Distinguishing nouns and verbs: A Tagalog case study[*]

Henrison Hsieh
McGill University
`henrison.hsieh@mail.mcgill.ca`

Author's Accepted Manuscript

Abstract

Many standard tests for identifying syntactic category, such as distribution and morphological potential, fail to distinguish between nouns and verbs in Tagalog. This state of affairs has led some scholars to propose that Tagalog does not treat verbs as a separate syntactic category from nouns. One such proposal by Kaufman (2009) further argues that adopting this so-called nominalist view of Tagalog allows us to straightforwardly understand a number of facts about the syntax of this language that have so far resisted consensus in analysis. This paper argues that the nominalist view cannot be maintained for Tagalog by showing differential behavior between nominal and verbal constituents in this language. It will furthermore be argued that the nominal behavior exhibited by verbal constituents can be attributed to a pervasive process of nominalization, broadly construed, neutralizing the aforementioned differential behavior. The result of this neutralization is reminiscent of phenomena found in other languages, and raises interesting questions regarding the range of nominalization processes observed cross-linguistically.

Keywords: Syntactic category, Relativization, Nominalization, Tagalog, Austronesian

1 Introduction

It has been observed (see Himmelmann 2008, Kaufman 2009 and references therein) that Tagalog (Austronesian, Western Malayo-Polynesian) exhibits a striking flexibility in the kinds of constructions that appear in predicate and argument positions. Notably, apparent nouns and apparent verbs exhibit this flexibility, as Kaufman (2009) shows with the following pair of examples (1). We find that both *aso* 'dog' and *nag-ingay* 'made noise' can appear in either sentence-initial predicate position or argument position (marked below by the determiner *ang*).[1]

---

[*] An earlier version of part of this paper was published in the *Proceedings of the Forty-Second Annual Meeting of the Berkeley Linguistics Society* with the title *Distinguishing nouns and verbs: Against nominalism for Tagalog*.

[1] This paper follows the Leipzig Glossing Rules (version of May 31, 2015) with the following additional abbreviations: AN = Agent Nominalizer, AIA = Ability and Involuntary Action (see Schachter and Otanes 1972, Ch. 5.13), AV = Actor Voice, CV = Conveyance Voice, GER = Gerund, LK = Linker, LN = Locative Nominalizer, LV = Locative Voice, PV = Patient Voice, PN = Patient Nominalizer. Tagalog examples from other authors have been edited to be consistent with the glossing conventions

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

    b. Aso ang nag-ingay.
         dog NOM PFV.AV-noise
         'What made noise was a dog.'

<div align="right">(Kaufman 2009, 19)</div>

One view of the data above might take (1a) to be a run-of-the-mill sentence involving a verbal predicate and a nominal subject, as sketched in (2a). For (1b) on the other hand, the subject *ang nag-ingay* 'what made noise' might be analyzed as derived from verbal structure (e.g., as a headless relative) whereas the predicate would be nominal (possibly with a null copula), as reflected in the free translation and sketched in (2b). Under this view, the flexibility of nouns and verbs is due to syntactic processes that have category-changing effects. Let us refer to such a position as a *verbal* view of Tagalog.

(2) a. [$_{V/VP}$ Nag-ingay   ][$_{DP\text{-}Subj}$ ang   aso].
             AV.PFV-noise   NOM dog
         'The dog made noise.'

    b. [$_{NP}$ Aso][$_{DP\text{-}Subj}$ ang   [$_{CP}$ Op_i [$_{TP}$ nag-ingay t_i]]].
             dog   NOM   AV.PFV-noise
         'What made noise was a dog.'

Another view might take the pattern exemplified above as evidence for true categorial flexibility due to the lack of a syntactic distinction between nouns and verbs. Kaufman (2009) provides a concrete instantitation of this view, arguing that seemingly verbal constructions like *nag-ingay* are better analyzed not as headless relatives but as something closer to participant nominals (akin to those formed with English *-er*). (3) sketches the stuctures for (1) following this view and provides free translations that more closely reflect this structure.

(3) a. [$_{DP}$ Nag-ingay   ][$_{DP\text{-}Subj}$ ang   aso].
             AV.PFV-noise   NOM dog
         The dog was a noise-maker.

    b. [$_{DP}$ Aso][$_{DP\text{-}Subj}$ ang   nag-ingay].
             dog   NOM AV.PFV-noise
         The noise-maker was a dog.

Kaufman further argues that many aspects of Tagalog syntax that have been the subject of much debate in the literature can be straightforwardly accounted for under this view. Such aspects include Tagalog's well-known extraction restriction, its so-called "voice system", and its case marking patterns, as discussed in Section 2. I follow Kaufman in referring to such a position as a *nominalist* view. In this paper I argue against nominalist views of Tagalog on both empirical and theoretical grounds. For concreteness, I focus on Kaufman's (2009) implementation, which will be referred to here as the Participant Nominal Analysis, or PNA.

The main goal of this paper is to present data showing that, despite the apparent flexibility of Tagalog nouns and verbs shown in (1), there still exists behavior in this language that is best explained by a dif-

---

used in this paper. All uncited examples are from my own elicitation work or native speaker intuitions. To the best of my knowledge, all speakers consulted (including myself) speak the dialect of Tagalog used in Manila. All errors and misquotations are my own.

ference in syntactic category. In particular, I compare constructions like *nag-ingay* 'made noise' above to morphologically similar constructions that I will argue better fit the characterization of participant nominal.

For example, the (in some sense) more "verbal" *lagyan* 'put (LV)' and the more "nominal " *lagayan* 'container' both derive from the root *lagay*. Both exhibit the flexibility illustrated in (1): they are able to appear as predicates or arguments (4).

(4) a. { Lagy-an / Lagay-an } ng   pera   ang kaho=ng ito.
         put-LV   put-LN   GEN money NOM box=LK this
    lagyan: 'Put money in this box.'
    lagayan: 'This box is a money container.'

  b. Ang kaho=ng ito ang   { lagyan / lagayan } ng   pera.
     NOM box=ng this NOM   put-LV   put-LN   GEN money
    lagyan: 'Put money in THIS BOX.'
    lagayan: 'The money container is THIS BOX.'

In certain contexts, this parallelism is broken. For example, (5) shows a context where an element *maging* is ungrammatical with the more verbal *lagyan* but is in contrast obligatory with the more nominal *lagayan*.

(5) a. Gusto ko=ng   (*****maging**) **lagy-an** ng   pera   ang kaho=ng ito.
         want 1SG.GEN=LK AV.be   put-LV   GEN money NOM box=LK this
    'I want to put money in this box.'

  b. Gusto ko=ng   **(**maging**) **lagay-an** ng   pera   ang kaho=ng ito.
     want 1SG.GEN=LK AV.be   put-LN   GEN money NOM box=LK this
    'I want this box to become a money container.'

The asymmetry in (5) can, however, be neutralized when elements from the extended nominal projection (in the sense of Grimshaw 2000, e.g., determiners, plural markers, etc.) mark the relevant constructions. For example, the numeral *isa* 'one' causes *maging* to be obligatory in both cases.

(6) Gusto ko=ng   **(maging) **isa=ng** { nila~lagy-an / lagay-an } ng   pera   ang kaho=ng ito.
    want 1SG.GEN=LK AV.be   one=LK IPFV~put-LV   put-LN   GEN money NOM box=LK this
    'I want {to put money in this box / this box to become a money container }.'

I argue that this pattern of asymmetry and neutralization is best understood as neutralization of an initial difference in syntactic category, since this neutralization occurs specifically in the presence of a variety of nominal functional elements. More specifically, I adopt a structure along the lines of that employed for event nominalizations and deverbalizations (in the sense of e.g., Kornfilt and Whitman 2011), whereby verbal projections are selected by nominal functional heads, resulting in a nominal external distribution. This data then constitutes counterevidence to the claim that no independent evidence exists for such a process of nominalization in Tagalog (see Himmelmann 2008, Sec. 3.1), and casts doubt on the foundational assumption of nominalist analyses.

Secondarily, I discuss in detail a significant problem that constructions with long-distance A-bar dependencies pose specifically for the PNA. Some authors, such as Aldridge (2009), point out that this implementation cannot readily generate these dependencies because it predicts the presence of an intervening

phase boundary (DP, as we will see) blocking extraction. I will show that while the nominalist view makes the wrong predictions regarding long-distance dependencies for verbal constructions (such as *nag-ingay* 'made noise'), these predictions are consistent with the behavior of truly nominal constructions in Tagalog. This highlights another distinction between nominal and verbal elements in Tagalog that cannot be captured without an underlying distinction in syntactic category, since without such a distinction, we predict identical behavior for extraction (whether well- or ill-formed).

This paper is organized as follows. Section 2 presents some background for Tagalog and the nominalist hypothesis (as instantiated in the PNA), including its motivations and advantages. Section 3 introduces three morphologically defined classes of constructions which display the flexible behavior discussed in the previous section. These construction classes become relevant in Section 4, which discusses the aforementioned asymmetry and neutralization, and argues that this pattern is best understood as asymmetry and neutralization of syntactic category. Section 5 discusses the problem that long-distance dependencies pose for the PNA. Finally, Section 6 concludes and considers the potential cross-linguistic implications of the Tagalog data.

2   The nominalist view and Tagalog

The notion that Tagalog does not distinguish syntactically between verbs and nouns is not a new one. Previous literature (e.g., Starosta et al. 1982) has argued for various formulations of this position, drawing on the flexibility of apparent nouns to appear in verbal position and vice versa. Kaufman's (2009) Participant Nominal Analysis takes this view further and proposes that there is more at stake here than the labels we use to refer to classes of syntactic objects. A core argument of the PNA is that the nominalist view in fact helps to explain many of the otherwise puzzling facts of Tagalog syntax.

This section is primarily concerned with presenting the major empirical facts of Tagalog that bear on this discussion and considering them in the context of the PNA. I first give an overview of some of the major points of Tagalog syntax. Then I discuss some of the evidence argued to support and motivate nominalist analyses for this language. The goal of this discussion is to show that there is indeed serious reason to suspect, at least initially, that Tagalog may not have a noun-verb distinction in the same way that, say, English does. I then give an overview of how the PNA is able to account for the major aspects of Tagalog syntax raised.

2.1   Tagalog basics

Tagalog is a predicate-initial language where arguments are marked for case. Full DPs functioning as core arguments bear one of the determiners *ang* or *ng*[2] (*si* and *ni* respectively for [+proper, +animate] DPs) while pronouns appear in different case forms.[3] There is also a third marker for oblique or adjunct-like constituents, *sa* (*kay* for [+proper, +animate]), that has some characteristics of both prepositions and determiners (see Gerassimova and Sells 2008, Himmelmann 2016 for further discussion). Examples of basic transitive sentences are given below.[4]

---

[2]This is one of two words that have an abbreviated spelling in Tagalog. It is pronounced "nang" [naŋ].

[3]I follow Kaufman's (2009) glossing of *ang* as NOM and *ng* as GEN, but I will refer to them as *ang* and *ng* in running text.

[4]Here, I take the term transitive to refer loosely to sentences that contain two arguments. There is much discussion in the literature as to which sentence types are basic transitive sentences and which ones are derived in some sense, or even if a syntactic notion of transitivity as is familiar from other languages is meaningful. See Guilfoyle et al. 1992, Aldridge 2004, Kaufman 2009

(7) a. **Mag**ba~basa **ang bata** ng libro.
        FUT.AV~read NOM child GEN book

        'The child is going to read a book.'

    b. Ba~basah-**in** ng bata **ang libro**.
        FUT~read-PV GEN child NOM book

        'The child is going to read the book.'

Verbs typically bear so-called voice morphology that correlates (roughly) with the thematic role of the argument marked *ang*.[5] In (7a), *ang*-marking appears on the agent, and the verb is marked with one of the actor voice (AV) prefixes *mag-*. In (7b), the patient is *ang*-marked, so the verb bears the patient voice (PV) suffix *-in*. Tagalog's voice system is distinct from other (non-Austronesian) voice alternations, such as the active-passive alternation in English, in that it extends beyond the core arguments.[6] There are at least two other voices that are typically acknowledged in the literature: locative voice (LV, suffix *-an*; e.g., locations and goals) and conveyance voice (CV, prefix *i-*; e.g., benefactors). Below is a sample paradigm from Rackowski and Richards (2005, 565).

(8) a. B<**um**>ili   **ang bata** ng tela sa   palengke para sa   nanay.
        <AV.PFV>buy NOM child GEN cloth OBL market   for OBL Mother

        'The child bought cloth at the market for Mother.'   *Actor Voice*

    b. B<in>ili-**Ø**   ng bata **ang tela** sa   palengke para sa   nanay.
        <PFV>buy-PV GEN child NOM cloth OBL market   for OBL Mother

        'The child bought the cloth at the market for Mother.'[7]   *Patient Voice*

    c. B<in>ilh-**an** ng bata ng tela **ang palengke** para sa   nanay.
        <PFV>buy-LV GEN child GEN cloth NOM market   for OBL Mother

        'The child bought cloth at the market for Mother.'   *Locative Voice*

    d. **I**-b<in>ili   ng bata ng tela sa   palengke **ang nanay**.
        CV-<PFV>buy GEN child GEN cloth OBL market   NOM Mother

        'The child bought cloth at the market for Mother.'   *Conveyance Voice*

The voice system also interacts with a restriction on A-bar extraction. Generally speaking, *ang*-marked DPs (but not *ng*-marked ones) may be targeted by processes like relativization, *wh*-question formation, topicalization, and clefting.[8] The examples below show this restriction for relativization. We see in (9) that relativizing the agent requires the verb to be in actor voice, and in (10) that relativizing the theme requires patient voice. Note that for the grammatical examples, a relative clause head can be present (with the

---

for a sampling of the various views on this issue.

[5]This is a simplification, but it should serve as a good approximation for the purposes of this paper. For example, actor voice can also be found in cases where the *ang*-marked argument is an experiencer, as a reviewer points out.

[6]As such, the term *voice* is used pretheoretically in this paper, and the parallel to voice alternations such as active-passive is made for the purpose of exposition. As with its phrase structure, many analyses have been proposed for the verbal morphology of Tagalog. Examples include *wh-* or Case-agreement (Rackowski and Richards 2005; Pearson 2005 for Malagasy, a related language), flavors of v^0 (Aldridge 2004), and, as will be discussed in this paper, participant nominal morphology (Kaufman 2009).

[7]The patient voice suffix *-in* is null in the presence of the infix *<in>*. This infix appears on its own in perfective aspect and with stem-initial CV reduplication in imperfective aspect for non-AV forms.

[8]Adjuncts, including *sa*-marked phrases, generally freely undergo A-bar extraction, although such extraction almost always differs in surface form from extraction of *ang*-DPs. For example, adjunct relatives do not involve the linker morpheme, and instead make use of a complementizer *kung* and a *wh*-element.

linker morpheme[9]) yielding a typical headed relative clause construction. Alternatively, the head (and the linker) can be absent, resulting in a headless relative clause paraphrased here as an English relative clause headed by *one*. The nature of these headless relatives is one of the main issues the PNA addresses.

(9) a. (**bata**=ng) b<**um**>ili   ng tela sa   palengke para sa   nanay
         child=LK <AV.PFV>buy GEN cloth OBL market   for OBL Mother

         '{ child / one } who bought cloth at the market for Mother'

   b. ?? **bata**=ng b<in>ili-Ø   **ang tela** sa   palengke para sa   nanay
       child=LK <PFV>buy-PV NOM cloth OBL market   for OBL Mother

       'child who bought cloth at the market for Mother'[10]

(10) a. * **tela**=ng b<**um**>ili   **ang bata** sa   palengke para sa   nanay
          cloth=LK <AV.PFV>buy NOM child OBL market   for OBL Mother

          Intended: 'cloth that the child bought at the market for Mother'

   b. (**tela**=ng) b<in>ili-Ø   ng   bata sa   palengke para sa   nanay
       cloth=LK <PFV>buy-PV NOM child OBL market   for OBL Mother

       '{ cloth / one } that the child bought at the market for Mother'

The Tagalog voice system, argument marking patterns, and extraction restrictions are some of the major points of focus in analyses of Tagalog syntax. These analyses have so far been varied to the extent that there is no general consensus in the literature (see e.g., Aldridge 2004, Guilfoyle et al. 1992, Kroeger 1993, Rackowski and Richards 2005 for some sense of the variability among analyses). Kaufman's (2009) PNA also endeavors to capture these facts by proposing an analysis where verbs are not a distinct syntactic category from nouns. Evidence and motivation for this noun-verb categorial union comes not only from historical reconstructions of the voice morphemes as nominalizers co-existing alongside a separate series of verbal morphology in Proto-Austronesian (Wolff 1973, Starosta et al. 1982, Ross 2009), but also from synchronically attested behavior in this language, which I turn to now.

2.2 Motivations for nominalism

There is good reason to suspect that nouns and verbs are actually part of a single syntactic category in Tagalog. These elements exhibit significant overlaps in their behavior, most saliently in their distribution and morphological potential. With regards to distribution, nominal and verbal constituents may both appear in prototypically verbal positions, like as predicates. Notice that we find no overt material (e.g., a copula) that might help distinguish the two categories in this position.

