Extraction and licensing in Toba Batak

Michael Yoshitaka Erlewine, December 2017

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

Keywords: Toba Batak, Austronesian, Austronesian voice, movement, extraction asymmetries, nominal licensing, licensing by adjacency, C and T, head bundling, composite probes

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

Contents

1 Introduction

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

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

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

I will argue for three core conclusions. First, I propose that the heads C and T exist in Toba Batak, with their common division of labor — T responsible for Case-licensing and moving the subject and C responsible for *wh*/focus-fronting — but movement to the clause 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, as in (1b).[2]

[1] See Ramchand and Svenonius 2014 and Wiltschko 2014 (especially chapters 2 and 3) for two recent discussions of the apparent universality of this hierarchical organization within clauses, as well as discussion of their possible sources.

[2] The following abbreviations are used in examples. ACT(IVE) and PASS(IVE) are voice morphemes, described in section 2.1. PN is a marker which precedes proper names, *si*. NEG = negation, AUX = auxiliary, PERF = perfective; PROX = proximal, MED = medial, DIST = distal (see footnote 24); RC = relative clause. 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. For Southern Tiwa: AGR = agreement. For Tongan: ERG = ergative, ABS = absolutive.

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

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

'What did Poltak buy?'

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

'When did Poltak buy a book?'

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

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

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

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

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

Second, I argue that the distinct functions of C and T are sometimes combined on a single head. Following Martinović 2015, 2017, I will call such a head CT.CT probes for targets which are simultaneously *wh* or focused (the requirement of C) and nominal (the requirement of T). This bundled

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

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

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

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:**

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

'Who bought a book?'

CT also inherits from C the ability to probe and attract multiple matching targets. Using CT, multiple nominals can be fronted, as long as they each also satisfy being a *wh* or focus, as in (4) below. Notice that (4) minimally contrasts with 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:**

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

'What did only Poltak buy?'

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

Third and finally, I propose that the CT composite probe in Toba Batak is unable to skip a partially matching goal. The CT head proposed for (4) seeks a goal which is simultaneously nominal and *wh* or focused. This composite probe cannot skip a partially matching goal. Concretely, this disallows the composite probe on CT from skipping the subject and Agreeing with a *wh*/focused nominal lower in the structure. This is one component of my explanation for the basic "subject-only" extraction restriction observed in the language.

This paper is organized as follows. I begin in section 2 with a brief introduction to Toba Batak word order, the voice system, and *wh*- and focus-fronting. Section 3 presents new data on the clause

[4]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* for the proposed organization of the clause periphery.

2 Preliminaries

2.1 Voice and word order

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

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

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

Following previous literature (Van der Tuuk 1864/1971; Nababan 1966, 1981; Percival 1981; Schachter 1984; Cole and Hermon 2008 a.o.), I refer to the prefix *maN-* in (5a) as *active* and *di-* in (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.[6] Postverbal word order will be discussed again in detail in section 4.2.

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

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

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

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

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

If a single DP is fronted, it must be the subject. This is true in the examples in (6) above and is also reflected in the *wh*-fronting contrasts in (7–8) below. Keenan 1972:181–182 and Keenan and Comrie 1977:68–69 famously note that relativization is limited to the subject in Toba Batak, as well as in a number of other Austronesian languages. They describe this as a "subject-only" restriction on extraction.

(7) **Agent *wh*-question ⇒ ACTIVE:**

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

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

'Who ate pork?'

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

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

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

'What did Poltak buy?'

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, in (9), with corresponding changes in postverbal core argument word order.

(9) **Extraction of non-DPs does not interact with voice:**[7]

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

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

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

[7]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 in Clark 1984, 1985:664–665 and Cole and Hermon 2008:150. 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.[8] A similar paradigm 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. In particular, the in-situ *wh*-questions are not echo questions, which cannot be embedded. (11) below

[8]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 advocates for such a cleft analysis for Toba Batak while also invoking *wh*-movement elsewhere; see footnote 42. 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.

shows that all three options in (10) can be embedded under the question-embedding verb *boto* 'know.' The possibility of embedding here shows that *wh*-movement is truly optional.

(11) **Embedded questions with in-situ or moved *wh*-phrases:**[9]

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 seen by the embedded 'when' questions in (12). Both linear positions of *andigan* 'when' in (12b) and (12d) are grammatical. Examples (12a,b) are active, whereas (12c,d) are passive, with corresponding changes to the word order of postverbal core arguments.

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

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

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

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

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

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

[9] Toba Batak has two suppletive passive voice prefixes which encode agent φ-features: first singular *hu-* and first plural inclusive *ta-*.

The picture is similar for focused constituents with particles such as the 'only' particle *holan*: they can be fronted or in-situ, but are very often fronted.[10]

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

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

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

'Only Poltak ate rice.'

It's worth stepping back here and noting that, at this point, we have no clear evidence for the existence of distinct processes of "*wh*-" or "focus-fronting" in Toba Batak. Recall that the language independently allows for the fronting of subjects; see e.g. (6). The facts presented thus far are 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*-phrases and constituents with focus particles as commanding a privileged status in the language. For convenience, I will refer to both *wh*-phrases and constituents modified by *holan* 'only' or *pe* 'even' as FORMALLY FOCUSED, formalized as [+FOC].

[10] DPs with focus particles additionally exhibit a strong preference to be the subject, if possible. This is reflected in the ungrammaticality of the passive variants of (13), where 'only Poltak' is the non-subject passive agent. We will see below that there is not an absolute ban on focus particles on non-subject DPs; focus particles are grammatical on non-subjects when the subject itself has a focus particle or is a *wh*-phrase, as in examples in examples (18–20) and (44) below.

3 Multiple fronting in Toba Batak

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

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

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

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

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

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

'Who ate rice?'

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

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

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

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

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

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

'What did Poltak buy?'

The opposite order — a non-focused DP followed by a *wh* DP — is grammatical as a matrix *wh*-question:

(16) **A hanging topic can precede a fronted *wh*:**

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

≈ 'Poltak, what did he eat?'

However, there are reasons to believe that (16) involves a hanging topic and should be distinguished from true multiple fronting. First, this topic requires a following prosodic break, indicated by #, unlike other preverbal constituents that are studied here. Second, such topics can be resumed by an optional pronoun, unlike other cases of fronting that I discuss here. In what follows, I will disregard such configurations involving hanging topics.

Now consider the examples in (17) below. These examples contrast minimally from the ungrammatical (14c) and (15c) above, in the non-*wh* DP being focused with *holan* 'only,' and are grammatical. Two DP arguments can be simultaneously fronted if they are both formally focused — *wh* or focused with a focus particle.[11]

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

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

Intended: 'Who ate only rice?' (=17b)

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

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

'Who ate only rice?'

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

'What did only Poltak eat?'

(18–20) below show similar examples of multiple fronting with two focused DPs, with supporting contexts which make them felicitous and true. The interpretations of (18) and (19) are unambiguous, reflecting the fact that preverbal quantificational material must be interpreted with surface scope. I have thus far not been able to elicit any multiple *wh*-questions.

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

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

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

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

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

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

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

The degraded status of (i) can be thought of as the result of a so-called semantic "intervention effect" on the interrogative *wh*-word. (Not to be confused with syntactic "intervention" affecting probing, discussed in section 4.5.) Roughly, *wh*-words must be interpreted with widest scope at LF, making the intended interrogative interpretation impossible when certain quantificational material (here, the 'only') intervenes. Note that quantifiers in preverbal positions must be interpreted with surface scope, as we will see in (18–20) below, causing the intervention effect in (i). See Beck 1996, 2006; Grohmann 2006; Mayr 2014; Kotek to appear for empirical and theoretical discussion of such intervention effects.

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

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

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

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

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

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

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

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

Notice that in these examples in (17–20), where two core argument DPs are simultaneously extracted, only one choice of voice morphology is grammatical. The generalization is that the subject DP must be in IMMEDIATELY PREVERBAL position, with the non-subject moved to a position in front of the subject.

