Extraction and licensing in Toba Batak

Michael Yoshitaka Erlewine, July 2017

Comments welcome: mitcho@mitcho.com

I investigate patterns of preverbal fronting in Toba Batak, a predicate-initial Austronesian language of northern Sumatra. Contrary to the claims of previous work on the language, I show that multiple constituents can be simultaneously fronted, though only in limited configurations. I argue that the distinct heads C and T are present in Toba Batak, with their common division of labor, but extraction patterns are restricted by the limited means of nominal licensing (abstract Case) in the language. In addition, the features of C and T have the option of being bundled together on a single head, inheriting properties of both C and T and probing together for the joint satisfaction of their probes. Such complex probes are unable to skip partially matching goals. This study sheds light on the relationship between western Austronesian voice system languages and the clause periphery in other language families.

Keywords: Toba Batak, Austronesian, Austronesian voice, movement, extraction asymmetries, nominal licensing, licensing by adjacency, C and T, probing, bundling, complex probing

Acknowledgements: This project would not be possible without my Batak teachers, Paris Lubis and Richard Siburian. I thank them for sharing their language with me. I also thank Reinold Limbong and Sopar Amrol Parulian Manik for discussion of judgments. For valuable comments and discussion, I thank Edith Aldridge, Wayan Arka, Louisa Bielig, Mary Dalrymple, Amy Rose Deal, Hadas Kotek, Martina Martinović, David Pesetsky, Norvin Richards, Yosuke Sato, Dylan Tsai, Michelle Yuan, and especially Theodore Levin and Coppe van Urk. Portions of this work were presented at AFLA 23 at the Tokyo University of Foreign Studies, the Australian Linguistics Society meeting at Monash University, GLOW 40 at the University of Leiden, as well as at MIT, the University of Hong Kong, Rutgers University, the University of California Berkeley, and National Tsing Hua University, and benefited from comments and questions there. I also thank Nora Samosir, Hannah Choi, and the Choi-Sihombing family for continued encouragement and support of my study of Toba Batak. All errors are mine.

Contents

1 Introduction

Work on comparative formal syntax has identified two positions in the clause periphery, often called C and T—traditionally for *complementizer* and *tense*—which are associated with two very different sets of syntactic functions (Chomsky, 1986, a.o.). The T head is commonly associated with properties of subjects, including φ-agreement and nominative case assignment, and in many languages triggers movement of the subject to its specifier (the EPP property). In contrast, the C head is the trigger of information-structural movements such as *wh*-movement. This division of labor between C and T is remarkably common across language families of the world.[1]

However, in many languages of the Austronesian language family, such a clear division of labor between the canonical functions of C and T is not immediately apparent. Many Austronesian languages exhibit a "voice" system where one argument of the verb is chosen to be the subject, with $\overline{\text{A}}$-movement limited to this subject argument (see e.g. Keenan and Comrie, 1977). Such extraction asymmetries suggest that the notion of "subject" in many Austronesian languages combines properties often associated with both C and T.

In this paper, I investigate the clausal periphery and patterns of extraction in Toba Batak, a predicate-initial Austronesian language of northern Sumatra. My work here is based on elicitations with four speakers who were born and grew up in Sumatra and currently live in Singapore. Evidence from Toba Batak will shed light on the possible organizations of functional heads in the clause periphery and their roles in probing and left peripheral movement, while also highlighting the role of nominal licensing (Case) in Austronesian voice system languages.

I will argue for three core conclusions. First, I propose that the heads C and T exist in Toba Batak, with their common division of labor—T responsible for Case-licensing and attracting the subject and C responsible for *wh*/focus-fronting—but movement to the clausal periphery is restricted due to limitations of nominal-licensing (Case) in the language. Toba Batak does not have morphological case alternations, so this licensing is abstract Case. *Wh*/focus-movement of a DP to Spec,CP cannot cooccur with the fronting of the subject to Spec,TP, as seen in (1a). However, the movement of a non-DP constituent to Spec,CP can freely cooccur with a DP in Spec,TP (1b).[2]

[1] See Ramchand and Svenonius 2014 and Wiltschko 2014 for two recent discussions of the apparent universality of this hierarchical organization within clauses, as well as discussion of their possible sources.

[2] The following abbreviations are used in examples. ACT(IVE) and PASS(IVE) are "voice" morphemes, described in section 2.1. PN is a marker which precedes proper names, *si*. NEG = negation, AUX = auxiliary, PERF = perfective; PROX = proximal, MED = medial, DIST = distal (see footnote 28); RC = relative clause. NA is used as the gloss for the particle *na* and

(1) **Only non-DPs can move to Spec,CP with the subject in Spec,TP:**

a. *[$_{CP}$ Aha [$_{TP}$ si Poltak [man-uhor ___ ___]]]? (*maN-tuhor* > *manuhor*)
what PN Poltak ACT-buy

'What did Poltak buy?'

b. ✓[$_{CP}$ Andigan [$_{TP}$ si Poltak [man-uhor buku ___ ___]]]?
when PN Poltak ACT-buy book

'When did Poltak buy a book?'

We contrast this pattern in (1) to the familiar patterns of movement in a language such as English, where *wh*-movement to Spec,CP cooccurs with subject movement to Spec,TP and is not limited to non-DPs (2).[3]

(2) **Both DPs and non-DPs can move to English Spec,CP:**

a. ✓[$_{CP}$ What will [$_{TP}$ Stephanie [be [$_{vP}$ ___ buying ___?

b. ✓[$_{CP}$ When will [$_{TP}$ Stephanie [be [$_{vP}$ ___ buying the book ___?

I will argue that the ungrammaticality of the *wh*-fronting in (1a) is due to a lack of Case-licensing for *aha* 'what' in Spec,CP. Toba Batak lacks mechanisms of structural Case-licensing in the lower part of the clause, with the exception of licensing via adjacency to the verb, which licenses *buku* 'book' in (1b). The object *aha* 'what' does not receive structural Case in its base position, nor can it receive Case in Spec,CP, leading to the ungrammaticality of (1a). In contrast, the English *what* in (2a) receives accusative case in its base position before moving to Spec,CP. The *wh*-word *andigan* 'when' in (1b) is not nominal and therefore does not require Case-licensing.

FOC is the gloss for the focus enclitic *do*; both will be discussed in section 5. For Southern Tiwa: AGR = agreement. For Tongan: ERG = ergative, ABS = absolutive.

Most Toba Batak examples here, without preverbal aspectual auxiliaries, could be given past tense or habitual present tense English translations. Auxiliaries will be discussed in sections 4.1 and 5.1.

In many examples, including (1), I will indicate gaps associated with fronted constituents, but I will not indicate which gap corresponds to which fronted constituent, due to the relative freedom of base word order amongst postverbal constituents, as discussed in section 4.2.

[3] I assume here that the subject in English originates in a predicate-internal position and moves to Spec,TP (see e.g. McCloskey, 1997). T-to-C movement of the auxiliary *will* is not illustrated in (2).

Second, I argue that the distinct functions of C and T are sometimes combined on a single head. Following Martinović (2015), I will call such a head *CT*.[4] CT probes for targets which are simultaneously *wh* or focused (the requirement of C) and nominal (the requirement of T). This bundled CT head is used to front nominal *wh*-phrases as in (3) or focused nominals. CT inherits from T the ability to Case-license its target.

(3) **CT attracts a *wh*/focused nominal and Case-licenses it:**

[$_{\text{CTP}}$ Ise [man-uhor buku ___]]?
who ACT-buy book

'Who bought a book?'

CT also inherits from C the ability to probe and attract multiple matching targets. Using CT, multiple nominals can be fronted, as long as they each also satisfy being a *wh* or focus, as in (4) below. Notice that (4) minimally contrasts from the ungrammatical (1a) above in the subject being focused with *holan* 'only.' Both fronted DPs in (4) are Case-licensed by CT.

(4) **CT can front multiple *wh*/focused nominals and Case-license them:**

[$_{\text{CTP}}$ Aha [holan si Poltak] [man-uhor ___ ___]]?
what only PN Poltak ACT-buy

'What did only Poltak buy?'

Examples of multiple DP fronting as in (4) have never before been described and are unpredicted by previous theories of Toba Batak clause structure and *wh*-movement, as in Clark 1992, Baldridge 2002, Cole and Hermon 2008, and Hermon 2009. The grammatical multiple fronting in (4) is also surprising in that it seems to violate the "subject-only" extraction restriction of Austronesian voice system languages, as will be discussed below.[5]

Third and finally, I propose that a complex probe (such as CT) is unable to skip a partially matching goal in Toba Batak. The CT head proposed for (4) seeks a goal which is simultaneously nominal and *wh* or focused. This complex probe cannot skip a partially matching goal. Concretely, this disallows the complex probe on CT from skipping the subject and Agreeing with *wh*/focused nomi-

[4]This relationship could also be modeled through feature inheritance (Chomsky, 2008; Ouali, 2008; Fortuny, 2008; Legate, 2011, a.o.). See footnote 38 below.

[5]I tend to use the terms *fronting* and *extraction* interchangeably to refer to the movement of constituents to a preverbal position.

nal lower in the structure. This is one component for my explanation for the basic "subject-only" extraction restriction observed in the language.

This paper is organized as follows. I begin in section 2 with a brief introduction to Toba Batak word order, the voice system, and *wh-* and focus-fronting. Section 3 presents new data on the clause periphery of Toba Batak. I develop my proposal in section 4 before presenting additional evidence from the particle *na* for the proposed organization of the clause periphery.

2 Preliminaries

2.1 Voice and word order

Toba Batak has a symmetric two-way "voice" system similar to that of neighboring Malayic languages. Consider the examples in (5) below, which are two ways of saying 'Poltak read the book.' The "subject" argument in each sentence is in bold.[6] The prefix on the verb (also in bold) correlates with the choice of subject argument.

(5) a. **Man**-jaha buku **si Poltak**.
ACT-read book PN Poltak
'Poltak read a book.'

b. **Di**-jaha si Poltak **buku**.
PASS-read PN Poltak book

Following previous literature (Van der Tuuk, 1864/1971; Nababan, 1966, 1981; Percival, 1981; Schachter, 1984a; Cole and Hermon, 2008, a.o.), I refer to the prefix *maN-* (5a) as *active* and *di-* (5b) as *passive*, though I should warn against conflation with Indo-European active/passive alternations. In particular, the passive is not valence-decreasing: for example, the agent *si Poltak* continues to be a non-oblique core argument in the passive (5b). Non-subject DPs—the active theme *buku* in (5a) and the passive agent *si Poltak* in (5b)—must be adjacent to the verb, but postverbal word order is otherwise free.[7] Postverbal word order will be discussed again in detail in section 4.2.

The canonical declarative order is predicate-initial, but subject-initial clauses as in (6) are common in elicitation.

[6] What I call "subject" here has also been called the "pivot" or "trigger" in some Austronesianist literature.

[7] Schachter (1984b, 145ff) and Cole and Hermon (2008, 151ff) do report various preferences between different postverbal word orders, but the requirement that the non-subject DP be verb adjacent is the only categorical constraint on postverbal non-clausal arguments. The only other constraint on postverbal word order that I am aware of is that complement clauses are obligatorily extraposed to the right, but this fact is not relevant in this paper.

(6) a. **Si Poltak** [**man**-jaha buku ___].
PN Poltak ACT-read book
'Poltak read a book.'

b. **Buku** [**di**-jaha si Poltak ___].
book PASS-read PN Poltak

If a single DP is fronted, it must be the subject. This is true in the examples in (6) above and is also reflected in the *wh*-fronting contrasts in (7–8) below. Keenan and Comrie (1977) famously noted that relativization is limited to the subject in Toba Batak, as well as in a number of other Austronesian languages. They describe this as a "subject-only" restriction on extraction.

(7) **Agent *wh*-question ⇒ ACTIVE:**

a. ✓**Ise** [**mang**-allang babi ___]?
who ACT-eat pork
'Who ate pork?'

b. *Ise [**di**-allang ___ **babi**]?
who PASS-eat pork

(8) **Theme *wh*-question ⇒ PASSIVE:**

a. *Aha [**man**-uhor ___ **si Poltak**]?
what ACT-buy PN Poltak
'What did Poltak buy?'

b. ✓**Aha** [**di**-tuhor si Poltak ___]?
what PASS-buy PN Poltak

Non-DP constituents do not participate in the voice alternation. In contrast to DPs, the fronting of non-DPs is independent of the choice of voice. For example, the benefactive PP 'for who' can be *wh*-fronted out of both active and passive clauses (9), with corresponding changes in postverbal core argument word order.

(9) **Extraction of non-DPs does not interact with voice:**[8]

a. ✓[$_{PP}$ Tu ise] [**man**-uhor buku **si Poltak** ___]? (*maN-tuhor* > *manuhor*)
for who ACT-buy book PN Poltak

[8]The gaps indicated here are only one option for the base linear position of the PP. *Tu ise* 'for who' could be before or after the subject in the corresponding *wh*-in-situ questions.

b. ✓[$_{PP}$ Tu ise] [**di**-tuhor si Poltak **buku** ___]?
for who PASS-buy PN Poltak book

'[For who] did Poltak buy the book?'

Examples (7–9) here are from my own elicitations but these same patterns are also described by Clark (1984, 1985) and Cole and Hermon (2008). Extraction of DPs is limited to the subject argument, whose choice is cross-referenced by voice morphology, whereas the extraction of non-DPs is independent of the choice of voice.

2.2 Optionality of *wh*- and focus-fronting

Many of the examples that I will discuss below involve the fronting of *wh*-phrases or focused phrases, so I will take a moment to discuss their status in Toba Batak. We have seen examples of *wh*-questions with fronting and this is the preferred strategy in elicitation. However, Toba Batak also allows for *wh*-in-situ. The examples in (10) are three ways of asking the same question: 'Who ate this pork?'

(10) **DP *wh*-movement is optional but preferred:**

a. Ise [mang-allang babi on ___]?
who ACT-eat pork PROX

b. Mang-allang babi on ise?
ACT-eat pork PROX who

c. Di-allang ise babi on?
PASS-eat who pork PROX

'Who ate this pork?'

The in-situ *wh*-word 'who' could be the subject as in (10b) or a non-subject argument as in (10c). For the nominal *wh*-phrase to be fronted as in (10a), it must be subject, as we saw in (7–8) above.[9] A similar paradigm of examples is presented in Silitonga (1973, 102–105).

[9]For many other Austronesian languages, *wh*-questions with fronting similar to (10a) have been analyzed as pseudoclefts; see Potsdam (2009) and citations there. For example, an example such as (10a) would be analyzed with 'who' being a matrix predicate with *mangallang babi on* being a headless relative 'the person that ate the pork.' Hermon (2009)

It's worth noting that the fronted and in-situ *wh*-questions in (10) do not differ in their use conditions. In particular, the in-situ *wh*-questions are not echo questions or so-called "declarative syntax questions" (Bobaljik and Wurmbrand, 2015), neither of which can be embedded. (11) below shows that all three options in (10) can indeed be embedded under the question-embedding verb *boto* 'know.' The possibility of embedding here shows that *wh*-movement is truly optional in Toba Batak.

(11) **Embedded questions can have *wh*-phrases in-situ or moved:**[10]

a. Hu-boto [ise [mang-allang babi on ___]].
PASS.1sg-know who ACT-eat pork PROX

b. Hu-boto [mang-allang babi on ise].
PASS.1sg-know ACT-eat pork PROX who

c. Hu-boto [di-allang ise babi on].
PASS.1sg-know PASS-eat who pork PROX

'I know [who ate this pork].'

The optionality of *wh*-movement also extends to non-DP and adjunct *wh*-words as well, as seen by the embedded 'when' questions in (12). Both linear positions of *andigan* 'when' in (12b) and (12d) are grammatical. Examples (12a,b) are active, whereas (12c,d) are passive, with corresponding changes to the word order of postverbal core arguments.

(12) **Non-DP *wh*-movement is also optional:**

a. Hu-boto [andigan [man-uhor buku ho]].
PASS.1sg-know when ACT-buy book 2sg

b. Hu-boto [man-uhor buku {andigan} ho {andigan}].
PASS.1sg-know ACT-buy book when 2sg when

advocates for such a cleft analysis for Toba Batak while also invoking *wh*-movement elsewhere; see footnote 49. This is relevant to the question, raised at the end of this section, of whether or not the language has a process of *wh*-movement per se. I will describe such examples as straightforwardly involving *wh*-movement from the gap position, with (10a) corresponding to a base structure as in (10b), and I will explicitly argue against the pseudocleft analysis for Toba Batak in section 5.2.

