1: Tracking the travellers*, eds. Amran Halim, Lois Carrington, and S. A. Wurm, 145-170. Canberra: Pacific Linguistics.

Stowell, Tim. 1991. Determiners in NP and DP. *Views on phrase structure*, eds. Katherine Leffel and Denis Bouchard, 37-56. Dordrecht: Kluwer.

Tonoike, Shigeo. 2008. DP movement analysis of relativization. Ms., University of Hawaiʻi, Honolulu.

Tryon, Darrell T. 1970. *Conversational Tahitian*: *An introduction to the Tahitian language of French Polynesia*. Canberra: Australian National University Press.

CONTROL INTO NOMINALIZED COMPLEMENTS IN MALAGASY*

Eric Potsdam   Maria Polinsky
University of Florida   Harvard University
potsdam@ufl.edu   polinsky@fas.harvard.edu

We consider the analysis of control structures in Malagasy in which the complement clause is introduced by the morpheme *ny*. *Ny* is the default determiner in Malagasy and we argue that such structures instantiate control into nominalized clauses. We consider and reject the dominant position in the literature that *ny* in such control structures has been reanalyzed as a complementizer.

1.   Introduction

Control phenomena have figured prominently in the generative syntax literature, going back to Rosenbaum 1967. One development in the history of obligatory control in particular is the inclusion of other languages and phenomena beyond the canonical case of obligatory control into nonfinite clauses seen in English. For example, recent work has documented and analyzed control into finite complement clauses (Landau 2004 and others). A phenomenon that is known from English and other languages but which has received relatively little attention is control into nominalized clauses (see Stiebels 2007:32–33 and Landau 2013:43–46). Examples from Arabic and Q'eqchi' are in (1), where nominalization of the complement clause is variously indicated by case morphology, a nominalization affix, and/or a determiner. We will call this NOMINAL CONTROL.

(1)   a.   STANDARD ARABIC
           Ziyaad   qarrara   **l**-raħiil-**a**.
           Ziyad   decided.3MSG   **DEF**-leaving-**ACC**
           'Ziyad decided to leave.'

---

b. Q'EQCHI'

N-in-lub	chi	x-mesunk-**il**	li	kabl.
PRS-1SG.ABS-tire	COMP	3SG.ERG-sweep-**NMLZ**	the	house

'I am tired of sweeping the house.'   (Kockelman 2003:(13))

The goal of this paper is to argue for the existence of nominal control in Malagasy, a VOS Austronesian language spoken on the island of Madagascar. A representative example is given in (2b), alongside ordinary control in (2a). It is the presence of the determiner *ny* which suggests a nominal control analysis.

(2) a. Nanandrana   hisambotra   ilay akoho   ny   ankizy.
        PAST.try.AT   IRR.catch.AT   DEM   chicken DET children

    b. Nanandrana   **ny**   hisambotra   ilay akoho   ny   ankizy.
        PAST.try.AT   **DET**   IRR.catch.AT   DEM   chicken DET children
        'The children tried to catch that chicken.'

The paper is structured as follows. We begin in section 2 with a discussion of Malagasy complement control patterns. Existing analyses of (2b) claim that it is not in fact nominal control but ordinary control with the formative *ny* having been reanalyzed as a complementizer (Randriamasimanana 1986, 2007, Ntelitheos 2012, 2013). Section 3 argues against this approach, attempting to establish that *ny* in such examples is a determiner. Section 4 presents our conclusions.

2.   Malagasy Control

Malagasy is an Austronesian language of the Malayo-Polynesian subgroup. It is typically described as having VOS word order with a Philippine-style symmetrical voice system. The precise analysis of Malagasy word order and clause structure is actually of some debate (see Pearson 2005 and references therein for discussion) so for concreteness we will assume the following picture: the core of a clause is a fully saturated predicate. The order of elements within this predicate is V + SUBJECT + OBJECT + OBLIQUE + ADJUNCT. From within the predicate, one element, called the TRIGGER, externalizes to the clause-final position. Following Pearson 2005, this externalization is A' movement. Verbal "voice" morphology registers the underlying grammatical role of the trigger. There are three voices: actor topic (AT) in which the subject is the trigger, (3a), theme topic (TT) in which the object is the trigger, (3b), and circumstantial topic (CT) in which an oblique element is the trigger, (3c). Subjects in non-actor topic clauses appear immediately after the verb, phonologically "bonded" to it. This is indicated in the orthography by a hyphen or apostrophe.

(3) a. N-i-antso   mpiasa   i   Mery.
           PAST-AT-call   worker   DET Mary
           'Mary called the worker.'

      b. N-antso-in'   i   Mery   ny   mpiasa.
           PAST-call-TT   DET Mary   DET worker
           'Mary called the worker.'

      c. N-i-antso-an'   i   Mery   mpiasa   ny   kiririoka.
           PAST-CT-call-CT DET Mary   worker   DET whistle
           'Mary called the worker with the whistle.'

Control verbs in Malagasy allow a number of patterns and have been much discussed in the generative syntax literature (Keenan 1976, 1995, Randriamasimanana 1986, 2007, Law 1995, Paul and Ranaivoson 1998, Pearson 2001, Polinsky and Potsdam 2002, 2003, 2005, Potsdam 2006, 2009, Ntelitheos 2006, 2012, 2013, Potsdam and Polinsky 2007). A typical paradigm for a subject control verb is given below.

(4)   Nanandrana   ø/mba/ny   hisambotra   akoho   Rasoa.
            tried.AT   Ø/COMP/NY   IRR.catch.AT   chicken Rasoa
            'Rasoa tried to catch a chicken.'

In (4), the actor topic control verb *manandrana* 'try' selects a clause-like complement.[1] This clause may be introduced paratactically, a construction we call BARE CONTROL. The complement clause may also be introduced by the complementizer *mba* (Randriamasimanana 1986, Potsdam and Polinsky 2007), what we will call *MBA*-CONTROL. Finally, the complement clause may be introduced by the formative *ny*, which we will argue below is a determiner and thus instantiates nominal control. To not prejudge the situation, we will for now call this *NY*-CONTROL. The complements introduced by these elements will be called bare clauses, *mba*-clauses, and *ny*-clauses, respectively.

The goal of this paper is to investigate more closely the *ny*-control pattern. As far as we have been able to determine, the control structure with *ny* is allowed with all control verbs, as is the bare control option. This contrasts with *mba*-control, which is only available with certain verbs. For example, *mba* is permitted with 'remember' but not with 'refuse':

---

[1] All verbs in Malagasy show morphological tense: *n(o)-* 'PAST', *h(o)-* 'FUT/IRR', and *ø-* 'PRES' (Pearson 2001). To first approximation, tense marking in controlled clauses follows two patterns: For some control verbs, the embedded tense marking must be future/irrealis; for others, there must be tense matching with the matrix verb (Paul and Ranaivoson 1998:121; Randriamasimanana 1986, 2007; Ntelitheos 2006:309–317).

(5) a. Nahatadidy ø/ny/mba hanidy ny varavarana ianao.[2]
   remember ø/NY/COMP lock DET door 2SG
   'You remembered to lock the door.'

   b. Nanda ø/ny/*mba hamafa trano ny vehivavy.
   refuse ø/NY/COMP sweep house DET woman
   'The woman refused to clean the house.'

It is widely recognized since Williams 1980 that there are two types of control constructions: obligatory control (OC) and non-obligatory control (NOC). There are a number of well-known diagnostics for distinguishing the two (see Williams 1980, Hornstein 1999, Landau 2013, among others for discussion). They include the following:

(6)   characteristics of OC
   a. requires a linguistic controller
   b. requires a local controller
   c. requires a c-commanding controller
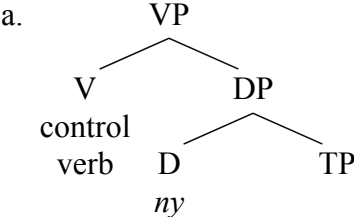   d. prohibits a strict reading under ellipsis

We have elsewhere shown that the other options in (4) are also OC (Polinsky and Potsdam 2003 for bare control and Potsdam and Polinsky 2007 for *mba*-control). The data in (7–10) confirm that *ny*-control instantiates OC as well.

(7)   Mikasa ny hanadio ny lapany ny andriana.
   intend.AT NY clean DET castle.3SG DET king
   'The king intends to clean his castle.'
   (lit.: intends the cleaning of his castle)
   *'The king intends for someone to clean his castle.'

(8)   Mihevitra Rasoa fa mikasa ny handao an' i Tana
   think Rasoa that intend NY leave ACC DET Tana
   ny governemanta.
   DET government
   'Rasoa thinks that the government intends to leave Tana.'
   *'Rasoa thinks that the government intends for her to leave Tana.'

(9)   Mikasa ny hanambady an- dRasoa ny fianakavian- dRabe.
   intend NY marry ACC Rasoa DET family Rabe
   '#Rabe's family intends to marry Rasoa.'
   *'Rabe's family intends for him to marry Rasoa.'
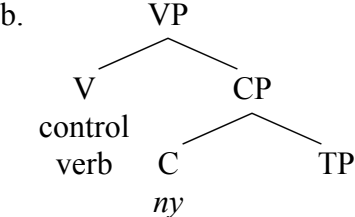
---

[2] Here and below, all verbs are in the actor topic form unless otherwise indicated.

(10) Nanaiky   ny   hividy   ity   trano   ity   ny   mpitsara.
      agree   NY   buy   DEM   house   DEM   DET judge
      Toa   izany   koa   ilay   mpampiasa.
      CONJ that   also DEM   employer
      'The judge agreed to buy this house and the employer also agreed to buy this house.'
      *'The judge agreed to buy this house and the employer also agreed for the judge to buy this house.'

The primary issue discussed in the literature regarding *ny*-control is the status of the formative *ny*. Ntelitheos 2006:284ff discusses *ny*-control and calls *ny* and the following clause a nominalization. He provides arguments, some of which we review in section 3, that *ny* and the following clause constitute DPs. In particular, he points out that *ny* is the default determiner in Malagasy (see Paul 2009). We will call this the DETERMINER HYPOTHESIS. On the other hand, Randriamasimanana 1986:498ff, 2007 and later Ntelitheos 2012, 2013 assert that *ny* is a complementizer. We will call this the COMPLEMENTIZER HYPOTHESIS. These two hypotheses for the structure of examples like (2b) are shown in (11). Under the determiner analysis, *ny* is a D° and projects a DP. Under the complementizer analysis, *ny* belongs to the lexical category C° and projects a CP.

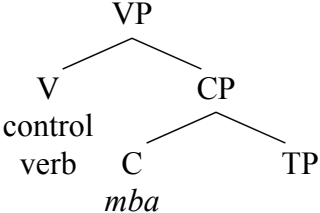(11)   DETERMINER HYPOTHESIS   COMPLEMENTIZER HYPOTHESIS
       a.   VP   b.   VP
             _____|_____   _____|_____
            V   DP   V   CP
         control   _____|_____   control   _____|_____
          verb   D   TP   verb   C   TP
                       ny   *ny*

Our primary goal is to reinforce Ntelitheos' original position that *ny*-control does involve nominal control, with the formative *ny* being a determiner. To do this, we will systematically compare *ny*-control and *mba*-control, as we take *mba* to be an uncontroversial complementizer:

(12)   VP
          _____|_____
         V   CP
      control   _____|_____
       verb   C   TP
                mba

3.   Evidence for Nominal Control

This section presents arguments in favor of the determiner hypothesis. We consider facts related to Malagasy's determiner inventory, complement selection, advancement to trigger, fronting, coordination, comparatives, and islandhood.

3.1.   Malagasy determiners

There is a well-known formal restriction in Malagasy that triggers must have a determiner (Keenan 1976, 2008, Pearson 2001, Paul 2000, 2009, Law 2006, and others). The basis for this restriction is not clear but it accounts for the data in (13).

(13)   Nihomehy   *(ny)   zaza.
         laugh   DET   child
         'The child laughed.'

Ny is the default determiner used in such cases, but its semantic contribution is not well-defined. Even in other positions, its semantic contribution, if any, is sometimes not evident (Law 2006, Keenan 2008, Paul 2009), as shown by (14). It is thus the expected determiner in the control context.