(11) *Verbal predicates*

   a. **K<um>ain ng tapa** si   Bong.   b. **T<um>akbo** ang pusa.
       <PFV.AV>eat GEN tapa NOM Bong   <PFV.AV>run NOM cat

       'Bong ate tapa.'   'The cat ran.'

---

[9]The linker (glossed LK) appears between the head and the relative clause modifier. It appears consistently in modification contexts, but also has a complementizer-like function, appearing in various embedding constructions.

[10]Some speakers (including myself) do not judge agent extractions out of non-actor voices to be totally ungrammatical. (9b) is not as well-formed as (9a), but is significantly better than the ungrammatical (10a). These judgements might be more typical of younger (around or below university-age) speakers (Nozomi Tanaka, p.c.). The analysis of such examples is the subject of ongoing work, so they will be treated here as ungrammatical for the purposes of discussion.

(12) *Nominal predicates*

      a. **Kaibigan ni Gina** si   Bong.
          friend   GEN Gina NOM Bong

          'Bong is <u>Gina's friend</u>.'

      b. **Aso** ang hayop na iyan.
          dog NOM animal LK that

          'That animal is <u>a dog</u>.'

These same constructions may furthermore also appear in prototypically nominal positions, like as the semantic head of a DP marked by the determiner *ang*. These DPs may in turn appear in argument positions as shown in (13-14). Notice again that no overt material (e.g., relativizing morphology) distinguishes the two types of objects, mirroring the picture we see in predicate position.

(13) *Verbal arguments*

      a. Na-kita   ko   **ang k<um>ain ng tapa**.
          PFV-see[PV] 1SG.GEN NOM <PFV.AV>eat GEN tapa

          'I saw <u>the one who ate a tapa</u>.'   *cf. (11a)*

      b. Pusa **ang t<um>akbo**.
          cat   NOM <PFV.AV>run

          '<u>The one that ran</u> is a cat.'   *cf. (11b)*

(14) *Nominal arguments*

      a. Nagtu~turo   **ang kaibigan ni   Gina**.
          IPFV.AV~teach NOM friend   GEN Gina

          '<u>Gina's friend</u> is teaching.'   *cf. (12a)*

      b. Matu~tulog **ang   aso**.
          FUT~sleep   NOM dog

          '<u>The dog</u> will sleep.'   *cf. (12b)*

For the examples above, we might point out that those labeled verbal bear voice and aspect morphology (specifically the infix *<um>*) whereas those labeled nominal do not. However, as previously mentioned, we find significant overlap in the morphological potential of putative nouns and verbs. Putatively verbal lexical items may appear bare (15-16) while putatively nominal ones may appear with voice and aspect morphology (17-18), in contrast to (11-14).[11] The distribution of these furthermore mirrors the data we have seen so far: both voice-marked nominal constructions and bare verbal constructions may appear in either predicate position or marked by *ang*.

(15) *Bare verbs in predicate position*

      a. **Takbo** na   ito   para sa   kanya.
          run   now this.NOM for OBL 3SG.OBL

          'This (pace) is already <u>a run</u> for him (even though it might be slow).'

      b. **Sulat ni   Sarah** ito.
          write GEN Sarah this.NOM

          'This is <u>Sarah's { writing / letter }</u>.'

---

[11]Note that we observe particular semantic effects with bare verbal roots and voice-marked nouns. Himmelmann (2008) points out that bare verbal roots often take on a small range of possible meanings, and Kroeger (1998) notes that voice-marked nouns often have meanings and argument structure that may be unpredictable from the root (although see Kaufman 2009, fn.4, for a response). This behavior might be taken to indicate that some kind of process of coercion has occurred in (15-18), instead of voice morphology simply picking out available argument-structural positions of the root (as argued by Kaufman).

(16)   *Bare verbs in argument position*

    a. Mabilis **ang takbo niya**.
       fast   NOM run   3SG.GEN

       '<u>Her running</u> is fast.'

    b. Na-hanap   ko   **ang sulat ni   Bong**.
       PFV.AIA-find[PV] 1SG.GEN NOM write GEN Bong

       'I found <u>Bong's letter</u>.'

(17)   *Voice marked nouns in predicate position*

    a. **K<in>aibigan   ni   Gina** ang   bata.
       <PFV>friend[PV] GEN Gina NOM child

       '<u>Gina befriended</u> the child.'

    b. **Magdo~doktor** si   Julian.
       FUT.AV~doctor NOM Julian

       'Julian <u>will become a doctor</u>.'

    c. **I-ta~taxi   ni   Armi** ang   mga gamit niya.
       CV-FUT~taxi GEN Armi NOM PL   thing 3SG.GEN

       'Armi <u>is going to transport</u> her things <u>by taxi</u>.'

(18)   *Voice marked nouns in argument position*

    a. H<in>abol   ng   bata **ang k<in>aibigan   ni   Gina**.
       <PFV>chase[PV] GEN child NOM <PFV>friend[PV] GEN Gina

       'The child chased <u>the one Gina befriended</u>.'

    b. K<in>ausap   ko   **ang magdo~doktor**.
       <PFV>speak.with[PV] 1SG.GEN NOM FUT.AV~doctor

       'I talked to <u>the one who will become a doctor</u>.'

    c. Ku~kun-in ko   **ang i-ta~taxi   ni   Armi**.
       FUT~get-PV 1SG.GEN NOM CV-FUT~taxi GEN Armi

       'I'll take <u>what Armi is going to transport by taxi</u>.'

The data given above only provides a brief overview of some of the Tagalog facts, but they serve to demonstrate that typical tests used to distinguish between nouns and verbs fail to identify the relevant categories in this language. Instead, it appears that a single class of syntactic objects in this language shows behavior that is both nominal (appearing morphologically bare or in argument position) and verbal (bearing "verbal" morphology, appearing in predicate position). Nominalist analyses of Tagalog, arguing that nouns and verbs should be grouped into a single syntactic category, are thus not without basis.[12] Under this view, any perceived distinction between nouns and verbs in Tagalog would simply stem from analogy with languages that do make a distinction, or from some other factor such as event structure.[13]

---

[12]The category label "verb" also appears compatible with the data presented so far. To my understanding, however, one reason for the nominal analysis is that the forms found in modern Tagalog (and many other Austronesian languages) derive from nominal constructions in Proto-Austronesian, which existed alongside distinct verbal constructions (Starosta et al. 1982).

[13]One might at this point wonder whether all syntactic objects in Tagalog exhibit this behavior, in which case the tests presented here might simply be unsuitable for this language. See Kaufman (2009, 26) for evidence that deictics (and more generally oblique phrases) exhibit a different distribution from nouns and verbs. Roughly speaking, deictics like *dito* 'here' take on a different form (i.e., *nandito*) when used predicatively instead of as an adjunct. We find similar behavior for oblique phrases marked *sa/kay*.

2.3 Theoretical advantages of nominalism

The Participant Nominal Analysis is an instantiation of the nominalist approach arguing that nominalism can go further than describing the distribution of the relevant constructions in Tagalog. At least two other major details of Tagalog syntax, which have so far resisted consensus in analysis, are argued to naturally fall out under this view. These are the voice marking patterns and A-bar extraction restrictions.

The PNA views the Tagalog voice morphemes as having a semantic function similar to participant nominalizers, like *-er* and *-ee* in English. That is, they create forms that denote some entity in the conceptual argument structure of the root they attach to. Basis for this claim can be found in historical reconstructions of the voice morphemes as nominalizers in Proto-Austronesian (Wolff 1973, Starosta et al. 1982, Ross 2009). Some evidence for this is also observable synchronically in the flexibility of the voice system, which has parallels to the behavior of English *-ee*. Specifically, Barker (1998) points out that forms like *amputee* do not necessarily denote entities that would appear as a syntactic argument to their roots. *Amputee* must denote a person even though the object of *amputate* must be some appendage.

Similar behavior is attested in the Tagalog voice system. Kaufman (2009, 7) observes that "any Tagalog lexical root can take any voice so long as the conceptual representation of the root provides for the relevant participant". For example, a canonically transitive root like *bili* 'buy' may appear not only in AV and PV forms as expected, but LV and CV form as well, as shown previously in (8), repeated below.

(19)   a.   **B<um>ili**   **ang bata** ng   tela   (sa   palengke) (para sa   nanay).
            <AV.PFV>buy NOM child GEN cloth OBL market   for   OBL Mother

            'The child bought cloth (at the market) (for Mother).'   *Actor Voice*

       b.   **B<in>ili-Ø**   ng   bata   **ang tela** (sa   palengke) (para sa   nanay).
            <PFV>buy-PV GEN child NOM cloth OBL market   for   OBL Mother

            'The child bought the cloth (at the market) (for Mother).'   *Patient Voice*

       c.   **B<in>ilh-an**   ng   bata   ng   tela   **ang palengke** (para sa   nanay).
            <PFV>buy-LV GEN child GEN cloth NOM market   for   OBL Mother

            'The child bought cloth at the market (for Mother).'   *Locative Voice*

       d.   **I-b<in>ili**   ng   bata ng   tela   (sa   palengke) **ang nanay**.
            CV-<PFV>buy GEN child GEN cloth OBL market   NOM Mother

            'The child bought cloth (at the market) for Mother.'   *Conveyance Voice*   (Rackowski and
            Richards 2005, 565)

Notice that the LV and CV morphemes in this particular set of examples (19c-19d) do not target syntactically required arguments of *bili* 'buy'. That is, the *ang*-marked DPs in LV (*palengke* 'market') and in CV (*nanay* 'mother') represent additional information typical of adjuncts; these can be omitted in examples where they are not *ang*-marked without affecting the grammaticality of the sentence. In comparison the *ang*-marked DPs for AV (*bata* 'child') and PV (*tela* 'cloth') in (19a-19b) behave as core arguments; omitting these in examples where they are not *ang*-marked results in an implicit argument reading in the best case and ungrammaticality in the worst.

To account for the voice marking behavior of Tagalog, other authors have proposed systems involving different applicative heads and "flavors" of *v*. In particular, such analyses tend to treat the locative and conveyance voice constructions as involving applicatives. However, Kaufman (2017) points out that such analyses are problematic, partially because the proposed applicatives for Tagalog in these analyses would be

strange from a crosslinguistic perspective. For example, he notes that these proposed applicative morphemes do not obviously interact with transitivity (a problem for Aldridge 2004) or are null and assign non-typical case values to their objects (a problem for Rackowski and Richards 2005).

The PNA has two major consequences for the syntax of Tagalog. The first concerns apparent relative clauses like those in (20), abbreviated from (9-10). Apparent headless relatives are treated as being formally more similar to participant nominals in that they do not have (much) internal verbal or clausal structure. Apparent headed relatives fall out as cases of nominal modification, which is surface-identical to adjectival modification (shown in 21). The intuition for this structure is given in the second set of translations below (in double angle brackets).
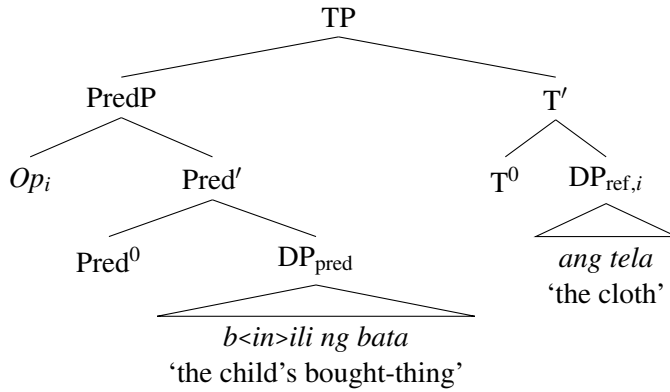
(20)   a.   (bata=ng) b<um>ili   ng   tela
            child=LK <PFV.AV>buy GEN cloth
            '{ child / one } that bought cloth'
            ⟪'(child who was a) buyer of cloth'⟫

       b.   (tela=ng) b<in>ili   ng   bata
            cloth=LK <PFV>buy[PV] GEN child
            '{ cloth / one } that the child bought'
            ⟪'(cloth that was the) child's bought-thing'⟫

(21)   a.   doktor na estudyante (rin)
            doctor LK student   also
            'doctor who is (also) a student'

       b.   doktor na matangkad
            doctor LK tall
            'tall doctor / doctor who is tall'

The second consequence concerns clause structure. The PNA uniformly treats Tagalog clauses as inherently copular and involving two DPs. The *ang*-marked DP in a clause (regardless of its thematic role) is treated as a sentential subject, while the rest of the clause is treated as a predicate DP. This approach contrasts with verbal analyses, which typically assign different structures to nominally and verbally predicated clauses, and also treat verbs and their internal arguments as constituents to the exclusion of the external argument.[14] The difference in constituency is illustrated in (22), which shows the clause structure the PNA adopts as applied to (23b) below. Here, the sentential subject, labeled DP_{ref}, is base-generated as the complement of T^0 and the DP predicate, labeled DP_{pred}, is contained within a PredP that is itself base-generated in Spec-TP.[15]

(22)   *Basic phrase structure under the PNA* (Kaufman 2009, 35)

---

[14]Although not all verbal analyses of Tagalog assign this kind of constituency across the board to regular declarative clauses. See e.g., Guilfoyle et al. 1992, Kroeger 1993.

[15]See Kaufman (2009, 35–36) for more details regarding this structure. As noted in the main text, the complement of T^0 is not a predicate-like constituent, but the sentential subject. For the Spec-TP position of PredP, Kaufman cites Massam's (2000) predicate fronting analysis as motivation, but crucially differs from her by base-generating PredP in this position.

```
                           TP
                 ┌─────────┴─────────┐
              PredP                  T′
          ┌────┴────┐            ┌───┴────┐
        Op_i      Pred′         T⁰     DP_ref,i
              ┌────┴────┐                 ╱╲
           Pred⁰      DP_pred          ang tela
                        ╱╲            'the cloth'
                 b<in>ili ng bata
              'the child's bought-thing'
```

Thus the apparent headless relatives in (20) appear as DP predicates in the following sentences.[16] Again, the bracketed translations are meant to capture the intuition behind the structure proposed by the PNA.

(23)   a.   [**B\<um>ili**   ng   tela]_DP [ang bata]_DP.
            \<AV.PFV>buy GEN cloth   NOM child
            'The child bought cloth.'

            ⟪'The child **was a buyer of cloth**.'⟫

       b.   [**B\<in>ili**   ng   bata]_DP [ang tela]_DP.
            \<PFV>buy[PV] GEN child   NOM cloth
            'The child bought the cloth.'

            ⟪'The cloth **was the child's bought-thing**.'⟫

The DP status of both predicate and *ang*-marked subject under the PNA can be used to make sense of the A-bar extraction restriction in Tagalog. The extractability of the *ang*-marked DP (i.e., DP_ref in (22)) is expected as it is a subject and we have no reason to believe that subjects of copular sentences cannot be extracted. The ban on *ng*-marked DP extraction falls out similarly readily under this analysis. In addition to marking agents and themes in non-actor and non-patient voices respectively, the determiner *ng* also marks possessors, as in (24).

(24)   B\<in>asa   ko   ang **libro ni   Juan**.
       \<PFV>read[PV] 1SG.GEN NOM book GEN Juan
       'I read Juan's book.'

The PNA treats all instances of *ng* as a genitive marker (in the same vein as Johns (1992) for the ergative and genitive markers in Inuktitut), so all *ng*-marked DPs in a predicate are genitive-marked DPs within a larger DP (i.e., DPs within DP_pred in (22)). From this, we can understand the ban on extracting *ng*-marked DPs simply as a ban on genitive extraction out of DPs, which is cross-linguistically widespread.[17]

---

[16]Note that this analysis does not propose anything specific to account for the variable word order possible among the postverbal constituents, although this gap seems to be shared among many analyses of Tagalog to varying degrees.

[17]Adjunct extraction, including that of *sa*-marked phrases, is derived by having these constituents adjoin to the sentential predicate constituent (PredP), outside of any DP constituent (Kaufman 2009, 40). For discussion of examples possibly showing genitive extraction, see Richards (2009a, Sec. 3) and fn. 10.

The fact that a nominalist approach can fairly straightforwardly capture some of the major points of contention in Tagalog syntax makes it a very appealing analysis for Tagalog. However, I will argue in the remainder of this paper that its application to the language in the first place is questionable. I will first show evidence arguing against the core assumption that Tagalog does not distinguish syntactically between verbs and nouns. Since many of the canonical tests for syntactic category fail for Tagalog (as shown above) or else operate off of assumptions that are not universally shared, I will base my argument on a general pattern of asymmetrical behavior between putative nouns and putative verbs. We will see that this asymmetry is subsequently neutralized in clearly nominal contexts, which I argue is evidence for the nature of the original asymmetry. I then consider some of the theoretical implications of the PNA for Tagalog. We will see that it makes specific predictions regarding the behavior of long-distance dependencies that are not empirically borne out.

3   Three kinds of "nominals"

The discussion in the rest of this paper will consider the behavior of three broad morphologically defined classes of constructions. These constructions all exhibit the distribution discussed previously in Section 2.2 – they appear equally productively in both argument and predicate positions. I refer to these three constructions as bare nominals, voice phrases, and derived nominals.