[10]Toba Batak has two suppletive passive voice prefixes which encode agent φ-features: first singular *hu-* and first plural inclusive *ta-*. Full pronouns can also be used for these agents, though the dedicated passive prefixes are preferred.

c. Hu-boto [andigan [di-tuhor ho buku]].
PASS.1sg-know when ACT-buy 2sg book

d. Hu-boto [di-tuhor ho {andigan} buku {andigan}].
PASS.1sg-know PASS-buy 2sg when book when

'I know [when you bought the book].'

The picture is similar for focused constituents with particles such as the 'only' particle *holan*: they can be fronted or in-situ, but are very often fronted.[11]

(13) **Optional but preferred focus-fronting with *holan* 'only':**

a. [Holan si Poltak] [mang-allang indahan ___].
only PN Poltak ACT-eat rice

b. Mang-allang indahan [holan si Poltak].
ACT-eat rice only PN Poltak

'Only Poltak ate rice.'

It's worth stepping back here and noting that, at this point, we have no clear evidence for the existence of distinct processes of "*wh*-" or "focus-fronting" in Toba Batak. Recall that the language independently allows for the fronting of subjects; see e.g. (6). The facts presented thus far are

[11] DPs with focus particles additionally exhibit a strong preference to be the subject, if possible. This is reflected in the ungrammaticality of the passive variants of (13) in (i) below, where 'only Poltak' is the non-subject passive agent. This contrasts from *wh*-DPs which are grammatical in this position, as in (10c/11c) above.

(i) a. *Indahan [di-allang [holan si Poltak] ___].
rice PASS-eat only PN Poltak

b. *Di-allang [holan si Poltak] indahan.
PASS-eat only PN Poltak rice

Intended: 'Only Poltak ate rice.' (=13)

An alternative hypothesis might be that focus particles are banned on non-subject DPs, but we will see later in examples such as (17–19) and (45) that non-subject DPs can bear focus particles. The crucial factor in (17–19) and (45) is that both DP core arguments are formally focused, in contrast to (i) where the subject is not formally focused. I will leave open the nature of this preference—and its application to focus particles but not *wh*-phrases—for future work.

compatible with the language being *wh*/focus-in-situ at its core, together with a general fronting process which can freely front subjects and non-DPs.

In the next section, I turn to patterns of multiple fronting in Toba Batak. One lesson will be that we ultimately must recognize *wh*/focus-fronting as a distinct process in the language, independent of the free fronting of subjects as in (6). For convenience, I will refer to both *wh*-phrases and constituents modified by *holan* 'only' and *pe* 'even' as "formally focused," formalized as [+FOC].

3 Multiple fronting in Toba Batak

As we have seen, Toba Batak syntax is predicate-initial but with many examples with one constituent fronted to a preverbal position. In this section I present new data on the simultaneous fronting of multiple constituents to preverbal positions. Save for one mention of "double fronting" in passing by Emmorey (1984, 45), no previous work has discussed the possibility of multiple fronting in Toba Batak. We will see that multiple fronting is possible in certain limited configurations. The empirical landscape presented here will motivate my proposal for the Toba Batak clause periphery in section 4.

I begin with the consideration of two DP arguments. The characterization given in the previous section and in all previous work on Toba Batak—that only the subject DP can be fronted—immediately predicts that the fronting of multiple core argument DPs should be impossible. And at first glance, this appears to be correct:

(14) ***Wh* agent, non-focused DP theme:**

a. Ise [mang-allang indahan ___]?
who ACT-eat rice

b. Indahan [di-allang ise ___]?
rice PASS-eat who

c. *Ise indahan [mang/di-allang ___ ___]?
who rice ACT/PASS-eat

'Who ate rice?'

Examples (14a,b) are two grammatical forms of the matrix question 'Who ate rice?' As noted above, Toba Batak allows for fronting of the *wh*-word, which must be the subject (14a), and also allows *wh*-in-situ and free fronting of subjects, resulting in (14b). As shown in shown in (14c), these two operations cannot cooccur to yield the *wh* DP followed by the non-focused DP, both in preverbal position, regardless of the choice of voice morphology.[12]

The contrast in (15) below is completely parallel to (14), but with a non-focused agent and *wh* theme. Cole and Hermon (2008, 183) discuss examples such as (14c, 15c) as support for their view that non-subject DPs are frozen and cannot move, which will be discussed in section 4.2.

(15) ***Wh* theme, non-focused DP agent:**

a. Aha [di-tuhor si Poltak ___]?
what PASS-buy PN Poltak

b. Si Poltak [man-uhor aha ___]? (*maN-tuhor* > *manuhor*)
PN Poltak ACT-buy what

c. *Aha si Poltak [maN/di-tuhor ___ ___]? (=1a)
what PN Poltak ACT/PASS-buy

'What did Poltak buy?'

This situation changes, however, if the two DPs are both formally focused—*wh* or focused with the particle *holan* 'only' or *pe* 'even.' Consider the examples in (16) below. These examples contrast minimally from the ungrammatical (14c) and (15c) above in the non-*wh* DP being focused with *holan* 'only,' and are grammatical.[13]

[12]The opposite order—a non-focused DP followed by a *wh* DP—is grammatical as a matrix *wh*-question (i). However, there are reasons to believe that (i) involves a hanging topic and should be distinguished from true multiple fronting. First, this topic requires a following prosodic break, indicated by # in (i), unlike other preverbal constituents that are studied here. Second, such topics can be resumed by an optional pronoun, unlike other cases of fronting that I discuss here. In what follows, I will disregard such configurations involving hanging topics.

(i) Si Poltak$_i$ *(#) aha di-allang (ibana$_i$)?
PN Poltak what PASS-eat 3sg.ANIMATE

≈ 'Poltak, what did he eat?'

[13]The opposite order, with the *holan*-marked DP above the *wh*-phrase as in (i) below, is judged as degraded at best:

(16) **Multiple fronting of a *wh* DP and an 'only' DP:**

a. Ise [holan indahan] [{*mang/✓di}-allang ___ ___]? cf (14c)
who only rice {*ACT/✓PASS}-eat

'Who ate only rice?'

b. Aha [holan si Poltak] [{✓mang/*di}-allang ___ ___]? (=4), cf (15c)
what only PN Poltak {✓ACT/*PASS}-eat

'What did only Poltak eat?'

(17–19) below show similar examples of multiple fronting with two focused DPs.[14] The interpretations of (17) and (18) are unambiguous, reflecting the fact that preverbal quantificational material must be interpreted with surface scope.

(i) [Holan indahan] ise [{??mang/*di}-allang ___ ___]?
only rice who {??ACT/*PASS}-eat

Intended: 'Who ate only rice?' (=16b)

The degraded status of (i) can be thought of as the result of a so-called semantic "intervention effect" on the interrogative *wh*-word. (Not to be confused with syntactic "intervention" affecting probing, discussed in section 4.5.) Roughly, *wh*-words must be interpreted with widest scope at LF, making the intended interrogative interpretation impossible when certain quantificational material (here, the 'only') intervenes. Note that quantifiers in preverbal positions must be interpreted with surface scope, as we will see in (17–19) below, causing the intervention effect in (i). See Beck 1996, 2006; Grohmann 2006; Mayr 2014; Kotek to appear for empirical and theoretical discussion of such intervention effects.

Note also that a parallel degradation is observed with a preverbal *holan* 'only' and a postverbal *wh*-phrase, in (ii). This too reflects the same intervention effect: quantifiers in preverbal position generally outscope those in postverbal position, so the *holan* 'only' scopes over the *wh* in (ii), blocking the intended interrogative interpretation. (I leave a full discussion of quantifier scope and the interpretation of *wh*-in-situ in Toba Batak for future research.) This effect in (ii) supports the view that the degradation in (i) is due to a general interpretational constraint on *wh*-phrases, rather than a syntactic restriction on the relative order of fronted focused DPs.

(ii) ??[Holan indahan] [di-allang ise ___]?
only rice PASS-eat who

Intended: 'Who ate only rice?'

[14] I have thus far not been able to elicit any multiple *wh*-questions.

(17) **Multiple fronting of two 'only' DPs:**

a. Context: All humans drink milk. Infants don't drink anything else, but all other humans drink other things too (e.g. at least water).

[Holan posoposo] [holan susu] [{*mang/✓di}-inum ___ ___].
only infant only milk {*ACT/✓PASS}-drink

'Only infants only drink milk.' (only infants > only milk; true)

b. Context: There's a party and most dishes were tried by many people. But there was exactly one dish that only Poltak ate: this pork.

[Holan babi on] [holan si Poltak] [{✓mang/*di}-allang ___ ___].
only pork PROX only PN Poltak {✓ACT/*PASS}-eat

'Only this pork was eaten by only Poltak.' (only this pork > only Poltak; true)

(18) **Multiple fronting of an 'only' DP and an 'even' DP:**

Context: We all eat lots of different foods, even many unusual things. But only one person goes so far as to eat worms: Poltak.

[Holan si Poltak] [gea pe] [{*mang/✓di}-allang ___ ___].
only PN Poltak worm even {*ACT/✓PASS}-eat

'Only Poltak eats even worms.' (only Poltak > even worms; true)

(19) **Multiple fronting of a 'even' DP and an 'only' DP:**

Context: Something is very strange at the party. No one's eating anything except for Poltak. In general, rice is the most likely thing for people to eat.

[Indahan pe] [holan si Poltak] [{✓mang/*di}-allang ___ ___].
rice even only PN Poltak {✓ACT/*PASS}-eat

'Even the rice was eaten by only Poltak.' (even rice > only Poltak; true)

Notice that in these examples in (16–19), where two core argument DPs are simultaneously extracted, only one choice of voice morphology is grammatical. The generalization is that the subject DP must be in *immediately preverbal* position, with the non-subject moved to a position in front of the subject.

To my knowledge, examples of the form in (16–19) have never before been described in Toba Batak or in any other Austronesian language. The availability of these multiple DP extractions has a number of implications for our understanding of Toba Batak syntax and Austronesian syntax more generally. First, contrary to all previous descriptions, the non-subject DP *can* be extracted, though only in this particular, limited configuration. Second, the contrast between (16) and the minimally contrasting examples in (14–15) show us that the grammar must distinguish *wh*/focus-fronting from any general purpose process of fronting, such as that described in (6) above.

The situation is different still with a DP and a non-DP. Examples (20) and (21) below show that the simultaneous fronting of a formally focused non-DP—a PP pied-piping 'who' in (20) and a *holan*-focused temporal adjunct in (21)—followed by the subject DP, both preceding the verb, is grammatical.[15] Example (20) comes from Emmorey (1984), who refers to such configurations in passing as "double fronting" (p. 45), and (21) is a naturally-occurring example from a text, reported by Cumming (1984).

(20) **Simultaneous fronting of non-DP *wh* and subject (Emmorey, 1984, 44):**

[$_{PP}$ Tu ise] mangga [di-lean hamu ___ ___]?
to who mango PASS-send 2pl

'To whom did you send the mango?'

(21) **Simultaneous fronting of non-DP focus and subject (Cumming, 1984, 27):**[16]

... [holan sa-hali sa-taon do] halak [man-uan eme ___ ___ di tano Batak].
only one-time one-year FOC people ACT-plant rice in land Batak

'...people plant rice only once a year in the Batak land.' *maN-suan* > *manuan*

[15]The opposite order can also be grammatical but involves a hanging topic; see e.g. (i). See footnote 12 above.

(i) Si Poltak, [sian dia] man-angko buku (ibana)?
PN Poltak from where ACT-steal book 3sg.ANIMATE

'Poltak, where did he steal the book from?'

[16]Cumming gives *satali* for 'once,' which should be *sahali*. The focus enclitic *do* is discussed in section 5.

Additional examples of this form are presented in (22). The data here shows that the fronted DP must be the subject as reflected in the choice of voice marking. Note also that the fronted subject in this configuration could itself be focused (here with *holan* 'only') or not.

(22) **Simultaneous fronting of non-DP *wh* and subject (aka "bodyguard") is grammatical:**

a. Andigan [(holan) indahan] [{*mang/✓di}-allang si Poltak ___ ___]?
when only rice {*ACT/✓PASS}-eat PN Poltak

'When did Poltak (only) eat rice?'

b. Andigan [(holan) si Poltak] [{✓maN/*di}-tuhor buku ___ ___]? (=1b)
when only PN Poltak {✓ACT/*PASS}-buy book

'When did (only) Poltak buy the book?' (*maN-tuhor* > *manuhor*)

The possibility of simultaneously extracting one non-DP and one DP in (20–22) is perhaps unsurprising, given that the fronting of non-DPs does not interact with voice, as reviewed in section 2.1. However, it is not simply the case that any simultaneous extraction of a DP and a non-DP is grammatical. Example (23) below shows that a *wh* DP and a non-focused non-DP cannot be simultaneously fronted:

(23) **Simultaneous fronting of *wh* DP and non-focused non-DP is ungrammatical:**

*Ise [PP sian toko buku] [man-angko buku ___ ___]?
who from store book ACT-steal book

The pattern in (20–22) of a non-DP extraction accompanied by subject movement to immediately preverbal position is also attested in Malagasy, another western Austronesian language, where it has been dubbed the "bodyguard" construction (Keenan, 1976). The idea is that the subject "guards" the clause behind it from the fronted adjunct. The construction has attracted some attention in the Malagasy literature, as Malagasy—like Toba Batak—is otherwise generally described as only allowing one constituent at a time to be fronted to a preverbal position. See e.g. Paul 2003; Sabel 2003; Potsdam 2006. I will return to this connection in section 5.

The evidence presented here shows that multiple fronting is possible in Toba Batak but only in particular configurations, summarized in (24) below. The data here shows an interaction between

being nominal or not ([±D]) and the presence or absence of formal focus (*wh* or focus with 'only' or 'even' [±FOC]).

(24) **Summary of multiple extraction patterns:**

a. *[+FOC, +D] [−FOC, +D] V... (14–15)

b. ✓[+FOC, +D] [+FOC, +D] V... (16–18)

c. ✓[+FOC, −D] [±FOC, +D] V... (20–22)

d. *[+FOC, +D] [−FOC, −D] V... (23)

The possible multiple extractions logically fall into two groups: the multiple [+FOC, +D] fronting in (24b) and what we might call the "bodyguard" pattern in (24c). The former pattern additionally teaches us that, when multiple DPs are fronted, the subject must be the immediately preverbal constituent. In the next section, I present my proposal for Case, voice, and the clause periphery of Toba Batak, which derives this distribution in (24).

4 Proposal

As with many other Austronesian languages, most previous work on Toba Batak has only ever described the preverbal fronting of a single constituent at a time. If this one fronted constituent is nominal, it must be the subject argument whose choice is cross-referenced by voice morphology on the verb; this is the famed "subject-only" Austronesian extraction requirement (Keenan and Comrie, 1977, a.o.). The data presented in the previous section shows that the empirical landscape of Toba Batak fronting is more complex than previously described, including certain grammatical configurations of multiple fronting.

In this section, I present my analysis for Toba Batak clause structure which derives the pattern of possible preverbal extractions in the language. I propose that Toba Batak clause structure includes the heads C and T with their familiar division of labor—T probing for a nominal, [+D], and C probing for *wh*/focus, [+FOC]—which can result in multiple extractions. The key additional consideration, I claim, is the limited means of nominal licensing (abstract Case) in the language. Accounting for Case licensing helps explain why the Toba Batak clause periphery seems so superficially limited and different from other, more familiar patterns of attraction by C and T. I begin by briefly presenting my working assumptions for voice in section 4.1 and discuss the role of nominal licensing

in Toba Batak in section 4.2. I then describe C and T and their probing in section 4.3. This derives the "bodyguard" pattern of multiple extraction in (24c).

Finally, after having argued that distinct C and T exist in Toba Batak, in section 4.4 I will argue that the functions associated with C and T also have the option of being bundled into a single head, which I refer to as CT. This bundled CT head can Case-license and attract multiple targets that are simultaneously formally focused and nominal ([+FOC, +D]), resulting in the grammatical pattern of multiple DP extraction in (24b).

As the characterization above indicates, "probe-goal" relationships (Chomsky, 2000, 2001) will play a large part in my analysis. Probes seek a target/goal in their c-command domain with which to *Agree*: establish a link and exchange information. A probe that seeks a target with the feature [F] is itself a formal feature, written as [PROBE:F].[17] I use probing here for Case-licensing nominals and for triggering movement. Probes must Agree with the structurally closest matching target[18]; if the probe can probe multiply, subsequent probing can look past earlier, closer goals.

I furthermore propose that the *use* or *invocation* of a probe is optional. For example, suppose the head X hosts a probe [PROBE:F]. Once X is introduced in the structure, [PROBE:F] can probe *or not*. (See also footnote 17.) This optionality of probing itself will be important for deriving the full set of Toba Batak facts. The details of probing will be discussed further in sections 4.3 and 4.4.

