

# Extraction and licensing in Toba Batak

Michael Yoshitaka ERLEWINE, December 2016

Comments welcome: [mitcho@nus.edu.sg](mailto:mitcho@nus.edu.sg)

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 heads C and T are present in Toba Batak, with their traditional 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. 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

**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 comments and discussion, I thank Edith Aldridge, Wayan Arka, Louisa Bielig, Mary Dalrymple, Amy Rose Deal, Hadas Kotek, Theodore Levin, Martina Martinović, David Pesetsky, Norvin Richards, Yosuke Sato, Coppe van Urk, Michelle Yuan, and audiences at AFLA 23 at the Tokyo University of Foreign Studies, the Australian Linguistics Society meeting at Monash University, MIT, the University of Hong Kong, and the National University of Singapore. 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

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Preliminaries</b>	<b>4</b>
2.1	Voice and word order . . . . .	4
2.2	Optionality of <i>wh</i> - and focus-fronting . . . . .	6
<b>3</b>	<b>Multiple fronting in Toba Batak</b>	<b>9</b>
<b>4</b>	<b>Proposal</b>	<b>14</b>
4.1	Clause structure and voice . . . . .	15
4.2	Nominal licensing . . . . .	19
4.3	Fronting with C and T . . . . .	26
4.4	Bundled CT . . . . .	31
4.5	Summary . . . . .	36
<b>5</b>	<b>Spelling out T and the pseudocleft analysis</b>	<b>37</b>
5.1	The distribution of particle <i>na</i> . . . . .	37
5.2	Against the pseudocleft analysis . . . . .	40
5.3	Lessons for the Malagasy bodyguard . . . . .	42
<b>6</b>	<b>Conclusion</b>	<b>43</b>

# 1 Introduction

Work on comparative formal syntax has identified two positions in the clause periphery, traditionally labeled C and T,<sup>1</sup> which are associated with two very different sets of properties (Chomsky, 1986, a.o.). T is commonly associated with properties of subjects, including  $\phi$ -agreement and nominative case assignment, and in many languages triggers movement of the subject to its specifier (the EPP property). In contrast, C 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.

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  $\bar{A}$ -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 traditionally associated with 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 elicitation 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 the heads C and T 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 two core conclusions. First, I propose that the heads C and T exist in Toba Batak, with their traditional functions—T 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).<sup>2</sup>

---

<sup>1</sup>C refers to *complementizer*, associated with clause-typing semantics, and T refers to *tense*, associated with finiteness and temporal interpretation. I simply refer to these heads as C and T here and concentrate on their functions as heads associated with certain syntactic processes. See also discussion in section 5.1.

<sup>2</sup>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 fn 25); RC = relative clause. For Southern Tiwa: AGR = agreement. For Niuean:

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

- a. \*<sub>[CP Aha <sub>[TP si Poltak [man-uhor \_\_\_\_ \_\_\_\_]]]</sub>]? (maN-tuhor > manuhor)</sub>
- what PN Poltak ACT-buy

'What did Poltak buy?'

- b. ✓[<sub>CP</sub> Andigan [<sub>TP</sub> 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).<sup>3</sup>

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

- a. ✓[<sub>CP</sub> What will [<sub>TP</sub> Stephanie [<sub>be</sub> [<sub>vP</sub> buying \_\_\_\_ ]? ]
- b. ✓[<sub>CP</sub> When will [<sub>TP</sub> Stephanie [<sub>be</sub> [<sub>vP</sub> \_\_\_\_ 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.

ERG = ergative, ABS = absolutive, LOC = locative. NA is used as the gloss for the particle *na* and FOC is the gloss for the focus enclitic *do*; both will be discussed in section 5. 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.

<sup>3</sup>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 which probes together. Following Martinović (2015), I will call such a head *CT*.<sup>4</sup> 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:**

[<sub>CTP</sub> 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:**

[<sub>CTP</sub> 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.<sup>5</sup>

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

---

<sup>4</sup>This relationship could also be modeled through feature inheritance (Chomsky, 2008; Ouali, 2008; Fortuny, 2008; Legate, 2011, a.o.). See footnote 36 below.

<sup>5</sup>I tend to use the terms *fronting* and *extraction* interchangeably to refer to the movement of constituents to a preverbal position.

periphery of Toba Batak. I develop my proposal in section 4 before presenting additional evidence from the particle *na* in section 5.

## 2 Preliminaries

### 2.1 Voice and word order

Toba Batak has a 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 in each sentence is in bold. The prefix on the verb (also in bold) correlates with the choice of subject argument.<sup>6</sup>

- |  |  |
|--|--|
| (5) a. <b>Man</b> -jaha buku <b>si</b> <b>Poltak</b> . | b. <b>Di</b> -jaha si Poltak <b>buku</b> . |
| ACT-read book PN Poltak                                | PASS-read PN Poltak book                   |
| ‘Poltak read a 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.<sup>7</sup>

The canonical declarative order is predicate-initial, but subject-initial clauses as in (6) are common in elicitation. Hermon (2009, 784) reports that subject-initial clauses comprise more than 50% of elicited sentences in her corpus. Cumming (1984, 17) reports that the proportion of sentences with fronted subjects ranges between 8% and 30% across a number of texts.

- |   |   |
|---|---|
| (6) a. <b>Si</b> <b>Poltak</b> [ <b>man</b> -jaha buku ____]. | b. <b>Buku</b> [ <b>di</b> -jaha si Poltak ____]. |
| PN Poltak ACT-read book                                       | book PASS-read PN Poltak                          |
| ‘Poltak read a book.’   |   |

<sup>6</sup>What I call “subject” here has also been called the “pivot” or “trigger” in some Austronesianist literature.

<sup>7</sup>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 effect.

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

(7) **Agent *wh*-question** ⇒ **ACTIVE:**

- |   |  |
|---|--|
| <p>a. ✓ <b>Ise</b> [ <b>mang</b>-allang babi ____ ]?<br/>               who ACT-eat       pork<br/>               ‘Who ate pork?’</p> | <p>b. *Ise [ <b>di</b>-allang ____ <b>babi</b> ]?<br/>               who PASS-eat       pork</p> |
|---|--|

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

- |  |   |
|--|---|
| <p>a. *Aha [ <b>man</b>-uhor ____ <b>si</b> <b>Poltak</b> ]?<br/>               what ACT-buy       PN Poltak<br/>               ‘What did Poltak buy?’</p> | <p>b. ✓ <b>Aha</b> [ <b>di</b>-tuhor si Poltak ____ ]?<br/>               what PASS-buy PN Poltak</p> |
|--|---|

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:**<sup>8</sup>

- |   |   |
|---|---|
| <p>a. ✓ [PP Tu ise] [ <b>man</b>-uhor buku <b>si</b> <b>Poltak</b> ____ ]?<br/>               for who ACT-buy   book PN Poltak</p> <p>b. ✓ [PP Tu ise] [ <b>di</b>-tuhor si Poltak <b>buku</b> ____ ]?<br/>               for who PASS-buy PN Poltak book</p> <p>      ‘[For who] did Poltak buy the book?’</p> | <p>(<i>maN-tuhor</i> &gt; <i>manuhor</i>)</p> |
|---|---|

---

<sup>8</sup>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.