(14)   Tia   (ny)   boky   frantsay   aho.
         like   DET   book   French   1SG
         'I like French books.'   (Paul 2009)

Ntelitheos 2012 observes that if *ny* is a determiner in *ny*-control, then given the structure in (11), one expects to see other determiners in the same position. Malagasy has a large number of demonstratives that encode number, distance from the speaker, and visibility (Rahajarizafy 1960:24, Rajemisa-Raolison 1969:53, Paul 2009). Demonstratives may be used alone in a prenominal position or as framing demonstratives that have an identical demonstrative both preceding and following the noun phrase:

(15) a. izany zaza   b. ireo   olona   ireo
        DEM   child   DEM   person   DEM
        'that child'   'those people'

Ntelitheos 2012:293 offers the following data showing that demonstratives may not replace *ny* in *ny*-control structures.

(16) a. *Mameno **ilay** handeha ho any Antsiranana Rabe.
       wish   DEM go   PREP LOC Antsiranana Rabe
      ('Rabe wished to go to Antsiranana.')

    b. *Nikasa **ity** hanoratra taratasy **ity** Rabe.
       intend DEM write   letter   DEM Rabe
      ('Rabe intended to write a letter.')

While our consultants also rejected the above examples, which would argue against the determiner hypothesis, they did accept control clauses with the determiner *izany*:

(17) a. Mikasa ny hanasa alika i Aina
      intend NY wash   dog   DET Aina
      'Aina intends to wash the dog.'

    b. Mikasa **izany** hanasa alika i Aina
      intend DEM wash   dog   DET Aina
      'Aina is contemplating washing the dog.'

Izany is the most vague demonstrative, encoding something that is invisible and indefinitely far from the speaker. If *izany* projects a DP in such examples it supports the determiner hypothesis. The impossibility of other demonstratives may arise because they are too lexically specified to be used with an event; for example, they may have an unwanted deictic interpretation.

3.2. Complement selection

A claim of the determiner hypothesis is that all verbs that participate in nominal control subcategorize for a DP complement and thus should allow a non-clausal DP complement. This appears to be the case. The examples in (18) are representative.

(18) a. Tsy nahatadidy ny pepetra aho.
      NEG   remember DET rule   1SG
      'I don't remember the rules.'

    b. Nitsahatra ny ady ny fahavalo.
      stop   DET war   DET enemy
      'The enemies stopped the war.'

    c. Mihevitra ny fiainana aho.
      think   DET life   1SG
      'I am thinking about life.'

    d. Mikasa   izany   ireo   mpiasa   ireo.
        intend   DEM   DEM   worker   DEM
        'The workers intend that.'

    e. Te   hanandrana   ny   sakafony   aho.
        want   try   DET   food.3SG   1SG
        'I want to try his food.'

The complementizer hypothesis, in contrast, predicts no correlation: we might expect *ny* to be lexically selected like *mba* 'COMP', as we saw above.[3]

3.3. Subject position

A test for nominal status in Malagasy is the ability to appear in trigger position. Only nominals can be triggers (Keenan 1976, Paul 2000:95–96, Potsdam 2011).[4] *Ny*-clauses in *ny*-control can advance to trigger position, (19b) (Paul and Ranaivoson 1998:117, Netlitheos 2013:7). Comparison to *mba*-clauses and bare infinitives suggests that *ny*-clauses are in fact nominal. *Mba*-clauses and bare infinitives cannot be triggers, (19c).

---

[3] We did find one verb that allows *ny*-control but does not seem to take a DP complement. The verb *mirohotra* means 'to rush to V, to do all at once/together'. It allows various control patterns but excludes a DP complement:

(i)   a.   Mirohotra   ø/ny/mba   mivoaka   ny   trano   ny   olona.
           rush   ø/NY/COMP exit   DET   building DET   people
           'The people are rushing to exit the building.'

        b.   *Mirohotra   {izany,   ny   tsinjaka   vaovao} ny   olona.
            rush   DEM   DET   dance   new   DET   people
           ('The people rushed to that/the new dance.')

        c.   Mirohotra   ho   amin'   izany   ny   olona.
           rush   PREP PREP   that   DET   people
           'The people are rushing to that.'

[4] One potential exception is argument clauses, which seem to show up as triggers, (i). As Keenan (1976:254, 285–286) indicates however, it is rather difficult to show that such clauses are triggers and not extraposed complements as suggested by the alternative extraposition translation.

(i)   Mazava/Antenaiko   [fa   efa   lasa   ny   mpianatra].
           clear/hope.TT.1SG   that   already   gone   DET   student
           'That the students already left is clear/hoped by me.'
           'It is clear/hoped by me that the students already left.'   (Keenan 1976:254)

(19) a. Mila [ø/ny/mba   hividy   sira]   ny   mpahandro.
      need   ø/NY/COMP buy   salt   DET cook

   b. Ilain'   ny   mpahandro [ny hividy   sira].
      need.TT DET cook   NY buy   salt

   c. *Ilain'   ny   mpahandro [(mba) hividy   sira].
       need.TT DET cook   COMP buy   salt
      'The cook needs to buy salt.'

The contrast suggests that *ny*-clauses and *mba*-clauses do not have the same categorial status. If they did, the above contrasts would be mysterious. The contrast makes sense if *ny*-clauses are DPs in trigger position and *mba*-clauses are CPs that cannot advance to trigger position.

    A similar pattern occurs with object control predicates. One might expect that *ny* control would be impossible with object control verbs if *ny* clauses are nominal because that would require the verb to select two DP complements. In fact, Malagasy is a double object language: ditransitive complement frames can be expressed with either a DP PP frame or a DP DP frame (Randriamasimanana 1986, Pearson 2000):

(20) a. Nanome   voankazo   ho   an' ny   gidro aho.
      give   fruit   PREP PREP DET lemur 1SG
      'I gave some fruit to the lemur.'

   b. Nanome   voankazo   ny   gidro aho.
      give   fruit   DET lemur 1SG
      'I gave the lemur some fruit.'

It is thus not surprising that *ny* clauses are possible with object control verbs, even if they are nominals:[5]

---

[5] Randriamasimanana notes the ungrammaticality of *ny*-clauses in the following object control examples, which our consultants also rejected. We have no explanation for this difference.

(i) a. *Niangavy an' i Jeanne   ny   handeha i   Marie.
      ask   ACC DET Jeanne   NY   go   DET Marie
      'Marie asked Jeanne to go.'   (Randriamasimanana 2007:(10a))

   b. *Nanery an' i Jeanne   ny   hanasa an' i Jaona i Paoly.
      force   ACC DET Jeanne   NY   wash   ACC DET John   DET Paul
      'Paul forced Jeanne to wash John.'   (Randriamasimanana 1986:536)

(21) a. Niteny   ahy   ø/ny/mba   hianatra   teny   gasy
        tell   1SG.ACC   ø/NY/COMP   learn   language   Malagasy
        ny   lehibeko.
        DET   boss.1SG
        'My boss told me to learn Malagasy.'

    b. Nampahatsiahy   ahy   ø/ny/mba   hanidy   ny   varavarana
        remind   1SG.ACC   ø/NY/COMP   lock   DET   door
        i   Soa.
        DET   Soa
        'Soa reminded me to lock the door.'

As with subject control verbs, the *ny*-clause can be promoted to subject position; the object control verb appears in its circumstantial topic form:

(22) a. Nitenenan'   ny   lehibeko   ahy   ny   hianatra
        tell.CT   DET   boss.1SG   1SG.ACC   NY   learn
        teny   gasy.
        language   Malagasy
        'To learn Malagasy was told to me by my boss.'

    b. Nampahatsiahivan'   i   Soa   ahy   ny hanidy ny   varavarana.
        remind   DET   Soa   1SG.ACC   NY lock   DET   door
        'To lock the door was reminded to me by Soa.'

The ability of *ny*-clauses to advance to trigger position strongly supports their status as DPs.

3.4.   Fronting

Malagasy has two fronting constructions which, to first approximation, target only triggers (Keenan 1976, Paul 2000, 2002, Sabel 2002, Pearson 2005, others). In each, the trigger is fronted and followed by a particle. The particle *no* is used to focus the fronted phrase and the particle *dia* is used to topicalize the fronted phrase:

(23) a. Manasa   lamba   Rasoa.
        wash   clothes   Rasoa
        'Rasoa is washing clothes.'

    b. Rasoa   **dia**   manasa   lamba.
        Rasoa   **TOP**   wash   clothes
        'Rasoa, she's washing clothes.'

c. Rasoa **no** manasa lamba.
   Rasoa **FOC** wash clothes
   'It's Rasoa who is washing clothes.'

The determiner hypothesis correctly predicts that *ny*-clauses can front since they can be triggers, (24) (see similar examples in Randriamasimanana 1986:507, Keenan 1995:196, and Ntelitheos 2013:290).

(24) a. Heverin- dRabe [ny hividy fiara].
   think.TT Rabe NY buy car
   'Rabe is thinking of buying a car.'

b. Ny hividy fiara no heverin- dRabe.
   NY buy car FOC think.TT Rabe
   'It's buy a car that Rabe is thinking of doing.'

c. Ny hividy fiara dia heverin- dRabe.
   NY buy car TOP think.TT Rabe
   'Buying a car, Rabe is thinking of doing that.'

The complementizer hypothesis would not allow such examples because such CPs could not first become triggers. As expected, *mba*-clauses and bare infinitives cannot front:

(25) a. * (Mba) hividy fiara no heverin- dRabe.
   COMP buy car FOC think.TT Rabe
   ('It's buy a car that Rabe is thinking of doing.')

b. * (Mba) hividy fiara dia heverin- dRabe.
   COMP buy car TOP think.TT Rabe
   ('Buying a car, Rabe is thinking of doing that.')

3.5. Comparatives

Comparatives provide a test for DP status. Potsdam 2011 argues that the standard of comparison in comparatives must be a DP with an overt determiner, as suggested by the representative examples in (26).

(26) a. Nividy laoranjy betsaka [noho *(ny) akondro] Rabe.
   bought orange many than DET banana Rabe
   'Rabe bought more oranges than bananas.'

b. Betsaka   kokoa   ny   vola   nangalarin' ny   olon-dratsy
   many   more   DET   money   steal.TT   DET thief
   tamin- dRasoa [noho *(ny) tamin' ny   sakaizany].
   PREP   Rasoa   than   DET   PREP DET friend.3SG
   'The thief stole more money from Rasoa than from her friend.'

The determiner hypothesis correctly predicts that *ny*-clauses will appear as the standard of comparison, (27). The complementizer hypothesis would rule out such examples.

(27)   Mankahala (kokoa) ny   mipasoka lamba [noho   ny mamafa
        hate   more   NY   iron   clothes than   NY sweep
        trano]   Rasoa.
        house   Rasoa
        'Rasoa hates to iron clothes more than to sweep the house.'

3.6.   Double passives

(28a) illustrates what Randriamasimanana 1986 calls the DOUBLE PASSIVE CONTROL construction (see also Polinsky and Potsdam 2005). Both the control verb and the embedded verb are in the theme topic form and the embedded clause object is the matrix trigger. The construction is used to front the object of the embedded clause, (28b), because, as we saw, only triggers can be fronted.

(28) a. Nandraman- dRasoa hosamborana   ny   akoho.
         tried.TT-   Rasoa IRR.catch.TT   DET chicken
         lit. "The chicken was tried by Rasoa to be killed.'
         'Rasoa tried to catch the chicken.'

     b. Inona   no   nandraman- dRasoa   hosamborana?
         what   FOC   tried.TT-   Rasoa   catch.TT
         'What did Rasoa try to catch?'

Ny is excluded from the double passive construction while *mba* is not, (29). We provide an account for this different below but it already suggests a difference between *ny*-clauses and bare/*mba*-clauses.

(29)   Nandraman- dRasoa ø/mba/*ny   hosamborana   ny   akoho.
        tried.TT-   Rasoa ø/COMP/NY   catch.TT   DET chicken
        'Rasoa tried to catch the chicken.'