4.1 Clause structure and voice

In this section I present my working assumptions for the basic clause structure of Toba Batak, including the role of voice morphology. My key desideratum of the active vs passive distinction is for the subject (the active agent and the passive theme) to be the highest DP in the lower part of the clause, which I call VoiceP. This accords with much previous work on Austronesian voice systems where the choice of voice reflects or entails a particular argument being structurally highest (Guilfoyle, Hung, and Travis, 1992; Aldridge, 2004; Rackowski and Richards, 2005, a.o.). The precise details of voice morphology and voice alternations are orthogonal to the core questions of

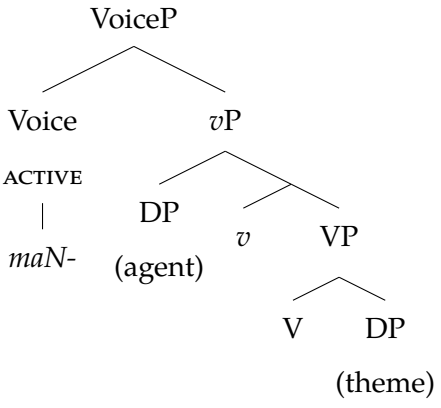
[17] This notation diverges from the more common [uF] notation for probes for [F] goals, where *u* may stand for "uninterpretable," indicating that [uF] must successfully Agree with a goal for the derivation to converge. (For some authors, *u* stands for "unvalued"; see Pesetsky and Torrego 2007 for discussion of the relationship between "uninterpretable" and "unvalued.") I eschew the [uF] notation because I do not want to allude to such a requirement for probes to successfully Agree. As noted below, I argue that the invocation of these probes in Toba Batak are optional.

[18] For similar ideas, see e.g. Relativized Minimality (Rizzi, 1990, 2001), the Minimal Link Condition (Chomsky, 1995, 2000), Shortest Move (Chomsky, 1993), and Attract Closest (Pesetsky, 2000).

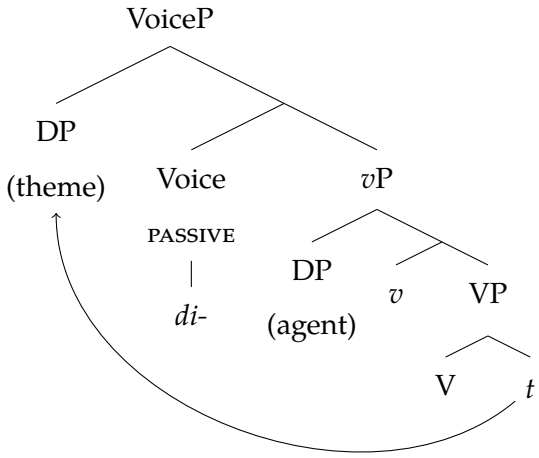
this paper, so long as this desideratum is met.[19] For concreteness, however, I will now spell out one particular approach.

I assume that the theme argument is generated as the complement of the verb (V) and the functional head *v* introduces the agent as its specifier. I propose a head which I call Voice as the locus of the active and passive morphemes, which takes *v*P as its complement.[20] The passive *di-* variant of Voice triggers fronting of the theme DP, whereas the active *maN-* variant triggers no such fronting. The structure of active and passive clauses are sketched in (25) below. Here I concentrate on transitive clauses, but in the general case I assume all arguments are base-generated within VoiceP.

(25) a. Active:



b. Passive:



[19] I will point out, however, that there is one family of proposals for Austronesian voice which cannot be adopted for Toba Batak: so-called *wh*-agreement or case-agreement approaches, which have been proposed for Chamorro (Chung, 1982, 1994, 1998), Palauan (Georgopoulos, 1991), Tagalag (Rackowski, 2002; Rackowski and Richards, 2005), and Malagasy (Pearson, 2001, 2005). These approaches take different voice morphemes to be the realization of agreement with the "subject" DP in its case value (e.g. nominative vs accusative). For the proposal that I put forth here, it is crucial that DP core arguments are not yet structurally Case-licensed within the VoiceP; see section 4.2. The Voice head therefore cannot meaningfully Agree with the subject in Case features: both core argument DPs in VoiceP would be identical, lacking Case values. In fact, although my analysis for Toba Batak involves the assignment of abstract Case as a means of nominal licensing, I do not propose any differentiated case values (e.g. nominative vs accusative), even abstractly. This accords with the fact that Toba Batak has no morphological case alternations. Agreement in grammatical function (thematic role), as proposed in Sternefeld 1995, would however still be an available analytic option similar in spirit to case agreement.

[20] This basic organization of Voice and *v* is the same as that proposed for Toba Batak in Cole and Hermon 2008, but my proposal here differs slightly in the function of active Voice. In particular, Cole and Hermon (2008, footnote 3) propose that the active Voice head moves the agent to its specifier. However, for my proposal here, it is only necessary for the agent to be the highest DP in VoiceP in active clauses. Assuming that the external argument agent is uniformly base-generated as the highest DP argument (i.e. in Spec,*v*P), the active Voice head may be inert.

The derivation of verb-initial (or more accurately, predicate-initial) word orders often involves head movement of the verb or (remnant) VP-fronting. See Clemens and Polinsky (to appear) for a recent overview. For concreteness, I take the former approach here: V head-moves cyclicly up through *v* to Voice, resulting in the pronunciation of the voice prefix and the verb root in a single word. The alternative, VP-fronting approach is often invoked in predicate-initial languages with extraction asymmetries, including by Cole and Hermon (2008) in their analysis of Toba Batak. I discuss the Cole and Hermon 2008 analysis after section 4.2.

With this proposal for the hierarchical structure of VoiceP in place, I turn to a brief discussion of word order. Recall that the word order of postverbal constituents in Toba Batak is entirely free, with the exception of a strict requirement that postverbal active themes and passive agents be immediately adjacent to the verb (Schachter, 1984b; Cole and Hermon, 2008). (See also footnote 7.) I therefore propose that all constituents in VoiceP can be scrambled postverbally; i.e., all and only linearizations of VoiceP with the Voice+*v*+V head leftmost are generated. Word orders where the active theme or passive agent are not verb-adjacent will be independently ruled out, due to considerations of nominal licensing presented in the following section, 4.2. Scrambling of postverbal constituents has been independently proposed for a number of Austronesian languages, including Malagasy (Paul, 2000; Pearson, 2000), Tagalog (Kroeger, 1991/1993; Richards, 1993; Wegmüller, 1998; Rackowski, 2002), Tongan (Otsuka, 2002, 2005), and Niuean (Clemens, 2014). This scrambling could be A'-scrambling as in many of these previous works, or a post-syntactic rearrangement at PF (Clemens, 2014).[21]

The clausal spine above VoiceP can optionally include negation and aspectual/modal auxiliaries, which are free morphemes.[22] I assume that AspP is always projected in the clausal spine, whether Asp is pronounced or not. If any constituent is fronted to a preverbal position, it necessarily pre-

Legate 2014 also proposes the sequence of heads Voice and *v* in the clausal spine of Acehnese, which is spoken just north of the Bataks at the northern tip of Sumatra. However, the division of labor of Legate's Voice and *v* differs from that of my proposal and that in Cole and Hermon 2008. For Legate, Voice is the external-argument-introducing head, in addition to being the locus of voice morphology, and *v* is associated solely with causative semantics. In contrast, Cole and Hermon and I take *v* to be the external-argument-introducing head, with Voice dedicated to the morphosyntax of voice alternations.

[21]This scrambling may involve additional syntactic structure above VoiceP, together with further verb-movement to ensure that the verbal complex is always leftmost. For presentational purposes, however, I will simply refer to this projection with variable word order as VoiceP.

[22]Aspectual/modal auxiliaries in (26a,b) are glossed AUX here as I cannot yet describe their precise semantics. On *nunga* (26c), see Mordechay 1984. See also Percival 1981 (pp. 86ff) on the inventory of auxiliaries.

cedes these negative markers and auxiliaries, as seen in (26).[23] I will brief discuss aspectual auxiliaries again in section 5.

(26) **Fronted constituents precede negation and auxiliaries:**

a. Si Poltak dang olo mang-allang babi.
PN Poltak NEG AUX ACT-eat pork

'Poltak {will not / does not want to} eat pork.'

b. Ibana naeng mang-allang babi.
3sg.ANIMATE AUX ACT-eat pork

'S/he is {going to / about to} eat pork.'

c. Ise nunga mang-allang indahan?
who PERF ACT-eat rice

'Who ate the rice?'

Above AspP (and optionally negation) is TP. The organization of T and C will be discussed in detail in the following sections 4.3 and 4.4.

[23] This is true even in examples of multiple extraction: negation and auxiliaries follow all fronted constituents and are immediately preverbal; see e.g. (i–ii). This clearly distinguishes multiple extraction in Toba Batak from the so-called *Object Voice* or *Passive Type 2* in Malay and Indonesian (Chung, 1976; Sneddon, 1996; Cole, Hermon, and Yanti, 2008, and many others) and Acehnese (Legate, 2014), where both the theme and agent are simultaneously preverbal, in that order, but auxiliaries and negation must precede the preverbal agent (iii–iv), suggesting that the agent is at the edge of VoiceP rather than actually in the higher clause periphery. See especially Legate 2014.

(i) Andigan {*nunga} si Poltak {nunga} man-uhor buku i? based on (1b/22b)
when *PERF PN Poltak PERF ACT-buy book MED

'When did Poltak buy that book?'

(ii) Ise {*nunga} holan indahan {nunga} di-allang? based on (16a)
who *PERF only rice PERF PASS-eat

'Who ate only rice?'

(iii) Auxiliary position in Standard Indonesian object voice (Cole, Hermon, and Yanti, 2008, 1506):

Topi ini {sudah} saya {*sudah} beli.
hat this PERF 1sg *PERF buy

'This hat has been bought by me.'

4.2 Nominal licensing

I now turn to the role of nominal licensing in Toba Batak. Nominals cross-linguistically require a form of licensing which in many languages is satisfied by—or correlates with—morphological case marking. Although Toba Batak does not have morphological case alternations, I follow the view that this licensing requirement of nominals holds in the language. This is the idea of *abstract Case* (Vergnaud, 1977/2008; Chomsky, 1980, 1981, a.o.).

My proposal for nominal licensing in Toba Batak comes in three parts. First, I follow Erlewine, Levin, and Van Urk (2015, 2017) in taking a core property of Austronesian-type voice systems to be that subject DPs receive their Case/licensing from a higher, clause-peripheral probe which can also be involved in their movement.[24] This is the function of T. The head T bears a [PROBE:D] probe which Case-licenses its target and can optionally attract it to its specifier.[25] Based on our derivations for Toba Batak clauses from the previous section, probing by [PROBE:D] on T will necessarily target the subject (active agent or passive theme), as it is structurally highest in VoiceP and therefore necessarily the closest matching [+D] target. This is illustrated in (27) below.

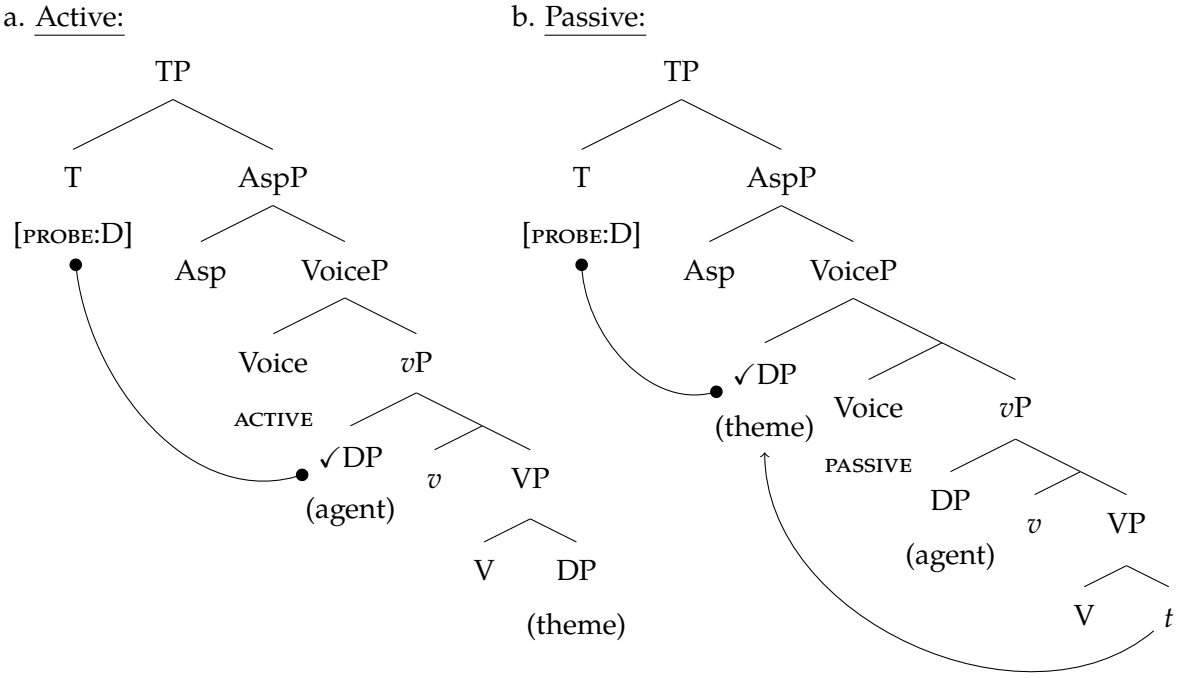
(iv) Auxiliary position in Acehnese object voice (Legate, 2014, 51):

Sie {akan} Fatimah {*akan} tagun keu lôn bak dapu.
meat will Fatimah *will cook to 1sg at kitchen

'Meat will be cooked by Fatimah for me in the kitchen.'

[24] I say "Austronesian-type" because, as Erlewine, Levin, and Van Urk (2015, 2017) argue, the core properties of Austronesian voice systems—one argument (the subject) with a particular case form; voice morphology cross-referencing the choice of subject; effects of voice on the case marking (or linear position, as in Toba Batak) of non-subject arguments; extraction limited to the subject—can also be observed in other, non-Austronesian languages. Erlewine et al. present evidence to this effect from Dinka, a Nilotic language of South Sudan.

(27) **Probing by [PROBE:D] licenses the subject DP:**



I use ✓ to indicate that the subject DPs in (27) have been licensed. Optional fronting of the subject by [PROBE:D] on T results in sentences with fronted subjects, as in (6) above. [PROBE:D] on T is not able to probe multiply to license an additional DP target; this will become important below.[26]

Second, I propose that there is no other structural Case-licensor in the Toba Batak clausal spine. In other words, if we relate the proposal so far for Toba Batak to the syntax of English, this is like saying that the mechanism of nominative case assignment by T is the only licensor for DP arguments; in particular, there is no source of licensing akin to accusative case assignment which could license an additional DP core argument. As a result, there is no source of structural Case-

[25]I think of the nominal probe [PROBE:D] and its Case-licensing ability as the essence of the head labeled T, rather than any tense semantics. This is related in spirit to Pesetsky and Torrego's (2001) idea that what we call nominative case is a [T] feature on nominals. As we have seen in (26), Toba Batak has preverbal aspectual auxiliaries but I will argue in section 5.1 that they are not realizations of T.

[26]Whether or not a probe is allowed to engage multiple goals must be a point of variation. For example, Hiraiwa 2001 argues that Japanese T is able to simultaneously assign nominative case to multiple DPs, but to my knowledge there is no other convincing example in the literature where T licenses multiple DPs with the same case feature value. Instead, T's Case-licensing probe being unable to license multiple DPs—as in Toba Batak—seems to be the more common option. I thank Theodore Levin and Omer Preminger (p.c.) for discussion of this point.

As for information-structural probing by C, here too there must be variation. In the standard description of V2 languages, C may only attract one goal, but even in V2 languages there are some cases of exceptional multiple fronting, yielding V3 order, subject to cross-linguistic variation. See e.g. Hsu 2017 for recent discussion.

licensing for the active theme in (27a) or the passive agent in (27b). Without an additional means of nominal licensing, these structures in (27) will crash.

This brings us to the third and final component of my proposal for nominal licensing in Toba Batak: *nominals can be licensed under linear adjacency with the verb at PF*.[27] The idea that nominals can be made immune to the Case Filter by forming a tight connection with a verb goes back to Baker's (1988, et seq) work on incorporation. Consider the pair of synonymous Southern Tiwa sentences in (28). The subject 'men,' in bold, is a free-standing word in (28a) but incorporated into the verb in (28b).

(28) **Noun incorporation in Southern Tiwa (Allen, 1986, 390):**

a. **Seuanin** i-wan-ban hliawrade-'ay.
man.PL AGR$_{3pl}$-come-PAST lady-to

b. Am-**seuan**-wan-ban hliawrade.
AGR$_{3sg}$-man-come-PAST lady

'The men came to the lady.'