To my knowledge, examples of the form in (17–20) have never before been described in Toba Batak or in any other Austronesian language. The possibility of simultaneously extracting multiple DPs has a number of implications for our understanding of Toba Batak syntax and Austronesian syntax more generally. First, contrary to all previous descriptions, the non-subject DP CAN be extracted, though only in this particular, limited configuration. Second, the contrast between (17) and the minimally contrasting examples in (14–15) show us that the grammar privileges formally focused constituents ([+FOC]): *wh*-phrases and constituents with focus particles.

The situation is different still with a DP and a non-DP. Examples (21) and (22) below show that a formally focused non-DP and a non-focused subject DP can be simultaneously fronted to preverbal positions, in that order.[12] The formally focused non-DP in (21) is a PP pied-piping 'who' whereas it is a *holan*-focused temporal adjunct in (22). Example (21) comes from Emmorey 1984, where it is

described in passing as "double fronting" (p. 45), and (22) is a naturally-occurring example from a text, reported in Cumming 1984.

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

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

'To whom did you send the mango?'

(22) **Simultaneous fronting of non-DP focus and subject (Cumming 1984:27):**[13]

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

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

Additional examples of this form are presented in (23). 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.

(23) **Simultaneous fronting of non-DP *wh* and subject 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*)

[12]The opposite order can also be grammatical but involves a hanging topic. See example (16) above and discussion there.

(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?'

[13]Cumming gives *satali* for 'once,' which should be *sahali*.

The possibility of simultaneously extracting one non-DP and one DP in (21–23) 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. (24) below shows that a *wh* DP and a non-focused non-DP cannot be simultaneously fronted, in either order:

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

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

b. *[PP Sian toko buku] ise [man-angko buku ___ ___]?
from store book who ACT-steal book

The evidence presented here shows that multiple fronting is possible in Toba Batak but only in particular configurations, summarized in (25) 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]).

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

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

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

c. ✓[+FOC, −D] [±FOC, +D] V... (21–23)

d. *[+FOC, +D] [−FOC, −D] V... (24a)

e. *[−FOC, −D] [+FOC, +D] V... (24b)

The possible multiple extractions logically fall into two groups: the multiple [+FOC, +D] fronting in (25b) and the combination of a [+FOC] non-DP preceding the subject, both before the verb, in (25c). 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 (25).

4 Proposal

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

In this section, I present my analysis for Toba Batak clause structure which derives the pattern of possible preverbal extractions in the language. I propose that Toba Batak clause structure includes the heads C and T with their familiar division of labor — T probing for a nominal, [+D], and C probing for *wh*/focus, [+FOC] — which can result in multiple extractions. A key additional consideration, I claim, is the limited means of nominal licensing (abstract Case) in the language. Considerations of 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 pattern of a focused non-DP and the subject being fronted simultaneously, in (25c).

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 (25b).

As the characterization above indicates, PROBE-GOAL relationships (Chomsky 2000, 2001) will play a large part in my analysis. Probes trigger the operation AGREE, which seeks a target/goal in the probe's c-command domain with which to establish a link and exchange information. A probe that seeks a target with the feature [F] is itself a formal feature, written as [PROBE:F].[14] I use probing here for Case-licensing nominals and for triggering movement. Probes must Agree with the structurally

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

closest matching target[15]; if the probe can probe multiply, subsequent probing can look past earlier, closer goals.

I furthermore propose that the relevant probes in Toba Batak OPTIONALLY invoke the operation Agree. For example, suppose the head X has the feature [PROBE:F]. Once X is introduced in the structure, the [PROBE:F] feature allows for — but does not require — the corresponding Agree operation to take place. This optionality of probing will be important for deriving the full set of Toba Batak facts. 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 for 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 *v*P. This accords with much previous work on Austronesian voice systems where the choice of voice reflects or entails a particular argument being structurally highest (Guilfoyle, Hung, and Travis 1992; Aldridge 2004; Rackowski and Richards 2005 a.o.). The precise details of voice morphology and voice alternations are orthogonal to the core questions of this paper, so long as this desideratum is met.[16] For concreteness, however, I will now spell out one particular approach.

I assume that the theme argument is generated as the complement of the verb (V) and the functional head *v* introduces the agent as its specifier. I propose following Aldridge 2004, 2008 that the active and passive morphemes correspond to two versions of the head *v* which vary in the presence or absence of a movement-triggering probe.[17] The passive *di-* variant of *v* has a movement-triggering [PROBE:D], which leads to movement of the theme to an outer specifier of *v*P. The movement-triggering [PROBE:D] on *di-* cannot be satisfied by the agent, as the agent is base-

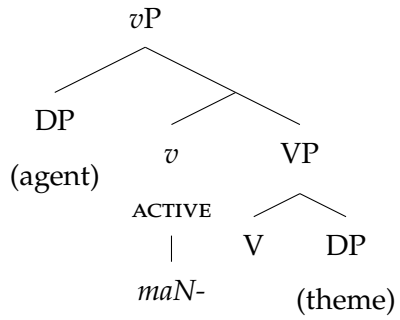
[15]For similar ideas, see e.g. Relativized Minimality (Rizzi 1990, 2001), the Minimal Link Condition (Chomsky 1995, 2000), Shortest Move (Chomsky 1993), and Attract Closest (Pesetsky 2000).

[16]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), and Tagalog (Rackowski 2002; Rackowski and Richards 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 *v*P; see section 4.2. Voice morphology cannot be the result of Agreement with the subject in Case features: both core argument DPs in *v*P would be identical, lacking Case values. In fact, although my analysis for Toba Batak involves the assignment of abstract Case as a means of nominal licensing, I do not propose any differentiated case values (e.g. nominative vs accusative), even abstractly. This accords with the fact that Toba Batak has no morphological case alternations. Agreement in grammatical function (thematic role), as proposed in Sternefeld 1995, would however still be an available analytic option similar in spirit to case agreement.

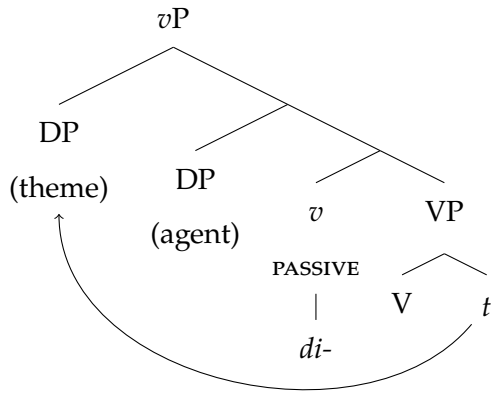
[17]I thank Theodore Levin (p.c.) for discussion here.

generated as a specifier of *v*. In contrast, the active *maN-* variant of *v* triggers no such movement of the theme. The structure of active and passive *v*P are sketched in (26) below. Here I concentrate on transitive clauses, but in the general case I assume all arguments are base-generated within *v*P.

(26) a. Active *v*P:



b. Passive *v*P:



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

With this proposal for the hierarchical structure of *v*P in place, I turn to a brief discussion of word order. I noted above 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 (see Schachter 1984:125, Cole and Hermon 2008:149–152; also footnote 6). I therefore propose that all constituents in *v*P can be scrambled postverbally; i.e., all and only linearizations of *v*P with *v*+V 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 2000:44, Pearson 2000), Tagalog (Kroeger 1991/1993 ch. 5, Richards 1993, Wegmüller 1998, Rackowski 2002:22–27), Tongan (Otsuka 2002, 2005), and Niuean (Clemens 2014 ch. 4). This scram-

bling could be A'-scrambling as in many of these previous works, or a post-syntactic rearrangement at PF (Clemens 2014 ch. 4).[18]

The clausal spine above *v*P can optionally include negation and aspectual/modal auxiliaries, which are free morphemes.[19] If any constituent is fronted to a preverbal position, it necessarily precedes these negative markers and auxiliaries:[20]

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

a. Si Poltak dang olo mang-allang babi.
PN Poltak NEG AUX ACT-eat pork
'Poltak {will not / does not want to} eat pork.'

b. Ibana naeng mang-allang babi.
3sg.ANIMATE AUX ACT-eat pork
'S/he is {going to / about to} eat pork.'

c. Ise nunga mang-allang indahan?
who PERF ACT-eat rice
'Who ate the rice?'