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.<sup>9</sup> A similar paradigm of examples is presented in Silitonga (1973, 102–105).

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 syn-

---

<sup>9</sup>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) briefly advocates for such a cleft analysis for Toba Batak. 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.



tax questions” (Bobaljik and Wurmbrand, 2015), both of which cannot 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:**

- 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 you
  - b. Hu-boto            [man-uhor buku {andigan} ho {andigan}].  
                              PASS.1sg-know   ACT-buy    book    when    you    when
  - c. Hu-boto            [andigan [di-tuhor ho    buku]].  
                              PASS.1sg-know   when        ACT-buy    you book
  - d. Hu-boto            [di-tuhor ho    {andigan} buku {andigan}].  
                              PASS.1sg-know   PASS-buy    you    when    book    when
- ‘I know [when you bought the book].’

The picture is similar for focused constituents with the ‘only’ particle *holan* and the ‘even’ enclitic *pe*: they can be fronted or in-situ, but are very often fronted. One quirk is that an in-situ non-subject DP cannot be ‘only’ *holan*-marked—I will return to this fact in section 4.2.

(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
- c. \*Di-allang [holan si Poltak] indahan.  
           PASS-eat    only    PN Poltak    rice  
           ‘Only POLTAK ate rice.’

(14) **Optional but preferred focus-fronting with *pe* ‘even’:**

- a. [Naniura pe] [di-allang ahu \_\_\_\_].  
           naniura even   PASS-eat   1sg
- b. Di-allang ahu [naniura pe].  
           PASS-eat   1sg   naniura even
- c. Mang-allang [naniura pe] ahu.  
           ACT-eat        naniura even 1sg  
           ‘I eat even NANIURA.’<sup>10</sup>

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 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, indepen-

---

<sup>10</sup>*Naniura* is special preparation of raw fish pickled in lime juice and spices.

dent 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 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:

(15) ***Wh* agent, referential 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 (15a,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 (15a), and also allows *wh*-in-situ and free fronting of subjects, resulting in (15b). As shown in shown in (15c), these two

operations cannot cooccur to yield the *wh* DP followed the referential DP, both in preverbal position, regardless of the choice of voice morphology.<sup>11</sup>

The contrast in (16) below is completely parallel to (15), but with a referential agent and *wh* theme. Cole and Hermon (2008, 183) discuss examples such as (15c, 16c) as support for their view that non-subject DPs are frozen and cannot move, which will be discussed in section 4.2.

(16) ***Wh* theme, referential 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 (17) below. These examples contrast minimally from the ungrammatical (15c) and (16c) above in the non-*wh* DP being focused with *holan* ‘only,’ and are grammatical.<sup>12</sup>

---

<sup>11</sup>The opposite order—a referential 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. Third, this word order is disallowed in embedded clauses (ii).

- (i) Si Poltak<sub>i</sub> \*(#) aha di-allang (ibana<sub>i</sub>)?  
     PN Poltak       what PASS-eat 3sg.ANIMATE  
     ≈ ‘Poltak, what did he eat?’
- (ii) \*Hu-boto [si Poltak (#) aha di-allang (ibana)].  
     PASS.1sg-know PN Poltak       what PASS-eat 3sg.ANIMATE  
     Intended: ‘I know [Poltak, what he ate].’

In what follows, I will disregard such configurations involving hanging topics.

<sup>12</sup>The opposite order, with the *holan*-marked DP above the *wh*-phrase as in (i) below, is judged as degraded at best, which I take to be for reasons of semantic interpretation; see e.g. Beck 2006.

(17) **Multiple fronting of a *wh* DP and an ‘only’ DP:**

- a. Ise [holan indahan] [{\*mang/√di}-allang \_\_\_\_ \_\_\_\_]? cf (15c)  
 who only rice {\*ACT/√PASS}-eat  
 ‘Who ate only rice?’
- b. Aha [holan si Poltak] [{√mang/\*di}-allang \_\_\_\_ \_\_\_\_]? (=4), cf (16c)  
 what only PN Poltak {√ACT/\*PASS}-eat  
 ‘What did only Poltak eat?’

(18) and (19) below show similar examples of multiple fronting with two focused DPs:<sup>13</sup>

(18) **Multiple fronting of two ‘only’ DPs:**

[Holan posoposo] [holan susu] [{\*mang/√di}-inum \_\_\_\_ \_\_\_\_].  
 only infant only milk {\*ACT/√PASS}-drink  
 ‘Only infants only drink milk.’

(19) **Multiple fronting of an ‘only’ DP and an ‘even’ DP:**

[Holan si Poltak] [naniura pe] [{\*mang/√di}-allang \_\_\_\_ \_\_\_\_].  
 only PN Poltak naniura even {\*ACT/√PASS}-eat  
 ‘Only Poltak even eats naniura.’

Notice that in these examples in (17–19), 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 fronted past the subject.

To my knowledge, examples of the form in (17–19) have never before been described. The availability of these multiple DP extractions has a number of implications for our understanding of Toba Batak syntax. First, contrary to all previous descriptions, the non-subject DP *can* be extracted,

---

(i) [Holan indahan] ise [{<sup>??</sup>mang/\*di}-allang \_\_\_\_ \_\_\_\_]?  
 only rice who {<sup>??</sup>ACT/\*PASS}-eat

Intended: ‘Who ate only rice?’ (=17b)

<sup>13</sup>I have thus far not been able to elicit any multiple *wh*-questions.

though only in this particular, limited configuration. Second, the contrast between (17) and the minimally contrasting examples in (15–16) 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)—and the subject DP is grammatical in that order.<sup>14</sup> 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 from Emmorey 1984, 44:**

[PP Tu ise] mangga [di-lean hamu \_\_\_\_ \_\_\_\_]?  
to who mango PASS-send 2sg

‘To whom did you send the mango?’

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

... [holan sa-hali sa-taon do] halak [manuan 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*

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.

<sup>14</sup>The opposite order can also be grammatical but involves a hanging topic; see e.g. (i). See footnote 11 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?’

(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... (15–16)

b. ✓[+FOC, +D] [+FOC, +D] V... (17–19)

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

Like 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 [uF]. I use probing here for Case-licensing nominals and for triggering movement. Probes must Agree with the structurally closest matching target<sup>15</sup>; if the probe can probe multiply, subsequent probing can look past earlier, closer goals.