The overall rationale behind this particular partitioning is to consider constructions in various derivational states. One of the central claims of the PNA is that Tagalog simply does not have the category of verb. Stated differently, it maintains that at no point is there any verbal structure (e.g., a *v*P/VP layer) in the derivation of any Tagalog DP; they are nominal throughout.[18] What we will see in the next section is that this claim does not hold. Crucially, voice phrases pattern separately from (sometimes minimally different) bare and derived nominals in a way that I will argue is best explained by syntactic category. This behavior is especially striking given the fact that there is heterogeneity within each of these three classes of constructions.

While I will devote some discussion to the issue of how the three types of nominals below might be derived, I will ultimately not make any strong claims regarding this. That is, I will ignore details like internal structure and focus instead on the external distribution of these constructions. The general argument is thus that some subset of these nominals must in fact be verbal at the relevant derivational step, and that some of the nominal properties we observe are due to further syntactic operations.

3.1   Bare nominals

The first class of constructions considered is the most morphologically simple one, consisting of words that do not show any overt morphology on the root. Examples can be found corresponding to more time-stable notions (in the sense of e.g., Givón 2001), typically lexicalized as nouns in languages that clearly have such a distinction, such as *guro* 'teacher' in (25) and *baboy* 'pig' in (26), shown in predicate and argument

---

[18]To be more precise, Kaufman (2009) argues that most of what we might think of as verbal is actually nominal. The constructions introduced in this section fall under this nominal umbrella. However, implicit in this argument is the possibility that there are still instances of true verbs in the language. This seems to be the case with the imperative forms from Batangas Tagalog discussed (p.24–26), and might also be the right way under this analysis to think about other types of constructions, like the recent perfective. Recent perfective clauses have no *ang*-marked argument, and are conventionally viewed as not being marked for voice (see Schachter and Otanes 1972, Ch. 5.23; Rackowski 2002).

positions. As an aid to the reader, the heads of these constructions will be indicated with double underlines in the remainder of the paper.

(25) a. <u>Guro</u> si Christine.
teacher NOM Christine
'Christine is a teacher.'

    b. B<in>ati ni Kiko ang <u>guro</u>.
    <PFV>greet[PV] GEN Kiko NOM teacher
    'Kiko greeted the teacher.'

(26) a. <u>Baboy</u> iyan.
pig that.NOM
'That's a pig.'

    b. K<in>ain ng <u>baboy</u> ang bulaklak.
    <PFV>eat[PV] GEN pig NOM flower
    'The pig ate the flower.'

We also find examples that tend to more closely correspond to less time-stable (or more volatile) notions typically lexicalized as verbs crosslinguistically. These can also appear in both argument and predicate positions. The examples below show *luto* 'cook' in (27) and *aral* 'learn' in (28).

(27) a. <u>Luto</u> ni Boyet ang adobo=ng iyan.
cook GEN Boyet NOM adobo=LK that.NOM
'That adobo is Boyet's cooking.'

    b. K<in>ain ko ang <u>luto</u> ni Boyet.
    <PFV>eat[PV] 1SG.GEN NOM cook GEN Boyet
    'I ate Boyet's cooking.'

(28) a. <u>Aral</u> ng kwento=ng ito ang pagiging masunurin.
learn GEN story=LK this NOM being obedient
'Being obedient is the moral of this story.'

    b. S<in>abi ko sa bata ang <u>aral</u> ng kwento=ng ito.
    <PFV>say[PV] 1SG.GEN OBL child NOM learn GEN story=LK this
    'I told the child the moral of this story.'

Analyses may differ on how they derive different bare nominals. Under the PNA for example, the more "nominal" examples in (25-26) would have roughly the same derivations as the more "verbal" ones in (27-28). Alternatively, we might imagine that some of these examples are formed by some zero-derivation process, resulting in potentially more complex structure that is not overtly realized. As stated at the beginning of this section, I will largely ignore the internal structure of these constructions and focus on their external distribution and behavior.[19]

3.2 Voice phrases

The second class of constructions I define consists of forms that bear voice morphology. Again, as shown earlier in Section 2, this class includes constructions that have both apparently nominal and apparently verbal roots. I will refer to these as voice phrases, and indicate them with wavy underlining.[20]

---

[19]It should be noted that this morphologically defined class is more heterogeneous than it might appear from the data presented so far. In particular, I am setting aside a group of lexical items that take on a stative (verbal) meaning and obligatorily appear bare. The example below shows *alam* 'know'. For a discussion of some of the behavior of these constructions, see Richards (2009b).

(i) <u>Alam</u> ko na um-alis si Juan.
know 1SG.GEN LK PFV.AV-leave NOM Juan
'I know that Juan left.'

(ii) S<in>abi ko ang <u>alam</u> ko.
<PFV>say[PV] 1SG.GEN NOM know 1SG.GEN
'I said what I know.'

[20]Himmelmann (2008) also discusses such constructions, calling them *V-words*, where *V* means *voice*.

As with bare nominals, voice phrases can appear both in predicate and argument positions. Recall that under the PNA, these constructions should be thought of as participant nominals. This is different from the more intuitive (or perhaps naïve) interpretation of these constructions. When in predicate position, they may appear without additional functional elements (e.g., determiners, etc) and receive a "verbal" interpretation, creating the prototypical Tagalog verb-initial sentence. On the other hand in argument positions where such functional elements are obligatory, they are most naturally translated into English as free relatives.

The difference in behavior between voice phrases in predicate versus argument position will be crucial for the discussion later on. When disambiguation is required, I use the term "bare voice phrase" to refer to those that are not marked with additional morphology, and "headless relative" to refer those with nominal functional elements, as exemplified by the examples in argument position. Again, the free translations in double angle brackets more closely correspond to the intuition behind the PNA.

(29) a. [H<in>abol   ng   bata]$_{Pred}$ [ang nag-doktor]$_{Subj}$.
          <PFV>chase[PV] GEN child   NOM PFV.AV-doctor
          'The child chased the one who became a doctor.'
          ⟪'The doctor-er was the child's chased-one.'⟫

      b. [Nag-doktor]$_{Pred}$ [ang h<in>abol   ng   bata]$_{Subj}$.
          PFV.AV-doctor   NOM <PFV>chase[PV] GEN child
          'The one that the child chased became a doctor.'
          ⟪'The child's chased-one was a doctor-er.'⟫

(30) a. [B<in>ili   ko]$_{Pred}$ [ang i-ta~taxi   ni   Armi]$_{Subj}$.
          <PFV>buy[PV] 1SG.GEN NOM CV-FUT~taxi GEN Armi
          'I bought what Armi is going to send by taxi.'
          ⟪'Armi's by-taxi-thing was my bought-thing.'⟫

      b. [I-ta~taxi   ni   Armi]$_{Pred}$ [ang b<in>ili   ko]$_{Subj}$.
          CV-FUT~taxi GEN Armi   NOM <PFV>buy[PV] 1SG.GEN
          'Armi is going to send what I bought by taxi.'
          ⟪'My bought-thing will be Armi's by-taxi-thing.'⟫

The examples above are marked for both aspect and voice, but constructions without aspect also exist. Below are some examples of such aspectless voice phrases appearing as embedded predicates of certain verbs. For some embedding predicates like *gusto* 'want', an additional property that we find is that the lower agent DP can be omitted, producing a control-like reading where the embedded agent gap corefers with some matrix argument (here, the experiencer). Simpler baseline examples are given in (31), and embedded versions of (29) are given in (32) for comparison.

(31) a. Gusto niya=ng   [**mag-basa** ng   libro]$_{Pred}$ ([si   May]$_{Subj}$).
          want 3SG.GEN=LK AV-read   GEN book   NOM May
          'She wants (May) to read books.'
          ⟪'She wants (May) to be a reader of books.'⟫

      b. Gusto niya=ng   [**basah-in** (ni   May)]$_{Pred}$ [ang libro]$_{Subj}$.
          want 3SG.GEN=LK read-PV   GEN May   NOM book
          'She wants (May) to read the book.'
          ⟪'She$_i$ wants the book to be {May's/her$_i$} read-thing.'⟫

(32)   a. Gusto niya=ng   [**habul-in** (ng bata)]$_{Prd}$ [ang nag-doktor]$_{Sbj}$.
           want 3SG.GEN=LK chase-PV GEN child   NOM PFV.AV-doctor

           'He wants (the child) to chase the one who became a doctor.'

           ⟪'He$_i$ wants the doctor-er to be {the child's/his$_i$} chased-one.'⟫

      b. Gusto niya=ng   [**mag-doktor**]$_{Prd}$ ([ang h<in>abol   ng bata]$_{Sbj}$).
           want   3SG.GEN=LK   AV-doctor   NOM   <PFV>chase[PV] GEN child

           'He wants (the one who chased the child) to become a doctor.'

           ⟪'He wants (the child's chased-one) to be a doctor-er.'⟫

As with the aspect-marked examples, aspectless voice phrases may appear in (embedded) predicate positions, and show the same patterns in terms of argument marking. Compare the aspect-marked sentences in (29) to the aspectless embeddings in (32). In both (a) examples, the patient *nag-doktor* is *ang*-marked, matching the PV form of the relevant voice phrase, and in both (b) examples, the agent *hinabol ng bata* is *ang*-marked, matching AV on the relevant voice phrase. Aspectless voice phrases may also appear as matrix predicates and in argument-like positions, although their distribution in these positions is more limited. Examples of environments where this form is found include imperatives, as in (33), and other non-indicative clause types such as complements to *upang* 'in order to', *baka* 'maybe', or the hortative/optative particle *sana*. Section 4.2 discusses these aspectless voice phrases in more detail, devoting some attention to the difference in behavior from their aspect-marked counterparts.

(33)   a. Mag-doktor kayo.   b. Kayo   ang mag-doktor.
          AV-doctor   2PL.NOM   2PL.NOM NOM AV-doctor

          '(You) Be doctors.'   '(<u>YOU</u>) be doctors.'

     We will see later on that this class of constructions, the voice phrases, behaves differently from the other two, the bare nominals discussed previously and the derived nominals to be discussed next. This difference in behavior takes on a particular shape: while bare voice phrases behave differently from the two kinds of nominals, headless relatives do not. I will ultimately argue that bare voice phrases are verbal, and the nominal properties we observe are found with headless relatives, which are the result of a syntactically high process of nominalization (in the sense of Kornfilt and Whitman 2011) applying to bare voice phrases, contra the PNA.

3.3   Derived nominals

The third and final class of constructions considered consists of forms appearing with morphemes which I analyze as the true participant nominalization morphemes in Tagalog (i.e., analogous to English *-er* and *-ee*). These will be marked with dashed underlining. Constructions with three such morphemes will be relevant in this paper. First is an agent nominalizer *taga-* in (34), which describes a person who has been given the responsibility to perform some action.[21]

(34)   a. Taga-bantay ng   tindahan si   Junjun.
          AN-guard   GEN store   NOM Junjun

---

[21]Another prefix, *maN*-RED or *mag*-RED, also forms agent nominalizations such as *manu~nulat* 'writer/author' from *sulat* 'write' or *magna~nakaw* 'thief' from *nakaw* 'steal'. These differ from the *taga-* forms in that they do not carry the meaning of assigned responsibility, but a brief survey shows that they exhibit the same behavior with respect to the asymmetry-neutralization pattern discussed in Sec. 4. For more examples, see Schachter and Otanes (1972, Ch. 3.7)

'Junjun is the guard for the store.'

b. Na-galit   si   Nene sa   taga-bantay.
    PFV-anger NOM Nene OBL AN-guard

'Nene got angry at the guard.'

Second is the patient nominalizer *-in* in (35), which is more semantically neutral than *taga-* in the sense that it simply picks out the patient of the particular root it attaches to without adding more semantic content.

(35)   a. Gawa-in ni   Chit ang   mag-bantay ng   paninda sa   tindahan.
          do-PN   GEN Chit NOM AV-guard   GEN wares   OBL store

          'Chit's job is to keep an eye on the wares in the store.'

       b. Hindi na-tapos   ang   gawa-in ni   Chit.
          NEG   PFV.AIA-finish[PV] NOM do-PN   GEN Chit

          'Chit's task did not get finished.'

Third is the locative nominalizer *-an* in (36), which picks out the location where an event takes place.

(36)   a. Lagay-an ng   asin ang   kaho=ng iyan.
          put-LN   GEN salt NOM box=LK   that

          'That box is a salt container.'

       b. B<um>ili   si   Pepe ng   lagay-an ng   asin.
          <PFV.AV>buy NOM Pepe GEN put-LN   GEN salt

          'Pepe bought a salt container.'

It is worth pointing out here that the surface forms of the theme and locative nominalizers are identical to the patient voice *-in* and locative voice *-an* morphemes respectively. The difference lies in the kinds of effects we find on the root. The nominalizers preserve the form of the root, while the voice morphemes can trigger a process of syncope, which deletes certain segments in the final syllable of the root and has effects on stress assignment. The following tables show this.

The table in (37) shows *gawâ* 'do', which has final stress and ends in a glottal stop. The voice morphemes trigger deletion of the final vowel (and glottal stop), and the entire form receives final stress. On the other hand, the nominalizers preserve the final two segments, and stress is assigned to the final syllable of the root, resulting in overall penultimate stress.[22,23]

(37)

gawâ [gaˈwaʔ] 'do'	Patient (*-in*)	Locative (*-an*)
VOICE PHRASE	gawín [gaˈwin]	gawán [gaˈwan]
DERIVED NOMINAL	gawáin [gaˈwaːʔin] ('activity')	gawáan [gaˈwaːʔan] ('factory')

---

[22]I use the following conventions for diacritics when stress information is important. Acute accents mark stressed vowels (á) and circumflex accents mark stressed vowels adjacent to final glottal stops (â).

[23]Stressed penultimate vowels are typically long. Himmelmann (2005) notes that "stress in Tagalog appears to be closely interrelated with vowel length. For some authors ... stress is the primary phenomenon while vowel length is an epiphenomenon. Others ... consider vowel length to be primary." The analysis of the suprasegmental effects of the two sets of morphemes turns out to be non-trivial for both stress- and length-primary approaches, so I instead indicate both surface-level stress and vowel length in the transcriptions here for completeness.

For penultimately stressed roots like *súlat* 'write' in (38), no deletion occurs, but stress placement is slightly different. The voice forms have (now penultimate) stress again assigned based on the final form, whereas the nominalizations bear word-final stress.

(38)

| *súlat* [ˈsuːlat] 'write'	Patient (*-in*)	Locative (*-an*)
VOICE PHRASE	sulátin [suˈlaːtin]	sulátan [suˈlaːtan]
DERIVED NOMINAL	sulatín [sulatˈin] ('written work')	sulatán [sulatˈan] ('writing surface')

The similarity of derived nominals to the voice-marked forms provides an interesting point of comparison, especially in light of the proposed nature of the voice morphemes under the PNA. We will see that these constructions pattern differently from voice phrases despite being minimally different from them. Crucially, they behave like bare nominals for the relevant tests, suggesting that it is this class of constructions that constitute the true participant nominalizations in Tagalog.
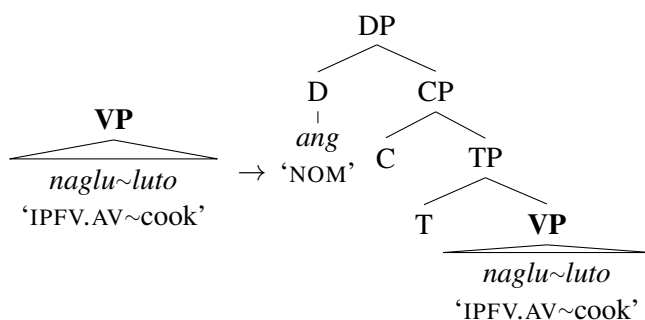
4 Asymmetry and neutralization

Let us now take a closer look at the three kinds of constructions outlined previously and see how they behave in various environments. In this section, we will see an asymmetry in behavior between voice phrases and the (bare and derived) nominals, showing that the voice phrases differ from the other two constructions with respect to some property. Furthermore, this asymmetry, and thus the difference with respect to the property in question, is consistently neutralized in the presence of functional material from the extended nominal projection. I argue that the nominal nature of the elements triggering the aforementioned neutralization gives us evidence that the cause of the asymmetry in the first place must be syntactic category.

Such a difference in behavior is unexpected under a nominalist view, where apparently verbal constructions are in fact inherently nominal (i.e., were never verbal at any point in their derivation). On the other hand, we can readily understand this difference under the verbal view, where the nominal distribution of verbal elements would stem from these elements appearing in larger nominal structures. The advantage of verbal approaches then is that they provide a unified account of the asymmetry and neutralization facts, whereas nominalist approaches would have to appeal to several properties to account for the same behavior. The difference between the two approaches is schematized below, showing the determiner *ang*.