Baker (1988, 106ff) discusses such examples. Because the verb in (28) is the unaccusative 'come,' Baker argues that there is only one source of structural Case in these sentences: T (his Infl). In (28a), the subject 'man' is Case-licensed by T and 'lady' is Case-licensed by the goal postposition 'to' *-'ay*. In (28b), 'men' has incorporated into the verb and the goal 'lady' is now Case-licensed by T, shedding its postposition. Notice too that the agreement on the verb changes from plural in (28a) to singular in (28b), now cross-referencing the singular 'lady.' What is important here is that there is no additional Case-assigner for 'man' in (28b) but the example is grammatical. Adjacency with the verb in the form of incorporation—Baker argues—obviates the need for the nominal to receive structural Case licensing.

Some languages allow a form of incorporation of entire noun phrases, not just single words, which has been called *Pseudo Noun Incorporation* or PNI. Consider the pair of examples in (29) below. Tongan nominal arguments exhibit an ergative/absolutive pattern of case marking. The

[27] Clark 1992 suggests that Toba Batak active themes and passive agents are assigned Case by the verb, with an adjacency requirement on structural case assignment (Stowell, 1981). As suggested by the following incorporation and Pseudo Noun Incorporation examples, however, the idea here is that these nominals can be licensed by adjacency as an *alternative* to structural Case licensing.

bolded theme 'cassava' is absolutive-marked in (29a), with the agent *Sione* being ergative-marked. The postverbal word order in (29a) is free. In contrast, the indefinite 'good cassava' in (29b) takes no determiner and no material can intervene between it and the verb. This is PNI. The resulting clause in (29b) has only one Case-marked argument, the agent 'Sione,' which is now absolutive, reflecting the fact that the NP 'good cassava' is exempt from structural Case licensing.

(29) **Pseudo Noun Incorporation in Tongan (Ball, 2005, 12):**

a. Na'e tō 'e Sione **'a e manioke**.
PAST plant ERG Sione ABS the cassava

'Sione planted the cassava.'

b. Na'e tō **manioke kano lelei** 'a Sione.
PAST plant cassava good ABS Sione

'Sione planted good cassava.'

Based on the study of the related Niuean language, Massam 2001 claims that PNI objects as in (29b) are NPs that must stay in their base positions as the complement of V, thereby trivially explaining the inability of any material to intervene between the verb and PNI object. But Baker 2014 argues that PNI in general exhibits a stricter, *head-head adjacency requirement*. Consider for example the minimal pair of Tongan PNI examples in (30):

(30) **Tongan PNI head noun must be verb-adjacent (Ball, 2005, 12–13):**

a. *Na'e tō **ki'i manioke** 'a Sione.
PAST plant small cassava ABS Sione

b. Na'e tō **manioke iiki** 'a Sione.
PAST plant cassava small ABS Sione

'Sione planted a small amount of cassava.'

While most modifiers are postnominal in Tongan, there are two 'small's: a prenominal *ki'i* and postnominal *iiki*. The contrast in (30) shows that the head noun *manioke* must be linearly adjacent to the verb, not just the entire NP argument. Contrasts such as (29) and similar effects in Chamorro,

Catalan, and Spanish led Baker 2014 to argue that PNI is not simply a NP argument that is the complement of V, contra Massam 2001. Adjacency of the verb with the head of the object (N) is necessary to license PNI objects under adjacency with the verb.

The western Austronesian languages of Malagasy and Balinese help us further refine the notion of adjacency relevant for licensing by adjacency. Erlewine, Levin, and Van Urk 2017 propose that non-subject agents in these languages lack a source for structural Case, and therefore must be licensed by adjacency with the verb. In addition to NP size agents, for which Balinese shows restrictions on modifier position parallel to (30) (Levin, 2015), both languages also allow for full definite DPs to be licensed by adjacency. By comparing patterns of definite DPs in this position in Malagasy and Balinese, Levin 2015 proposes that the head of the highest extended projection of the nominal (e.g. N or D) must be linearly adjacent to the verb for the nominal to be licensed by adjacency.[28]

Both Baker 2014 and Levin 2015 argue that a syntactic operation actively takes place in such examples to shield the nominal from the Case Filter. For Levin, this is Local Dislocation (Embick and Noyer, 2001, a.o.), a postsyntactic operation that adjoins two linearly adjacent heads. Levin proposes that this postsyntactic adjunction allows for the nominal to count as part of the verbal extended projection, obviating its need to be Case-licensed.[29] I refer readers to these works for further discussion of the theory of licensing by adjacency and its empirical motivation.

[28]I assume that Toba Batak arguments are uniformly DPs. The Toba Batak demonstratives markers *on* (proximal), *i* (medial), *an* (distal)—which can also be used independently as inanimate pronouns—are postnominal. However, the personal name marker *si* precedes names. Numerals are also prenominal. There is no clear definite determiner: *i* is glossed as DEF in Cole and Hermon 2008 and at first glance is a good candidate, but see Fox 1984 and Percival 1981 (p. 94) for discussion that suggests that its function differs from that of definiteness. Here I follow Tuller 1984 (p. 184) in treating *i* as a medial demonstrative.

Here I will assume that D in Toba Batak—which *si* may be an instance of—is head-initial and will leave the detailed syntax of Toba Batak DPs a goal for future research. Following Levin 2015, licensing by adjacency is therefore possible between a verb and an immediately *postverbal* DP, but not between a verb and an immediately *preverbal* DP.

Note also that DPs preceded by the focus particle *holan* can be licensed by adjacency with the verb, even though *holan* seems to intervene between the verb and the highest head of the DP. See (45) below for examples. One solution would be to say that *holan* in such cases itself acts as the head of the highest projection in the nominal extended projection. See e.g. Barbiers 2010 for arguments that focus particles can behave as heads of the constituent that they "modify."

[29]Pearson 2001 (p. 46) similarly proposes that a morphological operation combines the linearly adjacent heads, in order to account for phonological effects on non-subject agents in Malagasy, as in (31) below. Note that the operation Pearson 2001 refers to, citing Halle and Marantz 1993, is equivalent to Local Dislocation, which is simply Embick's term for morphological merger applying after linearization, over linearly adjacent heads.

For Baker, the relevant operation is a string-vacuous, covert head-movement, which marks the nominal as part of a complex predicate and thus not subject to the Case Filter. It is unclear whether Baker's approach would predict

Local Dislocation not only ensures the adjacency of the verb and the licensed nominal, but it also ensures that they will form a tight phonological unit at PF, including possible compound-like phonology between the verb and the head of the adjacent nominal. This effect is particularly clear in Malagasy. In example (31a), the non-subject agent is a definite description 'the farmer.' The D head *ny* and the verb together form a morphophonological word which is written together in Malagasy as *vonoin'ny*, as shown at right. In example (31b) below, the non-subject agent is a proper name, *Rabe*, and in this case the verb and name form a single word, written *novidin-dRabe*. As noted by Pearson (2005, 391), this form reflects the fusion of *n* and *r* into the prenasalized retroflexive plosive *ndr*, which is a sound change that is otherwise observed only word-internally.

(31) **Effects of Local Dislocation on Malagasy non-subject agents:**

a. Vono-ina **ny** **mpamboly** amin'-ny antsy ny akoho. *Vono-ina + ny > vonoin'ny*
kill-PASS the farmer with-the knife the chicken

'The farmer is killing the chickens with the knife.' (Pearson, 2001, 23)

b. No-vidy-ina **Rabe** hoan-dRasoa ilay satroka. *No-vidy-ina + Rabe > novidin-dRabe*
PAST-buy-PASS Rabe for-Rasoa that hat

'Rabe bought that hat for Rasoa.' (Keenan, 2000, 27)

Pearson 2001 notes that this apparent compounding also affects the calculation of phonological stress. Similar effects are also observed in Toba Batak, which suggest that the verb and immediately postverbal non-subject DPs form a phonological unit (Emmorey, 1984).

My proposal is that licensing by adjacency through Local Dislocation is the source of licensing for postverbal active themes and passive agents in Toba Batak. Toba Batak only has one source of structural Case licensing for DPs in the clause: the [PROBE:D] probe on T. In transitive clauses,

the compounding-like phonological effects of licensing by adjacency. See especially Levin 2015 (pp. 145–148) for a comparison of these two approaches.

As an anonymous reviewer correctly notes, licensing by adjacency should be thought of as a last-resort strategy for nominals which have not received structural licensing. This is explicit in Levin's proposal which relies on Local Dislocation, which by definition takes place after linearization at PF (Embick and Noyer, 2001), after all structural Case-licensing operations.

Another reviewer asks whether there are any other Case-licensing strategies available to the grammar besides probing by a Case-licensing head and Local Dislocation (adjacency). I am unable to give a definite, universal answer, but to my knowledge in the grammar of Toba Batak, these are the only two strategies available. See also footnote 39 on Case-licensing by probing.

postverbal non-subject DPs necessarily lack a structural Case licensor.[30] As noted in the previous section, these DPs are not necessarily verb-adjacent according to the structures in (25–27), but Toba Batak allows for the free scrambling of postverbal constituents. This allows for the generation of word orders where the non-subject DP is immediately postverbal, allowing for Local Dislocation to apply, licensing the DP.

This licensing by adjacency proposal immediately predicts a strict adjacency requirement between the verb and any non-subject DP. As noted above, such an effect has been observed previously by Schachter (1984a, 125). Consider for example the possible postverbal placement of the temporal adjunct *nantoari* 'yesterday' in (32). We see that the adverb *nantoari* 'yesterday' can be placed freely, with the exception of the position between the verb and the non-subject DP argument. Under my account here, breaking this linear adjacency leads to the non-subject DP failing to be licensed. This adjacency requirement is true of both active themes as in (32a) and passive agents as in (32b), reflecting the deep symmetry between the two voices in Toba Batak.

(32) **Adding *nantoari* 'yesterday' to (5a,b), based on Schachter (1984a, 125):**

a. Man-jaha {*nantoari} buku {nantoari} si Poltak {nantoari}.
ACT-read *yesterday book yesterday PN Poltak yesterday

b. Di-jaha {*nantoari} si Poltak {nantoari} buku {nantoari}.
ACT-read *yesterday PN Poltak yesterday book yesterday

'Poltak read a book yesterday.'

This same adjacency requirement can also be observed with ditransitive verbs. (33) below gives all six postverbal word orders of the subject agent DP, non-subject theme DP, and goal PP of the active verb *manga-lehon* 'give':

[30] The situation differs for fronted non-subjects, as in the case of multiple focused DP extraction, due to the different organization of the probes on C and T. I will discuss this in section 4.4. Objects of prepositions are Case-licensed directly by their preposition, explaining the availability of PP arguments and adjuncts.

(33) **Word order with three arguments:**

a. Manga-lehon buku tu si Uli si Poltak.
ACT-give book to PN Uli PN Poltak

'Poltak gave a book to Uli.'

b. Manga-lehon buku si Poltak tu si Uli.

c. *Manga-lehon tu si Uli {buku si Poltak / si Poltak buku}.

d. #Manga-lehon si Poltak {tu si Uli buku / buku tu si Uli}.
'The book gave Poltak to Uli.'

Only word orders (33a–b) with the non-subject theme DP *buku* 'book' immediately adjacent to the verb are grammatical. Note that the relative order of the subject DP and goal PP are free in (33a–b). The word orders in (33c) with the PP in immediately postverbal position are simply ungrammatical: there are two DPs in the clause but T can only license one, leaving one DP unlicensed. The word orders in (33d) are ungrammatical with the intended interpretation, but does have the nonsensical interpretation 'The book gave Poltak to Uli.' The immediately postverbal DP *si Poltak* is necessarily interpreted as the non-subject and therefore a theme in this active clause.

A further prediction of my proposal is that a clause can only include at most two DP arguments: one DP can be licensed structurally by T and another can be licensed by adjacency with the verb. This prediction is borne out in my speakers' grammars. All grammatical verb frames that I have elicited include at most two DP core arguments, with additional arguments being expressed as obliques such as the prepositional goal of 'give' in (33).[31]

[31] Interestingly, Schachter 1984a (pp. 136ff) describes a productive dative alternation in Toba Batak, giving pairs of examples such as in (i) below. Schachter reports both (ia) and (ib) as grammatical and having the same interpretation, but my speakers consistently reject (ib) and similar three-DP examples of Schachter's, and all different word order combinations thereof. I can only speculate that Schachter's speaker must have had an additional source of Case licensing in examples such as (ib), not available to my speakers.

(i) Dative alternation pair from Schachter (1984a, 137), with my speakers' judgments:

a. ✓Manga-lean biang si Torus tu si Ria.
ACT-give dog PN Torus to PN Ria

'Torus gives a dog to Ria.'

b. *Manga-lean si Ria si Torus biang.
ACT-give PN Ria PN Torus dog

To summarize, I have proposed that nominals in Toba Batak require a form of formal licensing (abstract Case), but Toba Batak clauses have only one structural Case licensor, the [PROBE:D] probe on T. Due to the syntax of voice, presented in section 4.1 above, [PROBE:D] will Case-license the subject, leaving postverbal non-subjects without licensing. In such a situation, if the non-subject is immediately postverbal, Local Dislocation can apply between the verb and non-subject, exempting it from (Levin, 2015; see also Baker, 2014). This offers a new account for the adjacency requirement on non-subjects discussed by Schachter 1984a and Cole and Hermon 2008.

Aside: Cole and Hermon's (2008) VP-fronting analysis

The adjacency requirement on postverbal non-subject DPs is a key point of discussion in Cole and Hermon 2008, the only contemporary syntactic analysis of Toba Batak clause structure. Cole and Hermon present a "VP-fronting" (technically VoiceP-fronting) analysis for Toba Batak: all arguments in the VoiceP except the passive agent and active theme necessarily evacuate the VoiceP, followed by fronting and freezing of the remnant VoiceP ("ACT-V theme" or "PASS-V agent"). Their analysis is explicitly designed to derive two effects. First, it explains the adjacency requirement observed in (32). Second, it predicts that non-subject DPs cannot ever be extracted, because they are frozen through the movement of VoiceP—a welcome consequence, according to Cole and Hermon.[32]

As Cole and Hermon discuss in their paper, the potential appeal of such "VP-fronting" analyses is to simultaneously derive both the observed word order restrictions and extraction asymmetry: passive agents and active themes are precisely the set of DPs that have fixed word order and cannot be extracted. However, I have shown in section 3 that the extraction of non-subject DPs is in fact possible, albeit in the limited configuration of fronting multiple formally focused DPs. Proponents of a VP-fronting analysis for Toba Batak then must abandon the biconditional between fixed word order and unextractability, and also provide an alternative explanation for the limited extractability of non-subject DPs outside of the multiple focused DP fronting configurations.

As noted by an anonymous reviewer, even if the freezing explanation for the extraction asymmetry is abandoned, Cole and Hermon's VoiceP-fronting approach could still derive the postverbal

[32] Clark 1992 and Baldridge 2002 also present accounts for Toba Batak that directly tie the inability of non-subjects to be extracted to their adjacency requirement. For Baldridge 2002, in a flavor of Combinatory Categorial Grammar, both facts are accounted for together by restricting abstraction over non-subject arguments. The account in Clark 1992 is most similar to my proposal here, deriving these two facts from limitations of case assignment (see footnote 27), predicting as I do that non-subject extraction could become possible if an independent source of licensing becomes available.

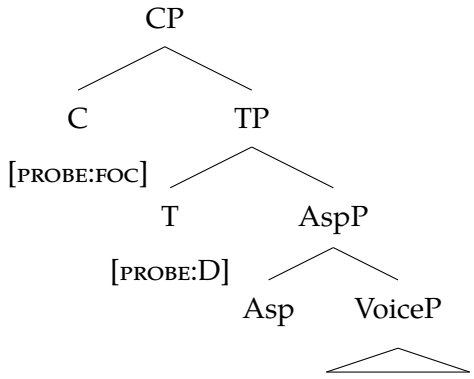
adjacency of non-subject DPs. I note, however, that this result of Cole and Hermon's proposal is little more than a stipulation: all material except passive agents and active themes must evacuate VoiceP before VoiceP-fronting, but there is no independent motivation for these movements. As they note in the last page of their paper (p. 195), they have "no explanation" for "why it should be the case that it is the direct object and passive agent that do not raise out of VoiceP prior to VoiceP raising."

In contrast to Cole and Hermon's approach, under my approach there is no freezing and no absolute ban on the movement of non-subjects. The adjacency of postverbal non-subject DPs is due to their inability to be structurally Case licensed, forcing them to be immediately postverbal for licensing by adjacency. Unlike the VoiceP analysis, there is no need to stipulate passive agents and active themes as the locus of exceptional behavior: the calculus of Case-licensing proposed for the language predicts that it is precisely these arguments that must be immediately postverbal.