Our explanation for the contrast in (29) crucially relies on the DP status of *ny*-clauses. It is based on the novel claim regarding Malagasy syntax in (30).

DPs are islands and movement cannot take place from within a DP. This includes advancement to trigger.

(30)   DPs are islands

This claim is supported by a number of constructions. First, we suggest that Malagasy *Tough*-Movement (Keenan 1976) involves actual movement and displaces a constituent from the complement clause of a *tough*-predicate to the matrix trigger position:

(31) a. Sarotra   vakina   ilay   boky.
         difficult   read.TT DEM   book
         lit. "That book is difficult to be read.'
         'That book is difficult to read.'

    b. Sarotra [vakina ~~ilay boky~~]   ilay boky.
        difficult read.TT   DEM   book

Evidence that the final DP is the matrix trigger comes from the fact that it can be fronted using the constructions discussed in section 3.4 and can be preceded by the question particle *ve*, which appears between the predicate and the trigger (Keenan 1995):

(32) a. Ilay   boky   no   sarotra   vakina.
         DEM   book   FOC   difficult   read.TT
         'It's that book that is difficult to read.'

    b. Sarotra   vakina   ve   ilay   boky?
        difficult   read.TT Q   DEM   book
        'Is that book difficult to read?'

If DPs are islands, then we correctly predict that the complement to *tough*-predicates cannot contain *ny*:

(33)   *Sarotra   ny   vakina   ilay   boky.
          difficult NY   read.TT   DEM   book
         lit. "That book is difficult to be read.'
         ('That book is difficult to read.')

   A second instance of movement in Malagasy is the Subject-to-Object Raising (SOR) construction shown in (34b) (Keenan 1976, Paul and Rabaovololona 1998, Pearson 2001).

(34) a. Nanantena Rabe [fa nanasa ny zaza Rasoa].
 hope Rabe that wash DET child Rasoa
 'Rabe hoped that Rasoa washed the child.'

 b. Nanantena an-dRasoa [ho nanasa ny zaza] Rabe.
 hope ACC-Rasoa COMP wash DET child Rabe
 'Rabe hoped Rasoa washed the child.' (Keenan 1976:283)

Pearson 2001:150ff argues that the embedded subject undergoes A' movement:

(35) Nanantena an-dRasoa [ho nanasa ny zaza ~~Rasoa~~] Rabe.
 hope ACC-Rasoa COMP wash DET child Rabe

As with *tough*-movement, *ny* is not possible in complement clauses with SOR:

(36) *Nanantena an-dRasoa [ny nanasa ny zaza ~~Rasoa~~] Rabe.
 hope ACC-Rasoa NY wash DET child Rabe
 ('Rabe hoped Rasoa washed the child.')

 Two other potential movement environments in Malagasy are Subject-to-Subject Raising (Flegg and Paul 2003), shown in (37), and Possessor Raising (Keenan 1976, Keenan and Ralalaoherivony 1998) shown in (38). In neither case can movement take place if the domain of extraction is introduced by *ny*.

(37) Manomboka [(*ny) avy ~~ny orana~~] ny orana.
 begin NY come DET rain
 'It's beginning to rain.'

(38) Maty [(*ny) vady ~~Rabe~~] Rabe.
 dead NY spouse Rabe
 'Rabe was widowed.' (Keenan and Ralalaoherivony 1998:69)

 Given this generalization, the determiner hypothesis correctly predicts that *ny*-clauses will be islands for advancement to trigger. This accounts for the impossibility of double passives with *ny*, (29). The embedded object cannot advance to trigger position out of the DP. It is able to do so if the embedded clause is introduced by a complementizer, either *mba* or zero, as CPs are not islands.

3.7. Summary

We have presented a number of phenomena that identify *ny*-clauses as DPs and *mba*-clauses as CPs, supporting the determiner hypothesis. This is in agreement

with Ntelitheos 2006 but contra Randriamasimanana 1986, 2007 and Ntelitheos 2012, 2013.[6]

Although the data in the previous sections strongly support the claim that *ny*-clauses can be DPs and cannot only be CPs, they do not in fact rule out a dual analysis in which *ny*-clauses are ambiguous between DP and CP. That is, allowing *ny*-clauses to <u>also</u> be CPs is also compatible with the data. This is in line with Ntelitheos's (2013) claim that *ny* is undergoing reanalysis from determiner to complementizer. The only decisive argument against a dual analysis is based on the double passive and islandhood of DP. If *ny*-clauses were also CPs, one would not expect the contrast between *ny* and *mba/ø* with respect to the double passive. There should be an analysis of clauses with *ny* that pattern with *mba*. At the same time, this is the most theory-internal argument.

Before concluding, we consider a semantic argument for the complementizer analysis from Ntelitheos 2013. It is based on the observation that *ny* does not seem to make a semantic contribution to *ny*-control structures. Speakers generally indicate that examples are synonymous with and without *ny*. Ntelitheos does note one weak semantic difference, in (39), however.

(39) a. Nikasa   (?ny)   hanasa   ny   lamba   Rasoa
       intend   NY   wash   DET   clothes   Rasoa
       fa   narary   tampoka   izy.
       but   ill   suddenly   3SG
       'Rasoa intended to wash the clothes but she suddenly became ill.'

   b. Nikasa   ?(ny)   hanasa   ny   lamba   Rasoa
       intend   NY   wash   DET   clothes   Rasoa
       fa   tsy   vitany   intsony   izany.
       but   NEG   complete.3SG   anymore   DEM
       'Rasoa intended to wash the clothes but they weren't finished by her.'
                            (Ntelitheos 2013:(45, 46))

In (39a), an event of washing did not take place and *ny* is dispreferred on this unrealized event. In (39b), in contrast, an event of washing was initiated and *ny* is preferable. This contrast is understandable if *ny* is contributing specificity or definiteness to an event of washing clothes. At the same time, he notes that the contrast is not systematic and not all speakers are sensitive to it. Some of our consultants did not perceive this difference. Ntelitheos concludes that "*ny* in control complements does not imply definiteness or specificity as in normal

---

[6] Ntelitheos 2006 provides two further arguments in favor of the determiner hypothesis from coordination and object shift. We were not able to fully replicate the supporting data. For reasons of space, we do not discuss them here.

noun phrases, exhibiting semantic bleaching of the sort illustrated with the English complementizer *that* (Ntelitheos 2013:8)".

While his observation about the lack of a semantic contribution from *ny* may be well-founded, the conclusion that it is therefore a complementizer is not necessarily warranted. Law 2006 and Paul 2009 discuss the interpretation of *ny* and suggest that it signals familiarity not definiteness/specificity; however, this interpretation can be overridden in certain contexts.

4.   Conclusion

We conclude that Malagasy does instantiate obligatory nominal control, as in the example below:

(40)   Nanandrana [DP ny   hisambotra ilay akoho] ny   ankizy.
         tried   DET catch   DEM chicken DET children
         'The children tried to catch the chicken.'
         (lit.: tried the catching of the chicken)

Ny in such examples can be a determiner that heads a DP. We were not able to rule out a complementizer analysis for *ny* suggesting that *ny* may in fact be undergoing a category change in this environment (Ntelitheos 2012, 2013). The one decisive argument against the complementizer hypothesis was based on the novel claim that DPs in Malagasy are islands, a claim that requires further investigation.

Nominal control in Malagasy instantiates obligatory control, as we showed in section 2. English nominal control, in contrast, is non-obligatory control (Landau 2013:43–46). It is not yet clear what the relationship is between control and nominalized clausal complements cross-linguistically. This is another issue in need of further investigation.

References

Flegg, Jill Heather, and Ileana Paul. 2003. On the differences between raising and control. *Cornell Working Papers in Linguistics* 19, 50–63. Department of Linguistics, Cornell University, Ithaca, NY.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30:69–96.

Keenan, Edward. 1976. Remarkable subjects in Malagasy. *Subject and topic*, ed. Charles Li, 247–301. New York: Academic Press.

Keenan, Edward L. 1995. Predicate-argument structure in Malagasy. *Grammatical relations: Theoretical approaches to empirical questions*, eds. Clifford S. Burgess, Katarzyna Dziwirek, and Donna B. Gerdts, 171–216. Stanford: CSLI.

Keenan, Edward. 2008. The definiteness of subjects and objects in Malagasy. *Case and grammatical relations*, eds. Greville Corbett and Michael Noonan, 241–261. Amsterdam: John Benjamins.

Keenan, Edward, and Baholisoa Ralalaoherivony. 2000. Raising from NP in Malagasy. *Lingvisticae Investigationes* 23:1–44.

Kockelman, Paul. 2003. The interclausal relations hierarchy in Q'eqchi' Maya. *IJAL* 29:25–48.

Landau, Idan. 2000. *Elements of control: Structure and meaning in infinitival constructions*. Dordrecht: Kluwer Academic Publishers.

Landau, Idan, 2004. The scale of finiteness and the calculus of control. *Natural Language and Linguistic Theory* 22:811–877.

Landau, Idan. 2013. *Control in generative grammar: A research companion.* Cambridge: Cambridge University Press.

Law, Paul. 1995. On grammatical relations in Malagasy control structures. *Grammatical relations: Theoretical approaches to empirical questions*, eds. Clifford S. Burgess, Katarzyna Dziwirek, and Donna B. Gerdts, 271–290. Stanford: CSLI.

Law, Paul. 2006. Argument marking and the distribution of wh-phrases in Malagasy, Tagalog, and Tsou. *Oceanic Linguistics* 45:153–190.

Ntelitheos, Dmitrios. 2006. The morphosyntax of nominalizations: A case study. Doctoral dissertation, University of California, Los Angeles.

Ntelitheos, Dmitrios. 2012. *Deriving nominals: A syntactic account of Malagasy nominalizations*. Leiden: Brill Publications.

Ntelitheos, Dimitrios. 2013. Subordination through nominalization: Det as Comp in Malagasy control complements. *Morphologie, syntaxe et sémantique des subordonnant*, 151–166. Clermont Ferrand, France: Presses universitaires Blaise Pascal.

Paul, Ileana. 2000. Malagasy clause structure. Doctoral dissertation, McGill University.

Paul, Ileana. 2002. On extraction asymmetries. *MIT Working Papers in Linguistics* 44, 211–224. Department of Linguistics and Philosophy, MIT, Cambridge, Ma.

Paul, Ileana. 2009. The presence versus absence of determiners in Malagasy. *Determiners: Universals and variation*, eds. Jila Ghomeshi, Ileana Paul, and Martina Wiltschko, 215–242. Amsterdam: John Benjamins.

Paul, Ileana, and Lucie Rabaovololona. 1998. Raising to object in Malagasy. *UCLA Occasional Papers in Linguistics* 20, 50–64. Department of Linguistics, UCLA, Los Angeles, Ca.

Paul, Ileana, and Jeannot F. Ranaivoson. 1998. Complex verbal constructions in Malagasy. *UCLA Occasional Papers in Linguistics* 20, 111–125. Department of Linguistics, UCLA, Los Angeles, Ca.

Pearson, Matthew. 2001. The clause structure of Malagasy: A minimalist approach. Doctoral dissertation, University of California, Los Angeles.

Pearson, Matthew. 2005. The Malagasy subject/topic as an A' element. *Natural Language and Linguistic Theory* 23:381–457.

Polinsky, Maria, and Eric Potsdam. 2002. Backward control: Evidence from Malagasy. *MIT Working Papers in Linguistics* 44, 257–272. Department of Linguistics and Philosophy, MIT, Cambridge, Ma.

Polinsky, Maria, and Eric Potsdam. 2003. Control in Malagasy. *Cornell Working Papers in Linguistics* 19, 173–187. Department of Linguistics, Cornell University, Ithaca, NY.

Polinsky, Maria, and Eric Potsdam. 2005. Malagasy control and its theoretical implications. *Berkeley Linguistics Society* 30, 365–376. Berkeley Linguistics Society, Berkeley, Ca.

Potsdam, Eric. 2006. Backward object control in Malagasy: Against an empty category analysis. *The Proceedings of WCCFL 25*, 328–336. Cascadilla Proceedings Project, Somerville, Ma.