[18] This scrambling may involve additional syntactic structure above *v*P, 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 *v*P.

[19] Aspectual/modal auxiliaries in (27a,b) are glossed AUX here as I cannot yet describe their precise semantics. On *nunga* as in (27c), see Mordechay 1984. See also Percival 1981:86ff on the inventory of auxiliaries.

[20] This is true even in examples of multiple extraction: auxiliaries follow all fronted constituents and are immediately preverbal:

(i) All fronted constituents precede auxiliaries:

a. Andigan {*nunga} si Poltak {nunga} man-uhor buku i? based on (1b/23b)
when *PERF PN Poltak PERF ACT-buy book MED
'When did Poltak buy that book?'

b. Ise {*nunga} holan indahan {nunga} di-allang? based on (17a)
who *PERF only rice PERF PASS-eat
'Who ate only rice?'

This clearly distinguishes multiple extraction in Toba Batak from the so-called *Object Voice* or *Passive Type 2* in Malay and Indonesian (Chung 1976, Sneddon 1996, Cole, Hermon, and Yanti 2008, and many others) and Acehnese (Legate 2014), where both the theme and agent are simultaneously preverbal, in that order, but auxiliaries and negation must precede the preverbal agent (ii–iii). See Legate 2014 chapter 3 for recent discussion of object voice constructions.

I assume that AspP is always projected in the clausal spine, whether Asp is pronounced or not. I will brief discuss aspectual auxiliaries again in section 5.

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

4.2 Nominal licensing

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

My proposal for nominal licensing in Toba Batak comes in three parts. First, I follow Erlewine, Levin, and Van Urk 2015, 2017 in taking a core property of Austronesian-type voice systems to be that subject DPs receive their Case/licensing via Agree with a higher, clause-peripheral probe which can also be involved in their movement. This is the function of T. The head T bears [PROBE:D] which Case-licenses its target and can optionally attract it to its specifier.[21] Based on our derivations for Toba Batak clauses from the previous section, probing by [PROBE:D] on T will necessarily target the subject (active agent or passive theme), as it is the structurally highest DP in *v*P and therefore necessarily the closest matching [+D] target. This is illustrated in (28) below.

(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.'

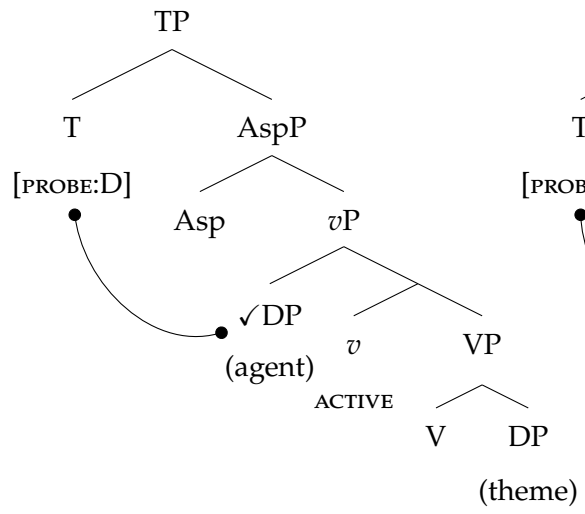
(iii) Auxiliary position in Acehnese object voice (Legate 2014:51):

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

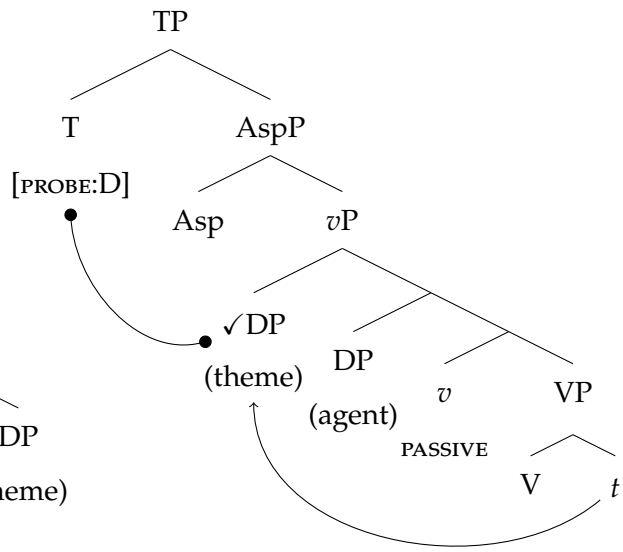
'Meat will be cooked by Fatimah for me in the kitchen.'

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

a. Active:



b. Passive:



I use ✓ to indicate that the subject DPs in (28) have been licensed. Optional fronting of the subject to Spec,TP by [PROBE:D] on T results in sentences with preverbal subjects as in (6) above. [PROBE:D] on T is not able to probe multiply to license an additional DP target; this will become important below.[22]

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 structural Case 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 for the active theme in (28a) or the passive agent in (28b). Without an additional means of nominal licensing, these structures in (28) will crash.

[21] I think of the nominal probe [PROBE:D] and its Case-licensing ability as the essence of the head labeled T, rather than any tense semantics. This is related in spirit to Pesetsky and Torrego's (2001) idea that what we call nominative case is a [T] feature on nominals. I will continue to call this head T in Toba Batak, in order to recognize the parallel to T/Infl in other languages in their association with subjecthood properties including nominative case. As we have seen in (27), Toba Batak has preverbal aspectual auxiliaries but I will argue in section 5.1 that they are not realizations of T.

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

The availability of multiple probing by C must also be subject to variation. See e.g. Hsu 2017 for recent discussion of subtle variations in V2 requirements, including the derivation of some V3 patterns.

This brings us to the third and final component of my proposal for nominal licensing in Toba Batak: NOMINALS CAN BE LICENSED UNDER LINEAR ADJACENCY WITH THE VERB AT PF.[23] The idea that nominals can be made immune to the Case Filter by forming a tight connection with a verb goes back to Baker's (1988, et seq) work on incorporation and is further motivated by similar constructions involving whole noun phrases, called PSEUDO NOUN INCORPORATION (PNI). Recent work on such LICENSING BY ADJACENCY attribute the effect to a PF operation that requires linear adjacency between the nominal and the verb at PF, post-linearization (Baker 2014; Levin 2015). I will briefly review licensing by adjacency effects in Austronesian languages, before discussing the nature of the syntactic operation involved and demonstrating their use in Toba Batak.

Consider first the pair of Tongan examples in (29) below. Nominal arguments in Tongan generally exhibit an ergative/absolutive pattern of case marking. The bolded theme 'cassava' is absolutive-marked in (29a), with the agent *Sione* being ergative-marked. The postverbal word order in (29a) is free. In contrast, the indefinite 'good cassava' in (29b) takes no determiner and no material can intervene between it and the verb. This is PNI. The resulting clause in (29b) has only one Case-marked argument, the agent 'Sione,' which is now absolutive, reflecting the fact that the NP 'good cassava' is exempt from structural Case licensing.

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

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

'Sione planted the cassava.'

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

'Sione planted good cassava.'

Based on the study of the related Niuean language, Massam 2001 claims that PNI objects as in (29b) are NPs that must stay in their base positions as the complement of V, thereby trivially explaining the inability of any material to intervene between the verb and PNI object. But Baker

[23] Clark 1992:59 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 a last-resort alternative to structural Case licensing.

2014 argues that PNI in general exhibits a stricter, HEAD-HEAD ADJACENCY REQUIREMENT. Consider for example the minimal pair of Tongan PNI examples in (30):

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

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

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

‘Sione planted a small amount of cassava.’

