CHAPTER 24

....

LEXICAL CATEGORY AND ALIGNMENT IN AUSTRONESIAN

....

DANIEL KAUFMAN

24.1 THE ALIGNMENT OF PHILIPPINE-TYPE LANGUAGES

....

PHILIPPINE-TYPE languages are often cited as exemplifying a cross-linguistically unique voice system, in which verb morphology can select not only an agent or patient, but also locative, instrumental, and other adjunct type relations as subject.[1] Current syntactic treatments characterize this phenomenon alternatively as a rich voice system, a rich applicative system, "case agreement" or thematic nominalization. Relatedly, there is disagreement as to whether Philippine languages are best analyzed as ergative, accusative, active, or symmetrical. The uncomfortable position of Philippine languages in regard to more common alignment systems has been a long-standing topic in the typological literature (some earlier treatments include DeWolf 1988; Shibatani 1988; Himmelmann 1991).[2] Here, I present a critical review of three current syntactic analyses of Tagalog: the ergative analysis (Gerdts 1988b; De Guzman 1988; Liao 2004; Aldridge 2004), the case-agreement approach (Richards 2000; Rackowski 2002; Rackowski and

---

[1] I use the term "Philippine-type language" here to refer to a subset of Austronesian languages spoken in the Philippines as well as northern Sulawesi and Borneo. Proto-Malayo-Polynesian, the ancestor of all Austronesian languages spoken outside of Taiwan is reconstructed as a Philippine-type language although some Austronesian languages of Taiwan may also be considered Philippine type (e.g. Amis, Paiwan, Seediq). See Himmelmann (2005) for discussion.

[2] Richards (2013:fn.2), for instance, states, "I suspect that the debate about whether Tagalog is ergative will prove to be a terminological one; Tagalog resembles ergative languages in some respects, and differs from them in others, and the only question is how vague we want the technical term 'ergative' to be."

Richards 2005) and a nominalism analysis (Starosta et al. 1982; Kaufman 2009a,b), while advocating for the latter. I also present new data from Mamuju (South Sulawesi subgroup, Indonesia) to demonstrate that *canonical* ergative languages also exist within the Western Malayo-Polynesian subgroup of Austronesian. I argue that this retreat to the canonical ergative type is a result of the historical re-emergence of verbal predication.

The typical Austronesian "voice" paradigm is exemplified in (1) and (2) with Tagalog.[3] Note that the use of each morpheme correlates with the selection of a different argument or adjunct as the "*ang* phrase."

(1) a. B<um>ili   ng bulaklak ang bátà   para sa   bilanggô
       um.BEG-buy *ng* flower   *ang* child   for   *sa* prisoner
       'The child bought a flower for the prisoner.'

    b. B<in>ili-∅   ng bátà   ang bulaklak para sa   bilanggô
       <BEG>buy-*in* *ng* Juan   *ang* flower   for   *sa* prisoner
       'A/the child bought the flower for the prisoner.'

    c. I-b<in>ili   ng bátà   ng bulaklak ang bilanggô
       i-<BEG>give   *ng* child *ng* flower   *ang* prisoner
       'A/the child bought the prisoner a flower.'

    d. B<in>il-han   ng bátà   ng bulaklak ang bilanggô
       <BEG>give-*an*   *ng* child *ng* flower   *ang* prisoner
       'A/the child bought a flower from the prisoner.'

(2) a. Nag-bigay   ng bulaklak sa   bilanggô ang bátà
       mag.BEG-give *ng* flower   *sa* prisoner   *ang* child
       'The child gave a flower to the prisoner.'

    b. I-b<in>igay ng bátà   sa   bilanggô ang bulaklak
       i-<BEG>give *ng* child *sa* prisoner   *ang* flower
       'A/the child gave the flower to the prisoner.'

---

[3] The different verbal morphemes glossed ACT for ACTOR VOICE/AGREEMENT, PAT for PATIENT VOICE/AGREEMENT, CONV for CONVEYANCE VOICE/AGREEMENT and LOC for LOCATIVE VOICE/AGREEMENT. The glosses should be self-explanatory except for the conveyance voice. The types of arguments and adjuncts selected by the conveyance voice do not seem to form a natural semantic class. They include instruments, benefactives, and themes. I adopt the term conveyance voice from Wolff (1973) based on its selection of objects that are conveyed away from the agent. It is also referred to as instrumental, benefactive and circumstantial voice in the literature. Note that the actor voice is indicated by <*um*> in (i) but by *nag-* in (ii). The latter form is best analyzed as a combination of <*um*> and *pag-*, a transitivity related prefix, but this will not concern us here.

Regarding spelling conventions, the Tagalog genitive case marker *ng* is an abbreviation for /naŋ/. I follow the accentual conventions of (Wolff et al. 1991) where final stress is treated as the unmarked default and penultimate stress/length is marked with an acute accent. The grave accent is used to indicate word-final glottal stop. Thus, *bátà* indicates /ˈbaːtaʔ/.

    c. B<in>igy-an   ng bátà   ng bulaklak ang bilanggô
       <BEG>give-*an* *ng* child *ng* flower   *ang* prisoner
       'A/the child gave the prisoner a flower.'

The proper analysis of the alternating morphology on the predicates (1) and (2) (glossed neutrally as *<um>, nag-, i-, -an,* etc.) as well as the markers introducing the arguments (*ang, ng, sa*) are at the crux of the debate on Philippine-type languages. As emphasized by Foley (2008), what makes the system atypical is its morphologically symmetrical nature; note that in each of the examples (1) and (2), one and only one voice marker appears.[4] Following the typologically oriented work of Himmelmann (1991) and Foley (1998), I trace this symmetry to the root level, specifically, the lack of a clear contrast between entity-denoting and event-denoting roots. Crucially, when this contrast develops, as in many languages of Indonesia, more canonical alignment patterns emerge as well (Kaufman 2009b).[5]

24.1.1 The Ergative Analysis of Philippine Languages

Under Aldridge's (2004) ergative analysis of Tagalog, our example sentences in (1) would be analyzed as in (3), following the interpretation of the morphology presented in Table 24.1.

(3)   a. B<um>ili   ng bulaklak ang bátà   para sa   bilanggô
         ANTIPASS-buy OBL flower   ABS child for   PREP prisoner
         'The child bought a flower the prisoner.'

     b. B<in>ili-∅   ng bátà   ang bulaklak para sa   bilanggô
         <BEG>buy-TR ERG child ABS flower   for   PREP prisoner
         'A/the child bought the flower for the prisoner.'

     c. I-b<in>ili   ng bátà   ng bulaklak ang bilanggô
         APPL-<BEG>give ERG child OBL flower   ABS prisoner
         'A/the child bought the prisoner a flower.'

     d. B<in>il-han   ng bátà   ng bulaklak ang bilanggô
         <BEG>give-APPL ERG child OBL flower   ABS prisoner
         'A/the child bought a flower from the prisoner.'

---

   [4] Note that the symmetry referred to here is primarily morphological. The different "voices" are clearly not interchangeable in the majority of cases as they correspond to different possibilities for the definiteness of arguments. The term "symmetrical" has also been used in the sense of (Bresnan and Moshi 1990) to refer to the treatment of ditransitive objects. This is also unrelated to the morphological symmetry discussed here.

   [5] In later historical developments, this ergativity has also given way to a more accusative syntax, as it does in Indonesian and Malagasy. See Aldridge (2008a) for a formal treatment of accusative features in Austronesian.

The sentence in (3b) contains a canonical transitive clause with the zero allomorph of the transitive suffix -*in*.[6] In this clause, the agent is marked with ergative case and the patient is marked with absolutive case. The sentence in (3a) must then be analyzed as an antipassive construction in which the agent is marked with absolutive and the patient is marked with oblique. In Tagalog, there is no formal difference between oblique and ergative case marking. Other Philippine languages, such as Ivatan (Reid 1966), distinguish these functions with separate case markers.

Table 24.1 Ergative analysis of Tagalog morphology

Argument marking	*ang*	ABSOLUTIVE
	ng₁	ERGATIVE
	ng₂	OBLIQUE
	sa	PREPOSITION
Predicate marking	*<um>*	ANTIPASSIVE₁
	mag-	ANTIPASSIVE₂
	-in	TRANSITIVE
	-an	APPLICATIVE₁ (directional)
	i-	APPLICATIVE₂ (instrumental, benefactive, etc.)

In support of the antipassive analysis of (3a), we can note the indefinite/non-specific interpretation of the object, a fact which has been commented upon in all descriptions and analyses of Tagalog (Bloomfield 1917; Schachter and Otanes 1982: 76; Wolff et al. 1991; Kroeger 1993; Maclachlan and Nakamura 1997: 310; Richards 2000; Rackowski 2002; Kaufman 2005; as well as older descriptive works). As pointed out by proponents of the ergative analysis, this is a typical property of antipassive patients and is seen clearly in numerous Inuit and Mayan languages. In Tagalog, the objects of putative antipassive verbs are robustly indefinite (except when the agent is extracted, see Adams and Manaster-Ramer 1988). For instance, in (4a) we see that a definite demonstrative is not felicitous (without a partitive reading) in the object position of an antipassive verb. There is no such constraint on the clause in (4b), with what is a canonical transitive verb on the ergative analysis.

(4) a. *?K<um>áin   nito   ang   bátà
         <ANTIPASS>eat OBL.this   ABS   child
         (For, 'The child ate this.' OK for 'The child ate from this.')

---

[6] The zero allomorph of -*in* is used in conjunction with the aspectual infix *<in>*. The *<um>* infix also has a zero allomorph which is used in the prospective aspect. Because both of these voice markers have zero allomorphs in their aspectual paradigm, neither can be said to be less marked than the other.

b. K<in>áin-∅   ng   bátà   ito
   <BEG>eat-TR   ERG   child   ABS.this
   'A/the child ate this.'

According to the criteria of case and transitivity, Tagalog, and the vast majority of Philippine languages, are clearly ergative. The problem, as noted by Foley (1998, 2008) and Himmelmann (1991, 2005) among others, is that there are few if any ergative languages outside the Austronesian family that use the same morphology for putative antipassives as well as monadic (underlyingly intransitive) predicates. In Philippine languages, not only is *<um>* used in antipassive contexts as in (5a), it is used in underlyingly intransitive predicates as in (5b). The same problem is noted by Paul and Travis (2006: 321) for Malagasy *man-* and holds throughout Philippine-type languages.

(5)   a. P<um>atay   ng   kambing   si   Galvan
      <ANTIPASS.BEG>kill   OBL   goat   ABS   Galvan
      'Galvan killed a goat.'

    b. L<um>ákad   si   Galvan
      <ANTIPASS.BEG>walk   ABS   Galvan
      'Galvan walked.'

Furthermore, *<um>* is used on meteorological verbs, which differ from canonical intransitives in disallowing an overt absolutive argument, as shown in (6). In this case then it seems that *<um>* simply functions to create an event-denoting predicate rather than reducing valency or relating a particular thematic role to the *ang* phrase.

(6)   a. <Um>áraw   (*ang araw)
      <ANTIPASS.BEG>sun   ABS sun
      'It was sunny.'

    b. <Um>ulan   (*ang ulan)
      <ANTIPASS.BEG>rain   ABS rain
      'It rained.'

    c. L<um>indol   (*ang lindol)
      <ANTIPASS.BEG>earthquake   ABS earthquake
      'There was an earthquake.'