Potsdam, Eric. 2009. Malagasy backward object control. *Language* 85:754-784.

Potsdam, Eric. 2011. A direct analysis of Malagasy comparatives. *Tampa Papers in Linguistics* 2, 1–24. University of South Florida, Tampa, Fl.

Potsdam, Eric, and Maria Polinsky. 2007. Missing complement clause subjects in Malagasy. *Oceanic Linguistics* 46:277–303.

Rahajarizafy, Antoine. 1960. *Essai sur la grammaire malgache*. Antananarivo: Imprimerie Catholique.

Rajemisa-Raolison, Régis. 1969. *Grammaire malgache*. Fianarantsoa: Librairie Ambozotany.

Randriamasimanana, Charles. 1986. *Causatives of Malagasy*. Honolulu: University of Hawaii Press.

Randriamasimanana, Charles. 2007. Malagasy control structures. *Concentric: Studies in Linguistics* 33:91–120.

Rosenbaum, Peter. 1967. *The grammar of English predicate complement constructions*. Cambridge, Ma.: MIT Press.

Sabel, Joachim. 2002. Wh-questions and extraction asymmetries in Malagasy. *MIT Working Papers in Linguistics* 44, 309–323. Department of Linguistics and Philosophy, MIT, Cambridge, Ma.

Stiebels, Barbara. 2007. Towards a typology of complement control. *ZAS Papers in Linguistics* 47, 1–80. ZAS, Berlin.

Williams, Edwin. 1980. Predication. *Linguistic Inquiry* 11:203–238.

SPLIT FOCUS PREFERENCES IN TAGALOG: EVIDENCE FROM CHILD LANGUAGE [*]

Nozomi Tanaka
University of Hawaiʻi at Mānoa
nozomit@hawaii.edu

William O'Grady
University of Hawaiʻi at Mānoa
ogrady@hawaii.edu

Kamil Deen
University of Hawaiʻi at Mānoa
kamil@hawaii.edu

Chae-Eun Kim
University of Hawaiʻi at Mānoa
chaeeun@hawaii.edu

Ryoko Hattori
University of Hawaiʻi at Mānoa
rhattori@hawaii.edu

Ivan Paul M. Bondoc
University of the Philippines Manila
imbondoc@up.edu.ph

Jennifer U. Soriano
University of the Philippines Manila
jusoriano@up.edu.ph

This study makes use of an elicited production task to investigate child and adult preferences for particular focus patterns in Tagalog declarative sentences and relative clauses. Our findings point to a general preference for theme focus over agent focus in declarative sentences on the part of both children and adults, but a preference among children for relative clauses with agent focus.

1. Introduction

A signature feature of Tagalog is its so-called focus system,[1] in which the verb is marked with an affix indicating the thematic role of one of its NP arguments, which in turn is marked by the case prefix *aŋ* and often said to be 'focused.' In this study, we will investigate two core patterns of this sort: agent focus (AF)

---

[1] Although the terms "voice" or "trigger" are also sometimes used for this phenomenon, we choose the term "focus" in this paper.

and theme focus (TF). The examples in (1)–(3)[2] help illustrate how the system works.

(1)   T<um>akbo   aŋ   lalake.
        <AF.PFV>run   +FOC   boy
        'The boy ran.'

(2)   B<um>ili   naŋ   libro   aŋ   lalake.
        <AF.PFV>buy   -FOC   book   +FOC   boy
        'The boy bought a book.'

 (3)   B<in>ili   naŋ lalake   aŋ   libro.
        <TF.PFV>buy   -FOC boy   +FOC   book
        'A boy bought the book.'

In the case of the intransitive sentence in (1), the verb carries the agent focus infix *-um-*, reflecting the fact that its sole argument (the agent *lalake* 'boy') is focused, as indicated by the case prefix *aŋ*. Whereas intransitive sentences allow just this one option, patterns containing a two-place predicate permit the alternation exemplified in (2) and (3). In (2), the agent argument is focused, as shown by its occurrence with *aŋ* and by the presence of agent focus infix *-um-* on the verb. In (3), in contrast, the theme argument is focused: it is marked by *aŋ* and triggers the theme focus infix *-in-* on the verb.

      Relative clauses (RC) in Tagalog are introduced by the 'linker' *-ŋ* and follow the noun that they modify.[3]

(4)   lalaki=[RC ŋ   t<um>akbo   __]
        boy=LNK   <AF.PFV>run
        'the boy that ran'

A key feature of Tagalog is that only focused arguments can be relativized. Thus, as illustrated below, only the agent argument can be relativized in agent focus patterns, and only the theme argument can be relativized in theme focus patterns.

(5) a. Relativization of the agent in an agent focus pattern (acceptable):
        lalaki=[RC ŋ   b<um>ili   naŋ   libro   __]
        boy=LNK   <AF.PFV>buy   -FOC book
        'the book that the boy bought'

---

[2] Abbreviations: AF = agent focus; +FOC = focus marker; -FOC = non-focus marker; IPFV = imperfective; LNK = linker; PFV = perfective; RC = relative clause; TF = theme focus
[3] We only show head-initial relative clauses here; however, Tagalog also has head-final and head-internal relative clauses (Aldridge 2004).

b. Relativization of the agent in a theme focus pattern (unacceptable):

*lalaki=[RC ŋ	b<in>ili	__	aŋ	libro]
boy=LNK	<TF.PFV>buy		+FOC	book

'the book that the boy bought'

(6) a. Relativization of the theme in a theme focus pattern (acceptable):

libro=[RC ŋ	b<in>ili	naŋ	lalake __]
book=LNK	<TF.PFV> buy	-FOC	boy

'the book that the boy bought'

b. Relativization of the theme in an agent focus pattern (unacceptable):

*libro=[RC ŋ	b<um>ili	__	aŋ	lalake]
book=LNK	<TF.PFV> buy		+FOC	boy

'the book that the boy bought'

Despite its unusual and intriguing features, the Tagalog focus system has received little attention in the literature on language acquisition. The few studies that have been done on declarative sentences (Tucker 1971, Segalowitz and Galang 1978) suffer from complications that make their results difficult to interpret, and there have been no studies at all on relative clause patterns. We attempt to remedy this situation by addressing the following two questions: (i) What is the preferred focus pattern in declarative sentences for children and adults? (ii) What is the preferred relative clause pattern for children and adults? We report on our findings in the next two sections.

2. Experiment 1—declarative clauses

Experiment 1 aimed to answer the first research question by using a picture-based elicited production task to examine adults' and children's production of declarative clauses.

2.1. Method

2.1.1. Participants

We studied 14 children aged 4;2 to 5;3 (mean 4;6) as well as 31 adult controls.

2.1.2. Materials

The experiment consisted of four conditions crossing the definiteness of agents and themes (definite agent and definite theme, definite agent and indefinite theme, and so on). Definiteness was established with the help of a context that included a picture and a background sentence (see Figures 1 and 2 below).

Half the items within each condition contained an animate agent and an animate theme involved in a semantically reversible event: hugging, carrying, pinching, pulling, and pushing.[4] The other half involved an animate agent and an inanimate theme involved in a semantically nonreversible event: cutting, eating, picking, washing, and reading.

For the purposes of this paper, we will restrict our attention to test items in which both agent and theme are definite and animate. Because focused arguments tend to be definite in Tagalog (Reid and Liao 2001), this condition offers participants an opportunity to cast their response as either an agent focus or a theme focus pattern, thereby revealing any preferences that they might have. A further advantage of this sort of test item is that it permits a direct comparison with relative clauses, the second structure type with which we are concerned. As we will see in section 3, the clause type normally used in studies of relative clauses also contains two animate definite arguments.

2.1.3. Procedure

Each test item began with presentation of a picture such as the one in Figure 1, accompanied by the background sentence (in Tagalog) "Here is a family: father, brother, sister, and mother," to ensure that each character would be familiar to the participants (and hence definite).

Figure 1. Context picture of the father, brother, sister, and mother in Condition 2.

Participants were then presented with the actual test item, which consisted of a picture such as the one illustrated in Figure 2, along with the query "What is happening here?" In this particular case, participants are expected to respond by

---

[4] An event is semantically reversible if it is equally plausible for either participant to function as agent or theme. Thus a sentence such as *The boy hugged the girl* is semantically reversible since, in principle, a boy could hug a girl or a girl could hug a boy. In contrast, *The boy read the book* is not reversible, since a book cannot read a boy.

reporting that the father is carrying the brother, giving us the opportunity to observe a possible preference for either agent focus or theme focus.

Figure 2. Target picture in Condition 2. The target answer is a Tagalog equivalent of "The father is carrying the brother."

2.2. Results

After removing irrelevant answers, we were left with 31 responses from adults and 19 from children. As shown in Figure 3, both groups manifested an overwhelming preference for theme focus (96.8% for adults, 84.2% for children).
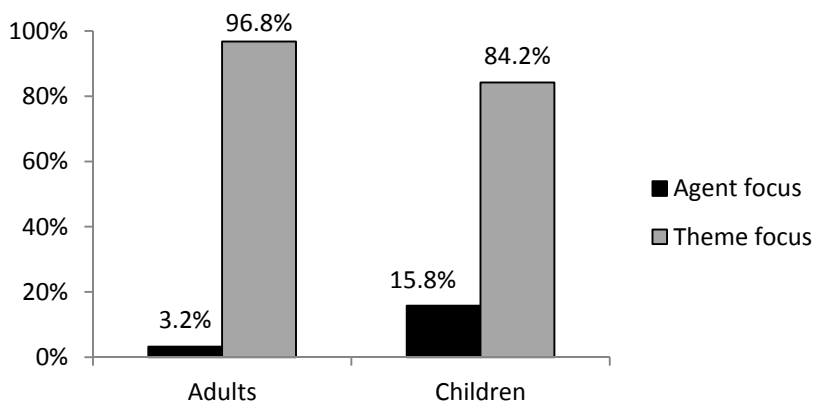
Figure 3. Adults' and children's focus preference when agent and theme are both [+animate] [+definite].

This asymmetry strongly suggests that theme focus is overall the preferred pattern in both adult and child production, consistent with reports in the literature that theme focus in the default in Tagalog (Payne 1982, De Guzman 1988, Aldridge 2012). We turn next to the question of whether a comparable preference is found in relative clauses.

3.   Experiment 2—relative clauses

Experiment 2 examines preferences in the use of relative clauses with the help a picture-based elicited production.

3.1. Method

3.1.1. Participants

Fourteen children aged 4;1 to 5;5 (mean 4;10) participated in our second experiment, along with seven adult controls.

3.1.2. Materials

We made use of 13 pairs of black-and-white pictures, which consisted of 3 practice items and 10 target items. The target items involved reversible events with an animate agent and an animate theme, both of which were made definite with the help of a context. There were two conditions—one that elicited agent focus relative clauses and one that elicited theme focus relative clauses.

3.1.3. Procedure
Each test item consisted of a two-picture panel depicting the same action using different characters; an accompanying pair of sentences provided a background context.
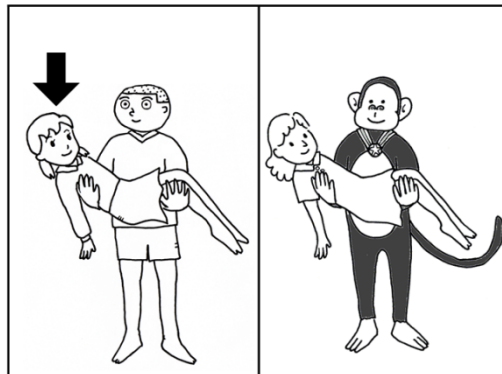
Figure 4. Sample agent focus relative clause item for the target response, "the boy who is carrying the girl."

In the case of Figure 4, for instance, participants were told (in Tagalog): "A boy is carrying the monkey. Another boy is carrying the girl." Because the characters in the picture were made familiar to the participants by the lead-in sentence, they can be considered definite, as was also the case with the characters in the test items used in our first experiment. They were then asked

"Who has the arrow mark?" The targeted response in this case is an agent relative clause, as shown in (7).