4.3 Fronting with C and T

With my analysis of voice and nominal licensing in place, I now present my proposal for preverbal fronting in Toba Batak. The multiple extraction data that I introduced above in section 3 necessitates a clausal architecture that can front multiple constituents to preverbal positions. For this, I propose that Toba Batak has the two functional heads C and T, as in the syntax of many other language families, with their common division of labor: T probes for a single DP and Case-licenses it, optionally fronting it, while C can probe for formally focused constituents and front them. The proposed structure of the clause periphery is sketched in (34). I will treat both C and T as simply unpronounced in this section but will return to the question of their pronunciation in section 5.

(34) **The organization of the Toba Batak clause periphery:**

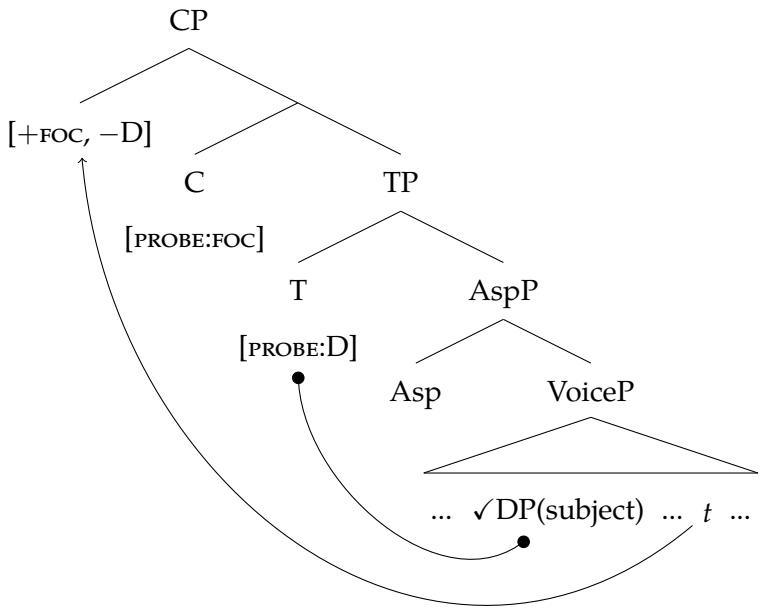


Given this familiar organization of C and T which can both independently attract constituents to their specifiers, why are multiple extractions in Toba Batak so limited? The key, I propose, is the limited means of nominal licensing in the language, as discussed in the previous section. The [PROBE:FOC] probe on C can attract [+FOC] targets, but if the target is nominal, it may then be in danger of having no source of licensing. This leads to the limited inventory of multiple extractions.

Let's consider the effects of these [PROBE:D] and [PROBE:FOC] probes, step by step. As discussed in the previous section, the [PROBE:D] probe on T will probe down and find the subject, which is the highest DP in VoiceP. This Case-licenses the subject. Now consider the effect of the [PROBE:FOC] probe on C. Assuming that there is a matching [+FOC] target in VoiceP, there are two possibilities: either the target is nominal or not.

Consider first the case where [PROBE:FOC] finds a [+FOC, −D] target: a formally focused non-DP. In contrast to [PROBE:D] on T, I take [PROBE:FOC] to obligatorily front any target that it Agrees with to Spec,CP.[33] Because the target is not nominal, we do not have to worry about its licensing. This structure is illustrated in (35) below.

(35) **Fronting [+FOC, −D] to Spec,CP:**



This derivation corresponds to our examples of *wh* or focused non-DP fronting. Because [PROBE:FOC] simply targets the highest [+FOC] constituent in VoiceP, which by assumption in this case is not nominal, we predict no interaction with voice morphology: it doesn't matter which DP is the sub-

[33] Recall, however, that the *invocation* of the probe is optional, though, allowing for optionality of the *wh*-movement in (35/36).

ject (highest in VoiceP). This is exactly what we have observed: the extraction of non-DPs does not interact with voice, as exemplified in (9), repeated here in (36):

(36) **Extraction of non-DPs does not interact with voice:** (=9)

a. [$_{CP}$ [$_{PP}$ Tu ise] [$_{VoiceP}$ man-uhor buku si Poltak ___]]?
for who ACT-buy book PN Poltak

b. [$_{CP}$ [$_{PP}$ Tu ise] [$_{VoiceP}$ di-tuhor si Poltak buku ___]]?
for who PASS-buy PN Poltak book

'[For who] did Poltak buy the book?'

The [PROBE:FOC] probe on C has the ability to probe and attract multiple targets, which is yet another pattern of multiple fronting. This is exemplified in (37) below.[34]

(37) **Multiple fronting of two [+FOC, −D] targets:**

[$_{CP}$ Boasa [holan [$_{PP}$ sian toko buku]] [$_{VoiceP}$ manangko buku ho ___ ___]]?
why only from store book ACT-steal book 2sg

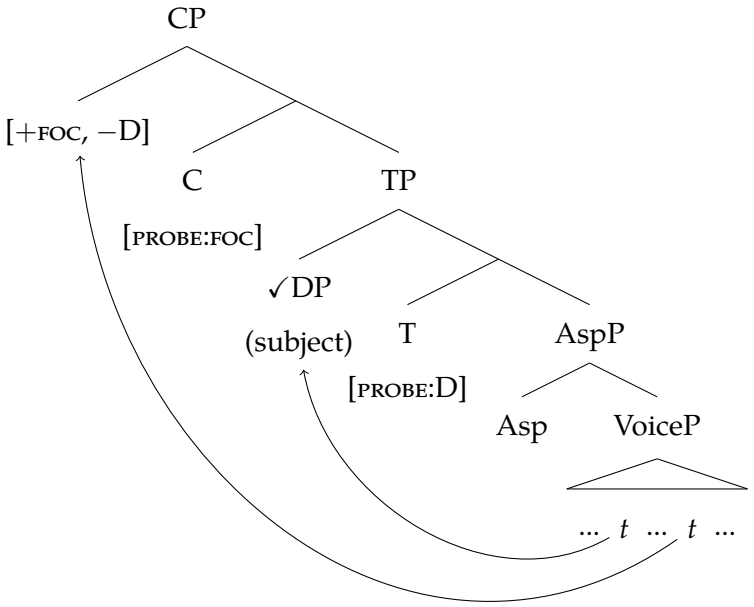
'Why do you only steal books from the BOOK STORE?'
(≈ Why don't you steal books from other places?)

Note that this ability to probe and attract multiply is a property of the [PROBE:FOC] probe, not shared by [PROBE:D]. If [PROBE:D] could probe multiply, it could Case-license multiple DPs in-situ and we would no longer predict the requirement that non-subject DPs must be adjacent to the verb for licensing, discussed in the previous section.

Now recall that [PROBE:D] has the ability to optionally front the subject DP that it has Agreed with and Case-licensed. This subject fronting could occur at the same time as the fronting of a [+FOC, −D] constituent to Spec,CP, resulting in the structure in (38). We have now successfully derived the so-called "bodyguard" configuration of multiple fronting: a formally focused non-DP followed by the subject, both in preverbal position.

[34]There is a question here in (37) regarding the timing of the two movements. I will propose that multiple movements triggered by the same probe do not "tuck in," *pace* Richards 1997; i.e. in (37), *holan sian toko buku* moves first, followed by movement of *boasa* to a higher position. See section 4.4 for the specific proposal and section 5.1 for supporting evidence.

(38) **Fronting [+FOC, −D] to Spec,CP and subject to Spec,TP:**



A grammatical example of this "bodyguard" configuration is repeated in (39) below. As indicated by the voice morphology in (39), the fronted DP in this configuration must be the subject. This is explained by the proposal here because the DP fronting is the result of probing by [PROBE:D] on T, which will necessarily target the highest DP in VoiceP, the subject. Note that the fronted subject in this configuration could itself also be formally focused.[35]

(39) **Simultaneous fronting of non-DP *wh* and the subject (aka "bodyguard"):** (=1b/22b)

[$_{CP}$ Andigan [$_{TP}$ [(holan) si Poltak] [$_{VoiceP}$ {✓maN/*di}-tuhor buku ___ ___]]]?
when only PN Poltak {✓ACT/*PASS}-buy book

'When did (only) Poltak buy the book?' (*maN-tuhor* > *manuhor*)

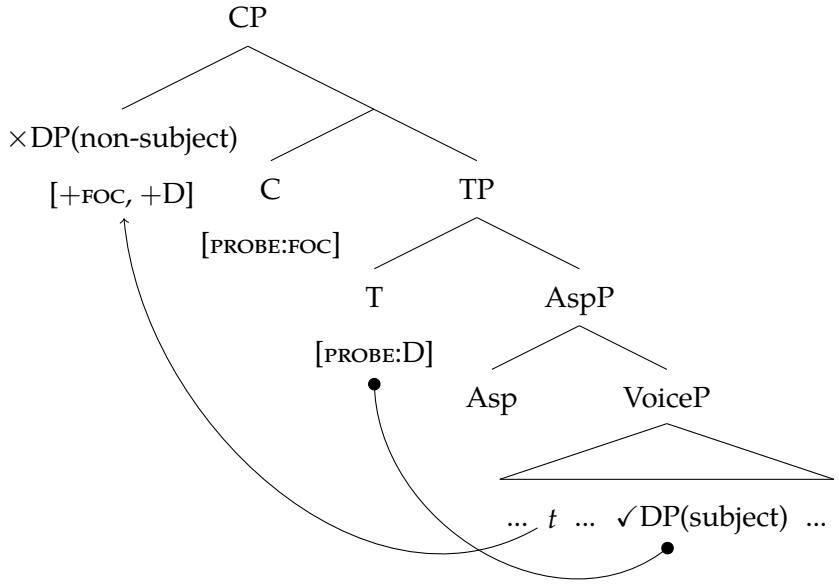
From these derivations presented above, we see that Toba Batak syntax is best modeled with two heads in the clause periphery associated with [PROBE:FOC] and [PROBE:D], corresponding neatly to C and T in proposals for clause structure in many other languages. Movement to Spec,TP is

[35]In the case where the subject is [+FOC], I assume the [PROBE:D] probe on T moves the subject to Spec,TP first. Subsequent probing by [PROBE:FOC] on C must be able to skip the subject in Spec,TP, probing past it to find another [+FOC] target in VoiceP. There are at least two possible responses to this issue. One is that the the movement of the subject to Spec,TP put it in a criterial position, from which it is unable to move further (Rizzi, 2006; Rizzi and Shlonsky, 2007), so it is overlooked by higher probing. Another is that there is an anti-locality constraint banning the attraction of the specifier of the probe's complement (Erlewine, 2016; Bošković, 2015, 2016), again leading to the subject not being visible for higher probing (Deal, 2016).

driven by [PROBE:D] and is therefore limited to the highest DP in VoiceP, i.e. the subject. In contrast, movement to Spec,CP is driven by a [PROBE:FOC] probe, which simply attracts the closest [+FOC] constituent in VoiceP.[36] Finally, the independence of these movements to Spec,TP and Spec,CP predicts the availability of simultaneously fronting a [+FOC] non-DP to Spec,CP and the subject to Spec,TP, resulting in the so-called "bodyguard" configuration (39). At the same time, the strict ordering of C over T means that a *wh*/focus cannot be fronted together with a preverbal *non*-DP, as T does not attract non-DPs, explaining the ungrammaticality of example (23) above.

These examples so far all involve [PROBE:FOC] attracting a non-DP target. But what happens if probing by [PROBE:FOC] finds a DP? Here the limitations of nominal licensing in Toba Batak rear their head. Suppose that the highest [+FOC] constituent in VoiceP is a non-subject DP. (I discuss further cases with [+FOC] subjects in the following section.) Let the [PROBE:D] probe on T probe down and Case-license the subject. Next, let the [PROBE:FOC] probe on C probe down and find the [+FOC] non-subject DP and front it. This results in the structure in (40).

(40) **A non-subject [+FOC, +D] in Spec,CP has no means of licensing:**



[36] I presume no Phase Impenetrability effects arising from a VoiceP or *v*P phase. Practically speaking, there is no evidence for a lower phase edge in Toba Batak and my analysis here acts as a demonstration that all extraction facts can be explained without reference to a lower phase edge.

Recall that another crucial aspect of my proposal is that there is no structural Case-assigner within VoiceP in Toba Batak. Based on the interaction of case-marking and quantificational scope in Japanese, Takahashi (2010) proposes that phasehood is related to Case-assignment. Specifically, he predicts that "*v*P will constitute a phase only when *v* values Case of an internal argument" (p. 335). Toba Batak may be an extreme example compatible with Takahashi's conjecture, in completely lacking a lower structural Case assigner as well as a lower phase boundary. I thank Lyn Tieu (p.c.) for pointing me to Takahashi's work.

The problem with the configuration in (40) has to do with the licensing of the non-subject DP. Recall that there is no structural Case licensor in the clause except for the [PROBE:D] probe on T, but [PROBE:D] here is licensing the subject and [PROBE:D] cannot probe multiply. If the non-subject stayed in postverbal position, it could be licensed by adjacency with the verb, thereby obviating its requirement for Case-licensing. But the fronting of the DP to Spec,CP in (40) bleeds the possibility of licensing by adjacency.[37] The end result is that there is no way to Case-license the non-subject in Spec,CP and the derivation does not converge.

The unavailability of the structure in (40) is part of the derivation of the famed "subject-only" extraction restriction (Keenan and Comrie, 1977, a.o.): the requirement that, if a single DP is extracted in Toba Batak, it must be the subject. T cannot attract a non-subject because it would have to skip the subject and (40) shows that C could attract a [+FOC] non-subject DP but the DP would then lack licensing. Note that the subject could also be fronted to Spec,TP in (40), but this does not change the fact that the non-subject DP in Spec,CP is unlicensed. This explains the ungrammaticality of simultaneously fronting a [+FOC] DP and a [−FOC] DP, as exemplified in (41):

(41) **Multiple extraction of [+FOC, +D] [−FOC, +D] is ungrammatical:** (=1a/15c)

*[$_{CP}$ Aha [$_{TP}$ si Poltak [$_{VoiceP}$ man-uhor ___ ___]]]?
what PN Poltak ACT-buy

Intended: 'What did Poltak buy?'

4.4 Bundled CT

In order to account for patterns of multiple extraction in Toba Batak, I've argued that Toba Batak clause structure involves the two heads C and T, which can independently probe and front formally focused and nominal constituents, respectively. In this section, I propose that the features of C and T also have the option of being *bundled* together on a single head, which I call *CT*. Overt morphological evidence for this bundling proposal will be presented in section 5.

A direct precursor of my proposal here is Legate 2011, who proposes that the features of C and T are bundled together in cases of DP fronting in Acehnese—an Austronesian language with a

[37] The non-subject could be immediately *pre*-verbal in (40), but this is not sufficient to license the non-subject, due to the directionality requirement of licensing by adjacency. See footnote 28.

very similar extraction restriction, spoken just north of the Bataks—as well as in Germanic subject V2. See also Legate 2014, pp. 83–84, 152–153. The idea that formal features can be bundled together on a single head or distributed across separate heads has also been proposed to account for observed variation—both within and between languages—in the organization of tense, aspect, and mood (Giorgi and Pianesi, 1996), tense and agreement (Bobaljik, 1995; Thráinsson, 1996; Bobaljik and Thráinsson, 1998, a.o.), complementizer systems (Bianchi, 1999), voice and little v (Pylkkänen, 2002, 2008; Harley, 2017), and V2 requirements (Hsu, 2016a, 2017). In addition to Legate's work, the bundling of C and T features onto a single head has been proposed to account for subject/nonsubject extraction asymmetries in Defaka (Ijoid Niger-Congo) (Bennett, 2009; Bennett, Akinlabi, and Connell, 2012) and Wolof (Atlantic Niger-Congo) (Martinović, 2015). Gallego to appear also recently proposed that languages vary in whether C and T are bundled or not, as a way to account for cross-linguistic differences in subject/nonsubject extraction asymmetries.[38]

When CT is bundled, the [PROBE:FOC] probe of C and the [PROBE:D] probe of T operate as a single probe that seeks targets that are simultaneously [+FOC, +D], i.e. matching the specifications of [PROBE:FOC] and [PROBE:D] at the same time. Such composite probing for combinations of features is also discussed in Coon and Bale 2014 and Van Urk 2015. Probes are additionally associated with specifications such as whether or not they can or must front their targets, whether they are Case-licensing, etc. This bundled [PROBE:FOC+D] probe on CT will inherit the marked ('yes') values for these specifications from [PROBE:D] and [PROBE:FOC]. These specifications are summarized in (42). In the rest of this section, I will demonstrate each of these aspects of the [PROBE:FOC+D] probe and their use.