While most modifiers are postnominal in Tongan, there are two ‘small’s: a prenominal *ki‘i* and postnominal *iiki*. The contrast in (30) shows that the NP’s head noun *manioke* must be linearly adjacent to the verb; it is not enough for the entire NP to be adjacent to the verb. Contrasts such as (30) and similar effects in Chamorro, Catalan, and Spanish led Baker 2014 to argue that PNI is not simply a NP argument that is the complement of V, contra Massam 2001. Adjacency of the verb with the head of the object (N) is necessary to license PNI objects under adjacency with the verb.

The western Austronesian languages of Malagasy and Balinese help us further refine the notion of adjacency relevant for licensing by adjacency. Erlewine, Levin, and Van Urk 2017 propose that non-subject agents in these languages lack a source for structural Case, and therefore must be licensed by adjacency with the verb. In addition to NP-sized agents, for which Balinese shows restrictions on modifier position parallel to (30) (Levin 2015:76–79), both languages also allow for full definite DPs to be licensed by adjacency to the verb. By comparing patterns of definite DPs in this position in Malagasy and Balinese, Levin 2015 (ch. 4) proposes that the head of the highest extended projection of the nominal (e.g. N or D) must be linearly adjacent to the verb for the nominal to be licensed by adjacency.[24]

[24]I assume that Toba Batak arguments are uniformly DPs. The Toba Batak demonstrative markers *on* (proximal), *i* (medial), and *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 DP is head-initial in Toba Batak, with *si* potentially being an instance of the head D, 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.

Both Baker 2014 and Levin 2015 argue that a syntactic operation applies in such examples to shield the nominal from the Case Filter. For Levin, this is Local Dislocation, a postsyntactic operation that adjoins two linearly adjacent heads; see Embick and Noyer 2001:562ff. Levin proposes that this postsyntactic adjunction allows for the nominal to count as part of the verbal extended projection, obviating its need to be Case-licensed.[25] I refer readers to these works for further discussion of the theory of licensing by adjacency and its empirical motivation.

My proposal is that licensing by adjacency through Local Dislocation is the source of licensing for postverbal active themes and passive agents in Toba Batak. Toba Batak only has one source of structural Case licensing for DP arguments in the clause: [probe:D] on T. In transitive clauses, postverbal non-subject DPs necessarily lack a structural Case licensor. As noted in the previous section, these DPs are not necessarily verb-adjacent according to the structures in (26) and (28), but Toba Batak allows for the free scrambling of postverbal constituents. This allows for the generation of word orders where the non-subject DP is immediately postverbal, allowing for Local Dislocation to apply, licensing the DP.

It is important to note that Local Dislocation applies late at PF, post-linearization (Embick and Noyer 2001, Levin 2015 ch. 4), rather than taking place during the narrow syntax. In particular, it is not possible to invoke Local Dislocation and license the nominal and then move the nominal. We can think of this licensing by adjacency as a last-resort licensing mechanism at PF.

Local Dislocation not only ensures the adjacency of the verb and the licensed nominal, but it also ensures that they will form a tight phonological unit at PF, including possible compound-like phonology between the verb and the adjacent nominal. Emmorey 1984 has shown that the verb and postverbal non-subject DP in Toba Batak form a phonological or intonational unit that is relevant for the the default position of nuclear pitch accent in Toba Batak sentences. This can be thought of as enforced by or a result of Local Dislocation between the verb and postverbal non-

Note also that DPs preceded by the focus particle *holan* can be licensed by adjacency with the verb, even though *holan* seems to intervene between the verb and the highest head of the DP. See (44) below for examples. One solution would be to say that *holan* in such cases itself acts as the head of the highest projection in the nominal extended projection. See e.g. Barbiers 2010 for arguments that focus particles can behave as heads of the constituent that they "modify."

[25] For Baker, the relevant operation is a string-vacuous, covert head-movement, which marks the nominal as part of a complex predicate and thus not subject to the Case Filter. See especially Levin 2015:145–148 for a comparison of these two approaches.

Pearson 2001:46 similarly proposes that a morphological operation that combines linearly adjacent heads applies between verbs and postverbal non-subjects in Malagasy. The operation Pearson refers to, citing Halle and Marantz 1993, is equivalent to Local Dislocation, which is simply Embick's term for morphological merger applying after linearization, over linearly adjacent heads.

subject DP. I refer the interested reader to Emmorey 1984. See also Pearson 2001 for discussion of similar phonological effects on postverbal non-subject agents in Malagasy.

This licensing by adjacency proposal immediately predicts a strict adjacency requirement between the verb and any non-subject DP. As noted above, such an effect has been observed previously by Schachter (1984:125). Consider for example the possible postverbal placement of the temporal adjunct *nantoari* 'yesterday' in (31). 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 (31a) and passive agents as in (31b), reflecting the deep symmetry between the two voices in Toba Batak.

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

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

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

'Poltak read a book yesterday.'

This same adjacency requirement can also be observed with ditransitive verbs. (32) below gives all six postverbal word orders of the subject agent DP, non-subject theme DP, and goal PP of the active verb *manga-lehon* 'give':

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

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

'Poltak gave a book to Uli.'

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

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

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

Only word orders (32a–b) with the non-subject theme DP *buku* 'book' immediately adjacent to the verb are grammatical. Note that the relative order of the subject DP and goal PP are free in (32a–b). The word orders in (32c) with the PP in immediately postverbal position are simply ungrammatical: there are two DPs in the clause but T can only license one, leaving one DP to be licensed by adjacency at PF. The word orders in (32d) are ungrammatical with the intended interpretation, but does have the nonsensical interpretation 'The book gave Poltak to Uli.' The immediately postverbal DP *si Poltak* is necessarily interpreted as the non-subject and therefore a theme in this active clause.

A further prediction of my proposal is that a clause can include at most two DP arguments: one DP can be licensed structurally by T and another can be licensed by linear adjacency with the verb at PF. This prediction is borne out in my speakers' grammars. All grammatical verb frames that I have elicited include at most two DP core arguments, with additional arguments being expressed as obliques such as the prepositional goal of 'give' in (32).[26]

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

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

The adjacency requirement on postverbal non-subject DPs is a key point of discussion in Cole and Hermon 2008, the only contemporary syntactic analysis of Toba Batak clause structure. Cole and Hermon present a VP-fronting analysis for Toba Batak: all arguments in the lower domain of the clause — what they call VoiceP, corresponding to *v*P here — except the passive agent and active theme necessarily evacuate VoiceP, followed by fronting and freezing of the remnant VoiceP. Their analysis is explicitly designed to derive two effects. First, it explains the adjacency requirement observed in (31). Second, it predicts that non-subject DPs cannot ever be extracted, because they

[26] Interestingly, Schachter 1984:136ff describes a productive dative alternation in Toba Batak, resulting in ditransitives with three DP arguments. However, my speakers consistently reject these examples of Schachter's (e.g. Schachter 1984:137 exx 33b, 34b) and different word order combinations thereof. I can only speculate that Schachter's speaker must have had an additional source of Case licensing in these examples, not available to my speakers.

are frozen through the movement of VoiceP — a welcome consequence, according to Cole and Hermon.[27]

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

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

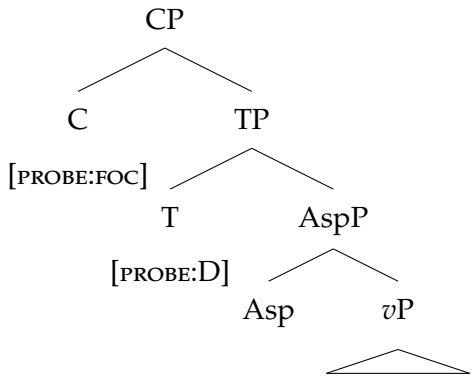
In contrast to Cole and Hermon's approach, under my approach there is no freezing and no absolute ban on the movement of non-subjects. The adjacency of postverbal non-subject DPs is due to their inability to be structurally Case licensed, forcing them to be immediately postverbal for licensing by adjacency. There is no need to stipulate passive agents and active themes as the locus of exceptional behavior: the lack of any structural Case-licensor in vP and the availability of licensing by adjacency together predict that it is precisely these arguments that must be immediately postverbal.