One additional assumption that I add is that the *use* or *invocation* of a probe is optional. For example, suppose the head X hosts a probe [uF]. Once X is introduced in the structure, [uF] can probe *or not*. 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, a.o.). The precise details of voice morphology and voice alternations are orthogonal to the core questions of this paper, so long as this desideratum is met.<sup>16</sup> For concreteness, however, I will now spell out one particular approach.

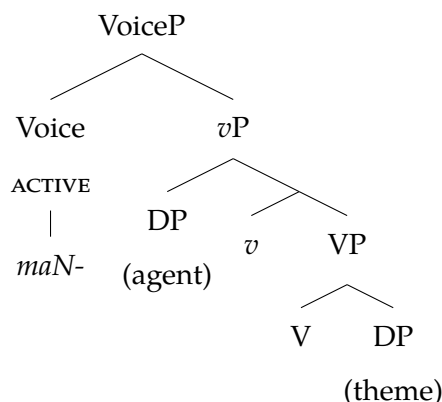
---

<sup>15</sup>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).

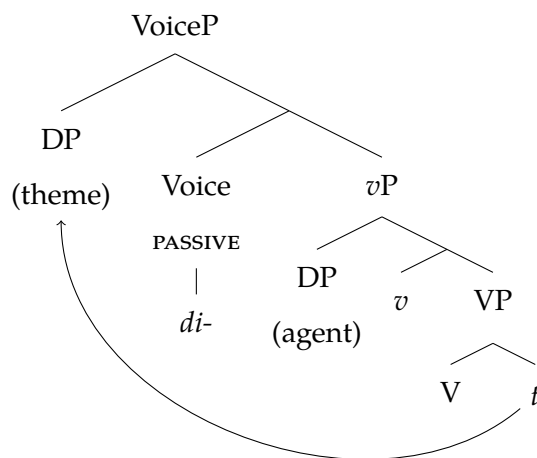
<sup>16</sup>I will point out, however, that there is one family of proposals for Austronesian voice which cannot be adopted for Toba Batak: so-called *wh*-agreement or case-agreement approaches, which have been proposed for Chamorro (Chung, 1982, 1994, 1998), Palauan (Georgopoulos, 1991), Tagalaog (Rackowski, 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,

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 *vP* as its complement.<sup>17</sup> The active *maN-* variant of Voice is functionally inert, whereas the passive *di-* variant triggers fronting of the theme DP.<sup>18</sup> 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:



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 thematic role, as proposed in Sternefeld 1995, would however still be an analytic option similar in spirit to case agreement.

<sup>17</sup>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,*vP*), the active Voice head may be inert.

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.

<sup>18</sup>The movement of the theme DP cooccurring with passive Voice could be thought of as the result of an obligatory nominal probe on Voice that triggers movement (strong/EPP [uD]), together with the effect of an anti-locality constraint which bans the movement of the agent from Spec,*vP* to Spec,VoiceP. An anti-locality constraint of this form, banning movement from one specifier to the immediately dominating projection's specifier, has been proposed by Erlewine (2016) and Bošković (2015, 2016): Spec-to-Spec Anti-Locality. Deal (2016) proposes that Spec-to-Spec Anti-Locality can result in probes (here, passive Voice) skipping the specifiers of complement phrases (here, Spec,*vP*), leading to attraction of a

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. I will 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 will discuss their approach in the following 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), 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).<sup>19</sup>

The clausal spine above VoiceP can optionally include negation and aspectual/modal auxiliaries, which are free morphemes.<sup>20</sup> 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, they necessarily precede these negative markers and auxiliaries, as seen in (26).<sup>21</sup> I will briefly discuss aspectual auxiliaries again in section 5.

---

lower target. But again, the exact relationship between this fronting of the theme subject and passive Voice is orthogonal to the core questions of the current paper.

<sup>19</sup>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.

<sup>20</sup>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, 86ff on the inventory of auxiliaries.

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

- a. Ahu dang olo mang-allang babi.  
 1sg NEG AUX ACT-eat pork  
 'I {do not want / am not willing} to eat pork.'
- b. Iban 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 any optional negation) is TP. The organization of T and C will be discussed in detail in the following sections 4.3 and 4.4.

---

<sup>21</sup>This is true even in examples of multiple extraction: negation and auxiliaries follow all fronted constituents and are immediately preverbal; see e.g. (i). 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, a.o.) and Acehnese (Legate, 2014), where both the theme and agent are simultaneously fronted, in that order, but auxiliaries and negation must precede the preverbal agent (ii–iii), 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?  
 when \*PERF PN Poltak PERF ACT-buy book MED  
 'When did Poltak buy that book?'
- (ii) 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.'
- (iii) 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.'

## 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, to appear) 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.<sup>22</sup> This is the function of T. The head T bears a [uD] probe which Case-licenses its target and can optionally attract it to its specifier.<sup>23</sup> Based on our derivations for Toba Batak clauses from the previous section, probing by [uD] 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.

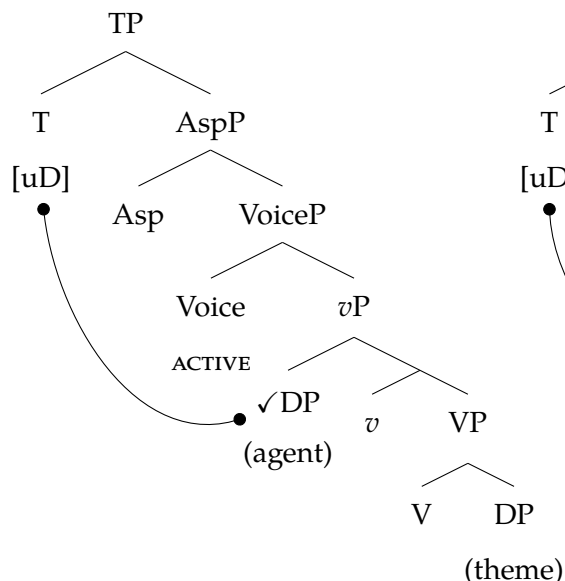
---

<sup>22</sup>I say “Austronesian-type” because, as Erlewine et al. (2015, to appear) 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.

