# Bikol clefts and topics and the Austronesian extraction restriction

Michael Yoshitaka Erlewine and Cheryl Lim, National University of Singapore

Draft, December 2019

Comments welcome: mitcho@nus.edu.sg

Many Austronesian languages exhibit an extraction restriction whereby only one particular DP — the "pivot" argument, the choice of which is reflected by morphology on the verb — can be $\overline{\text{A}}$-extracted. We show that such extraction restrictions can vary between different $\overline{\text{A}}$-constructions in Bikol: local clefting is limited to the pivot, whereas topicalization can target pivots and non-pivot agents, but not other non-pivot DPs. Following the phase-theoretic, locality-based approach to such extraction asymmetries in related Austronesian languages, we propose that clefting and topicalization differ in the featural specifications of their probes, but must always attract their closest matching goal. Evidence for this approach comes from interactions between clefting, topicalization, and hanging topic left dislocation in long-distance configurations. Such data motivates the view that the classic Austronesian pivot-only extraction restriction is best characterized in terms of syntactic locality, rather than as a restriction on the grammatical function or morphological case of movement targets.

**Keywords** Bikol, Austronesian voice systems, extraction asymmetry, probing, locality, clefting, topicalization, syntactic ergativity

# Contents

**1 Introduction** **1**

**2 Case and voice in Bikol** **4**

**3 Local clefts and topics** **9**

3.1 Clefts . . . . . 9

3.2 Topics . . . . . 12

3.3 Summary . . . . . 16

**4 Proposal** **17**

4.1 Voice and case in Bikol . . . . . 17

4.2 Topicalization and hanging topic left dislocation . . . . . 24

4.3 Multiple topic constructions . . . . . 29

4.4 Clefts . . . . . 33

4.5 Summary . . . . . 39

**5 Long-distance clefts and the Austronesian extraction restriction** **40**

5.1 Background: Voice systems and long-distance extraction . . . . . 41

5.2 Long-distance clefting . . . . . 44

5.3 Long-distance clefting and embedded topics . . . . . 48

**6 Conclusion** **53**

**Acknowledgements** We thank John Raff Sebastian and Kidjie Saguin for discussion of judgments. Earlier versions of this work were presented at the Syntax/Semantics Reading Group at NUS, the 25th meeting of the Austronesian Formal Linguistics Association at Academia Sinica, Asian Junior Linguists 3 at the University of Hong Kong, Kwansei Gakuin University, IIT Delhi, and the University of Pennsylvania. For helpful comments and discussion, we thank those audiences, Kenyon Branan, Tingchun Chen, Keely New, our *NLLT* editor Julie Anne Legate and three anonymous reviewers, and especially Theodore Levin. This research is supported by the Singapore Ministry of Education under the grants "Theory and variation in extraction marking and subject extraction asymmetries" (R-103-000-142-115) and "Subjecthood in Southeast Asia: Description and theory" (R-103-000-160-112) awarded to the first author, which are gratefully acknowledged.

# 1 Introduction

A major topic in Austronesian syntax has been the study of so-called "Austronesian-(type) voice systems" or "Philippine alignment." In languages of this form, each clause has one argument which we call the "pivot," and only this designated pivot argument can be targeted for $\overline{\text{A}}$-extraction. The nature of this extraction restriction — which has also been described by some authors as a "subject-only" or "absolutive-only" restriction — has been a focal point for typological and theoretical discussions of extraction asymmetries (Keenan and Comrie, 1977; Schachter and Otanes, 1972; Aldridge, 2004; Rackowski and Richards, 2005, a.o.) and is also central to discussions of the notion of "subjecthood" in Austronesian and beyond (Keenan, 1976; Schachter, 1976, 1996; Guilfoyle, Hung, and Travis, 1992; Kroeger, 1993, a.o.).

In this paper, we describe patterns of $\overline{\text{A}}$-extractions — specifically, clefting and topicalization — in Bikol, an Austronesian language of the central Philippines closely related to Tagalog.[1] At first glance, Bikol exhibits a familiar Philippine voice system. In example (1a), the theme *lalaki* 'man' has been chosen as the designated pivot and therefore is in nominative case. Patient Voice morphology on the verb reflects that the nominative argument is the verb's theme. Local clefting is limited to this pivot argument, as in (1b,c). Clefting of the non-pivot agent *eskwela* 'student' in (1c) is ungrammatical both when retaining its original genitive case marker or changing it to nominative case. Local clefting thus manifests the basic pivot-only extraction asymmetry predicted of Philippine voice system languages.

(1) **Local clefting is limited to the pivot:**

a. Baseline Patient Voice (PV) clause:

G<in>adan su lalaki kaso eskwela.
PV-kill NOM man GEN student
'The student killed the man.'

[1]The following abbreviations are used in glosses for Bikol examples: AV = Actor Voice, PV = Patient Voice, LV = Locative Voice, BV = Benefactive Voice (see §2); NOM = nominative, GEN = genitive, DAT = dative; DEM = demonstrative; NEG = negation. All uncredited data come from our elicitation with two speakers of Bikol in Singapore, led by the second author.

b. Grammatical pivot (theme) cleft:

Su lalaki su [g<in>adan ___ kaso eskwela ].

NOM man NOM PV-kill GEN student

'It's the man that the student killed.'

c. Ungrammatical non-pivot agent cleft:

*Su / kaso eskwela su [g<in>adan su lalaki ___ ].

NOM / GEN student NOM PV-kill NOM man

Intended: 'It's the student that killed the man.'

In contrast, we observe that local topicalization in Bikol can target both pivots and non-pivot agents. Examples (2a) and (2b) are both grammatical and express the same proposition that 'The student killed the man.' In (2a), the pivot *lalaki* 'man' is topicalized to a pre-verbal position, whereas in (2b), the non-pivot agent *eskwela* 'student' is topicalized to a pre-verbal position. When the non-pivot agent is topicalized in (2b), its case marking changes to be in nominative case, resulting in a clause with two nominative phrases; however, (2b) unambiguously means 'The student killed the man,' and not 'The man killed the student.'

(2) **Grammatical topicalization of theme pivot and non-pivot agent:**

a. Su lalaki [g<in>adan ___ kaso eskwela ].

NOM man PV-kill GEN student

b. Su eskwela [g<in>adan su lalaki ___ ].

NOM student PV-kill NOM man

'The student killed the man.'

The availability of non-pivot agent topicalization in (2b) is surprising against the backdrop of the widely-discussed pivot-only restriction on $\overline{\text{A}}$-extraction in these languages.[2] In addition,

[2]Examples similar to (2b) in other, related Philippine languages have been observed in passing by some authors. However, such examples are few and far between and they have not received serious attention. We return to this point in the conclusion.

the availability of two nominative-marked arguments in (2b) raises questions for the nature of nominative case and the interaction of voice marking and case in these languages, which we will address.

The core of our proposal will be that clefting and topicalization involve probes with different featural specifications: Clefting involves a head which probes for a [D] feature — following Aldridge's (2004; 2017) proposals of a [$\phi$] feature probe for all $\overline{\text{A}}$-extractions in related voice system languages — whereas topicalization probes for a [TOP] feature. Both probes must target the structurally closest matching goal (Rizzi, 1990, 2001; Chomsky, 1995, 2000) and then move it. Following the work of Aldridge (2004, 2008), Rackowski and Richards (2005), and others, the pivot argument in Austronesian voice system languages is the highest argument in the lower phase, in a (outer) specifier of $v$P. Due to their differing featural specifications, clefting cannot attract another DP past the pivot, whereas topicalization can skip the highest DP (the pivot) and attract a non-pivot agent occupying the inner specifier of $v$P. When the agent is itself the pivot, in Actor Voice (AV), it is the only DP at the edge of the $v$P phase. Probing obeys Phase Impenetrability (Chomsky, 2000), explaining the unavailability of non-pivot theme topicalization, illustrated in (3).

(3) **Non-pivot themes cannot be topicalized:**

*Su eskwela [nag-gadan su lalaki ___ ].
NOM student AV-kill NOM man
Intended: 'The man killed the student.'

Support for our locality-based approach will come from the behavior of long-distance clefting. In contrast to local clefting which is restricted to pivots, as in (1) above, long-distance clefting can target non-pivot agents, as seen in (4). We argue that such examples involve a step of non-pivot agent topicalization within the embedded CP which makes the embedded non-pivot agent the highest DP within the embedded clause, which is then the closest target for matrix clefting.

(4) **Long-distance cleft of embedded non-pivot agent:**

Su eskwela su [pig-balita ning radyo [$_{CP}$ na g<in>adan su lalaki ___ ]].
NOM student NOM PV-report GEN radio that PV-kill NOM man
'It's the student that the radio reported *t* killed the man.'

This and additional data inform our description of the nature of the basic Austronesian pivot-only extraction restriction, obeyed in Bikol by local clefting. We argue that the observed "pivot-only" extraction restriction must be characterized in terms of syntactic locality, reflecting the attraction of the structurally closest DP target, rather than any requirement to attract pivots or even nominative DPs.

Additionally, we will also discuss hanging topic left dislocation (HTLD), a non-movement-derived form of topic with an obligatory corresponding pronoun. As a non-movement construction, HTLD can be used for any DP argument: pivots, non-pivot agents, as well as non-pivot themes. Just as long-distance clefts can be fed by movement topicalization as in (4), allowing for non-pivot agent clefts, long-distance clefts can be fed by embedded HTLD. This results in long-distance clefts with corresponding embedded pronouns, with no "extraction" restriction, unlike gapped clefts.

This paper is structured as follows. Section 2 introduces background on the interaction of case marking and voice morphology in Austronesian voice systems and how these properties manifest in Bikol. Local $\overline{\text{A}}$-extraction facts are presented in section 3, followed by our core analysis in section 4. In section 5, we discuss long distance clefting, which will provide the motivation for a locality-based characterization of the Austronesian extraction restriction. Along the way, we will describe two different types of topicalization, the organization of the $v$P phase edge, and the determination of morphological case in Bikol.

## 2 Case and voice in Bikol

In this section we will introduce basic properties of Bikol morphosyntax which will be relevant for the subsequent study. Many Austronesian languages, including Bikol, exhibit a particular

constellation of case marking, verbal morphology, and extraction interactions which have been termed a "voice system." A summary of these key properties is reproduced in (5):

(5) **Characteristics of Austronesian-type voice systems:**

(Erlewine, Levin, and Van Urk, 2017: 376)

a. A privileged argument: One argument is designated the "pivot," and is realized in a particular morphological form and/or structural position, regardless of its original thematic role or grammatical function.

b. Articulated voice morphology: Morphology on the verb varies with the choice of pivot, including options for taking certain oblique arguments as pivot.

c. Extraction restriction: $\overline{\text{A}}$-extraction (*wh*-movement, relativization, topicalization, etc.) is limited to the pivot argument.

d. Marking of non-pivot agents: Non-pivot agents are morphologically marked, often coinciding with the form of possessors (i.e. genitive case).

Voice systems were made famous through the study of Philippine languages — especially Bikol's sister language Tagalog — and therefore are sometimes referred to as "Philippine-type" alignments. However, such voice systems are attested in many other Austronesian languages beyond the Philippines. Chen 2017 describes it as "a typologically unique grammatical system found in nine of the ten Austronesian primary branches" (p. 1).[3] See Erlewine, Levin, and Van Urk 2017 and Chen 2017 for recent overviews of the properties of such voice systems and their analysis. As is noted in these works, there is significant variation in the terms used for such systems in previous literature.[4]

[3]Erlewine, Levin, and Van Urk (2015; 2017; in prep.) observe that the Nilotic language of Dinka (South Sudan) also exhibits all of the characteristic properties of voice systems in (5), leading them to refer to these voice systems as "Austronesian-type."

[4]Other terms for what we call the "pivot" here include terms such as "subject," "trigger," "topic," and "focus." We especially avoid the latter terms, as they suggest particular information-structural notions. As we will see, the pivot argument is often an information-structural topic or focus, but under certain circumstances, non-pivot arguments can be associated with these categories. Many authors refer to the pivot as "subject," while acknowledging that some diagnostics appear to track thematic base positions in many of these languages; see e.g. Kroeger 1991, 1993.

The core voice system properties in (5) are all readily observed in Bikol, although in the rest of the paper we will show that the facts surrounding $\overline{\text{A}}$-extraction (5c) are in reality more complicated. In the rest of this section, properties (5a), (5b), and (5d) of the Bikol voice system will be presented. Data on $\overline{\text{A}}$-extraction which partially supports the characterization in (5c) will be presented in the following section.

Canonical word order in Bikol is predicate-initial. Consider the examples in (6) below, which all express the basic proposition that 'The woman bought cheese at a shop for Andrew.' In each example, there is one pivot DP in nominative case, in bold in (6), and voice morphology on the verb which correlates with this choice of pivot argument. The pivot can be the thematic agent (6a) or theme (6b), but can also be a non-core thematic argument such as a location (6c) or a beneficiary (6d) which is otherwise expressed as an oblique. Post-verbal word order is free; only one word order is given for each example here.

(6) **Voice alternation in Bikol:**

a. Actor Voice (AV):

**Nag**-bakal **su** **babayi** ning keso sa tindahan para ki Andrew.
AV-buy NOM woman GEN cheese DAT shop for DAT Andrew

b. Patient Voice (PV):

**Pig**-bakal kaso babayi **su** **keso** sa tindahan para ki Andrew.
PV-buy GEN woman NOM cheese DAT shop for DAT Andrew

c. Locative Voice (LV):

**Pig**-bakal-**an** kaso babayi ning keso **su** **tindahan** para ki Andrew.
LV-buy GEN woman GEN cheese NOM shop for DAT Andrew

d. Benefactive Voice (BV):

**I**-b<**in**>akal kaso babayi ning keso sa tindahan **si** **Andrew**.
BV-buy GEN woman GEN cheese DAT store NOM Andrew
'The woman bought (the) cheese at a/the shop for Andrew.'

Bikol distinguishes three different cases — nominative, genitive, and dative — with a rich

inventory of surface forms that vary based on animacy and number. The table in (7) covers all case marker forms in examples that we will discuss, involving singular noun phrases, as well as the corresponding third-singular animate pronouns and demonstrative pronouns, which are used for inanimate referents. In examples throughout, we will simply gloss these markers as NOM, GEN, or DAT respectively. Genitive and nominative animate pronouns are second position clitics; see also Erlewine and Levin to appear. See Mintz 1973: ch. 2 and McFarland 1974 for more detailed descriptions of these inventories.

(7) **Case markers for singular noun phrases and corresponding pronouns:**

| | case markers | | pronouns | |
|---|---|---|---|---|
| | proper name | common noun | 3sg animate | demonstrative |
| nominative | *si* | *su* | *=siya* | *ito* |
| genitive | *ni* | *kaso/ning*[5] | *=niya* | *kaito*[6] |
| dative | *ki* | *sa* | *sainya* | *kaito* |

Non-pivot core arguments are generally in genitive case. In addition, specific non-pivot themes appear in dative case rather than genitive case as in (8), but all non-pivot agents are in genitive case.[7]

(8) **Specific non-pivot themes are dative-marked:**

Nag-hiling **sa** babayi si Andrew.
AV-see DAT woman NOM Andrew
'Andrew saw the woman.'

[5]Genitive inanimate common nouns are introduced by *ning*. Genitive animate common nouns can be preceded by *kaso* or *ning*.

[6]*Kaito* is both the genitive and dative form of the demonstrative (Mintz, 2019: 184).

