Notes on "noun phrase structure" in Tagalog

Nikolaus P. Himmelmann

Abstract

This paper presents some observations on the syntax and semantics of the Tagalog phrase marking particles *ang*, *ng*, and *sa*. While there is some evidence for the widely held view that the phrase marking particles form a kind of paradigm in that they are at least in partial complementary distribution, they differ significantly in their distributional characteristics. Consequently, it will be argued that *sa* heads prepositional phrases, while *ang* and *ng* head higher-level phrases (i.e. phrases where PPs occur as complements or adjuncts). These phrases may be considered DPs, although they differ in a number of regards from DPs in European languages. Because of these differences, their status as determiners may be open to questions, but there can be little doubt that *ang* and *ng* provide examples par excellence for functional elements displaying (syntactic) head characteristics.

Analyzing *ang* and *ng* as determiners raises the issue of how they relate to other elements which are usually considered determiners, in particular demonstratives. This problem is taken up in the second main part of the article. It is proposed that demonstratives in fact may occur in two different phrase-structural positions, i.e. they occur both as alternate heads instead of *ang* and *ng* and as their complements.

1. Introduction[1]

With a few exceptions (e.g. some temporal adverbials), all non-pronominal arguments and adjuncts in Tagalog are marked by one of the three clitic particles *ang*, *ng* or *sa*.[2] Typical uses of these markers are seen in the following example involving a 3-place predicate in patient voice where *ang* marks the subject, *ng* the non-subject actor and *sa* the recipient:[3]

---

[1]   This paper was originally presented at the special panel session *Noun Phrase Structures: Functional Elements and Reference Tracking* at the *Tenth International Conference on Austronesian Linguistics* 2006 in Palawan/Philippines. It has been updated sligthly, but no attempt has been made to cover more recent developments in the analysis of phrase structure in general, and noun phrase structure in particular.
   I am grateful to the organisers of the panel, Simon Musgrave and Michael Ewing, for inviting me to this panel. I would also like to thank Jan Strunk for very helpful comments on a pre-conference version of this paper. And I owe very special thanks to Dan Kaufman for detailed, rigorous and challenging comments on a written draft which have helped to clarify some issues and prevented some lapses. Unfortunately, it was not possible to deal with all the challenges in sufficient detail here, a task I therefore will have to leave for the future.

[2]   The major exception are personal names (*Pedro, Maria* etc.) which occur with the markers *si, ni* and *kay* (plural *sina, nina, kina*). The distribution of personal name phrases is similar to that of *ang, ng* and *sa*-phrases, but there are a number of important differences which preclude the option of simply extending the analysis proposed here for *ang, ng* ang *sa* to these markers. The syntax of the personal name markers is not further investigated here, and unless explicitly noted otherwise, the claims made for *ang, ng* and *sa* do not apply to them.
   Another set of exceptions involves arguments connected to the predicate with the linking particle *=ng/na* as in *pumuntá=ng Manila* (AV:go=LK Manila) 'went to Manila'.

[3]   Apart from a few simple phrases used to illustrate basic phrase structure, all examples in this paper are taken from natural discourse. Sources are the author's own corpus of spontaneous spoken narratives which includes stories from Wolff et al.'s (1991) textbook, Tagalog websites (coded as *www*) and the texts in Bloomfield (1917). The examples from spoken narratives retain features of the spoken language (in particular common reductions). Glosses for content words are from English (1986). Orthographic conventions follow the standard norm. This is relevant in particular with regard to how the proclitic particles are represented. As they form

```
(1)  iniabót          ng  manggagamot  sa  sundalo  ang  itlóg
     handed:PV.RLS GEN doctor          LOC soldier  SPEC egg
     The physician handed the egg to the soldier, …
```

The grammatical category and function of these particles is a matter of debate and there are many different terms in use for referring to them, including *case markers*, *relation markers*, *determiners* and *prepositions*.[4] Most analyses, however, agree with regard to the assumption that these markers form a kind of paradigm. There are a number of observations that support this assumption. Most importantly perhaps, as just noted, all non-pronominal argument and adjunct expressions have to have one of these markers. Personal pronouns and demonstratives, which typically are not marked with these markers, occur in three different forms which are known as the *ang*, *ng* and *sa*-form because they have roughly the same distribution as the expressions marked by these clitics.[5] This provides further support for the assumption that they form a kind of paradigm. Furthermore, the markers determine the syntactic distribution of the phrase introduced by them, a point to which we will return shortly.

Nevertheless, there are important differences between *sa*-phrases on the one hand and *ang* and *ng*-phrases on the other. Most importantly, *sa*-phrases can be direct complements of *ang* and *ng*. Consequently, it will be argued in section 2 that they occur in different types of phrases while still sharing the essential property of being the syntactic heads of their respective phrases: *sa* heads prepositional phrases, while *ang* and *ng* head determiner phrases.

