Nouns, verbs, and hidden structure in Tagalog

NORVIN RICHARDS

Kaufman's article is rich in interesting facts about Tagalog, many of them gathered from underexplored corners of the language. Space constraints prevent me from discussing at length many of his insights; I will instead devote this space to some concerns I have with some of his conclusions.

Consider the Tagalog sentences in (1)[1]:

(1) a. Nag-ingay   ang aso
         AV.BEG-noise   NOM dog
         'The dog made noise'
     b. Aso ang nag-ingay
         dog NOM AV.BEG-noise
         'The one that made noise was a dog'

As Kaufman notes, the examples in (1) contain the same three words in different orders. Both *aso* 'dog' and *nag-ingay* 'make noise' appear to be capable of being either the subject or the predicate of the sentence.

On one type of theory, the availability of the word orders in (1) constitutes evidence that Tagalog does not distinguish between as many kinds of lexical categories as English does, or perhaps that functional structure can select for more lexical categories in Tagalog than it can in English. On this kind of account, there are held to be no selectional differences between nouns and verbs in Tagalog, perhaps because the distinction

---

[1] I have retained Kaufman's conventions for glosses, but have reverted to regular Tagalog orthography, removing indications of stress, cliticization, and infixation.

between nouns and verbs does not exist in this language; the words *nag-ingay* 'made noise' and *aso* 'dog' are the same kind of word, and the fact that each can occupy the other's position in sentences like the ones in (1) is unsurprising. Kaufman endorses a version of this theory.

Another type of approach posits phonologically null structure in (1b), giving the Tagalog sentence a structure not unlike that of its English translation. On this kind of theory, the subject of (1b) is actually a null nominal head, modified by a relative clause containing the predicate *nag-ingay* 'made noise'. Such a theory might also posit a null copula in (1b).

In section 1, I will offer arguments for distinguishing between lexical categories in Tagalog, and for positing phonologically null structure in examples like (1b). In section 2, I will address Kaufman's arguments against such a theory. Finally, in section 3, I will consider some of Kaufman's more general claims about the language.

1.   Tagalog nouns and verbs

As mentioned above, a theory which distinguishes between nouns and verbs in Tagalog might posit several kinds of null structure in an example like (1b), including a null copula and a null nominal head to serve as the subject. In this section I will offer arguments for each of these kinds of null structure.

Examples like the ones in (1) and (2) seem to support the idea that Tagalog lacks distinctions between classes of lexical items. Any lexical category may apparently be a predicate in this language (for reasons of space, I limit discussion to apparent nouns and verbs):

(2)   a.   Lumalangoy   ako
            AT.BEG.IMPF-swim 1S.NOM
            'I'm swimming'
      b.   Doktor ako
            doctor 1S.NOM
            'I'm a doctor'

In fact, there are contexts in which the behavior of verbal and non-verbal predicates diverge. For instance, in infinitival clauses, verbal predicates

are simply put into the infinitival form, while non-verbal predicates must acquire a verbal copula *maging*[2]:

(3) a. Ayoko   na-ng   lumangoy
       don't.want-1s.GEN now-LNK AT.INF-swim
       'I don't want to swim any more'

   b. Ayoko   na-ng   *(**maging**) doktor
       don't.want-1s.GEN now-LNK AT.INF-be doctor
       'I don't want to be a doctor any more'

Here is a context in which Tagalog predicates are required to contain a verb: if the predicate is non-verbal, a copular verb *maging* is introduced. In other words, a complete description of Tagalog grammar must be able to make reference to the category 'verb'; verbs and nouns cannot be identical.

   In fact, we can find evidence that even examples like (2b), which have no overt copula, contain a null copula. In dialects of English with null copulas (such as African American Vernacular English), null copulas are subject to a well-known restriction; they cannot be immediately followed by an extraction site (Wolfram 1969):

(4) a.   You Ø beautiful.
   b. * How beautiful you Ø!