(7)   lalaki=[$_{RC}$ ŋ   b<um>u~buhat   naŋ babae   __]
        boy=LNK   <AF>IPFV~carry -FOC girl
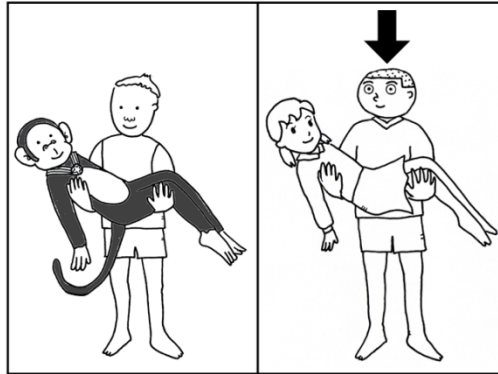   'the boy who is carrying the girl'

Figure 5. Sample theme focus relative clause item for the target response, "the girl who the boy is carrying."

Figure 5 exemplifies a test item designed to elicit a theme relative clause, as illustrated in (8).

(8)   babai=[$_{RC}$ ŋ   b<in>u~buhat   naŋ   lalake __]
        girl=LNK   <TF>IPFV~carry   -FOC   boy
        'the girl who the boy is carrying.'

3.2. Results

We analyzed 65 responses from adults (32 from the agent focus relative clause condition; 33 from the theme focus relative clause condition) and 48 responses from children (25 from the agent focus relative clause condition; 23 from the theme focus relative clause condition). Figure 6 summarizes our findings.

     As can be seen in Figure 6, adult participants performed at ceiling in both conditions; however, results from children showed a significantly higher success rate on agent focus relative clauses over theme focus relative clauses ($\beta = 1.54 \pm 0.62$, $p < .05$).[5] This suggests that children find agent focus relative clauses easier to produce than their theme focus counterparts—a result that runs counter to the strong preference for theme focus observed in declarative clauses.

---

[5] I report fixed effect coefficients, β, from mixed effects logistic regression, which includes participants and items as random effects. The estimates of the two random effects were close to zero, indicating that the probability of targeted responses does not depend on individual participants or items.
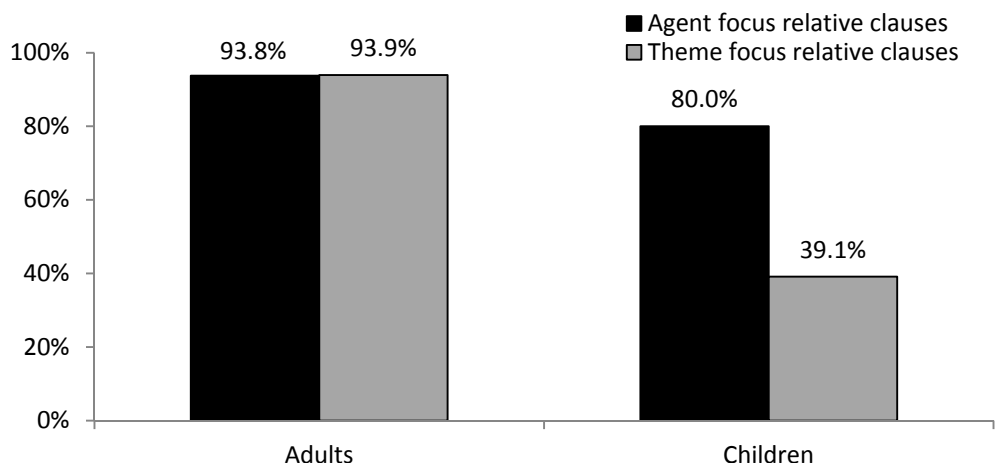
Figure 6. Rate of target responses in both relative clause types.

4. General discussion

In sum, our results reveal a mystery of sorts. On the one hand, we have been able to experimentally confirm a standard generalization about Tagalog: theme focus is preferred to agent focus. Moreover, this seems to be true for children, almost to the same overwhelming degree as it is for adults.

On the other hand, matters are very different in the case of relative clauses, where children favor agent focus patterns over their theme focus counterparts by a large and statistically significant margin. In this regard, the children's performance reflects a much broader cross-linguistic tendency for learners and speakers to favor relativization of a verb's external argument over internal arguments and obliques, as reported for Ch'ol (Clemens et al. to appear), Dutch (Frazier 1987), English (Diessel and Tomasello 2005), French (Holmes and O'Regan 1981), German (Schriefers, Friederici, and Kühn 1995), Greek (Stavrakaki 2001), Hebrew (Friedmann, Belletti, and Rizzi 2009), Hungarian (MacWhinney and Pléh 1988), Japanese (Kawashima 1980), Kaqchikel (Heaton 2015), Korean (Cho 1999), Mandarin Chinese (Hsu, Hermon, and Zukowski 2009), Persian (Rahmany, Marefat and Kidd 2011), Q'anjob'al (Clemens et al. to appear), and Swedish (Håkansson and Hansson 2000).

Our findings thus raise a series of new questions about Tagalog pertaining to the nature of focus, the preferences that are manifested in different constructions, and the manner in which those preferences emerge in the course of language acquisition. These questions are at the forefront of additional research that we are currently conducting on Tagalog.

References

Aldridge, Edith. 2004. Internally headed relative clauses in Austronesian languages. *Language and Linguistics* 5:99–129.

Aldridge, Edith. 2012. Antipassive and ergativity in Tagalog. *Lingua* 122:192–203.

Cho, Sookeun. 1999. The acquisition of relative clauses: Experimental studies on Korean. Doctoral dissertation, University of Hawaiʻi at Mānoa.

Clemens, Lauren Eby, Jessica Coon, Pedro Mateo Pedro, Adam Milton Morgan, Maria Polinsky, Gabrielle Tandet, and Matthew Wagers. To appear. Ergativity and the complexity of extraction: A view from Mayan. *Natural Language and Linguistic Theory*.

Diessel, Holger, and Michael Tomasello. 2005. A new look at the acquisition of relative clauses. *Language* 81:882–906.

Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5:519–559.

Friedmann, Naama, Adriana Belletti, and Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119:67–88.

Håkansson, Gisela, and Kristina Hansson. 2000. Comprehension and production of relative clauses: A comparison between Swedish impaired and unimpaired children. *Journal of Child Language* 27:313–33.

Heaton, Raina. 2015. The status of syntactic ergativity in Kaqchikel. Paper presented at the Society for the Study of the Indigenous Languages of the Americas (SSILA) Annual Meeting 2015, Portland, Oregon.

Holmes, V. M., and J. K. O'Regan. 1981. Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior* 20:417–430.

Hsu, Chun-Chieh Natalie, Gabriella Hermon, and Andrea Zukowski. 2009. Young children's production of head-final relative clauses: Elicited production data from Chinese children. *Journal of East Asian Linguistics* 18:323–60.

Kawashima, Michiko. 1980. The acquisition of Japanese relative clauses. Master's thesis, University of Alberta.

MacWhinney, Brian, and Casaba Pléh. 1988. The processing of restrictive relative clauses in Hungarian. *Cognition* 29:95–141.

Rahmany, Ramin, Hamideh Marefat, and Evan Kidd. 2011. Persian speaking children's acquisition of relative clauses. *European Journal of Developmental Psychology* 8:367–388.

Reid, Lawrence A., and Hsiu-Chuan Liao. 2001. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5:433–490.

Schriefers, Herbert, Angela D. Friederici, and Katja Kühn. 1995. The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language* 34:499–520.

Segalowitz, Norman S., and Rosita G. Galang. 1978. Agent-patient word-order preference in the acquisition of Tagalog. *Journal of child language* 5:47–64.

Stavrakaki, Stavroula. 2001. Comprehension of reversible relative clauses in specifically language impaired and normally developing Greek children. *Brain and Language* 77:419–31.

Tucker, G. Richard. 1971. Focus acquisition by Filipino children. *Philippine Journal of Psychology* 4:21–24.

DIAGNOSTICS FOR *ALREADY* VS. PERFECT ASPECT: A CASE STUDY OF JAVANESE *WIS* *

Jozina Vander Klok
University of British Columbia
jozina.vander@ubc.ca

Lisa Matthewson
University of British Columbia
lisa.matthewson@ubc.ca

Already and the perfect aspect are both acceptable in many of the same environments: both can express the recent past, a result, or an experiential reading. In this paper we develop a set of cross-linguistic diagnostics to distinguish *already* from the perfect aspect, using the auxiliary *wis* in Javanese as a case study. *Wis* has been characterized as *already*, a (present) perfect, a past tense, or a perfective. We argue that Javanese *wis* is a focus operator which presupposes that the focus is a maximal element among a set of ordered alternatives, following Krifka's (2000) analysis of English *already*.

1.   Introduction

Already and the perfect aspect are acceptable in many of the same environments, since both refer to an event prior to the utterance time without relying on a specific past reference time. Our main goal in this paper is to establish a set of cross-linguistically applicable diagnostics to distinguish *already* from the perfect. The auxiliary *wis* in Javanese (Western Malayo-Polynesian) presents an ideal case study, as it has been variously characterized as *already* (Robson 2002; Robson and Wibisono 2002; Ewing 2005; Cole, Hara, and Yap 2008; Conners 2008), a (present) perfect (Horne 1961; Dahl 1985), a perfective (Conners 2008; Hoogervorst 2010; Vander Klok 2012), and a past tense marker (Favre 1866; Robson and Wibisono 2002). The varying characterizations highlight how easy it is to conflate distinctions which, while similar, can be proven to be distinct in their semantics.

This paper is the first targeted investigation of the semantics of *wis* in Javanese. We present new data on this auxiliary from primary fieldwork on a Javanese dialect spoken in Paciran, East Java, Indonesia.[1] We investigate the two potential hypotheses that *wis* expresses perfect aspect or that it expresses *already*. We argue that based on our set of diagnostics, *wis* can only have the

---

[1] Paciran is part of the Pesisir continuum along the northern coast of Java (Hoogervorst 2010). The data from Paciran Javanese are primarily in *Ngoko* 'Low Javanese', the everyday speech of the villagers. In this paper, we adopt the Standard Javanese form in the text (*wis*), but use *wes* in the examples from Paciran Javanese to reflect their preferred spelling.

semantics of *already*. We provide an analysis of it broadly following the semantics that Krifka (2000) proposes for English *already*.[2]

2.   Similarities between *already* and the perfect aspect

In this section, we first focus on similarities between the perfect and *already* to show how these two hypotheses can easily be conflated. First, *already* and the perfect both refer to an event prior to the speech time without a specific past reference time, as suggested by the translations in (1).

(1)   Context: Bu Z. talking to Bu S. about K.'s background.
    Wes   bel-ajar   nek Jogja nem ulan.
    already INTR-study at   Jogja six   month
    'She **has studied** / **already studied** in Jogja for six months.'

In English, perfect aspect can express the recent past, a result, or an experiential reading (e.g., Comrie 1976; Smith 1997). *Already* can express these exact same readings (cf. Mittwoch 1988; Michaelis 1992:324), as illustrated in (2).

(2) a.   Jordan {has (just) / already} left.   (RECENT PAST)
    b.   Andrea {has / already} arrived in London.   (RESULT)
    c.   Bethany {has / already} visited Edinburgh (before). (EXPERIENTIAL)

Because of these similarities, there is a risk of misanalyzing elements which meet this temporal configuration. The possibility for misdiagnosis is reinforced in questionnaires for cross-linguistic semantic use such as that of Dahl (1985). Many of the questions that are classified by Dahl as hallmark examples of perfect aspect could equally target *already*, cf. (2). Consider the examples in (3) from Dahl's questionnaire (Dahl's #64, 42, 53 respectively).

(3) a.   Child: Can I go now?
         Mother: You BRUSH your teeth?   (RECENT PAST)
    b.   A: Is the King still alive?
         B: (No,) he DIE.   (RESULT)
    c.   You MEET my brother (at any time in your life until now)?
                                                   (EXPERIENTIAL)

In each of these examples for English, the uninflected verb (in capital letters) is equally compatible with three different conjugations: the (present) perfect, the simple past, or the simple past with the adverb *already*. This shows how misdiagnosis of a marker is easily possible.