The question of the Tagalog antipassive is discussed at length by Aldridge (2012b), who argues that antipassives are not derived but rather combine with intransitive verbal morphology. Aldridge (2012b: 198-200) claims that a demotion analysis of antipassives requires downward movement of the demoted object, a possibility that is excluded by Chomsky (1995, 2005). While this is easy to argue from a Philippine perspective, it

remains to be explained why antipassive morphology only attaches to transitive verbs in other well-known ergative languages.[7]

The two other verbal alternations involving *i-* and *-an* are treated as applicatives on the ergative approach (Aldridge 2004). The first promotes benefactives, instrumentals and conveyance objects to absolutive while the latter promotes directional and locative arguments. There are two difficulties with treating these morphemes as applicatives. The first is that, unlike traditional applicatives, they are always required to introduce arguments that correspond to their functions, whether selected by the predicate or not. For example, a predicate like *bigay* 'give' requires applicatives to promote both the theme and the recipient to absolutive position, as shown in (7).

(7) a. Bigy-an   mo   ng   mangga   ang   bátà
       give-APPL   2S.ERG   OBL   mango   ABS   child
       'Give the child a mango.'

   b. I-bigay   mo   ang   mango   sa   bátà
       APPL-give   2S.ERG   ABS   mangga   PREP   child
       'Give the mango to the child.'

This is typologically unusual, as one of the two alternations is generally treated as basic, but in Tagalog and other Philippine languages all such arguments are derived as there is no unmarked verb. The second difficulty is that we expect applicative morphology to co-occur with *-in*, which under the ergative analysis must be treated as a simple transitive marker. Yet no Philippine language employs combinations such as *i*-STEM-*in* or STEM-*in-an* for these putative applicative constructions.

On the other hand, a strong piece of supporting evidence for the ergative view is the well-known restriction on extraction in Philippine and other Austronesian languages (as first discussed in a generative context by Keenan 1972). *Ng*-marked arguments in Tagalog cannot be topicalized, relativized, or clefted. This is demonstrated via ungrammatical topicalization of an ergative argument in (8) and an antipassive patient in (9). In both cases, topicalization of the corresponding absolutive argument is fully acceptable, as shown in the (a) sentences.

(8) a. Ang   bulaklak   ay   b<in>ili-∅   ng   bátà
       ABS   flower   TOP   <BEG>buy-TR   ERG   child
       'The flower, the child bought.'

   b. *Ng   bátà   ay   b<in>ili-∅   ang   bulaklak
       ERG   child   TOP   <BEG>buy-TR   ABS   flower
       (For, 'The child, bought the flower.')

---

[7] The Chol intransitive marker *-i* and transitive marker *-V* may, however, be relevant here. The *-i* suffix appears with underlying monadic predicates as well as derived intransitives, such as passives. Antipassives are formed with a transitive light verb and thus do not display *-i*. See Coon (2012) for discussion.

(9)   a. Ang   bátà   ay   b<um>ili   ng   bulaklak
         ABS   child   TOP   <ANTIPASS.BEG>buy   OBL   flower
         'The child, bought a flower.'

      b. *Ng   bulaklak   ay   b<um>ili   ang   bátà
         OBL   flower   TOP   <ANTIPASS.BEG>buy   ABS   child
         (For, 'A flower, the child bought.')

Similar restrictions are attested in Mayan and Inuit languages among others. It is these facts, together with the reduced transitivity of "actor voice" verbs marked with *<um>* and *mag-*, that form the strongest arguments for the ergativity analysis of Philippine-type languages. To review, the most basic arguments for the ergative view are shown in (10) and those against it, in (11).
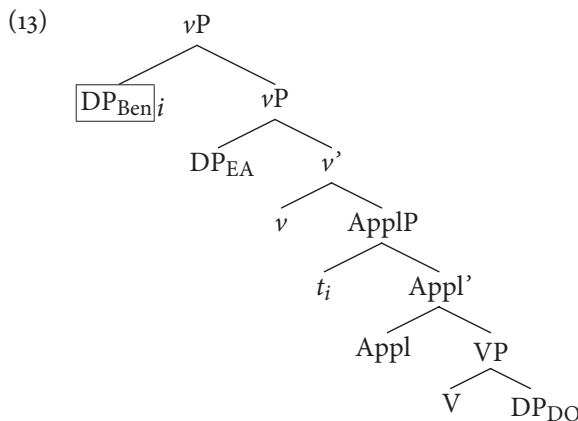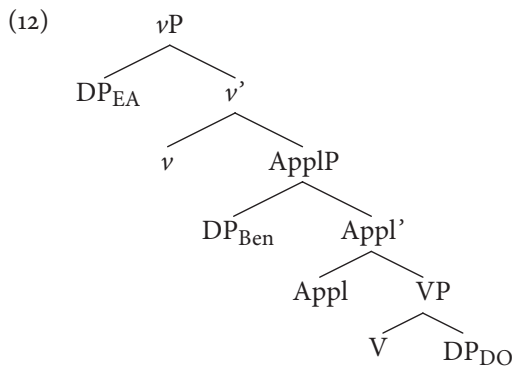
(10)   a. indefiniteness (and low scope) of "actor voice" objects
       b. restrictions on the extraction of "actor voice" objects and transitive agents
       c. case marking pattern (e.g. gen/erg syncretism) common to other ergative languages

(11)   a. absence of unmarked transitive and intransitive predicates
       b. complementary distribution of "transitive" *-in*, and "applicatives" *i-* conv and *-an* LOC.
       c. absence of a true de-transitivizing antipassive

24.1.2 The Case Agreement Analysis of Philippine Languages

Chung (1998), Richards (2000), Rackowski (2002) and Pearson (2001) present variations of an agreement approach to the Austronesian verbal alternations seen above in which the morphology instantiates agreement with an argument in an A-bar position (see also Erlewine, Levin, and Van Urk, Chapter 16, this volume, for a similar approach employing an A-position). I restrict my focus here to the analyses of Richards (2000), Rackowski (2002) and Rackowski and Richards (2005) as the languages investigated by Chung (1998) and Pearson (2001) (Chamorro and Malagasy, respectively) differ from Tagalog in some important respects.

   In case agreement approaches, all arguments have their case checked in their respective A-positions after which object shift may occur. Object shift moves a DP to an outer specifier of vP above the merged position of the external argument. The verb then agrees with the highest argument, not for canonical agreement features, but rather for its case features. Rackowski (2002) proposes that complements of verbs receive accusative case and verbs register accusative agreement with *-in* (or its zero allomorph). External arguments receive nominative case from T with the corresponding verbal agreement being

<um> or *mag-*. Nominative agreement with the external argument only takes place when no other argument has moved to a higher position. This is the case when the direct object is indefinite and there are no applicatives present. On this analysis, Tagalog also has null high and low applicatives (cf. Pylkkänen 2002) which promote benefactives, instrumentals, locatives, and other types of arguments and adjuncts. The high applicative assigns oblique case and the verb displays oblique agreement with the prefix *i-*. The low applicative assigns dative case and the verb displays dative agreement with the suffix *-an*. The merged structure of a benefactive (high) applicative construction is shown in (12) and the result of subsequent movement in (13). The applicative argument ends up on the edge of vP and enters into an agreement relationship with the verb which is spelled out as *i-*.

(12)

```
            vP
           /  \
        DP_EA  v'
              /  \
             v   ApplP
                 /   \
            DP_Ben   Appl'
                     /   \
                  Appl    VP
                         /  \
                        V  DP_DO
```

(13)

```
               vP
              /  \
        [DP_Ben]_i  vP
                   /  \
                DP_EA  v'
                      /  \
                     v   ApplP
                         /   \
                       t_i   Appl'
                             /   \
                          Appl    VP
                                 /  \
                                V  DP_DO
```

This analysis is significantly more abstract than that of the ergative analysis, in which voice morphology is interpreted directly as applicatives and transitivity markers. Here, the actual applicatives are only detectable indirectly through distinct agreement on the predicate. Note also that the cases with which the predicate agrees are abstract. As seen earlier, there are only three phonologically distinct argument markers in Tagalog but for the case agreement analysis to hold, there must be distinct oblique and dative cases underlyingly that the verb agrees with. The interpretation of Tagalog's core functional morphology under this analysis is shown in Table 24.2.

Table 24.2 Case agreement analysis of Tagalog morphology

Argument marking	*ang*	AGREEMENT TRIGGER
	ng	DEFAULT CASE
	sa	PREPOSITION
Predicate marking	*<um>*	NOMINATIVE AGREEMENT$_1$
	mag-	NOMINATIVE AGREEMENT$_2$
	-in	ACCUSATIVE AGREEMENT
	-an	DATIVE AGREEMENT (directional)
	i-	OBLIQUE AGREEMENT (instrumental, benefactive, etc.)

The *ang* marker, interpreted as absolute case on the ergative analysis, is treated as a topic marker of sorts under case agreement analyses.[8] The *ng*-case, which instantiates both ergative case and oblique case on the ergative analysis is simply glossed as case in the case-agreement analysis. It is not clear what kind of case it is and the details of its assignment remain to be explained.

As with the ergative analysis, there are also difficult typological and theoretical questions that arise with ostensive applicatives in the case agreement approach. First of all, applicative objects must somehow retain their underlying (oblique and dative) case for case agreement to take place but in familiar nominative-accusative languages applicatives promote arguments to a position where they receive accusative case and behave as canonical direct objects.[9] Second, it remains unclear why applicative objects must shift to the edge of vP and block agreement with the external argument. Without such movement, we expect to find structures like that in (14b), where the *ang* phrase

---

[8] However, the use of "topic" cannot be taken too literally, as the *ang* phrase has been shown conclusively to not be an actual discourse topic (Kroeger 1993; Adams and Manaster-Ramer 1988; Kaufman 2005). For instance, in the following mini-dialogue, an *ang* phrase appears within the focus of a completely felicitous response. But see Richards (2000) for arguments that the *ang* phrase is similar to the topic position in Icelandic and other Germanic languages.

(i) A: Ano ang g<in>á~gawà   ni   Jojo
         what *ang* <BEG>IMPRF~do   CASE   Jojo?
         'What is Jojo doing?'

    B: H<in>ú~hugás-an   niya   ang   mga   pinggan.
        <BEG>IMPRF~wash   CASE.3S   *ang*   PL   dish
        'He's washing the dishes.'

[9] The literature on Bantu applicatives bears this out as well as studies of non-accusative languages such as those of the Salish family. Kiyosawa and Gerdts (2010:41), for instance, shows that applicative objects across Salish languages are extracted in exactly the same manner as direct objects, not as obliques, which require nominalization of the clause.

benefactive fails to trigger case agreement on the verb. As pointed out by Rackowski and Richards (2005), such promotion without further movement is ungrammatical. They offer potential analogues from other syntactic domains but little if any cross-linguistic support exists for applicative objects blocking agreement with an external argument in nominative-accusative languages.[10]

(14) a. Ni-lútò-Ø   ni   Romeo   ang   adóbo   para   sa   babáe
          BEG-cook-ACC   CASE   Romeo   *ang* adobo   for   PREP   woman
          'Romeo cooked the adobo for the woman.'

                                                            (Rackowski and Richards 2005:ex.13a)

      b. Ni-lútò-Ø   ni   Romeo   ng   babáe   ang   adóbo
          BEG-cook-ACC   CASE   Romeo   CASE   woman   *ang*   adobo
          'Romeo cooked the adobo for a woman.'

                                                            (Rackowski and Richards 2005:ex.11b)

The theory makes crucial use of Pylkkanen's (2002) distinction between low and high applicatives but the Tagalog facts do not accord well with the theory's prediction. Low applicatives relate a promoted argument to a direct object and are crucially dependent on the presence of a direct object. Rackowski (2002) argues that this is the case based on the data in (15).