As Wolfram notes, the same constraint holds of the contracted copula in Standard English:

(5) a.   You**'re** beautiful.
   b. * How beautiful you**'re**!

---

2   The verb *maging* can also mean 'become':

(i) Naging   doktor ako   noong   1989
     AT.BEG-be doctor 1s.NOM when-LNK 1989
     'I became a doctor in 1989'

The examples in (3) are constructed to exclude this meaning; (3b), for instance, means that I currently am a doctor, and want to stop being one. See Richards (2009) for further discussion.

Thus, there is a class of objects, including null copulas in AAVE and contracted copulas in Standard English, which cannot be followed by an extraction site. The same is true of the Tagalog null copula:

(6) a.   Ano-ng   klase-ng   doktor ang sinabi   ng tatay niya
          what-LNK kind-LNK doctor NOM PT.BEG-say GEN father 3S.GEN
          na   gusto niya-ng   **maging __**?
          LNK want 3S.GEN-LNK AT.INF-be
          'What kind of doctor did her father say she wanted to be?'
     b. * Ano-ng   klase-ng   doktor ang sinabi   ng tatay niya
          what-LNK kind-LNK doctor NOM PT.BEG-say GEN father 3S.GEN
          na   **Ø __**   siya?
          LNK   3S.NOM
          'What kind of doctor did her father say she is?'

In (6a), the most deeply embedded predicate, *anong klaseng doktor* 'what kind of doctor', begins the derivation in a position preceded by the overt copula *maging*, and it may undergo wh-extraction. In (6b), by contrast, this predicate is preceded by a null copula, and is impossible to extract[3]. We can assimilate these facts to the AAVE facts above, as long as we are willing to posit null copulas in Tagalog.

A potential problem for this approach arises in examples like (7):

(7) Ano-ng   klase-ng   doktor siya?
     what-LNK kind-LNK   doctor 3S.NOM
     'What kind of doctor is she?'

Here the copula is null, and yet the predicate can be wh-extracted. Because extraction does not cross clause boundaries, however, it is difficult to be certain that overt wh-movement has taken place at all; the predicate might simply be in situ. If we consider versions of (7) with an overt copula, we find evidence that this is the correct analysis:

(8) **Naging**   ano-ng   klase-ng   doktor siya?
     AT.BEG-be   what-LNK kind-LNK doctor 3S.NOM
     'What kind of doctor did she become?'

---

[3]   The problem with (6b) is unrelated to the presence of the second-position clitic *siya* at the end of the sentence; replacing this clitic with an ordinary DP like *si Juan* 'NOM Juan' does not improve the sentence.

If the null copula in (7) is in the same position as its overt counterpart *naging* in (8), then it is not followed by an extraction site, and the general condition on null copulas is satisfied.

I have tried so far to argue that Tagalog examples like (1b) (repeated below as (9b)) contain a null copula, and that Tagalog's apparent freedom to use any kind of lexical category as a predicate is in fact only apparent:

(9)   a.   Nag-ingay   ang aso
            AV.BEGIN-noise NOM dog
            'The dog made noise'

      b.   Aso ang nag-ingay
            dog NOM AV.BEG-noise
            'The one that made noise was a dog'

If the reasoning outlined above is right, then the noun *aso* can only be a predicate with the help of a verbal copula, though this copula is often phonologically null. On this theory, Tagalog is like English in requiring copulas with non-verbal predicates, and therefore makes use of a distinction between verbs and non-verbs.

Let us turn to the second type of null structure which is commonly posited in an example like (9b). Linguists who argue that Tagalog distinguishes between nouns and verbs typically claim that the subject of (9b) is a null nominal head, modified by a relative clause containing the verb *nag-ingay* 'made noise'.