[38]The relationship between C and T on the one hand and the bundled head CT could be thought of in one of two ways. One option is for the lexicon to begin with the atomic features/heads C = [PROBE:FOC] and T = [PROBE:D] which then bundle presyntactically into a single CT = [PROBE:FOC+D] head. The possibility of such "presyntactic bundling" is discussed sometimes under the banner of "fusion": see Matushansky 2006, footnote 23 for discussion. See also Hsu 2016a,b for further technical discussion of such bundling operations.

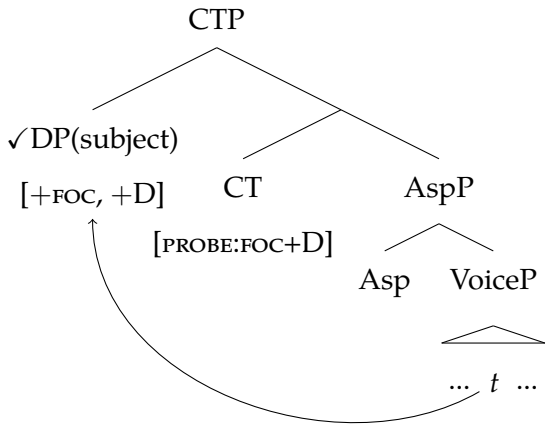
Alternatively, Giorgi and Pianesi 1996 and Martinović 2015 propose that certain features enter the derivation bundled but can then "scatter" or "split," respectively, under certain circumstances. Legate (2011, 2014) also starts with a bundled head, but relates splitting to the idea of feature inheritance, Chomsky's (2008) proposal that the features of T originate on C and are passed down to T, resulting in the familiar C and T heads (see also Ouali, 2008; Fortuny, 2008). Legate raises the possibility of *under-inheritance*, where all features stay on C and none are passed down to T, as a means of preserving the bundled head (CT). However, Legate's under-inheritance approach to bundling is strictly speaking not isomorphic to the structures I adopt here, as it predicts the presence of a functionally inert T head below C when under-inheritance takes place. See also Gallego 2014, to appear for a comparison of feature inheritance to feature bundling/scattering approaches.

(42) **A calculus of probe bundling:**

	C = [PROBE:FOC]	+	T = [PROBE:D]	=	CT = [PROBE:FOC+D]
Case-licenses target?	no		yes		yes[39]
Must front target?	yes		no		yes
Can probe multiply?	yes		no		yes

I begin by discussing the case where the subject DP is formally focused: [+FOC, +D]. Probing by the bundled [PROBE:FOC+D] will find the focused subject and front it. This is illustrated in (43).

(43) **Fronting a focused subject with CT = [PROBE:FOC+D]:**



[PROBE:FOC+D] inherits [PROBE:D]'s Case-assigning ability (42), licensing the subject DP. This structure corresponds to examples with a single fronted *wh* or focused DP.

Now suppose [PROBE:FOC+D] probes again. This is possible because it inherits [PROBE:FOC]'s ability to probe multiply (42). If there is another [+FOC, +D] constituent past the subject—i.e. a formally focused non-subject DP—[PROBE:FOC+D] will Case-license it and front it. I propose that this is the source of the grammatical multiple focused DP extractions, as in the examples in (16), repeated here in (44):

[39]There may be a principled reason why the [PROBE:D] probe on T and [PROBE:FOC+D] on bundled CT Case-license their targets, but [PROBE:FOC] on C does not: I tentatively suggest that Case-licensing probes are those that specifically probe for DPs or a subset thereof. This generalization is compatible with the proposals for probing and Case-licensing in recent work on languages with Austronesian-type voice systems as in Van Urk 2015 and Erlewine et al. 2017. I thank the reviewers for raising this issue.

(44) **Multiple extraction of two [+FOC, +D] constituents:** (=16)

a. [$_{CTP}$ Ise [holan indahan] [$_{VoiceP}$ {*mang/$^{\checkmark}$di}-allang ___ ___]]?
who only rice {*ACT/$^{\checkmark}$PASS}-eat

'Who ate only rice?'

b. [$_{CTP}$ Aha [holan si Poltak] [$_{VoiceP}$ {$^{\checkmark}$mang/*di}-allang ___ ___]]?
what only PN Poltak {$^{\checkmark}$ACT/*PASS}-eat

'What did only Poltak eat?'

As indicated by the voice markers in (44), the two focused DPs must be fronted with the subject in immediately preverbal position. I propose that when a probe attracts multiple constituents, the later target moves to a higher position, resulting in nesting rather than crossing dependencies (see Pesetsky, 1982, also footnote 34); i.e. these probes do not "tuck in," *pace* Richards 1997. Probing by [PROBE:FOC+D] will first find the [+FOC] subject DP and front it, followed by attraction of the [+FOC] non-subject DP to a higher position.

It's worth noting that this second invocation of the [PROBE:FOC+D] probe on C is itself optional. It is thus grammatical for CT to attract a focused subject but leave another formally focused DP in-situ below. The in-situ DPs will be licensed by adjacency with the verb; see footnote 28 above for relevant discussion. The example in (45) attests to this possibility.

(45) **Extraction of one [+FOC, +D], with another [+FOC, +D] left in-situ:**

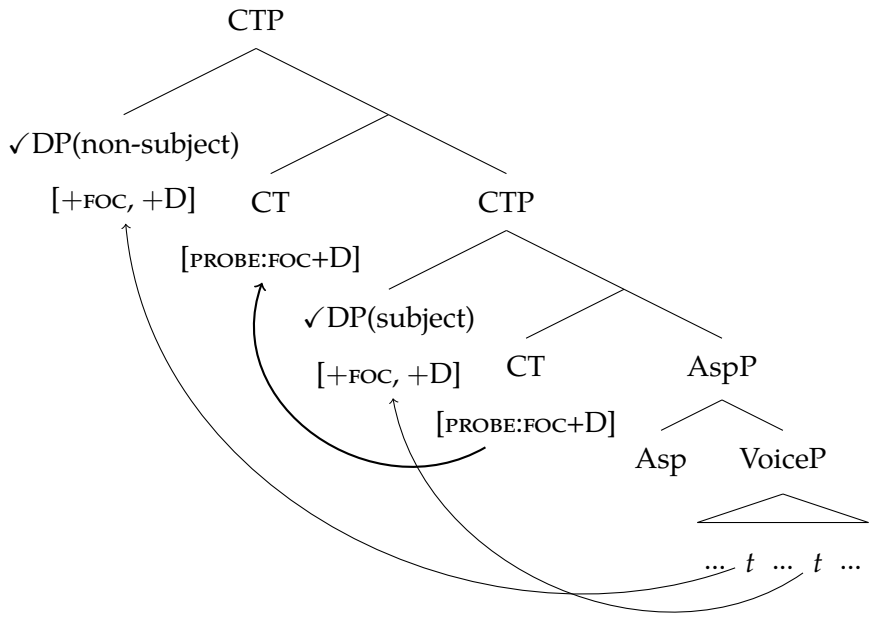
[$_{CTP}$ Ise [$_{VoiceP}$ mang-allang [holan indahan] ___]]?
who ACT-eat only rice

'Who ate only rice?'

When CT attracts multiple targets as in (44) above, I propose that CT hosts these multiple specifiers through *reprojection*: CT head-moves to merge with its own CTP projection, projecting a higher CTP.[40] This is illustrated in (46) below, with the thick arrow for reprojection of CT. I will present morphological evidence for this reprojection of the CT head in section 5.

[40] See Iatridou and Kroch 1992, Watanabe 1992, Browning 1996, and references there on so-called *CP-recursion*. See also more general discussion of head-reprojection in Surányi 2005, Georgi and Müller 2010, and references there.

(46) **CT reprojection for multiple extraction of [+FOC, +D] constituents:**



Note that when [PROBE:FOC+D] finds a matching target, it then must move it—a property inherited from [PROBE:FOC] on C (42). This property is crucial for accounting for the ungrammaticality of examples such as (47). Note that the active *mang-allang* version of (47) without *nantoari* was grammatical, in (45).[41]

(47) **CT cannot Case-license a DP and not front it:**

*Ise {mang/di}-allang nantoari [holan indahan]?
who {ACT/PASS}-eat yesterday only rice

Intended: 'Who ate only rice yesterday?'

Assume for contradiction that [PROBE:FOC+D] attracts its targets *optionally*. Consider first the active *mang-allang* variant of (47). The [PROBE:FOC+D] probe finds the [+FOC] subject *ise* 'who,' Case-licenses it, and fronts it. Now notice that, due to the adverb *nantoari* intervening between the non-

[41] The same contrast between (45) and (47) holds with the thematic roles reversed:

(i) Aha di-allang (*nantoari) [holan si Poltak]?
what PASS-eat yesterday only PN Poltak

'What did only Poltak eat (yesterday)?'

subject DP and the verb, the non-subject cannot be licensed by adjacency with the verb. However, in this hypothetical derivation [PROBE:FOC+D] could probe again to find the [+FOC] non-subject and Case-license it, but *without* moving it. This would incorrectly predict the availability of example (47) with active *mang-allang*. Next, consider the passive *di-allang* variant of (47). Let [PROBE:FOC+D] on CT probe down, find the [+FOC] subject DP *holan buku* 'only book' and Case-license it, but *not* move it. Subsequent probing by [PROBE:FOC+D] could then license and front the formally focused non-subject DP. This derivation incorrectly predicts example (47) with passive *di-allang* to be grammatical. The examples in (47) thus show that the bundled [PROBE:FOC+D] probe must attract the targets that it Agrees with, as in (42). Recall, however, that the initial invocation of these probes is still optional, allowing for optionality in fronting: see (44) vs (45). What is necessary is to front every target that the probe Agrees with.

4.5 Complex probes and partial matches

One additional configuration must be discussed is the case where the subject DP is [−FOC] but there is a lower [+FOC, +D] constituent in the clause. If [PROBE:FOC+D] could probe down and Case-license and attract the [+FOC, +D] non-subject, across the [−FOC] subject, we would end up with a structure as in (48) below. The in-situ subject could be licensed by adjacency with the verb, so this derivation would incorrectly predict (48) to be grammatical.

(48) **Ungrammatical example generated by [PROBE:FOC+D] probing across the subject:**

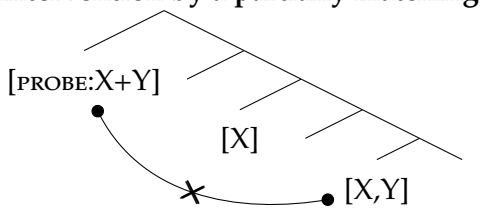
*[$_{CTP}$ [$_{non\text{-}subject}$ Aha] [$_{VoiceP}$ man-uhor ___ [$_{subject}$ si Poltak]]]]? (=8a)
what ACT-buy PN Poltak

Intended: 'What did Poltak buy?'

What went wrong here? The answer has to do with the behavior of complex probes such as [PROBE:FOC+D]. Although it is true that the subject in (48) is not a valid goal for [PROBE:FOC+D]—because it is a [D] but lacks the [FOC] feature—the subject is a *partial* match for the probe. In recent work, Deal 2015 shows convincingly that complex probes may interact with partially-matching goals, even if they do not fully satisfy the probe. We can imagine different resolutions for what happens after a complex probe encounters a partially-matching goal.

When a complex probe encounters a partial match in Toba Batak, I propose that it then stops its search procedure. Because the partial match does not satisfy the probe, the probe does not Agree with any target.[42] The end result is that probing in this configuration will fail. This is schematized in (49).[43]

(49) **Intervention by a partially matching goal:**



Concretely, probing by the [PROBE:FOC+D] probe on CT cannot probe past the subject in (48). Probing by CT in (48) will fail, even though there is a matching [+FOC,+D] goal lower in the structure. As there is no other route to grammatically derive the surface representation in (48), the example is judged as ungrammatical. In this way, the proposed inability for a complex probe to probe past a partial match (49) derives the basic "subject-only" extraction asymmetry observed in the language.[44]

[42] An alternative conception is that when the CT [PROBE:FOC+D] probe encounters a [−FOC,+D] subject, it *will* trigger Agree, but with the behavior associated with a standalone [PROBE:D] probe. In this case, no Case-licensing probe will be able to probe again. See section 4.3 for the behavior of [PROBE:D] (unbundled T).

[43] This may be a point of cross-linguistic variation. For additional discussion of the behavior of complex probes—including cases in other languages where it is proposed that partial matches can be safely skipped—see e.g. Coon and Bale 2014; Van Urk 2015; Deal 2015.

[44] This configuration of intervention by partial matches has also been discussed extensively in the literature on processing and acquisition. It has long been observed that mastery of object relative clauses comes much later than for subject relative clauses in the course of acquisition. Friedmann, Belletti, and Rizzi 2009 describes these effects as the child's inability to to establish a relationship with a [+D,+REL] target across an intervening [+D] in the configuration in (49). On the other hand, in subject relatives, the highest DP is relativized, avoiding this configuration. I refer the interested reader to Rizzi 2013, sec. 9 for a recent summary of this line of work.

This work suggests a possible explanation for the status of this constraint and its possible variation. Suppose this inability for complex probes to probe past partial matches is subject to cross-linguistic variation, as I suggest in footnote 43. In languages which have non-subject relatives, children come to learn that their languages do allow complex probing across partial matches, due to the many non-subject relatives in their linguistic input. But if a child is acquiring a language where extraction overwhelmingly targets subjects, their grammar will preserve the ban on complex probing across partial matches (49).

4.6 Summary

Motivated by the multiple extraction data presented in section 3, in this section I presented a new analysis of voice, case, and extraction in Toba Batak. Central to this analysis is the role of nominal licensing: Toba Batak lacks a structural Case-licensor lower in the clause, in VoiceP, but allows for licensing by adjacency with the verb. It's therefore impossible to move a non-subject DP to Spec,CP, as it will fail to be licensed. In contrast, non-DP constituents can be safely moved to Spec,CP as they are not subject to licensing. This derives the contrast from (1), repeated here as (50). The licensing by adjacency of non-subjects also accounts for restrictions on postverbal word order.

(50) **Only non-DPs can move to Spec,CP with the subject in Spec,TP:** (=1)

a. *[$_{CP}$ Aha [$_{TP}$ si Poltak [$_{VoiceP}$ man-uhor ___ ___]]]? (*maN-tuhor* > *manuhor*)
what PN Poltak ACT-buy

'What did Poltak buy?'

b. ✓[$_{CP}$ Andigan [$_{TP}$ si Poltak [$_{VoiceP}$ man-uhor buku ___ ___]]]?
when PN Poltak ACT-buy book

'When did Poltak buy a book?'

The existence of structures such as (50b) in Toba Batak has been observed briefly by Cumming (1984) and Emmorey (1984), but has not been seriously investigated. My account here straightforwardly derives it using a familiar organization of C and T in the Toba Batak clause periphery. This is the so-called "bodyguard" configuration discussed in the Malagasy literature (Keenan, 1976, a.o.), a connection I return to in section 5.3.

I further proposed that C and T have the option of being *bundled* into a single head, CT, with a bundled [PROBE:FOC+D] probe that combines properties of C's [PROBE:FOC] probe and T's [PROBE:D] probe. Standard properties of probing such as the requirement to Agree with the structurally closest target (see footnote 18), together with the inability of complex probes to skip partially matching goals (49), derive the famous "subject-only" extraction restriction, which limits $\overline{\text{A}}$-movement of DPs to the subject (Keenan and Comrie, 1977, a.o.), while also allowing for the multiple extraction of focused DPs as in (51):

(51) **CT can front multiple *wh*/focused nominals and Case-license them:** (=4/44b)

[$_{CTP}$ Aha [$_{CTP}$ [holan si Poltak] [$_{VoiceP}$ man-uhor ___ ___]]]?
what only PN Poltak ACT-buy

'What did only Poltak buy?'

According to this proposal, the organization of the clause periphery in Toba Batak could involve separate C and T heads or a bundled CT head, with the option of CT probing multiply and reprojecting, as in (51). In the next section, I present overt morphological evidence for these features of the organization of the Toba Batak clause periphery that I have proposed.

5 Spelling out T and the pseudocleft analysis

In this section, I introduce an additional empirical consideration: the distribution of the particle *na*. The particle *na* can be added optionally to many of the examples with fronting that I have considered, with some interspeaker variation in where it is allowed. I will show that the distribution of *na*—including the pattern of observed interspeaker variation—can be naturally captured under the proposal here, with *na* being the optional realization of either T or the bundled CT. The use of *na* will also offer explicit evidence for the CT reprojection proposal above, as well as an argument against a pseudocleft analysis for Toba Batak *wh*-questions, suggested by Hermon (2009).

5.1 The distribution of particle *na*

Silitonga 1973 claims that there are two distinct but homophonous free morphemes *na* in the language: a complementizer *na* and a *na* introducing relative clauses.[45] One immediately noticeable difference is that the *na* introducing embedded clauses is optional (52) but the *na* introducing relative clauses is obligatory (53).[46]

[45] Silitonga 1973, p. 132: "there is a complementizer in this language whose morphological form is identical with that of the relative pronoun: *na*." In response, Tuller 1984 notes "Our consultant feels strongly that there are 'two *na*'s' as well" (p. 191). There is also a bound morpheme *-na* which is a third-person possessive marker or pronoun.