---

[2] For reasons of space, we omit here the proof that *wis* cannot be a past tense marker (Favre 1866; Robson and Wibisono 2002) or a perfective marker (Conners 2008; Hoogervorst 2010; Vander Klok 2012).

Dahl (1985) reports that based on his questionnaire, Javanese *wis* expresses perfect aspect because it occurs in all of his prototypical examples for the perfect. We reran this questionnaire with two speakers of Paciran Javanese, and found similar results to Dahl's. Specifically, the questions which have recent past or resultative readings were translated with the auxiliary *wis*, as shown in (4) (corresponding to Dahl's #64, 42, 39).[3]

(4) a. Child: 'Can I go now?'
     Mother: Opo awakmu **wes** sikat-an?   (RECENT PAST)
               Q   2SG   already brush-AN
               'You BRUSH your teeth?'
  b. A: 'Is the King still alive?'
     B: Ora, rojo **wes** mangkat   (RESULT)
               NEG king already pass.away
               '(No,) he DIE.'
  c. Context: 'Do you know my brother?'   (EXPERIENTIAL)
               Yo, aku **wes** ke-temu dulur-mu sepisan
               Yes, 1SG already KE-meet sibling-your once
               pirang-pirang taun kepungkor
               RED-some   year ago
               '(Yes,) I MEET him (once) several years ago.'

Without further evidence, the results of this questionnaire for Javanese would lead us to conclude that *wis* is a perfect, as *wis* is acceptable in recent past, resultative, and experiential contexts. This is what Dahl (1985) concludes. However, as shown above in (2) for English, *already* is equally compatible in each of these environments. This fact renders Dahl's questionnaire results inconclusive. What is lacking, then, are diagnostics that differentiate the perfect from *already*.

3.   Diagnostics to distinguish *already* from the perfect

We identify the following diagnostics to distinguish *already* from the perfect:[4]
(i) morphological transparency and truth-conditional equivalency in interactions

---

[3] Dahl (1985) argues that a separate lexical item, *tau*, expresses experiential perfect aspect (4c). Our fieldwork on Paciran Javanese confirms that *tau* is used for experiential readings, but *wis* also allows experiential readings as shown in (4c).

[4] We do not discuss the incompatibility of *already* with downward-entailing quantifiers (e.g., *only, less than*) as a possible diagnostic. (See Soh and Gao 2008 on English *already*, Soh and Gao 2008; Soh 2008, 2009 on sentence-final Mandarin –*le*, Soh 2012 on post-verbal Colloquial Malay *dah*, Olsson 2013 on other South East Asian languages.) We leave this potential diagnostic for future research due to a number of factors that ameliorate the acceptability of English *already* with downward-entailing quantifiers. These include (i) the scope of *already*, (ii) the use of prosody in indicating corrective or contrastive focus, and (iii) the use of different types of predicates.

with negation (duality of *already*) (König 1977, 1991; Löbner 1989, 1999; Krifka 2000); (ii) an 'earliness' implication (Löbner 1989; Mittwoch 1993; Michaelis 1992, 1996; Krifka 2000); (iii) inchoative interpretation with stative predicates; (iv) (in)compatibility with adverbs specifying a past time interval (Giorgi and Pianesi 1997; Portner 2003); and (v) 'Extended Now' interpretation (McCoard 1978).

A marker which expresses *already* could involve (i) duality, (ii) an 'earliness' implication, and (iii) an inchoative interpretation with statives; but disallow (v) an 'Extended Now' interpretation. Evidence that a marker expresses the perfect could involve (iv) incompatibility with past-time adverbs and (v) an 'Extended Now' interpretation. Further, we expect that a perfect aspect will not have (i) duality or (ii) an 'earliness' implication. We apply these diagnostics to Javanese *wis,* arguing that *wis* can only express *already*.

3.1.   Duality of *already*

A first diagnostic that differentiates *already* from the perfect aspect is the interaction with negation. *Already* can be grouped with *still, not yet,* and *no longer/not anymore* based on (i) possible morphological transparency with these related items; and (ii) equivalent truth-conditions in interaction with negation (König 1991; Löbner 1989, 1999; Krifka 2000; among others). Importantly, these properties are not upheld with the perfect.

Morphological transparency is illustrated with Hebrew in (5). *Kvar* 'already' and *'adayin* 'still' are related by duality, in that the inner negation of one is equivalent to the outer negation of the other (cf. Löbner 1989, 1999). The external negation of the *kvar* 'already' sentence in (5a) is lexically expressed by the inner negation of *'adayin* 'still', as in (5b). Similarly, the external negation of the *'adayin* 'still' sentence in (5c) is expressed by the internal negation of *kvar* 'already' in (5d). In other words, *already* means '*not still not'* and *still* means '*not already not'*.

(5) a. **kvar**   yored geshem.   b. **'adayin lo**   yored geshem.
         already rain   is   still   not rain   is
         'It is already raining.'   'It is not yet raining.'

    c. **'adayin** yored geshem.   d. **kvar**   lo yored geshem.
         still   rain   is   already not rain   is
         'It is still raining.'   'It is not raining anymore.'

                         (Krifka 2000:401, Hebrew, *glosses and translation added*)

This system has been noticed in other languages as well, such as Dutch, French, English (Löbner 1989:170), Spanish, and Czech (Krifka 2000). Not all languages have morphologically transparent systems: English *already* and *still* have suppletive forms that are negative polarity items, *yet* and *anymore/longer* respectively (Krifka 2000:401).

Turning to Javanese, we can examine whether forms related to *wis* are morphologically transparent with markers that express *still, not yet, no longer*. First, we observe that the outer negation of *isek* 'still' is expressed by the inner negation of *wis*.[5] This morphological transparency is parallel to the forms in Hebrew in (5c-d), and can only be understood under the hypothesis that these forms are duals.

(6) Mas Mawon **wes** **gak** ndandan-i   jareng-e.
     Mr. Mawon already NEG AV.fix-APPL   fishing.net-DEF
     'Mawon is no longer fixing the fishing net.'

For the external negation of Javanese *wis*, speakers invariably offer sentences with a suppletive form *durung* 'not yet' (Robson and Wibisono 2002:203). While this relation is not morphologically transparent, speakers comment that *durung* 'not yet' is like *ora wis* 'NEG already' (the external negation of *wis*, but which is non-existent).

(7) Mas Mawon **durung** ndandan-i   jareng-e.
     Mr. Mawon not.yet AV.fix-APPL   fishing.net-DEF
     'Mawon didn't fix the fishing net yet.'

With respect to equivalent truth-conditions based on interaction with negation, Löbner (1989, 1999) presents evidence that German *schon* 'already' and *noch* 'still' as well as English *already* and *still* are duals based on their truth-conditional interaction with respect to negation. This relation parallels the interaction with negation seen with duality in quantifiers (e.g. $\forall$, $\exists$). For instance, the internal negation of the universal quantifier $\forall$ is $\forall x[\neg P(x)]$, which is truth-conditionally equivalent to the external negation of the existential quantifier $\exists$: $\neg\exists x[P(x)]$. Thus, *Every light is off* is equivalent to *It is not the case that some light is on*. Conversely, *Not every light is off* is equivalent to *Some light is on*.

This equivalency is observed in English with *already*. In (8a,b), the external negation of *already* ($\neg already\ [p]$) is truth-conditionally equivalent to the internal negation of *still* (*still* $[\neg p]$). Similarly, in (8c,d), the external negation of *still* ($\neg still\ [p]$) is truth-conditionally equivalent to the internal negation of *already* (*already* $[\neg p]$).[6]

(8) a. $\neg already\ [p]$ = b. *still* $[\neg p]$
        It is **not yet** [raining].   It is **still** [not raining].

---

[5] Inner negation of *wis* with *ora* 'NEG' is also possible. *Ora* and *gak* are interchangeable in most environments in Paciran Javanese.

[6] van der Auwera (1993:616) gives additional equivalency tests such as 'Peter is already in Madrid' = It is no longer the case that Peter is not in Madrid' / 'It is not the case that Peter is not yet in Madrid.' These examples proved too difficult to properly elicit in Javanese.

c. ¬*still* [*p*]   =   d. *already* [¬*p*]
   It is **not** [raining] **anymore**.   It is **already** [not raining].

Turning to Javanese, if *wis* expresses *already*, then we expect to find truth-conditional equivalencies with negation with an item that expresses *still*. We expect that the external negation of *wis* (¬*wis* [*p*]) is truth-conditionally equivalent to the internal negation of *isek* 'still' (*isek* [¬*p*]). This prediction is borne out, as shown by the equivalencies in (9).

(9) a.   ¬*wis* [*p*]   =   b. *isek* [¬*p*]
         durung [udan].   **isek** [gak   udan].
         not.yet   rain   still   NEG   rain
         'It is not yet raining.'   'It is still not raining.'

And vice versa, we predict that the external negation of *isek* 'still' (¬*isek* [*p*]) is truth-conditionally equivalent to the internal negation of *wis* 'already' (*wis* [¬*p*]). However, we cannot test this in Javanese since external negation of *isek* 'still' is not grammatically possible (*gak isek udan 'it is no longer raining').

In this section, we have discussed two properties which pertain to the duality diagnostic distinguishing *already* from the perfect aspect: (i) possible morphological transparency with these related items; and (ii) equivalent truth-conditions based on interaction with negation. We showed that Javanese has both these properties, providing evidence that *wis* expresses *already*.

3.2.   'Earliness' of *already*

Another property that distinguishes *already* from perfect aspect relates to an implicature of 'earliness' (Mittwoch 1993; Löbner 1989, 1999; Krifka 2000). This earliness factor is illustrated in (10) with German *schon* 'already'.

(10)a. Es ist **schon** zwei – nicht erst eins.
        'It is already two – not (still) one.'
    b. Sie kommt **schon** um zwei – nicht erst um drei.
        'She is already coming at two – not at three.'   (Löbner 1989:193)

Löbner (1989:194) points out that "the event or state under consideration occurs earlier than in the contrasting case" in both examples above: in (10a) time has passed by earlier than the speaker expects and in (10b) the arrival occurs earlier than expected. Earliness equally holds for English *already*. In (11), *already* "serves to assert that the state of inebriation has come about at a point prior to the time at which it might be expected to eventuate" (Michaelis 1992:326).

(11)   When we arrived, before noon, Huey was **already** drunk.
                            (Michaelis 1992:326, citing *Ken Kelley, 'Huey Newton',*
                            California Magazine 8/90)

In contrast, in (12), the perfect aspect asserts that Huey's drunkenness occurred before the reference time of before noon, but does not express that the inebriation is earlier than expected.

(12)   When we arrived, before noon, Huey **had been** drunk.

We now turn to Javanese *wis*. *Wis* expresses that the state obtained occurs earlier than expected, behaving like *already* and not like a perfect aspect. An example from a dialogue is given in (13), where Bu G. reports that she said the following to her grandmother:

(13)   mbok   **wes**   jam setengah wolu   ndak-an engko
        grandmother already hour half   eight   to-AN   later
        kari   reng pasar.
        left.behind at   market
        'Grandmother, it's already 7:30a.m. so there won't be anything at the market soon.'

In Indonesia, the morning market starts approximately at dawn (5:30a.m.) and ends at 8a.m. at the latest. This example is similar to that with *schon* in (10a): *wis* expresses that time has passed by earlier than the speaker expects, and she now has to hurry before the market closes. This meaning would not be expressed by the perfect '*It has been 7:30a.m.*'.
        Consider also (14), which expresses that the baby can walk earlier than expected.

(14)   bayi-ne   **wes**   iso   melaku
        baby-DEF already CIRC.POS walk
        'The baby can already walk.'

These examples with *wis* exhibit an earliness factor in comparison to some contextually relevant point. If *wis* were to express the perfect aspect, the earliness component would not be captured.