To see the arguments for this conclusion, consider the pair of examples in (10):

(10) a.   Sumayaw   diyan ang mga pinsan ko
            AV.BEG-dance there NOM PL   cousin 1S.GEN
            'My cousins danced there'

      b.   Iyo-ng dalawa-ng sumayaw   diyan ang mga   pinsan
            that-LNK two-LNK   AV.BEG-dance there NOM PL   cousin
            ko
            1S.GEN
            'Those two who danced over there are my cousins'

Kaufman intends this pair of examples to show that the word *sumayaw* 'danced' is capable of combining with material associated with nominal projections, like demonstratives and number; the intent is support his

claim that Tagalog draws no distinction between nouns and verbs. In his footnote 34, he notes that the two sentences differ in meaning, but that he is "not aware of any evidence for substantial differences in their underlying structure."

If we are to distinguish between nouns and verbs in Tagalog, by contrast, we will have to posit different structures for the sentences in (10). On this theory, *sumayaw* 'danced' in (10a) is an ordinary verbal predicate; in (10b), by contrast, *sumayaw* 'danced' is contained in a relative clause, which modifies a null nominal head that is also modified by *iyong* 'those' and *dalawang* 'two'. This theory posits a clause boundary which is absent in Kaufman's theory; (10b) is biclausal, while (10a) consists only of the matrix clause.

Once piece of evidence that examples like (10b) are biclausal involves wh-extraction. The locative expression *diyan* 'there' may be converted to a wh-phrase and extracted in (10a):

(11)   Saan sumayaw   ang mga pinsan mo?
       where AV.BEG-dance NOM PL   cousin 2S.GEN
       'Where did your cousins dance?'

No such wh-extraction is possible in (10b), even if we change the example slightly to make the wh-question more pragmatically plausible:

(12)   a.   *Saan [iyo-ng   dalawa-ng $\emptyset_N$ [sumayaw   ]]ang mga
                 where that-LNK two-LNK   AV.BEG-dance NOM PL
                 pinsan mo?
                 cousin 2S.GEN
                 'Where were those two who danced your cousins?'
       b.   *Saan [ang dalawa-ng $\emptyset_N$ [sumayaw   ]]ang mga
                 where NOM two-LNK   AV.BEG-dance NOM PL
                 pinsan mo?
                 cousin 2S.GEN
                 'Where were the two who danced your cousins?'

On the account which posits a headless relative as a predicate for examples like (10b), the ill-formedness of the examples in (12) follows from the same conditions on islands that rule out the English translations in (12); *saan* 'where' cannot be extracted from a relative clause. If the examples in (10) are to be structurally identical, it is hard to see how this contrast can be made to follow.

A second argument for distinguishing between the two predicates in (10) has to do with the conditions on copulas described earlier in this section. We saw there that nominal predicates require an overt copula *maging* in infinitival clauses, while verbal predicates do not (and in fact cannot combine with *maging*). By this test, the predicate in (10a) acts like a verb, while (10b) behaves as though it has a non-verbal predicate[4]:

(13) a. Ayaw   nami-ng
　　　don't.want 1PLEXCL.GEN-LNK
　　　[(***maging**) sumayaw   diyan]
　　　　AV.INF-be AV.INF-dance there
　　　'We don't want to dance there'

　　 b. Ayaw   nami-ng
　　　don't.want 1PL.EXCL.GEN-LNK
　　　[*(**maging**) iyo-ng dalawa-ng sumayaw   diyan]
　　　　AV.INF-be that-LNK two-LNK AV.BEG-dance there
　　　'We don't want to be those two who danced there'

A third argument for a structural difference between the examples in (10) has to do with the distribution of negation. If (10b) is biclausal, then there should be two positions for structural negation, with corresponding differences in meaning. This is the case:

(14) a. **Hindi** iyo-ng   dalawa-ng [sumayaw   diyan]ang mga
　　　NEG   that-LNK two-LNK   AV.BEG-dance there NOM PL
　　　pinsan ko
　　　cousin 1S.GEN
　　　'It's not the case that those two who danced over there are my cousins'

　　 b. Iyo-ng   dalawa-ng [**hindi** sumayaw   diyan]ang mga
　　　that-LNK two-LNK   NEG AV.BEG-dance there NOM PL
　　　pinsan ko
　　　cousin 1S.GEN
　　　'Those two who didn't dance over there are my cousins'

---

4 (13b) without *maging* does have an irrelevant reading, "We don't want those two who danced there". On this reading, the complement of *maging* is the DP *iyong dalawang sumayaw diyan* 'those two who danced there', rather than an infinitival clause.