[7]The Tagalog equivalent of this Differential Object Marking is well-studied; see e.g. Schachter and Otanes 1972 and, more recently, Latrouite 2011a and Sabbagh 2016. From a broader, pan-Austronesian perspective, Chen 2017 describes such markers as a variant of the case marker for non-pivot themes (in our terms, genitive) which encodes definiteness or specificity, rather than as a distinct case.

Although the voice system allows for different arguments to be the pivot and hence nominative, in the canonical, predicate-initial word order, it is not possible for two arguments of the clause to simultaneously be nominative. This explains the ungrammaticality of (9) below, in contrast to (6b) above. The ungrammaticality of (9) even with *sa tindahan* intervening between the two core arguments shows that the ungrammaticality of (9) is not due to a simple ban on adjacent nominatives.

(9) **Only one (post-verbal) argument may be nominative:**

*Pig-bakal **su** babayi (sa tindahan) **su** keso.
PV-buy NOM woman DAT house NOM cheese
Intended: 'The woman bought the cheese (at the store).'

It's worth noting that this "voice" system descriptively differs from familiar "voice" alternations in European (and other) language families. First, neither the Actor Voice nor Patient Voice appears to be morphologically or syntactically simpler on the surface, leading some authors to refer to such systems as "symmetric" voice systems; see especially Foley 2008. Second, in the Non-Actor Voices (NAV) — which some authors refer to as "passives" — the agent argument continues to be a DP core argument of the clause, rather than a demoted oblique. The present paper will in fact offer further support for the view that NAV agents are full-fledged DP arguments.

Finally, we note that there is a not insignificant tradition of describing Philippine languages as exhibiting ergative/absolutive alignment. See for example Payne 1982, DeGuzman 1988, Gerdts 1988, Mithun 1994, and Aldridge 2004. Under this view, Actor Voice clauses are formally intransitive, with an oblique theme, and Non-Actor Voice clauses are formally transitive. Pivots are absolutive and NAV agents are ergative, with the case on non-specific AV themes then being a homophonous oblique. On this point, see especially Aldridge's (2004; 2012) ergative analysis for Tagalog, whose voice morphology and case facts parallel the Bikol facts above. The pivot-only $\overline{\text{A}}$-extraction restriction is then an absolutive-only extraction restriction, which is also attested in other language families where the "ergative" designation is less controversial, such as Inuit, Mayan, and Salishan. See Deal 2016 and Polinsky 2017 for two recent overviews of such syntac-

tic ergativity cross-linguistically. For two recent, critical reviews of the ergative hypothesis for Austronesian-type voice system languages, see Erlewine, Levin, and Van Urk 2017 and Chen 2017.

In this paper we use the terms nominative and genitive for the two core cases in Bikol, as in the earlier examples in this section, and later present an analysis for Bikol case and voice in these terms. However, the empirical contribution of our paper as well as its theoretical import is logically separable from this choice. Our core proposal for Bikol extraction facts, in section 4, in fact largely follows the syntax for Austronesian voice systems proposed in Aldridge's work. Lessons for the analysis of syntactic ergativity — to the extent that Philippine voice system languages should be described as ergative — will be presented at the end of section 5.

## 3 Local clefts and topics

In this paper we discuss the clefting of DPs and two types of DP topic constructions in Bikol, which we refer to as topicalization and hanging topic left dislocation (HTLD). We limit our attention to dependencies with DPs, as the movement of non-DPs behaves quite differently in Philippine languages.[8] In this section, we specifically consider local clefts and topics. As in many other Austronesian languages, DP *wh*-questions are formally clefts — see Potsdam 2009 for an overview — and therefore will not be described separately. In the interest of space, we will concentrate on extractions of agent and theme arguments from Actor Voice (AV), Patient Voice (PV), and Locative Voice (LV) clauses.

### 3.1 Clefts

As we have noted above, it is often claimed that only the pivot can be $\overline{\text{A}}$-extracted in voice system languages — famously described as a "subject-only" restriction by Keenan and Comrie (1977) (see footnote 4) — including in closely related Philippine languages (Kroeger, 1991; Reid and Liao, 2004; Aldridge, 2004; Rackowski and Richards, 2005). This characterization indeed holds for local clefting, our first $\overline{\text{A}}$-construction. Clefts have two parts: the exhaustive focus, or focus-containing

[8]See for example the discussion of Tagalog "adjunct fronting" in Kroeger 1991, as well as Hsieh in prep. We do discuss the behavior of some pre-verbal PPs briefly in section 4.4.

phrase, and the background (a gapped clause), separated by a nominative case marker.[9] Example (10) shows that only the agent pivot can be clefted out of an AV clause. Clefting the non-pivot theme in (10b) is ungrammatical, whether retaining the original genitive case marker *ning* or changing the focus to be nominative. Example (11) similarly shows that only the theme pivot can be clefted from a PV clause, as we also saw in (1) above.

(10) **Local clefting from an AV clause:**

a. Grammatical agent pivot cleft:

[$_{\text{focus}}$ Su babayi ] su [$_{\text{background}}$ nag-kaon (*=siya) ning keso ].
NOM woman NOM AV-eat NOM.3sg GEN cheese

'It's the woman that ate the cheese.'

b. Ungrammatical theme cleft:

*[$_{\text{focus}}$ Su/ning keso ] su [$_{\text{background}}$ nag-kaon su babayi ].
NOM/GEN cheese NOM AV-eat NOM woman

Intended: 'It's (the) cheese that the woman ate.'

(11) **Local clefting from a PV clause:**

a. Grammatical theme pivot cleft:

[$_{\text{focus}}$ Su lalaki] su [$_{\text{background}}$ g<in>adan (*=siya) kaso eskwela ].
NOM man NOM PV-kill NOM.3sg GEN student

'It's the man that the student killed.'

b. Ungrammatical agent cleft:

*[$_{\text{focus}}$ Su/kaso eskwela] su [$_{\text{background}}$ g<in>adan su lalaki ].
NOM/GEN student NOM PV-kill NOM man

Intended: 'It's the student that killed the man.'

[9]The fronted constituent in a cleft is also sometimes called a "pivot," but we avoid this term here. "Pivot" in this paper uniformly refers to the argument cross-referenced by voice.

The ungrammaticality of the nominative clitic pronoun =*siya* in both examples shows that local clefts must be gapped; i.e. they must have a post-verbal pivot gap, not a corresponding pronoun. Post-verbal gaps will generally not be indicated in examples, due to the flexible post-verbal word order mentioned in section 2.

We also present clefts from Locative Voice (LV) in (12) below. (12) shows that only the locative pivot can be clefted. Clefting the non-pivot agent as in (12b) or the non-pivot theme as in (12c) is ungrammatical.

(12) **Local clefting from a LV clause:**

a. Grammatical locative pivot cleft:

[$_{\text{focus}}$ Su tindahan] su [$_{\text{background}}$ pig-bakal-an kaso babayi ning keso ].
NOM store NOM LV-buy GEN woman GEN cheese

'It's at the store that the woman bought cheese.'

b. Ungrammatical agent cleft:

*[$_{\text{focus}}$ Su/kaso babayi ] su [$_{\text{background}}$ pig-bakal-an su tindahan ning keso ].
NOM/GEN woman NOM LV-buy NOM store GEN cheese

Intended: 'It's the woman that bought cheese at the store.'

c. Ungrammatical theme cleft:

*[$_{\text{focus}}$ Su/ning keso ] su [$_{\text{background}}$ pig-bakal-an kaso babayi su tindahan ].
NOM/GEN cheese NOM LV-buy GEN woman NOM store

Intended: 'It's (the) cheese that the woman bought at the store.'

From these examples, we see that cleft-formation can only target the pivot, the argument in nominative case and cross-referenced by voice morphology on the verb. Local clefting thus follows the claimed pivot-only restriction on $\overline{\text{A}}$-extraction (5c). As noted above, DP *wh*-questions are also formed using clefts and therefore follow the extraction restriction observed in (10–11).

### 3.2 Topics

Next, we turn to topics in Bikol. We use the term "topic" to refer to DP arguments in pre-verbal position without an exhaustive focus interpretation. Here we will leave the description of the precise discourse effects of these topics for future work. Topics can be formed in two different ways in Bikol: topicalization and hanging topic left dislocation (HTLD). We will argue that topicalization involves movement, whereas hanging topics are base-generated high.

The examples in (13) involve topicalization of their pivots. (13a) has topicalized an agent pivot from an AV clause and (13b) has topicalized a theme pivot from a PV clause. Topicalization is associated with no intonational break and cannot be resumed by corresponding pronouns. In the following examples, topics — and corresponding pronouns, if any — are in bold.

(13) **Pivot topicalization:**

a. **Su babayi** nag-kaon ning keso.
NOM woman AV-eat GEN cheese

b. **Su keso** k<in>aon kaso babayi.
NOM cheese PV-eat GEN woman
'The woman ate (the) cheese.'

In contrast, hanging topics are followed by an obligatory intonational break and have a corresponding post-verbal pronoun. Consider the examples in (14) below. In (14a), the agent pivot 'woman' is topicalized from an AV clause, followed by an intonational break — indicated by a # — with a corresponding post-verbal nominative pronoun *=siya*, which encliticizes to the verb. In (14b), the theme pivot 'cheese' is topicalized from a PV clause, with a following pause and corresponding full pronoun.

(14) **Pivot HTLD, with prosodic break and pronoun:**

a. **Su babayi #** nag-kaon *(**=siya**) ning keso.
NOM woman AV-eat NOM.3sg GEN cheese
'The woman$_i$, she$_i$ ate cheese.'

b. **Su keso # k<in>aon kaso babayi *(**ito**).**

NOM cheese PV-eat GEN woman NOM.DEM

'The cheese$_i$, the woman ate it$_i$.'

The obligatoriness of the intonational break and corresponding pronoun are generally one-to-one. (An exception is discussed in footnote 20 below.) Throughout this paper, we will give English translations with canonical word order for Bikol examples with topicalization, as in (13) above, whereas we give English translations with hanging topics with corresponding pronouns for Bikol HTLD, as in (14). We have chosen to do this in order to highlight the presence or absence of the corresponding pronoun in the Bikol sentences through their English translations. We should however reiterate that we are making no claims regarding the discourse status of these two constructions which here we call "topics" and, in particular, we make no claim that the information structural properties of these Bikol sentences match those of their English translations.

With these basic descriptions of the two forms of topics in place, we now consider which arguments can be targeted for topicalization and HTLD. Examples (13) and (14) above showed that both topicalization and HTLD can target pivots. Topicalization can additionally target the non-pivot agent of Non-Actor Voice clauses. This is observed in the PV example (15), and we will see the same generalization extend to LV below.

(15) **Non-pivot agent topicalization:**

***Su* babayi** k<in>aon su keso.

NOM woman PV-eat NOM cheese

i. 'The woman ate the cheese.'

ii. * 'The cheese ate the woman.'

Note that the topic in (15) must be in nominative case, even though the corresponding post-verbal position is a genitive case position. Example (16) retains the original genitive case marker on the topic *babayi* 'woman' in (15), resulting in ungrammaticality. Recall that multiple post-verbal arguments cannot be in nominative case; see (9) above.

(16) **Ungrammatical topicalization of non-pivot agent with original case marker:** cf (15)

*__Kaso babayi__ k<in>aon su keso.
GEN woman PV-eat NOM cheese

On the surface, topicalizing a non-pivot agent as in (15) results in a string with two nominative phrases: the pre-verbal topic 'woman' and the post-verbal 'cheese.' However, (15) is unambiguous in its interpretation: the post-verbal nominative phrase is unambiguously the pivot of this PV clause and therefore the verb's theme, whereas the pre-verbal nominative topic is unambiguously the non-pivot agent.[10] Our proposal will account for this restriction.

Although non-pivot agents can be topicalized, non-pivot themes cannot. This is illustrated in (17) below, which attempts to topicalize the non-pivot theme *keso* 'cheese' from an AV clause. The sentence is ungrammatical with *keso* in nominative or its original genitive case.

(17) **Topicalization of non-pivot theme is ungrammatical:**

*__Su/ning keso__ nag-kaon su babayi.
NOM/GEN cheese AV-eat NOM woman
Intended: 'The woman ate (the) cheese.'

The generalization that topicalization can target pivots and non-pivot agents (15) but not non-pivot themes (17) also extends to additional voices. Consider the options for topicalization from a Locative Voice (LV) clause in (18). Interestingly, in LV, where both the agent and theme are non-pivot arguments as determined by the choice of voice morphology, we continue to observe an asymmetry: topicalization can target the non-pivot agent (18b), again resulting in a structure with two nominative phrases, but cannot target the non-pivot theme (18c). The locative pivot can also naturally be topicalized, as in (18a).

[10]This is not simply an effect of animacy. For example, example (2b) in section 1, which is structurally parallel to (15), is similarly unambiguous.

(18) **Topicalization from a LV clause:**

a. Grammatical locative pivot topicalization:

**Su tindahan** pig-bakal-an kaso babayi ning keso.
NOM store LV-buy GEN woman GEN cheese

b. Grammatical non-pivot agent topicalization:

***Su*/*kaso babayi** pig-bakal-an su tindahan ning keso.
NOM/GEN woman LV-buy NOM store GEN cheese

c. Ungrammatical non-pivot theme topicalization:

***Su/ning keso** pig-bakal-an kaso babayi su tindahan.
NOM/GEN cheese LV-buy GEN woman NOM store
'The woman bought cheese at the store.'

In summary, topicalization — which we will argue below to involve movement — does not follow a pivot-only restriction, unlike clefting in (10–12). In particular, clefting of non-pivots was ungrammatical even if the focused constituent was changed to be in nominative case; see (10b), (11b), and (12b,c) above. At the same time, topicalization is not unrestricted: non-pivot agents can be topicalized but not non-pivot themes.

Next we turn to hanging topic left dislocation (HTLD). We saw in example (14) above that HTLD can target pivots. In addition, HTLD can target non-pivot agents as well as non-pivot themes as in (19–20) below. These examples each correspond to the topicalization examples in (15) and (17) above, where we saw that non-pivot agents but not non-pivot themes can be topicalized.

(19) **Non-pivot agent HTLD from PV and LV:**

a. ***Su babayi #*** k<in>aon **=niya** su keso.
NOM woman PV-eat GEN.3sg NOM cheese
'The woman$_i$, she$_i$ ate the cheese.'

b. ***Su babayi #*** pig-bakal-an **=niya** su tindahan ning keso.
NOM woman LV-buy GEN.3sg NOM store GEN cheese
'The woman$_i$, she$_i$ bought cheese at the store.'

(20) **Non-pivot theme HTLD from AV and LV:**

a. ***Su*** **keso** **#** nag-kaon su babayi **kaito**.
NOM cheese AV-eat NOM woman DAT.DEM
'The cheese$_i$, the woman ate it$_i$.'

b. ***Su*** **keso** **#** pig-bakal-an kaso babayi **kaito** su tindahan.
NOM cheese LV-buy GEN woman DAT.DEM NOM store
'The cheese$_i$, the woman bought it$_i$ at the store.'

In these examples of non-pivot HTLD (19–20), the topics themselves are in nominative case, even though their corresponding pronouns are in genitive or dative case.[11] Like (15) and (18b) above, the resulting string has two nominative phrases, but each is unambiguous in its interpretation. The pre-verbal hanging topic must correspond to the post-verbal pronoun.

We conclude that there is no restriction on the DP arguments that can be targeted by HTLD. Below we will argue that this is because HTLD does not involve movement, in contrast to topicalization.

### 3.3 Summary

In this section, we presented data on clefting and two types of topics from local clauses in Bikol. Local clefting obeys the pivot-only extraction restriction. Topicalization can target pivots and non-pivot agents. Hanging topic left dislocation can target any core argument, including non-pivot themes. These possibilities are summarized in (21) below.