[27] Clark 1992 and Baldridge 2002 also present accounts for Toba Batak that directly tie the inability of non-subjects to be extracted to their adjacency requirement. For Baldridge 2002:179–184, in a flavor of Combinatory Categorial Grammar, both facts are accounted for together by restricting abstraction over non-subject arguments. The account in Clark 1992:59–60 is most similar to my proposal here, deriving these two facts from limitations of case assignment (see footnote 23), predicting as I do that non-subject extraction could become possible if an independent source of licensing becomes available.

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 necessitate a clausal architecture that can front multiple constituents to preverbal positions. For this, I propose that Toba Batak has the two functional heads C and T, as in the syntax of many other language families, with their common division of labor: T probes for a single DP and Case-licenses it, optionally fronting it, while C can probe for formally focused constituents and front them. The proposed structure of the clause periphery is sketched in (33). I will treat both C and T as simply unpronounced in this section but will return to the question of their pronunciation in section 5.

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



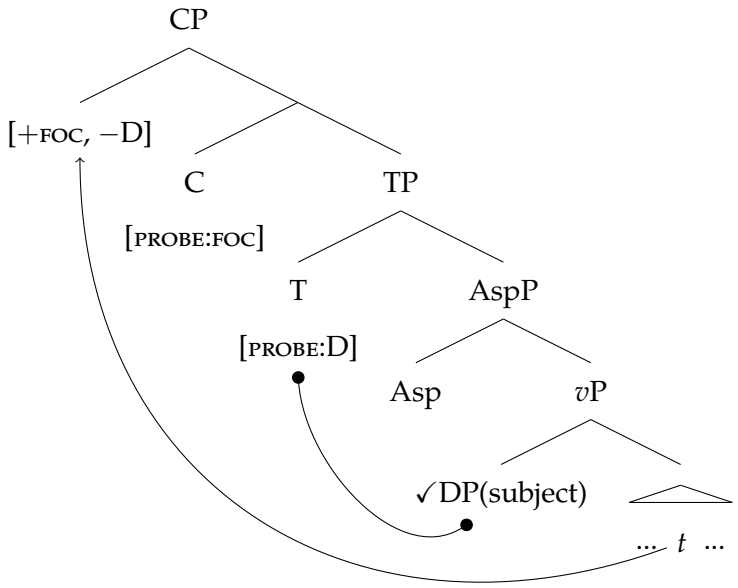
Given this familiar organization of C and T which can both independently attract constituents to their specifiers, why are multiple extractions in Toba Batak so limited? The key, I propose, is the limited means of nominal licensing in the language, as discussed in the previous section. [PROBE:FOC] on 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 [PROBE:D] and [PROBE:FOC], step by step. As discussed in the previous section, [PROBE:D] on T will probe down and find the subject, which is the highest DP in *v*P. This Case-licenses the subject. Now consider the effect of probing by [PROBE:FOC] on C. Assuming that there is a matching [+FOC] target in *v*P, there are two possibilities: either the target is nominal or not.

Consider first the case where [PROBE:FOC] triggers Agree and finds a [+FOC, −D] target: a formally focused non-DP. [PROBE:FOC] then optionally fronts the target that it Agrees with to Spec,CP.[28]

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 [PROBE:FOC] simply targets the highest [+FOC] constituent in *v*P, 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 *v*P). 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):

[28] My proposal is however also compatible with [+FOC] necessarily fronting any target that it successfully Agrees with. Recall that I have proposed that the invocation of Agree by a probe on C or T is itself optional and there is no observable effect of Agree with [PROBE:FOC] for [+FOC] targets which are in-situ.

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

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

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

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

[PROBE:FOC] 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.[29]

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

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

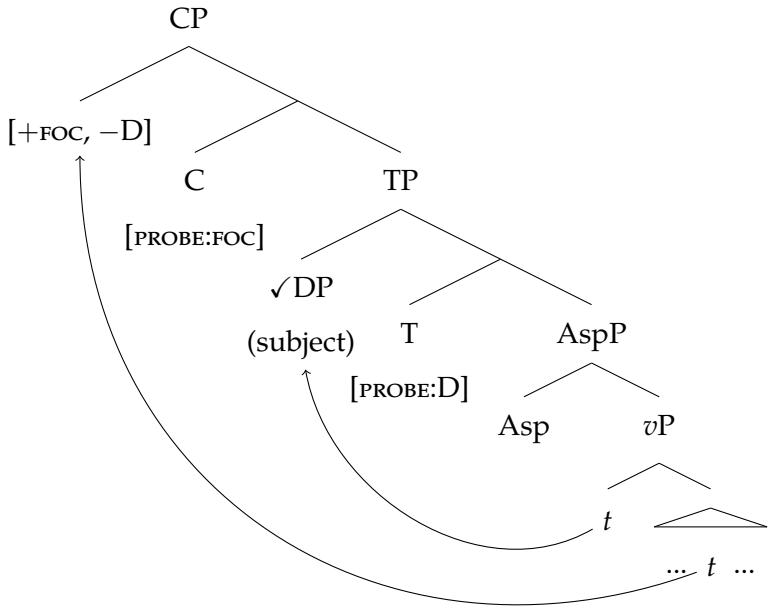
'Why do you only steal books from the BOOK STORE?'
(≈ Why don't you steal books from other places?)

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

Now recall that [PROBE:D] has the ability to optionally front the subject DP that it has Agreed with and Case-licensed. This subject fronting could occur at the same time as the fronting of a [+FOC, −D] constituent to Spec,CP, resulting in the structure in (37). We have now successfully derived the pattern of multiple fronting with a formally focused non-DP followed by the subject, both in preverbal position.

[29] There is a question here in (36) regarding the timing of the two movements. I will propose that multiple movements triggered by the same probe do not "tuck in," *pace* Richards 1997. That is, 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 for supporting evidence.

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



A grammatical example of this pattern of multiple fronting 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 proposal here because the DP fronting is the result of probing by [PROBE:D] on T, which will necessarily target the highest DP in *v*P, the subject.

(38) **Simultaneous fronting of non-DP *wh* and the subject:** (=1b/23b)

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

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

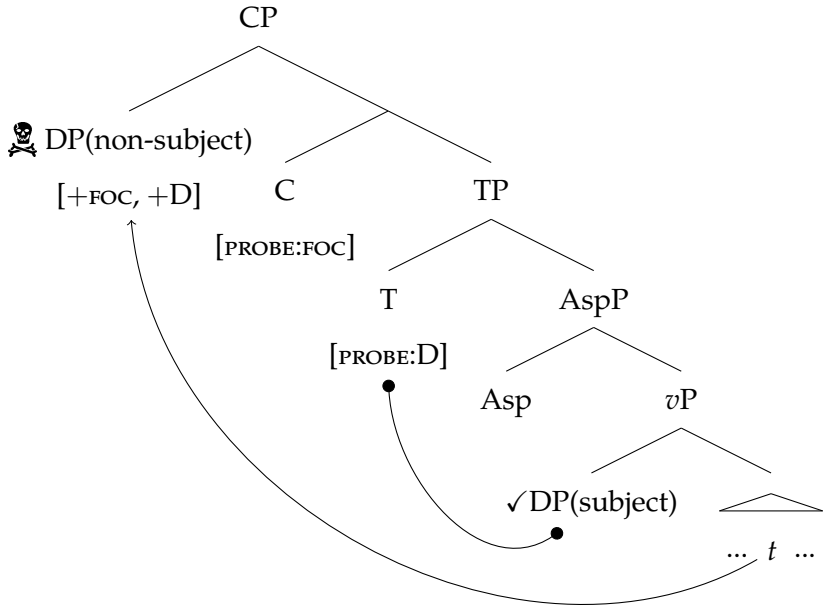
Note that the fronted subject in this configuration could itself also be formally focused, as in the *holan* variant in 38. In the derivation of such examples, [PROBE:D] on T first moves the subject to Spec,TP. Subsequent probing by [PROBE:FOC] on C must skip the subject in Spec,TP, even though the subject has a matching [+FOC] feature, in order to move the focused non-DP to Spec,CP.[30]

[30]There are at least two ways to think about this. One is that the movement of the subject to Spec,TP put it in a criterial position, from which it is unable to move further (Rizzi 2006; Rizzi and Shlonsky 2007), so it is overlooked by higher probing. Another is that there is an anti-locality constraint banning the attraction of the specifier of the probe's complement (Erlewine 2016; Bošković 2015, 2016), leading to the subject in Spec,TP not being visible for probing by [PROBE:FOC] on C (Deal 2016).