The proposal that *ang* and *ng* are determiners is not without problems. Among other things, this proposal raises the issue of how they are related to the other main candidates for determiner status in Tagalog, i.e., the demonstratives. Section 3 attends to this issue.
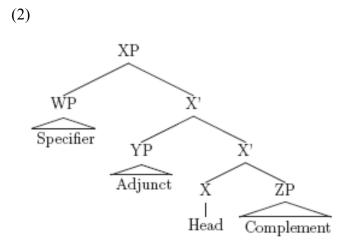
In exploring Tagalog phrase structure, X-Bar theory will be used as a research heuristic, and X-Bar schemata of the type shown in (2) are used as representational devices. The use of X-Bar theory is motivated by the fact that it is a useful tool for investigating hierarchical phrase structure. Furthermore, it provides a representational format which is widely understood. However, using X-Bar theory as a research heuristic does not mean that all universalist assumptions underlying its 'orthodox' uses are adopted here as well. That is, it is not assumed that all major phrases in all languages involve all the positions and functions shown in (2). Instead, every position and function needs to be supported by language-specific, typically distributional evidence. Importantly, no use is made of empty categories and positions simply in order to preserve the putatively universal structure depicted in (2).[6]

---

phonological words with the following item, representations such as *angitlóg* or *ang=itlóg* rather than *ang itlóg* would be more appropriate.

[4] See Reid (2002: 296f) for a fuller list of terms used for the elements.

[5] As in the case of personal name phrases, however, there are a few important differences which preclude a simple extension of the analysis for *ang*, *ng* and *sa*-phrases.
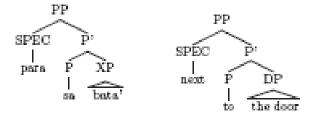
[6] See Kornai & Pullum (1990) for some of the problems created by the unrestrained proliferation of empty categories in X-bar analyses. Note also that much of the following analysis and argument becomes void once it is assumed that the Tagalog phrase markers may be followed by empty nominal heads in all those instances where their co-constituents do not appear to be nominals syntactically and semantically.

(2)

2. On the phrase-structural position of *sa*, and *ang* and *ng*

In form and function, the marker *sa* behaves very much like a (locative) preposition in better known European languages. Apart from some temporal expressions which may occur without any phrase marker, most adjuncts are introduced with this marker. It also occurs with a number of non-subject arguments, including beneficiaries and recipients. One major difference between European-style prepositions and Tagalog *sa* pertains to the fact that *sa* in Tagalog is the *only* preposition, while European languages typically allow for a broader set of items to function as prepositions. Consequently, *sa* is an obligatory constituent in Tagalog prepositional phrases. Different prepositional meanings and functions are expressed by combining *sa* with a specifier as in *para sa* 'for', *galing sa* 'from', *dahil sa* 'because of', *hanggang sa* 'until', *tungkol sa* 'about', *ukol sa* 'about', or *alinsunod sa* 'according to'. In short, it appears unproblematic to analyse phrases with *sa* very much like prepositional phrases in English, as shown in (3) for the phrase *para sa bata'* 'for the child'.[7]

(3) The structure of PP in Tagalog and English

Another major difference between Tagalog and English pertains to the category of the complement (the XP in example (3)). In English, this usually has to be a DP (or, if one
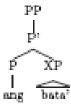
---

7 I am aware of the fact that SPEC is a highly problematic category (cp. the classic squib by McCawley 1989 and the more recent 'attack' by Starke 2004, inter alia) and that current generative analyses no longer analyse modifiers of prepositions such as *para* and *next* as specifiers (at the cost of an enormous inflation of phrasal heads and categories). The main point here is that despite some differences, phrases marked with *sa* can be analysed very much along the same lines as prepositional phrases in English and similarly structured languages.

prefers, an NP).[8] In Tagalog, this is a considerably more complex issue we will return to below.

    A third difference between Tagalog *sa* and English prepositions is the fact that *sa* does not allow for intransitive uses (i.e. there are no verb-particle constructions in Tagalog). Consequently, Tagalog *sa* is unequivocally a function word, and it is rather tempting to view it as part of a paradigm of phrase marking function words which would also include the other two phrase-marking clitics *ang* and *ng*. If one assumes that *ang*, *ng* and *sa* are in a paradigmatic relationship, it would follow that analogous analyses are assumed for *ang* and *ng*. Thus, *ang bata'* 'the/a child' would be analysed as shown in (4).

(4)

However, the assumption of paradigmatic organisation would imply that *sa* and *ang/ng* are in complementary distribution and occupy the same phrase-structural position. This implication is clearly wrong. While *ang* and *ng* are in complementary distribution, both of them may immediately precede a phrase marked by *sa*, as seen in the following examples:[9]

```
(5)   ginagawa                na  ang  sa  Barangay Catmon
      -in-      RDP1-gawá'    na  ang  sa  Barangay Catmon
      -RLS(UG)-RDP1-made  now SPEC LOC Barangay Catmon
```
    (the clinic in Barangay Cay Prombo has already been erected,) the one in Barangay Catmon is currently under construction [www]

```
(6)   yamang ang  sa  pagóng ay tumubo'    hanggáng sa   magbunga.
      yamang ang  sa  pagóng ay -um-tubo'  hanggáng sa   mag-bunga
      while  SPEC LOC turtle PM -AV-growth until     LOC AV-flower
```
    while that of the turtle grew until it bore fruit.

```
(7)   ang kanilang     lagáy    ay gaya ng  sa  isang     busabos
      ang kanilá =ng  lagáy    ay gaya ng  sa  isá =ng  busabos
      SPEC 3.PL.DAT=LK position PM like GEN LOC one =LK slave

      o  alipin at  isang     panginoón o  hare'.
      o  alipin at  isá=ng panginoón o  hari'
      or slave  and one=LK master     or king
```
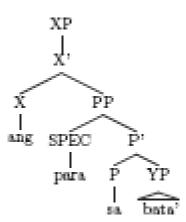   their position was like that of a slave or thrall and a lord or king.