(21) **Possible DP targets for local dependencies:**

| | pivot | non-pivot agent | non-pivot theme |
|---|---|---|---|
| clefting | ◯ (10a, 11a, 12a) | × (11b, 12b) | × (10b, 12c) |
| topicalization | ◯ (13, 18a) | ◯ (15, 18b) | × (17, 18c) |
| hanging topic left dislocation | ◯ (14) | ◯ (19a,b) | ◯ (20a,b) |

[11] We gloss the pronouns in (20) as dative as they are specific non-pivot themes, which appear in dative case; see e.g. (8). However, recall that *kaito* is both a genitive and dative form, as noted in footnote 6.

Non-pivot topics involve an apparent mismatch in case marking: the pre-verbal topic is in nominative case, instead of the genitive or dative case of its corresponding post-verbal gap or pronoun. In the next section, we present our analysis for Bikol voice and case, as well as the specific analyses for clefting, topicalization, and HTLD, with additional supporting data.

# 4 Proposal

In this section we present our analysis for the patterns of voice, case, and local dependencies in Bikol introduced in the previous section. A key point which we account for is the ability of topicalization to target non-pivot agents as well as pivots, but not non-pivot themes, in contrast to clefting which is strictly pivot-only and hanging topic left dislocation (HTLD) which is unrestricted. To preview our account, we will propose that topicalization is a movement construction which involves probing for a discourse feature [TOP], whereas clefting involves probing for the feature [D]. The pivot and non-pivot agent are the only DPs at the $v$P phase edge, and so there is no way to target non-pivot themes, even with a [TOP] probe. In contrast, HTLD is a non-movement construction, unrestricted by Phase Impenetrability. We will also discuss the determination of morphological case in Bikol, explaining the simultaneous appearance of two nominative phrases in resulting topic constructions.

Our proposal is presented in three parts. Section 4.1 presents our proposal for case and voice in Bikol. We present our analysis for the two topic constructions in section 4.2, followed by additional supporting data from patterns of multiple topics in section 4.3. Section 4.4 then presents our analysis for Bikol clefts. Note that all dependencies in this section will be local, accounting for the patterns presented in section 3 above. We then discuss long-distance clefting in section 5.

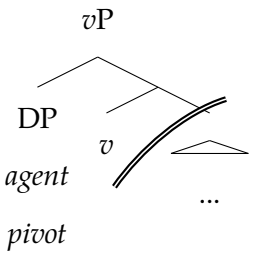
## 4.1 Voice and case in Bikol

We begin by presenting our framework for the voice system and morphological case in Bikol. For the voice system, we will follow the spirit of widely-adopted and influential phase-based approaches to voice systems in Philippine languages, drawing especially on the work of Aldridge (2004, 2008) and Rackowski and Richards (2005). Under such approaches, the pivot DP is distin-

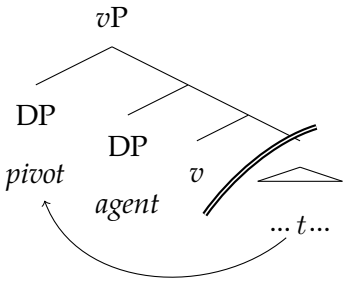
guished by being the highest DP in *v*P — the lower phase of the clause — in a (outer) specifier of *v*P. Agents are base-generated in Spec,*v*P. In Actor Voice (AV) clauses, there is no movement to the edge of the *v*P phase; the agent pivot is base-generated as the only specifier of *v*P and remains the highest DP in the *v*P; see (22a). In Non-Actor Voice (NAV) clauses, a non-agent DP is moved to the outer specifier of *v*P, above the agent DP (22b).[12] Specifiers of *v*P are illustrated on the left in trees, but this does not reflect their word order, which will be discussed below.

(22) **The *v*P phase in AV and NAV clauses:**

a. Actor Voice:




b. Non-Actor Voices:




*v*P is a phase and therefore material within the complement of the phase head *v* will be inaccessible for syntactic operations from above (Phase Impenetrability; Chomsky, 2000). In (22), this domain of impenetrability is illustrated with a double line. This approach predicts a basic asymmetry between AV and NAV clauses: in AV clauses, the *v*P phase edge has only one DP which may be accessible for syntactic operations from above, whereas in NAV clauses, there are two. In the following subsection, we will propose that this is precisely what allows for topicalization to target only pivots (in AV and NAV clauses) and non-pivot agents (in NAV); these are the only DP constituents of the lower phase which can move out. See also Erlewine and Levin to appear for a recent, additional argument for precisely this organization of the *v*P phase edge, based on the inventory of clitic pronouns in Philippine-type voice system languages.

Voice morphology is the realization of the head *v*, which the lexical verb V head-moves to. Aldridge 2004 and Rackowski and Richards 2005 differ in the precise mechanisms that relate the

[12]Complement CPs may also function as the pivot and move to the outer specifier of *v*. This will be important in the analysis of long-distance dependencies, discussed in section 5.

realization of voice morphology to movement of the pivot in NAV clauses.[13] However, both of these approaches agree on the basic geometry for the *v*P phase edge in AV vs NAV clauses, reviewed in (22) above. We adopt this common proposal here. NAV clauses involve movement of the pivot to an outer specifier of *v*P whereas AV clauses involve no such movement, leaving the agent to be the highest DP in the phase and the only DP at in the *v*P phase edge.

Post-verbal word order in Bikol is free, except for a requirement that complement clauses be rightmost. Following Erlewine's (2018) proposal for Toba Batak, another predicate-initial Austronesian voice system language, and its more general extension in Erlewine, Levin, and Van Urk to appear, we propose that all constituents in *v*P are subject to scrambling, such that all linearizations of *v*P with the verbal complex (*v*+V) as the leftmost constituent can be generated. Scrambling of post-verbal constituents has been independently proposed for the sister language Tagalog by many authors: see Kroeger 1991: ch. 5, Richards 1993, Wegmüller 1998, and Rackowski 2002: 22–27.[14]

Next we turn to our proposal for morphological case determination in Bikol. Following Marantz 1991, we propose that morphological case in Bikol may be structurally assigned or realized with context-sensitive defaults. Nominative can be assigned structurally by T, via Agree, as in Aldridge's (2004) analysis of absolutive in "T-type" languages. Marantz (p. 247) proposes that "unmarked case may be sensitive to the syntactic environment; for example, in a language GEN may be the unmarked case for NPs inside NPs (or DPs) while NOM may be the unmarked case inside IPs."[15] In particular, we propose that default case in the *v*P phase is genitive and default

[13] Aldridge proposes that the movement reflects a [D] probe on NAV *v* with the EPP property, which is absent on AV *v*. This is related to her analysis of morphological case in these languages as realizing an ergative-absolutive alignment. In contrast, Rackowski & Richards describe voice morphology as reflecting the choice of the outermost specifier of *v*P, with object shift being one possible motivation for such a movement; see also Rackowski 2002. This latter approach builds on previous accounts of Austronesian voice morphology as so-called *wh*-agreement or case agreement, as in Chung 1982, 1994, 1998 and Pearson 2001, 2005. For our purposes, it is not necessary to identify the underlying mechanism which relates the choice of pivot and choice of voice morphology.

[14] This scrambling could involve syntactic movements to positions above *v*P, as many of these authors propose, together with further verb movement to ensure that the verbal complex is leftmost. Alternatively, it could be a post-syntactic rearrangement of the constituents, as in Clemens 2014: ch. 4. Here, for presentational purposes, we simply consider different linearizations of *v*P without committing ourselves to a particular mechanism for this scrambling.

[15] See Levin 2015: ch. 6 for recent discussion of Marantz 1991. Relevant for us is Levin's discussion which concludes

case in the CP phase is nominative.

This idea that different syntactic domains are associated with morphologically distinct default cases has been recently proposed for both Austronesian and non-Austronesian languages in Baker 2015, Chen 2018, and Erlewine, Levin, and Van Urk to appear. Baker (ch. 4) argues that in Finnish, default case for the Spell-Out of the lower, *v*P phase is partitive, whereas nominative is the default case associated with the higher phase. Chen 2018 documents intricate patterns of case-stacking in Amis, a Formosan Austronesian language, which allows her to explicitly establish default and dependent cases associated with distinct domains. In particular, Amis nominals can undergo case determination multiply, allowing nominals that have moved out of the lower phase to bear both genitive (lower default) and nominative (higher default). From a different perspective, based on a comparative typological study of argument marking and word order patterns across Austronesian voice system languages, Erlewine, Levin, and Van Urk (to appear) similarly argue for genitive as a default case within the lower phase for Puyuma (Formosan) and Tagalog, which is a sister language to Bikol. An anonymous reviewer points out that this choice of genitive as the lower default case may have a diachronic origin, due to the lower portion of clauses in many Austronesian languages having its origin in nominal structures. This idea dovetails neatly with Chen's (2018) discussion, where she explicitly argues that the lower phase of clauses in Amis *is* synchronically nominal in nature.

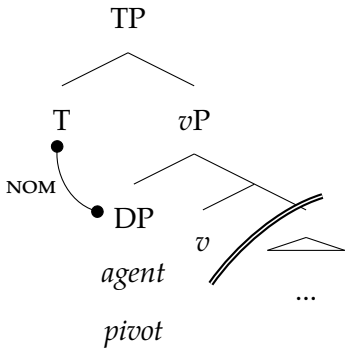
The derivation of AV and NAV clauses as well as the determination of morphological case will be illustrated below. We begin with the AV clause derivation in (23). Following the voice system proposal above in (22), the agent is base-generated in Spec,*v*P and no other argument is moved to the *v*P phase edge. We propose that T bears [PROBE:D] which assigns structural nominative case to its target.[16] As the agent is the highest DP in the *v*P — and, in this case, the only one accessible by Phase Impenetrability — [PROBE:D] on T necessarily targets the agent pivot, which receives nominative case.

---

that "[Marantz's] categories of unmarked and default case must be collapsed" (p. 212). We use the term "default" here.

[16]For explicitness, we use the notation [PROBE:F] to represent a probe for the feature [F], instead of the more common [uF] notation, which different linguists take to stand for "unvalued" or "uninterpretable"; see Pesetsky and Torrego 2007 for discussion. See Erlewine 2018 footnote 13 for more discussion of [PROBE:F] notation.

(23) **Actor Voice clause derivation:**




Any DP which is realized in the *v*P phase and lacks structural case-marking will receive default genitive case (Erlewine, Levin, and Van Urk, to appear, in prep.). This accounts for the genitive case on non-pivot themes in AV clauses. In addition, as noted above, specific non-pivot themes receive dative case through a separate process (see footnote 7) and therefore will not receive default genitive. The surface form computed for an AV clause with a non-specific theme is presented in (24). Recall that the linear order of constituents in the *v*P is subject to scrambling, with the only constraint being that the verbal complex be leftmost.

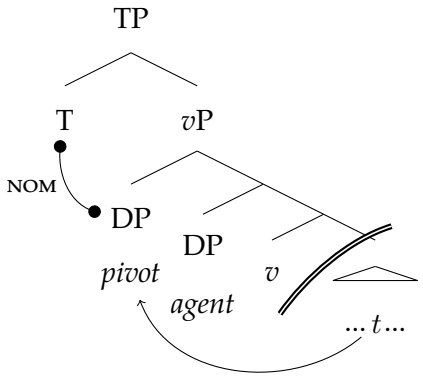
(24) **An Actor Voice clause at PF:**

"AV-V NOM=$DP_{ag/pivot}$ GEN=$DP_{th}$" *or* "AV-V GEN=$DP_{th}$ NOM=$DP_{ag/pivot}$"

Next we turn to the derivation of Non-Actor Voice clauses. This is illustrated with the tree in (25). As we introduced above, in NAV clauses, a non-agent DP moves above the agent to an outer specifier of *v*P. [PROBE:D] on T will find the highest DP, which is the pivot, and assign it structural nominative case.[17] The non-pivot agent has not received structural case, so it will receive default genitive as it is in the *v*P.

[17]We assume that pivots of other voices such as Locative Voice do not receive structural case lower, just as the theme pivot in PV does not in (26). Such pivots could originate in an applicative layer, as proposed by Rackowski 2002, Aldridge 2004, and Rackowski and Richards 2005, or could have started with a preposition or oblique case which is then undone through a process such as Baker's (1988) Preposition Incorporation, as suggested in Guilfoyle, Hung, and Travis 1992: fn. 7 and Kroeger 1990. See also Erlewine, Levin, and Van Urk in prep. However, for our current purposes, we will concentrate on PV examples as illustrative of NAV clauses more generally.

(25) **Non-Actor Voice clause derivation:**



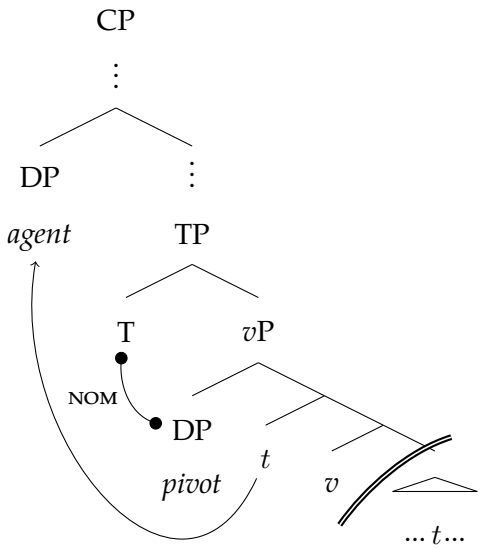

(26) **A Non-Actor Voice clause at PF:**

"NAV-V NOM=$DP_{pivot}$ GEN=$DP_{ag}$" *or* "NAV-V GEN=$DP_{ag}$ NOM=$DP_{pivot}$"

It's worth highlighting that the *v*P phase boundary is relevant here in two distinct senses. For purposes of probing and movement, the complement of the phase head *v* constitutes a distinct domain, inaccessible for higher probing (Phase Impenetrability). This boundary is indicated by the double line in the trees above. However, for purposes of linearization (scrambling) and default case calculation, it is the entire *v*P maximal projection, including its specifiers, that behaves as a unit. Unless moved higher, specifiers of *v*P are linearized post-verbally and subject to scrambling together with all other *v*P-internal constituent. Non-pivot agents receive default genitive case, just as (non-specific) non-pivot themes do. We suggest that this distinction correlates with the timing of the relevant operations: probing is a narrow-syntactic operation and is sensitive to the double line (Phase Impenetrability), whereas linearization and default case determination takes place post-syntax, at PF, where the entire *v*P behaves as one unit.

Finally, we discuss the calculation of morphological case for a constituent which moves out of the *v*P phase. First consider the movement of pivots. Pivots receive structural nominative and will retain this structural case when moved. However, the situation is more complicated when a non-pivot DP moves. Due to the organization of the *v*P phase edge, this can only occur with non-pivot agents. (We discuss the mechanisms of probing which lead to movement of non-pivot agents in the following section.)

(27) **Movement of a non-pivot agent:**




We propose that any DP without structural case which is pronounced in the CP phase will receive default nominative. Non-pivot agents have no source of structural case, so their morphological case realization will depend on the phase in which they are pronounced. If the agent stays within the *v*P phase, it appears with default genitive. But if an agent moves out into the CP phase, as in (27), it will appear in nominative case. The PF realization of a structure as in (27) is sketched in (28).