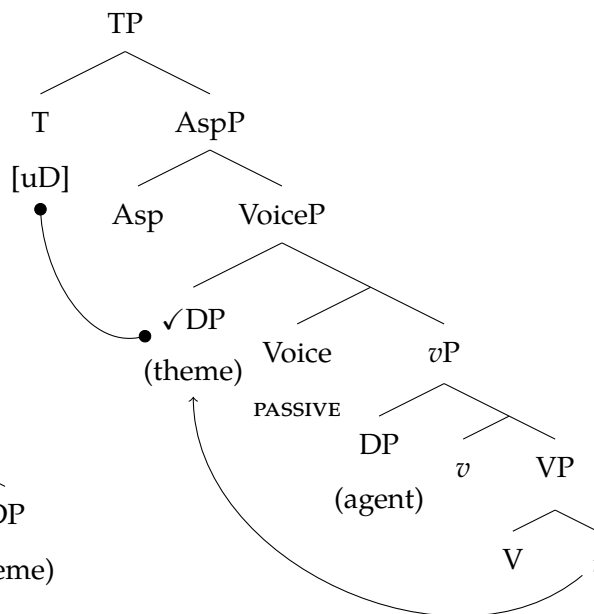
<sup>23</sup>I think of the nominal probe [uD] and its Case-licensing ability as the essence of the head labeled T, rather than any tense semantics. This is related in spirit to Pesetsky and Torrego’s (2001) idea that what we call nominative case is a [T] feature on nominals. As we have seen in (26), Toba Batak has preverbal aspectual auxiliaries but I will argue that they are not realizations of T, as I argue in section 5.1.

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

a. Active:



b. Passive:



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

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-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 by linear adjacency with the verb at PF*.<sup>24</sup> The idea that nominals can avoid the wrath of 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

<sup>24</sup>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.

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<sub>3pl</sub>-COME-PAST lady-to
- b. Am-**seuan**-wan-ban    hliawrade.  
       AGR<sub>3sg</sub>-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.

Based on the distribution of certain caseless arguments in Niuean, Massam (2001) proposed that some languages allow a form of incorporation of entire noun phrases, not just single words. This is called *Pseudo Noun Incorporation* or PNI. Consider the pair of examples in (29) below, which are synonymous. Niuean nominal arguments generally exhibit an ergative/absolutive pattern of case marking. The bolded theme ‘taro shoots’ is absolutive-marked in (29a), with the agent ‘family’ being ergative-marked, but undergoes PNI in (29b), shedding its case and number marking and moving to immediately postverbal position. The resulting clause in (29b) has only one Case-marked argument, the agent ‘family,’ which is now absolutive. The noun phrase ‘taro shoot’ is licensed under adjacency with the verb, rather than through case assignment.

(29) **Pseudo Noun Incorporation in Niuean (Clemens, 2014, 90):**

- a. Kua tō he magafaoa e tau huli talo he māla.  
 PERF plant ERG family ABS PL shoot taro LOC farm
- b. Kua tō huli talo e magafaoa he māla.  
 PERF plant shoot taro ABS family LOC farm

‘The family planted taro shoots at the farm.’

Similar cases of Pseudo Noun Incorporation, narrowly defined as applying to internal arguments, have been described for Chol (Mayan) (Coon, 2010) as well as in Sakha (Turkic) and Tamil (Dravidian) (Baker, 2014). Levin 2015 additionally relates PNI to patterns of in-situ agents in the Austronesian languages of Balinese and Malagasy, which must be immediately postverbal. Levin 2015 proposes that these nominals undergo *Local Dislocation* (Embick and Noyer, 2001, a.o.), a post-syntactic adjunction operation, allowing them to join the verbal extended projection and obviating the need for their nominal extended projections to be Case-licensed.<sup>25</sup>

My proposal is that this PF adjacency and 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 [uD] probe on T. In transitive clauses, postverbal non-subject DPs necessarily lack a structural Case licenser.<sup>26</sup> 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.

---

<sup>25</sup>Levin (2015) argues that the relevant form of adjacency is *head-head adjacency*; that is, the V head must be linearly adjacent to the highest head in the nominal’s extended projection, which I take to be D. This requires that the D head in Toba Batak be head-initial. 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, 94 for discussion that suggests that its function differs from that of definiteness. Here I follow Tuller 1984, 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.

<sup>26</sup>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.



In the remainder of this section, I will give two arguments for my proposal that Toba Batak employs such licensing by adjacency for non-subject DP arguments. Along the way, I will discuss and argue against the VP-fronting proposal for Toba Batak clause structure in Cole and Hermon 2008. I then return to my proposal for patterns of fronting in Toba Batak in section 4.3.

My proposal that non-subject DPs are licensed under adjacency with the verb 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 (30). 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 (30a) and passive agents as in (30b), reflecting the deep symmetry between the two voices in Toba Batak.<sup>27</sup>

(30) **Adding *nantoari* ‘yesterday’ to (5a,b):**

- 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 adjacency effect 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 pas-

---

<sup>27</sup>Clemens 2014 proposes that the verb adjacency requirement of Niuean PNI, observed in (29), is prosodically motivated. In brief, the head-complement relationship between the verb and a PNI-ed argument must be reflected in the prosodic structure with the verb and argument being adjacent constituents in a shared phonological ( $\varphi$ ) phrase (p. 130). This forces a linearization in Niuean where the PNI argument and the verb are adjacent. This approach could extend to the adjacency effect for Toba Batak active themes but does not extend to passive agents, which also obey the adjacency requirement (30b). Assuming the Uniformity of Theta Alignment Hypothesis (Baker, 1988), passive agents and active themes will be in different structural relationships with the verb, making it difficult to formulate a uniform prosodic motivation for these and only these arguments necessarily forming a tight prosodic unit with the verb. In contrast, the account here which necessitates adjacency by considerations of nominal licensing is successful in treating active themes and passive agents as a natural class.

sive agent and active theme necessarily evacuate the VoiceP,<sup>28</sup> 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 (30). Second, it predicts that non-subject DPs cannot ever be extracted, because it is frozen through the movement of VoiceP—a welcome consequence, according to Cole and Hermon.<sup>29</sup>

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. The relevant examples are repeated below in (31). The grammaticality of the examples in (31), repeated from (17), teaches us that it is false that non-subject DPs cannot be fronted and it is also false that non-subject DPs must always stay adjacent to the verb.

(31) **Multiple fronting of a *wh* DP and an ‘only’ DP:** (=17)

- a. Ise [holan indahan] [di-allang \_\_\_\_ \_\_\_\_]?  
     who only rice PASS-eat  
     ‘Who ate only rice?’
- b. Aha [holan si Poltak] [mang-allang \_\_\_\_ \_\_\_\_]?  
     what only PN Poltak ACT-eat  
     ‘What did only Poltak eat?’

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 the lack of any alternative licensing strategy for these DPs. In the case of multiple focused DP extraction as in (31), I will propose that the probe involved with the extraction of the DPs itself Case-licenses its targets. In this limited configuration, then, both core argument DPs will be structurally Case-licensed in tandem with their fronting. Details of fronting and probing in this

---

<sup>28</sup>In my reading of Cole and Hermon 2008, the technical mechanism for triggering the evacuation of all material except this one non-subject DP (if any) is unclear. See also Chung 2008 for additional discussion of Cole and Hermon 2008 from the perspective of its possible application to the clause structure of Indonesian.