There are no phrases where *sa* immediately precedes *ang* or *ng* (*sa ang, *sa ng). Consequently, one has to assume a phrase structural position for *ang* and *ng* which is 'higher' than the one for *sa* (i.e. which c-commands *sa*), as shown for the phrase *ang para sa bata'*

---

[8] I ignore here examples such as *from under the bed* which could be argued to consist of a preposition (*from*) taking a PP (*under the bed*) as complement.

[9] See Reid (2002:209-211) for a similar argument.

'the one for the child' in (8). Recall the remark at the end of section 1 that in this paper no use is made of empty categories in order to preserve putatively universal phrase structures. Hence, given the fact that *ang* and *ng* (like *sa*) cannot occur on their own, the most straightforward assumption is that the PP in examples (5)-(7) is indeed a complement and not an X' or XP-adjunct.

(8)

There are other differences between *sa* and the other two makers which make it clear that these indeed belong to different categories. Thus, for example, only *sa* may be affixed with the stative prefix *na-*, thereby expressing the meaning 'be in/at/on etc.' as in:

```
(9)  semantalang  syá  'y  nasa        tabí ng  ilog
     samantala=ng siyá ay  na-    sa   tabí ng  ilog
     meanwhile=LK 3.SG PM  RLS.ST-LOC  side GEN river
         When he was close to the riverside,
```

There is no **na-ang* or any other combination of an affix + *ang* or *ng*.

Another difference pertains to the possibility of being replaced by a corresponding form of the demonstrative. Both *ang* and *ng* freely allow for this possibility. In the following example, the *ang*-form of the proximal demonstrative *itó* (plus enclitic linker *–ng*) takes the place of *ang*:

```
(10) nung          mágising   itong  iná
     noón     :LK  má-gising  itó=ng iná
     DIST.GEN:LK ST-awake     PRX=LK mother
         When the mother woke up,
```

But for *sa*, replacement by the *sa*-form of a demonstrative is impossible. There is no **dito(ng)/diyan(g) Maynila* 'here/there in Manila' as an alternative to *sa Maynila* 'in Manila'.[10] The *sa*-form of the deictic always has to be juxtaposed with a full *sa*-phrase as in:

---

[10] Daniel Kaufman (pc) draws my attention to colloquial examples attested on the Internet where *ditong Maynila/Pilipinas* occurs in construction with directional predicates such as *puntá* 'go to' or *balík'* 'return to'. However, such examples are probably best analysed as involving a clitic positioning of *ditó,* the linker linking the predicate to its directional complement (cp. the example *pumuntá=ng Manila* given in footnote 2 above). That is, *kelan balik muh ditong maynila* (when return 2s.GEN PRX.LOC=LK Maynila) 'when will you return here to Manila' (original spelling retained) involves the predicate phrase *balik na maynila* 'return to Manila' with two second position clitics (*mu* and *ditó*) occurring in between the two elements of this phrase.

```
(11) dito      sa  kahariang  itó ay merong                isang  sultán
     PRX.LOC LOC kingdom:LK PRX PM EXIST.DIST.LOC.LK one:LK sultán
       here in this kingdom there was a sultan
```

This also holds for complex prepositions consisting of a specifier and *sa*: **para ditong X* is ungrammatical. It has to be *para dito sa X*.

Furthermore, while *sa* co-occurs with specifying elements such as *para* in the preceding example, there are no such elements which could precede *ang* or *ng*.

Taking now a closer look at *ang* and *ng*, these two markers are identical with regard to their phrase-internal properties. They are clearly in complementary distribution, they always occur at the left edge of the phrase they belong to, and they can be replaced by corresponding forms of the demonstratives. They differ with regard to their external distribution: Phrases with *ang* occur in subject function (e.g. *ang itlóg* in (1)), in topic function (see also *ang sa pagong* in (6)):

```
(12) ang kuba'      ay ma-hina'   ang katawán
     SPEC hunchback PM ST-weakness SPEC body
       the hunchback was weak of body (lit. as for the hunchback, the body was weak)
```

and as predicates:

```
(13) ang langgám rin  ang  tumulong sa  mga bata'
     SPEC ant      also SPEC AV:help  LOC PL  child
       The ants also helped the children (lit. The ones who helped the children were also the
       ants).
```

The marker *ng*, on the other hand, marks non-subject complements such as *ng manggagamot* in (1) and possessors such as *ng ilog* in (9). When marking non-subject undergoers, *ng* alternates with *sa* in a definiteness alternation: non-subject undergoers marked with *sa* are usually definite; for those marked with *ng* an indefinite reading is preferred, but not obligatory, as demonstrated in (15).

```
(14) itó ang  pusa=ng k<um>ain sa  dagá'
     PRX SPEC cat=LK  <AV>eat   LOC rat
       This is the cat that ate the rat. (McFarland 1978:157)
```

(15) a. itó ang pusang kumain **sa** dagá'   unambiguously definite = (14)

     b. itó ang pusang kumain **ng** dagá'   indefinite or non-specific preferred, but definite also possible

     c. itó ang pusang kumain **ng isang** dagá'   unambiguously indefinite (*isá* = 'one')

     d. itó ang pusang kumain **ng** dagáng **iyón**   unambiguously definite (*iyón* = DIST)

Pronouns and personal names always occur in *sa*-form when functioning as non-subject undergoers.

Table 1 summarizes this brief (and not fully exhaustive) survey of the external distribution of *ang*, *ng* and *sa* phrases.