From these derivations presented above, we see that Toba Batak syntax is best modeled with two heads in the clause periphery associated with [PROBE:FOC] and [PROBE:D], corresponding neatly to C and T in proposals for clause structure in many other languages. Movement to Spec,TP is driven by [PROBE:D] and is therefore limited to the highest DP in *v*P, i.e. the subject. In contrast, movement to Spec,CP is driven by [PROBE:FOC], which simply attracts the closest [+FOC] constituent in *v*P.[31] Finally, the independence of these movements to Spec,TP and Spec,CP predicts the availability of simultaneously fronting a [+FOC] non-DP to Spec,CP and the subject to Spec,TP as in (38).

These examples so far all involve [PROBE:FOC] attracting a non-DP target. But what happens if probing by [PROBE:FOC] finds a DP? Here the limitations of nominal licensing in Toba Batak rear their head. Suppose that the highest [+FOC] constituent in *v*P is a non-subject DP. (I discuss cases with [+FOC] subjects in the following section.) Let [PROBE:D] on T probe down and Case-license the subject. Next, let [PROBE:FOC] 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 licensor in the clause except for [PROBE:D] on T, but [PROBE:D] here is licensing the subject and [PROBE:D] cannot probe multiply. If the non-subject stayed in postverbal

[31] I presume no Phase Impenetrability effects arising from a *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.

position, it could be licensed by adjacency with the verb, rescuing it from violating the Case Filter at PF. But the fronting of the DP to Spec,CP in (39) bleeds the possibility of licensing by adjacency.[32] 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 the DP would then be unlicensed. 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/15c)

*[$_{CP}$ Aha [$_{TP}$ si Poltak [$_{vP}$ 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 have 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, and form a single, composite probe. 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

[32] The non-subject could be immediately preverbal in (39), but this is not sufficient to license the non-subject, due to the directionality requirement of licensing by adjacency. See footnote 24.

for observed variation — both within and between languages — in the organization of tense, aspect, and mood (Giorgi and Pianesi 1996), tense and agreement (Bobaljik 1995; Thráinsson 1996; Bobaljik and Thráinsson 1998 a.o.), complementizer systems (Bianchi 1999), voice and little v (Pylkkänen 2002, 2008; Harley 2017), and V2 requirements (Hsu 2016a, 2017). In addition to Legate's work, the bundling of C and T features onto a single head has been proposed to account for subject/non-subject extraction asymmetries in Defaka (Ijoid Niger-Congo) (Bennett 2009; Bennett, Akinlabi, and Connell 2012) and Wolof (Atlantic Niger-Congo) (Martinović 2015, 2017); see also Gallego 2017.[33]

I propose that when CT is bundled, [PROBE:FOC] of C and [PROBE:D] of T operate as a single, COMPOSITE PROBE that seeks targets that are simultaneously [+FOC, +D], i.e. matching the specifications of [PROBE:FOC] and [PROBE:D] at the same time. Such composite probing for combinations of features is also discussed in Coon and Bale 2014, Van Urk 2015, and Deal 2015. Probes are additionally associated with specifications such as whether or not they can or must front their targets, whether they are Case-licensing, etc. These specifications for [PROBE:FOC] on C, [PROBE:D] on T, and the composite [PROBE:FOC+D] on CT are given in (41).

(41) **Properties of probes on C, T, and CT:**

	C = [PROBE:FOC]	T = [PROBE:D]	CT = [PROBE:
Case-licenses target?	no	yes	yes
Must front target?	??[34]	no	yes
Can probe multiply?	yes	no	yes

For the composite [PROBE:FOC+D], I propose that it Case-licenses its target (like [PROBE:D]), must front any target it finds, and has the ability to optionally probe multiply (like [PROBE:FOC]). Ultimately, we would like to know whether the properties of a probe are predictable by their featural

[33]The relationship between C and T on the one hand and the bundled head CT could be thought of in one of two ways. One option is for the lexicon to begin with the atomic features/heads C = [PROBE:FOC] and T = [PROBE:D] which then bundle presyntactically into a single CT = [PROBE:FOC+D] head. The possibility of such "presyntactic bundling" is discussed sometimes under the banner of "fusion": see e.g. Matushansky 2006:87 footnote 23 and Coon and Bale 2014:97ff. See also chapters 4 and 5 of Hsu 2016a and also Hsu 2016b for further technical discussion of such bundling operations. Here I will informally describe these heads and their probes as being "bundled."

Alternatively, Giorgi and Pianesi 1996:13–17, 231–232 and Martinović 2015:64–77, 2017 propose that certain features enter the derivation bundled but can then "scatter" or "split," respectively, under certain circumstances. Legate 2011, 2014:83–84 also starts with a bundled head, but relates splitting to the idea of feature inheritance, Chomsky's (2008) proposal that the features of T originate on C and are passed down to T, resulting in the familiar C and T heads. (See also Ouali 2008; Fortuny 2008.) Legate raises the possibility of UNDER-INHERITANCE, where all features stay on C and none are passed down to T, as a means of preserving the bundled head (CT). However, Legate's under-inheritance approach to bundling is strictly speaking not isomorphic to the structures I adopt here, as it predicts the presence of a functionally inert T head below C when under-inheritance takes place. See also Gallego 2014, 2017 for a comparison of feature inheritance with feature bundling/scattering approaches.

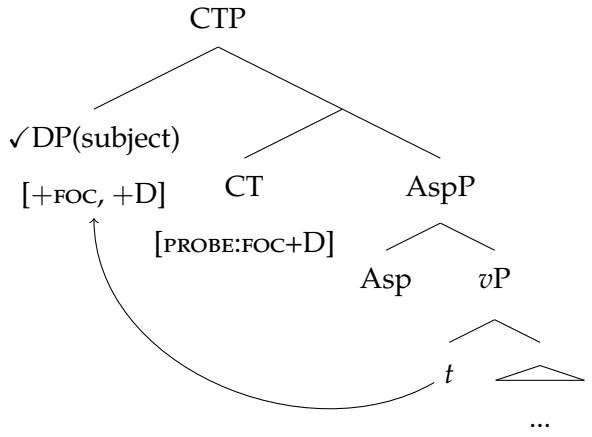
[34]It is not possible to determine whether or not [PROBE:FOC] necessarily fronts targets that it Agrees with. See footnote 28 above.

specification or, in the case of a composite probe on a bundled head, by the properties of their component probes.[35] Here I will leave these issues open for future work and will concentrate on motivating these specifications for these probes in Toba Batak.

I furthermore propose that the bundling of heads C and T necessarily entails the bundling of their probes into the composite [PROBE:FOC+D]. That is, it is not possible for Toba Batak C and T to be bundled into a single head CT with their two distinct probes, [PROBE:FOC] and [PROBE:D], kept separate. While I do not deny the general availability of a single head hosting multiple, distinct probes, evidence will be presented in section 5 that the bundling of C and T in Toba Batak must be one-to-one with the bundling of their probes into [PROBE:FOC+D].

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

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



Like [PROBE:D], [PROBE:FOC+D] Case-licenses its target, licensing the subject DP. (See footnote 35.) This structure corresponds to examples with a single fronted *wh* or focused DP.