(28) **A moved non-pivot agent at PF:**

"**NOM**=$\text{DP}_{\text{agent}}$ ... [$_{v\text{P}}$ NAV-V NOM=$\text{DP}_{\text{pivot}}$]"

There are therefore two sources of surface nominative case in our proposal: structural nominative via Agree with T and default nominative in the CP phase. In (28), the post-verbal pivot DP bears structural nominative whereas the pre-verbal non-pivot agent bears default nominative by virtue of its position in the CP phase. As noted by Schütze (2001), identity between structural nominative and a default case in a higher domain of the clause (e.g. on topics) is cross-linguistically common.[18]

[18] An alternative approach would be to claim that all nominative is the result of default nominative in CP, and that Agree with T has the effect of making a DP target behave as if it is in the CP for the purposes of default case calculation.

A consequence of this proposal is that, despite there being multiple sources of nominative case, only one DP can bear nominative case and appear in a post-verbal position. This is the pivot DP which receives structural nominative from T, which can stay in the *v*P and therefore be linearized post-verbally. For any other DP to bear nominative case, it must move out of the *v*P into the CP phase and therefore be in a pre-verbal linear position. This explains the impossibility of multiple post-verbal nominatives, as illustrated in (9) above.

The analysis for Bikol voice and case presented in this section derives the surface morphosyntax for basic AV and NAV clauses in Bikol that we saw in section 2. In addition, two features of this approach will be important for the analysis of Bikol topics and clefts, which we turn to in the following sections. First, the new proposal that nominals in the CP receive default nominative will be important for deriving the case marking observed on topics. Second, two DPs are at the *v*P phase edge in NAV clauses — the pivot and the non-pivot agent — whereas only the agent pivot is at the phase edge in AV clauses. While this is a feature of previous phase-based accounts for voice system syntax in Rackowski 2002, Aldridge 2004, and Rackowski and Richards 2005, its consequences have not been fully discussed in previous work (except recently in Erlewine and Levin to appear). This organization of the *v*P phase edge will be crucial for explaining the differing extraction restrictions on clefting vs topicalization in Bikol.

### 4.2 Topicalization and hanging topic left dislocation

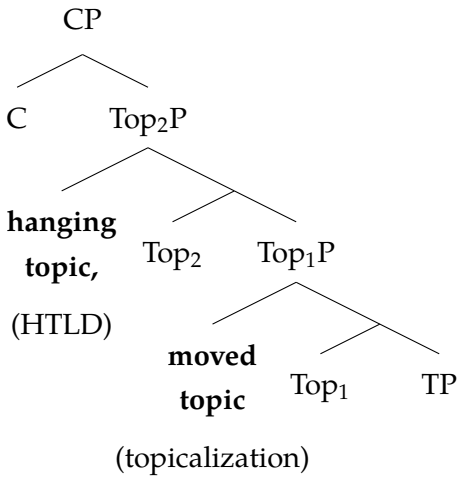
Recall that there are two topic constructions in Bikol: topicalization, which involves a gap and no prosodic break, and hanging topic left dislocation (HTLD), which has a corresponding pronoun and a prosodic break. Topicalization can target non-pivot agents as well as pivot DPs, but not non-pivot themes, whereas HTLD can target any DP argument. In this section we present our analysis for these facts.

We propose two functional heads in the clause periphery, which we simply label $\text{Top}_2$ and

That is, even in cases where the pivot stays in the *v*P, the pivot will receive the CP's default nominative, due to its Agree relationship with the CP phase's T head. This alternative proposal has the conceptual advantage of claiming that all surface forms of nominative case are due to the same mechanism — default case in the CP — but with a new proposal regarding the effect of Agree with T. We will leave the full consideration of this alternative description for future work.

$Top_1$, with $Top_2$ c-commanding $Top_1$. In Rizzi 1997 terms, these can be thought of as heads in a split CP. This organization is illustrated schematically in (29):

(29) **Topic layers in the Bikol clause periphery:**




Topicalization is the result of $Top_1$: [PROBE:TOP] on $Top_1$ fronts any [TOP] goal it finds to Spec,$Top_1$P. $Top_2$ generates hanging topics: a DP is base-generated in Spec,$Top_2$P and binds a pronoun in its scope. Any constituent in Spec,$Top_2$P is followed by a prosodic break. In the following section, we present data from multiple topicalization which supports the higher position for hanging topics.

The claim that topicalization involves movement while HTLD involves base-generation and binding is supported by island-sensitivity (Ross, 1967). Examples (30–31) below show that topicalization but not HTLD is sensitive to islands, as diagnosed by examples with attempted topic dependencies into an adjunct island (a) or relative clause island (b).

(30) **Topicalization is island-sensitive:**

a. *Su babayi pig-uran [$_{island}$ bagu pig-hiling ni Andrew ].
NOM woman PV-rain before PV-see GEN Andrew
Intended: 'It rained [$_{island}$ before Andrew saw the woman].'

b. *Su eskwela na-hiling =ko su lalaki [$_{island}$ na nag-gadan ].
NOM student PV-see GEN.1sg NOM man that AV-kill
Intended: 'I saw the man [$_{island}$ that killed the student].'

(31) **HTLD is island-insensitive:**

a. Su babayi # pig-uran [$_{island}$ bagu pig-hiling =siya ni Andrew ].
NOM woman PV-rain before PV-see NOM.3sg GEN Andrew
'The woman$_i$, it rained [$_{island}$ before Andrew saw her$_i$].'

b. Su eskwela # na-hiling =ko su lalaki [$_{island}$ na nag-gadan sainya ].
NOM student PV-see GEN.1sg NOM man that AV-kill DAT.3sg
'The student$_i$, I saw the man [$_{island}$ that killed them$_i$].'

Further evidence for this movement / non-movement contrast comes from the interpretation of verb-argument idiom chunks (Marantz, 1984). Here we use two idioms for 'mumbling' and 'being a coward':

(32) **Baselines with idiomatic interpretations:**

a. Piga-kaon ni Andrew su tataramon.
PV-eat GEN Andrew NOM words/language
'Andrew mumbles / doesn't speak clearly.' (literally 'Andrew eats words/language')

b. Ma-luya su buot =ko.
AV-weak NOM feelings GEN.1sg
'I am a coward / lack courage.' (literally 'my feelings are weak')

Topicalization retains these idiomatic interpretations, in (33), but HTLD does not, leaving only their literal interpretations available, in (34). This is explained by the topics in (33) being generated together with their predicates and then subsequently moved, whereas the hanging topics in (34) are base-generated high and thus never in a local relationship with their predicates.

(33) **Topicalization retains idiomatic interpretation:**

a. Su tataramon piga-kaon ni Andrew.
NOM words/language PV-eat GEN Andrew
'Andrew mumbles / doesn't speak clearly.'

b. Su buot =ko ma-luya.
NOM feelings GEN.1sg AV-weak
'I am a coward / lack courage.'

(34) **HTLD blocks idiomatic interpretations:**

a. #Su tataramon # piga-kaon kaito ni Andrew.
NOM words/language PV-eat DAT.DEM GEN Andrew
'Words/language$_i$, Andrew eats them/it$_i$.'

b. #Su buot =ko # ma-luya ito.
NOM feelings GEN.1sg AV-weak NOM.DEM
'My feelings$_i$, they$_i$ are weak.'

We now turn to the explanation for the possible targets of topicalization. Probing is subject to the Phase Impenetrability Condition (Chomsky, 2000); therefore, [PROBE:TOP] on Top$_1$ cannot probe into the complement of $v$ and attract a matching goal. In AV clauses, this means that only one DP can be topicalized: the pivot agent; see (22a). In NAV clauses, two DPs are potentially accessible for probing: the pivot and the non-pivot agent, which are both specifiers of $v$P; see (22b). Non-pivot themes are not accessible for topicalization because of Phase Impenetrability.[19] This accounts for

[19] An anonymous reviewer asks whether a theme with a [TOP] feature could be moved to the $v$P phase edge through an intermediate movement mechanism, making it accessible for probing from above. Such a derivation may be possible, but we would then describe the theme as the pivot and the clause would be PV; the result would not be the topicalization of a non-pivot theme from an AV clause. As noted in footnote 13 above, we adopt the shared intuition of Aldridge 2004 and Rackowski and Richards 2005 that there is a correlation between the choice of voice morphology and the choice of DP moved to an outer specifier of $v$P, if any, but do not commit to a particular view of the possible motivation(s) for movement to pivot position. Our analysis is compatible with altruistic intermediate movements being one such motivation.

the patterns of topicalizability documented in section 3: pivots and non-pivot agents are the only DPs which can be topicalized.

[PROBE:TOP] will find the closest accessible target with the [TOP] feature. In cases of non-pivot agent topicalization observed above, the non-pivot agent bears a [TOP] feature but the pivot does not. Because the pivot does not bear the feature that the probe seeks, it does not intervene for the topicalization of the non-pivot agent.

Once a DP occupies Spec,Top$_1$P or Spec,Top$_2$P, if it has not received structural case, it will be realized with default nominative case; see (28) above. This explains the appearance of nominative case on non-pivot topics, as in (35), which correspond to a post-verbal gap or pronoun which receives genitive or dative case.

(35) **Nominative case on non-pivot topics:**

a. **Su babayi** k<in>aon su keso. = (15)
NOM woman PV-eat NOM cheese
'The woman ate the cheese.'

b. **Su babayi** # k<in>aon =niya su keso. = (19a)
NOM woman PV-eat GEN.3sg NOM cheese
'The woman$_i$, she$_i$ ate the cheese.'

c. **Su keso** # nag-kaon su babayi kaito. = (20b)
NOM cheese AV-eat NOM woman DAT.DEM
'The cheese$_i$, the woman ate it$_i$.'

Our analysis also accounts for the unambiguous interpretation of examples such as those in (35) with two nominative arguments. As noted above, the only argument that can receive nominative case and be in a post-verbal position (and therefore within $v$P) is the pivot, which receives structural nominative from T. Therefore, the post-verbal nominative 'cheese' in (35a,b) is necessarily the pivot theme of the PV verb and the post-verbal nominative 'woman' in (35c) is necessarily the pivot agent of the AV verb. The pre-verbal topic then is unambiguously interpreted as the other core argument, corresponding to the post-verbal gap or pronoun.

### 4.3 Multiple topic constructions

The analysis presented above for the two topic constructions is supported by examples with multiple pre-verbal topics. We first consider the two grammatical PV examples in (36). Both topics in (36) are in nominative case, as is independently predicted for each topic construction.

(36) **PV clauses with multiple topics:**

a. Si Pedro # su babayi g<in>adan (=niya).
NOM Pedro NOM woman PV-kill GEN.3sg
i. 'Pedro$_i$, he$_i$ killed the woman.'
ii. * 'Pedro$_i$, the woman killed him$_i$.'

b. Si Pedro # su babayi g<in>adan =siya.
NOM Pedro NOM woman PV-kill NOM.3sg
i. * 'Pedro$_i$, he$_i$ killed the woman.'
ii. 'Pedro$_i$, the woman killed him$_i$.'

Both examples in (36) are PV clauses with two pre-verbal DPs — Pedro and *babayi* 'woman' — but they differ in their interpretation, depending on the choice of post-verbal pronoun. In (36a), Pedro is the agent, corresponding to the post-verbal genitive pronoun, while *babayi* is the pivot theme.[20] In (36b), Pedro is the pivot theme, corresponding to the post-verbal nominative pronoun, while *babayi* 'woman' is the agent. Both examples are unambiguous in their interpretation.

The generalization is as follows. In these sequences of two topics, the first topic is a hanging topic, with a prosodic break and corresponding post-verbal pronoun, whereas the second topic is the result of topicalization. Example (37) below shows that it is not possible to add a prosodic break after the second topic, with or without a break after the first topic, and regardless of the choice of post-verbal pronoun.

[20]But note that the genitive clitic pronoun is optional in this example. We nonetheless analyze Pedro in (36a) as the result of HTLD, as it must be followed by its characteristic prosodic break, but with the pronunciation of its bound pronoun being optional. At this point we are unable to describe the precise conditions which govern this optionality, and leave this question open for future work.

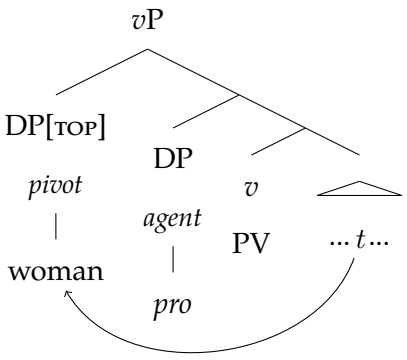
(37) **No prosodic break possible after the second DP:**

*Si Pedro (#) su babayi # g<in>adan (=niya/=siya).
NOM Pedro NOM woman PV-kill GEN.3sg/NOM.3sg

This data in (36–37) supports our proposal that topics with a prosodic break and corresponding pronoun (hanging topics in Spec,Top$_2$P) are structurally higher than topics with no break and no corresponding pronoun (movement-derived topics in Spec,Top$_1$P).

Let's consider the derivation of each of these PV multiple topic examples in (36). We first consider the derivation of (36a). Here there is a hanging topic binding an agent pronoun and a topicalized theme pivot. We therefore begin by constructing a PV clause with the full DP 'woman' with a [TOP] feature as the theme and a pronoun as the agent. Following movement of the pivot theme to an outer specifier of the *v*P, we result in a *v*P organized as in (38):

(38) ***v*P for (36a):**




The rest of the clausal spine is built following the hierarchy in (29), beginning with the merger of T. [PROBE:D] on T will Agree with the closest DP, assigning *babayi* 'woman' nominative case. The agent pronoun is in *v*P so it receives default genitive case. Top$_1$ is merged and its [PROBE:TOP] Agrees with the pivot DP 'woman' and fronts it to Spec,Top$_1$P. Top$_2$ is then merged together with another DP, Pedro, as its specifier, which binds the lower agent pronoun.[21] The clause is complete

[21]We can think of the Top$_2$ head itself as having the semantics of a binder, as in Büring 2005 or similar to Constant 2014's functional head for contrastive topics.

once we merge the C head to form the root CP. The resulting hierarchical structure is as in (39a) below, together with its final linearized structure in (39b).

(39) **Final structure for (36a):**

a. [$_{CP}$ [$_{Top2P}$ Pedro$_i$ Top$_2$ [$_{Top1P}$ woman Top$_1$ [$_{TP}$ T [$_{vP}$ $t_{th/pivot}$ [ *pro*$_{ag,i}$ [ $v$ ...

b. "NOM=DP$_i$ # NOM=DP$_{th/pivot}$ PV-V =GEN.3sg$_{ag,i}$"

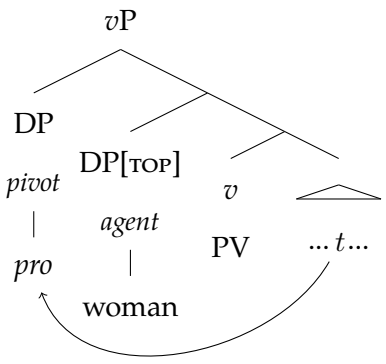
⇒ "si Pedro$_i$, su babayi ginadan=niya$_i$"

'Pedro$_i$, he$_i$ killed the woman.'

Both topics are realized in nominative case: *babayi* 'woman' bears structural nominative from T whereas Pedro receives default nominative in the CP. The hanging topic in Spec,Top$_2$P is followed by a prosodic break. The post-verbal pronoun is genitive and thus appears in the *=niya* form. This results in the correct surface form attested in (36a), and also derives the correct, unambiguous interpretation for this string.

Next we turn to the derivation of example (36b). This example is superficially similar to (36a) but with a post-verbal nominative pronoun in place of the genitive pronoun in (36a), resulting in a markedly different interpretation, 'Pedro, the woman killed him.' We begin by building a PV *v*P with a pronoun theme pivot moving to its outer specifier, above the [TOP]-marked agent DP *babayi*.