Table 1: Phrase markers and syntactic functions

ANG	NG	SA
SUBJECT PREDICATE TOPIC	NON-SUBJECT COMPLEMENTS (CLAUSE-LEVEL), POSSESSIVE COMPLEMENTS (PHRASE-LEVEL)	ADJUNCTS, SOME NON-SUBJECT COMPLEMENTS (usually in alternation with *ng*), PREDICATE[11]

Importantly, the distribution of *ang*, *ng* and *sa*-phrases holds *regardless* of what follows the marker in the phrase. As should be obvious from the examples discussed so far, the markers occur with co-constituents of various types and complexity. Thus, for example, *ang* occurs with simple content words such as *itlóg* in (1) or *kuba'* in (12) and with prepositional phrases as in (5) and (6) above. But co-constituents of all three phrase-marking clitics can also be more complex. Thus, they may introduce a verb[12] accompanied by all its arguments and adjuncts except the subject argument as in (cp. also example (13) above):

(16) at   hulih-in **ang   nag-sá-sabuy   sa   kanyá**
     and catch-PV SPEC RLS.AV-RDP1-spatter LOC 3.SG.DAT
     ng   buhangin
     GEN sand
        and catch the one who was throwing sand on him

They may even introduce a complete non-finite clause as in the following two examples.[13] Here the constituents in parentheses constitute full clauses, consisting of a predicate (*manalo* and *talunin*, respectively) and a subject (*tayo* and *ang isang kaaway*, respectively). The predicates are in subjunctive form which is used in non-finite clauses and imperatives. In (18), the clause marked with *ang* (i.e. *ang talunin ang isang kaaway*) itself functions as a (negated) predicate in matrix construction with a topicalised (= inverted) subject (i.e. *ang hangád ng nagsísipaglaro'*).

(17) talagang   nakákatuwá   **ang [manalo   tayo]**
     talagá=ng naka-   RDP1-tuwá' ang   maN-talo   tayo
     really=LK RLS.ST.AV-RDP1-joy   SPEC AV-surpassed 1.PL.IN
        It is really fun to win (lit. when we win).

(18) Sa   sipa' ang   hangád   ng   nagsísipaglaro'
     LOC kick SPEC ambition GEN player
     ay hindí **ang [talun-in   ang   isang   kaaway]**
     PM NEG   SPEC defeated-PV SPEC one:LK opponent
        In *sipa* the aim of the players is not to defeat an opponent, …

---

[11] Predicate uses of *sa*-phrases are not illustrated in the preceding discussion and are not directly relevant for present concerns, but only included for reasons of completeness. Here is a simple example: *sa akin ang Pinas* 'Pinas is mine'.
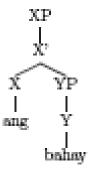
[12] Verbs are morphologically defined as those content words which are inflected for aspect and mood.

[13] To the best of my knowledge, examples of this type have so far not been discussed in the literature. They are quite rare, but not at all problematical with regard to acceptability in all registers (the first example is from Wolff et al. 1991, the second from Bloomfield 1917)). It seems likely that similar examples are possible with *ng* and *sa*, though this has to be investigated in more detail.

The important point for our current concern is that, regardless of the complexity of the constituent following *ang*, a phrase headed by *ang* can function, and can *only* function, as subject, topic or predicate. That is, the syntactic distribution of the phrase is fully determined by *ang*. Similarly, the syntactic distribution of *ng* and *sa*-phrases is fully determined by *ng* and *sa*, except that in some of their uses they regularly alternate in accordance with definiteness distinctions.

Consequently, there can be little doubt as to the fact that *ang* and *ng* like *sa* are the heads of their respective phrases, at least with regard to being the "external representative" (Zwicky 1993) of the phrase, a core characteristic of syntactic heads. Strictly speaking, and unlike demonstratives in both English and Tagalog, these markers are *not* distributional equivalents of their phrases in the sense of Bloomfield (1933) because they cannot form a phrase all by themselves. They minimally need one further co-constituent. Hence, for a phrase such as *ang bahay* we can assume the constituent structure given in (19).
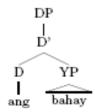
(19)

```
        XP
        |
        X'
       /  \
      X    YP
      |     |
     ang    Y
            |
          bahay
```

Strictly speaking, *ang* or *ng* as phrase heads instantiate the X-Bar schema only incompletely in that they do not allow for specifiers (which, as noted above, is one of the differences between them and the preposition *sa*).

While it is widely accepted that *sa* is a preposition and hence a phrase headed by *sa* is a PP, there is little agreement as to the category of *ang* and *ng*. I have argued elsewhere (Himmelmann 1984, 1991, 1998:333-336) that *ang* is best analysed as a specific article and *ng* as its genitive form. If that is accepted, *ang* and *ng* can be considered determiners, and phrases headed by them as DPs, as shown in (20) (and done henceforth in this paper). This categorization would also appear to be supported by the fact noted above that they may be replaced by the corresponding form of a demonstrative. However, this is not quite straightforward support because demonstratives may also co-occur with *ang* and *ng* in what appears to be a single phrase. We will return to this issue in the following section.