(15) a. B<in>igy-an   ko   ang   ama   ng   anak.
          <BEG>give-DAT   1S   *ang*   father   CASE   child
          'I gave the father his child.'

      b. *B<in>igy-an   ko   *ang*   ama.
          <BEG>give-DAT   1S   *ang*   father
          'I gave (to) the father.'

However, the judgment reported for in (15b) seems better attributed to pragmatic rather than syntactic factors. There can be no act of giving without a theme and there is thus an expectation that the theme will be included in an out-of-the-blue context. Similar examples could be produced without a theme but more conclusive is the data shown in (16), which demonstrates that monadic predicates of all types can take -*an*, counter to the expectations of a low applicative in Pylkkanen's framework.

---

[10] An analogy is made to the *wager*-class of verbs in English, in which an argument must move from its underlying position with no obvious motivation. Collins (2002) and Epstein and Seely (2006:83-5) question the judgments surrounding this phenomenon in English and it remains to be seen if the proposed link can be further substantiated.

(i) a. *John wagered Mary to have won the race.
     b. Mary$_i$ was wagered t_i to have won the race.

(16) a. Ni-lakár-an   ng   bátà   ang   daan
          BEG-walk-DAT   CASE   child   *ang*   road
          'The child walked the road.'

      b. S<in>igaw-an   ni   Romeo   si   Jojo
          <BEG>shout-DAT   CASE   Romeo   *ang*   Jojo
          'Romeo shouted at Jojo.'

      c. <In>ubu-han   ni   Romeo   si   Jojo
          <BEG>cough-DAT   CASE   Romeo   *ang*   Jojo
          'Romeo coughed at Jojo.'

      d. Na-ulan-an   si   Jojo
          STA.BEG-rain-DAT   *ang*   Jojo
          'Jojo was rained on.'

The role of specificity shift is crucial in both the ergative and case agreement approaches to Tagalog. Both predict, in their own ways, that a definite/specific object should not be possible for a matrix clause verb marked with <*um*> or *mag-*. On the ergative analysis, these verbs are antipassives whose objects are assigned inherent oblique case from *v* and remain in VP. In VP, they are subject to existential closure (Diesing 1992) and receive an existential indefinite interpretation. In the case agreement analysis, specificity triggers object shift, which puts a shifted object above the external argument and thereby triggers accusative agreement rather than nominative agreement with <*um*> or *mag-*. Neither approach, however, seems to predict all the specificity facts of applicative constructions correctly. Rackowski (2002) and Rackowski and Richards (2005) explicitly claim that the underlying direct object in applicative constructions can be specific, offering the example in (17) as evidence.

(17) I-pinaglútò   ni   Romeo   ng   adóbo   ang   babáe
      OBL-cook   CASE   Romeo   CASE   adobo   *ang*   woman
      'Romeo cooked (the) adobo for the woman.' (Rackowski and Richards 2005: 570)

However, there is no independent evidence for the definite interpretation of the object in structures like (17). Nothing in (17) forces a specific interpretation and, in fact, evidence points to definite direct objects being just as marked in structures such as (17) as they are as object of verbs marked with <*um*> or *mag-*. This becomes clear by using pronominal arguments as diagnostics. As Rackowski and Richards (2005) show with (18), a pronominal object of an <*um*> marked verb is ungrammatical. Unexpectedly, the same ungrammaticality appears with direct objects of *-an* and *i-* marked verbs, as in (19).

(18) *S<um>ampal   ko   ang   mandurukot.
       <NOM.ASP>slap   1S.CASE   *ang*   pickpocket
       (For, 'The pickpocket slapped me.')   (Rackowski and Richards 2005: 568, ex.4b)

(19) B<in>igy-an   ni   Maria   (ng   pera   /*ni   Romeo)   si
     <BEG>cook-DAT   CASE   Maria   CASE   money   / CASE Romeo   *ang*
     Juliette
     Juliette
     'Maria gave money/*Romeo to Juliette.'

In a regular matrix clause with a definite object and an oblique, the object must surface with *ang* case and the oblique must surface as an oblique/prepositional object, as shown in (20).

(20) I-b<in>igay   ni   Maria   si   Romeo   kay   Juliette
     OBL-<BEG>cook   CASE   Maria   *ang*   Romeo   OBL   Juliette
     'Maria gave Romeo to Juliette.'

Rackowski and Richards (2005: 568) follow Chomsky (2001) in positing that:

> object shift occurs as the result of an EPP-feature on v that is present only when it has an effect on semantic outcome. There is an effect on semantic outcome because the position at the edge of the vP is assigned a specific interpretation, while everything internal to vP is assigned a nonspecific interpretation.

But there is an issue of derivational look-ahead in this account as well as an empirical problem. As *ng*-marked objects cannot be specific/definite in a regular matrix clauses, we must assume that they are barred from leaving vP. This means that the presence of an applicative object must arbitrarily preclude an EPP-feature on *v* on this account. Barring any larger generalization which can be extracted from this coincidence, it merely describes the facts.

Regarding *ng*-marked agents, Rackowski & Richards recognize the need to account for both an existential indefinite as well as a specific/definite reading for them. As seen in (21), a *ng*-marked agent can felicitously be a proper name or an indefinite pronoun.

(21) Hindi   b<in>igy-an   (ni   Jose/ninuman)   si   Maria   ng
     NEG   <BEG>give-DAT   CASE   Jose/CASE.anyone *ang* Maria   CASE
     pera
     money
     'Jose/nobody gave Maria money.'

The reasoning for this ambiguity is that only internal arguments have the *opportunity* to undergo object shift. For arguments that do not have such an opportunity, specificity is unpredictable. But as we have seen in (19), this empirical claim cannot be upheld. A separate treatment is required for *ng*-marked objects, which cannot be definite/specific, and *ng*-marked agents, which can. This partially supports the ergative analysis in that ergative arguments are not known to show the kind of definiteness effects attested for antipassive patients. However, the restricted interpretation of direct objects of *-an* and

i- verbs is equally problematic. On the ergative analysis, non-specific readings are only forced for antipassive patients but something extra must be said for *-an* and *i-* verbs, which can only be analyzed as transitive.

We find that the case-agreement analysis captures several important truths about Philippine voice systems, enumerated in (22). On the other hand, there remain theoretical issues with the analysis, some of which are summarized in (23).

(22)   a.   no unmarked verbs, (virtually) all inflected verbs show "voice morphology"
       b.   "voice morphology" is only marked once per verb
       c.   range of meanings associated with "voice morphology" resembles that of case marking

(23)   a.   incorrect predictions regarding definiteness/specificity
       b.   unusual properties of (null) applicatives
       c.   unclear what morphological case is and how it is assigned

24.1.3   The Nominalization Analysis of Philippine Languages

Yet another interpretation of Philippine-type verbal alternations views them as participant nominalizations. This was first proposed by Starosta et al. (1982) on a historical basis and further developed by Ross (2002, 2009) and, in synchronic terms, by Kaufman (2009a).[11] On this approach, each "voice" indicates a different participant nominalization, as shown in Table 24.3.[12] The three argument markers are interpreted as case, as in traditional descriptions. The *ang* marker indicates subject case, which can be termed either nominative or absolutive. The *ng*-case is first and foremost genitive case, that is, the case of possessors, but is recruited for other purposes due to the regular use of nominalized predicates. Finally, *sa* is an oblique case marker (and not a preposition as in the two approaches reviewed above).[13]

---

[11]   This also builds on an important body of work beginning with Bloomfield's (1917) Tagalog grammar and furthered by Capell (1964), Naylor (1980), DeWolf (1988), Himmelmann (1987, 1991, 2008) and Gil (1993, 1995, 2000) who propose precategorial analyses for Tagalog, or, in the case of Himmelmann (2008), a distinction that cross-cuts traditional categories. The technical side of the analysis has a close analogue in Johns' (1992) nominalization-based approach to Inuktitut. See also Coon (2014) for a similarly *v*-less analysis of partially overlapping facts in the morphosyntax of Chol, a Mayan language.

[12]   For reasons of space, I do not discuss the difference between <*um*> and *mag-* although I treat *mag-* as containing an inner causative as suggested by Travis (2000). This is supported in Kaufman 2009c where I show comparative evidence for *mag-* being composed historically, and potentially synchronically, of the morphemes *p*<*um*>*a*<*R*>- <AV>CAUS<MIDDLE>-.

[13]   For all locality based approaches to the extraction restriction, including case agreement and the ergative analysis, there must in fact only be two cases. The oblique case, marked by *sa*, is treated as a preposition by these approaches because *sa* phrases can typically be fronted and topicalized regardless of voice. In other words, oblique phrases are effectively invisible for calculations of locality under previous approaches. But there are many reasons to believe that *sa* is a case marker rather than a preposition. It is in complementary distribution with the other case markers and is selected by bona fide prepositions, e.g.

Table 24.3 Nominalist analysis of Tagalog morphology		
Argument marking	*ang*	NOMINATIVE/ABSOLUTIVE
	ng	GENITIVE
	sa	OBLIQUE
Predicate marking	*<um>*	ACTOR NOMINALIZATION
	mag-	INNER CAUSATIVE + ACTOR NOMINALIZATION
	-in	PATIENT NOMINALIZATION
	-an	LOCATIVE NOMINALIZATION
	i-	CIRCUMSTANTIAL NOMINALIZATION

This view differs from the previous two approaches in treating all predication in Philippine languages as inherently copular. The interpretation of the *ang* phrase is not derived through the use of applicatives but rather through copular identification with the predicate itself. To compare with Johns' (1992) analysis of Inuktitut, she proposes that predications are formed compositionally in the manner shown in (24).

(24) a. kapi-jaq
          stab-PASS.PART
          'the stabbed one'

    b. anguti-up   kapi-ja-a
          man-ERG   stab-PASS.PART-3S/3S
          'the man's stabbed one' OR 'the one that the man stabbed'

    c. anguti-up   nanuq   kapi-ja-a
          man-ERG   polar   bear.ABS   stab-PASS.PART-3S/3S
          'The polar bear is the man's stabbed one.' OR 'The man stabbed the polar bear.'
                                                                            (Johns 1992)

Applying this approach to Tagalog, we arrive at the following literal translations of for the basic "voice" alternations:

(25) a. K<um>áin   ng   dagà   ang   púsà
          <ACT.BEG>eat   GEN   rat   NOM   cat
          'The cat was the eater of a rat.'

    b. K<in>áin-∅   ng   púsà   ang   dagà
          <BEG>eat-PAT   GEN   cat   NOM   rat
          'The rat was the eaten one of the cat.'

c. K<in>aín-an   ng   púsà   ng   dagà   ang   pinggan
   <BEG>eat-LOC   GEN   cat   GEN   rat   NOM   plate
   'The plate was the cat's eating place of the rat.'

d. I-k<in>áin   ng   púsà   ng   dagà   ang   áso
   CONV-<BEG>eat   GEN   cat   GEN   rat   NOM   dog
   'The dog was the cat's "eating benefactor" of the rat.'   (Kaufman 2009a: 6)

This avoids the difficulties of theta-linking approaches discussed by Rackowski (2002) as nominalization, like case, is well known to reflect thematic roles imprecisely. More importantly, nominalization also makes sense of the fact that Philippine predicate morphology does not distinguish between argument and adjunct. Barker (1998: 714) discusses at length the way in which English *-ee* nominalizations select participants that are not part of the argument structure of the corresponding verb.[14] This offers an excellent analogue to one of the more typologically difficult aspects of Philippine-type voice.