Since (10a) is monoclausal, there is only one possible position for clausal negation:

(15)   **Hindi** sumayaw   diyan ang mga pinsan ko
       NEG   AV.BEG-dance there NOM PL   cousin 1S.GEN
       'It's not the case that my cousins danced there'

   A fourth argument for a structural difference between the examples in (10) has to do with clitic placement. It will be easiest to demonstrate the argument by modifying the examples in (10) as in (16), allowing us to insert the second-position clitic *niya* '3S.GEN':

(16)   a.   Nakita   *niya*   ang mga pinsan   ko
            PV.BEG-see 3S.GEN NOM PL   cousin   1S.GEN
            'He saw my cousins'
       b.   Iyo-ng   dalawa-ng [nakita   *niya*]   ang mga   pinsan
            that-LNK two-LNK   PV.BEG-see 3S.GEN   NOM PL   cousin
            ko
            1S.GEN
            'Those two that he saw are my cousins'

In both of the examples in (16), *niya* attaches to the first word of its clause. In (16a), this is the first word of the sentence, since the sentence is monoclausal; in (16b), it is the first word of the relative clause *nakita niya* 'that he saw'. We can investigate more closely the domains in which these clitics attach by adding negation to the sentences, providing them with more potential clitic hosts. Since, as we just saw, negation can appear in two different places in examples like (16b), we have three examples to consider:

(17)   a.   **Hindi** *niya*   nakita   ang mga   pinsan   ko
            NEG   3S.GEN PV.BEG-see NOM PL   cousin   1S.GEN
            'He didn't see my cousins'
       b.   Iyo-ng   dalawa-ng [**hindi** *niya*   nakita]   ang   mga
            that-LNK two-LNK   NEG   3S.GEN PV.BEG-see NOM   PL
            pinsan ko
            cousin 1S.GEN
            'Those two that he didn't see are my cousins'

c. **Hindi** iyo-ng   dalawa-ng [nakita   *niya*]   ang mga
   NEG   that-LNK two-LNK   PV.BEG-see 3S.GEN NOM PL
   pinsan ko
   cousin 1S.GEN
   'It's not the case that those two that he saw are my cousins'

In the examples in (17), *niya* attaches to the first word of its clause. (17a), again, is monoclausal, and the clitic therefore attaches to *hindi* 'NEG', the first word of the sentence. In (17b–c), *niya* begins in the derivation inside a relative clause, in which the verb is *nakita* 'saw'; consequently, the clitic in these examples must attach to the first word of the relative clause. In (17b), the first word of the relative clause is again the negative marker *hindi*, which negates the embedded verb. In (17c), negation has been attached to the main clause, and the first word in the relative clause is again the verb *nakita*, to which the clitic must therefore attach.

Kaufman wishes to defend the traditional claim that apparent 'verbs' may be heads of nominal phrases, with the same distribution that nouns have. I have argued against this analysis; examples in which a verb immediately follows unambiguously nominal material such as demonstratives, I claimed, are in fact examples in which the verb is contained in a relative clause, modifying a null nominal head. We have now seen four arguments for this conclusion: the structures in question are islands for extraction (because, I claimed, they contain relative clauses), they behave like nominals for purposes of the distribution of the copula, and phenomena like the placement of negation and the behavior of clitics indicate the presence of a clause boundary which would be unexpected on Kaufman's theory.

2.   Against null structure

Kaufman offers two main arguments against positing null structure in Tagalog, which I will address briefly here. The first has to do with conditions on the Tagalog linker; the second, with the distribution of genitive nominals.