(20)

In many ways, Tagalog *ang* and *ng*-phrases are much clearer instances of a DP than the kinds of nominal expressions in European languages that have been hypothesized to instantiate this phrase type. Most importantly, and unlike articles in most European languages, the Tagalog functional elements are obligatory – they occur without exception in all phrases of this type – and they *fully* determine the distribution of the phrase they head. Note that the question of what the actual category of *ang* and *ng* is, is at least in part independent of whether they are phrasal (co-)heads.[14]

I consider it an unresolved issue whether it is necessary and useful to attribute a case function to *ang* and *ng*-phrases in addition to their function as determiners. Calling *ng* the "genitive form of *ang*", as I just did, obviously invites the inference that at least *ng* is a case marker. However, while this is a convenient gloss giving a rough, though incomplete and in some ways also misleading idea of the distribution of *ng*-phrases (see Table 1 above), it is far from clear whether this form is in any relevant sense similar to genitive case forms in Latin, German or Icelandic, or to phrases marked by *of* in English. Both historically and synchronically, there are good reasons to assume that *ng* consists of the linker *na* plus the specific article *ang*, i.e. that it marks 'linked referential phrases' and thus is but one of the many types of modifiers marked with a linker in Tagalog.[15]
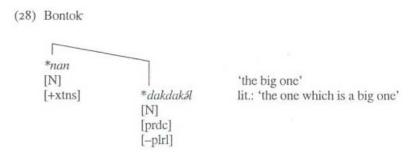
Both *ang* and *ng*-phrases are thus prototypical instances of what Van Valin (2008:168) calls a "reference phrase" (RP). A major advantage of this concept is the fact that it remains noncommittal as to the lexical category of the constituent(s) appearing within such a phrase, thereby avoiding the well-entrenched confusion between lexical categories and syntactic functions enshrined in the classic phrase structure rule S → NP + VP.

The analysis proposed here largely agrees with the analysis in Reid (2002) who also considers *ang* to be a syntactic head. However, Reid provides an analysis in terms of dependency rather than constituency which makes it difficult to compare the analyses in all details. According to Reid, phrase marking clitics such as Tagalog *ang* or Bontok *nan* are *nominal* heads of their phrases, roughly meaning something like 'the one'. The fact that they cannot form phrases by themselves is accounted for by the feature [+xtns] which means that they obligatorily require a dependent predicate to form grammatical phrases. This is

---

[14] It is clear that functional elements functioning as phrase heads do not show all the head properties usually assigned to phrase heads. Most importantly, they are not heads in semantic terms. I consider it an open issue how the special head status of functional elements is best captured. Possibly, the kind of co-head analysis used in LFG (cp., for example, Bresnan 2001) is more adequate than simply applying the standard phrase structure schema to functional elements as done here.

[15] The linker itself in all likelihood derives from a (not case-marked!) demonstrative (Himmelmann 1997:159-188, 2001:834f), which is one reason for not considering *ng* a genitive case marker.

illustrated with the following stemma for the Bontok phrase equivalent to Tagalog *ang malakí* 'the big one' (= example 28 from Reid 2002).[16]

(28) Bontok

```
        ┌─────────────┐
   *nan            │
   [N]             │
   [+xtns]    *dakdakə́l        'the big one'
              [N]              lit.: 'the one which is a big one'
              [prdc]
              [–plrl]
```

Much of Reid's argument as to why *ang* is a noun rather than a determiner depends on definitions and assumptions specific to the lexicase framework, which need not be further discussed here. Still, even abstracting away from the specifics of the theoretical framework used, there appears to be a major problem with this analysis relating to the fact that it fails adequately to capture the highly grammaticised status of phrase marking clitics such as *ang* and *ng* (but see Reid (2000:36-42) for a more dynamic version of this analysis which provides for the possibility that the phrase marking clitics no longer head the constructions but rather become dependents themselves). While the assumption that they are nominals meaning 'the one' may make sense in cases such as *ang malakí* 'the one which is a big one' or *ang bahay*, which in terms of this analysis would have to mean 'the one which is a house', it is difficult to see how one could account for examples such as (17) and (18). In these examples, the complement of *ang* is a complete clause, and it is not clear how this clause could function as the dependent predicate required by the phrase-marking clitic.

Furthermore, in Reid's analysis, it would appear that the phrase-marking clitics are very similar to demonstratives, except that the latter are additionally marked for deictic distinctions. Reid does not discuss demonstratives explicitly, but according to his stemmata 20, 22, 23 and 27, demonstratives are also analysed as the nominal heads of their phrases. As already mentioned above, phrase-marking clitics and demonstratives indeed appear to share some essential characteristics as the latter may replace the former (cp. example (10)). But the relationship between the two classes of elements and the phrase-structural position of demonstratives is more complex than this, as further discussed in the next section.

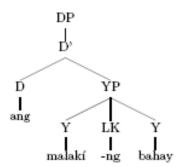## 3. On the phrase-structural position of demonstratives

Demonstratives in Tagalog may be used pronominally, i.e., forming a major constituent all by themselves. An example is the use of *itó* in (14) above where it functions as the predicate in an equational clause. They may also be used "adnominally", that is, as a co-constituent in a nominal expression. In this use, they have to be linked to their co-constituents with a linker as in *itó-ng bahay* 'this house'.