Following this first and successful probing in (42), like [PROBE:FOC], [PROBE:FOC+D] has the ability to probe multiply. Suppose [PROBE:FOC+D] probes again. If there is another [+FOC, +D] constituent past the subject — i.e. a formally focused non-subject DP — [PROBE:FOC+D] will Case-license it and front it. I propose that this is the source of the grammatical multiple focused DP extractions, as

[35] For example, there may be a principled reason why [PROBE:D] on T and [PROBE:FOC+D] on bundled CT Case-license their targets, but [PROBE:FOC] on C does not: I tentatively suggest that Case-licensing probes are those that specifically probe for DPs or a subset thereof.

in the examples in (17), repeated here in (43). I will argue below that [PROBE:FOC+D] obligatorily fronts any target it Agrees with.

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

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

'Who ate only rice?'

b. [$_{CTP}$ Aha [holan si Poltak] [$_{vP}$ {$^{\checkmark}$mang/*di}-allang ___ ___]]?
what only PN Poltak {$^{\checkmark}$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 1982:267ff and also footnote 29. That is, these probes do not "tuck in," *pace* Richards 1997. Probing by [PROBE:FOC+D] will first find the [+FOC] subject DP and front it, followed by attraction of the [+FOC] non-subject DP to a higher position.

It's worth noting that this second invocation of Agree by [PROBE:FOC+D] on C is also optional. It is thus grammatical for CT to attract a focused subject but leave another formally focused DP in-situ below. The in-situ DPs will be licensed by adjacency with the verb; see footnote 24 above for relevant discussion. The example in (44) attests to this possibility.

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

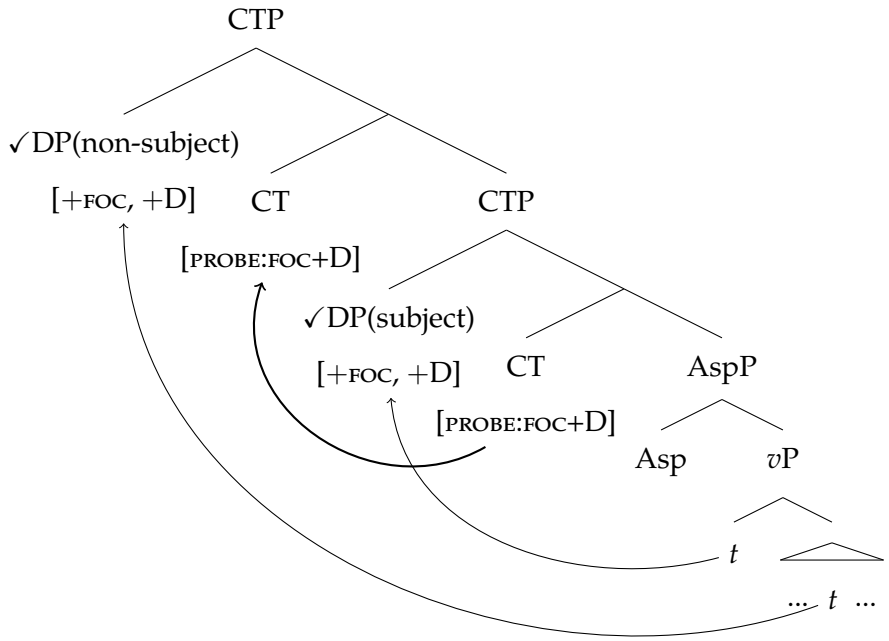
[$_{CTP}$ Ise [$_{vP}$ mang-allang [holan indahan] ___]]?
who ACT-eat only rice

'Who ate only rice?'

When CT attracts multiple targets as in (43) above, I propose that CT hosts these multiple specifiers through REPROJECTION: CT head-moves to merge with its own CTP projection, projecting a

higher CTP.[36] This is illustrated in (45) below, with the thick arrow for reprojection of CT. I will present morphological evidence for this reprojection of the CT head in section 5.

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



I propose that when [PROBE:FOC+D] finds a matching target, it then must move it. This property is crucial for accounting for the ungrammaticality of examples such as (46). Note that the active *mang-allang* version of (46) without *nantoari* was grammatical, in (44).[37]

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

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

Intended: 'Who ate only rice yesterday?'

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

[37] The same contrast between (44) and (46) holds with the thematic roles reversed:

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

'What did only Poltak eat (yesterday)?'

Assume for contradiction that [PROBE:FOC+D] can Agree with a target and leave it in-situ. Consider first the active *mang-allang* variant of (46). [PROBE:FOC+D] finds the [+FOC] subject *ise* 'who,' Case-licenses it, and fronts it. Now notice that, due to the adverb *nantoari* intervening between the non-subject DP and the verb, the non-subject cannot be licensed by adjacency with the verb. However, in this hypothetical derivation [PROBE:FOC+D] could probe again to find the [+FOC] non-subject and Case-license it, without moving it. This would incorrectly predict the availability of example (46) with active *mang-allang*. Next, consider the passive *di-allang* variant of (46). Let [PROBE:FOC+D] on CT probe down, find the [+FOC] subject DP *holan buku* 'only book' and Case-license it, but not move it. Subsequent probing by [PROBE:FOC+D] could then license and front the formally focused non-subject DP. This derivation incorrectly predicts example (46) with passive *di-allang* to be grammatical. The examples in (46) thus show that [PROBE:FOC+D] must attract the targets that it Agrees with. Recall, however, that each invocation of Agree by this probe is optional, allowing for optionality in fronting: see (43) vs (44). What is necessary is to front every target that the probe Agrees with.

4.5 Composite probes and partial matches

One additional configuration that must be discussed is the case where the subject DP is [−FOC] but there is a lower [+FOC, +D] constituent in the clause. If [PROBE:FOC+D] could probe down and Case-license and attract the [+FOC, +D] non-subject, across the [−FOC] subject, we would end up with a structure as in (47) below. We would predict that the in-situ subject could be licensed by adjacency with the verb, incorrectly predicting (47) to be grammatical.

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

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

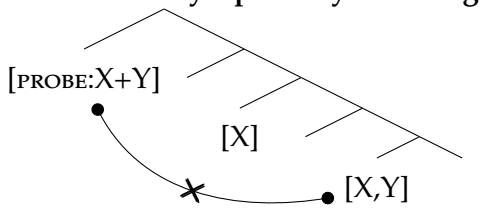
Intended: 'What did Poltak buy?'

What went wrong here? The answer has to do with the behavior of composite probes such as [PROBE:FOC+D]. Although it is true that the [−FOC] subject in (47) is not a valid goal for [PROBE:FOC+D], the subject is a PARTIAL MATCH for the probe. In recent work, Deal 2015 shows convincingly that composite probes may interact with partially-matching goals, even if they do not fully satisfy the

probe. We can imagine different resolutions for what happens after a composite probe encounters a partially-matching goal, and for this to be a point of cross-linguistic variation.

When a composite probe encounters a partial match in Toba Batak, I propose that it then stops its search procedure. Because the partial match does not satisfy the probe, the probe does not Agree with any target. The end result is that probing in this configuration will fail. This is schematized in (48).[38]

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



Concretely, [PROBE:FOC+D] on CT cannot probe past the subject in (47). Probing by CT in (47) will fail, even though there is a matching [+FOC, +D] goal lower in the structure. As there is no other route to grammatically derive the surface representation in (47), the example is judged as ungrammatical. In this way, the proposed inability for a composite probe to probe past a partial match, as in (48), contributes to deriving the basic "subject-only" extraction asymmetry observed in the language.

4.6 Summary

Motivated by the multiple extraction data presented in section 3, in this section I presented a new analysis of voice, case, and extraction in Toba Batak. Key to this analysis is the role of nominal licensing: Toba Batak lacks a structural Case-licensor lower in the clause, in *v*P, but allows for last-resort licensing by adjacency with the verb at PF. 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 in (1), repeated here as (49). The licensing by adjacency of non-subjects also accounts for restrictions on postverbal word order.