(39) *Sketch of the nominalist approach*
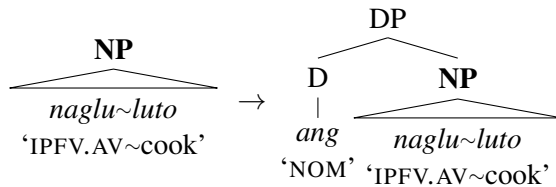
(40) *Sketch of the verbal approach*

```
                              DP
                          ╱        ╲
                         D          CP
       VP                │        ╱    ╲
    ╱──────╲      →     ang      C      TP
  naglu~luto           'NOM'          ╱   ╲
 'IPFV.AV~cook'                      T     VP
                                        ╱──────╲
                                      naglu~luto
                                     'IPFV.AV~cook'
```

To show this pattern of asymmetry and neutralization, I focus only on examples where the relevant constructions appear as clausal predicates. In this environment, phrases may unambiguously and naturally appear without overt nominal (or prepositional) functional morphology. In other environments, such as in argument position, there is a tendency or even a requirement to mark phrases with at least a determiner-like element, making it more difficult to observe any pre-neutralization behavior.[24]

4.1 Question-answer pairs

One context where we find an asymmetry between voice phrases and bare/derived nominals is in question-answer parallelism. In many cases of *wh*-questions, we expect parallelism of some sort between the question and the response. In English for example, there is parallelism between *wh*-words and the corresponding phrases in the response – *what* generally corresponds to DPs and *where* to PPs.

(41) Q: **What** is he eating?

A: (He is eating) **Some fish**.

(42) Q: **Where** did she find the book?

A: (She found the book) **Under the couch**.

Similar behavior can be found in Tagalog, shown in the examples below. A *wh*-word and its corresponding answer must match to be felicitous. In Western Malayo-Polynesian languages, argument questions have been argued to involve pseudocleft structures (Aldridge 2002 for Tagalog; Paul 2001, Potsdam 2006 for Malagasy). In (43) for example, an *ang*-marked *wh*-word appears as a predicate and an *ang*-marked headless relative appears in subject position. These questions require DPs as answers. On the other hand, oblique *wh*-questions like (44) have been argued to involve a different sort of movement, such as regular *wh*-movement (Aldridge 2002) or some other kind of fronting (Kroeger 1993).[25] These in turn require oblique-marked responses. Mismatches are judged as infelicitous, especially when answers are fragments (boldface). That is, (44b) is an infelicitous answer to (43a), and (43b) to (44a).

---

[24]In addition to the two environments discussed here, earlier versions of this paper also discussed coordination as an environment where the relevant asymmetry and neutralization was observable. However, as an anonymous reviewer points out, there is a significant amount of evidence to suggest that some distinction along the lines of viewpoint aspect might better capture the Tagalog facts. Nevertheless, the Tagalog coordination facts turn out to be independently interesting, especially in a cross-linguistic context, and so they are left for future work.

[25]Kaufman (2009, 34) also implicitly adopts a pseudocleft vs. movement split between argument and oblique questions for the PNA.

(43)   a.   Q: **Sino**   ang   b<in>igy-an   niya   ng   susi?
                  who.NOM   NOM   <PFV>give-LV   3SG.GEN   GEN   key

                  'Who did he give a key to?' —→ 'The one he gave a key to is who?'

       b.   A: **Si**   **Rica** (ang   b<in>igy-an   niya   ng   susi).
                  NOM   Rica   NOM   <PFV>give-LV   3SG.GEN   GEN   key

                  '(It was) Rica (who he gave a key to).'

(44)   a.   Q: **Kanino**   niya   i-b<in>igay   ang   susi?
                  who.OBL   3SG.GEN   CV-<PFV>give   NOM   key

                  'To whom did he give the key?'

       b.   A: **Kay Rica** (niya   i-b<in>igay   ang   susi).
                  OBL   Rica   3SG.GEN   CV-<PFV>give   NOM   key

                  '(It was) to Rica (that he gave the key).'

Here, we focus on what will be referred to as *fully congruent answers*. These answers are those that differ from their questions only with respect to the constituent corresponding to the *wh*-element (i.e., the sentence-initial predicate phrase).

Under a nominalist view, questions asking *what* should potentially have (at least) three kinds of fully congruent answers corresponding to the three construction types under discussion. Since voice phrases and bare and derived nominals are all nominal, a reasonable expectation under this view would be that these three construction types should be well-formed answers to *what*-questions.

A verbal view of Tagalog on the other hand predicts that voice phrases should be ill-formed as part of a fully congruent answer to a *what*-question. That is, a verbal answer to a question requiring a nominal answer should be ill-formed.[26]

The examples in (45-46) below show that the behavior predicted by the verbal view is what we find in Tagalog. As before, free translations that reflect the intuition behind the PNA are provided in an effort to highlight the predictions it makes. Predicates are typeset in boldface.

(45)   Q: **Ano**   ang   kina~kain   ni   Kim?
            what.NOM   NOM   IPFV~eat[PV]   GEN   Kim

            'The thing that Kim is eating **is what**?' (i.e., 'What is Kim eating?')

            ⟪'Kim's eaten-thing is what?'⟫

(46)   a.   A: <u>**Luto** ni</u>   **Harvey** ang   kina~kain   ni   Kim.
                  cook   GEN   Harvey   NOM   IPFV~eat[PV]   GEN   Kim

                  'What Kim is eating is Harvey's cooking.'   *Bare Nominal*

                  ⟪'Kim's eaten-thing is Harvey's cooking.'⟫

       b.   A: <u>**Lutu-in** ni</u>   **Harvey** ang   kina~kain   ni   Kim.
                  cook-PN   GEN   Harvey   NOM   IPFV~eat[PV]   GEN   Kim

                  'What Kim is eating is Harvey's cuisine.'   *Derived Nominal*

                  ⟪'Kim's eaten-thing is Harvey's cuisine.'⟫

---

[26]An alternative explanation available to the verbal view would be that answers with voice phrase predicates have a different information structure from those with bare or derived nominal predicates, likely due to differences in clause structure. Such an explanation less clearly available to the PNA, as the explicit assumption is that most clauses in Tagalog have the same inherently copular structure, allowing for the potential exceptions mentioned earlier in fn. 18.

    c. # A: **Ni-luto**   **ni** **Harvey** ang kina~kain   ni   Kim.
                 PFV-cook[PV] GEN Harvey NOM IPFV~eat[PV] GEN Kim

                'Harvey cooked what Kim is eating.'   *Voice Phrase*

                ⟪'Kim's eaten-thing is Harvey's cooked-thing.'⟫

    The fully congruent answers predicated by both bare and derived nominals (46a-b) are well-formed responses to the question asking *ano* 'what' in (45). On the other hand, a similar answer containing a bare voice phrase (46c) is infelicitous.[27] With fragment answers (boldface), bare and derived nominals remain felicitous, but the voice phrase is even more infelicitous than its full clause counterpart.[28]

    That the asymmetry in (46) is due to a categorial difference is further supported by the way in which this asymmetry can be neutralized. As outlined at the beginning of this section, the contrast between voice phrases and bare and derived nominals shown in (46) is neutralized when these constructions appear in clearly nominal contexts. The examples below show another set of answers for the question in (45), repeated as (47). Note that all the responses in (48) are felicitous (as full clauses or as fragment answers), despite being minimally different from (46) in terms of the presence of *ang*, *mga*, or *isa* marking the predicate.[29]

(47) Q: **Ano**   ang kina~kain   ni   Kim?
          what.NOM NOM IPFV~eat[PV] GEN Kim

         'What is Kim eating?'

         ⟪'Kim's eaten-thing is what?'⟫

(48) a. A: {**Ang /Mga /Isa=ng**} <u>luto</u> **ni** **H.** ang kina~kain   ni   K.
             NOM PL   one=LK cook GEN H. NOM IPFV~eat[PV] GEN K.

           'What Kim is eating is Harvey's cooking.'   *(cf. 46a)*

          ⟪'Kim's eaten-thing is Harvey's cooking.'⟫

    b. A: {**Ang /Mga /Isa=ng**} <u>lutu-in</u> **ni**   **H.** ang kina~kain   ni   K.
             NOM PL   one=LK cook-PN GEN H. NOM IPFV~eat[PV] GEN K.

           'What Kim is eating is Harvey's cuisine.'   *(cf. 46b)*

          ⟪'Kim's eaten-thing is Harvey's cuisine.'⟫

    c. A: {**Ang /Mga /Isa=ng**} <u>ni-luto</u>   **ni** **Harvey** ang kina~kain   ni   Kim.
             NOM PL   one=LK PFV-cook[PV] GEN Harvey NOM IPFV~eat[PV] GEN Kim

---

[27]Speakers' intuitions regarding answers like (46c) are that they somehow do not answer or otherwise ignore the question. These intuitions line up with the respective non-bracketed English translations, and suggest that these sentences that do not match up information-structurally with the respective questions.

[28]Note, however that the question below is fine with derived nominal and voice phrase answers. I tentatively take this example to be similar to examples like English *What did you do?*, which are less restrictive in the kinds of answers they can accept. Interestingly, the voice phrase but not the derived nominal answer is infelicitous in a fragment answer (omitting *iyan* 'that'), mirroring the pattern in (46).

(iii) Q: Ano   ito?   (iv) A: <u>Lagáy-an</u> ko   (iyan)   ng asukal.
          what.NOM this.NOM   put-LN   1SG.GEN that.NOM GEN sugar

        'What is this?'   'That's my container of sugar.'

                                    (v) A: <u>Nila~lagy-án</u> ko   #(iyan)   ng asukal.
                                            IPFV~put-LV 1SG.GEN that.NOM GEN sugar

                                      'I put sugar in that.'

[29]The difference in meaning caused by the choice of nominal element is subtle, so in the interest of space, I simply give rough translations into English. (48c) should give a general sense of what goes on with the other two.

'What K. is eating {is what/are ones that/is one that} H. cooked.'

⟨⟨'Kim's eaten-thing is Harvey's cooked-thing.'⟩⟩   *(cf. 46c)*

Under the verbal view, a straightforward way to explain these facts is to assume that *ang*, *mga*, and *isa* are part of the nominal extended projection (in the sense of Grimshaw 2000). The fact that we find no change in felicitousness for bare and derived nominals with or without the relevant functional material ((46a-b) vs (48a-b)) is consistent with these constructions being nominal in structure. That is, bare and derived nominals should remain nominal when selected by nominal extended heads (such as D^0). On the other hand, we see a change in felicitousness for voice phrases with the addition of the same nominal extended heads ((46c) vs (48c)). In this case, I argue that the verbal structure of the voice phrase is embedded within a larger nominal structure as previously discussed in (40).[30] The predicates in (48) are thus uniformly nominal overall regardless of their internal structure, and the original difference in felicitousness is eliminated.

For a nominalist approach, it is not immediately clear how this pattern of asymmetry and neutralization can be captured. To illustrate the problem: the data in (46) suggests that the asymmetry must be tied somehow to the choice between bare and derived nominal on one hand and voice phrase on the other. Given that nominalism holds that there is no categorial difference between nominal and verbal elements in Tagalog, syntactic category cannot be used to explain the relevant behavior. Furthermore, an explanation appealing to larger clause-structural differences between the sentences is unlikely. As syntactic potential is partially determined by category, a natural consequence for nominalism is to have similar or identical clause structure between putatively nominally and verbally predicated clauses (as is made explicit in the PNA). The difference between sentences in (46) is therefore most likely tied to the semantics of the predicates themselves. However, the sentences in (48) show us that potential semantic differences must be neutralized by *ang*, *mga*, and *isa*.

A reviewer suggests that one potential semantic difference that can be used to explain the asymmetry is the distinction between individual- and event-denoting predicates. The logic is that (45) requires an individual-denoting answer, which is provided by (46a-b) but not (46c). In support of this claim, the reviewer notes that the felicity judgements in (46,48) are flipped in response to the question in (49). That is, the voice phrase answer is felicitous, while the bare and derived nominal answers and the answers with *ang*, *mga*, and *isa* are not.

(49)   **Ano**   ang nangyari?
       what.NOM NOM happen.PFV
       'What happened?'

This alternative is problematic for a few reasons. The first is that the answers in (46,48) are not fully congruent answers (in the sense defined here) to (49). It is thus not clear that the felicity judgements in the answers to the original question (45) are due to this proposed event–individual distinction.

If we consider congruent answers such as those below, we see the second problem. (50a) shows that the voice phrase is clearly individual-denoting, as it results in strange meanings when used with *nangyari* 'happened'. The voice phrase therefore cannot be felicitous in a fully congruent answer. In comparison, a clearly event-denoting gerund-like construction in the same environment is felicitous (50b).

---

[30]One concrete implementation of this might be an extension of Kornfilt and Whitman's (2011) Functional Nominalization Thesis. Concrete details of how this might work are left for future research, as the focus of this paper is to argue that the noun-verb distinction in Tagalog is active.

(50)   a.   # **(Ang)** <u>**Ni-luto**</u>   **ni**   **Harvey** ang   nangyari.
            NOM PFV-cook[PV] GEN Harvey NOM happen.PFV

            'Harvey cooked what happened.'

            'What happened is what Harvey cooked.'

            ⟨⟨'What happened is Harvey's cooked-thing.'⟩⟩

       b.   **(Ang) Paglu~luto ni**   **Harvey** (**ng**   **manok**) ang   nangyari.
            NOM GER~cook GEN Harvey   GEN chicken NOM happen.PFV

            'What happened was Harvey's cooking (of chicken).'

While there is something intuitively event-like about voice phrases (specifically bare ones), it is not immediately clear how such a property can be captured in an analysis such as the PNA, which explicitly argues that they are individual-denoting. One possibility might be to have the property responsible for the event properties of voice phrases be somehow internal to the semantics or otherwise marked with a feature (say [+event]). In this way, the construction as a whole can retain its individual-denoting status while having properties relating to events at the same time. However, such a proposal needs to be spelled out more completely, particularly so that it is clear that the relevant property is not simply a surrogate for a noun-verb distinction.

In summary, this section provided the first piece of evidence in support of a noun-verb distinction in Tagalog by showing an environment in which a particular pattern of asymmetry and neutralization can be found. We saw that bare and derived nominals are grammatical as predicates of responses to questions asking *what*, while bare voice phrases are not. We further saw that placing these three constructions in clearly nominal structures neutralizes this asymmetry, making them uniformly grammatical as the predicates of said answers. The next subsection shows another environment where we find this pattern, suggesting that it is a more general phenomenon.

4.2   Embedded infinitives

Continuing the general strategy of finding environments that distinguish between the three different construction types, we turn to some data from Richards (2009b). Certain constructions in Tagalog embed clause-like objects whose verbs bear voice marking but not aspect marking.[31] I pretheoretically refer to these forms as infinitives, despite the fact that Tagalog marks aspect and not tense. Initial examples are given below with the matrix verb *gusto* 'want'.[32]

(51)   a.   <u>**Nagba~basa**</u> si   Mark ng   libro.
            IPFV.AV~read NOM Mark GEN book

            'Mark is reading books.'

---

[31]Alternatively, the verbs in (51b, 52b) can be viewed as bearing some dependent value of aspect, in the sense that this value is not generally found in independent clauses, outside of a few cases, like imperatives.

[32]These examples are given with embedded agents indicated, but the remaining discussion will focus on control cases (as discussed in Sec. 3.2), as judgements for these are clearer. Potentially significant details about the non-control examples (those with overt embedded agents) will be provided in footnotes. Note also that for the control cases the controlled gap may correspond to the *ang*-marked DP, as with (51b). Such cases require additional explanation under the PNA, as it is not clear for these what would serve as the subject (i.e., the *ang*-marked DP) of the copular clause, nor is it clear what predicate(s) would hold of this subject. This is particularly a problem for the approach to multi-clausal structure discussed in Sec. 5.2.

    b. Gusto niya=ng **mag-basa** (si Mark) ng libro.
       want 3SG.GEN=LK AV-read NOM Mark GEN book
       'She wants (Mark) to read books.'

(52) a. **K\<in\>ain** ni Miriam ang manok.
       \<PFV\>eat[PV] GEN Miriam NOM chicken
       'Miriam ate the chicken.'

    b. Gusto niya=ng **kain-in** (ni Miriam) ang manok.
       want 3SG.GEN=LK eat-PV GEN Miriam NOM chicken
       'He wants (Miriam) to eat the chicken.'

Richards (2009b) notes that clauses with putatively non-verbal predicates can also be embedded in these contexts. The major difference is that non-verbally predicated embeddings must appear with the element *maging*, which is ungrammatical with verbally predicated embeddings, as shown in (55).[33]

(53) a. Guro si Fe.   b. Gusto ko=ng ***(maging)** guro.
       teacher NOM Fe   want 1SG.GEN=LK AV.be teacher
       'Fe is a teacher.'   'I want to be a teacher.'

(54) a. Payat si Jun.   b. Gusto ko=ng ***(maging)** payat.
       thin NOM Jun   want 1SG.GEN=LK AV.be thin
       'Jun is thin.'   'I want to be thin.'

(55) Gusto ko=ng **(*maging)** mag-basa ng libro.
     want 1SG.GEN=LK AV.be AV-read GEN book

    'I want to read books.'[34]

---

[33]The judgements for the non-control versions of (53b, 54b) are slightly more complicated. Omitting *maging* in these examples is not ungrammatical for some speakers, although doing so causes a change in meaning, such that the sentence conveys wanting some state without referring to any change of state. The paraphrases with *to be the case* attempt to capture this reading.

(vi) Gusto ko=ng   [?](maging) guro si Fe.
     want 1SG.GEN=LK AV.be teacher NOM Fe

    'I want {it to be the case that Fe is / Fe to be} a teacher.'