3.3.   Inchoative interpretation with stative predicates

Our third diagnostic for distinguishing the perfect aspect from *already* involves the interpretation when a marker combines with a stative predicate.[7] Unlike a

---

[7] This diagnostic does not necessarily distinguish *already* from perfective aspect. Smith (1997) argues that in some languages, the perfective also induces inchoativity with states (e.g.,

sentence containing the present perfect, a stative sentence containing *already* conveys a change *into* the state denoted by the predicate. This is illustrated in (15) and (16) for stage-level (temporary) and individual-level (permanent) states respectively. In each case, the (b) examples with *already* seem to commit the speaker to the claim that the subject did not previously possess the relevant attribute.

(15) a. Paula has been tired / disappointed / pregnant.
     b. Paula is already tired / disappointed / pregnant.   STAGE-LEVEL STATES

(16) a. The child has been tall / intelligent / fat.
     b. The child is already tall / intelligent / fat.   INDIVIDUAL-LEVEL STATES

The inchoative effect is particularly striking with the individual-level states, since these inherently do not typically convey an initial change into the relevant state.

     As pointed out by Mittwoch (1993), the inchoative effect of *already* is cancelable in certain environments, such as in (17)-(18). Nevertheless, *already* does seem to at least imply a change-of-state meaning.

(17)   Peter's eyes were already brown when he was born. (Mittwoch 1993:76)

(18)   A: I've applied for American citizenship.
        B: Is your husband also applying?
        A: He is already American, for he was born in America.
       (Mittwoch 1993:74)

Applying this diagnostic to Javanese *wis,* we see a strong inchoative effect, which is especially detectable with individual-level states. The data are given in (19)-(21).

(19)   Context: I haven't seen Kana in one year. When I left before, she was still short.
       Kana (sa'iki) kok **wes**   gedhe / dhuwur.
       Kana now   PRT already big   tall
       'Kana is already big now!'

Mandarin verb-final *-le*). This diagnostic is only valid in addition to (i) duality and (ii) the earliness implicature of *already* which the perfective does not share.

(20) Pak Bambang **wes**   ngerti   cara-ne   ndandan-i   montor
      Mr. Bambang already AV.know way-DEF AV.repair-APPL car
      'Pak Bambang already knows how to repair cars.'
      Consultant's comments: *Bolahe kursus*. 'Because [he took] a course.'
      Sa'durunge durung ngerti. 'Before, he did not understand yet.'

(21) Siti, mata-ne **wes**   biru
      Siti eye-DEF already blue
      'Siti, her eyes have become blue.'
      Contexts offered by consultant: (i) *Operasi, matane dadi biru*. 'She had
      an operation, her eyes became blue.'; (ii) *Dike'i kontak lens* 'She was
      given contact lenses.'

Wis is actually infelicitous with an individual-level state in a non-inchoative
discourse situation, as in (22), in comparison to the inchoative discourse
situation in (19) above.

(22)   Context: Ever since birth, Miss Ulum has been big.
      # Mbak Ulum **wes**   gedhe.
        Miss   Ulum already   big

This infelicity is further illustrated with states that do not allow a $\neg p$ state
before the reference time, as in (23). These states are felicitous with the perfect,
as it has no such requirements.

(23) # She is already young.   (Löbner 1989:181)

This prediction is also borne out with Javanese *wis*, showing that *wis* is
restricted to events/states that satisfy that $\neg p$ is true at a time before the
reference time.

(24) # dik   Tomo umur-e lima-ng taun.   Tomo **wes**   **enom**.
        yg.brother Tomo age-DEF five-LNK year   Tomo already   young

Finally, just like in English, the change-of-state effect is cancelable in Javanese
in certain discourse contexts, as shown in (25).

(25) srikoyo   **wes**   legi.   Gak perlu namba   gulo.
      sugar.apple already   sweet. NEG need AV.add   sugar
      'Sugar apples are (already) sweet. [You] don't need to add sugar.'

(25) seems to contrast with (22), where a non-inchoative discourse context led
to *wis* being judged as infelicitous. Interestingly, the cases where the non-
inchoative readings are licensed all seem to involve an expectation that the

hearer assumes that the state does *not* hold. Thus, (17) contrasts the facts with a potential situation where Peter's eyes have turned brown since his birth; (18) involves a correction of B's assumption that the husband is not yet American. Similarly, in Javanese, (25) corrects the hearer in their apparent belief that sugar apples are not sweet.[8]

3.4.   Compatibility with past time adverbials

Compatibility with past time adverbials is a fourth diagnostic for empirically distinguishing *already* from the present perfect aspect. In languages such as English, the present perfect is unacceptable with adverbs expressing a definite past time, as illustrated in (26) with *yesterday*.

(26) * I have written yesterday.   (McCoard 1978:123, citing Pickbourn 1789)

The present perfect contrasts with the past perfect or tenseless perfect forms in (27a, b) respectively, in which past time adverbials are acceptable (McCawley 1971). Past time adverbials are also compatible with *already* in English (*I already wrote yesterday.*).[9]

(27) a. Mary had arrived the day before.
     b. Having arrived yesterday, Mary can answer our questions.
                                                    (Portner 2003:465)

For languages like English or Mainland Scandinavian, specific past-time adverbials distinguish the present perfect from *already* (Giorgi and Pianesi 1997). However, Giorgi and Pianesi (1997) observe that in other languages like Italian, German, Dutch, and Icelandic, such adverbials *are* acceptable with the present perfect. This is illustrated in (28) for Dutch.[10]

(28)   Jan is om vier uur weggegaan.
         'John left at four.' (Lit. 'John has left at four.')
         (Giorgi and Pianesi 1997:87)

Giorgi and Pianesi tie the absence of past-time adverbial effects in these languages to the different semantics of the present tense from English-type

---

[8] Some languages do have perfects which induce inchoativity effects; see e.g., Koontz-Garboden (2007) on Tongan, Matthewson et al. (2012) on Niuean. However, these effects differ from the inchoative feature of *already:* inchoative perfects also allow inchoative readings with activity predicates, while *already* does not (and Javanese *wis* does not).

[9] See Portner (2003) for a pragmatic approach to this distinction based on presuppositions of the present tense in the present perfect; see Giorgi and Pianesi (1997) for a syntactic approach.

[10] The adverbial *om vier* 'at four' can refer to past, present, or future time. A better example would be one in which pastness is entailed by the adverbial itself (e.g. *yesterday*) (cf. Portner 2003:465).

languages. While the present tense is syntactically realized in Italian, German, Dutch, and Icelandic, semantically it is vacuous, so that the present is simply the absence of the past. It is therefore important to be able to recognize whether the present tense is syntactically and semantically realized in the language under study for this diagnostic. If it is and the marker under question is incompatible with past time adverbials with the present tense, this suggests that the marker is a perfect. If the marker is compatible with past time adverbials, this is evidence that it expresses *already*. In a language where the present tense is not syntactically or semantically realized, this diagnostic will not distinguish *already* from the perfect: both could be compatible with past time adverbials.

In Javanese, *wis* is compatible with adverbs specifying a past time interval, as in (29) from a recorded conversation:

(29) Context: *Two women, Bu Z. and Bu S., talking together:*
gek ngi aku **wes** ngomong [...] sik pak Arif iku loh....
just yesterday 1SG already AV.speak Mr. Mr. Arif DEM PRT
'Yesterday, I have already spoken to the Mr. Arif.' (translation offered by consultant)

However, we cannot conclude that *wis* expresses *already* because in Javanese, there is no syntactically overt present tense form (e.g., Horne 1961; Robson 2002). There is also no semantic realization of present tense (that is absent in the overt syntax): any event or state can be interpreted in the present, past or future, depending on the context. Furthermore, predicates can be modified by adverbs expressing future, present, or past reference time, as in (30), suggesting that the predicate does not have any null tense specifications.

(30) sego pecel-e bu Maula **wingi / sa'iki / sesok** di-murah-no.
rice *pecel*-DEF Mrs. Maula yesterday / now / tomorrow PASS-cheap-APPL
'Bu Maula's *pecel* rice was made cheaper yesterday.'
'Bu Maula's *pecel* rice is made cheaper now.'
'Bu Maula's *pecel* rice will be made cheaper tomorrow.'

Javanese differs in these respects from St'át'imcets (Lillooet Salish), for instance, in which bare predicates cannot be interpreted as future and cannot co-occur with adverbs specifying future reference time (Matthewson 2006:677). Matthewson argues on the basis of these data that there is a null tense morpheme indicating non-past in St'át'imcets. Javanese, on the other hand, has no null tense specifications. There is therefore no semantic realization of present tense in Javanese. Consequently, the fact that there is no restriction with past time adverbials is inconclusive for distinguishing whether *wis* is best analyzed as a perfect or as *already*. Cross-linguistically, however, if a language has syntactically and semantically realized present tense, this diagnostic can serve to

distinguish *already* from the perfect: infelicity with an adverb specifying a past time reference would suggest the marker under study expresses perfect aspect.

3.5.   Extended Now of the perfect

Our fifth diagnostic to distinguish perfect aspect from *already* concerns 'Extended Now' or 'lifetime' effects, which relate specifically to the present perfect. In (31a), the present perfect is infelicitous (despite the fact that the printing press is still currently relevant). The (b) example with *already,* however, is felicitous, and does not incur a lifetime effect. The lifetime effects also disappear with the past perfect (*had discovered).*

(31) a. ?? Gutenberg has discovered the art of printing.

                                  (McCoard 1978, citing Dietrich 1955)
     b. Gutenberg already discovered the art of printing (in the 15th century).

Portner (2003) argues that these effects fall out from the Extended Now presupposition of the perfect, which derives from the temporal semantics of the present tense. More specifically, the Gutenberg example is odd because the "Extended Now [requirement of the present tense] would not include the past event of Gutenberg's discovery in any context" (Portner 2003:506). Turning to Javanese, Javanese counterparts to these types of sentences are felicitous, as shown in (32).

(32)   Columbus **wes**   nemok-no   Amerika (taun   1492).
        Columbus already   AV.find-APPL America year   1492
        'Columbus already discovered America (in 1492).'

Similarly, there is no lifetime effect in that the subject must be still living in order for these sentences to be acceptable (cf. (31)); (33) is judged as perfectly acceptable.

(33)   Context: Kartini (1879-1904)
        Kartini **wes**   nules   surat bongso kondisi   wong wedhok nok
        Kartini already AV.write letter about   condition person woman   in
        Jawa.
        Java.
        'Kartini has written letters about women's conditions in Java.' (offered)

However, the fact that lifetime effects only arise in the present perfect, along with the fact that Javanese is a tenseless language, raise the question whether the above cases could be interpreted with past reference time, as a past perfect. This would suggest that the data are compatible with *wis* expressing the perfect aspect.

Some cross-linguistic evidence against this alternative proposal comes from other languages which also do not have obligatory overt marking for past reference time, such as Niuean (Polynesian) and St'át'imcets. If the acceptability of (32)-(33) derives from their proposed status as past perfects, we would predict that corresponding examples in Niuean and St'át'imcets containing a perfect aspect, but no marking for pastness, would be acceptable. This is however not the case; Matthewson (2013) shows that both Niuean and St'át'imcets perfects induce lifetime effects. The Niuean and St'át'imcets facts suggest that lifetime effects are not dependent on a 'real' present tense, but rather the Extended Now presupposition of the perfect. If Javanese *wis* were a perfect marker, despite the fact that Javanese is tenseless, we would still expect lifetime effects to arise under the Extended Now presupposition of the perfect. Since lifetime effects are non-existent in Javanese with *wis*, this leads to the conclusion that *wis* is best interpreted as *already*.

3.6.   Summary of diagnostics: Javanese *wis* as 'already'

We have shown that Javanese *wis* has the property of duality with *isek/ijek* 'still', an implication of earliness, and an inchoativity effect, but lacks lifetime effects. These results strongly suggest that *wis* expresses *already*. The remaining diagnostic, compatibility with a past temporal adverb, was inconclusive due to independent features of the temporal system of Javanese; namely, that Javanese does not semantically or syntactically realize present tense. In the next section we sketch a semantic analysis of Javanese *wis*.