The Tagalog linker has two allomorphs, a velar nasal and a free-standing syllable *na*, which are partly conditioned by phonotactics. As Kaufman notes, however, there are further conditions on the allomorphy

which are poorly understood; an adjective modifying a noun, for example, must use the velar nasal allomorph of the linker if this is phonotactically possible, but a complement clause following a verb may be preceded by either allomorph of the linker, with a slight preference for the *na* allomorph, even if the velar nasal would be phonologically permissible.

Kaufman notes that the linkers in the examples in (18) obey the same conditions on their allomorphs:

(18) a. Ito ang dalawa (-ng/#na) guro
          this NOM two   LNK   teacher
          'These are the two teachers'
      b. Ito ang dalawa (-ng/#na) nagtuturo
          this NOM two   LNK   AV.BEG-INCM-teach
          'These are the two who are teaching'

Kaufman claims that the identical behavior of the linkers in these examples supports the claim that *guro* 'teacher' and *nagtuturo* 'is teaching' are the same kind of word. Alternatively, we might take these examples as evidence for the presence of a null nominal head in (18b); on this view, the linkers in both examples are followed by a nominal (which is null in (18b)). The argument seems not to distinguish between the theories.

Kaufman's second argument has to do with the contrast in (19):

(19) a. ang basag (*ng   babae)
          NOM broken GEN woman
          'the woman's broken one'
      b. ang basag na   bintana (ng babae)
          NOM broken LNK   window   GEN woman
          'the woman's broken window'

The contrast in (19), Kaufman argues, is mysterious on a theory in which both the examples in (19) contain nominal heads. If there is a null nominal head in (19a), why can it not license a possessor, like the overt nominal head in (19b)?

Kaufman notes in his footnote 24 that the ill-formedness of (19a) may be more pragmatic than syntactic, since there are structurally parallel examples which are well-formed, given an appropriate context. Given this, it is unclear that the contrast in (19) should influence our beliefs about the syntax of Tagalog. Even if we take the contrast in (19) at face value, we

already know from languages like English that null nominals are more re-stricted in the material with which they can combine than their overt counterparts:

(20)   a.   John's Ø
       b.   * the Ø

If we did decide to construct a syntactic account of the facts in (19), then, Tagalog would not be the first language to require such an account.

3.   Larger issues

Kaufman discusses consequences for his theory from two contentious do-mains of Tagalog syntax: conditions on extraction, and binding theory. Considerations of space prevent me from discussing these phenomena at any length, but I will comment briefly here on each.

Kaufman offers two kinds of arguments for his theory from the condi-tions on extraction in Tagalog, some of which are exemplified in (21):

(21)   a.   Sino ang bumili   ng   tela?
            who NOM AV.BEG-buy GEN cloth
            'Who bought the cloth?'
       b.   * Sino ang binili   ang tela?
            who NOM PV.BEG-buy NOM cloth

On Kaufman's theory, the predicates in (21) are both nominal, and the wh-extracted phrase in (21b) is the possessor of the nominal predicate *binili* (which should be given a translation, on Kaufman's theory, some-thing like 'bought thing'). Kaufman suggests two possible problems with this type of extraction. One is morphological; the extracted wh-phrase does not have the case morphology of a possessor. The other is more syn-tactic; extraction of possessors is often impossible.

The morphological part of this proposal cannot cover all the facts about wh-extraction in Tagalog. In particular, consider the conditions on long-distance extraction:

(22)   a.   Sino ang sinabi   ni Maria [na bumili   ng tela]?
            who NOM PV.BEG-say GEN Maria LNK AV.BEG-buy GEN cloth
            'Who did Maria say bought the cloth?'

b. *Sino ang sinabi   ni Maria [na **binili**   ang tela]?
   who NOM PV.BEG-say GEN Maria LNK PV.BEG-buy NOM cloth

c. *Sino ang **nagsabi**   si Maria [na bumili   ng tela]?
   who NOM AV.BEG-say NOM Maria LNK AV.BEG-buy GEN cloth

d. *Sino ang **nagsabi**   si Maria [na **binili**   ang tela]?
   who NOM AV.BEG-say NOM Maria LNK PV.BEG-buy NOM cloth