(40) ***v*P for (36b):**




We now build the higher phase. T is merged and [PROBE:D] Agrees with the closest DP target, which is the pivot pronoun, assigning it nominative case. Next, Top$_1$ is merged and its [PROBE:TOP]

finds the agent DP and moves it to Spec,Top$_1$P. The Top$_2$ head is merged with its specifier, Pedro, which binds the theme pivot pronoun. After merging C, we yield the structure in (41):

(41) **Final structure for (36b):**

a. [$_{CP}$ [$_{Top2P}$ Pedro$_i$ Top$_2$ [$_{Top1P}$ woman Top$_1$ [$_{TP}$ T [$_{vP}$ *pro* $_{th/pivot,i}$ [ $t_{ag}$ [ $v$ ...

b. "NOM=DP$_i$ # NOM=DP$_{ag}$ PV-V =NOM.3sg$_{th/pivot,i}$"

⇒ "si Pedro$_i$, su babayi ginadan=siya$_i$"

'Pedro$_i$, the woman killed him$_i$.'

Both topic DPs are in nominative case because they are pronounced in the CP phase. In addition, the pronoun also received structural nominative case, resulting in the post-verbal clitic form *=siya*. This results in the correct surface form in (36b), with the correct interpretation.

So far we've looked at multiple topics in a PV clause. Under our proposal both the pivot and non-pivot agent in a NAV clause are at the *v*P phase edge and thus accessible for topicalization, and both arguments can be targeted for HTLD as well. This allowed for the two minimally contrasting examples in (36) above which are both grammatical but with differing interpretations. But now consider multiple topics in an AV clause. Here we observe an asymmetry: Example (42a) is grammatical with its post-verbal dative pronoun, whereas (42b) is ungrammatical with its post-verbal nominative pronoun.

(42) **AV clauses with multiple topics:**

a. Su eskwela # su lalaki nag-gadan sainya.
NOM student NOM man AV-kill DAT.3sg

i. 'The student$_i$, the man killed them$_i$.'

ii. * 'The student$_i$, they$_i$ killed the man.'

b. *Su eskwela # su lalaki nag-gadan =siya.
NOM student NOM man AV-kill NOM.3sg

Intended: 'The student$_i$, they$_i$ killed the man.'

This asymmetry is predicted by our account. Following our proposal and the discussion of the PV examples in (36) above, the outer, hanging topic *eskwela* 'student' in (42) must bind the post-verbal pronoun, with the inner topic *lalaki* 'man' being moved from its base position. In an AV clause, only the agent pivot is at the *v*P phase edge and thus available for topicalization. In contrast, HTLD is not similarly limited as it does not involve movement. This together explains the grammaticality of example (42a). Example (42b) is ungrammatical because the non-pivot theme *lalaki* 'man' would have to be moved from within the lower phase, in violation of Phase Impenetrability. This asymmetry observed in AV clauses with multiple topics in (42) thus further supports both our analysis for the difference between topicalization and HTLD as well as our proposal for the syntax of the *v*P phase edge in AV and NAV clauses, following Rackowski 2002, Aldridge 2004, Rackowski and Richards 2005, and Erlewine and Levin to appear.

Our proposal for the derivation of topics via movement (topicalization) and base-generation (HTLD) thus accounts for basic patterns of Bikol topics, but also more complex multiple topic patterns. Topicalization occurs when $\text{Top}_1$'s [PROBE:TOP] moves its target. Pivots and non-pivot agents are the only arguments at the *v*P phase edge, accessible for [TOP] probing from the higher phase. Non-pivot themes are not accessible for topicalization, due to Phase Impenetrability. In contrast, hanging topics are base-generated high in Spec,$\text{Top}_2$P, involving no movement and thus no sensitivity to the phase edge. Finally, both types of topics are necessarily in nominative case, either by being the pivot which has received structural nominative from T, or by lacking structural case and bearing default nominative in the CP.

### 4.4 Clefts

Recall that, unlike local topicalization, local clefting is limited to the pivot:

(43) **Grammatical local pivot cleft:** = (11a)

[$_{\text{focus}}$ Su lalaki] su [$_{\text{background}}$ g<in>adan kaso eskwela ].
NOM man NOM PV-kill GEN student

'It's the man that the student killed.'

(44) **Ungrammatical local non-pivot agent cleft:** = (11b)

*[$_{\text{focus}}$ Su eskwela ] su [$_{\text{background}}$ g<in>adan su lalaki ].
NOM student NOM PV-kill NOM man

Intended: 'It's the student that killed the man.'

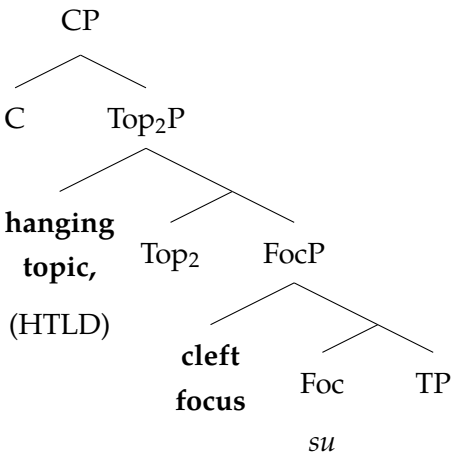
We propose that clefting involves a Foc (Focus) head (Rizzi, 1997) with a probe which attracts the closest [D], [PROBE:D].[22] This is functionally equivalent to Aldridge's (2004; 2017) proposal where, in her terms, extraction involves a [u$\phi$] probe. We propose that the marker *su* that appears in clefts is synchronically the realization of this Foc head, rather than being a true nominative case marker.[23] [PROBE:D] will necessarily target the closest constituent with a [D] feature. We argue that the cleft background — i.e. the sister of Foc — is not a full clause (e.g. CP) and in particular cannot include the Top$_1$ and Top$_2$ projections introduced in the previous section. Concretely, we will identify the sister of Foc as TP, as in (45).

[22] An anonymous reviewer questions the conceptual validity of clefting, associated with exhaustive focus semantics, probing for [D] rather than a corresponding information-structural feature such as [FOC]. First, we note that probing for [FOC] alone will make incorrect predictions, as we describe below. But there is another alternative. Following the Erlewine 2018 analysis for focus fronting in Toba Batak, clefting in Bikol could alternatively be described as due to a *composite* probe on Foc which seeks targets which simultaneously bear both [FOC] and [D] features, [PROBE:FOC+D], together with a restriction that such composite probes in the language are unable to probe past partial matches. Effectively, such a probe would match the highest DP if it bears [FOC], but will not match a non-[FOC] highest DP and also cannot probe past it for a better match. See further discussion in Erlewine 2018 §4.5..

For Toba Batak, Erlewine 2018 specifically motivates this composite probing analysis over conceptually simpler alternatives. For Bikol, however, we are not aware of any configurations where these two analyses would make divergent predictions. We therefore adopt the technically simpler [PROBE:D] analysis for Bikol clefting here, despite the conceptual challenge raised by our reviewer, but leave open the composite probing analysis as another possible alternative.

[23] Clefts in many languages are analyzed as biclausal structures, with the gapped background clause being a type of relative clause. Such an analysis offers an explanation for this nominative case marker: the relative clause structure may be the nominal argument of a higher copular clause. Although such an analysis is likely to be diachronically related to the cleft structures here, we will not adopt this biclausal approach because relativization possibilities for our Bikol speakers do not line up with the restricted possibilities for clefting as in (43–44). We leave a full description and investigation of these relativization options for future work.

(45) **The Bikol clause periphery for clefts:**




Although the background clause does not include topic layers, we will show below that hanging topics can be hosted above the cleft focus, as reflected by the $Top_2$ head above Foc in (45).

We first show that the cleft background cannot include a hanging topic nor a movement-derived topic. As local clefting is limited to the pivot DP and topicalization can target the pivot or non-pivot agent, the most plausible configuration to test would be with a pivot focus and non-pivot agent topic, as in (46) below. Example (46) is ungrammatical, with or without the pause and corresponding pronoun to make the topic a hanging topic.

(46) **Cleft background cannot include a topic:**

*[$_{\text{focus}}$ Su keso ] su [$_{\text{background}}$ **su** **babayi** (#) k<in>aon (**=niya**) ].
NOM cheese NOM NOM woman PV-eat GEN.3sg

Intended: 'It's cheese that the woman (, she) ate.'

Additional, independent evidence for the cleft's background not itself being a full clause comes from high adverbials. Consider the speaker-oriented modifier 'unfortunately' and epistemic adverb 'maybe' in (47). The examples in (48) show that they can appear before or after a topicalized pivot.

---

Although our analysis here is that this *su* in clefts is synchronically the realization of the Foc head (45) rather than a case marker, we will continue to gloss it as NOM throughout.

(47) **High adverbials in sentence-initial position:**

{Sa kamalasan / siguro} nag-inom ning hilo su babayi.
DAT bad.luck / maybe AV-drink GEN poison NOM woman
'{Unfortunately / Maybe} the woman drank poison.'

(48) **High adverbials available before or after movement-derived topic:**

{Sa kamalasan / Siguro} su babayi {sa kamalasan / siguro} nag-inom ning hilo.
DAT bad.luck / maybe NOM woman DAT bad.luck / maybe AV-drink GEN poison
'{Unfortunately / Maybe} the woman drank poison.'

Neither can appear inside a cleft background, as in (49). They can, however, appear above the cleft focus as in (50).

(49) **High adverbials cannot appear inside the cleft background:**

*[$_{\text{focus}}$ Su babayi ] su [$_{\text{background}}$ {sa kamalasan / siguro} nag-inom ning hilo ].
NOM woman NOM DAT bad.luck / maybe AV-drink GEN poison
Intended: 'It's the woman that {unfortunately / maybe} drank poison.'

(50) **High adverbials can appear above the cleft focus:**

{Sa kamalasan / Siguro} [$_{\text{focus}}$ su babayi ] su [$_{\text{background}}$ nag-inom ning hilo ].
DAT bad.luck maybe NOM woman NOM AV-drink GEN poison
'{Unfortunately / Maybe} it's the woman that drank poison.'

This further supports our proposal that cleft backgrounds are not full clauses.

In contrast, it is possible to have a hanging topic above the cleft focus, as in (51). In both examples in (51), the cleft focus is the pivot, according with the generalization that local clefting is limited to the pivot, with the hanging topic corresponding to a non-pivot core argument.

(51) **Hanging topic can precede the cleft focus:**

a. **Su babayi** # [$_{\text{focus}}$ su keso ] su [$_{\text{background}}$ k<in>aon **=niya** ].
NOM woman NOM cheese NOM PV-eat GEN.3sg
'The woman$_i$, it's the cheese that she$_i$ ate.'

b. **Su keso** # [$_{\text{focus}}$ su babayi ] su [$_{\text{background}}$ nag-kaon **kaito** ].
NOM cheese NOM woman NOM AV-eat DAT.DEM
'The cheese$_i$, it's the woman that ate it$_i$.'

This supports our proposal above in (45), whereby a Top$_2$ head for hanging topics can be projected above the cleft's Foc head.

We propose that clefting involves movement of the focused constituent, predicting clefting to be island-sensitive. This is demonstrated with the adjunct island and relative clause island data in (52). The island-sensitivity of clefting here can be contrasted to the island-insensitivity of HTLD in (31) above.

(52) **Clefting is island-sensitive:**

a. *[$_{\text{focus}}$ Su babayi ] su [$_{\text{background}}$ pig-uran [$_{\text{island}}$ bagu pig-hiling ni Andrew ]].
NOM woman NOM PV-rain before PV-see GEN Andrew
Intended: 'It's the woman that it rained [$_{\text{island}}$ before Andrew saw *t*].'

b. *[$_{\text{focus}}$ Su eskwela ] su [$_{\text{background}}$ na-hiling =ko su lalaki [$_{\text{island}}$ na nag-gadan ]].
NOM student NOM PV-see GEN.1sg NOM man that AV-kill
Intended: 'It's the student that I saw the man [$_{\text{island}}$ that killed *t*].'

Recall too that local clefts must have a post-verbal gap in the background clause, corresponding to the focus, further supporting their derivation via movement. (See examples (10–11) above.) This detail will become important in section 5, where we will see that long-distance clefts can have a pronoun in place of a gap.

Finally, our description of clefts as triggered by [PROBE:D] on Foc predicts that non-DP categories such as PPs cannot undergo clefting. Although PPs can also be focus-fronted, as in (53), it is clear that this does not involve the same structure as the clefts with DP focus phrases described

here. First, a pre-verbal focused PP cannot be followed by the *su* marker which is required for DP focus clefts. Second, second-position clitics such as the pronoun *=ako* in (53) can be hosted on these focused pre-verbal PPs, but second position clitics do not climb up to the DP focus of clefts, as has also been described for Tagalog (Kroeger, 1991: 123–125).

(53) **Pre-verbal PP focus is not a cleft:**

Sa eskwelahan **=ako** (***su**) nag-basa ning libro, bako sa halong.
DAT school NOM.1sg NOM AV-read GEN book NEG DAT home
'It's at school that I read books, not at home.'

We will therefore use the term "cleft" here to specifically refer to the DP focus-fronting construction and leave the detailed structure of non-DP focus-fronting in the language for future work.

A second prediction that our theory of clefting makes, as noted by an anonymous reviewer, is that clefting via [PROBE:D] should be able to skip intervening non-DP categories. This prediction is also borne out. Consider example (54), based on the cleft in (43) above, which shows that it is possible to have a locative PP before the verb within a cleft background:

(54) **Clefting skips intervening PPs:**

[$_{\text{focus}}$ Su lalaki ] su [$_{\text{background}}$ **sa halong** g<in>adan kaso eskwela ].
NOM man NOM at home PV-kill GEN student
'It's the man that the student killed at home.'

Recall from (46) above that it is not possible for the cleft background to include any DP topics. Therefore the pivot 'man' must have moved from a post-verbal position within the background clause, within *v*P. The pre-verbal PP is necessarily outside of *v*P. The grammaticality of (54) thus shows that clefting is unaffected by intervening PPs, as is predicted by our proposal that clefts are derived using [PROBE:D] on Foc.[24]

[24]Example (54) also shows that, despite the unavailability of high adverbials and DP topics in cleft backgrounds (46–49), there is space to host locative PPs. This again reinforces the idea that the positions available for pre-verbal DPs and pre-verbal PPs in Bikol fundamentally differ, as we also noted above in relation to focused DPs vs focused PPs.

We note that both aspects of our proposal for the structure of clefts — attraction by [PROBE:D] and the lack of topic projections within the cleft background (46) — are necessary in order to derive their strict pivot-only behavior, reflected in (43–44). If a more selective probe such as a hypothetical [PROBE:FOC] were utilized, we would predict that a non-[FOC] pivot DP could be skipped, allowing the cleft to attract a [FOC] non-pivot agent instead, just as we proposed that topicalization involves [PROBE:TOP].

At the same time, if the cleft background contained $\text{Top}_1$ or $\text{Top}_2$, a topic could be built first, making a non-pivot argument the highest DP within the background clause. Subsequent clefting with [PROBE:D] would be predicted to be able to attract that non-pivot argument, fed by topicalization or HTLD within the background clause. One might imagine that this hypothetical possibility is independently ruled out, for example due to some inherent information-structural incompatibility between topics and foci. However, as we will see in the following section, precisely this interaction — topicalization feeding higher clefting — is possible with full CP embedded clauses in Bikol. Therefore, in order to derive the pivot-only restriction on local clefts, we must additionally propose that the background clause is not a full CP.