(vii) Gusto ko=ng   [?](maging) payat si Jun.
     want 1SG.GEN=LK AV.be thin NOM Jun

    'I want {it to be the case that Jun is / Jun to be} thin.'

Richards (2009b) notes a similar effect with imperfective aspect on the verbal examples (51b, 52b). He notes that examples like (viii) can be hard for speakers to get, although adding certain adverbials may help.

(viii) Ayo=ko na=ng natu~tulog si Mark (tuwing umu~uwi ako).
     NEG.want=1SG.GEN now=LK IPFV~sleep NOM Mark whenever IPFV.AV~go.home 1SG.NOM

    'I no longer want Mark to be sleeping (whenever I {come home / am on my way home}).'

                                              (Richards 2009b, 186, modified)

A reviewer suggests that one natural way of understanding this alternation with respect to *maging* and aspect is in terms of the size of the embedded constituent (e.g., CP vs *v*P). With larger embeddings, the embedded predicate is not directly selected, and so does not need to be infinitival or occur with *maging*.

[34]A potential alternative explanation for the ungrammaticality in this example might be the presence of two infinitival verbs. This does not seem to be the issue here, as replacing *mag-basa* with an aspect-marked form (*nag-basa*, *nagba~basa*, or *magba~basa*) does not alleviate the ungrammaticality.

Building on this initial picture from Richards (2009b), I show in what follows that the behavior of *maging*[35] follows the same pattern of asymmetry and neutralization presented previously. I then discuss other evidence from this environment that supports a view where the point at which these constructions acquire their nominal distribution (i.e., as headless relatives) must be syntactically high.

First, I extend Richards (2009b) by showing that derived nominals behave like bare nominals and not voice phrases. This is shown in the pairs of examples below (56-57). As with bare nominals (53b), *maging* is obligatory with derived nominals (56b, 57b), replicating what we have seen thus far. Furthermore, the fact that the voice phrase examples are not marked with aspect[36] makes the difference between voice phrases and the corresponding derived nominals even more strikingly small. Whereas one might reasonably suspect that aspect has some role to play in the asymmetry found in the question-answer pairs, such a position would be harder to defend in this case. It thus appears that the difference in the grammaticality of *maging* is tied to the difference between two morphemes that perform the same function conceptually: *taga-* vs *mag-*, picking out agents, and *-an* (LN) vs *-an* (LV), picking out locations.

(56)   a.   Gusto ko=ng   **(*maging)** <u>mag-luto</u> ng   gulay.
            want   1SG.GEN=LK   AV.be   AV-cook GEN vegetable
            'I want to cook vegetables.'

      b.   Gusto ko=ng   ***(maging)** <u>taga-luto</u> ng   gulay.
            want   1SG.GEN=LK   AV.be   AN-cook GEN vegetable
            'I want to be a vegetable cook.'

(57)   a.   Gusto ko=ng   **(*maging)** <u>lagy-án</u> ng   pera   ang   kaho=ng ito.
            want   1SG.GEN=LK   AV.be   put-LV   GEN money NOM box=LK   this
            'I want to put money in this box.'

      b.   Gusto ko=ng   ***(maging)** <u>lagáy-an</u> ng   pera   ang   kaho=ng ito.
            want   1SG.GEN=LK   AV.be   put-LN   GEN money NOM box=LK   this
            'I want this box to be a money container.'

We also find neutralization under a clearly nominal context, similar to what we have seen previously. When these embedded predicates are marked with *isa* 'one', *maging* becomes uniformly required across the three construction types. Compare (58-59) to (56-57). A bare nominal example (60) is provided for completeness.[37],[38]

---

[35]Richards (2009b) provides an analysis of *maging* as the overt infinitival form of the Tagalog copula whose form in matrix contexts is null. While *maging* is typically translated in isolation as 'become', I follow Richards in glossing *maging* as 'be'. That said, the discussion in this section will remain agnostic as to the analysis of this element. Specifically, I do not assume that Tagalog makes use of a null (verbal) copula. This assumption would mean that nominally- and adjectivally-predicated clauses are in fact verbally predicated, in turn invalidating the explanation proposed for the asymmetry and neutralization discussed. I instead assume (tentatively) that nominal and adjectival constituents are themselves able to serve as clausal predicates. Reconciling the analysis Richards proposes with the proposal here is left for future work.

[36]An alternative view is that they are marked for some deficient or semantically non-contentful value of aspect. For example, Himmelmann (2008) refers to this form of the verb as irrealis perfective.

[37]*Isa* 'one' is the only neutralizer of the three that is available in this environment. Determiners mark DP arguments, not embedded predicates, of verbs like *gusto* 'want'. Compare (53b) to (ix) below. It is also generally strange to mark such predicates as plural with *mga*, even if the relevant other parts of the sentence match in number. This time, compare (53b) to (x).

(ix)   Gusto ko   {ang / ng} <u>guro</u>.   (x)   Gusto nila=ng   maging ([?]mga) <u>guro</u>.
      want   1SG.GEN NOM/GEN teacher   want   3PL.GEN=LK AV.be   PL   teacher
      'I want {the / a} teacher.'   'They want to be teachers.'

[38]At least one speaker consulted judges (58a, 59a) as bad, regardless of whether or not *maging* is present. I suspect that these

(58)   a.   Gusto ko=ng   ***(maging) isa=ng** n̰aglu~luto   ng   gulay.
            want 1SG.GEN=LK   AV.be   one=LK IPFV.AV~cook GEN vegetable

            'I want to be one who cooks vegetables.'   *(cf. 56a)*

       b.   Gusto ko=ng   ***(maging) isa=ng** t̰aga-luto ng   gulay.
            want 1SG.GEN=LK   AV.be   one=LK AN-cook GEN vegetable

            'I want to be a vegetable cook.'   *(cf. 56b)*

(59)   a.   Gusto ko=ng   ***(maging) isa=ng** n̰ila~lagy-án ng   pera   ang kaho=ng ito.
            want 1SG.GEN=LK   AV.be   one=LK IPFV~put-LV GEN money NOM box=LK this

            'I want this box to be something in which money is put.'   *(cf. 57a)*

       b.   Gusto ko=ng   ***(maging) isa=ng** l̰agáy-an ng   pera   ang kaho=ng ito.
            want 1SG.GEN=LK   AV.be   one=LK put-LN   GEN money NOM box=LK this

            'I want this box to be a money container.'   *(cf. 57b)*

(60)   Gusto ko=ng   ***(maging) isa=ng** d̰oktor.
       want 1SG.GEN=LK   AV.be   one=LK doctor

       'I want to be a doctor.'

This instance of asymmetry and neutralization is again easily understood as a neutralization of syntactic category from verbal to nominal. The addition of *isa* 'one' also triggers another change in embedded voice phrases that I argue further supports a view where these are verbal constructions that become nominal high in the structure. Whereas aspect cannot appear on the embedded voice phrases in (56a, 57a) without *isa*, aspect must appear on the ones in (58a, 59a), where *isa* is present. This can be straightforwardly understood as *isa* requiring a complement of a particular size, in this case, minimally containing an Aspect projection. Another way of framing this is that in the presence of *isa*, the matrix verb no longer selects the voice phrase as its complement, leading to the infinitive form no longer being licensed on the latter. As a result, the voice phrase must be marked for a non-dependent value of aspect.[39]

In comparison, this data is problematic for a nominalist analysis like the PNA, as we again see the asymmetry and neutralization pattern familiar from the question-answer pairs in the previous section. As argued, it is hard to see how a nominalist analysis can account for this pattern in a general way.

One possibility suggested by Dan Kaufman (p.c.) is that the data discussed in this section should instead be understood as reflecting a distinction between stative (i.e., bare/derived nominal) and non-stative (i.e., voice phrase) predicates. The crucial observation here is that *maging* must mean non-stative 'become' (instead of stative 'be') in matrix contexts (see Richards 2009b). The idea is then that predicates like *gusto* 'want' have a selectional requirement for their complements to be non-stative, and marking bare and derived nominals with *maging* changes them from stative to non-stative.[40]

_____

sentences may come across as unnecessarily verbose without the proper context.

   [39]While not addressed directly here, the data presented in this portion of the paper is consistent with a verbal analysis of *maging* (as in Richards 2009b). This then predicts that it should pattern like the voice phrases in this section. This is what we find. Compare to (56b) the grammatical, if verbose and redundant, sentence below, where *isa* marks the entire embedded constituent, including *maging*. The original *maging* (italics) appears with aspect marking, and another obligatory *maging* (boldface) surfaces.

   (xi)   Gusto ko=ng   ***(maging)** isa=ng *magiging* t̰aga-luto ng   gulay   si   Martin.
          want 1SG.GEN=LK AV.be   one=LK FUT.AV.be AN-cook GEN vegetable NOM Martin

          'I want Martin to be someone who will become a vegetable cook.'

   [40]Although compare fn. 33. Also, this requirement presumably applies to other cases that take infinitive voice phrase forms in Tagalog, of which examples were given in Sec. 3.2.

While such an approach does capture the distribution of *maging* in the data considered so far, it does not explain the distribution of aspect forms, as shown in (61). If voice phrases and *maging* are non-stative by assumption, they should fulfill the proposed selectional requirement regardless of the specific aspectual marking they bear, contrary to what we see in (61a) and (61b), respectively. This data thus shows that a non-finiteness requirement must be postulated anyway in addition to non-stativity.

(61) a. Gusto ko=ng   { mag-basa / *nag-basa / *nagba~basa} ng   libro.
          want 1SG.GEN=LK AV-read   AV.PFV-read AV.IPFV~read GEN book
          'I want to read a book.'

    b. Gusto ko=ng   {maging / *naging   / *nagiging} bumbero.
          want 1SG.GEN=LK AV.be   AV.PFV.be AV.IPFV.be firefighter
          'I want to be a firefighter.'

Furthermore, the stativity approach must still make reference to syntactic category to explain the behavior of another type of predicate: bare stative verbs. Some stative predicates in Tagalog can appear without aspect or voice morphology in matrix declaratives (Himmelmann 2008, Richards 2009b). When such predicates do bear voice morphology, they take on a non-stative meaning. This alternation is shown in (62) for *alam* 'know' vs *nalaman* 'found out'.[41] Crucial for the discussion here is that such predicates must appear in the non-stative (or voice-marked non-finite) form when embedded under *gusto*, as in (63).

(62) a. Alam ko   ang sagot.
          know 1SG.GEN NOM answer
          'I know the answer.'   (Richards 2009b, 186)

    b. N-alam-an   ko   ang sagot.
          PFV.AIA-know-LV 1SG.GEN NOM answer
          'I found out the answer.'   (Richards 2009b, 187)

(63) Gusto ko=ng   {m-alam-an   / *alam} ang sagot.
       want 1SG.GEN=LK AIA-know-LV   know NOM answer
       'I want to know the answer.'   (Richards 2009b, 186, modified)

While the appearance of a non-stative form in the complement of *gusto* in (63) appears to support the idea that these environments crucially require non-stative predicates, we see that there are two ways to fulfill this requirement: with *maging* as in (61b) or with a differently marked form as in (63). These two strategies are, generally speaking, in complementary distribution. For example, (64) shows that *alam* cannot take *maging* without a change in meaning and the loss of the control structure.

(64) * Gusto ko=ng   maging alam ang   sagot.
        want 1SG.GEN=LK AV.be   know NOM answer
        Intended: 'I want to know the answer.'
        Grammatical as: 'I want the answer to become known.'

Thus, capturing the distribution of *maging* with respect to predicates like *doktor* and those like *alam*, requires reference to a property other than stativity. The most straightforward such property is syntactic category,

---

[41]Note that *gusto* 'want' also falls under this class of predicates.

but appealing to such a difference in this case is fundamentally incompatible with nominalist analyses of Tagalog.[42]

In contrast, adopting a verbal analysis allows us to explain this distribution fairly straightforwardly. I assume that predicates like *gusto* 'want' (and other similar environments) require non-finite complement clauses, and that non-finiteness must be morphologically expressed. Verbally predicated clauses (i.e., bare voice phrases and bare stative verbs) can straightforwardly satisfy these requirements since they have non-finite forms, as shown in (61a) and (63). On the other hand, nominally predicated clauses (i.e., bare/derived nominals and headless relatives) cannot themselves satisfy these requirements, since these predicates do not have non-finite forms. Instead, such cases require the insertion of a semantically vacuous dummy *maging*, akin to *do*-support as suggested by Richards (2009b).[43] Note that assuming morphologically marked non-finite forms is independently needed to account for the bare stative verbs.

To summarize this section, I have built on data from Richards (2009b) and shown that as the predicates of embedded infinitives, voice phrases behave differently from bare and derived nominals with respect to the element *maging*. As with the question-answer pairs, this difference in behavior becomes neutralized when the relevant constructions are marked with nominal functional morphology. I argued that this behavior is evidence that syntactic category is the specific property that *maging* is sensitive to. This was further supported by the behavior of bare stative verbs.

Overall, then, these environments provide support for a counterargument to the nominalist approach by showing that voice phrases, specifically *bare* ones, show differences in their external distribution as compared to bare and derived nominals. This data also illustrates that the widely reported nominal distribution of voice phrases can be attributed to the presence of nominal functional morphology, creating what I refer to here as a headless relative.

It should be clear from this discussion that the core disagreement with a nominalist approach is not the (eventual) nominal nature of the relevant voice phrases, but the point at which such constructions become nominal. Under the PNA specifically, voice phrases must be nominal throughout their structure. On the other hand, I argue here that voice phrases must have internal verbal structure, and are instead "nominalized" (take on nominal properties) later in their derivation. I take this "nominalization" to be caused by the introduction of nominal functional morphology in a manner similar to that proposed for event nominalizations and deverbal nouns (e.g., Kornfilt and Whitman 2011), where nominal heads directly select verbal projections.

Allowing nominal heads to directly select verbal projections accounts for the categorial change argued for in this section. A lingering question that is arguably separate from the issue of syntactic category,

---

[42]A reviewer argues that *maging* is in fact an affix bearing a value for voice. The reviewer notes that (a) *maging* does not consistently behave like a separate word (e.g., for second position clitic placement), and that (b) other voice forms of *maging* are attested (e.g., CV *i-paging*, see Schachter and Otanes 1972, Ch.5.27). Given this, the distribution of *maging* might have a morphological explanation, whereby a constraint against double voice marking prevents *maging* from marking voice phrases. However, data like (63-64) remains problematic, since there is no morphological ban *per se* on *maging* marking at least some bare stative verbs.

[43]As discussed in fn. 35, this crucially cannot be the non-finite form of a null copula. The proposed semantic vacuity may be supported by the ambiguity of data like (xii), which is true if the speaker is *either* a doctor wishing to quit (stative, dummy *maging*) *or* a medical student wishing to drop out (non-stative, contentful *maging*). Notably, preliminary investigation into the behavior of other stative predicates, like *alam* 'know' in (63), suggests that this ambiguity is strongest, if not only available, with *maging*.

(xii) Ayo=ko   na=ng   maging doktor.
       NEG.want=1SG.GEN now=LK AV.be   doctor

       'I don't want to {be/become} a doctor anymore.'   (Richards 2009b, 183)

but is worth considering briefly, is how nominalized voice phrases get the correct denotation. The PNA posits that voice morphology fulfills this role. In comparison, verbal approaches are more amenable to using movement, as with relativization. However, Tagalog's lack of clear relativizers or relative pronouns necessitates positing null material, which is a general criticism against verbal analyses of Tagalog. The issue of null material is admittedly not addressed here in much detail, but it is worth noting that this is not a problem specific to Tagalog. Other languages that do not have overt relative pronouns exist (e.g., Japanese), so positing null relative operators or pronouns does not seem unmotivated. Null material notwithstanding, the issue of movement is discussed in Section 5.[44]

5   The problem of long distance dependencies

In this section, I consider in more detail the implications of a major theoretical difference highlighted in this paper between the nominalist and verbal views. Specifically, we will see that the point in the derivation where Tagalog voice phrases acquire their nominal characteristics has consequences for constructions involving long-distance A-bar dependencies. I follow up on responses by Aldridge (2009) and Richards (2009a) to Kaufman's (2009) Paricipant Nominal Analysis arguing that this analysis makes incorrect predictions regarding long-distance A-bar dependencies. Thus while a major strength claimed of the PNA is that it naturally derives many major points of Tagalog syntax (recall Sec. 2.3), this approach appears to not fit as naturally when we consider more complex phenomena.

I will first show that long-distance dependencies are indeed attested in Tagalog, and that these involve movement. I then consider a few possibilities for accounting for embedded clauses in general under the PNA, as the proposal by Kaufman (2009) does not explicitly spell out how to analyze such constructions.

Three structures, schematized in (66-68), will be considered as potential analyses of embedded clauses. These are the structures on which predictions regarding long distance dependencies will be evaluated. As an aid to the reader, the matrix predicate in each structure is underlined, and an example sentence (65) is given with parts relevant to the schemata labeled 1, 2, and 3.

(65) Na-balita-an(**1**)   ni   Gina na b<um>ili(**2**)   ang   pulis(**3**) ng   tela.
        PFV.AIA-news-LV GEN Gina LK <PFV.AV>buy NOM police   GEN cloth
        'Gina heard that the police officer bought cloth.'