<sup>29</sup>Clark 1992 and Baldrige 2002 also present accounts for Toba Batak that directly tie the inability of non-subjects to be extracted to their adjacency requirement. For Baldrige 2002, in a flavor of Combinatory Categorical 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 24), predicting as I do that non-subject extraction could become possible if an independent source of licensing becomes available.

configuration in (31) will be presented in detail in section 4.4. But at this point it suffices to say that the approach in Cole and Hermon 2008—the only existing modern analysis of Toba Batak clause structure—is unable to account for the data presented in section 3, such as (31).

The licensing by adjacency account also explains a peculiar contrast observed above between the focus particles *holan* ‘only’ and *pe* ‘even.’ I showed above that DPs modified with either of these focus particles prefer to front, but they could also be in-situ as the subject or non-subject, with one exception: *holan*-marked DPs cannot be postverbal non-subjects. Examples (13c) and (14c) are repeated here in (32).

(32) **A contrast between *holan*- and *pe*-marked non-subject DPs:**

- a. \*Di-allang [holan [DP si Poltak]] indahan. (=13c)  
 PASS-eat only PN Poltak rice  
 ‘Only POLTAK ate rice.’
- b. Mang-allang [[DP naniura] pe] ahu. (=14c)  
 ACT-eat naniura even 1sg  
 ‘I eat even NANIURA.’

This contrast is explained straightforwardly by the licensing by adjacency account. The difference between (32a) and (32b) is simply that *holan* precedes its focus whereas *pe* follows its focus. I assume focus particles adjoin to DPs as illustrated. Licensing by adjacency requires the highest layer of the DP to be linearly adjacent to the verb (see footnote 25). In example (32a), the ‘only’ particle *holan* itself interrupts the linear adjacency between the verb and *si Poltak* necessary for its licensing, resulting in its ungrammaticality. In contrast, the ‘even’ particle *pe* follows the DP in (32b) and therefore does not interrupt the linear adjacency condition.

Note that this contrast is not accounted for by Cole and Hermon’s (2008) proposal: in their approach, the non-subject DP is stranded within VoiceP while everything else evacuates, leading to its linear adjacency with the verb. Modifiers on non-subject DPs—let alone their linear position—are not predicted to affect the convergence of the derivation.<sup>30</sup>

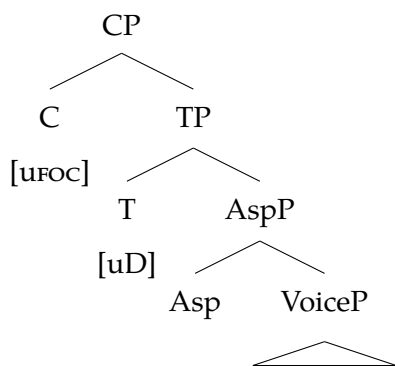
<sup>30</sup>One might be tempted to stipulate an alternative difference between *holan* and *pe*: suppose *holan*-marked DPs are generally required to be the subject argument, but *pe*-marked DPs do not share this requirement. This alternative cannot be correct. The grammatical multiple extraction examples in (18) and (19) show that *holan* DPs can be non-subjects, so long as they are fronted and structurally Case-licensed in a multiple extraction configuration.

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 licenser, the [uD] probe on T. Due to the syntax of voice, presented in section 4.1 above, [uD] will Case-license the subject, leaving postverbal non-subjects without licensing. In such a situation, the non-subject can be licensed by adjacency with the verb (Levin, 2015). This offers a new account for the adjacency requirement on non-subjects discussed by Schachter 1984a and Cole and Hermon 2008. This account is preferable to Cole and Hermon’s (2008) VP-fronting analysis that derives the adjacency effect but predicts that non-subjects are completely immobile, contrary to fact, and which cannot account for the interaction with focus particles presented here.

### 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 traditional 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 fronts them. The proposed structure of the clause periphery is sketched in (33). I will treat both C and T as simply unpronounced in this section but will return to this question in section 5.

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



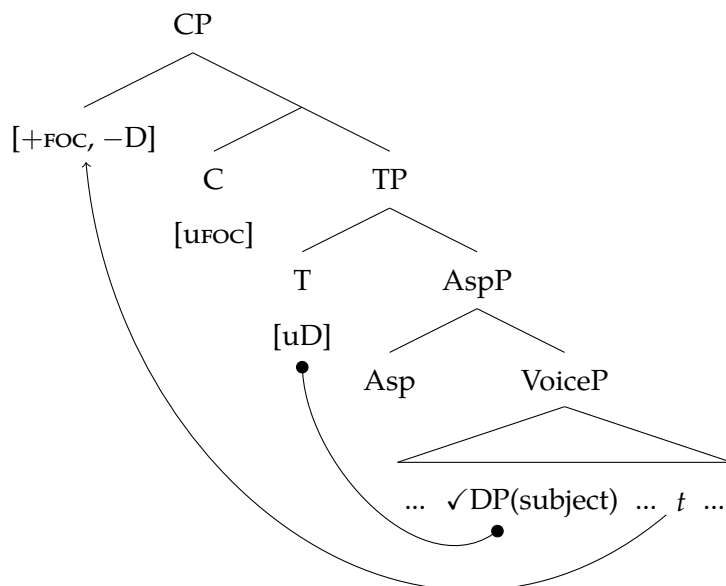
Given this familiar organization of C and T which can both independently attract constituents to their specifiers, why are multiple extractions in Toba Batak so limited? The key, I propose, is the limited means of nominal licensing in the language, as discussed in the previous section. The

[uFOC] 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 [uD] and [uFOC] probes, step by step. As discussed in the previous section, the [uD] 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 [uFOC] 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 [uFOC] finds a [+FOC, -D] target: a formally focused non-DP. In contrast to [uD] on T, I take [uFOC] to obligatorily front any target that it Agrees with to Spec,CP.<sup>31</sup> Because the target is not nominal, we do not have to worry about its licensing. This structure is illustrated in (34) below.

(34) **Fronting [+FOC, -D] to Spec,CP:**



This derivation corresponds to our examples of *wh* or focused non-DP fronting. Because [uFOC] 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 subject (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 (35):

<sup>31</sup>Recall, however, that the *invocation* of the probe is optional, though, allowing for optionality of the *wh*-movement in (34/35).