As argued in Kaufman (2009a), the nominal nature of event-denoting predicates in Tagalog can also go a long way in explaining other curiosities of Philippine syntax, including:

(26) a. Near identical syntactic distribution of canonical event-denoting predicates (i.e. "verbs") and entity-denoting predicates (i.e. "nouns").
      b. Identity between the case of ergative agents and possessors across Philippine languages.
      c. Ungrammaticality of extracting genitive marked arguments corresponds with the difficulties of extracting possessors and other nominal dependents cross-linguistically.
      d. Symmetric nature of Philippine "voice" – only one instance per predicate, as would be expected of nominalization morphology.

gáling 'from', *patúngo* 'towards.' The oblique also displays a human/non-human distinction (*sa* for non-humans vs. *kay* for humans) which is a unique feature of the two other case markers and not found on any of the bona fide prepositions. These facts alone should dispel any notions that the oblique could be a preposition but see Kaufman (2009a:40) and Gerassimova and Sells (2008:196-197) for further arguments.

[14] Barker utilizes the notion of "episodic linking" rather than argument structure to account for how participant nominalizations in English identify their referent.

> … the meaning of the verb amputate guarantees the existence of a person undergoing amputation, even though there is no syntactic argument that corresponds to this participant… the fact that the person undergoing amputation is a participant of every amputation event is sufficient to enable a set of amputation events to characterize the *-ee* noun amputee: for each amputation event *e*, there exists an individual *x* which is a participant in *e* such that *x* is (becomes) an amputee. Thus amputee is episodically linked to the meaning of amputate despite the fact that there is no corresponding syntactic argument position.

Similarly, Aronoff (1980) discusses the role of Gricean principles in certain morphosyntactic alternations.

Event-denoting predicates in Philippine-type languages can always serve as the direct complement to case markers.[15] There is no strong evidence from Philippine languages for null complementizers or headless relatives in (27b) and (c).[16]

(27) a. Ang áso
          NOM dog
          'the dog'

      b. Ang lú~lutú-in
          NOM IMPRF~cook-PAT
          'the thing to be cooked'

      c. Ang mag lú~lutò
          NOM ACT-IMPRF~cook
          'the one who will cook'

The syntax of *ng*-marked possessors and *ng*-marked agents appears identical. That is, there is no reason to believe that the addition of the morphology in (28b) leads to a substantial difference in syntactic structure between (28a) and (b).

(28) a. Súlat   ni   Juan   iyan
          write   GEN   Juan   that:NOM
          'That is Juan's letter.'

      b. S<in>úlat-∅   ni   Juan   iyan
          <BEG>write-PAT   GEN   Juan   that:NOM
          'Juan wrote that.'

Johns (1992) makes the same claim for Inuktitut pairs such as those in (29). Inuktitut *ja* in (29b) functions similarly to the Tagalog patient nominalizer *-in*/∅. In both cases, a NP or DP predicate combines with a structural subject through a copular structure to yield a predication such as (28b).[17]

(29) a. anguti-up   qimmi-a
          man-GEN   dog-3s/3s
          'the man's dog'

---

[15] Kaufman (2009a:25) points out an important exception to this pattern with the descendants of the original (non-nominalized) verbs in certain dialects of Tagalog.

[16] But see Richards (2009) for some arguments to this effect and Kaufman (forthcoming) for a defense of the symmetry.

[17] This is similar to Pearson's 2005 proposal of base generating the equivalent of the *ang*-phrase in Malagasy in an A-bar position located in a high rightwards specifier and co-indexed with a lower null operator in an A-position.

b. anguti-up   kapi-ja-a
    man-GEN   stab-PASS.PART-3S/3S
    'the man's stabbed one'
    OR 'the one that the man stabbed'

There are two potentially independent conditions conspiring against fronting genitive arguments. On one hand, extracting ergatives should bear similarity to sub-extraction from NP in other languages, a highly constrained operation. Indeed, in Tagalog itself extraction of possessors is subject to precisely the same constraints as ergative extraction as shown by the illicit topicalizations in (30).[18]

(30)   a. B<in>ili-∅   ni   Maria   ang   kótse   ni   Juan.
         <BEG>buy-PAT   GEN   Maria   NOM   car   GEN   Juan
         'Maria bought Juan's car.'

     b. *Ni   Juan   ay   b<in>ili-∅   ~~ko~~   ang   kótse _____.
        GEN   Juan   TOP   <BEG>buy-PAT   ~~3S.GEN~~   NOM   car
        (For, 'Juan, Maria bought (his) car.')

     c. *Ni   Maria   ay   b<in>ili-∅   _____   ang   kótse   ni   Juan.
        GEN   Maria   TOP   <BEG>buy-PAT   NOM   car   GEN   Juan
        (For, 'Maria, (she) bought Juan's car.')

---

[18] Richards cites Cena's (1979) examples in (i) of apparent possessor extraction as an argument against constraints on linking ergative extraction to possessor extraction.

    (i)   a. Kasama   ng   doktor   ang   anak
         companion   GEN   doctor   NOM   child
         'The child is with the doctor'

      b. ang   doktor   [na   kasama   ang   anak]
        NOM   doctor   LNK   companion   NOM   child
        'the doctor that the child is with'

                                        (Cena 1979; Richards 2009)

As shown in (30), possessor extraction of the normal type is subject to exactly the same constraints as ergative extraction. A minimal pair for Cena's example with an ostensibly verbal predicate, shown in (ii), demonstrates clearly that relativization of the ergative argument is no more marked than relativization of the possessor. I leave the analysis of this construction to further work noting only that Cena's examples do not entail different treatments for possessors and ergative arguments.

    (ii)   a. s<in>a~sama-han   ng   doktor   ang   anak
         <BEG>IMPRF~accompany-LOC   GEN   doctor   NOM   child
         'the doctor accompanies his/her child'

      b. ang   doktor   [na   s<in>a~sama-han   ang   anak]
        NOM   doctor   LNK   <BEG>IMPRF~accompany-LOC   NOM   child
        'the doctor that accompanies his/her child'

Potentially separate from the issue of extraction from NP is the general ban on genitive predicates. This is independently necessary in Tagalog as shown in (31) and has clear cognates cross-linguistically. Genitive phrases can be modifiers, as in (31a) but not predicates, as shown in (31b). As in Hungarian and many other languages (Szabolcsi 1983), predicate and extracted possessors must be expressed as obliques or datives, as in (32).

(31) a. Ang   súlat   ni   Juan
         NOM   write   GEN   Juan
         'The letter of Juan's'

      b. *Ni   Juan   ang   súlat
         GEN   Juan   NOM   write
         (For, 'The letter is Juan's.')

(32) Kay   Juan   ang   súlat
      OBL   Juan   NOM   write
      'The letter is Juan's.'

Argument questions in Philippine-type languages have been widely analyzed with the interrogative phrase as the predicate of the clause (Paul 2001; Aldridge 2002; Oda 2002; Massam 2003; Potsdam 2006, 2009; Gerassimova and Sells 2008). As we have already seen that genitive predicates are banned, we also correctly predict that genitive case interrogatives should be ruled out, as seen in (33). Note that the genitive interrogative *is* allowable in an in-situ post-nominal position, as shown in (34) (see also Erlewine, Levin, and Van Urk, this volume, Chapter 16, n.21).

(33) a. Kaníno   ang   súlat?
         OBL.who   NOM   write
         'Whose is the letter?'

      b. *Nino   ang   súlat?
         GEN.who   NOM   write
         (For, 'Whose is the letter?')

(34) Ang   súlat   nino?!?
      NOM   write   GEN.who
      'The letter of who?!?'

Similarly, if event-denoting predicates in Philippine-type languages are nominalized, we can rule out interrogatives like those in (35b) and (c), which attempt to make a predicate out of a genitive agent. The ungrammaticality in (35b) ensues from the constraint against genitive predicates and (35c) is ruled out on the simple basis of case preservation. What should be a genitive marked agent, as in (35a), cannot be expressed as an interrogative in the nominative (*ang*) case. A copular predication has at most two nominative marked

arguments, one in the predicate position and one in the subject position. The nominal approach to Philippine-type languages is thus uniquely able to unify the ungrammaticality behind (34) and (35), which is common to the vast majority of Philippine-type languages.

(35) a. Sú~sulát-in   ni   Jojo   ang   libro
         IMPRF~write-PAT   GEN   Jojo   NOM   book
         'Jojo will write the book.'

      b. *Nino   ang   sú~sulát-in   ang   libro?
         GEN.who   NOM   IMPRF~write-PAT   NOM   book
         (For, 'Who will write the book?')

      c. *Sino   ang sú~sulát-in   ang   libro?
         NOM.who   NOM IMPRF~write-PAT   NOM   book
         (For, 'Who will write the book?')

On this approach, genitive marked arguments in Philippine-type languages bear a certain resemblance to English *of*-phrases. Just like *ng*-phrases, *of*-phrases are highly restricted in predicate position (36), and also display restrictions on their extraction from NP, as shown in (37).[19]

(36)   The team is (*of) Juan's.

(37)   a. Juan was an employee of Rizal.
       b. *?Of whom was Juan an employee?

In (38), we see that these constraints do not hold for instrumental agents of passive verbs.

(38)   a. Juan was employed by Rizal.
       b. By whom was Juan employed?

As pointed out in Kaufman 2009a, this could hold the key to why some ergative languages show a restriction on extracting the ergative argument while others do not (see Manning 1996 and Dixon 1994 for discussion). If the ergative argument is treated as a nominal dependent, we expect it to share a morphological case with possessors and resist extraction. If the ergative argument is a verbal dependent, we

---

[19] See Davies and Dubinsky (2003) for a review of approaches to extraction from NP. Davies and Dubinsky rely on participant structure to derive the complex pattern of extractability from English NPs but they do not consider extraction from participant nominalizations. Other approaches have implicated the left branch condition (Ross 1967), the ECP (Chomsky 1981), the case filter (Huang 1982), movement of non-constituents (Bošković 2005), and most recently, the treatment of *n*P or DP as a strong phase (Chomsky 2001; Svenonius 2004).

expect it to share a morphological case with instrumentals or obliques and to allow extraction more freely. In particular, it seems that a genitive-ergative syncretism in combination with a pseudo-cleft strategy for interrogatives is what conspires to constrain extraction in ergative languages. In languages with the same syncretism but in-situ interrogatives, as Inuktitut, constraints against ergative interrogatives are not attested, as evidenced by (39). Indeed, even in the rare contexts that Tagalog allows wh-in-situ, as in (40), genitive interrogatives become acceptable (compare above).

(39) Kia   Alaana   kii-ja-nga?
     who.GEN   Alana.NOM   bite-PASS.PART-3S/3S
     'Who bit Alana?'   (Yuan 2013)

(40) G<in>awà-∅   nino   ang   sapatos   na   iyon?
     <BEG>make-PAT   GEN.who   NOM   shoe   LNK   that
     'Who made those shoes?'   (Schachter and Otanes 1982: 512)

The typological correlation is worth investigating on a larger scale but here we are most concerned with what happens when a bona fide *v* category develops from *n* and its consequences for ergativity. The crucial properties of the *v* versus *n* heads which determine lexical category are the following.

(41) *n* properties:
      a. possessor is projected [Spec, n]
      b. GENITIVE case to *n*P-internal phrases
      c. strong island properties

(42) *v* properties:
      a. agent is projected in [Spec, v]
      b. ACCUSATIVE case to object
      c. islandhood dependent on *v* features

The *v* category projects an agent and more generally, *v* is associated argument structure that is arguably not present in low nominalizations. While apparent arguments can nonetheless be expressed in low nominalizations, they are not distinguished by a distinct object case nor are they obligatory. The general assignment of genitive case within *n*P can be seen in (43), where the agent, instrument and a manner adverbial are all marked with *ng*-case.