### 4.5 Summary

In this section we presented our proposal for Bikol clause structure, morphological case, topics, and clefts, following in large part previous work on the analysis of Philippine voice as in Rackowski 2002, Aldridge 2004, and Rackowski and Richards 2005. Concentrating on the salient difference between the two movement operations of topicalization and clefting, we proposed a locality-based account for the differing extraction restrictions.

Our analysis builds on common Minimalist assumptions regarding the locality of syntactic operations. In particular, movement is subject to Phase Impenetrability and is triggered by a probe which must target its closest goal (Chomsky, 2000, 2001 and many others). Following Aldridge 2004, 2017, the pivot-only restriction on clefting is due to the probe involved — [PROBE:D], or possibly the composite [PROBE:FOC+D] as in footnote 22 — necessarily targeting the highest DP. Topicalization instead involves an information-structurally articulated probe, [PROBE:TOP], which

can skip a non-[TOP] pivot DP to attract a non-pivot agent topic. Phase Impenetrability explains the inability of topicalizing non-pivot themes, which are not at the *v*P phase edge.

We additionally note that the distinct behaviors of local clefting and topicalization forms an argument against the Equidistance hypothesis. Equidistance (Chomsky 2000: 122, 130, 2001: 27) is a hypothesis that predicts that multiple specifiers of a single projection are "equidistant" from higher probes for the purposes of locality. With Equidistance, if both the pivot and non-pivot agent of a NAV clause featurally match a higher probe (for example, [PROBE:D]), the probe could then Agree with either specifier of *v*P without incurring a violation. In Bikol, adopting Equidistance would incorrectly predict that non-pivot agents could be clefted or receive nominative case. In contrast, we adopt a strict view of probe locality without Equidistance, where an outer specifier counts as closer to a higher probe than an inner specifier, but with clefting and topicalization probing for different features. This successfully derives the attested patterns of clefting and topicalization in the language. See also Doggett 2004 for further discussion and an independent critique of Equidistance.

Finally, we note that under our view, nothing about clefting is inherently linked to pivot-hood. As proposed in section 4.4 above, background clauses of a local cleft are simply structured so that the pivot is necessarily the highest DP. We predict that if there is a strategy for making a non-pivot DP the closest to the cleft's probe, clefting would target this non-pivot DP instead. We will see that this is the case in the next section, where we consider long-distance clefts.

## 5 Long-distance clefts and the Austronesian extraction restriction

In this section, we take a closer look at the nature of the famed Austronesian pivot-only extraction restriction. We have seen that, in Bikol, this restriction is obeyed by local clefting but not by local topicalization or HTLD, so our approach will be to further study clefting in Bikol. At first glance, there are a number of different ways to characterize this type of extraction restriction:

(55) **Three characterizations of the classic Austronesian extraction restriction:**

i. Pivot-only: Only arguments cross-referenced by voice ("pivots") can be extracted.

ii. Nominative-only: Only nominative arguments can be extracted.

iii. Locality-based: Only the structurally highest argument can be extracted.

The challenge is to distinguish between these three different descriptions. Every clause has only one pivot, which is in nominative case. Assuming that a topic cannot be formed first (see section 4.4 above), every clause also only has one nominative argument, which is the pivot. And assuming the basic proposal for the hierarchical structure of voice system languages (section 4.1 above), the highest argument in every clause will be the pivot, in nominative case. Therefore, in basic examples of local clefting, these three descriptions in (55) are extensionally equivalent: The pivot is the only nominative argument, and is structurally highest. The study of local clefts alone does not allow us to determine the correct characterization for this restriction.

For this reason, in this section we will study long-distance clefting in Bikol. We begin in section 5.1 with some preliminary discussion of long-distance extraction in voice system languages. The core data on long-distance clefting will be presented in section 5.2. Unlike local clefts, long-distance clefting can target embedded non-pivot agents as well as embedded pivots, which forms an argument against the "pivot-only" characterization of clefting in (55i). We propose that, in such examples, embedded topicalization takes place first and feeds clefting. We support this approach, in section 5.3, with additional data from the interaction of long-distance clefting and embedded topics. In the end, we will also be able to tease apart the "nominative-only" (55ii) and locality-based (55iii) approaches, solidifying our argument that the Austronesian extraction restriction exemplified by Bikol clefting must be described in terms of hierarchical structural configurations and the locality of syntactic operations.

## 5.1 Background: Voice systems and long-distance extraction

Just as $\overline{\text{A}}$-extraction from local clauses is limited in languages with Austronesian-type voice systems, long-distance extraction is also constrained in a similar fashion. Descriptively, extraction out of an embedded clause in Bikol requires that the embedded clause itself be the pivot of the higher

clause. In other words, long-distance $\overline{\text{A}}$-movement is always subextraction from a clausal pivot. This pattern has been well-documented in Tagalog since Kroeger 1991: ch. 7, and is also a major point of the Rackowski and Richards 2005 analysis of Tagalog, discussed above.

Here we consider the long-distance clefting of embedded pivots, as in the grammatical examples in (56). The pivots of the embedded clauses are the theme Andrew in (56a), where the embedded clause is PV, and the agent 'man' in (56b), where the embedded clause is AV.

(56) **Long-distance clefting is grammatical across a PV verb:**

a. Si Andrew su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na g<in>adan kaso lalaki.
NOM Andrew NOM PV-report GEN radio that PV-kill GEN man
'It's Andrew that the radio reported that the man killed *t*.'

b. Su lalaki su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na nag-gadan ki Andrew.
NOM man NOM PV-report GEN radio that AV-kill DAT Andrew
'It's the man that the radio reported that *t* killed Andrew.'

Notice that in both cases the higher verb 'report' is in PV, with its agent 'radio' in genitive case as expected. We can think of the complement clause 'that the man killed Andrew' as the pivot of the verb 'report' in PV, although CPs do not exhibit morphological case marking. The embedded clause's pivot is then subextracted to yield the grammatical cleft in (56).

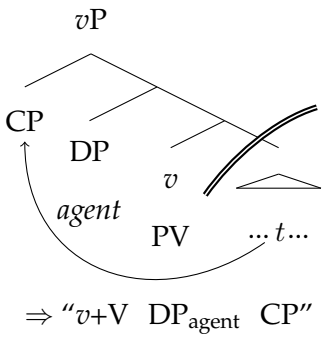
The higher verb must be PV for this long-distance extraction to take place. Example (57) below minimally contrasts from (56a), with the higher 'report' clause in AV, and the result is ungrammatical. The pivot of this higher clause is the agent 'radio,' instead of the complement clause.

(57) **Long-distance clefting is ungrammatical across an AV matrix clause:**

*Si Andrew su [$_{TP}$ **nag**-balita su radyo [$_{CP}$ na g<in>adan kaso lalaki.
NOM Andrew NOM AV-report NOM radio that PV-kill GEN man
Intended: 'It's Andrew that the radio reported that the man killed *t*.'

Under the Minimalist probe-driven conception of movement adopted here, what is important for our purposes is that the highest DP within the embedded CP will count as the "closest" for the cleft's [PROBE:D] in (56), instead of the agent DP of the verb 'report.' Different approaches could be taken, but for concreteness here we briefly present and follow the analysis of long-distance extraction from Rackowski and Richards' study of Tagalog. In grammatical cases of long-distance extraction as in (56), the complement CP itself moves to an outer Spec,*v*P above any agent DP. The verb is in the PV form, correlating with this movement of the theme to Spec,*v*P. This structure is illustrated in (29). Recall that *v*P will be linearized with the verbal complex leftmost, explaining the final word order. CPs are generally rightmost, due either to extraposition or their relative weight.

(58) **An embedded CP in Spec,*v*P (PV), as in (56):**




⇒ "*v*+V $DP_{agent}$ CP"

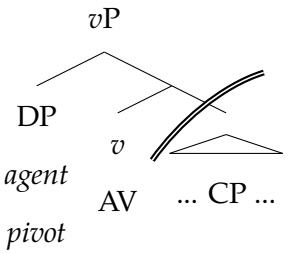
Rackowski & Richards propose that the movement of CP to Spec,*v*P involves the establishment of an Agree relationship with the CP, which makes it transparent for probing from above.[25] The cleft's [PROBE:D] will search into the pivot CP first, matching with the highest DP goal within.[26] The result will be that [PROBE:D] will target the highest DP within the embedded CP rather than the matrix agent DP.

In contrast, if the higher verb is in AV as in (57), the complement CP will not move to Spec,*v*P. Due to Phase Impenetrability, it is impossible to probe into the CP which is inside the lower VP.

[25]Independent evidence from Zulu (Halpert, 2012, 2016: ch. 6, Halpert, 2019), Dinka (Van Urk and Richards, 2015), and Hungarian (Den Dikken, 2018: ch. 4) support the idea that an Agree relationship with a CP is necessary for probing into it.

[26]A related and important detail in this configuration in (29) is that [PROBE:D] on T cannot skip the pivot CP and Agree with the agent DP, which would otherwise predict the availability of nominative case on the agent DP in (56).

(59) **An embedded CP in the complement of $v$ (AV), as in (57):**




The licit and illicit patterns of probing from above for a goal in the lower phase are summarized in (60). In simple cases of probing for a local goal, the goal must be in Spec,$v$P to be accessible for probing from above (60a–b) due to Phase Impenetrability, making pivots and non-pivot agents uniquely visible for probing from above. In cases where the goal is embedded within a CP, that CP itself must move to Spec,$v$P in order to escape Phase Impenetrability (60c–d) and to be made transparent for probing (see footnote 25).[27]

(60) **Patterns of probing:**

| | | | |
|---|---|---|---|
| a. | *PROBE ... [$_{vP}$ | $v_{AV}$ [$_{VP}$ ... GOAL | |
| b. | PROBE ... [$_{vP}$ GOAL | $v_{PV}$ [$_{VP}$ ... | |
| c. | *PROBE ... [$_{vP}$ | $v_{AV}$ [$_{VP}$ ... [$_{CP}$ ... GOAL ... ] | (57/59) |
| d. | PROBE ... [$_{vP}$ [$_{CP}$ ... GOAL ... ] | $v_{PV}$ [$_{VP}$ ... | (56/58) |

For these reasons, in all subsequent examples of long-distance clefting, the higher verb will be in PV. Such examples become ungrammatical with a different choice of voice marking, as in (57) above.

## 5.2 Long-distance clefting

As noted briefly above, long-distance clefting in Bikol differs from local clefting in being able to target non-pivot agents as well as pivots. We have seen in example (56) above that embedded pivots can be clefted long-distance. Example (61), repeated from (4) above, shows that embedded

[27] In their study of the Nilotic language Dinka, Van Urk and Richards (2015) have similarly argued that movement of an embedded CP to Spec,$v$P is necessary in order to extract out of that CP.

non-pivot agents can also be clefted long-distance. This clearly contrasts from the behavior of local clefts, which can only target the pivot.

(61) **Long-distance cleft of embedded non-pivot agent:** = (4)

Su eskwela su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na g<in>adan su lalaki.
NOM student NOM PV-report GEN radio that PV-kill NOM man
'It's the student that the radio reported *t* killed the man.'

We propose that long-distance clefting of non-pivot agents as in (61) involves a first step of embedded topicalization, followed by long-distance clefting. First, we note that topicalization can take place within embedded clauses, moving a non-pivot agent to the embedded CP clause edge. Just as in topicalization in local matrix clauses, the non-pivot agent topic *eskwela* 'student' appears in nominative case in (62).

(62) **Embedded topicalization of a non-pivot agent:**

Pig-balita ning radyo [$_{CP}$ na **su** **eskwela** g<in>adan su lalaki.
PV-report GEN radio that NOM student PV-kill NOM man
'The radio reported that the student killed the man.'

The embedded non-pivot agent *eskwela* 'student' is now the highest DP in the embedded CP in (62). If we cleft from (62), [PROBE:D] will search into the embedded CP, as the higher verb 'report' is PV, and attract the highest DP in the embedded clause. This allows for the successful derivation of the long-distance non-pivot agent cleft in (61).

Now recall that Bikol also has another way to form topics, HTLD, associated with a prosodic break and a corresponding pronoun. HTLD can also target embedded CP edges, as demonstrated with an embedded non-pivot agent hanging topic in (63). Clefting using [PROBE:D] based on this structure in (63) will yield a long-distance non-pivot agent cleft with a corresponding lower pronoun, which is indeed grammatical, in (64). (Topicalization and HTLD can in fact co-occur at an embedded clause edge, which will be discussed in the following section.)

(63) **Embedded HTLD of a non-pivot agent:**

Pig-balita ning radyo [$_{CP}$ na **su eskwela #** g<in>adan **=niya** su lalaki.
PV-report GEN radio that NOM student PV-kill GEN.3sg NOM man
'The radio reported that, the student$_i$, they$_i$ killed the man.'

(64) **Long-distance non-pivot agent cleft with pronoun:**

Su eskwela su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na g<in>adan **=niya** su lalaki.
NOM student NOM PV-report GEN radio that PV-kill GEN.3sg NOM man
'It's the student$_i$ that the radio reported that *they*$_i$ killed the man.' = (61)

Note that the cleft focus in (64) is not followed by the prosodic break associated with hanging topics (63). This is, however, predicted by our account, where the prosodic break associated with HTLD is tied to the pronunciation of a constituent in Spec,Top$_2$P.

Furthermore, in contrast to topicalization, HTLD can also target non-pivot themes. This predicts that an embedded non-pivot theme can be clefted as long as it is fed by HTLD, not topicalization, making the corresponding embedded pronoun obligatory. This is borne out in (65).

(65) **Long-distance non-pivot theme cleft requires pronoun:**

Su eskwela su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na nag-gadan su lalaki *(**sainya**).
NOM student NOM PV-report GEN radio that AV-kill NOM man DAT.3sg
'It's the student$_i$ that the radio reported that the man killed *them*$_i$.'

In this section we've concentrated on the possibility of topicalization or HTLD feeding clefting as a means of clefting embedded non-pivot arguments, but the same approach can also yield long-distance clefts of an embedded pivot DP. As predicted by this approach, long-distance pivot clefts as in (56) can also be accompanied by corresponding lower pronouns, which we propose reflects embedded HTLD followed by clefting.[28]

(66) **Long-distance pivot cleft with pronoun, reflecting embedded HTLD:**

Si Andrew su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na g<in>adan **=siya** kaso lalaki.
NOM Andrew NOM PV-report GEN radio that PV-kill NOM.3sg GEN man
'It's Andrew$_i$ that the radio reported that the man killed *him*$_i$.' = (56)

The patterns of possible long-distance clefting with and without a corresponding pronoun are summarized in (67) below, together with the possibilities for different local dependencies from section 3 above. Local and long-distance clefting differ in two ways: long-distance clefting can have a corresponding pronoun, while local clefting cannot, and long-distance clefting can target a greater range of possible DP arguments as its focus, also dependent upon the presence or absence of a pronoun.

(67) **Possible DP targets for local and long-distance dependencies, based on (21):**

| | pivot | non-pivot agent | non-pivot theme |
|---|---|---|---|
| local clefting (gap) | ◯ | × | × |
| local topicalization (gap) | ◯ | ◯ | × |
| local HTLD (pronoun) | ◯ | ◯ | ◯ |
| long-distance clefting (gap) | ◯ (56) | ◯ (61) | × (65) |
| long-distance clefting (pronoun) | ◯ (66) | ◯ (64) | ◯ (65) |

Our proposal that embedded topicalization and HTLD can feed long-distance clefting predicts precisely this pattern in (67). Unlike the edge of a cleft background which is a TP (45), full CPs include the Top$_1$ and Top$_2$ projections, which can make a non-pivot the highest DP in the embedded CP, which can then be clefted.