Before looking more closely at the phrase-structural position of demonstratives, it will be useful briefly to look at complex nominal expressions without a demonstrative such as *ang*

---

[16] Note that phrase marking in Bontok, and more generally in northern Philippine languages, is quite distinct from the one found in Tagalog. Reid (2006) provides a detailed survey of the systems encountered in the northern languages.

malakíng bahay 'the big house'. The major observation with regard to these expressions is the fact that there is no straightforward distributional evidence with regard to their heads. Importantly, the order of the co-constituents of *ang* is variable (both *malakíng bahay* and *bahay na malakí* are possible), there is always a linker in between co-constituents of these phrases, and no constituent is obligatory in the sense that only one of them has to be present (i.e. both *ang malakí* and *ang bahay* are well-formed phrases). Note that all of this does not hold true for *ang* (or *ng*): change of its position results in ungrammatical phrases (**bahay ang malaki*, **malaking bahay ang*), and *ang* cannot freely be omitted or occur by itself. Consequently, in a first approximation, we may hypothesize that the structure of phrases such as *malaking bahay* is flat, as shown in

(21)

This is almost certainly not the complete story because, on the one hand, very little is known so far with regard to possible constraints on the variable ordering of constituents in these expressions and the pragmatic or semantic correlates of different orders.[17] On the other hand, there may be distributional evidence for identifying the head of such a phrase provided by constraints on the placement of second-position clitics and the plural word *mga*.[18] This, however, is a rather complex issue which cannot be further explored here. As a consequence, no attempt is being made to assign the complement of the determiner in these phrases to a specific category. Instead, the non-specific label YP is used throughout this article.

Returning to demonstratives, there are two major features which have to be accounted for. First, like most other elements in Tagalog nominal expressions, the position of demonstratives is variable. They can appear at the very beginning and at the very end of such expressions, as seen in the following examples:

---

[17] Kaufman (2005:192f) proposes that different orders here correlate with differences in information structure in that the last element in the phrase tends to receive a phrasal accent and to constitute the most salient or contrastive element. A similar suggestion is made by Schachter & Otanes (1972) and Schachter (1987:944) who, however, limit their claims to demonstratives as further discussed below.

[18] See Kroeger (1993:153f) for some preliminary observations regarding second-position clitics, and Kolmer (1998:11-18) on *mga* in nominal expressions.

(22) **ito**=ng ma-laki=ng bahay   'this big house'
     PRX=LK ST-size=LK house

    also possible:
    itong bahay na malakí
    malaking bahay na **itó**
    bahay na malaki na **itó**

However, it is not possible to place the demonstrative in between other constituents of a nominal expression:

(23)   a. ?*bahay na **itong** malakí
      b. ??*malakíng **itong** bahay

Some native speakers consider these example marginally acceptable (hence the ?) but they then have a clearly different structure: the demonstrative forms a constituent with the word following it, and this sub-constituent is in apposition to the first word. Hence (23)a would mean something like 'a/the house, this big one' and (b) sth. like 'a/the big one, this house'. The latter obviously is pragmatically highly marked and thus not liked at all by native speakers (to date, no examples for either (a) or (b) have been found in texts). We will return to this point below.

    The second point to take note of pertains to the fact that there can be two demonstratives in what appears to be a single nominal expression, as in:

(24) **itong** bahay na **itó**   'this house'

(25) **ito=ng** amáng   sultang   **itó**
     PRX=LK father:LK sultan:LK PRX
      this sultan-father

In line with the first observation, both demonstratives have to be at the outermost periphery of the expression. Obviously, it is somewhat unusual to have what appears to be the same element twice in a single expression. I will now try to show that this is in fact not the case. Rather, the two demonstratives in this construction occupy different hierarchical positions and also have somewhat different functions.[19]

    There is evidence that demonstratives in the left periphery occur in the same position as *ang* because, in more formal registers at least, *ang* and demonstratives in leftmost position are in complementary distribution. Hence, a phrase such as

(26)   ??ang itong bahay

is considered ungrammatical in Standard Tagalog (as opposed to *ang bahay na itó* which is fine). In informal registers, including chat room communication, however, examples such as the following occur frequently enough that one probably has to grant them some acceptability:

---

[19] Kaufman (2010:217f) also argues that there are two structural positions for demonstratives, based on the fact that only demonstratives in the left periphery can be case-marked. Demonstratives in the right periphery always take the default *ang*-form, regardless of the case marking of the phrase they appear in (i.e. it is *sa bahay na itó* and *ng bahay na itó*, not **sa bahay na ditó* or **ng bahay na nitó*).
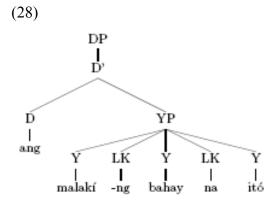
```
(27) Importanteng-importante ang  ito=ng ebidensiya.
     DUP.LK-importante        SPEC PRX=LK evidence
```
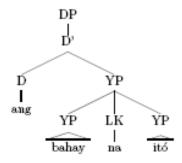 This evidence is very important. [www]

Nevertheless, the fact that *ang* and *ng* are usually missing when a phrase is introduced by a demonstrative suggests that demonstratives in the left periphery in fact occupy the same structural position as *ang* (and *ng* when they occur in *ng*-form). This is not very surprising on the assumption that both *ang/ng* and demonstratives are determiners. In further support of this assumption, it can be noted that a reduced form of the distal demonstrative, i.e., *yung* (< *iyón-ng*), is replacing *ang* in many of its uses in colloquial speech (i.e., it is being grammaticised as a new determiner). Importantly, *yung* shares two of the core characteristics of *ang*, i.e., it cannot form a phrase by itself and it has to occur in leftmost position.

   What about demonstratives in rightmost position then? An initial hypothesis would be that they are hierarchically on the same level as the other constituents in a nominal expression, as illustrated in (28):

(28)

However, such an analysis would wrongly predict that the demonstratives are freely convertible within YP. One way to ensure that the demonstrative occurs in rightmost position would be to analyse it as being in apposition to the other members of a nominal expression. A possible structure is given in (29).
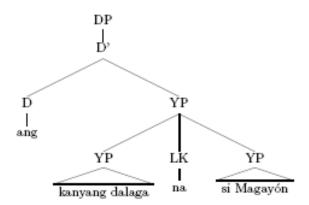
(29)

This analysis would also predict that it is possible to reverse the order of the two adjoined YPs in (29), hence creating phrases of the type *ang itong bahay*. As already mentioned in connection with examples (26) and (27) above, such structures are possible in colloquial style, but highly marked in terms of the standard language.