(66) $[_{TP}$ $[_{DP\text{-}pred}$ **Pred1** $[_{CP}$ *na* $[_{TP}$ **Pred2** . . . T **DP3** $]$ $]$ $]$ T *pro$_{subj}$* $]$ *Complex NP*

        $\approx$'It was Gina's news that the police officer bought cloth.'

(67) $[_{TP}$ $[_{DP\text{-}pred}$ $[_{NP}$ **Pred1** . . . $]$ *na* $[_{NP}$ **Pred2** . . . $]$ $]$ T **DP3** $]$ *Non-hierarchical*

        $\approx$'The police officer was Gina's heard-(about-)thing and the buyer of cloth.'

(68) $[_{TP}$ $[_{DP\text{-}pred}$ **Pred1** . . . $]$ T $[_{CP\text{-}subj}$ *na* $[_{TP}$ **Pred2** . . . T **DP3** $]$ $]$ $]$ *CP Subject*

        $\approx$'That the police officer bought cloth was heard by Gina.'

(66) is analogous to the analysis of embedded clauses simply as CP complements of their selecting verbs, modulo the selecting head being nominal. Section 5.1 shows that this structure is immediately ruled out for reasons both internal to the PNA (no extraction out of DPs) as well as crosslinguistic (complex NPs are islands). We thus consider alternatives. (67) follows subsequent work by Kaufman (2011) proposing that

---

[44]Regarding similar issues with positing a null copula, see fn. 35.

Tagalog does not have true long-distance dependencies involving DPs, and instead forms complex predicates (e.g., combining *Gina heard x* and *x bought cloth* in (65)).[45] Section 5.2 shows that while this structure both allows for (apparent) long-distance dependencies and is motivated by a number of syntactic facts, it predicts the wrong semantics for such constructions. On the other hand, (68) takes an intuition behind Rackowski and Richards's 2005 proposal—that CPs in Tagalog must have the same syntactic status as subjects (i.e., *ang*-marked DPs) to be transparent for extraction—and translates it to the PNA. Section 5.3 argues that this structure is the least problematic for deriving long-distance dependencies under the PNA. However, I discuss in section 5.4 an overgeneration problem that this alternative faces: while voice phrases are transparent for long-distance extraction, bare/derived nominals are not.

5.1 Long-distance dependencies

Long-distance A-bar dependencies are attested in Tagalog. In the examples below, there is a dependency between a gap and an associated phrase (indicated with subscripts) that crosses at least one clause boundary.

(69)   a.   Ano$_i$ ang   s<in>abi   mo   [$_{CP}$ na b<in>ili   ni   Vic t_i]?
           what.NOM NOM <PFV>say[PV] 2SG.GEN LK <PFV>buy[PV] GEN Vic

           'What$_i$ did you say [$_{CP}$ that Vic bought t_i]?'

   b.   ang   libro$_i$=ng s<in>abi   mo   [$_{CP}$ na b<in>ili   ni   Vic t_i]
           NOM book=LK <PFV>say[PV] 2SG.GEN   LK <PFV>buy[PV] GEN Vic

           'the book$_i$ that you said [$_{CP}$ that Vic bought t_i]'

According to scholars such as Aldridge (2009), Richards (2009a), the derivation of these constructions poses problems for the PNA. These problems are rooted in movement being blocked due to phase edges: either from a DP, corresponding to Tagalog voice morphology being syntactically low (under the PNA), or from a CP, as we will see for some possible extensions to the nominalist account. To illustrate this, I will focus on relative clauses.

First, it should be established that such examples involve movement of some kind. Evidence for this comes from their being subject to island effects. The examples in (70) contain embedded *wh*-questions as compared to the embedded declarative CPs in (69) above. As in English, *wh*-questions are islands for extraction in Tagalog. An example without the long-distance dependency is given in (71) as a baseline.[46]

(70)   a.   * Ano$_i$ ang   s<in>abi   mo   [$_{CP}$ kung kailan b<in>ili   ni   Vic t_i]?
           what.NOM NOM <PFV>say[PV] 2SG.GEN   COMP when <PFV>buy[PV] GEN Vic

           'What$_i$ did you say [$_{CP}$ when Vic bought t_i]?'

   b.   * ang   libro$_i$=ng s<in>abi   mo   [$_{CP}$ kung kailan b<in>ili   ni   Vic t_i]
           NOM book=LK <PFV>say[PV] 2SG.GEN   COMP when <PFV>buy[PV] GEN Vic

---

[45] See Kaufman 2006 for discussion on long-distance dependencies of adjuncts and oblique phrases.

[46] A reviewer points out that some examples similar to (70), like (xiii), seem acceptable (perhaps because of D-linking or because (70) is bad for reasons other than islandhood). A quick check-in with consultants suggests that examples like (xiii) are generally hard to accept, but having a specific context may improve acceptability for some speakers. Detailed investigation of this seems better left for a separate paper.

(xiii)   Alin$_i$ ang alam mo   kung kanino   dapat i-bigay t_i?
       which NOM know 2SG.GEN COMP who.OBL should CV-give

       'Which$_i$ do you know [to whom t_i should be given]?'

'the book$_i$ that you said [$_{CP}$ when Vic bought t_i]'

(71)   <u>S<in>abi</u>   mo   [$_{CP}$ kung kailan <u>b<in>ili</u>   ni   Vic ang <u>libro</u>].
       <PFV>say[PV] 2SG.GEN COMP when <PFV>buy[PV] GEN Vic NOM book

       'You said [$_{CP}$ when Vic bought the book].'

Next, let us consider relative clauses with local dependencies (i.e., those where no intermediate clause boundary is crossed), such as the one below in (72), and compare the derivations under the PNA and under a verbal approach. For concreteness, I assume movement of a null relative pronoun, but any movement analysis should suffice for the following discussion.[47]
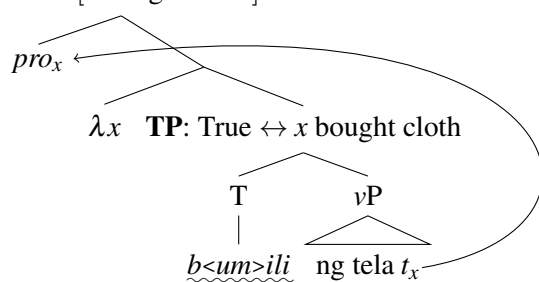
(72)   <u>pulis$_i$</u> [na <u>b<um>ili</u>   ng   tela t_i]
       police   LK <PFV.AV>buy GEN cloth

       'police officer who bought cloth'

       ⟪'police officer who was a buyer of cloth'⟫

On both approaches, we can assume that the relative pronoun originates from the same position that the corresponding full DP argument would (indicated with t_i above). For a verbal analysis, such as that given by Aldridge (2009), this position would be within *v*P, sketched in (73).[48] For the PNA on the other hand, the DP in a clause that can undergo extraction is given the status of subject (recall the basic structure given in (22)), and so the relative pronoun must originate from this position, shown in (74).[49]

Here, I adopt an approach to movement where a trace is interpreted semantically as a variable, which is subsequently bound at the landing site of the moved constituent via Predicate Abstraction (Heim and Kratzer 1998). This process produces the same end result semantically for both approaches: a predicate of individuals (semantic type $\langle e,t \rangle$) that is true of an individual x if and only if x bought cloth. The difference lies in the role of movement in generating the denotation of the construction as a whole.

(73)   *Verbal derivation of a local relative clause*

       CP: $\lambda x.[x$ bought cloth$]$

           pro$_x$

               λx **TP**: True $\leftrightarrow x$ bought cloth

                       T   *v*P

                   <u>b<um>ili</u>   ng tela t_x

(74)   *(Modified) PNA derivation of a local relative clause*

---

[47]For example, see Aldridge (2003a,b) for a head raising analysis for Tagalog relative clauses.

[48]Again for concreteness, I follow Aldridge's (2009) derivation of V-initial word order by head-moving V to T and keeping all arguments within *v*P. The crucial part (that movement originates within *v*P) should be true across most verbal analyses.

[49]The problems I discuss here should hold even if we posit a more traditional Spec-TP subject position.

CP: $\lambda x.[x$ was a buyer of cloth$]$

pro_x

λx **TP**: True $\leftrightarrow x$ was a buyer of cloth

PredP: $\lambda y.[y$ was a buyer of cloth$]$

b<um>ili ng tela   T   DP$_{Subj}$

t_x
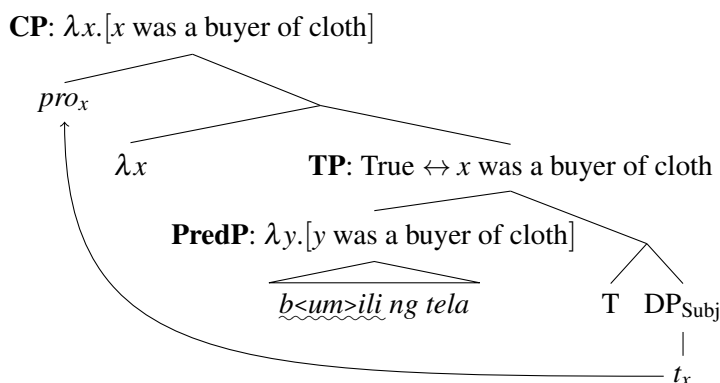
For the verbal approach in (73), movement and Predicate Abstraction is what "leaves open" one of the argument slots of the verb, in a sense. This allows the whole expression to denote the relevant participant in the event (in this case the agent). On the other hand, with the nominalist derivation in (74), the denotation of the entire expression is already independently derived lower in the structure, in the DP dominated by PredP. In this case, the relevant event participant is picked out by the semantics of the AV morpheme *<um>*, which operates close to the root *bili* 'buy', similar to how we might analyze English *-er* (see e.g., Baker and Vinokurova 2009, Keenan 2000). Note that the movement indicated in (74) is a departure from the original formulation of the PNA, and is effectively semantically vacuous: PredP and CP will always have identical denotations. Nevertheless, I assume movement for the moment to keep in line with the island effects in (70). An alternative without movement is considered in Section 5.2.

Now we can consider more complex examples where the movement originates from within an embedded clause. Under the verbal view, we can make the standard assumptions that embedded clauses are CP complements of matrix V, and that A-bar movement proceeds successive cyclically. Thus, the relative pronoun would pass through the lower Spec-CP position before landing in the matrix Spec-CP position. For the example in (75), we would have the structure in (76).

(75) pulis$_i$ na na-balita-an   ko[$_{CP}$=ng   b<um>ili   ng   tela t_i]
      police LK PFV.AIA-news-LV 1SG.GEN=LK <PFV.AV>buy GEN cloth

      'the police officer who I heard bought cloth'

(76) *Verbal derivation of a long-distance relative clause*
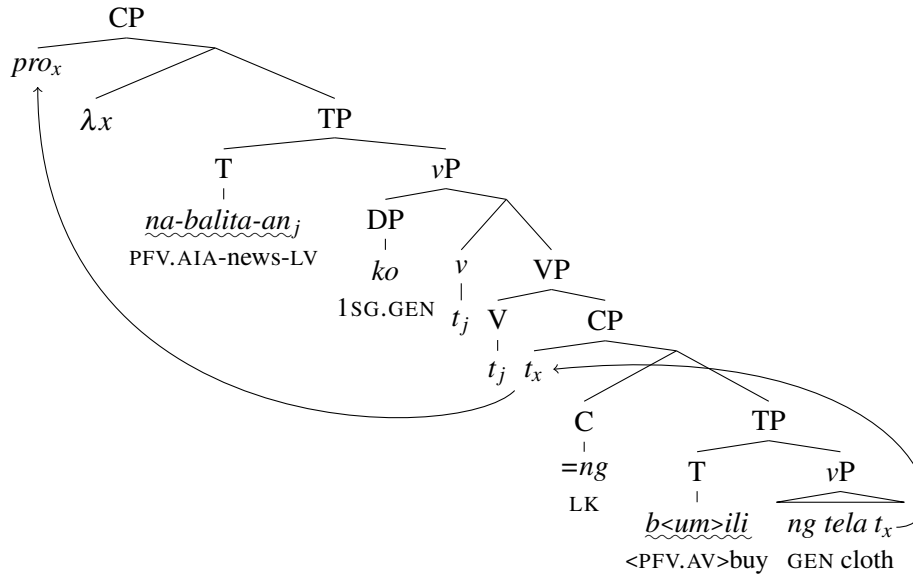
Semantically, the derivation also goes similarly, modulo the need to shift to an intensional semantics to handle embedded clauses. In (77a), the trace of the moved pronoun is interpreted as a variable (x), as before. The difference now is that this variable is bound higher in the structure, above the matrix TP (77c), since the relative pronoun has moved higher. The result is a predicate that is true of an individual x if x bought cloth in all possible worlds that are compatible with my hearsay in some reference world (usually the actual world).

(77)   a. Embedded CP: *=ng b<um>ili ng tela t_x*
       $\lambda w.[x$ bought cloth in $w]$

   b. Matrix TP: *na-balita-an ko[$_{CP}$=ng b<um>ili ng tela t_x]*
       $\lambda w.[\forall w' \in W[w'$ is compatible with what I heard in $w \rightarrow$

       x bought cloth in $w']]$

   c. Relative Clause: *(pulis) [na na-balita-an ko[$_{CP}$=ng b<um>ili ng tela t_x]]*
       $\lambda x.[\lambda w.[\forall w' \in W[w'$ is compatible with what I heard in $w \rightarrow$

       x bought cloth in $w']]]$

I now turn to a derivation of (75) following the PNA. As with the verbal analysis, I assume that the relative pronoun must originate from within the embedded CP. Kaufman (2009) does not discuss how clausal arguments are handled in his analysis, so a few possibilities will be considered here.

The first option to consider is parallel to standard assumptions about complement clauses: that they are CP complements of their selecting head. In the case of the PNA, this would place the CP close to the root *balita* (specifically inside *n*P) in a configuration similar to a complex noun phrase. (79) provides a paraphrase roughly capturing the intuition behind this structure as well as a more fleshed out tree for the structure previously schematized in (66). Certain details have been omitted from the tree in (79), but the crucial point is that the embedded CP is the complement of $\sqrt{\text{ROOT}}^0$, the base-generation site of *balita* 'news'.[50]

---

[50]See Kaufman 2009, 32–34 for discussion of the DP-internal structure shown here.

(78)   <u>Na-balita-an</u>   ko[CP=ng   <u>b<um>ili</u>   ng tela ang <u>pulis</u>].
         PFV.AIA-news-LV 1SG.GEN=LK <PFV.AV>buy GEN cloth NOM police

         'I heard that the police officer bought cloth.'

(79)   ⟪'It is [DP my news [CP that the police officer was a buyer of cloth]].'⟫
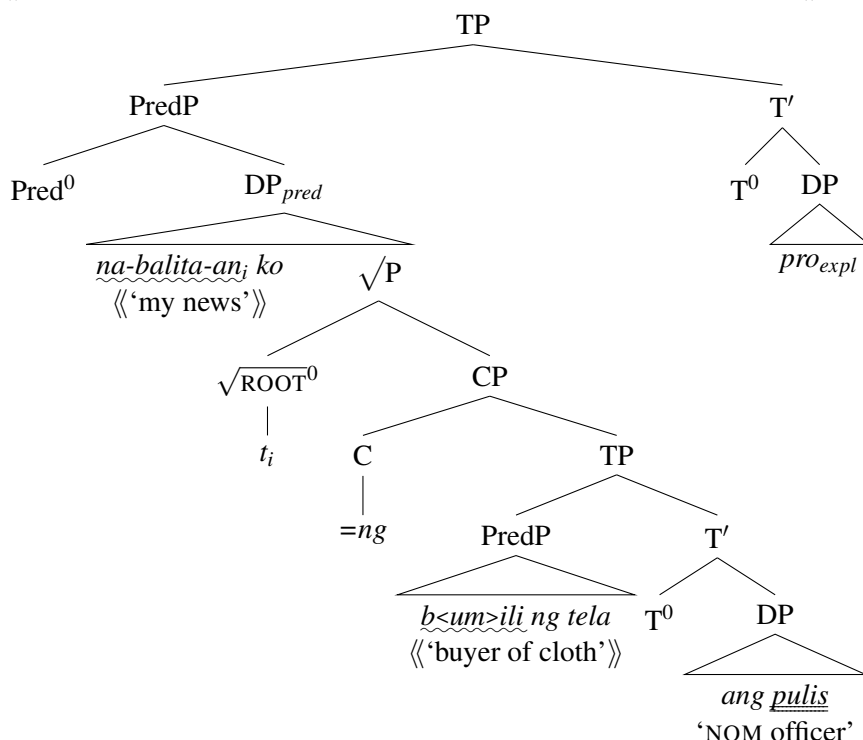
In this structure, the CP is contained within a larger DP (i.e., DP*pred*), so the relative pronoun, originating from the same position as *ang pulis*, should not be able to move to matrix Spec-CP by virtue of being in a complex NP island. Also recall from Section 2.3 that the PNA derives the Tagalog A-bar extraction restriction as a ban on extracting DPs from within larger DP structures. Given that the structure sketched in (79) incorrectly prevents the formation of long-distance dependencies by movement, let us consider some potential alternatives where the embedded CP appears structurally higher, making the extraction position more accessible. I argue that these alternatives run into the opposite problem of overgenerating unattested behavior.