(68) **Long-distance clefting fed by topicalization and HTLD:**

a. [$_{FocP}$ DP$_{focus}$ *su*$_{Foc}$ [$_{TP}$ ... [$_{CP}$ *na* [$_{Top2}$ [$_{Top1}$ ___ [ V ... (56, 61)

[28]In the case of long-distance pivot clefting with a gap, as in (56) above, we hypothesize that its derivation may or may not involve a first step of embedded topicalization of the pivot. These derivations may be effectively indistinguishable.

b. $[_{\text{FocP}}\ \text{DP}_{\text{focus}}\ \ su_{\text{Foc}}\ [_{\text{TP}}\ \ldots\ [_{\text{CP}}\ na\ [_{\text{Top2}}\ \_\_\__{\text{i}}\ [_{\text{Top1}}\ \ [\text{V}\ pro_{\text{i}}\ \ldots$ (64–66)

We have already argued in section 4 that topicalization involves movement, explaining why it is limited to pivots and non-pivot agents and leaves a gap, whereas HTLD involves base-generation, explaining why it is not limited to particular arguments and involves a pronoun. Long-distance clefting with a gap can be derived by a first step of embedded topicalization, explaining why it is not strictly pivot-oriented, unlike local clefting. Clefting with a pronoun is only possible long-distance, because it is fed by HTLD, and consequently can target any DP argument. This derives the pattern in (67).

Finally, we note that there is a potential conceptual complication to this proposal. We have argued that a single constituent can first be made an embedded topic and then subsequently clefted, yielding exhaustive focus semantics. Under some approaches to information structure, topic-hood and focus-hood are expected to be mutually incompatible. However, here we reiterate that we use the term "topic" (and the corresponding feature label [TOP] for movement topicalization) descriptively to refer to fronting which is not interpreted as an exhaustive focus. In particular, our core findings here would be unaffected if the movement here described as "topicalization" were instead thought of as a purely optional movement.[29]

### 5.3 Long-distance clefting and embedded topics

We have argued that long-distance clefting can involve a first step of embedded topicalization or HTLD. This approach then predicts non-trivial interactions between long-distance clefting and embedded topics. We will discuss such patterns in this section.

First, as a preliminary, we observe that topicalization and HTLD can simultaneously target an embedded clause edge, just as they can simultaneously target the edge of a simplex clause.[30] For

[29]Following the common Minimalist assumption that all movement is feature-driven, this "optional" movement could be implemented by positing an optional feature [X] of no information-structural import, with the head we have described as "$\text{Top}_1$" in (29) bearing [PROBE:X] instead of [PROBE:TOP].

[30]In particular, recall that there are greater possibilities for multiple simultaneous topics in a Non-Actor Voice clause than in an Actor Voice clause. (See especially (36) vs (42) above.) We therefore use examples with embedded PV clauses

ease of presentation, in this section we will use single and double underlines respectively for outer, base-generated hanging topics and inner, movement-derived topics, as well as their corresponding gaps. The two examples in (69) below are string-identical except for the choice of pronoun in the embedded clause and this corresponds to their different interpretations.[31]

(69) **Multiple topicalization at an embedded clause edge:**

a. Pig-balita ning radyo [$_{CP}$ na su lalaki # su eskwela g<in>adan ?(=niya).
PV-report GEN radio that NOM man NOM student PV-kill GEN.3sg

i. * 'The radio reported that the man$_i$, the student killed him$_i$.'

ii. 'The radio reported that the man$_i$, he$_i$ killed the student.'

b. Pig-balita ning radyo [$_{CP}$ na su lalaki # su eskwela g<in>adan =siya.
PV-report GEN radio that NOM man NOM student PV-kill NOM.3sg

i. 'The radio reported that the man$_i$, the student killed him$_i$.'

ii. * 'The radio reported that the man$_i$, he$_i$ killed the student.'

In (69a), the hanging topic *lalaki* 'man' is interpreted as the non-pivot agent, whereas in (69b), it is interpreted as the pivot theme. Just as we established above in section 4.3 for unembedded multiple topic constructions, the generalization is that the post-verbal pronoun unambiguously corresponds to the higher, hanging topic. See (38–41) above for the derivation of these patterns, which also apply to the embedded clauses in (69).

The question now is what options are possible when we build clefts from these structures in (69). On the surface, the resulting clefts in (70) appear as long-distance clefts of *lalaki* 'man' with a single topic at the edge of the embedded clause. The two examples in (70) again differ only in the choice of pronoun after the embedded verb, and each example is unambiguous in its interpretation. Descriptively, the embedded post-verbal pronoun tracks the interpretation of the

---

here.

[31]The optionality of the genitive clitic pronoun in (69a) and (70a) below parallels its optionality in unembedded contexts as well, as we observed in section 4.3. In these examples, however, the use of the pronoun is preferred, as we would expect of HTLD. As noted in footnote 20, we analyze the variant with no pronoun as involving an unpronounced variant of the genitive pronoun.

fronted cleft focus.

(70) **Long-distance clefts with an embedded topic:**

a. Su lalaki su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na ___ su eskwela g<in>adan ?(=niya).
NOM man NOM PV-report GEN radio that NOM student PV-kill GEN.3sg

i. * 'It's the man$_i$ that the radio reported that the student killed him$_i$.'

ii. 'It's the man$_i$ that the radio reported that he$_i$ killed the student.'

b. Su lalaki su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na ___ su eskwela g<in>adan =siya.
NOM man NOM PV-report GEN radio that NOM student PV-kill NOM.3sg

i. 'It's the man$_i$ that the radio reported that the student killed him$_i$.'

ii. * 'It's the man$_i$ that the radio reported that he$_i$ killed the student.'

Notice that the interpretations of (70a,b) correspond one-to-one to the interpretations of examples (69a,b) above. That is, each example in (70) is unambiguously interpreted as a cleft of the embedded hanging topic 'man' from (69). We indicate this in (70) with corresponding gaps in the embedded hanging topic positions.

We can also be certain that the 'man' in the grammatical (70a,b) has indeed moved from the embedded clause as indicated, rather than being base-generated at the top. As we described in section 5.1 above, movement out of an embedded clause is only possible if the embedded clause functions as the pivot of the higher verb. Long-distance movement in (70a,b) was possible because the higher verb 'report' is in PV. If the higher clause is instead in AV, both (70a,b) become ungrammatical:

(71) **Variants of (70) with AV embedding verb are all ungrammatical:**

*Su lalaki su [$_{TP}$ **nag**-balita su radyo [$_{CP}$ na su eskwela g<in>adan (=niya/=siya).
NOM man NOM AV-report NOM radio that NOM student PV-kill GEN.3sg/NOM.3sg

The unavailability of the (i) interpretation for the string in (70a) also teaches us that it is not possible to extract a post-verbal pivot across a pre-verbal topic. Imagine a derivation where we begin with the embedded clause in (72a). If we were able to cleft the post-verbal pivot 'man'

out of this embedded clause, across the pre-verbal agent topic 'student,' we would predict the availability of the structure in (72b), as a type of long-distance theme pivot cleft. This result in (72b) is string-identical to (70a) and would be predicted to have the unattested (70ai) interpretation.

(72) **Extraction of a post-verbal pivot across a topic:**

a. Embedded clause with topicalization:

... [$_{CP}$ na su eskwela g<in>adan su lalaki

that NOM student PV-kill NOM man

'... that the student killed the man.'

b. Hypothetical cleft of *lalaki* 'man' from (a), across the topic *eskwela* 'student':

*Su lalaki su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na su eskwela g<in>adan ___.

NOM man NOM PV-report GEN radio that NOM student PV-kill

Intended: 'It's the man that the radio reported that the student killed *t*.' = (70ai)

We return now to the derivation of clefts from the embedded multiple topic structures in (69). The examples in (70) showed that clefting of the outer, hanging topics from (69) is possible. What about clefting the inner, movement-derived topics from (69)? This would result in long-distance clefts with embedded hanging topics, marked by their characteristic prosodic gap, with the cleft foci corresponding to gaps in the embedded inner, movement-derived topic positions. These hypothetical structures with their predicted interpretations are given (73). They are judged as ungrammatical.

(73) **Long-distance clefting across an embedded hanging topic:**

a. *Su eskwela su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na su lalaki # ___ g<in>adan (=niya).

NOM student NOM PV-report GEN radio that NOM man PV-kill GEN.3sg

Intended: 'It's the student that the radio reported that the man$_i$, he$_i$ killed *t*.'

b. *Su eskwela su [$_{TP}$ pig-balita ning radyo [$_{CP}$ na su lalaki # ___ g<in>adan =siya.

NOM student NOM PV-report GEN radio that NOM man PV-kill NOM.3sg

Intended: 'It's the student that the radio reported that the man$_i$, *t* killed him$_i$.'

The pattern observed in (70) vs (73) is presented schematically in (74) below. The data presented here strengthens the argument that clefting necessarily attracts the DP which is highest and therefore structurally closest to the probe. From a structure with multiple embedded topics as in (69), it is only possible to cleft the higher, embedded hanging topic (70) and not possible to cleft the lower, embedded movement-derived topic (73).

(74) **Long-distance cleft from an embedded clause with two topics:**

a. $[_{\text{FocP}}\ \underline{\underline{\text{DP}}}_{\text{focus}}\ \ su_{\text{Foc}}\ \ [_{\text{TP}}\ \ldots\ [_{\text{CP}}\ na\ \ [_{\text{Top2}}\ \ \_\_\__{\text{i}}\ \ [_{\text{Top1}}\ \ \underline{\underline{\text{DP}}}\ \ [\ \text{V}\ pro_{\text{i}}\ \ldots$ (70)

b. $*[_{\text{FocP}}\ \underline{\underline{\text{DP}}}_{\text{focus}}\ \ su_{\text{Foc}}\ \ [_{\text{TP}}\ \ldots\ [_{\text{CP}}\ na\ \ [_{\text{Top2}}\ \ \underline{\underline{\text{DP}}}_{\text{i}}\ \ [_{\text{Top1}}\ \ \_\_\_\_\_\ \ [\ \text{V}\ pro_{\text{i}}\ \ldots$ (73)

Similarly, from a clause with one pre-verbal DP topic and one post-verbal pivot, it is not possible to extract the pivot across the topic (72).[32]

Moreover, recall from the previous section that topicalization can feed clefting in cases where there is no embedded hanging topic. This configuration is repeated here in (75) from (68a) above. The ungrammaticality of (74b) therefore cannot be attributed to a general immobility of inner, movement-derived topics.

(75) **Long-distance clefting can target a movement-derived topic:** = (68a)

$[_{\text{FocP}}\ \underline{\underline{\text{DP}}}_{\text{focus}}\ \ su_{\text{Foc}}\ \ [_{\text{TP}}\ \ldots\ [_{\text{CP}}\ na\ \ [_{\text{Top2}}\ \ \ [_{\text{Top1}}\ \ \_\_\_\_\_\ \ [\ \text{V}\ \ldots$ (56, 61)

Such data helps us to distinguish between the "nominative-only" and locality-based characterizations of the Austronesian extraction restriction in (55) above. Although it is true that all DPs which can be clefted are nominative in their lower positions (pivots or topics), being nominative is not a *sufficient* condition to be clefted. That is, the proper characterization of the restriction on clefting cannot be that *any* nominative phrase can be clefted. Even if being nominative is a prerequisite for clefting — see e.g. Deal 2017 on deriving extraction asymmetries through case-discriminating

[32]In contrast, recall that a post-verbal DP can be extracted across a pre-verbal PP; see example (54) above. This possibility is explained in our analysis by the fact that [PROBE:D] must attract the highest DP, but does not interact with intermediate non-DP categories such as PPs.

probing — only the highest nominative DP can be clefted. In fact, because the structurally highest DP within any clause will necessarily be in nominative case (either a pivot which receives nominative case from T, or a topic which has moved into the CP phase), a restriction of the cleft's probe to nominative goals is unnecessary. The extraction restriction inevitably must make reference to locality, and any characterization additionally making reference to nominative case targets must be rejected on grounds of theoretical parsimony.

Finally, we note that this concluding lesson can also be translated into ergative hypothesis terms. As noted in section 2 above, there are many works which describe Philippine voice system languages such as Bikol as exhibiting ergative/absolutive alignment. This includes Aldridge 2004, whose influential approach to the basic clause structure of voice systems, we have adopted here. In brief, for these authors, what we have described here as nominative case is better described as absolutive, and $\overline{\text{A}}$-extraction in these languages exhibit "syntactic ergativity": in particular, an "absolutive-only" extraction restriction. If we were to adopt the ergative hypothesis as a mode of description, we would conclude that the syntactic ergativity observed in Austronesian voice system languages — evidenced in clefts in Bikol — in fact should not be described as an "absolutive-only" extraction restriction (pace e.g. (Deal, 2017)). The appearance of an "absolutive-only" requirement on local clefts is due to the absolutive pivot argument being structurally highest in a cleft clause. The source of this "syntactic ergativity" then is, again, best described as a locality-based effect.

## 6 Conclusion

In this paper we've described and analyzed patterns of clefting and topic formation in Bikol, an Austronesian language of the central Philippines. Our analysis supports the view that the basic Austronesian "pivot-only" extraction restriction is best analyzed in terms of hierarchical configurations and the locality of syntactic operations. $\overline{\text{A}}$-constructions which exhibit this "pivot-only" extraction restriction, such as cleft-formation in Bikol, simply involve probing for [D] (Aldridge, 2004, 2017). In local clefts in Bikol, the pivot is necessarily the highest DP, but in long-distance clefting, embedded topicalization or hanging topic left dislocation (HTLD) can apply first to feed the cleft with a pivot or non-pivot argument as its closest DP target. Furthermore, based on the

interactions of long-distance clefting and embedded topics studied in section 5.3, we conclude that considerations of syntactic locality is a necessary and sufficient condition for explaining the possible patterns for clefting in Bikol. Not only is there no preference for clefts to attract a "pivot," but it is both insufficient and unnecessary to describe clefting as subject to a case-discriminating (e.g. "nominative-only" or "absolutive-only") extraction restriction.

At the same time, we observed that topicalization — which is not bound by the basic pivot-only restriction, even for local topics — is not completely unconstrained. The movement-derived construction of topicalization can target non-pivot agents as well as pivots but not non-pivot themes. This is explained by the organization of the $v$P phase edge in Austronesian voice system languages: the agent is the only specifier of $v$P in Actor Voice, whereas in Non-Actor Voices, the pivot moves to an outer specifier of $v$P, resulting in two specifiers (Rackowski, 2002; Aldridge, 2004; Rackowski and Richards, 2005; Erlewine and Levin, to appear). $v$P is a phase, with these specifiers of $v$P being the only possible targets for syntactic operations from above. This contrasts with the behavior of HTLD, which can target any DP argument, including non-pivot themes. Evidence from island-sensitivity and idiom interpretation motivates the view that HTLD does not involve movement, unlike topicalization and clefting.

Our investigation here also contributes to the theory of case determination, offering new evidence for the domain-sensitive approach to morphological case determination as in Baker 2015, where "default" case (Marantz, 1991) can take different forms in different domains of a single clause. In particular, we claim that genitive is the default case for DPs within the $v$P phase and nominative is the default case for DPs outside of $v$P in Bikol. Our evidence for this claim comes in particular from the behavior of movement-derived non-pivot agent topics: non-pivot agents are genitive if post-verbal and nominative if a pre-verbal topic. Independently, we observe that genitive behaves as the default case in the $v$P domain, corresponding linearly to the post-verbal field, as Erlewine, Levin, and Van Urk (to appear) recently argue to be the case for many related languages. The behavior of hanging topics, which are base-generated high, shows that nominative is the default case in the higher domain of the clause, which accords with the observation that nominative is often the default case for similar hanging topic constructions cross-linguistically (Schütze, 2001).