(22a) represents the only grammatical way of asking this question in Tagalog; in (22b–d), we can see that changing the voice of either of the verbs in the sentence makes it ill-formed (the offending verbs are bold-faced). If a DP is extracted in Tagalog, not only must the verb of the DP's clause take the voice which would mark the DP as Nominative, but verbs of higher clauses must be in the voice which would mark the clause from which extraction takes place as Nominative. The problem for a theory which blames the condition on extraction entirely on the morphology of the extracted phrase is that it cannot handle the second half of this generalization. In (22c) above, the verb of the embedded clause has the same form as its counterpart in the well-formed (22a), and we would expect the morphology of the extracted wh-phrase to be the same; nevertheless, the result is ill-formed.

Kaufman's second approach to extraction is more promising; perhaps it is simply that possessors are difficult to extract (or, as in (22), to extract from). It is interesting, in this regard, that possessors of nominal predicates do seem to be extractable in Tagalog, a fact first discussed by Cena (1979):

(23) 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'

In (23b), the possessor of the nominal predicate *kasama* 'companion' can be relativized.

Still, if this problem could be circumvented (along with the various arguments in the previous section against declaring all Tagalog predicates to be in some sense nominal), then Kaufman's account of the conditions on extraction would be attractive. In this form, the account would closely

resemble the locality-based approaches which, as he notes, are often offered as explanations for these facts; Kaufman would in effect be claiming that the conditions on extraction represent an island effect of a particular sort.

Finally, Kaufman has a brief discussion of binding facts in Tagalog, which begins by mentioning their ''lack of clarity''. As in any language, there are binding facts in Tagalog which are unclear, but I am not aware of any controversy about the facts that Schachter originally described; in particular, the Actor invariably binds an anaphor in Patient position, and the reverse is impossible, regardless of which voice we find on the verb:

(24) a.   Kumagat ang aso sa sarili niya
               AV.BEG-bite NOM dog DAT self   3s.GEN
               'The dog bit himself'
     b.   Kinagat   ng aso ang sarili   niya
               PV.BEG-bite GEN dog NOM self   3s.GEN
     c.   *Kumagat   ng aso ang sarili niya
               AV.BEG-bite   GEN Juan NOM self 3s.GEN
               'Himself bit the dog'
     d.   *Kinagat   ang aso ng sarili niya
               PV.BEG-bite   NOM dog GEN self 3s.GEN

Kaufman discusses a well-formed example similar to (24b), comparing it with the well-formed example in (25):

(25) Kaaway ni   Tyson   ang sarili niya
       enemy   GEN   Tyson   NOM self 3s.GEN
       'Tyson's enemy is himself'

The well-formedness of (25) is indeed striking. Still, if we are to take seriously the parallel between (25) and (24b), what we expect, on Kaufman's theory, is that any possessor (that is, any nominal marked with GEN) will be able to bind any nominative argument. But this is not what we find. (24c) shows one counterexample, in which a possessor is unable to bind the nominative argument. In fact, anaphor binding seems to be generally unaffected by case marking, contrary to what Kaufman's theory would lead us to expect. This is one area where Kaufman is careful to state that he is prevented by space constraints from describing his whole theory, so perhaps he has an account of these facts in mind.

4.   Conclusion

Kaufman's proposal is original and interesting, and he discusses many underexplored topics in Tagalog, uncovering a variety of intriguing facts. I have tried to argue here against some of his conclusions, but he has performed a service to Austronesianists, and to the field more generally, though his careful and creative work on the language.

Massachusetts Institute of Technology
norvin@mit.edu

References

Cena, Resty. 1979. Tagalog counterexamples to the Accessibility Hierarchy. *Studies in Philippine Linguistics* 31.119–124.
Richards, Norvin. 2009. The Tagalog copula. Talk given at AFLA 16.
Wolfram, Walt. 1969. *A sociolinguistic description of Detroit Negro speech*. Washington, D.C.: Center for Applied Linguistics.