5.2   Long-distance via non-hierarchical structure

One alternative is sketched by Kaufman (2011) in subsequent work arguing that Tagalog does not form true long-distance A-bar dependencies. Instead, the claim is that apparent long-distance extraction involves a flatter, non-hierarchical relationship between two NP predicates. Under this view, a long-distance relative clause would have the structure bracketed in (80).

(80)   <u>libro</u>=ng [[s<in>abi   mo   ] na [b<in>ili   ni   Vic]]
         book=LK <PFV>say[PV] 2SG.GEN   LK <PFV>buy[PV] GEN Vic

         'book that you said Vic bought'

⟪'book that was your said-thing that was Vic's bought-thing'⟫

In this structure, the non-hierarchical syntactic relationship between the "embedding" predicate *sinabi mo* and the "embedded" predicate *binili ni Vic* is also reflected in the semantics. Kaufman argues that the two predicates are of type ⟨*e,t*⟩ and compose semantically via Predicate Modification (Heim and Kratzer 1998). This approach is partially motivated by the presence of the linker morpheme *na*, which also appears in clear cases of nominal and adjectival modification, as in (21). A paraphrase reflecting the intuition behind this is given above in double angle brackets.

A number of additional facts about Tagalog are claimed to support this view. First, embedding predicates in Tagalog can take arguments that are clearly DPs, as shown in (81a). This gives plausibility to the idea that such predicates can denote objects of type ⟨*e,t*⟩. Second, embedding predicates may appear linearly after, rather than before, their embedded predicate, as shown in (81b). This is taken to be evidence for the lack of hierarchical structure between the two predicates.[51]

(81) a. Ikaw   ang s<in>abi   ko.
         2SG.NOM NOM <PFV>say[PV] 1SG.GEN

         'You're the one I said.'

      b. libro=ng [[b<in>ili   ni Vic] na [s<in>abi   mo   ]]
         book=LK <PFV>buy[PV] GEN Vic LK <PFV>say[PV] 2SG.GEN

         'book that Vic bought that you said/mentioned'

         ⟪'book that was Vic's bought-thing that was your said-thing'⟫

These observations do not seem to generalize well to other clause-embedding predicates, however. The predicate *inakala* '(mistakenly) thought' also takes a clausal complement and allows long-distance extraction, as shown in (82).

(82) a. <In>akala   mo   na b<in>ili   ni Vic ang libro.
         <PFV>think[PV] 2SG.GEN LK <PFV>buy[PV] GEN Vic NOM book

         'You thought (mistakenly) that Vic bought the book.'

      b. libro=ng [[<in>akala   mo   ] na [b<in>ili   ni Vic]]
         book=LK <PFV>think[PV] 2SG.GEN   LK <PFV>buy[PV] GEN Vic

         'book that you thought (mistakenly) that Vic bought'

         ⟪'book that was your thought-thing that was Vic's bought-thing'⟫

---

[51]In relation to the island effects pointed out in (70), the reviewer notes that *kung*-marked phrases do not function as predicates of individuals. Thus, they cannot function as a copular predicate (xiv) nor as a nominal modifier with the linker morpheme (xv), in contrast to regular voice phrases. This is taken to indicate that the examples in (70) are not instances of true island effects.

(xiv)   * [Kung kailan b<in>ili   ni Vic ang libro]Pred (ang) kahapon.
           COMP when <PFV>buy[PV] GEN Vic NOM book   NOM yesterday

        Intended: 'When Vic bought the book is/was yesterday.'

(xv)   (ang) araw (*na) [kung kailan b<in>ili   ni Vic ang libro]
       NOM day   LK   COMP when <PFV>buy[PV] GEN Vic NOM book

       '(the) day when Vic bought the book.'

However, while *inakala* allows apparent DP arguments (83a), reversing linear order with embedded material is not as well-formed (83b). Assuming the non-hierarchical structure discussed here, we predict at least that the grammaticality of (82b), whether it is well- or ill-formed, should track the grammaticality of (83b).

(83) a. Ikaw   ang <u><in>akala</u>   ko.
           2SG.NOM NOM <PFV>think[PV] 1SG.GEN
           'You're the one I thought (mistakenly).'

      b. ?* <u>libro</u>=ng [[<u>b<in>ili</u>   ni Vic] na [<u><in>akala</u>   mo   ]]
           book=LK <PFV>buy[PV] GEN Vic LK <PFV>think[PV] 2SG.GEN
           'book that Vic bought that you thought (mistakenly)'
           ⟨⟨'book that was Vic's bought-thing that was your thought-thing'⟩⟩

More generally, this approach is also problematic from a semantic perspective, as it does not place the embedded predicate within the semantic scope of the matrix predicate. To see why this is a problem, consider the baseline in (84).

(84) <u><In>akala</u>   mo   [na <u>b<in>ili</u>   ni Vic ang libro].
      <PFV>think[PV] 2SG.GEN LK <PFV>buy[PV] GEN Vic NOM book
      'You thought (mistakenly) that Vic bought the book.'

Standard analyses of clause-embedding (or more generally intensional) predicates (von Fintel and Heim 2011, Heim and Kratzer 1998) derive their semantics by assuming semantic scope of the matrix predicate over its complement. This scope allows the semantic content of *inakala* 'thought (mistakenly)' to operate on the proposition denoted by *binili ni Vic ang libro* 'Vic bought the book' producing truth conditions which are true only if you (the addressee) have a belief that Vic bought the book. This scope relationship should be relatively unaffected by A′-movement (recall the derivation in (77)).

The non-hierarchical analysis, however, assumes a flat structure for apparent long-distance A′-dependencies, where the "matrix" and "embedded" predicates are argued to compose via Predicate Modification (or similar operation). Taking this semantics at face value and applying it to (80), we get the result shown in the derivation in (85).

(85) a. "Embedded" Predicate: <u>*b<in>ili ni Vic*</u>
        $\lambda x.[\lambda w.[\text{Vic bought } x \text{ in } w]]$

      b. "Matrix" Predicate: <u>*<in>akala mo*</u>[52]
        $\lambda x.[\lambda w.[\text{You thought/made an assumption about } x \text{ in } w]]$

      c. Predicate Modification: *[<u><in>akala mo</u>] na [<u>b<in>ili</u> ni Vic]*
        $\lambda x.[\lambda w.[\text{You thought/made an assumption about } x \text{ in } w \land$

                                        Vic bought x in $w]]$

---

[52]Note that a more typical denotation for a predicate like this would be something like the following:

$$\lambda P.[\lambda w.[\forall w' \in W[w' \text{ is compatible with what you think in } w \to P(w') = 1]]],$$

where P is of type $\langle s,t \rangle$ (overall type $\langle st,st \rangle$). However, such a denotation would cause problems for the application of Predicate Modification with the denotation of *b<in>ili ni Vic*, which is of type $\langle e,st \rangle$. The denotation used here is inferred from sentences like (81a), where *inakala* appears to take a DP complement.

   d. Predicate Modification: *libro=ng [[<in>akala mo] na [b<in>ili ni Vic]]*

     $\lambda x.[\lambda w.[x$ is a book in $w \wedge$ Vic bought x in $w \wedge$

                                    you thought/made an assumption about x in $w]]$

The final step shows that the construction as a whole holds true of entities about which you (the addressee) made an assumption, and does not require you (the addressee) to hold beliefs regarding Vic's buying of this entity. The denotation also holds true of entities such that Vic actually bought it in the same possible world as the one in which you made some assumption about it. In other words, it also predicts *inakala* '(mistakenly) thought' to be a factive predicate. These are incorrect results that arise because the matrix predicate does not have semantic scope over its complement.[53]
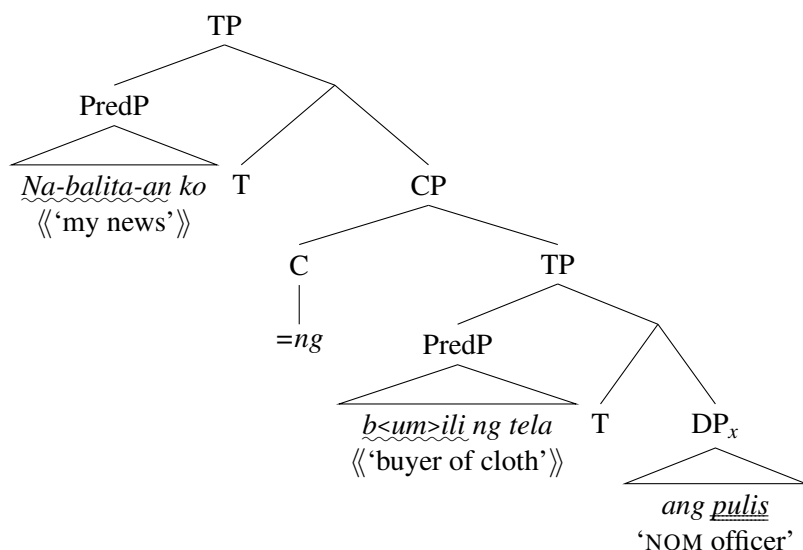
     While it may ultimately be possible to adopt a flat structure for cases with apparent long-distance dependencies in Tagalog while also correctly deriving semantic scope, this remains to be fleshed out. These problems notwithstanding, the non-hierarchical analysis faces another problem in that it predicts that voice phrases should behave identically to bare and derived nominals with respect to (apparent) long distance extraction constructions. Barring semantic restrictions, it should be possible to replace the matrix voice phrase with a bare or derived nominal in these long distance extraction contexts. I will argue later on that this behavior is not attested. Before this, I consider a second alternative for generating long-distance dependencies.

5.3 Long-distance out of subjects

A second approach is proposed by Aldridge (2009, 56), based on the behavior of embedded finite clauses under (non-long-distance) extraction. She suggests, following the PNA, that "the embedded CP is presumably the subject" (i.e., has the same status as the *ang*-marked DP) of an embedding predicate like *nabalitaan ko* because "the embedded clause expresses what was [heard]". This approach is supported by the fact that embedded CPs participate in the Tagalog extraction restriction system. That is, A′-extracting an embedded CP (e.g., *What did you say?*) requires a specific voice form on the matrix verb, parallel to questioning other DPs (e.g., *Who said...?*). A tree for (78) under this view is given in (86), with an intuitive paraphrase given in (87). Recall from (22) that under the PNA, subjects are base-generated as complements of T^0. Thus, similarly to how *ang pulis* occupies the subject position of the lower TP, the CP occupies the subject position of the higher TP, which is the position targeted for local clausal extraction, as Aldridge discusses.Applying this now to long-distance extraction, we see that the intended extraction site (DP$_x$) is no longer contained within a larger DP, which was the problem for the low CP approach discussed in Sec. 5.1.

(86)   *Possible extrapolation of the PNA for clausal subjects*

---

[53]Interestingly, the problem seems to be less acute for *sinabi* 'said'. (80) can, with the right prosodic breaks, have the non-scopal interpretation sketched in the truth conditions in (85). Reversing the predicates, as in (81b), appears to reinforce this interpretation, suggesting that *sinabi* does have an available individual-denoting reading (along the lines of *mentioned*). Note that the different aspects of the non-scopal interpretation seem to occur together: if the truth conditions require the book to actually have been bought, then they must be *compatible* with the addressee not having said anything about the buying.

```
                              TP
                    _____/_____
                PredP                        \
            _____/\_____         _____CP_____
     Na-balita-an ko   T        C                          TP
     《'my news'》        |      ____/\____      _____/_____
                       =ng    PredP          T              DP_x
                        ____/\____                    _____/_____
                 b<um>ili ng tela   T               ang pulis
                 《'buyer of cloth'》                 'NOM officer'
```

(87)   《'[$_{CP}$ That the police officer was a buyer of cloth] is my news.'》

The appearance of the CP in subject position might lead one to think that long-distance extraction should still be impossible, due to Ross's (1967) Sentential Subject Constraint. However, there is a wealth of literature discussing how subjects in Tagalog (and other Austronesian languages) are divergent in a few ways from the traditional notion of subject, given that the usual subjecthood tests yield mixed results in this language (Guilfoyle et al. 1992, Kroeger 1993, Rackowski 2002, among others).

Taking the term "subject" to mean the syntactically prominent *ang*-marked constituent, there is some evidence suggesting that subjects in Tagalog are in fact transparent to extraction to some extent. For example, it has been observed that possessors of subjects may undergo extraction (Cena 1979, Kroeger 1993).

Thus, treating the complement CP as having the same status as an *ang*-marked subject DP as in (86) leaves open the possibility of this constituent being transparent to extraction. A concrete implementation of this is given by Rackowski and Richards (2005), who argue that long distance extraction in Tagalog requires that CPs have precisely this subject-like status. The formal nature of "subject" differs between the two analyses (a result of Object Shift and Agree for Rackowski and Richards vs a structural position for the PNA), but if we assume for the purposes of discussion that the subject position in the PNA is similar in terms of transparency to extraction, then long distance extraction in cases like (86) is at least not immediately ruled out.

This approach introduces another problem, however. It overgenerates by way of predicting that long distance extraction should be equally possible for constructions where the matrix predicate is a voice phrase and for those where the matrix predicate is a bare or derived nominal. This overgeneration problem is shared with the non-hierarchical alternative previously discussed, and will be demonstrated next.

5.4   Long-distance in true nominals

The two possible approaches to clausal complements discussed above (non-hierarchical and CP subject) both predict that long-distance dependencies should be well-formed not only for voice phrase predicates, but also for bare and derived nominal predicates. This is because for both approaches, the matrix predicate is syntactically distant in some sense from the embedded one. There should thus be no *syntactic* restriction

on the possibilities for forming long-distance A′-dependencies involving the three types of constructions discussed in this paper.

For the non-hierarchical approach, the sketch of which is repeated in (88), the apparent matrix and embedded predicates have a non-hierarchical relationship to each other. Furthermore, there is no movement from either predicate, by assumption. It should thus be possible in general to use a bare or derived nominal in place of the voice phrase for the "matrix" predicate, so long as both predicates are of the correct semantic type and are conceptually compatible with the semantic head of the construction. It is not obvious that the syntactic identity of either predicate should affect the well-formedness of the construction as a whole.

(88) <u>libro</u>=ng [[s<in>abi   mo   ] na [b<in>ili   ni   Vic]]
       book=LK <PFV>say[PV] 2SG.GEN   LK <PFV>buy[PV] GEN Vic

       'book that you said Vic bought'

       ⟨⟨'book [that was your said-thing] [that was Vic's bought-thing]'⟩⟩

For the CP subject approach, the matrix predicate DP appears in a PredP projection which is itself in the specifier of matrix TP, as in (86) previously. The syntactic relationship between the matrix predicate DP and the subject CP is thus indirect if at all present. Without a more direct relationship between the two constituents (e.g., some kind of selection) it is not obvious that movement out of the subject CP should be restricted by the matrix predicate.

The prediction we arrive at for both approaches is not borne out in the data. I now show that the true nominals (as argued previously in Section 4) do not generally allow the formation of long-distance A′-dependencies from their embedded clauses.

A number of bare and derived nominals can take clausal complements, roughly producing what in English would be complex noun phrases (e.g., *the claim (is) that...*, *the rumor (is) that...*). Some of these forms have at least a corresponding voice phrase construction, giving us a good comparison point for our present purposes. Two pairs of examples are given below, one pair with a bare nominal and one with a derived nominal. Both pairs have a minimally different voice phrase construction for comparison, with (89a) being familiar from the previous half of this section.

(89) a. <u>Na-balita-an</u>   ko=ng   [b<um>ili   ng tela ang pulis].
          PFV.AIA-news-LV 1SG.GEN=LK   <PFV.AV>buy GEN cloth NOM police

          'I heard that the police officer bought cloth.'   *(repeated from 78)*

      b. <u>Balita</u> ko=ng   [b<um>ili   ng tela ang pulis].
          news 1SG.GEN=LK   <PFV.AV>buy GEN cloth NOM police

          'I heard that the police officer bought cloth.'

          'That the police officer bought cloth is my news.'

(90) a. <u>Hina~hangad</u>   ko=ng   [<u>tapus-in</u> ang sanaysay na ito].
          IPFV~aspire[PV] 1SG.GEN=LK   finish-PV NOM essay   LK this

          'I aspire to finish this essay.'

      b. <u>Hangar-in</u> ko{=ng   / ang} [<u>tapus-in</u> ang sanaysay na ito].
          aspire-PN 1SG.GEN=LK NOM   finish-PV NOM essay   LK this

          'My aspiration is to finish this essay.'

          'To finish this essay is my aspiration.'