Baker 2015 argues for the idea that different domains of a clause can be associated with different choices of default case from an investigation of the Finnish partitive versus nominative. Our study here offers an important additional, in-depth argument for this theoretical possibility, on top of the additional recent evidence for this idea from patterns of Amis case-stacking in Chen 2018. Our analysis also offers a new, concrete theoretical approach to movement dependencies without case connectivity.

Finally, we conclude with a brief note on variation in the Austronesian extraction restriction(s), both across and within individual languages. We first note that some examples of non-pivot agent topics in other Philippine languages can be found in previous literature. The Tagalog example in (76) comes from De Guzman 1995 and shows that non-pivot agents can be topicalized — with or without a corresponding pronoun — but non-pivot themes cannot.[33] More recent experimental work by Pizarro-Guevara and Wagers 2018 has confirmed that Tagalog speakers accept non-pivot agent topics but not non-pivot theme topics, and that local non-pivot agent topics are much more acceptable than local non-pivot agent clefts.

(76) **Tagalog (De Guzman 1995: 56–57; reproduced in Latrouite 2011b: 69):**

a. **Ang nanay,** lulutu-in (**=niya**) ang isda.
NOM mother cook-PV GEN.3sg NOM fish
'The mother, (she) will cook the fish.'

b. ?? **Ang isda,** mag-luluto (**nito**) ang nanay.
NOM fish AV-cook GEN.DEM NOM mother
Intended: 'The fish, mother will cook (it).'

Example (77) comes from Sells's (2000) discussion of Kapampangan and shows a case of a grammatical non-pivot agent topic. Non-pivot theme topicalization is not attested in this work.

[33] Richards 2005 shows that non-pivot agents in Tagalog can also undergo another type of fronting called *ay*-topicalization, similarly resulting in structures with multiple *ang*-marked phrases, but non-pivot themes cannot.

Miller 1988: 40–41 gives a similar example of a non-pivot agent topic, which is also discussed in Kroeger 1991: 214–215, but the impossibility of non-pivot theme topics is not mentioned by either of those authors.

(77) **Kapampangan (Sells, 2000: 124):**

**Ing lalaki** e **=na** =ya seli ing mangga.
NOM man not GEN.3sg NOM.3sg bought NOM mango
'The man did not buy the mango.'

Notice that in each of these grammatical examples, (a) the agent topic is in nominative case, resulting in a sentence with two nominative phrases, but (b) the interpretation of the sentence is unambiguous, and (c) the pronoun corresponding to the pre-verbal topic is in genitive case. These properties are exactly what we have observed in Bikol non-pivot agent topicalization.

A similar phenomenon is also attested in Seediq, an Austronesian language of Taiwan (Atayalic), but where topics appear in a clause-final position instead of pre-verbally. Example (78) shows this alternation for a PV agent, with the agent in genitive case in the baseline (78a) but in nominative case in (78b) with a corresponding genitive clitic pronoun. In contrast, example (79) shows that such an alternation is unavailable for non-pivot themes. See also Erlewine 2014 for further discussion.

(78) **Seediq non-pivot agent topic (Aldridge, 2004: 44–45):**

a. Wada bube-un na Pawan ka dangi=na.
AUX hit-PV GEN Pawan NOM friend=GEN.3sg

b. Wada **=na** bube-un ka dangi=na **ka Pawan-ni**.
AUX =GEN.3sg hit-PV NOM friend=GEN.3sg NOM Pawan-DEF
'Pawan hit his friend.'

(79) **Seediq non-pivot theme topic (Aldridge, 2004: 45):**

a. M<n>atis patis ka seediq kiya.
AV-write book NOM person that

b. *M<n>atis ka seediq kiya **ka patis**.
AV-write NOM person that NOM book
'That person wrote a book.'

Such examples lead us to suspect that the availability of non-pivot topics, especially with non-pivot agents, may in fact be quite widespread across Philippine-type voice system languages. At the same time, this possibility appears not to be universal across these languages: an anonymous reviewer notes that non-pivot agent topicalization is not tolerated in Malagasy.[34] Within the framework for extraction restrictions developed and defended here, we hypothesize that such variation reflects different featural specifications on the probes involved: if topicalization and clefting both involve the same probe specification such as [PROBE:D] — as in the spirit of Aldridge's (2004) unified account of the Austronesian extraction restriction — we would expect both constructions to be subject to the same extraction restriction.

The possibility of this cross-linguistic variation parallels the variation observed between different $\overline{\text{A}}$-dependencies within Bikol, where clefting involves [PROBE:D] and topicalization involves [PROBE:TOP]. Although $\overline{\text{A}}$-dependencies are often described as a natural class (Chomsky, 1977), subsequent work has also shown that there are important distinctions in this space as well (see e.g. Cinque, 1990; Lasnik and Stowell, 1991; Postal, 1994). The featural specifications of the probes involved may be one way in which different $\overline{\text{A}}$-dependencies are distinguished.

Examples of the form in (76–79) have largely been ignored in previous discussions of Austronesian syntax, but we believe that they are important data points which show that the characterization of all $\overline{\text{A}}$-dependencies in these languages as strictly pivot-only is overly simplistic and ultimately incorrect. The careful investigation of such extraction restrictions — both between different languages as well as between different $\overline{\text{A}}$-constructions in individual languages — will contribute to our broader understanding of the shape of possible variation in $\overline{\text{A}}$-probing. Our investigation into Bikol here offers a first step in this direction and offers a template for the description and analysis of the restrictions on different $\overline{\text{A}}$-dependencies within an individual language.

[34]See Himmelmann 2002, Ross 2002, Blust 2010, and Erlewine and Levin to appear for discussion of the category of "Philippine-type" languages, which includes Malagasy.

## References


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

Aldridge, Edith. 2008. Generative approaches to ergativity. *Language and Linguistics Compass* 10:966–995.

Aldridge, Edith. 2012. Antipassive and ergativity in Tagalog. *Lingua* 122:192–203.

Aldridge, Edith. 2017. $\phi$-feature competition: A unified approach to the Austronesian extraction restriction. In *Proceedings of CLS 52*.

Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Baker, Mark C. 2015. *Case: Its principles and its parameters*. Cambridge University Press.

Blust, Robert. 2010. Review: The Austronesian languages of Asia and Madagascar by Alexander Adelaar and Nikolaus P. Himmelmann. *Oceanic Linguistics* 49:302–312.

Büring, Daniel. 2005. *Binding theory*. Cambridge University Press.

Chen, Tingchun. 2018. Multiple case assignment: An Amis case study. Doctoral Dissertation, Massachusetts Institute of Technology.

Chen, Victoria. 2017. A reexamination of the Philippine-type voice system and its implications for Austronesian primary-level subgrouping. Doctoral Dissertation, University of Hawai'i.

Chomsky, Noam. 1977. On *wh*-movement. In *Formal syntax*, ed. Peter Culicover, Thomas Wasow, and Adrian Akmajian, 71–132. New York: Academic Press.

Chomsky, Noam. 1995. *The minimalist program*. MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–156. MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1–52. MIT Press.

Chung, Sandra. 1982. Unbounded dependencies in Chamorro grammar. *Linguistic Inquiry* 13:39–78.

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

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

Cinque, Guglielmo. 1990. *Types of A'-dependencies*. MIT Press.

Clemens, Lauren Eby. 2014. Prosodic noun incorporation and verb-initial syntax. Doctoral Dissertation, Harvard.

Constant, Noah. 2014. Contrastive topic: Meanings and realizations. Doctoral Dissertation, University of Massachusetts Amherst.

Deal, Amy Rose. 2016. Syntactic ergativity: Analysis and identification. *Annual Review of Linguistics* 2:1–21.

Deal, Amy Rose. 2017. Syntactic ergativity as case discrimination. In *Proceedings of WCCFL 34*, ed. Aaron Kaplan, Abby Kaplan, Miranda K. McCarvel, and Edward J. Rubin, 141–150.

DeGuzman, Videa P. 1988. Ergative analysis for Philippine languages: an analysis. In *Studies in Austronesian linguistics*, ed. Richard McGinn, 323–345. Ohio University Press.

Den Dikken, Marcel. 2018. *Dependency and directionality*. Cambridge University Press.

Doggett, Teal Bissell. 2004. All things being unequal: Locality in movement. Doctoral Dissertation, Massachusetts Institute of Technology.

Erlewine, Michael Yoshitaka. 2014. Subject marking on non-subjects in Squliq Atayal. Manuscript, to appear in *Proceedings of AFLA 20*.

Erlewine, Michael Yoshitaka. 2018. Extraction and licensing in Toba Batak. *Language* 94:662–697.

Erlewine, Michael Yoshitaka, and Theodore Levin. to appear. Philippine clitic pronouns and the lower phase edge. *Linguistic Inquiry* .

Erlewine, Michael Yoshitaka, Theodore Levin, and Coppe van Urk. 2015. What makes a voice system? On the relationship between voice marking and case. In *AFLA 21: The Proceedings of the 21st Meeting of the Austronesian Formal Linguistics Association*, ed. Amber Camp, Yuko Otsuka, Claire Stabile, and Nozomi Tanaka, 51–68. Asia-Pacific Linguistics.

Erlewine, Michael Yoshitaka, Theodore Levin, and Coppe van Urk. 2017. Ergativity and Austronesian-type voice systems. In *Oxford Handbook of Ergativity*, ed. Jessica Coon, Diane Massam, and Lisa deMena Travis, 373–396. Oxford University Press.

Erlewine, Michael Yoshitaka, Theodore Levin, and Coppe van Urk. to appear. The typology of nominal licensing in Austronesian voice system languages. In *Proceedings of AFLA 26*.

Foley, William A. 2008. The place of Philippine languages in a typology of voice systems. In *Voice and grammatical relations in Austronesian languages*, ed. Peter K. Austin and Simon Musgrave, 22–44. CSLI Publications.

Gerdts, Donna B. 1988. Antipassives and causatives in Ilokano: evidence for an ergative analysis. In *Studies in Austronesian linguistics*, ed. Richard McGinn, 295–321. Ohio University Press.

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

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

Halpert, Claire. 2016. Raising parameters. In *Proceedings of WCCFL 33*, 186–195.

Halpert, Claire. 2019. Raising, unphased. *Natural Language & Linguistic Theory* 37:123–165.

Himmelmann, Nikolaus P. 2002. Voice in western Austronesian: an update. In Wouk and Ross (2002), 7–16.

Hsieh, Henrison. in prep. Beyond nominative: A broader view of A-bar dependencies in Tagalog. Doctoral Dissertation, McGill University.

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

Keenan, Edward L., and Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8:63–99.

Kroeger, Paul R. 1990. Case incorporation in Philippine languages. Manuscript, Stanford.

Kroeger, Paul R. 1991. Phrase structure and grammatical relations in Tagalog. Doctoral Dissertation, Stanford University.

Kroeger, Paul R. 1993. Another look at subjecthood in Tagalog. *Philippine Journal of Linguistics* 24:1–16.

Lasnik, Howard, and Tim Stowell. 1991. Weakest crossover. *Linguistic Inquiry* 22.

Latrouite, Anja. 2011a. Differential object marking in Tagalog. In *Proceedings of AFLA 18*, ed. Lauren Eby Clemens, Gregory Scontras, and Maria Polinsky, 94–109.

Latrouite, Anja. 2011b. Voice and case in Tagalog: The coding of prominence and orientation. Doctoral Dissertation, Heinrich-Heine Universität.

Levin, Theodore. 2015. Licensing without case. Doctoral Dissertation, Massachusetts Institute of Technology.

Marantz, Alec. 1984. *On the nature of grammatical relations*. MIT Press.

Marantz, Alec. 1991. Case and licensing. In *Proceedings of ESCOL 8*, ed. Germán F. Westphal, Benjamin Ao, and Hee-Rahk Chae, 234–253.

McFarland, Curtis Daniel. 1974. The dialects of the Bikol area. Doctoral Dissertation, Yale University.

Miller, Barry W. 1988. Non-configurationality in Tagalog. Doctoral Dissertation, University of Michigan.

Mintz, Malcolm W. 2019. *Bikol dictionary*. University of Hawai'i Press.

Mintz, Malcolm Warren. 1973. Case and semantic affixes of Bikol verbs. Doctoral Dissertation, University of Hawai'i.

Mithun, Marianne. 1994. The implications of ergativity for a Philippine voice system. In *Voice: Form and function*, ed. Barbara A. Fox and Paul J. Hopper, number 27 in Typological Studies in Language, 247–277. John Benjamins.

Payne, Thomas E. 1982. Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6:75–106.

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

Pearson, Matthew. 2005. The Malagasy subject/topic as an A′-element. *Natural Language & Linguistic Theory* 23:381–457.

Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Clausal and phrasal architecture: Syntactic derivation and interpretation*, ed. Simin Karimi, Vida Samiian, and Wendy K. Wilkins, 262–294. John Benjamins.

Pizarro-Guevara, Jed, and Matt Wagers. 2018. Agent extraction under patient voice in Tagalog is acceptable: Evidence from acceptability ratings in various A'-dependencies. Presented at AFLA 25.

Polinsky, Maria. 2017. Syntactic ergativity. In *The Blackwell companion to syntax, 2nd edition*, ed. Martin Everaert and Henk van Riemsdijk. Blackwell.

Postal, Paul M. 1994. Contrasting extraction types. *Journal of Linguistics* 30:159–186.

Potsdam, Eric. 2009. Austronesian verb-initial languages and *wh*-question strategies. *Natural Language & Linguistic Theory* 27:737–771.

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

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

Reid, Lawrence A., and Hsiu-chuan Liao. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5:433–490.

Richards, Norvin. 1993. Tagalog and the typology of scrambling. Honors thesis, Cornell University.

Richards, Norvin. 2005. Person-case effects in Tagalog and the nature of long-distance extraction. In *Proceedings of AFLA 12*, ed. Jeffrey Heinz and Dimitrios Ntelitheos, 383–394.

Rizzi, Luigi. 1990. *Relativized minimality*. MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*, ed. Liliane Haegeman, 281–337. Kluwer Academic Publishers.

Rizzi, Luigi. 2001. Relativized minimality effects. In *The handbook of contemporary syntactic theory*, ed. Mark R. Baltin and Chris Collins, 89–110. Wiley-Blackwell.

Ross, John Robert. 1967. Constraints on variables in syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice marking. In Wouk and Ross (2002), 17–62.

Sabbagh, Joseph. 2016. Specificity and objecthood in Tagalog. *Journal of Linguistics* 52:639–688.

Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. In *Subject and topic*, ed. Charles N. Li and Sandra A. Thompson, 491–518. Academic Press.

Schachter, Paul. 1996. The subject in Tagalog: still none of the above. In *UCLA Occasional Papers in Linguistics*, volume 15, 1–61. Los Angeles: Department of Linguistics, UCLA.

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

Schütze, Carson T. 2001. On the nature of default case. *Syntax* 2:205–238.

Sells, Peter. 2000. Raising and the order of clausal constituents in the Philippine languages. In

*Formal issues in Austronesian linguistics*, ed. Ileana Paul, Vivianne Phillips, and Lisa Travis, 117–143. Springer.

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

Wegmüller, Ursula. 1998. *Sentence structure and ergativity in Tagalog*. Number 36 in Arbeitspapiere des Instituts für Sprachwissenschaft der Universität Bern.

Wouk, Fay, and Malcolm Ross, ed. 2002. *The history and typology of Western Austronesian voice systems*. Canberra: Pacific Linguistics.