5.   Analysis

In this section, we sketch an analysis of *wis*, basing our discussion broadly on Krifka (2000). Krifka's core idea is that *already* is a focus-sensitive operator, which places a restriction on the alternatives to the focus. His denotation for *already* is given in (34):

(34)   ALREADY($< B, F, \leq_A >$) $\Leftrightarrow < B, F, \leq_A >$, presupposition: $\forall$ X$\in$ A[X $\leq_A$ F]

(Krifka 2000:4)

Already applies to a proposition which consists of a Background (B) and a Focus (F), and which relies on an ordering A (which may be temporal, numerical, etc.). *Already* does not change the truth conditions of the proposition (it outputs the same $< B, F, \leq_A >$); its only function is to introduce a presupposition that the asserted Focus is the highest-ranked salient alternative on the A-scale. We illustrate how this works for (35), where *three* is in focus, following Krifka (2000).

(35)   Lydia is already three_F months old.
       alternatives considered: {Lydia is 1 month old, Lydia is 2 months old,
                                  Lydia is 3 months old}
       alternative asserted: {Lydia is 3 months old} (Krifka 2000:5)

(35) asserts that Lydia is three months old, and presupposes that three is the highest-ranked salient alternative number of months for Lydia's age. In other words, Lydia is three months old, rather than the possible alternatives that could have been asserted, namely one or two months. The fact that Lydia's age is the greatest of those that are 'considered entertainable' leads to the implicature that her age is greater than one might have expected (because it is greater than the average of her reasonably possible ages) (Krifka 2000:5). If (35) did not contain *already,* the set of alternatives would include also higher ages (Lydia is 4 months old, etc.). The speaker would therefore assert that Lydia is 3 months old rather than any of the other salient alternative possible ages of 1, 2, 4 or 5 months.

       This also captures the contribution of *wis* in Javanese. An example is given in (36).

(36)   Context: *Bu G. reports that she said the following to her grandmother:*
       mbok   **wes**   **jam setengah wolu** ndak-an engko
       grandmother   already   hour half   eight   to-AN   later
       kari   reng pasar.
       left.behind at   market
       'Grandmother, it's already 7:30a.m. so there won't be anything at the market soon.'

       alternatives considered: 5:30a.m., 6:30a.m., 7:30a.m.
       alternative asserted: 7:30a.m.

PRESUPPOSITION: 7:30a.m. is the highest-ranked salient alternative among the market times.
ASSERTION: It is 7:30a.m.
IMPLICATURE: 7:30a.m. is later than the average of the reasonable market times

Another example from Javanese overtly shows that the time has passed earlier than expected. In the recorded conversation in (37), Bu Z. exclaims that time has passed faster than she expected.

(37)   Context: Ngelewat (Visiting a family to pay respects to the deceased.)
        Bu Z: 'How long has it been [since she passed away]?'
        Bu G: 'It's (already) seven days later, Mrs. Zum!'
        Bu Z: ya Allah... **wes**   pito-ng   ndino.... yo   **kok cepet loh**.
                 ya Allah   already   seven.LNK N.day   yes   PRT   fast   PRT
                 'Ya Allah, it's already been seven days. Wow, that's so fast!'

With respect to the inchoative effects of *already,* Krifka observes (2000:7-8) that his analysis has no problem with the 'already American' cases discussed by Mittwoch (1993) (see (17) and (18) above). This is because Krifka's analysis does not hardwire an inchoative semantics, unlike the analysis of Löbner (1989, 1999), which models the change-of-state semantics of German *schon* 'already' as a presupposition: *schon* Φ presupposes a time before the reference time for which ¬ Φ is true. However, it seems that incorporating no inchoativity effect at all would also be a mistake, since the 'already American' cases do require a very specific type of discourse context to be felicitous.

It seems to us that under a broadly Krifka-type analysis, the inchoative effect of *already* can be viewed as a conversational implicature which arises due to the following reasoning: if the speaker conveys that the predicate holds at an earlier time than was expected, then the speaker does not believe the predicate to be timelessly true. On the contrary, the speaker is aware of a previous time interval during which the predicate did *not* hold. From this, the hearer concludes that there was an immediately prior time interval at which the plain proposition was false. This captures the fact that the 'already American' cases require a specific type of discourse context to be felicitous where the speaker addresses the fact that the hearer believes the plain proposition to be false. This is true in both English (see (18)) and Javanese (see (25)).

6.   Conclusion

This paper puts forward five diagnostics that distinguish the perfect aspect from *already*: (i) duality of *already*, (ii) earliness implicature, (iii) inchoativity effects with statives, (iv) compatibility with past temporal adverbials, and (v) Extended Now/lifetime effects. We applied these diagnostics to Javanese *wis*, and established that *wis* is best analyzed as expressing *already*. We adopt the semantic analysis of English *already* proposed by Krifka (2000) for Javanese *wis,* whereby *wis* applies to a proposition containing a Background and a Focus and asserts that same proposition and presupposes that the asserted proposition has a faster development speed than other salient alternatives. This in turn leads to implicatures of earliness and of inchoativity.

Our findings have implications for the cross-linguistic study of aspect in under-studied languages. This study has the potential to be particularly useful for Austronesian languages. In Colloquial Malay, post-verbal/sentence-final *dah* is argued to have the semantics of *already* (Soh 2012) instead of completive or

perfective aspect (Koh 1990:202). Similarly, *sudah* in Standard Formal Malay is argued to express *already* (Kader 1981:36) or completive or perfective aspect (Soh 1994 and references therein) and Indonesian *sudah* is noted to express both properties of *already* and the perfect aspect (Sneddon et al. 2010; Grangé 2010; Kaswanti Purwo 1984, 2011; Olsson 2013). Our set of diagnostics can be used to help distinguish the closely similar markers *already* and the perfect, which could otherwise be left unnoticed or misanalyzed.

References

van der Auwera, Johan. 1993. 'Already' and 'still': Beyond duality. *Linguistics and Philosophy* 16:613-653.

Cole, Peter, Yurie Hara, and Ngee Thai Yap. 2008. Auxiliary fronting in Peranakan Javanese. *Linguistics* 44:1-43.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Conners, Thomas J. 2008. Tengger Javanese. Doctoral dissertation, Yale University.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford; New York: Blackwell.

Dietrich, Gerhard. 1955. *Erweiterte Form, Präteritum und Perfektum im Englischen. Eine Aspekt- und Tempusstudie*. Munich: Max Hueber Verlag.

Ewing, Michael C. 2005. *Grammar and inference in conversation: Identifying clause structure in spoken Javanese*. Philadelphia: John Benjamins.

Favre, L'Abbé P. 1866. *Grammaire javanaise*. Paris: L'imprimerie Impériale.

Giorgi, Alessandra, and Fabio Pianesi. 1997. *Tense and aspect: From semantics to morphosyntax*. New York: Oxford University Press.

Grangé, Philippe. 2010. Aspect and modality in Indonesian: The case of *sudah*, *telah*, *pernah*, and *sempat*. *Wacana* 12:243–268.

Hoogervorst, Tom Gunnar. 2010. Describing Surabaya's linguistic ecology. Master's thesis, Leiden University.

Horne, Elinor C. 1961. *Beginning Javanese*. Vol. 3: Yale Linguistic Series. New Haven, London: Yale University Press.

Kader, Mashudi. 1981. *The syntax of Malay interrogatives*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Kaswanti Purwo, Bambang. 1984. *Deiksis dalam Bahasa Indonesia* [Deixis in Indonesian]. Jakarta: Balai Pustaka.

Kaswanti Purwo, Bambang. 2011. *Sudah* in contemporary Indonesian. Paper presented at 15th International Symposium on Malay/Indonesian Linguistics. Malang, Indonesia.

Koh, Ann Sweesun. 1990. Topics in colloquial Malay. Doctoral dissertation, University of Melbourne.

König, Ekkehard. 1977. Temporal and non-temporal uses of 'noch' and 'schon' in German. *Linguistics and Philosophy* 1:173-198.

König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective*. London, New York: Routledge.

Koontz-Garboden, Andrew. 2007. Aspectual coercion and the typology of change of state predicates. *Journal of Linguistics* 43(1):115-152.

Krifka, Manfred. 2000. Alternatives for aspectual particles: Semantics of *still* and *already*. *Berkeley Linguistics Society* 26, 401-412. Berkeley Linguistics Society, Berkeley, Ca.

Löbner, Sebastian. 1989. German *schon-erst-noch*: An integrated analysis. *Linguistics and Philosophy* 12:167-212.

Löbner, Sebastian. 1999. Why German *schon* and *noch* are still duals: A reply to van der Auwera. *Linguistics and Philosophy* 22:45-107.

Matthewson, Lisa. 2006. Temporal semantics in a supposedly tenseless language. *Linguistics and Philosophy* 9:673-713.

Matthewson, Lisa. 2013. The building blocks of grammatical aspect: Evidence from Austronesian and Salish perfects. Paper presented at the Workshop on Semantic Variation. University of Chicago.

Matthewson, Lisa, Heidi Quinn, and Lynsey Talagi. 2012. Notes on the Niuean perfect. In *Theories of Everything: In Honor of Ed Keenan,* ed. Thomas Graf, Denis Paperno, Anna Szabolcsi, and Jos Tellings. UCLA Working Papers in Linguistics, Vol. 17, 222-233.

McCawley, James. 1971. Tense and time reference in English. *Studies in Linguistic Semantics*, ed. Charles J. Fillmore and D. Terence Langendoen, 96–113. New York: Holt Rinehart and Winston.

McCoard, Robert. 1978. *The English perfect: Tense-choice and pragmatic inference*. North-Holland: Amsterdam.

Michaelis, Laura A. 1992. Aspect and the semantics-pragmatics interface: The case of *already*. *Lingua* 87:321-339.

Michaelis, Laura A. 1996. On the use and meaning of *already*. *Linguistics and Philosophy* 19:477-502.

Mittwoch, Anita. 1993. The relationship between *schon/already* and *noch/still*: A reply to Löbner. *Natural Language Semantics* 2:71-82.

Olsson, Bruno. 2013. Iamitives: Perfects in Southeast Asia and beyond. Master's thesis, Stockholm University.

Pickbourn, James. 1789. *A dissertation on the English verb: Principally intended to ascertain the meaning of its tenses*. Menston, England: Scholar Press.

Portner, Paul. 2003. The (temporal) semantics and (modal) pragmatics of the perfect. *Linguistics and Philosophy* 26:459-510.

Robson, Stuart. 2002. *Javanese grammar for students, 2nd ed*. Glen Waverley: Monash Papers on Southeast Asia.

Robson, Stuart, and Singgih Wibisono. 2002. *Javanese-English Dictionary*. North Claredon: Tuttle Publishing.

Smith, Carlota. 1997. *The parameter of aspect*. Dordrecht: Kluwer Academic Publishers.

Sneddon, James N., K. Alexander Adelaar, Dwi Noverini Djenar, and Michael C. Ewing. 2010. *Indonesian: A comprehensive grammar*: Routledge comprehensive grammars. London; New York: Routledge.

Soh, Hooi Ling. 1994. Aspect and the organization of argument structure and phrase structure: Evidence from Malay. Master's thesis, University of Calgary.

Soh, Hooi Ling. 2008. The syntax and semantics of change/transition: Evidence from Mandarin Chinese. *Theoretical and cross-linguistic approaches to the semantics of aspect*, ed. Susan Rothstein, 387-419. Philadelphia: John Benjamins.

Soh, Hooi Ling. 2009. Speaker presupposition and Mandarin sentence-final *–le*: A unified analysis of the "change of state" and "contrary to expectation" reading. *Natural Language and Linguistic Theory* 27:623-657.

Soh, Hooi Ling. 2012. The aspectual marker 'dah' in colloquial Malay. Paper presented at 19th Austronesian Formal Linguistics Association. Taipei, Taiwan.

Soh, Hooi Ling, and Meijia Gao. 2008. Mandarin sentential *–le*, perfect and English *already*. *Event Structure in Linguistic Form and Interpretation*, eds. Johannes Dölling, Tatjana Heyde-Zybatow, and Martin Schäfer. Berlin: Mouton De Gruyter.

Vander Klok, Jozina. 2012. Tense, aspect, and modality in Paciran Javanese. Doctoral dissertation, McGill University.