An appositional structure of the type shown in (29) is needed for independent reasons to account for examples such as (30) where a personal name expression (marked by *si*) is in apposition to a common noun expression (*kanyang dalaga* 'his daughter'):
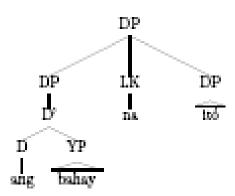
```
(30) ang  kanya=ng    dalaga       na si Magayón
     SPEC 3.SG.DAT=LK  young_woman  LK PN Magayón
       his daughter Magayón
```

(31)

The major alternative to the analysis in (29) would be to consider demonstratives in the right periphery (and also personal noun phrases such as *si Magayón* in (30)) to be in apposition 'one level higher up'. That is, rightmost *itó* could be considered to form a DP by itself which functions as an adjunct to the rest of the phrase, as shown in (32). Since the demonstratives can also be used pronominally, the big advantage of this analysis would be that one could generalize a '(pronominal) head of DP' analysis for all uses of the demonstratives.

(32)

However, there are a number of problems for this alternative analysis. The perhaps least important problem is that it does not allow for structures such as (27) where the demonstrative occurs immediately after *ang*. Furthermore, phrases such as *itóng bahay na itó* would consist of two adjoined DPs headed by the same element which, while not totally impossible, is not very plausible.

More importantly, demonstratives may form the only other constituent in a nominal expression headed by *ang*. This is necessarily so when demonstratives are pluralized with the

plural word *mangá* (conventionally spelled *mgá*) as in (33). But it also occurs when there is no other element in the nominal expression, as in (34).

(33) **ang mga itó** y panghule   lamang ng mga pare' ng kwalta
    ang mgá itó ay paN-huli   lamang ng mgá pare' ng kuwalta
    SPEC PL   PRX PM GER-a_catch only   GEN PL priest GEN money
      these (concepts) were merely a device of the priests for getting money

(34) isang   araw ay inimbita   niyá   **ang itó**
    isá=ng araw ay -in-   imbitá niya   ang itó
    one=LK day PM -RLS(UG)-invite 3.SG.POSS SPEC PRX

    na tumulóy   sa kanyang   bahay
    na -um-tulóy   sa kanyá   =ng bahay
    LK -AV-come_in LOC 3.SG.DAT=LK house
      One day she invited this (latter) one into her house. [www]

Analyzing these examples as appositional along the lines indicated in (32) would imply that *ang* occurs without a complement in these examples. This would be highly unusual since it is nowhere else attested.

    I assume that the demonstrative in these examples is identical to the demonstrative which occurs at the right periphery, since all major constituents in a nominal expression may function as the sole complement of *ang*. That is, each of the three main co-constituents of *ang* in *ang malakíng bahay na itó* can be the sole co-constituent of *ang*:

(35)   ang bahay   'the house'
           ang malakí   'the big one'
           ang itó   'this one'

This, to my mind, strongly suggests that demonstratives, apart from being heads of DPs, also may form phrases of the same type as content words such as *bahay* or *malakí*. Consequently, the analysis given in (29) is to be preferred to the one in (32) despite the fact that it requires a double categorization of demonstratives: they are both (pronominal) heads of DPs and deictic modifiers which occur as adjuncts in the periphery of nominal expressions, preferably in the rightmost position. As heads, they can form DPs on their own or by taking further complements. In the latter case, they are always the leftmost element in a DP.

    To further support this analysis, one would expect some semantic or pragmatic differences corresponding to the difference in phrase-structural position. Schachter (1987:944; see also Schachter & Otanes 1972: 120) notes in this regard

> Although a demonstrative and the noun it modifies may occur in either order, the alternative orderings are generally not in free variation, but are, rather, conditioned by discourse factors. The constituent that comes second typically represents the more salient information and may, for example, be contrastive.

He illustrates this with the following two examples (accents added and glossing modified in accordance with conventions used in this paper):

(36) Mahal   itong   galáng. (Pero mura   itong   singsíng.)
    mahál   itó =ng galáng   pero mura   itó=ng singsíng
    expensive PRX =LK bracelet but cheap PRX=LK ring
      This bracelet is expensive. (But this ring is cheap.)

```
(37) Mahál      ang galang     na itó.(Pero mura  ang  galang
     expensive SPEC bracelet  LK PRX  but cheap  SPEC bracelet
     na iyán.)
     LK MED
```
   This bracelet is expensive. (But that bracelet is cheap.)

In terms of the current analysis, one could further add that in preposed position, demonstratives play a more "determiner-like" role, taking on functions of the phrase-marking clitics *ang* and *ng*, while in postposed position, their function is more clearly deictic.

4. Conclusion

In this paper, it has been argued that the phrase-marking clitics *ang*, *ng* and *sa* in Tagalog are the syntactic heads of the phrases introduced by them. More specifically, *sa* is a local preposition heading a PP, while *ang* and *ng* are determiners heading DPs. While there are sufficient similarities between Tagalog PPs and DPs and their equivalents in European languages to warrant use of these category labels, it should be clearly understood that the former differ from the latter in that the nature of the complements of P and D in Tagalog is still in need of much further research and may turn out to differ substantially. As indicated in section 2, both functional elements appear to allow for a broader range of complements than is usually assumed for Ps and Ds. Most importantly, Tagalog Ds allow PPs and clauses among their complements.