(35) **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 [uFOC] probe on C has the ability to probe and attract multiple targets, which is yet another pattern of multiple fronting. This is exemplified in (36) below.<sup>32</sup>

(36) **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?’ (maN-angko > manangko)

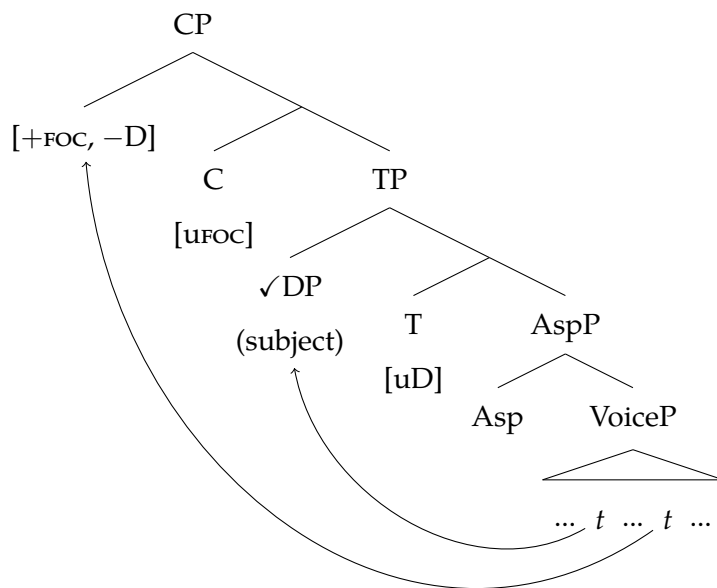
(≈ Why don’t you steal books from other places?)

Note that this ability to probe and attract multiply is a property of the [uFOC] probe, not shared by [uD]. If [uD] 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 [uD] 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 (37). We have now successfully derived the so-called “bodyguard” configuration of multiple fronting: a formally focused non-DP and the subject fronted together, in that order.

<sup>32</sup>There is a question here in (36) regarding the timing of the two movements. I will propose in 4.4 that multiple movements by the same probe do not “tuck in,” *pace* Richards 1997; i.e. in (36), *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.

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



A grammatical example of this “bodyguard” configuration is repeated in (38) below. As indicated by the voice morphology in (38), the fronted DP in this configuration must be the subject. This is explained by the approach here because the DP fronting is the result of probing by [uD] 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.<sup>33</sup>

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

[<sub>CP</sub> Andigan [<sub>TP</sub> [(holan) si Poltak] ... [<sub>VoiceP</sub> {<sup>√</sup>maN/\*di}-tuhor buku \_\_\_\_ \_\_\_\_]]]?  
when only PN Poltak {<sup>√</sup>ACT/\*PASS}-buy book

‘When did (only) Poltak buy the book?’

(*maN-tuhor* > *manuhor*)

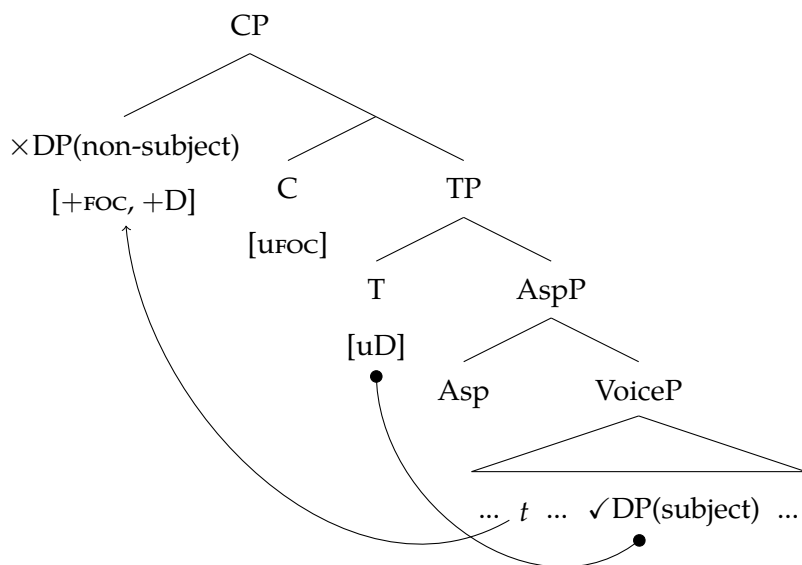
From these derivations presented above, we see that the independent effects of [uFOC] on C and [uD] on T are alive and well in Toba Batak. Movement to Spec,TP is driven by [uD] and is therefore limited to the highest DP in VoiceP, i.e. the subject. In contrast, movement to Spec,CP is driven by a [uFOC] probe, which simply attracts the closest [+FOC] constituent in VoiceP.<sup>34</sup> Finally, the independence of these movements to Spec,TP and Spec,CP predicts the availability of fronting a [+FOC]

<sup>33</sup>In the case where the subject is [+FOC], I assume the [uD] probe on T moves the subject to Spec,TP first. Subsequent probing by [uFOC] on C must be able to skip the subject in Spec,TP, probing past it to find another [+FOC] target in VoiceP. I assume that this is the effect of an anti-locality constraint banning the attraction of the specifier of the probe’s complement (Erlewine, 2016; Bošković, 2015, 2016; Deal, 2016). See also footnote 18 above.

non-DP and the subject simultaneously, in that order, resulting in the so-called “bodyguard” configuration (38). 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 [uFOC] attracting a non-DP target. But what happens if probing by [uFOC] 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 [uD] probe on T probe down and Case-license the subject. Next, let the [uFOC] probe on C probe down and find the [+FOC] non-subject DP and front it. This results in the structure in (39).

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



The problem with the configuration in (39) has to do with the licensing of the non-subject DP. Recall that there is no structural Case licenser in the clause except for the [uD] probe on T, but [uD] here is licensing the subject and [uD] cannot probe multiply. If the non-subject stayed in postverbal

<sup>34</sup>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.



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 (39) bleeds the possibility of licensing by adjacency.<sup>35</sup> 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 (39) 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 (39) shows that C could attract a [+FOC] non-subject DP but it would then lack licensing. Note that the subject could also be fronted to Spec,TP in (39), 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 (40):

(40) **Multiple extraction of [+FOC, +D] [−FOC, +D] is ungrammatical:** (=1a/16c)

\*[<sub>CP</sub> Aha [<sub>TP</sub> si Poltak ... [<sub>VoiceP</sub> 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 very similar extraction restriction, spoken just north of the Bataks—as well as in Germanic subject V2. See also Legate 2014, 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

<sup>35</sup>The non-subject could be immediately *pre-verbal* in (39), but this is not sufficient to license the non-subject, due to the directionality requirement of licensing by adjacency. See footnote 25.

mood (Giorgi and Pianesi, 1996), tense and agreement (Bobaljik, 1995; Thráinsson, 1996; Bobaljik and Thráinsson, 1998, a.o.), complementizer systems (Bianchi, 1999), causatives (Pylkkänen, 2002, 2008), and V2 requirements (Hsu, 2016, to appear b). 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).<sup>36</sup>

When CT is bundled, the [uFOC] probe of C and the [uD] probe of T operate as a single probe that seeks targets that are simultaneously [+FOC, +D], i.e. matching the specifications of [uFOC] and [uD] at the same time. 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 [uFOC+D] probe on CT will inherit the marked (‘yes’) values for these specifications from [uD] and [uFOC]. These specifications are summarized in (41). In the rest of this section, I will demonstrate each of these aspects of the [uFOC+D] probe and their use.