(43) *nP* [B<in>uks-an   ko   ng   súsi   ng   maingay]   ang   pintuan
        <BEG>open-LOC   1S.GEN   GEN   key   GEN   loud   NOM   door
        'I opened the door loudly with the key.'

Note also that the ergative argument is not obligatory nor must it be interpreted as *pro* when omitted in line with nominalizations cross-linguistically (Abney 1987; Alexiadou 2001; Himmelmann 1991: 22). The sentence in (44), for instance, can be uttered in an out-of-the-blue context without any implication regarding an agent. Optionality is expected for a possessor but not for the external argument of a transitive verbal projection.[20]

(44) Mukhang   k<in>ain-∅   ang   lahat
      seem:LNK   <BEG>eat-PAT   NOM   all
      'It seems that everything has been eaten.'

Relatedly, there are also patient voice predicates that function as adversatives and cannot take agents of any sort. Two such examples are shown in (45) and (46).[21]

(45) Ni-langgam-∅   ang   asúkal
      BEG-ant-PAT   NOM   sugar
      'The sugar was infested with ants.'

(46) In-ú~ubo-∅   ang   bátà
      BEG-cough-PAT   NOM   child
      'The child is (effected by) coughing.'

As discussed earlier, the full range of definiteness/specificity effects have presented difficulties to previous treatments of Tagalog. In the simplest case, which all theories have an account for, actor voice patients tend to be non-specific and strictly disallow pronominals, as seen in (47). More difficult to account for is the parallel constraint on themes of locative and circumstantial voice predicates, as seen earlier in (19).

(47) a. Nag-pi~pinta   ng   mukha   si   Juan
         ACT.BEG-IMPRF~paint   GEN   face   NOM   Juan
         'John paints faces.'

      b. *Nag-pi~pinta   ko   si   Juan
         ACT.BEG-IMPRF~paint ~~1s.~~ ~~GEN~~ NOM   Juan
         (For, 'John paints me.')

---

[20] As Lawrence Reid points out (p.c.), this is not the normal state of affairs in the Cordilleran languages, spoken in North Luzon, Philippines. In many of these languages it seems that an ergative agent is obligatory with transitive predicates. I take this property to be a later historical development based on its narrow distribution but nonetheless very important with regard to the viability of a nominal analysis for the Cordilleran subgroup.

[21] We can again draw a parallel to English patient nominalizations with *-ee* in that they appear to operate outside of argument structure (Barker 1998), attaching to monadic bases (*standee, escapee*) as well as nominal bases (*hoaxee*).

The nominalization approach suggests a parallel between these facts and similar constraints on definite *of*-phrase objects of nominals in English, as shown in (48).[22]

(48) a.   John is a painter of portraits
      b.   *John is a painter of me

In English, this effect extends to themes of recipient nominalizations, as shown in (49). While (49a) shows that a definite theme is possible, a pronominal theme is completely unacceptable. Definite genitive case themes are thus possible in both English and Tagalog (under the right circumstances), but both languages strictly disallow pronominals in this position. A formalization of these facts ~~must be postponed to later work~~ but ~~the~~ parallels ~~shown here~~ are clearly promising.

(49) a.   John was the awardee of the nobel prize.
      b.   *John was the awardee of it.

In regard to the semantics of the nominative/absolute phrase, it seems far more felicitous to attribute its referential properties to the case marker itself as an operator rather than strictly to object shift (as originally advocated by Himmelmann 1991). Out-of-the-blue exclamations such as those in (50) support such an approach. It would make little sense to derive the definiteness of (50b) via object shift or agreement with T. On the analysis suggested here, *ang* contains both case and definiteness features.[23]

(50) a.   Dagà!
           rat
           'A rat!'

      b.   Ang pangúlo!
           ang president
           'The president!'

In the next subsection, we look at root level phenomena to demonstrate that much of the syntax attributed to the voice/agreement system is already present at the first phase of word building, further suggesting the absence of *v* in the functional inventory.

---

[22] Note that *John is a painter of mine* is acceptable on the possessive reading where I have a painter. According to several speakers I have consulted, this reading is also acceptable for the Tagalog constructions as in (47b) although I have not found naturally occurring examples of an aspect-inflected actor voice predicate with a possessor.

[23] This approach is further supported by other Central Philippine languages which distinguish [±SPECIFIC] variants for each case marker (see McFarland 1974 and Zorc 1977 for examples). Of course, an analysis in which *ang* functions as a specific/definite determiner is not necessarily incompatible with object shift. As Edith Aldridge points out (p.c.), Austronesian languages which still show the same specificity/definiteness pattern without overt case markers suggest that a structural approach might still be necessary.

24.1.4 Ergativity on the root level

Foley (1998) argues that Tagalog roots are fundamentally different from those of English:

> The lexeme *give* in English is a verb, with a corresponding argument structure <actor, undergoer, locative>. The Tagalog root *bigay* 'give' however is precategorial; it lacks a true argument structure, but does have a precategorial semantic structure like 'the giving of something to someone by someone.'

On this view, it is the voice morphology which imbues words with argument structure rather than anything in the underlying semantic representation. This explains why notional valency has little consequence for the voice marking potential of roots in Philippine-type languages. There are, however, problems with the claim that roots lack category and argument structure altogether. In contrast to early claims made about Tagalog roots, they appear regularly in their bare form as both predicates and arguments. Unlike the predictions of the precategorial analysis, bare roots are restricted to particular readings and do not display flexibility in their distribution of thematic roles to genitive and nominative marked arguments. As discussed by Himmelmann (1991: 40) and Kaufman (2009a), the reading of a bare root is consistently that of a patient-oriented (or "proto-object") entity. In Table 24.4, we find a list of roots and their glosses in the first two columns. In the following columns we see the event denoting predicates formed from these roots through the use of voice morphology.

Table 24.4 Tagalog root meanings

	Root	Gloss	Actor voice form	Gloss
√WALK	lákad	'a walk, an errand'	l<um>ákad	'to walk'
√EAT	káin	'eating, a meal'	k<um>áin	'to eat'
√THINK	ísip	'thought, thinking'	mag-isip	'to think'
√KILL	patay	'corpse'	p<um>atay	'to kill'
√BREAK	básag	'a break'	b<um>ásag	'to break'
√TEACH	túrò	'lesson, teaching'	mag-túrò	'to teach'
√SAY	sábi	'what is said'	mag-sábi	'to say'
√BUY	bili	'price bought for'	b<um>ili	'to buy'
√TAKE	kúha	'taken object'	k<um>úha	'to take'

Most surprising is the interpretation of stems with the causative prefix, as exemplified in Table 24.5 (adapted from Schachter 1976: 105). Here, we find that even when embedded under a causative head, these roots maintain their patient-oriented interpretation.

Table 24.5 Tagalog causative meanings

	Causative stem	Gloss
CAUS - √HAND.OVER	pa-ábot	'something caused to be handed over'
CAUS - √BRING	pa-dala	'something caused to be brought'
CAUS - √MAKE	pa-gawà	'something caused to be made'
CAUS - √COOK	pa-lútò	'something caused to be cooked'
CAUS - √KEEP	pa-tágò	'something caused to be kept'

The patient-oriented nature of roots is projected to the clause level in predications, which is analyzed as essentially copular. Thus, even with a bare root predicate as in (51), the agentive argument surfaces with genitive case and the patient/theme argument surfaces with nominative/absolutive case.

(51) Háwak   ni   Jojo   ang   pera
      hold   GEN   Jojo   NOM   money
      'Jojo holds the money.'

Just as with fully inflected event-denoting predicates, extraction obeys the predicted pattern. The genitive argument is trapped within the domain of the nominal predicate and cannot be topicalized or otherwise extracted, as shown in (52).

(52) a. Ang pera   ay   háwak   ni   Jojo
          NOM money   TOP   hold   GEN   Jojo
          'The money, Jojo holds.'

      b. *Ni   Jojo   ay   háwak   ang   pera
          GEN   Jojo   TOP   hold   NOM   money
          (For, 'Jojo, holds the money.')

So what does nominalization add to an already nominal root? Primarily, it changes its reference to that of a potential participant. Again, we can compare English *-ee* in a word like *hoaxee* where it similarly attaches to a nominal stem to select a notional participant. Second, it also allows the root to combine with aspect inflection, which bestows an agentive reading to the genitive/ergative argument.[24] This is highly reminiscent of

---

[24] Aspect inflection is generally impossible without voice with one exception. Predicates inflected for the ~~immediate past~~ show no voice marking and assign genitive case to all arguments. This is probably related to the exclamative function of the ~~immediate past~~ in Tagalog, as discussed in Kaufman 2011.
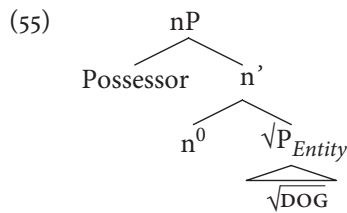
what has been described by Davis and Demirdache (2000: 100) for St'át'imcets, a Salish language.

In St'át'imcets, notionally bivalent roots surface as unaccusative predicates when used in their bare form. This is seen in (53), where the bare predicate *qam't* 'hit' assigns the patient role to the subject *ti sqáycw-a* 'the man'. Similar to Tagalog, an agent must be licensed by a voice-related morpheme, as seen in the contrast between (54a) and (b). In (54b), the suffix *-en* introduces an agent and creates a transitive predicate. For Davis and Demirdache (2000), roots like *qam't* and *máys* as well as "classic" unaccusative roots contain an underlying causer (following Chierchia 1989) but no agent. This accords relatively well with Tagalog bare root predicates in that they license causers but not true agents.[25]

(53) qam't   ti   sqáycw-a
      hit   DET   man-DET
      'The man was hit (with something thrown).'

(54) a. Mays ti   tsítcw-a
          built DET   house-DET
          'The house got built.'

      b. Máys-en   ti   tsítcw-a
          built-DIR-ERG   DET   house-DET
          '(They) built the house.'

Let us briefly consider one theory of how Tagalog predicates are built from the root up. The category determining head *n* introduces a referential index (in the sense of Baker 2003). When the inserted root is entity-denoting, as in (55), this is entirely straightforward. In (55), the merged n+√ stem simply takes on a referential index associated with an instance of ⟦dog⟧. But determining how the reference is set with an event-denoting root, as in (56), is more complex, as it is often a particular participant rather than the event which the bare stem denotes on the surface.

(55)

```
              nP
             /  \
    Possessor    n'
                /  \
              n⁰    √P_Entity
                     /\
                   √DOG
```

---

[25] The lack of an agentive reading with bare root predicates can be gleaned from (i), where the adverb *nang sindayà* 'intentionally' is ~~shown to be~~ infelicitous.

(i) Bigay   ko   ang   bulaklak   (?*nang   sinadyà)
     give   1S.GEN   NOM   flower   GEN   intentionally
     'The flower is my gift (?*intentionally).'