[46] Silitonga 1973 also describes a separate, reduced relative strategy without the particle *na* but where the relative clause predicate is introduced by a prefix *si-*, which replaces the voice prefixes. In contrast, all examples here are full relative clauses where verbs include their voice morphemes—the type for which Silitonga notes that "the relative pronoun *na* does not seem to be deletable" (p. 122). *Na* is also obligatory at the beginning of headless relatives as well.

(52) ***na* introducing an embedded clause:**

Hu-boto [(na) modom si Poltak].
PASS.1sg-know NA sleep PN Poltak

'I know that Poltak is sleeping.'

(53) ***na* introducing a relative clause:**

Hu-ida [baoa [$_{RC}$ *(na) modom] i].
PASS.1sg-see man NA sleep MED

'I saw the man who is sleeping.'

I will first concentrate on the behavior of the optional *na* introducing embedded clauses such as in (52) and return to the *na* in relative clauses in section 5.2 below.

The optional *na* at embedded clause edges is described as a complementizer by Silitonga (1973) and others. I will however propose that its distribution is better modeled as spelling out the head T or, for one speaker, the bundled head CT. For this, I turn to the distribution of *na* in fronting constructions.

First we observe that *na* appears optionally in fronted *wh*-questions such as in (54), right after the fronted *wh*-phrase. It cannot follow preverbal auxiliaries such as *nunga* in (54) which I take to be in Asp.

(54) ***na* in a fronted *wh*-question:**

Ise (na) nunga (*na) ro?
who NA PERF *NA come

'Who came?'

All four of my speakers agree on the pattern presented in (54). However, there are other configurations where judgments systematically split into two patterns. The symbol % in (55) below indicates grammaticality for three speakers (Pattern A) but ungrammaticality for one (Pattern B). There is no position where *na* is accepted by the B speaker but not by A speakers. I note that the judgments I report here are stable across sessions and each speaker's behavior is internally consistent.

(55) **Configurations with systematic variation in the availability of *na*:**

a. Andigan (%na) man-uhor buku si Poltak?
when NA ACT-buy book PN Poltak

b. Andigan (*na) si Poltak (%na) man-uhor buku?
when NA PN Poltak NA ACT-buy book

'When did Poltak buy a book?'

The consistent pattern of variation here can be straightforwardly captured under my proposal. The key is the organization of the heads C and T. Recall that my derivation for the fronting of a [+FOC] subject DP as in (54)—where all speakers' judgments are in agreement—involves the bundled head CT. See (43) above for the derivation of this structure. The examples in (55)—where two patterns emerge—are those where C and T must be separate heads. In particular, the fronted *wh*-phrases in (55) are [+FOC, −D] so they must have been fronted by C with its [PROBE:FOC] feature, not T's [PROBE:D] or the bundled CT's [PROBE:FOC+D]. See (35, 38) above for the structures of (55a,b), respectively.

Here I adopt Distributed Morphology, a Late Insertion model of morphology (see e.g. Halle and Marantz, 1994; Embick and Noyer, 2007). I propose that Pattern A speakers employ *na* as the optional realization of the feature bundle [T] in the context of a local [C] feature whereas the Pattern B speaker uses *na* to spell out the specific feature bundle [C, T]:

(56) **Vocabulary insertion rules for optional *na*:**

$$na \text{ or } \emptyset \quad \leftrightarrow \quad \begin{cases} \text{[T] / [C] \_\_\_} & \text{(Pattern A)} \\ \text{[C, T]} & \text{(Pattern B)} \end{cases}$$

The [T] rule for Pattern A speakers will apply to the bundled head CT as well as to T with a structurally adjacent C. Under (56), all speakers allow for the pronunciation of CT as *na*, explaining the uniform availability of *na* in (54). Only Pattern A speakers allow for the pronunciation of *na* immediately before the verb in (55), because this is the locus of a T head, in the context of a structurally adjacent C. No speaker allows for *na* in between the *wh* non-DP and the non-focused DP in (55b) because this is the position of the unbundled C head, which matches neither of the rules in (56).

The reference to a local [C] feature in the vocabulary insertion rule for Pattern A speakers in (56) blocks the realization of *na* in unembedded declarative clauses. We observe that *na* generally cannot be used to introduce a declarative matrix clause, as in (57). (An apparent exception is discussed in (59) below.) The particle *na* also cannot appear in a matrix clause with a [−FOC] subject in preverbal position, with nothing else fronted, as in (58).

(57) ***na* cannot introduce a discourse-initial matrix clause:**

Context: discourse-initial

(*Na) modom si Poltak.
NA sleep si Poltak

'Poltak is sleeping.'

(58) ***na* disallowed in matrix clause with free subject fronting:**

(*Na) si Poltak (*na) modom.
NA PN Poltak NA sleep

'Poltak is sleeping.'

I propose that functional features such as [C] and [T] are not included in the clausal spine unless necessary—either for selectional purposes or for the use of their probes.[47] In the matrix clauses in (57) and (58), there is no C projected in the structure so the structural description of the Pattern A rule in (56) is not met. In contrast, [C] is present in clauses with fronting triggered by [PROBE:FOC] such as in (54–55) above as well as for embedded clauses as in (52), where the embedding verb selects for a [C] complement.

[47] The idea that functional structure is only present when independently necessary for the derivation has been independently proposed in Rizzi 1997 (see p. 314–315 and references there), Starke 2001, and Adger 2003. See also Erlewine 2016 (p. 475) for independent empirical motivation.

Interestingly, questions can be introduced with *na*, as in (i–ii). This may indicate the inclusion of C for marked (non-declarative) clause-typing purposes. I thank Dylan Tsai (p.c.) for asking about such examples.

(i) Na ro ise?
NA come who

'Who came?'

An apparent exception to the rule that *na* does not introduce matrix clauses is observed in responses to questions such as *Why?* or *What is someone doing?* as in (59). I tentatively explain such uses as involving a type of presentational embedding, akin to the English *It's that...* or *The answer is that....* Such clauses are therefore formally embedded clauses.[48] Note that there is no overt copula in Toba Batak.

(59) **Utterance-initial *na* is possible for broad-focus answers and explanations:**

Q: Margua si Poltak saonari?
do.what PN Poltak now

'What is Poltak doing now?'

A: (Na) modom (si Poltak).
NA sleep PN Poltak

'(It's / The answer is) that Poltak is sleeping.'

With this understanding of *na* in place, consider the distribution of *na* in examples with the simultaneous fronting of multiple formally focused DPs. Here *na* is acceptable after either fronted DP and in fact can be pronounced in both positions simultaneously, for all speakers:

(60) **The particle *na* with two *wh*/focus-fronted DPs:**

Ise (na) holan indahan (na) di-allang?
who NA only rice NA PASS-eat

'Who eats only rice?' (*na...na* ok too)

This configuration is precisely where my proposal predicts that the CT head will reproject, resulting in a final syntactic representation with two CT heads. See (46) above. The availability of the

(ii) Na modom do si Poltak?
NA sleep FOC PN Poltak

'Is Poltak sleeping?'

[48] Alternatively, these clauses may be a distinct form of matrix clause that may have grammaticalized from such an embedding—a form of what Evans 2007 calls "insubordination."

particle *na* in both positions simultaneously in (60) supports this CT reprojection account, and is not predicted under alternative proposals such as the simple use of multiple specifiers on CTP. At the same time, the lack of speaker variation in this judgment reported in (60) further supports the proposal for the realization of *na* in (56) above, which predicts no variation in the realization of bundled CT heads.

I should note that, for these cases of optional *na*, I have not been able to discern any semantic difference reflected in the presence or absence of *na*. Although the heads C and T in my proposal accord with many other languages in hosting [PROBE:FOC] in C and [PROBE:D] in T, T in Toba Batak is not associated with any tense semantics. Temporal interpretation is controlled through preverbal temporal auxiliaries, which I take to be in Asp and necessarily below T. See (54) above.

5.2 Against the pseudocleft analysis

The distribution of *na* above also relates to an alternative, pseudocleft analysis of fronting in Toba Batak. For illustration purposes, I will use an example discussed in Hermon 2009. Hermon proposes that (61) is a *wh*-pseudocleft structure:[49] *ise* 'who' with the focus enclitic *do* is the matrix predicate and its argument is the headless relative *na mangantuk biang i* 'the one that hit the dog.' Recall that relative clauses are obligatorily introduced by *na* (53). This pseudocleft parse is sketched in (62). *Op* here represents the null operator involved in relativization.

(61) **A possible *wh*-pseudocleft structure (Hermon, 2009, 785):**

Ise do na mang-antuk biang i?
who FOC NA ACT-hit dog MED

'Who is the one that hit the dog?' (Hermon's translation)

(62) Ise do [$_{RC}$ *Op* na mang-antuk biang i ___]?
who FOC NA ACT-hit dog MED

[49] Hermon (2009) includes a variant of example (61) lacking both *do* and *na*—*Ise mangantuk biang?*—which she describes as derived through fronting of the *wh*-word and not a pseudocleft construction. However, earlier in the paper she suggests that Toba Batak does not have argument *wh*-movement, only pseudoclefts: "The five VP-raising languages reviewed by P [=Potsdam 2009] (Malagasy, Maori, Niuean, Seediq, and Toba Batak) indeed do not allow argument Wh-movement. They stick to Wh-in-situ or clefts" (p. 779).

Under the pseudocleft analysis, only relativization is involved in such examples, not movement of the *wh*-phrase from the postverbal gap position. Such pseudocleft analyses have been widely adopted for apparent *wh* fronting in many Austronesian languages; see Potsdam 2009 for discussion and references.

I argue that Toba Batak does have true fronting. In general, the examples that I have described in this paper cannot be reanalyzed as instances of pseudoclefting. First, note that in all of the fronting examples that I discuss, the particle *na* is optional wherever it is possible; it is never obligatory. This is also true in the presence of the focus enclitic *do* as in (61), which is itself optional.[50] See also Jackson 1984 on the focus enclitic *do*, which also includes many examples of *do* on clause-initial constituents with corresponding postverbal gaps, which are not followed by *na*. This clearly contrasts from the use of *na* in relative clauses which is obligatory, casting doubt on a possible reanalysis of these examples as involving relativization.

Second, I return to the example of two simultaneously fronted focused DPs from (60) above. Recall that this example allows for *na* to be pronounced after both of the fronted DPs at the same time. If each *na* indicates a separate relative clause edge, the example must have a parse as in (63) below.

(63) **A pseudocleft parse for (60):**

Ise [$_{RC1}$ *Op* na holan indahan [$_{RC2}$ *Op* na di-allang ___ ___]]?
who NA only rice NA PASS-eat

≈ '[$_{RC1}$ The person *x* such that [$_{RC2}$ the thing that *x* eats] is only rice] is who?'

[50] An analysis of *do na* constructions as a pseudocleft construction is suggested in Van der Tuuk 1864/1971, the pioneering 19th century grammar of Toba Batak, written in Dutch. I refer to the 1971 English translation here. On page 343 while discussing *do*, Van der Tuuk suggests the following:

> "If *do* stands after a substantive [nominal] used predicatively or after a pronoun functioning as such a substantive, then, in order that the verb may function as the subject, it must be made into a substantive by *na*, e.g., *au do na mangoli* - I am one who comes to buy a wife."

That is, the verb *mangoli* 'buy a wife' must be turned into a nominal—relativized—by the addition of *na*, in order for it to be the argument of the predicate nominal *au* '1sg.' Like in the *wh* cases discussed here, for my speakers, the use of *do* on a clause-initial constituent does not necessitate the use of *na* as suggested by Van der Tuuk of the 19th century language.

Notice that a pseudocleft analysis in (63) forces us to relativize (RC1) over a non-subject position inside RC2. The result should be a relative clause island violation (Ross, 1967, a.o.). The grammaticality of (60/63)—and more generally, the equal grammaticality of (60) with the first *na*, the second *na*, or both—further suggests the particle *na* in fronting constructions should not be taken as evidence for the use of relativization in these cases of fronting with optional *na*.

5.3 Lessons for the Malagasy bodyguard

Finally, I note that the distribution of extraction asymmetries in Toba Batak and of the particle *na* is strongly reminiscent of extraction facts in Malagasy, another western Austronesian language, spoken in Madagascar. As noted above, Malagasy also exhibits a subject-only extraction restriction for DPs but allows some non-DP constituents to be fronted regardless of voice marking (Keenan, 1972, 1976, a.o.). Relevant here is the derivation of apparent *wh*/focus-fronting in Malagasy, which involve the preverbal particle *no*. Example (64a) is such an example of apparent *wh*-fronting:

(64) **Malagasy adjunct *wh*-question and the "bodyguard" variant (Potsdam, 2006, 212):**

a. Aiza no mividy mofo Rasoa?
where NO buy.ACT bread Rasoa

b. Aiza Rasoa no mividy mofo?
where Rasoa NO buy.ACT bread

'Where does Rasoa buy bread?'

Famously, in cases of adjunct *wh* or focus, the subject DP can optionally also be in a preverbal position, acting as a "bodyguard" (64b). Compare the pattern in (64) to the Toba Batak examples in (55) above. The parallel in word order is exact and, furthermore, the positions of Malagasy *no* are exactly the positions where Toba Batak Pattern A speakers can optionally add the particle *na*.

The nature of such apparent fronting and the status of this particle *no* have been of significant debate in the Malagasy literature; see Potsdam 2006 for an overview. In particular, the current consensus opinion is that these *no* constructions in Malagasy are pseudoclefts rather than the result of *wh*/focus-fronting, with *no* being a definite determiner or a specialized relative clause marker (see e.g. Dahl, 1986; Paul, 2001, 2003; Potsdam, 2006; Kalin, 2009). The particle *no* is not a general relative clause marker in the language (see Potsdam, 2006, 220ff), making the purported connection

to relativization more tenuous than in Toba Batak, where a particle *na* exists as a (homophonous, obligatory) relative clause marker. Furthermore, the position of the subject in the bodyguard construction (64b) has been perceived as a further challenge. Paul 2003 for example proposes that the bodyguard subject is a Spec,DP possessor of the pseudocleft headed by *no* in D, although possessors in Malagasy are otherwise postnominal.

My analysis for Toba Batak derives a parallel DP extraction asymmetry and bodyguard-like pattern in (55) through a familiar organization of C and T in the clause periphery, not involving relativization. These parallels suggests a new approach to the Malagasy clause periphery and bodyguard construction. However, here I highlight three important differences between the behavior of Malagasy and Toba Batak brought to my attention by a knowledgeable anonymous reviewer. First, they note that there is no free fronting of non-focused subjects as in Toba Batak. This is important because the mechanism for this fronting (probing by [PROBE:D] on a split T) is what I propose as the source of the fronted subject in bodyguard-like structures as in (55b). Second, the bodyguard construction in Malagasy disallows focused subjects, as reflected in (65), which is ungrammatical regardless of the position of *no*.

(65) **The Malagasy "bodyguard" subject cannot be focused (example from reviewer; cf (64b)):**

*Aiza (no) Rasoa ihany (no) mividy mofo?
where NO Rasoa only NO buy.ACT bread

Intended: 'Where is only Rasoa buying bread?'

This contrasts from the behavior of similar, bodyguard-like structures in Toba Batak, where the fronted subject can be formally focused, as in (22). Third, the reviewer notes that there is no multiple focused DP extraction as in the many Toba Batak examples presented here.

These differences suggest that, even if the basic analysis for Toba Batak here can be extended to Malgasy *no* and the bodyguard construction, crucial details regarding the organization of probes on C and T must differ between Malagasy and Toba Batak. I will leave the full exploration of this possibility for future work.

6 Conclusion

Toba Batak has been discussed as an exemplar of the "subject-only" extraction restriction of many Austronesian languages, ever since the discussion of the language in Keenan 1972 and Keenan and Comrie 1977. If a DP is extracted, it must be the subject DP, whose choice is cross-referenced on the verb. Further work on Toba Batak has attempted to relate this extraction restriction to the verb adjacency requirement of non-subject DPs (Clark, 1992; Baldridge, 2002; Cole and Hermon, 2008). In this paper, I presented patterns of multiple extraction which are unpredicted by any previous account of Toba Batak. These patterns motivate a familiar organization of the left periphery—C associated with *wh*/focus-fronting above T associated with subject-licensing—above a lower domain (VoiceP) with generally free word order.