As we saw earlier, voice phrases allow for long-distance A-bar extraction (91a, 92a). In comparison, the same extraction is ungrammatical for the bare and derived nominal counterparts (91b, 92b), which is what the PNA incorrectly predicts for the voice phrases. Similar behavior can be observed with headless relatives and similarly "headless" bare/derived nominal constructions that only have a gap in their embedded clause, as shown in (93-94).[54]

(91) a.   pulis$_i$ na <u>na-balita-an</u>   ko=ng   [<u>b<um>ili</u>   ng   tela t_i]
           police LK PFV.AIA-news-LV 1SG.GEN=LK <PFV.AV>buy GEN cloth

           'police officer who I heard bought cloth'

      b. *? pulis$_i$ na <u>balita</u> ko=ng   [<u>b<um>ili</u>   ng   tela t_i]
           police LK news 1SG.GEN=LK   <PFV.AV>buy GEN cloth

           'police officer who my news is that he bought cloth'

(92) a.   sanaysay$_i$ na <u>hina~hangad</u>   ko=ng   [<u>tapus-in</u> t_i]
           essay   LK IPFV~aspire[PV] 1SG.GEN=LK   finish-PV

           'essay that I aspire to finish'

      b. *? sanaysay$_i$ na <u>hangar-in</u> ko{=ng   / ang} [<u>tapus-in</u> t_i]
           essay   LK aspire-PN 1SG.GEN=LK NOM   finish-PV

           'essay that my aspiration is to finish it'

(93)   ang   {<u>na-balita-an</u>   / * <u>balita</u>} ko=ng   [<u>b<um>ili</u>   ng   tela t_i]
         NOM PFV.AIA-news-LV news 1SG.GEN=LK <PFV.AV>buy GEN cloth

         'the one who I heard bought cloth'

(94)   ang   {<u>hina~hangad</u>   / *<u>hangar-in</u>} ko=ng   [<u>tapus-in</u> t_i]
         NOM IPFV~aspire[PV]   aspire-PN 1SG.GEN=LK   finish-PV

         'what I aspire to finish'

As previously discussed, it is unclear under the PNA how to restrict long-distance extraction out of bare and derived nominals while simultaneously allowing it out of voice phrases. Under a verbal approach, the above facts should be fairly unsurprising and fall out as being similar to crosslinguistically common phenomena (e.g., complex noun phrase islands). As a concrete example of how this might fall out, let us consider the analysis proposed by Rackowski and Richards (2005).

---

[54]A reviewer provides a naturally occurring datapoint taken from the internet (xvi) that is similar in form to (91b). It is currently unclear how to account for data like this, but a few observations are in order. First is that while the full sentence is grammatical, the "headless" relative version is ill-formed, parallel with the bare nominal version of (93). Second is that (xvi) is a (pseudo)cleft (so-called *ang*-inversion), while (91b) is a relative clause. Evidence exists suggesting that *ang*-inversion is generally less restricted than other types of A′-extraction. For example imperatives show similar behavior: they can undergo clefting, but not relativization, whether headed or headless. It might then be that these constructions are formed via different syntactic processes.

(xvi) *(si rasheed at si   mama asthy) ang balita ko=ng   malapit na   maging Contributor
         NOM Rasheed and NOM Mama Asthy NOM news 1SG.GEN=LK near   now AV.be   contributor

      '(Rasheed and Mama Asthy are) the ones I hear are close to becoming contributors.'

(xvii) Bilh-in mo   ang tela.   (xviii) Ang tela ang bilh-in mo.
        buy-PV 2SG.GEN NOM cloth   NOM cloth NOM buy-PV 2SG.GEN
        'Buy the cloth.'   'Buy <u>THE CLOTH</u>.'

                                               (xix)   * ang (tela=ng) bilh-in mo
                                                  NOM cloth=LK buy-PV 2SG.GEN
                                                  ~'the {cloth / one} you should buy'

Rackowski and Richards account for Tagalog voice morphology as the reflex of an Agree relationship involving a v^0 probe. They assume that possible goals include not only DPs but also CPs, as they also participate in the voice and extraction system in this language (recall Section 5.3). Following a version of Huang's (1982) Condition on Extraction Domain, they assume that for DPs and CPs to be transparent for *wh*-extraction, they must independently Agree with a phase head. Thus, Agree between a phase head v^0 and the c-commanded CP (which is complement to V) must be established for long-distance extraction to proceed. They thus derive long-distance extraction for clauses with voice phrase (verbal) predicates.

While Rackowski and Richards's proposal does not deal directly with the behavior of clauses with bare and derived nominal matrix predicates, the different behavior for these constructions is in principle derivable. The crucial assumption would have to be that the Agree relation proposed for clausal complements of voice phrases does not affect those of the nominals. This could be the case, for example, if we assume that the CPs in examples like (89b) and (90b) are complements of their respective N heads, resulting in complex NPs. These embedded CPs are consequently contained within a DP and are not valid goals for Agree with higher phase heads, such as a matrix clause v^0. Without this Agree relation, the these CPs are opaque to extraction as desired.

The data above thus highlights another area in which nouns and verbs show distinct behavior lining up with what is observed cross-linguistically. Assuming the nominal and verbal nature of the relevant constructions, we can straightforwardly account for the contrast in the availability of extraction out of their complement clauses. Specifically, we have recourse to accounts of independently observed phenomena such as complex NP islands. Conversely, a nominalist view would have no way to syntactically differentiate voice phrases from bare and derived nominals in a way that would derive this contrast in long-distance extraction.

6   Conclusion

This paper argues for the existence of a noun-verb distinction in Tagalog, contrary to the primary stance of nominalist views. This was done mainly by focusing on a recurring pattern of asymmetry between verbal and nominal constructions that was subsequently neutralized when these constructions were marked with nominal functional elements.

The picture that emerges from this pattern suggests that what I refer to as voice phrases come to acquire their nominal distribution high in their structure, at the introduction of the aforementioned functional elements. This is in contrast to what I refer to as bare and derived nominals, which have a nominal distribution form the beginning, or acquire it low in the derivation. Evidence from long-distance extraction constructions support this characterization, being consistent with previous work such as that by Wasow (1976) who shows that such low processes often result in internal structure that is opaque to further syntactic operations (e.g., extraction).

On the other hand, voice phrases only acquire a nominal distribution high in the structure and thus allow extraction. The fact that this switch in distribution occurs with in the presence of nominal functional material is reminiscent of the Functional Nominalization Thesis (Borsley and Kornfilt 2000). I speculate that this process is what underlies the pattern of asymmetry and neutralization discussed in Section 4.

(95)   *The Functional Nominalization Thesis* (FNT; via Kornfilt and Whitman 2011)
    Nominal properties of a nominalization are contributed by a nominal functional projection. The nominalization has verbal properties below the nominal functional projection, nominal properties above it.

The kind of nominalizations typically discussed in the context of the FNT are event-denoting (e.g., Abney 1987, Kornfilt and Whitman 2011), and not individual-denoting like the Tagalog examples discussed here. Work that deals with this issue (e.g., Alexiadou and Grimshaw 2008, Baker and Vinokurova 2009) usually ties the syntactic height of the nominalizers intimately with the resulting interpretation among other properties (e.g., argument structure, compatibility with adverbs, etc.). Thus, larger constructions containing verbal structure and denoting individuals have tended to be analyzed as relative clauses. However, recent work challenges this correlation between the syntax and semantics of these constructions, arguing that a process of nominalization in the spirit of the FNT can produce constructions that denote individuals but have demonstrably verbal internal structure.

For example, Toosarvandani (2014) discusses two morphemes *-na* and *-dɨ* in Northern Paiute that attach to verbs to form nominalizations. These nominalizations may be individual- or event-denoting depending on whether or not there is a gap in the construction corresponding to one of the verb's arguments (shown in (96) for the suffix *-dɨ*). Toosarvandani provides evidence that the nominalized constituent must be *v*P, so the construction as a whole is neither a relative clause (as it is smaller than a full clause) nor a low nominalization operating close to the root.

(96)   a.   Nɨ   **pauma-wini-dɨ** naka.
            1SG.NOM rain-IPFV-NMLZ hear

            'I hear **it raining**.'

       b.   **Su=kutsu**   **patsa-dɨ**   mia-hu.
            DEF.NOM=cow kill.SG-NMLZ go-PFV

            '**The one who killed the cow** left.'   (Toosarvandani 2014, 787)

Another example comes from Blackfoot, which is argued to have clausal nominalizations by Bliss (2014). In this language, clauses can be nominalized in two ways (so-called bare- and *hp*-nominalizations) to form individual-denoting constructions, shown in (97). Bliss notes that these constructions have full clausal structure and do not have dedicated nominalizing morphemes.

(97)   a.   áyo'kaiks   b.   Nitsáápi [otáóoyihpi].
            a-yo'kaa-iksi   nit-yaapi ot-a-ooyi-hp-yi
            IPFV-sleep.AI-3PL   1-see.AI 3-IPFV-eat-CN-INAN
            'the ones who sleep'   'I saw what s/he was eating.'

                                                                                          (Bliss 2014, 92)

She contrasts these constructions with abstract- and associated-instrument nominalizations, which are formed from smaller (non-clausal) constituents with dedicated morphemes *-(hsi)n* and *-a'tsis*. Interestingly, this makes the Blackfoot clausal nominalization system similar in some respects to the behavior of voice phrases in Tagalog. To account for their behavior, Bliss proposes that an [N] feature on a head high in the structure is responsible for nominal distribution of these constructions, with reference determined by an Agree-based mechanism.[55]

While the comparison at this point is preliminary, it is interesting to consider the behavior of Tagalog voice phrases (specifically headless relatives) in the context of these other high, individual-denoting nominal constructions. Exploration of these issues is left for future work.

---

[55]Note that Bliss (2014) does not explicitly argue against a relative clause analysis for the Blackfoot constructions. Johansson (2012) argues against a nominalization analysis, referring to Bruening (2001) for long-distance extraction and island effects as tests for relative clause-hood, which was some of the inspiration for this section.

References

Abney, Steven. 1987. The English noun phrase in its sentential aspect. Doctoral Dissertation, MIT.

Aldridge, Edith. 2002. Nominalization and *Wh*-movement in Seediq and Tagalog. *Language and Linguistics* 3:393–426.

Aldridge, Edith. 2003a. Remnant movement in Tagalog relative clause formation. *Linguistic Inquiry* 34:631–640.

Aldridge, Edith. 2003b. Wh-movement in Seediq and Tagalog. In *Proceedings of AFLA 8*, ed. Andrea Rackowski and Norvin Richards, volume 44 of *MIT Working Papers in Linguistics*, 1–28.

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

Aldridge, Edith. 2009. Minimalist questions for the nominalist analysis of Tagalog syntax. *Theoretical Linguistics* 35:51–62.

Alexiadou, Artemis, and Jane Grimshaw. 2008. Verbs, nouns and affixation. In *SinSpeC: Working Papers of the SFB 732*, ed. Florian Schäfer, volume 1, 1–16. Stuttgart: University of Stuttgart.

Baker, Mark, and Nadya Vinokurova. 2009. On agent nominalizations and why they are not like event nominalizations. *Language* 85:517–556.

Barker, Chris. 1998. Episodic *-ee* in English: A thematic role constraint on new word formation. *Language* 44:695–727.

Bliss, Heather. 2014. Assigning reference in clausal nominalizations. In *Cross-linguistic investigations of nominalization patterns*, ed. Ileana Paul, 85–117. Amsterdam: John Benjamins Publishing Company.

Borsley, Robert, and Jaklin Kornfilt. 2000. Mixed extended projections. In *The nature and function of syntactic categories*, ed. Robert Borsley, volume 32 of *Syntax and Semantics*, 101–131. Academic Press.

Bruening, Benjamin. 2001. Syntax at the edge: Cross-clausal phenomena and the syntax of Passamaquoddy. Doctoral Dissertation, MIT.

Cena, Resty M. 1979. Tagalog counterexamples to the Accessibility Hierarchy. *Studies in Philippine Linguistics* 3:119–124. URL `https://www.sil.org/resources/archives/25963`.

von Fintel, Kai, and Irene Heim. 2011. Intensional semantics. URL `http://mit.edu/fintel/fintel-heim-intensional.pdf`, MIT Lecture Notes.

Gerassimova, Veronica, and Peter Sells. 2008. Long-distance dependencies in Tagalog: The case for raising. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, ed. Charles B. Chang and Hannah J. Haynie, 190–198. Cascadilla Proceedings Project.

Givón, Talmy. 2001. *Syntax: An introduction*, volume 1. John Benjamins Publishing Company.

Grimshaw, Jane. 2000. Locality and extended projection. In *Lexical specification and insertion*, ed. Peter Coopmans, Martin Everaert, and Jane Grimshaw. John Benjamins Publishing Company.

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

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Blackwell textbooks in linguistics. Blackwell Publishers.

Himmelmann, Nikolaus. 2005. Tagalog. In *The Austronesian languages of Asia and Madagascar*, ed. Alexander Adelaar and Nikolaus Himmelmann, chapter 12, 350–376. Routledge.

Himmelmann, Nikolaus. 2008. Lexical categories and voice in Tagalog. In *Voice and grammatical functions in Austronesian languages*, ed. Peter K. Austin and Simon Musgrave, chapter 10. Stanford: CSLI.

Himmelmann, Nikolaus. 2016. Notes on "Noun Phrase Structure" in Tagalog. In *Explorations of the syntax-semantics interface*, ed. Jens Fleischhauer, Anja Latrouite, and Rainer Osswald, 319–342. Düsseldorf University Press. URL `http://dup.oa.hhu.de/id/eprint/526`.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Doctoral Dissertation, MIT.

Johansson, Sara. 2012. Relative clauses, or clause-sized nominalizations? a consideration of Blackfoot. *Working Papers of the Linguistics Circle of the University of Victoria* 21:1–15.

Johns, Alana. 1992. Deriving ergativity. *Linguistic Inquiry* 23:57–97.

Kaufman, Daniel. 2006. Rigidity versus Relativity in adverbial syntax: Evidence from Tagalog. In *Clause structure and adjuncts in Austronesian languages*, ed. Hans-Martin Gärtner, Paul Law, and Joachim Sabel, volume 87 of *Studies in Generative Grammar*, 151–194. Mouton de Gruyter.

Kaufman, Daniel. 2009. Austronesian nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35:1–49.

Kaufman, Daniel. 2011. The nominalist hypothesis in Austronesian: Problems and prospects. University of Delaware Linguistics Colloquium Handout.

Kaufman, Daniel. 2017. Lexical category and alignment in austronesian. In *Oxford handbook of ergativity*, ed. Jessica Coon, Diane Massam, and Lisa Travis. New York: Oxford University Press.

Keenan, Edward. 2000. Morphology is structure: A Malagasy test case. In *Formal issues in Austronesian linguistics*, ed. Ileana Paul, Vivianne Phillips, and Lisa deMena Travis. Springer.

Kornfilt, Jaklin, and John Whitman. 2011. Afterword: Nominalizations in syntactic theory. *Lingua* 121:1297–1313.

Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog*. Dissertations in Linguistics. Center for the Study of Language and Information.

Kroeger, Paul. 1998. Nouns and verbs in Tagalog: a reply to Foley. Paper presented at the 3rd Lexical-Functional Grammar Conference. URL `http://www.gial.edu/images/pdf/Kroeger-reply-Foley-LFG98.pdf`.

Massam, Diane. 2000. VSO and VOS: Aspects of Niuean word order. In *The syntax of verb-initial languages*, ed. Andrew Carnie and Eithne Guilfoyle, 97–116. Oxford University Press.

Paul, Ileana. 2001. Concealed pseudo-clefts. *Lingua* 111:707–727.

Pearson, Matt. 2005. Voice morphology, case, and argument structure in Malagasy. In *Proceedings of AFLA 11*, ed. Paul Law. Zentrum für Allgemeine Sprachwissenschaft.

Potsdam, Eric. 2006. More concealed pseudoclefts and the clausal typing hypothesis. *Lingua* 116:2154–2182.

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

Rackowski, Andrea, and Norvin Richards. 2005. Phase edge and extraction: A Tagalog case study. *Linguistic Inquiry* 36:565–599.

Richards, Norvin. 2009a. Nouns, verbs, and hidden structure in Tagalog. *Theoretical Linguistics* 35:139–152.

Richards, Norvin. 2009b. The Tagalog copula. In *Proceedings of AFLA 16*.

Ross, John R. 1967. Constraints on variables in syntax. Doctoral Dissertation, MIT.

Ross, Malcolm. 2009. Proto Austronesian verbal morphology: A reappraisal. In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, ed. Alexander Adelaar and Andrew Pawley, 295–326. Canberra: Pacific Linguistics.

Schachter, Paul, and Fe Otanes. 1972. *Tagalog reference grammar*. University of California Press.

Starosta, Stanley, Andrew Pawley, and Lawrence Reid. 1982. The evolution of focus in Austronesian. In *Papers from the Third International Conference on Austronesian Linguistics*, ed. Amran Halim, Lois Carrington, and Stephen Wurm, volume 2 of *Pacific Linguistics Series C*, 145–170. Canberra: Australian National University.

Toosarvandani, Maziar. 2014. Two types of deverbal nominalization in Northern Paiute. *Language* 90:786–833.

Wasow, Thomas. 1976. Transformations and the lexicon. In *Proceedings of the 1976 MSSB Irvine Conference on the Formal Syntax of Natural Language*, ed. Peter Culicover, Thomas Wasow, and Adrian Akmajian, 327–360. Academic Press.

Wolff, John. 1973. Verbal inflection in Proto-Austronesian. In *Parangal kay Cecilio Lopez*, ed. Andrew B. Gonzalez, number 4 in Philippine Journal of Linguistics Special Monograph, 71–91. Quezon City: Linguistic Society of the Philippines.