(56)

```
                    nP
                  /    \
          Possessor     n'
                      /    \
                    n⁰      √P_Event
                          /       \
                      Theme        √'
                                 /     \
                            √GIVE       Recipient
```
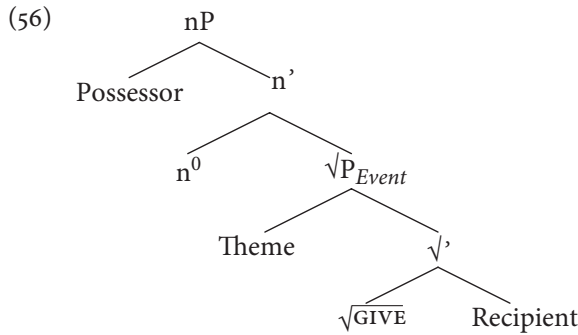
Despite lacking an argument structure, an event-denoting root projects associated thematic roles as Davidsonian event arguments (or "participants," in the terminology of Grimshaw 1990). In the case of $\sqrt{GIVE}$ this will include at least a theme and a recipient, as shown. Let us then posit that when $\sqrt{P}$ combines with n, the root raises to n and n probes downwards to select the nearest thematic role, which determines the ~~meaning~~ of the stem. This will naturally exclude an agentive interpretations of roots, as Agents are introduced by the higher functional projection, VoiceP (Kratzer 1996). It will also exclude possessor interpretations as the possessor role is projected to the specifier of nP and is thus outside the c-command domain of n. We thus have a mechanism for deriving a phrase such as that in (57), where the root *bigay* $\sqrt{GIVE}$ is identified with the theme of the event, i.e. a gift. The possessor and the recipient are expressed as genitive and oblique phrases, respectively.[26]

(57) Bigay   ni   Nonoy   kay   Neneng   ang   bulaklak
      give   GEN   Nonoy   OBL   Neneng   NOM   flower
      'The flower is a gift of Nonoy to Neneng'

The higher functional category VoiceP licenses an agent and hosts the voice marking in Voice⁰, which allows the predicate to be identified with one of several thematic roles via episodic linking (Barker 1998) as in participant nominalization. The AspP projection houses aspect morphology and certain aspectual adverbs. These two functional categories undergo morphological merger, as evidenced by several points of syncretism in the Voice/Aspect paradigm and by the fact that voice marking is a prerequisite for aspect marking. We see a fully inflected analogue of (57) in (58).

---

[26] Note that this predicts patient-orientation for all low nominalizations. This has been argued for independently by Salanova (2007) and is supported by the broad comparative study of Koptjevskaja-Tamm (1993). It may even be present in English bare nominalizations of bivalent roots (e.g. *take* as in 'The take was $500'), which seem to only refer to the patient and never the agent.

(58) I-b<in>igay   ni   Nonoy   kay   Neneng   ang   bulaklak
      CONV-<BEG>give   GEN   Nonoy   OBL   Neneng   NOM   flower
      'The flower was given by Nonoy to Neneng'

The peculiar meaning of causative stems seen earlier in Table 24.5 is also predicted here by the fact that the causative head lies beneath the voice head. This is corroborated by the order of prefixes: the conveyance voice prefix and actor voice prefix are always external to the causative marker: *i-pa-* CONV-CAUS-, *mag-pa-* ACT-CAUS-. The referential index is set by *n* and is not affected by further movement to the causative head, *pa-*. So while the causative projection introduces a causer, the causer is introduced too late to be identified with the meaning of the bare stem.

(59)

```
                    CausP
                   /    \
              Causer     Caus′
                        /    \
                     pa-      nP
                            /    \
                   Possessor      n′
                                /    \
                             n⁰       √P_Event
                                    /        \
                                Theme         √′
                                            /    \
                                        √GIVE     Recipient
```
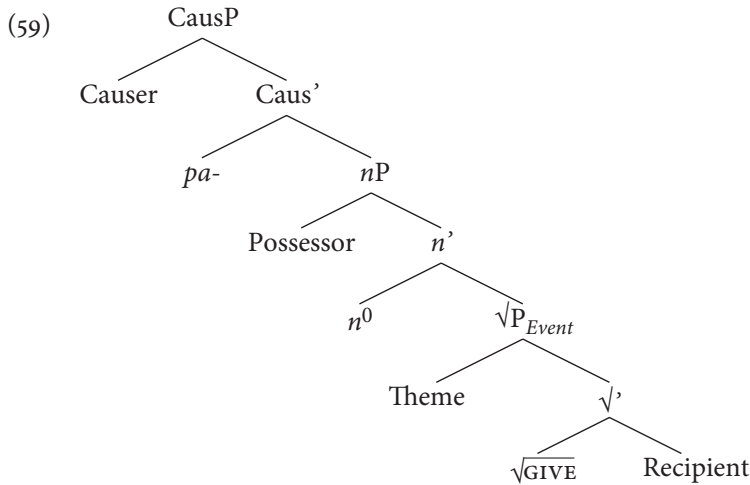
The structure in (59) underlies the predicate phrase in examples like (60).

(60) Pa-bigay   ni   Nonoy   ang   bulaklak
      CAUS-give   GEN   Nonoy   NOM   flower
      'The flowers are Nonoy's ~~to cause (someone) to give~~'

This section has shown that ergativity in Tagalog is present on the root level even before the attachment of voice morphology. This poses an additional challenge to both the case agreement and canonical ergativity approach reviewed above. Under those approaches it is left to explain why object shift should occur in the absence of a VP and, if it occurs, why it is not registered as agreement or transitivity marking on the predicate.

    With this brief introduction to Tagalog word structure, we are now ready to examine Mamuju, an Indonesian language which has maintained an ergative pattern while developing a strong noun-verb distinction.

24.2 MAMUJU AND THE RE-EMERGENCE OF *v*

In the following, we observe the basic morphosyntactic alternations of Mamuju, an Austronesian language of the South Sulawesi subgroup, to show the consequences of *v* on the syntax.[27] What differentiates Mamuju from Tagalog is the loss of the nominal properties of event denoting predicates and the development of a strong Noun/Verb contrast (Himmelmann 2005: 128-131). Two important consequences of this are that agreement and valency changing operations make reference to underlying argument structure and that constraints on ergative extraction are loosened.

Mamuju displays clear differences from Tagalog on the level of root and word. One simple distinction between nouns and verbs in Mamuju is that only verbs can be the complement of the tense marker *na* FUTURE. As seen in (61b), nominal predicates cannot follow *na*. In Tagalog, there are no tense/aspect markers whose distribution distinguishes root classes.

(61) a. na   menjari=aʔ   guru   jao   di   Udayana
          FUT   become=1SG.NOM   teacher   there   PREP   Udayana
          'I will be a teacher there at Udayana.'

      b. (*na)   guru=aʔ   jao   di   Udayana
          FUT   teacher=1SG.NOM   there   PREP   Udayana
          'I am a teacher there at Udayana.'

Conversely, verbal roots cannot combine directly with a possessor, unlike nouns, as shown in (62). An overt nominalizing head must combine with a verbal stem before prior to modification by a possessor. As seen earlier, this is not the case in Philippine-type languages.

(62) a. *langi-na
              swim-3.GEN

      b. bau-na
          fish-3.GEN
          'his/her fish'

If the symmetric nature of Philippine-type alignment systems are due to the nominal nature of the roots, we expect that the emergence of *v* will allow for bona fide

argument structure to be projected from the root and an absence of symmetry in the derivational morphology. Recall that in Tagalog, the notion of valency plays almost no role in determining a root's morphological potential. This is seen clearly in table 24.6, where a notionally monovalent root, *langoy* 'swim' is compared with a bivalent root *patay* 'kill.' Both roots can take the full range of voice morphology suggesting that they are essentially of the same type at the point where they merge with Voice.

Table 24.6 Tagalog word classes

	√LANGOY 'swim'		√PATAY 'kill'	
<um> ACTOR VOICE	l<um>angoy	'NOM to swim'	p<um>atay	'NOM to kill'
-in PATIENT VOICE	languy-in	'to swim to NOM'	patay-in	'to kill NOM'
-an LOCATIVE VOICE	languy-an	'to swim in NOM'	patay-an	'to kill from NOM'

This contrasts starkly with Mamuju and other South Sulawesi languages. In Table 24.7, we see that the cognate Mamuju roots are in complementary distribution with regard to their morphosyntactic potential. An unergative verb like *langi* 'swim,' requires prefixation with *mo* to form an intransitive predicate. In contrast, a bivalent root like *patei* 'kill,' cannot take the intransitive prefix but rather must be prefixed with antipassive *mang-* to enter into a intransitive predication. Bivalent stems must be prefixed with ergative agreement when forming transitive verbs, as seen with *ku-patei* 'I kill (X).' This agreement marking is impossible with monovalent roots like *langi*.

Table 24.7 Mamuju word classes

	√LANGI 'swim'		√PATEI 'kill'	
mo- ACTIVE	mo-langi	'ABS to swim'	*mo-patei	—
mang- ANTIPASSIVE	*man-langi	—	mam-patei	'ABS to kill'
ku- 1SG.ERG	*ku-langi	—	ku-patei	'ERG to kill ABS'

This development can be understood as part of the emergence of a robust *v* category which converts the thematic structure of an event-denoting root into actual argument structure. Note also that, unlike in Philippine-type languages, we now find an unmarked class of intransitive and transitive verbs. Unaccusatives, like *tama* 'enter' in (63), do not require derivational morphology when surfacing as intransitive verbs. Similarly, bivalent roots are only prefixed with ergative agreement when functioning as transitive

verbs, as seen in (64). All predicates can host second-position pronominal clitics reflecting the absolute argument.

(63) tama=do=ʔ   di   songi
      ENTER=ALREADY=1.ABS   PREP   room
      'I already entered the room'

(64) na-kita=ko
      3.ERG-see=2.ABS
      'S/he sees you.'

Note also the divergence in possessor and ergative agreement. The former is expressed via a set of suffixes and the latter through a set of verbal prefixes. In all Philippine-type languages, these two functions are expressed with the same set of pronouns, the typically second-position genitive clitics.

Importantly, Mamuju has a robust antipassive, which only combines with transitive predicates and is used to introduce indefinite objects, as can be seen in the comparisons in (65) and (66).[28] In the transitive clause, a missing patient argument would be interpreted as a null pronoun retrievable from discourse. The ergative argument typically follows the verb directly although scrambling may also occur to yield an ABS ERG order.

(65) a. na-kande   i   Husni   bau
          3.ERG-eat   ART   Husni   fish
          'Husni is eating the fish.'

      b. mang-kande (bau)   i   Husni
          ANTIPASS-eat fish   ART   Husni
          'Husni is eating (fish).'

(66) a. mu-kita=aʔ
          2.ERG-see=1.ABS
          'You see me.'

      b. mang-kita=aʔ   bau
          ANTIPASS-see=1.ABS   fish
          'I see a fish.'

In addition to the anti-passive, a limited number of verbs appear to take a "super-antipassive," which has also been described for the neighboring Seko Padang language by

---

[28] The antipassive *mang-* prefix derives historically from a distributive/pluractional infix *<ŋ> (Kaufman 2009c). The path from Proto-Malayo-Polynesian distributive marker to an antipassive marker in the South Sulawesi languages is straightforward. Just like antipassives, distributive verbs are bivalent but cannot take a specific or definite object. The Makassarese cognate *aN(N)-* is discussed by Jukes (2013).

Payne and Laskowske (1997). The super-antipassive also expresses the subject as an absolutive but does not allow for any objects whether definite or indefinite, as shown in (67).[29]

(67) k<um>ande=ko   (*bau)
      <SuperAntiPass>eat=2.ABS   fish
      'You eat.'

Mamuju also has a passive, marked on the verb with the prefix *ni-*, shown in (68b). Ergative agreement is obligatory on transitive verbs and thus the passive allows for backgrounded or impersonal agents.[30]

(68) a. Apa   na-kande   todapa?
          what   3.ERG-eat   people
          'What do the people eat?'

      b. Apa   ni-kande   (*todapa)?
          what   PASS-eat   people
          'What is being eaten?'