   Similarly, Tagalog demonstratives are not just one kind of determiner, interchangeable with the determiners *ang* and *ng*. Instead, they are also adjuncts which have to occur in the peripheral position of nominal expressions, typically in rightmost position, but in some registers also in leftmost position, immediately *after* the phrase-initial determiner.

Abbreviations

AV	ACTOR VOICE	PN	PERSONAL NAME
DAT	DATIVE	POSS	POSSESSIVE
DIST	DISTAL	PRX	PROXIMAL
DUP	DUPLICATION	PV	PATIENT VOICE
GEN	GENITIVE	RLS	REALIS
GER	GERUND	RDP	REDUPLICATION (numbers indicate different formal types of reduplication)
IN	INCLUSIVE		
LK	LINKER		
LOC	LOCATIVE	SG	SINGULAR
NEG	NEGATION	SPEC	SPECIFIC ARTICLE
PL	PLURAL	ST	STATIVE
PM	PREDICATE MARKER	UG	UNDERGOER

References

Arka, I Wayan & Malcolm Ross (eds), 2005, *The many faces of Austronesian voice systems*, Canberra: Pacific Linguistics

Bloomfield, Leonard (1917) *Tagalog texts with grammatical analysis*, Urbana: The University of Illinois.

---- (1933) *Language*, London: Allen & Unwin.

Bresnan, Joan (2001) *Lexical-Functional Syntax*, Oxford: Blackwell.

English, Leo J. (1986) *Tagalog-English dictionary*, Manila: National Book Store.

Himmelmann, Nikolaus P. (1984) *Linking im Tagalog*, unpublished M.A. thesis, Universität Müchen.

---- (1991) *The Philippine Challenge to Universal Grammar*, Arbeitspapier Nr. 15, Köln: Institut für Sprachwissenschaft.

---- (1997) *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*, Tübingen: Niemeyer.

---- (1998) 'Regularity in irregularity: Article use in adpositional phrases', *Linguistic Typology* 2:315-353.

---- (2001) 'Articles', in: Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds), *Language Typology and Language Universals*, 831-841, Berlin: de Gruyter.

Kaufman, Daniel A. (2005) 'Aspects of pragmatic focus in Tagalog', in: Arka & Ross (eds), 175-196

---- (2010) *The morphosyntax of Tagalog clitics: a typologically driven approach*, unpubl. PhD thesis Cornell University

Kolmer, Agnes (1998) *Pluralität im Tagalog*, Arbeitspapier Nr. 31 (Neue Folge), Köln: Institut für Sprachwissenschaft.

Kornai, András & Geoffrey K. Pullum (1990) 'The X-bar Theory of Phrase Structure', *Language* 66:24-50.

Kroeger, Paul R. (1993) *Phrase Structure and Grammatical Relations in Tagalog*, Stanford: Stanford University Press.

McCawley, James D., 1989, 'INFL, Spec, and other fabulous beasts', *Behavioral and Brain Sciences* 12:350-352

McFarland, Curtis D. (1978) 'Definite objects and subject selection in Philippine languages', in C. Edrial-Luzares and A. Hale (eds) *Studies in Philippine Linguistics*, vol. 2/1:139-182, Manila: Linguistic Society of the Philippines.

Reid, Lawrence A. (2000) 'Sources of Proto-Oceanic initial prenasalization: the view from outside Oceanic', in: Videa P. De Guzman and Byron Bender (eds), *Grammatical analysis: Morphology, syntax and semantics: Studies in honor of Stanley Starosta* (Oceanic Linguistics Special Publications No. 29), 30-45, Honolulu: University of Hawai'i Press.

---- (2002) 'Determiners, nouns or what? Problems in the analysis of some commonly occurring forms in Philippine languages', *Oceanic Linguistics* 41:295-309.

---- (2006),'On reconstructing the morphosyntax of Proto-Northern Luzon, Philippines', *Philippine Journal of Linguistics* 37:1-64.

Schachter, Paul (1987) 'Tagalog', in Bernard Comrie (ed.) *The world's major languages*, 936-958, London: Croom Helm.

Schachter, Paul and Otanes, Fay T. (1972) *Tagalog reference grammar*, Berkeley: University of California Press.

Starke, Michal (2004) 'On the inexistence of specifiers and the nature of heads', in Adriana Belletti (ed.), *The Cartography of Syntactic Structures*, vol. 3: *Structures and Beyond* (Oxford Studies in Comparative Syntax), 252-268, New York: Oxford University Press.

Van Valin Jr., Robert D. (2008) 'RPs and the nature of lexical and syntactic categories in Role and Reference Grammar', in R. D. Van Valin Jr. (ed.), *Investigations of the syntax-semantics-pragmatics interface*, 161-178, Amsterdam: Benjamins.

Wolff, John U. with Maria T.C. Centeno and Der-Hwa V. Rau (1991) *Pilipino through self-instruction*, 4 vols., Ithaca: Cornell Southeast Asia Program.

Zwicky, Arnold M. (1993) "Heads, bases and functors", in: Greville G. Corbett, Norman M. Fraser & Scott McGlashan (eds), *Heads in grammatical theory*, 292-315, Cambridge: Cambridge University Press.