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

	<u>C = [uFOC]</u>	+	<u>T = [uD]</u>	=	<u>CT = [uFOC+D]</u>
Case-licenses target?	no		yes		yes
Must front target?	yes		no		yes
Can probe multiply?	yes		no		yes

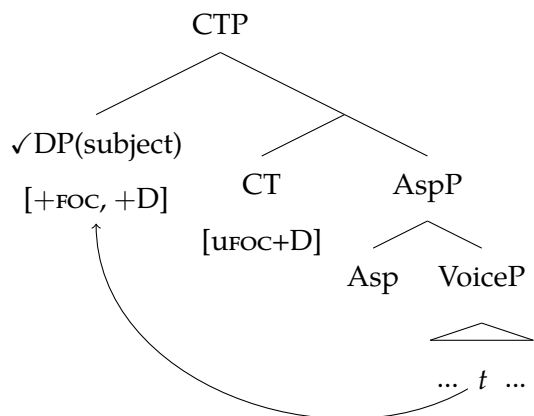
I begin by discussing the case where the subject DP is formally focused: [+FOC, +D]. Probing by the bundled [uFOC+D] will find the focused subject and front it. This is illustrated in (42).

---

<sup>36</sup>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 = [uFOC] and T = [uD] which then bundle presyntactically into a single CT = [uFOC+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 2016, to appear a for further technical discussion of such bundling operations.

Alternatively, Giorgi and Pianesi 1996 and Martinović 2015 conceive of features as originally bundled and 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.

(42) **Fronting a focused subject with CT = [uFOC+D]:**



The [uFOC+D] probe inherits [uD]’s Case-assigning ability (41), licensing the subject DP. This structure corresponds to examples with a single fronted *wh* or focused DP.

Now suppose the [uFOC+D] probes again. This is possible because it inherits [ufoc]’s ability to probe multiply (41). If there is another [+FOC, +D] constituent past the subject—i.e. a formally focused non-subject DP—[uFOC+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 (17), repeated here in (43):

(43) **Multiple extraction of two [+FOC, +D] constituents:** (=17/31)

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

‘Who ate only rice?’

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

‘What did only Poltak eat?’

As indicated by the voice markers in (43), 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 32); i.e. these probes do not “tuck in,” *pace* Richards 1997. Probing

I furthermore propose that CT hosts multiple specifiers through *reprojection*: CT head-moves to merge with its own CTP projection, projecting a higher CTP.<sup>37</sup> This is illustrated in (44) below, with the thick arrow for reprojection of CT. I will present morphological evidence for this reprojection of the CT head in section 5.

CTP

✓DP(non-subject)

[+FOC, +D]

CT

[uFOC+D]

CTP

✓DP(subject)

[+FOC, +D]

CT

[uFOC+D]

AspP

Asp

VoiceP

...

t

...

t

...

(45) A couple structures (correctly) not generated by the proposal:

- Intended: 'Who ate only rice?' (=43a)

34

b. \*[CTP [subject Aha] [VoiceP di-allang [non-subject holan [DP si Poltak]] \_\_\_\_]]?  
                   what                   PASS-eat                   only                   PN Poltak

Intended: ‘What did only Poltak eat?’ (=43b)

Assume for contradiction that [uFOC+D] attracts its targets *optionally*. In (45a), let [uFOC+D] on CT probe down, find the [+FOC] subject DP, and Case-license it, but *not* move it. Subsequent probing by [uFOC+D] could then front the formally focused non-subject DP. This would incorrectly predict example (45a) to be grammatical. Conversely, suppose [uFOC+D] finds a [+FOC] subject, Case-licenses it, and fronts it. Now notice that, due to the ‘only’ particle *holan* intervening between the non-subject DP and the verb, the non-subject cannot be licensed by adjacency with the verb. (See (32) above.) However, in this hypothetical derivation [uFOC+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 (45b). These examples show that the bundled [uFOC+D] probe must attract the targets that it Agrees with (41). Recall, however, that the initial invocation of these probes is still optional, allowing for optionality in fronting; what is necessary is to front every target that the probe interacts with.

Finally, consider the case where the subject DP is [−FOC] but there is a lower [+FOC, +D] constituent in the clause. If [uFOC+D] can probe down and Case-license and attract the [+FOC, +D] non-subject, we would end up with a structure as in (46) below. The in-situ subject could be licensed by adjacency with the verb, so this derivation would incorrectly predict (46) to be grammatical.

(46) **Ungrammatical example generated by [uFOC+D] probing across the subject:** (=8a)

\*[CTP [non-subject Aha] ... [VoiceP man-uhor \_\_\_\_ [subject si Poltak]]]]?  
                   what                   ACT-buy                   PN Poltak

Intended: ‘What did Poltak buy?’

I propose that in such cases, probing by [uFOC+D] will fail due to *defective intervention* (see e.g. Chomsky, 2000). Defective intervention occurs when the presence of a closer, imperfect target blocks successful probing of a lower, matching target. In this case, probing down by [FOC+D] will first encounter the pivot. Although the pivot is not a satisfactory target for the [uFOC+D] probe, it

is a partial match because it is [+D]. Due to defective intervention, the probe will not be able to engage the lower, matching target [+FOC, +D], explaining the ungrammaticality of (46).

## 4.5 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 (47). The licensing by adjacency of non-subjects also accounts for restrictions on postverbal word order.

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

- a. \*[<sub>CP</sub> Aha [<sub>TP</sub> si Poltak ... [<sub>VoiceP</sub> man-uhor \_\_\_\_ \_\_\_\_]]]? (maN-tuhor > manuhor)  
           what       PN Poltak                   ACT-buy

‘What did Poltak buy?’

- b. ✓[<sub>CP</sub> Andigan [<sub>TP</sub> si Poltak ... [<sub>VoiceP</sub> man-uhor buku \_\_\_\_ \_\_\_\_]]]?  
           when               PN Poltak                   ACT-buy   book

‘When did Poltak buy a book?’