I proposed here that nominal licensing is the key to explaining the distribution of extraction in Toba Batak, as well as the verb adjacency requirement of postverbal non-subject DPs. Although Toba Batak does not have morphological case, nominals nonetheless must be licensed through abstract structural Case assignment or by adjacency to the verb. This explains the inability to *wh*/focus-front a non-subject DP across a non-focused subject, as we can in English as in *What will Stephanie be buying?* (2a): the non-subject in Toba Batak will not be licensed, unlike the English accusative *what*.

At the same time, this account correctly allows for the exceptional extraction of non-subjects if they can be Case-licensed. This exceptional, additional licensor for non-subject DPs comes from the "bundling" of C and T together into CT, with a bundled [PROBE:FOC+D] head which inherits from both the basic C and T probes. This allows us to derive the simultaneous fronting of both core argument DPs when they are both formally focused, unlike the freezing account of Cole and Hermon 2008 which predicts non-subject fronting to be impossible. Furthermore, in section 5, I showed that my analysis for the organization of C, T, and bundled CT, can naturally explain the distribution of the optional particle *na*, as well as the shape of its interspeaker variation.

The work here represents the first steps towards a better understanding of Austronesian "voice" systems, which have long been discussed as a typologically distinct alignment system with a notable extraction restriction. Toba Batak shows us that Austronesian voice can be modeled—and in fact is best modeled—using a clausal organization familiar from many other language families, together with the careful consideration of nominal licensing.

References

Adger, David. 2003. *Core syntax: A minimalist approach*. Oxford University Press.

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

Allen, Barbara Jane. 1986. Goal advancement in Southern Tiwa. *International Journal of American Linguistics* 52:388–403.

Baker, Mark. 2014. Pseudo noun incorporation as covert noun incorporation: Linearization and crosslinguistic variation. *Language and Linguistics* 15:5–46.

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

Baldridge, Jason. 2002. Lexically specified derivational control in Combinatory Categorial Grammar. Doctoral Dissertation, University of Edinburgh.

Ball, Douglas. 2005. Tongan noun incorporation: Lexical sharing or argument inheritance. In *Proceedings of the 12th international conference on Head-Driven Phrase Structure Grammar*, 7–27. CSLI Publications.

Barbiers, Sjef. 2010. Focus particle doubling. In *Structure preserved: Studies in syntax for Jan Koster*, ed. Jan-Wouter Zwart and Mark de Vries, 21–30. John Benjamins.

Beck, Sigrid. 1996. Quantified structures as barriers for LF movement. *Natural Language Semantics* 4:1–56.

Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14:1–56.

Bennett, William G. 2009. Two subject asymmetries in Defaka focus extraction. Manuscript, Rutgers University.

Bennett, William G., Akinbiyi Akinlabi, and Bruce Connell. 2012. Two subject asymmetries in Defaka focus constructions. In *Proceedings of the 29th West Coast Conference on Formal Linguistics*, ed. Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz, and Alex Trueman, 294–302.

Bianchi, Valentina. 1999. *Consequences of antisymmetry: Headed relative clauses*. Mouton de Gruyter.

Bobaljik, Jonathan David. 1995. Morphosyntax: The syntax of verbal inflection. Doctoral Dissertation, Massachusetts Institute of Technology.

Bobaljik, Jonathan David, and Höskuldur Thráinsson. 1998. Two heads aren't always better than one. *Syntax* 1:37–71.

Bobaljik, Jonathan David, and Susi Wurmbrand. 2015. Questions with declarative syntax tell us what about selection? In *50 years later: Reflections on Chomsky's aspects*, 13–31. MIT Working Papers in Linguistics.

Bošković, Željko. 2015. From the Complex NP Constraint to everything: On deep extractions across categories. *The Linguistic Review* 32:603–669.

Bošković, Željko. 2016. On the timing of labeling: Deducing comp-trace effects, the subject condition, the adjunct condition, and tucking in from labeling. *The Linguistic Review* 33:17–66.

Browning, Margaret A. 1996. CP recursion and *that-t* effects. *Linguistic Inquiry* 27:237–255.

Cheng, Lisa Lai-Shen, and Norbert Corver, ed. 2006. *Wh-movement: Moving on*. MIT Press.

Chomsky, Noam. 1980. On binding. *Linguistic Inquiry* 11:1–46.

Chomsky, Noam. 1981. *Lectures on government and binding*. Foris.

Chomsky, Noam. 1986. *Barriers*. MIT Press.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20*, ed. Kenneth Hale and Samuel Jay Keyser, 1–52. MIT 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.

Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, ed. Robert Freidin, Carlos P. Otero, and Maria-Luisa Zubizarreta, 133–166. MIT Press.

Chung, Sandra. 1976. On the subject of two passives in Indonesian. In *Subject and topic*, ed. Charles N. Li and Sandra A. Thompson, 57–99. Academic 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.

Clark, Robin. 1984. The syntactic nature of Logical Form: Evidence from Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, number 5 in UCLA Occasional Papers in Linguistics, 9–16.

Clark, Robin. 1985. The syntactic nature of Logical Form: Evidence from Toba Batak. *Linguistic Inquiry* 16:663–669.

Clark, Robin. 1992. Towards a modular theory of coreference. In *Logical structure and linguistic structure*, ed. Cheng-Teh James Huang and Robert Carlen May, 49–78. Kluwer.

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

Clemens, Lauren Eby, and Maria Polinsky. to appear. Verb-initial word orders (primarily in Austronesian and Mayan languages). In *The Blackwell companion to syntax, 2nd edition*. Blackwell.

Cole, Peter, and Gabriella Hermon. 2008. VP raising in a VOS language. *Syntax* 11:144–197.

Cole, Peter, Gabriella Hermon, and Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118:1500–1553.

Coon, Jessica, and Alan Bale. 2014. The interaction of person and number in Mi'gmaq. *Nordlyd* 40:85–101.

Cumming, Susanna. 1984. The syntax and semantics of prepredicate word order in Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, number 5 in UCLA Occasional Papers in Linguistics, 17–36.

Dahl, Otto Christian. 1986. Focus in Malagasy and Proto-Austronesian. In *FOCAL 1*, ed. Paul Geraghty, Lois Carrington, and Stephen A. Wurm, 21–45. Pacific Linguistics.

Deal, Amy Rose. 2015. Interaction and satisfaction in φ-agreement. In *Proceedings of NELS 45*, ed. Thuy Bui and Deniz Özyıldız, 1–14.

Deal, Amy Rose. 2016. Raising to ergative: Remarks on applicatives of unaccusatives. Manuscript, University of California at Berkeley.

Embick, David, and Rolf Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32:555–595.

Embick, David, and Rolf Noyer. 2007. Distributed Morphology and the syntax/morphology interface. In *The Oxford handbook of linguistic interfaces*. Oxford University Press.

Emmorey, Karen. 1984. The intonation system of Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, number 5 in UCLA Occasional Papers in Linguistics, 37–58.

Erlewine, Michael Yoshitaka. 2016. Anti-locality and optimality in Kaqchikel Agent Focus. *Natural Language & Linguistic Theory* 34:429–479.

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.

Evans, Nicholas. 2007. Insubordination and its uses. In *Finiteness: Theoretical and empirical foundations*, ed. Irina Nikolaeva, 366–431. Oxford University Press.

Fortuny, Jordi. 2008. *The emergence of order in syntax*. John Benjamins.

Fox, Barbara A. 1984. Participant tracking in Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, number 5 in UCLA Occasional Papers in Linguistics, 59–79.

Friedmann, Naama, Adriana Belletti, and Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119:67–88.

Gallego, Ángel. 2014. Deriving feature inheritance from the Copy Theory of movement. *The Linguistic Review* 34:41–71.

Gallego, Ángel. to appear. The EPP in labeling theory: Evidence from Romance. *Syntax* .

Georgi, Doreen, and Gereon Müller. 2010. Noun-phrase structure by reprojection. *Syntax* 13:1–36.

Georgopoulos, Carol. 1991. *Syntactic variables: Resumptive pronouns and A' binding in Palauan*. Springer.

Giorgi, Alessandra, and Fabio Pianesi. 1996. *Tense and aspect: From semantics to morphosyntax*. Oxford University Press.

Grohmann, Kleanthes K. 2006. Top issues in questions: Topics—topicalization—topicalizability. In Cheng and Corver (2006), 249–288.

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.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In *The view from Building 20*, ed. Kenneth Hale and Samuel Jay Keyser, 111–176. MIT Press.

Halle, Morris, and Alec Marantz. 1994. Some key features of Distributed Morphology. In *MIT working papers in linguistics 21*, 275–288.

Harley, Heidi. 2017. The 'bundling' hypothesis and the disparate functions of little *v*. In *The verbal domain*, ed. Roberta D'Alessandro, Irene Franco, and Ángel Gallego, 3–28. Oxford University

Press.

Hermon, Gabriella. 2009. Language typology and universal grammar: A commentary on the paper by Eric Potsdam. *Natural Language & Linguistic Theory* 27:773–787.

Hiraiwa, Ken. 2001. Multiple agree and the defective intervention constraint in Japanese. In *Proceedings of HUMIT 2000*, number 40 in MIT Working Papers in Linguistics, 67–80.

Hsu, Brian. 2016a. Syntax-prosody interactions in the clausal domain: Head movement and coalescence. Doctoral Dissertation, University of Southern California.

Hsu, Brian. 2016b. Unification of feature scattering and M-merger as Coalescence. In *NELS 46: Proceedings of the 46th Meeting of the North East Linguistic Society*, ed. Christopher Hammerly and Brandon Prickett, volume 2, 133–146.

Hsu, Brian. 2017. Verb second and its deviations: An argument for feature scattering in the left periphery. *Glossa* 2:1–33.

Iatridou, Sabine, and Anthony Kroch. 1992. The licensing of CP-recursion and its relevance to the Germanic verb-second phenomenon. *Working Papers in Scandinavian Syntax* 50.

Jackson, Catherine A. 1984. Focus particles in Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, number 5 in UCLA Occasional Papers in Linguistics, 80–99.

Kalin, Laura. 2009. News about the *no*-phrase: Specificational pseudoclefts in Malagasy. In *Proceedings of AFLA 16*, 61–75.

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

Keenan, Edward L. 1972. Relative clause formation in Malagasy. In *The Chicago which hunt*, ed. P. M. Peranteau, J. N. Levi, and G. C. Phares, 169–189. Chicago Linguistic Society.

Keenan, Edward L. 2000. Morphology is structure: A Malagasy test case. In *Formal issues in Austronesian linguistics*, ed. Ileana Paul, Vivianne Phillips, and Lisa Travis, 27–48. Springer.

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

Kotek, Hadas. to appear. Intervention effects arise from scope-taking over alternatives. In *Proceedings of NELS 47*.

Kroeger, Paul. 1991/1993. Phrase structure and grammatical relations in Tagalog. Doctoral Dissertation, Stanford University.

Legate, Julie Anne. 2011. Under-inheritance. Presented at NELS 42.

Legate, Julie Anne. 2014. *Voice and* v*: Lessons from Acehnese*. MIT Press.

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

Martinović, Martina. 2015. Feature geometry and head-splitting: Evidence from the morphosyntax of the Wolof clausal periphery. Doctoral Dissertation, University of Chicago.

Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language & Linguistic Theory* 19:153–197.

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37:69–109.

Mayr, Clemens. 2014. Intervention effects and additivity. *Journal of Semantics* 31:513–554.

McCloskey, James. 1997. Subjecthood and subject positions. In *Elements of grammar*, ed. Liliane Haegeman, 197–235. Kluwer Academic Publishers.

Mordechay, Susan. 1984. The semantics and pragmatics of the "perfect" marker *nunga* in Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, number 5 in UCLA Occasional Papers in Linguistics, 100–121.

Nababan, Partabas Wilmar Joakin. 1966. Toba Batak, a grammatical description. Doctoral Dissertation, Cornell University.

Nababan, Partabas Wilmar Joakin. 1981. *A grammar of Toba-Batak*. Number 6 in Materials in languages of Indonesia. Pacific Linguistics.

Otsuka, Yuko. 2002. VOS in Tongan: passive or scrambling? In *Proceedings of AFLA 9*, 122–136.

Otsuka, Yuko. 2005. Two derivations of VSO: A comparative study of Niuean and Tongan. In *Verb first: On the syntax of verb initial languages*, ed. Andrew Carnie, Heidi Harley, and Sheila Ann Dooley, 65–90. John Benjamins.

Ouali, Hamid. 2008. On C-to-T feature transfer: the nature of agreement and anti-agreement in Berber. In *Agreement restrictions*, ed. Roberta D'Alessandro, 159–180. Mouton de Gruyter.

Paul, Ileana. 2000. Malagasy clause structure. Doctoral Dissertation, McGill University.

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

Paul, Ileana. 2003. Multiple topics: evidence from Malagasy. In *Proceedings of AFLA 9*, ed. Anastasia Riehl and Thess Savella, 137–148.

Pearson, Matthew. 2000. Two types of VO languages. In *The derivation of VO and OV*, ed. Peter Svenonius, number 31 in Linguistic Aktuell, 327–363. John Benjamins.

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.

Percival, W. Keith. 1981. *A grammar of the urbanized Toba-Batak of Medan*. Pacific Linguistics.

Pesetsky, David. 1982. Paths and categories. Doctoral Dissertation, Massachusetts Institute of Technology.

Pesetsky, David. 2000. *Phrasal movement and its kin*. MIT Press.

Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 355–425. MIT Press.

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.

Potsdam, Eric. 2006. The cleft structure of Malagasy *wh*-questions. In *Clause structure and adjuncts in Austronesian languages*, ed. Hans-Martin Gärtner, Paul Law, and Joachim Sabel, 195–232. Mouton de Gruyter.

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

Pylkkänen, Liina. 2002. Introducing arguments. Doctoral Dissertation, Massachusetts Institute of Technology.

Pylkkänen, Liina. 2008. *Introducing arguments*. MIT Press.

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.

Ramchand, Gillian, and Peter Svenonius. 2014. Deriving the functional hierarchy. *Language Sciences* 152–174.

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

Richards, Norvin. 1997. What moves where when in which language? Doctoral Dissertation, Massachusetts Institute of Technology.

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. Wiley-Blackwell.

Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In Cheng and Corver (2006), 97–133.

Rizzi, Luigi. 2013. Locality. *Lingua* 130:169–186.

Rizzi, Luigi, and Ur Shlonsky. 2007. Strategies of subject extraction. In *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*, ed. Uli Sauerland and Hans-Martin Gärtner, number 89 in Studies in Generative Grammar, 115–160. Mouton de Gruyter.

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

Sabel, Joachim. 2003. Malagasy as an optional multiple *wh*-fronting language. In *Multiple wh-fronting*, ed. Cedric Boeckx, 229–254. John Benjamins.

Schachter, Paul. 1984a. Semantic-role-based syntax in Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, number 5 in UCLA Occasional Papers in Linguistics, 122–149.

Schachter, Paul, ed. 1984b. *Studies in the structure of Toba Batak*. Number 5 in UCLA Occasional Papers in Linguistics.

Silitonga, Mangasa. 1973. Some rules reordering constituents and their constraints in Batak. Doctoral Dissertation, University of Illinois at Urbana-Champaign.

Sneddon, James Neil. 1996. *Indonesian: A comprehensive grammar*. Routledge.

Starke, Michal. 2001. Move dissolves into Merge: a theory of locality. Doctoral Dissertation, University of Geneva.

Sternefeld, Wolfgang. 1995. Voice phrases and their specifiers. In *Sfs report 05-95*.

Stowell, Timothy Agnus. 1981. Origins of phrase structure. Doctoral Dissertation, Massachusetts Institute of Technology.

Surányi, Balázs. 2005. Head movement and reprojection. *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae, Sectio Linguistica* 26:313–342.

Takahashi, Masahiko. 2010. Case, phases, and nominative/accusative conversion in Japanese. *Journal of East Asian Linguistics* 19:319–355.

Thráinsson, Höskuldur. 1996. On the (non)-universality of functional projections. In *Minimal ideas*, ed. Werner Abraham, Samuel David Epstein, Höskuldur Thráinsson, and Jan-Wouter Zwart, 253–281. John Benjamins.

Tuller, Laurice. 1984. Genitive case in Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, number 5 in UCLA Occasional Papers in Linguistics, 172–194.

van der Tuuk, Herman Neubronner. 1864/1971. *A grammar of Toba Batak*. Springer.

van Urk, Coppe. 2015. A uniform syntax for phrasal movement: A Dinka Bor case study. Doctoral Dissertation, Massachusetts Institute of Technology.

Vergnaud, Jean-Roger. 1977/2008. Letter to Noam Chomsky and Howard Lasnik. In *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, ed. Robert Freidin, Carlos P. Otero, and Maria-Luisa Zubizarreta, 3–15. MIT Press.

Watanabe, Akira. 1992. Wh-in-situ, subjacency, and chain formation. *MIT Occasional Papers in Linguistics* 2.

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

Wiltschko, Martina. 2014. *The universal structure of categories*. Cambridge.