Mamuju has bona fide applicatives that promote benefactives and oblique arguments to absolutive. Furthermore, the division between low and high applicatives does not run into the difficulties noted earlier for Tagalog. In (69a), the applicative *-ang* attaches low and transitivizes an intransitive verb by introducing a theme, thereby feeding ergative agreement. When *-ang* merges as a high applicative, as in (69b), it attaches to a transitive stem (*pa-lamme*) and introduces a benefactive.

(69) a. ku-lamme-ang   buku
          1S.ERG-~~CAUS~~-fall-APPL~~=2S.ABS~~   book
          'I dropped a book.'

---

[29] The super-antipassive is more useful than it appears at first sight. With the loss of case marking on DPs, the difference between antipassive and super-antipassive is all that distinguishes a post-verbal object from a post-verbal subject in sentences like (i).

(i) a. mangapa   bongi   itte   ampe'   k<um>ande   bau?
         why/when   night   DEM   CONJ   <SUPERANTIPASS>eat   fish
         'What time did the fish eat last night?'

     b. mangapa   bongi   itte   ampe'   mangande   bau?
         why/when   night   DEM   CONJ   ANTIPASS:eat   fish
         'What time did he eat fish last night?'

[30] The existence of a passive in Philippine-type languages is a somewhat difficult question. Reid and Liao (2004) and Tanangkingsing and Huang (2007) claim that *ma-* verbs are essentially passives in Bontok and Cebuano, respectively, because of their inability to license agents. On the other hand, the *ma-* prefix indicates non-volitionality and is obligatory on certain unaccusative predicates like 'to fall',

b. ku-pa-lamme-ang=ko   buku
    1S.ERG-CAUS-fall-APPL=2S.ABS   book
    'I dropped a book for you.'

Although the data on double applicatives is still unclear for Mamuju, other languages of Sulawesi allow for promotion of multiple adjuncts to arguments through applicative stacking. Tukang Besi, for instance, allows the comitative and benefactive applicative combination shown in (70). Among the South Sulawesi languages, Selayarese allows double applicatives when both the theme and recipient of a ditransitive verb are definite, as shown in (71) and Sirk (1996: 82) discusses an identical construction for Bugis. Recall that no Philippine-type language allows double applicatives of this type (on the analysis of the voice/nominalization morphology as applicative marking).[31]

(70) No-wila-**ngkene**-**ako**   te   ina-no   te   Wa   Ki'i.
     3.RL-go-COM-APPL   CORE   mother-3POSS   CORE   Wa   Ki'i
     'She went with Wa Ki'i for her mother.'   (Donohue 1999: 248)

(71) Ku-kiring-**i**-**ang**=ko   doe'-injo
     1S.ERG-send-APPL-APPL=2.ABS   money-DEF
     'I sent you the money.'   (Basri 1999: 313)

As shown in (72b), applicatives are incompatible with the antipassive. Aldridge (2012b) predicts this behavior via the requirement that applied objects require structural case, which is systematically lacking in antipassives.[32]

---

which makes it look quite different from a traditional passive. (See Kaufman 2012 for a discussion of the historical development of *ma-* verbs). I thank Laurie Reid for bringing this point to my attention.

[31] An apparent counterexample is found in the Cordilleran languages that employ the circumfix *i- -an* for the benefactive. While these were historically two morphemes (**i-* CONVEYANCE and **-an* LOCATIVE), they have been reanalyzed as one. This is clear from the syntax of benefactives in the relevant languages, in which only one argument, the benefactive, is "promoted" to the nominative/absolutive.

[32] Note that in Bajau, and potentially other Austronesian languages, antipassives can both host definite objects, as seen in (i), and freely allow combinations of actor voice and applicatives, as shown in (ii). The prefix must still be considered antipassive as it only attaches to transitive stems to make an absolutive argument from the agent.

(i) Nga-daka'   manu'   iru   aku   pugay   uwa'-ku.
    ANTIPASS-catch   chicken   that   1S   do.for   father-1S.GEN
    'I caught that chicken for my father.'

(ii) Nga-daka-an   uwa'-ku   manu'   iru   aku.
    ANTIPASS-catch   father-1S.GEN   chicken   that   1S
    'I caught that chicken for my father.'   (Donohue 1996: 789)

Bajau thus represents an intermediate stage between Mamuju and modern Indonesian, where the cognate *meng-* prefix has extended its domain to include many monadic verbs as well (e.g. Indonesian *menangis* /meN-tangis/ 'cry', *menyala* /meN-nyala/ 'to be on/alight').

(72) a. Mam-baca=aʔ   buku   bua   Husni
          ANTIPASS-read=1.ABS   book   for   Husni
          'I read a book for Husni.'

      b. *Mam-baca-ang=aʔ   buku   Husni
          ANTIPASS-read-APPL=1.ABS book   Husni
          (For, 'I read Husni a book.')

Antipassives are, however, allowed to combine freely with applicatives when the agent is extracted, as in (73). This is the same condition under which definite patients are possible for actor voice predicates in Philippine-type languages.

(73) Sema mam-baca-ang   Husni   buku?
      who ANTIPASS-read-APP   Husni   book
      'Who read Husni a book?'

The Mamuju data observed thus far suggests that a verb's underlying argument structure largely determines its morphosyntactic potential. The *v* category combines with event-denoting roots and projects a basic argument structure from the thematic structure. This structure can then be adjusted by transitivizing and detransitivizing morphology. In Philippine-type languages, to the extent that we can speak of a fixed argument structure, it is created by the nominalizing/voice morphology rather than inherited from the root.

I argued earlier that the simplest account of extraction constraints in Philippine-type languages unifies the patterns found across event-denoting and entity-denoting predicates. In Mamuju, where event-denoting predicates no longer have any nominal properties at all, we would expect a loosening of these extraction constraints. This expectation is borne out for topicalization, as shown in (74). Here, the ergative argument is extracted to a pre-verbal topic position, which was shown earlier in (8) to be generally ungrammatical in Tagalog.[33]

(74) Baco na-patei   Ali
      Baco 3.ERG-kill   Ali
      'Ali killed Baco.' OR 'Baco killed Ali.'

However, relativization and question formation are still restricted in the usual way, as seen in (75), which only has a single interpretation.

[33] Edith Aldridge (p.c.) points out that the extraction might be expected if *na-* functions as a resumptive pronominal clitic. From a morphological perspective, ergative person marking in Mamuju appears to be standard agreement in that it is obligatory, attaches consistently to the left edge of the verbal stem and can co-occur with a co-referential pronominal argument. Nonetheless, it is possible that agreement plays a role in licensing topicalization, especially given that antipassive objects, which do not trigger agreement, are more highly constrained. See Kaufman (2008) for details.

(75) Sema na-kita   Ali?
      who 3.ERG-see   Ali
      'Who did Ali see?' NOT 'Who saw Ali?'

It would seem then that extraction constraints can, historically speaking, outlast the nominal features of the predicate. Mamuju would seem to be an ideal candidate for an ergative analysis along the lines of Aldridge (2004) and would not pose any of the empirical hurdles found with Philippine-type languages enumerated in (11). The *mang-* prefix would be a true antipassive while the formatives *me-, mo-* and *mu-* could be treated as instantiations of intransitive *v*. Unlike as in Philippine languages, polyvalent roots would give rise to unmarked transitive verbs and the suffixes *-i* and *-ang* behave like true applicatives attaching to an unmarked transitive verb and, with agent extraction, to an antipassive. Nonetheless, I would argue that the restrictions on extraction in Mamuju obtain the best explanation as historical residue from Philippine-type morphosyntax.

Recall that two potentially independent factors account for extraction restrictions in Philippine-type languages: the constraints on extraction from NP and the ban on genitive predicates. As shown earlier in detail, Mamuju has developed a strong N/V distinction and thus constraints on extraction from NP have become irrelevant for ergative extraction. This explains why ergatives can be freely topicalized to the preverbal position as in (74). The second factor, however, remains firmly in place as seen in (76), which shows the only way a would-be genitive argument can be extracted. The "dummy" predicate *ampunna* 'owner' allows a possessor to be expressed as an absolutive argument.

(76) Sema   ampunna   ku'bur   itte   di   bao   di
      who   owner   grave   that   PREP   down   PREP
      Timbu me-loda   batu?
      Timbu AV.HAVE-roof   stone
      'Whose is that grave down in Timbu with the stone roof?'
                                              (Stromme 1991:Maradika Lasalaga)

Although there is no overt morphological case marking of DPs in any of the languages of the South Sulawesi subgroup, we find that the interrogative and relative marking elements themselves bear unambiguous traces of absolutive case. The interrogative pronoun *sema* 'who,' which is plausibly derived from Proto-Austronesian **si-ima* (Blust et al. 2010), contains a reflex of the personal nominative/absolutive case marker **si-*.[34]

---

[34] The presence of Proto-Malayo-Polynesian nominative/absolutive **si* can also be found in Indonesian/Malay *siapa* 'who,' among many other Austronesian languages. See Ross (2006) for a detailed reconstruction of the relevant Austronesian case markers. An alternative source of Mamuju *sema* is a hypothetical *sai-ma* 'who-RELT.' Blust et al. (2010) reconstructs **sai* 'who' for Proto-Malayo-Polynesian and *ma* is found in Bajau and Kambera as a relativizer (although not attested for Mamuju). The nominative case element on *sai* is not as easy to isolate but given the reconstructions of the nominative singular proper name marker **si* and its plural counterpart **sa*, it is not far-fetched to also relate the *s(a)* of *sai* to the nominative case function.

Crucially, no genitive/ergative variant (which we would expect to come out as *nema*) exists in Mamuju (nor in any of the South Sulawesi languages). Recall that genitive marked interrogative pronouns do exist in Philippine-type languages but can only be used with in-situ interrogative phrases, as seen earlier in (40), repeated here in (77).

(77) G<in>awà-∅   nino   ang   sapatos   na   iyon?
      <BEG>make-PAT   GEN.who   NOM   shoe   LNK   that
      'Who made those shoes?'   (Schachter and Otanes 1982: 512)

Similarly, the Mamuju relative marker *anu* is clearly cognate with the Tagalog nominative interrogative *ano* (underlyingly also /anu/). As the interrogatives themselves show morphological traces of absolute case, it follows that only absolute arguments can be clefted given case preservation.[35] The proper morphological analysis then of interrogative sentences such as (75) is shown in the glossing of (78), where the case marked interrogative constrains the types of arguments that can be extracted. Similarly, the grammatical voice of verbs in relative clauses is constrained by the case of the relativizer, as exemplified in (79).

(78) **Sema**   na-kita   Ali?
      ABS.who   3.ERG-see   Ali
      'Who did Ali see?' NOT 'Who saw Ali?'

(79) Ku-pe-ingarangng-i=mo   **apa**   **anu**   ku-so'na
      1.ERG-TR-remember-APP=EMPH   ABS.what   ABS.RELT   1.ERG-dream
      sambongi.
      last.night
      'I remember all that I dreamt last night.'   (Stromme 1991:Alibe Niso'na)

This analysis, although syntactically trivial, offers a morphologically well-founded basis for the difference between relativizations and question formation on one hand, which both require case marked operators, and topicalization, which does not. As we have seen, the former operations are just as constrained in Mamuju as they are in Philippine-type languages while the latter operation is freer, as constraints against extraction from NP are no longer relevant for event-denoting predicates.[36]

---

[35] The morphological argument is less clear with 'what' than 'who' as nominative 'what' can be treated as unmarked. Compare Tagalog *ano* 'what,' *ng ano* 'GEN what,' *saan* 'OBL.what/where.' However, the *a-* initial in the Mamuju relative marker *anu* and interrogative *apa* 'what' is found in a wide range of nominative/absolutive interrogatives throughout Philippine-type and Formosan languages. In Tagalog, we can compare the argument interrogatives *a-no, a-lin* 'which' and *si-no* 'NOM.who' which all begin with the putative nominative markers *a-* or *si-* (for proper names) to the adjunct interrogatives *ilan* 'how many,' *kailan* 'when,' *bakit* 'why,' *magkano* 'how much' none of which begin with *a-*. Ross (2006) presents the comparative data but tentatively reconstructs the **a-* formative with demonstrative features rather than case features.