The existence of structures such as (47b) 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 [uFOC+D] probe that combines properties of C’s [uFOC] probe and T’s [uD] probe. Standard properties of probing à la Chomsky 2000 such as the requirement to Agree with the structurally closest target and sensitivity to defective interveners derive the famous “subject-only” extraction restriction, which limits  $\bar{A}$ -movement of DPs to the subject (Keenan and Comrie, 1977, a.o.), while also allowing for the multiple extraction of focused DPs as in (48):

(48) **CT can front multiple *wh*/focused nominals and Case-license them:** (=4/43b)

[<sub>CTP</sub> Aha [<sub>CTP</sub> [holan si Poltak] ... [<sub>VoiceP</sub> 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 (48). In the next section, I present overt morphological evidence for these features of the organization of C and T 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.<sup>38</sup> One immediately noticeable difference is that the *na* introducing embedded clauses is optional (49) but the *na* introducing relative clauses is obligatory (50).<sup>39</sup>

<sup>38</sup>Silitonga 1973, 132: “there is a complementizer in this language whose morphological form is identical with that of the relative pronoun: *na*.” In response, Tuller 1984, 191 notes “Our consultant feels strongly that there are ‘two *na*’s’ as well.” There is also a bound morpheme *-na* which is a third-person possessive marker or pronoun.

<sup>39</sup>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 “The relative pronoun *na* does not seem to be deletable” (p. 122).

(49) *na* introducing an embedded clause:

Hu-boto [(na) modom si Poltak].

PASS.1sg-know NA sleep PN Poltak

‘I know that Poltak is sleeping.’

(50) *na* introducing a relative clause:

Hu-ida [baoa [<sub>RC</sub> \*(na) modom]].

PASS.1sg-see man NA sleep

‘I saw the man who is sleeping.’

I will first continue to discuss the *na* introducing embedded clauses such as in (49) 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 subsequent work. 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 (51), right after the fronted *wh*-phrase. It cannot follow preverbal auxiliaries such as *nunga* in (51) which I take to be in Asp.

(51) *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 (51). However, there are other configurations where judgments systematically split into two patterns. The symbol % in (52) 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.



(52) **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] pivot DP as in (51)—where all speakers’ judgments are in agreement—involves the bundled head CT. See (42) above for the derivation of this structure. The examples in (52)—where two patterns emerge—are those where C and T must be separate heads. In particular, the fronted *wh*-phrases in (52) are [+FOC, –D] so they must have been fronted by C with its [uFOC] feature, not T’s [uD] or the bundled CT’s [uFOC+D]. See (34, 37) above for the structures of (52a,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] whereas the Pattern B speaker uses *na* to spell out the more specific feature bundle [C, T]:

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

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

Note that the [T] rule for Pattern A speakers will apply to the bundled head CT as well as to T, in the absence of a more specific rule for CT. Therefore all speakers allow for the pronunciation of *na* in CT, explaining (51). Only Pattern A speakers allow for the pronunciation of *na* immediately before the verb in (52), because this is the locus of the T head. No speaker allows for *na* in between the *wh* non-DP and the unfocused DP in (52b) because this is the position of the unbundled C head, which matches neither [T] nor [C, T] in (53). This proposal also accounts for the optional inclusion of *na* at the edges of embedded clauses as in (49), though not as a complementizer per se.

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 added in both positions simultaneously, for all speakers:

(54) **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 (44) above. The availability of the particle *na* in both positions simultaneously in (54) supports this CT reprojection account here, 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 (54) further supports the proposal for the realization of *na* in (53) 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 [uFoc] in C and [uD] 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 is necessarily below T; see (51) 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 by Hermon 2009. Hermon proposes that (55) is a *wh*-pseudocleft structure:<sup>40</sup> *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* (50). This pseudocleft parse is sketched in (56). *Op* here represents the null operator involved in relativization.

(55) A possible *wh*-pseudocleft structure, from 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)

(56) Ise do [RC Op na mang-antuk biang i ]?  
 who FOC NA ACT-hit dog MED

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 citations.

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 (55), which is itself optional.<sup>41</sup> 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.

<sup>40</sup>Hermon 2009 includes a variant of example (55) 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).

<sup>41</sup>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.

Second, I return to the example of two simultaneously fronted focused DPs from (54) 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 (57) below.

(57) **A pseudocleft parse for (54):**

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?'

Notice that a pseudocleft analysis in (57) 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 (54/57)—and more generally, the equal grammaticality of (54) 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 the presentation of extraction asymmetries in Toba Batak and the distribution 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 (58a) is such an example of apparent *wh*-fronting:

(58) **Malagasy adjunct *wh*-question and the “bodyguard” variant (Potsdam, 2006, 212):**

- a. Aiza no mividy mofo Raso?
 

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

where Raso NO buy.ACT bread

‘Where does Raso 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” (58b). Compare the pattern in (58) to the Toba Batak examples in (52) 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 (58b) 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.

Here I will not review or take aim at the various arguments presented in this literature for the pseudocleft analysis in Malagasy as well as its extension to the bodyguard construction. However, I will note that my own analysis for Toba Batak here—deriving a parallel DP extraction asymmetry and bodyguard-like pattern in (52) through a familiar organization of C and T in the clause periphery, not relativization—naturally suggests an alternative account for these extraction patterns in Malagasy as well. 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 (Chomsky, 1981, a.o.)—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 licenser for non-subject DPs comes from the “bundling” of C and T together into CT, with a bundled [ $\text{uFOC}+\text{D}$ ] 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 strictly predicts non-subject 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

- 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 Categorical Grammar. Doctoral Dissertation, University of Edinburgh.
- 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.
- 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.
- Chung, Sandra. 2008. Indonesian clause structure from an Austronesian perspective. *Lingua* 118:1554–1582.
- 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. 2010. VOS as predicate fronting in Chol. *Lingua* 120:354–378.
- 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. 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. to appear. Ergativity and Austronesian-type voice systems. In *Oxford Handbook of Ergativity*, ed. Jessica Coon, Diane Masmam, and Lisa deMena Travis. 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.
- 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.
- 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. 1994. Some key features of Distributed Morphology. In *MIT working papers in linguistics* 21, 275–288.
- Hermon, Gabriella. 2009. Language typology and universal grammar: A commentary on the paper by Eric Potsdam. *Natural Language & Linguistic Theory* 27:773–787.
- Hsu, Brian. 2016. Syntax-prosody interactions in the clausal domain: Head movement and coalescence. Doctoral Dissertation, University of Southern California.
- Hsu, Brian. to appear a. Unification of feature scattering and M-merger as Coalescence. In *Proceedings of NELS* 46.
- Hsu, Brian. to appear b. Verb second and its deviations: An argument for feature scattering in the left periphery. *Glossa*.
- 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., and Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8:63–99.
- 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.
- 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. 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.
- 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.
- 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. 2001. Relativized minimality effects. In *The handbook of contemporary syntactic theory*, ed. Mark R. Baltin and Chris Collins. Wiley-Blackwell.
- 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.
- 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.
- Thráninsson, Höskuldur. 1996. On the (non)-universality of functional projections. In *Minimal ideas*, ed. Werner Abraham, Samuel David Epstein, Höskuldur Thráninsson, 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.
- 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.