[36] Note however that this approach requires a view of the grammar that attributes more power to the lexicon than is currently popular in generative theorizing. In particular, a lexical gap in the case paradigm for certain operators must be able to block syntactic movements. This would not be

Note that this analysis extends beyond Mamuju and can account for a similar pattern in Indonesian, a language which is generally considered to have adopted a nominative-accusative alignment pattern (Chung 1976, 2008; Cole et al. 2008; Aldridge 2008a). As can be seen in (80), actor voice objects can be topicalized but not relativized in Indonesian.

(80) a. Orang   itu,   saya   meng-ajak   ke   sini.
          person   that   1SG   ACT-invite   to   here
          'As for that person, I invited (him/her) here.'

      b. *Orang   itu   yang   saya   meng-ajak   ke   sini.
           person   that   RELT   1SG   ACT-invite   to   here
           (For, 'That person whom I invited here.'   (Arka and Manning 2008: 53))

Again, we can plausibly attribute (80b) to the relativizer, which Adelaar (1992) reconstructs as *ia-ng* 3S.NOM-LNK. Indeed, *ia* functions as a strictly nominative case pronoun in ~~the~~ modern Indonesian as well. Compare in (81) the distribution of nominative *ia* with *dia*, a case neutral pronoun which can function as either subject or object of an actor voice clause (Musgrave 2001a).

(81) Dia/ia   me-lihat   dia/*ia
      3S.NEUT/3S.NOM   ACT-see   3S.NEUT/3S.NOM
      'He sees him.'

Were Indonesian to have in its functional inventory a case neutral relativizer (hypothetically *diang*), we might expect that extraction could take place from a wider array of syntactic positions. The prediction here is that Austronesian's well-known "subjects only" condition on extraction can easily dissipate once the nominal features of event-denoting predication are lost, as it is then "relic" properties of functional items that keep the restriction in place rather than the strong island characteristics of Philippine-type predicates.

  While constraints on ergative extraction are relatively tenacious among Indonesian languages, we do find several ergative languages that allow ergative extraction in a manner unknown among Philippine-type languages. This is exemplified here by Sumbawa Besar (as described by Shiohara 2013) and Selayarese (as described by Basri and Finer 1987; Basri 1999).[37] The transitive verbal clause in Sumbawa employs an unmarked verb

countenanced in a framework such as Distributed Morphology where the lexicon is reduced to almost nothing and the putative lexical gap relied upon above would itself have to be derived synchronically in the syntax. On the other hand, it is a perfectly natural analysis in Combinatory Category Grammar (Steedman 2001) where Mamuju *anu* would be of the category: $(N\backslash N)/(S/NP_{ABS})$, that is, a predicate which seeks a clause with a nominative argument gap to its right and a noun to its left to yield another noun via predicate modification.

  [37] See also Erlewine, Levin, and Van Urk, Chapter 16, this volume, who discuss Balinese pre-verbal subjects in object voice.

stem with prefixal ergative agreement and introduces the ergative argument with the case marker *léng*, as seen in (82).[38]

(82) ka=ku=inóm   kawa   nan   léng   aku
      PAST=1SG=drink   coffee   that   by   1SG
      'I drank the coffee.'   (Shiohara 2013: 132)

Antipassives are formed with the nasal prefix *N-* and subject agreement (which is similar, but not identical to ergative agreement). As shown in (84), antipassives do not allow for objects.

(83) a. ka=ku=ng-inóm   aku.
          PAST=1SG=ANTIPASS-drink   1SG
          'I drink (something).'

      b. *ku=ng-inóm   kawa   (nan)
          1SG=ANTIPASS-drink   coffee (that)
          (For, 'I will drink (the) coffee.')
                                                         (Shiohara 2013: 132)

Crucially, the ~~restriction against~~ antipassive objects (much like Mamuju's super-antipassive in (67)) appears to be indefeasible and thus trumps the restriction against extraction of the ergative argument. The result is that agent extraction proceeds from a transitive clause as shown in (84a) in the presence of an object. The antipassive is used in subject questions only when no object is present, as in (84b).

(84) a. sai adè ka=tumpan'   jangan=ta
          who NOM PAST=get   fish=this
          'Who caught the fish?'

      b. sai   adè   ka=n-umpan'?
          who   NOM   PAST=ANTIPASS-get
          'Who already had a catch (in fishing)?'
                                                         (Shiohara 2013: 135)

Topicalization of the ergative argument to a preverbal position is also possible, as expected, although interestingly, Shiohara shows that the ergative marker must be omitted in this construction, as shown in (85) (see n. 38).

[38] The pattern of a *leng*-phrase agent triggering prefixal agreement appears very similar on the surface to that found in Acehnese (Durie 1985; Legate 2012b, 2014b). However, while Sumbawa clearly follows an ergative pattern, Acehnese is typically described as having an active-stative alignment type which Legate (2012b, 2014b) derives from properties of *v*P. Shiohara (2013) in fact glosses *leng* as 'by' but because *leng* seems obligatory on external arguments of transitive verbs, I re-gloss *leng* as ERG. Shibatani (2008) argues that there is a significant difference between the verb in agent extraction, as in (84a), and a canonical declarative like (82). Space does not allow a more in depth discussion of the complex Sumbawa facts.

(85) (*leng)   aku (ku=)inóm   kawa   nan.
      ERG   1SG 1SG=drink   coffee   that
      'I drink the coffee.'   (Shiohara 2013: 137)

Selayarese presents a similar state of affairs. The alternation in (86) shows that person marking and alignment are identical to Mamuju in the simple case. A definite object requires a transitive clause with ergative agreement and a second-position absolutive clitic. An indefinite object must be introduced with an antipassive/intransitive verb.

(86) a. Ku-halli'=i   sapo=ɲjo
          1.ERG-buy=3.ABS   house=DEF
          'I bought the house.'

      b. M-mali=a   sapo
          INTR-buy=1.ABS   house
          'I bought a house.'   (Mithun 1991a: 175)

As with Sumbawa and Mamuju, the external argument of both a transitive and intransitive predicate can be topicalized, as shown in (87).

(87) a. I   Baso'   la-alle=i   doe'=ɲjo
          PM   Baso   3.ERG-take=3.ABS   money=DEF
          'Baso took the money.'

      b. I   Baso'   (a)ng-alle(=i)   doe'
          PM   Baso   INTR-take=3.ABS   money
          'Baso took some money.'   (Finer 1994: 159)

Like Sumbawa, but unlike Mamuju, this extends to cases of agent extraction in transitive clauses, as seen in (88), where the verb still takes ergative agreement rather than antipassive/intransitive morphology. Just like in the dialect of Sumbawa described by Shiohara (2013), this is the only option when the object is definite.[39]

(88) Inai   la-sumbele=i   tedong=injo?
      who   3.ERG-slaughter=3.ABS   buffalo=DEF
      'Who slaughtered the buffalo?'   (Hasan Basri 2006 p.c.)

---

[39] Note that the extraction of antipassive objects in Selayarese is still restricted just as in more morphosyntactically conservative languages. Recall that in Indonesian, where the cognate prefix *meN*- marks transitive active voice verbs, topicalization (but not relativization) of the object is permissible, as shown earlier in (80).

(i)   *Doe   (a)ng-alle=i   i   Baso'
      money   INTR-take=3.ABS   PM   Baso
      (For, 'Money, Baso took.')   (Finer 1994)

This also extends to relative clauses, as seen in (89), which are optionally introduced by the relative marker *nu*.

(89) a. asu   (nu)   n-datalaʔ   bahi
          dog   RELT   INTR-chase   pig
          'the dog that chased a pig.'

      b. asu   (nu)   la-datala-ɲjo=i   bahi=ɲjo
          dog   RELT   3.ERG-chase-DEF=3.ABS   pig=DEF
          'the dog that chased the pig.'

                                                    (adapted from Basri 1999: 292)

While this diachronic analysis of case marked relativizers and interrogative operators can extend to many Austronesian languages transparently there are inevitably exceptions. Bajau as described by Donohue (1996), for instance, shows the classic extraction restrictions but has a relativizer *ma* that alternates with ∅. Bajau *ma* is homophonous with the oblique case marker/preposition rather than an absolutive/nominative one and yet it is only the absolutive argument which can be relativized. The converse problem is found in Malagasy, where a focus marker which almost certainly also derives from PMP *anu allows for clefting of prepositional phrases. (Note also that the Selayarese relativizer in (89) derives from *anu* yet allows relativization of the ergative argument.)

(90) T-amin-ny   antsy   **no**   nanapaka   ity   hazo   ity   Sahondra
      PST-with-DET   knife   FOC   PST.ACT.cut   this   tree   this   Sahondra
      'It was with the knife that Sahondra cut this tree.'   (Paul 2001)

Clearly, the etymology of relativizers and interrogatives is only suggestive of the solution proposed here. We cannot expect that the features and structures involved in questions and relative clauses will not diverge from their etymological roots. Nor should we expect that the features of a relative marker or interrogative may not be inherited by a lexical item that comes to replace it.

24.3 CONCLUSION

I have argued here that lexical categories play a critical role in accounting for the difference between Philippine-type language and related ergative languages of Indonesia. In particular, symmetrical voice systems in which every event-denoting predicate is marked with voice morphology emerge from ~~non-verbal predicates~~. True argument structure with unmarked intransitive and transitive event-denoting predicates are rooted in verbal categories. Mamuju was presented as a canonical ergative language with a highly developed N/V contrast. As a result, it displays canonical antipassives and

applicatives of a kind not found in Philippine-type languages. When event-denoting predicates lose their nominal properties and become verbalized, certain syntactic properties of ergative alignment (e.g. extraction constraints) are more likely to be lost as predicted by Manning's (1996: 21) hypothesis that "syntactic ergativity" results from nominalization, in contrast to more surface-oriented "morphological ergativity." Further research in the comparative syntax of Indonesian languages is necessary to better understand the correlates between alignment-type, lexical category and extractability. The bewildering assortment of agreement and argument marking patterns (see, for instance, Kikusawa (Chapter 23, this volume) as well as the papers in Adelaar and Himmelmann (2005), Wouk and Ross (2002), Arka and Ross (2005) and Adelaar (2013) for a sample) will likely require decades to fully sort out. One of the purposes of this chapter has been to suggest new potential correlations for further investigation as we expand our empirical scope to the many under-described Austronesian languages of Indonesia and beyond.

Abbreviations

BEG, begun aspect (a component of both the perfective and progressive); APPL, applicative; ABS, absolute; NOM, nominative; GEN, genitive; ERG, ergative; TR, transitivity marker; PREP, general preposition; TOP, topic marker; ANTIPASS, antipassive; SUPERANTIPASS, super antipassive; IMPRF, imperfective; DAT, dative; OBL, oblique; PASS.PART, passive participle; LNK, linker; DET, determiner; FUT, future; ART, article; PASS, passive; DEM, demonstrative; CONJ, conjunction; CAUS, causative; VRB, verbalizer; EMPH, emphatic particle; NVOL, non-volitional.