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

[38] This configuration of intervention by partial matches has also been discussed extensively in the literature on processing and acquisition. See Friedmann, Belletti, and Rizzi 2009:82–85 and Rizzi 2013:180–182.

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

'What did Poltak buy?'

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

'When did Poltak buy a book?'

The existence of structures such as (49b) 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.

I further proposed that C and T have the option of being BUNDLED into a single head, CT, with a composite [PROBE:FOC+D] that combines properties of C's [PROBE:FOC] and T's [PROBE:D]. Standard properties of probing such as the requirement to Agree with the structurally closest target (footnote 15), together with the inability of composite probes to skip partially matching goals, in (48), derive the famous "subject-only" extraction restriction, which limits $\overline{\text{A}}$-movement of DPs to the subject (Keenan and Comrie 1977 a.o.), while also allowing for the multiple extraction of focused DPs as in (50):

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

[$_{CTP}$ Aha [$_{CTP}$ [holan si Poltak] [$_{vP}$ 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 (50). In the next section, I present overt morphological evidence for these features of the organization of the Toba Batak clause periphery that I have proposed.

5 Spelling out T and the pseudocleft analysis

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

5.1 The distribution of particle *na*

Silitonga 1973:122ff claims that there are two distinct but homophonous free morphemes *na* in the language: a complementizer *na* and a *na* introducing relative clauses. One immediately noticeable difference is that the *na* introducing relative clauses is obligatory (Silitonga 1973:122), as in (51), whereas the *na* introducing embedded clauses is optional, as in (52). *Na* is also obligatory at the beginning of headless relatives as well.

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

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

'I saw the man who is sleeping.'

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

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

'I know that Poltak is sleeping.'

I will first concentrate on the behavior of the optional *na* introducing embedded clauses such as in (52) and return to the *na* in relative clauses in section 5.2 below. The optional *na* at embedded clause edges is described as a complementizer by Silitonga (1973:81) and others. I will however propose that its distribution is better modeled as spelling out the head T or, for one speaker, the bundled head CT. For this, I turn to the distribution of *na* in fronting constructions.

First we observe that *na* appears optionally in fronted *wh*-questions such as in (53), right after the fronted *wh*-phrase. It cannot follow preverbal auxiliaries such as *nunga* in (53) which I take to be in Asp.

(53) ***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 (53). However, there are other configurations where judgments systematically split into two patterns. The symbol % in (54) 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.

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

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

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

'When did Poltak buy a book?'

The consistent pattern of variation here can be straightforwardly captured under my proposal. The key is the organization of the heads C and T. Recall that my derivation for the fronting of a [+FOC] subject DP as in (53) — where all speakers' judgments are in agreement — involves the bundled head CT. See (42) above for the derivation of this structure. The examples in (54) — where two patterns emerge — are those where C and T must be separate heads. In particular, the fronted *wh*-phrases in (54) are [+FOC, −D] so they must have been fronted by C with its [PROBE:FOC] feature, not T's [PROBE:D] or the bundled CT's [PROBE:FOC+D]. See (34, 37) above for the structures of (54a,b), respectively.

Here I adopt Distributed Morphology, a Late Insertion model of morphology (see e.g. Halle and Marantz 1994; Embick and Noyer 2007; Bobaljik 2017). I propose that Pattern A speakers employ *na* as the optional realization of the feature bundle [T] in the context of a local [C] feature whereas the Pattern B speaker uses *na* to spell out the specific feature bundle [C, T]:

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

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

The [T] rule for Pattern A speakers will apply to the bundled head CT as well as to T with a structurally adjacent C.[39] Following (55), all speakers allow for the pronunciation of bundled CT as *na*, explaining the uniform availability of *na* in (53). Only Pattern A speakers allow for the pronunciation of *na* immediately before the verb in (54), because this is the locus of a T head, in the context of a structurally adjacent C. No speaker allows for *na* in between the *wh* non-DP and the non-focused DP in (54b) because this is the position of the unbundled C head, which matches neither of the rules in (55).

Note in particular that the behavior observed in (54a) supports the view that the bundling of the heads C and T in Toba Batak is one-to-one with the bundling of their probes into [probe:foc+D]. Suppose hypothetically that the heads C and T could be bundled into CT, retaining their separate probes [probe:foc] and [probe:D]. We would expect (54a) to be derivable using such a head: [probe:D] Agrees with and licenses the subject in-situ and [probe:foc] Agrees with and attracts the [+foc] *wh* non-DP *andigan* to Spec,CTP. But the behavior of the particle *na* documented here forms an argument against this possibility. The particle *na* is possible in (54a) only for Pattern A speakers, patterning together with example (54b) where the two probes result in their own specifiers, rather than with (53) where the composite probe is used.

The reference to a local [C] feature in the vocabulary insertion rule for Pattern A speakers in (55) blocks the realization of *na* in unembedded declarative clauses. We observe that *na* generally cannot be used to introduce a declarative matrix clause, as in (56). (An apparent exception is discussed in (58) below.) The particle *na* also cannot appear in a matrix clause with a [−foc] subject in preverbal position, with nothing else fronted, as in (57).

[39] The content of structurally adjacent heads has been argued to be relevant for the resolution of contextual allomorphy. See Bonet and Harbour 2012 and Bobaljik 2017 for recent overviews.

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

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

'Poltak is sleeping.' (discourse-initial)

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

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

'Poltak is sleeping.'

I propose that functional features such as [C] and [T] are not included in the clausal spine unless necessary — either for selectional purposes or for the use of their probes.[40] In the matrix clauses in (56) and (57), there is no C projected in the structure so the structural description of the Pattern A rule in (55) is not met. In contrast, [C] is present in clauses with fronting triggered by [PROBE:FOC] such as in (53–54) above as well as for embedded clauses as in (52), where the embedding verb selects for a [C] complement.

An apparent exception to the rule that *na* does not introduce matrix clauses is observed in responses to questions such as *Why?* or *What is someone doing?* as in (58). I tentatively explain such uses as involving a type of presentational embedding, akin to the English *It's that...* or *The answer is that....* Such clauses are therefore formally embedded clauses.[41]

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

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

(i) Na ro ise?
NA come who

'Who came?'

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

'Is Poltak sleeping?'

[41] Alternatively, these clauses may be an instance of what Evans 2007 calls INSUBORDINATION: "a conventionalized main clause use of what, on prima facie grounds, appears to be formally subordinate clauses" (p. 367).

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

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

'What is Poltak doing now?'

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

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

In an embedded clause with a fronted subject, *na* can appear between the fronted subject and the verb for my Pattern A speakers, but no speaker accepts it before the fronted subject; see 59. This too is predicted by the account here, with *na* as a realization of T, and is a strong argument that *na* is not a complementizer, contra Silitonga 1973.

(59) ***na* cannot precede a fronted subject in an embedded clause:**

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

'I know that Poltak is sleeping.'

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

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

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

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

This configuration is precisely where my proposal predicts that the CT head will reproject, resulting in a final syntactic representation with two CT heads. See (45) above. The availability of the particle *na* in both positions simultaneously in (60) supports this CT reprojection account, and is

not predicted under alternative proposals such as the simple use of multiple specifiers on CTP. At the same time, the lack of speaker variation in this judgment reported in (60) further supports the proposal for the realization of *na* in (55) above, which predicts no variation in the realization of bundled CT heads.

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

5.2 Against the pseudocleft analysis

The distribution of *na* above also relates to an alternative, pseudocleft analysis of fronting in Toba Batak. For illustration purposes, I will use an example discussed in Hermon 2009. Hermon proposes that (61) is a *wh*-pseudocleft structure:[42] *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* (51). This pseudocleft parse is sketched in (62). *Op* here represents the null operator involved in relativization.

(61) **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)

(62) 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

[42] Hermon 2009:785 includes a variant of example (61) lacking both *do* and *na* — *Ise mangantuk biang?* — which she describes as derived through fronting of the *wh*-word and not a pseudocleft construction. However, earlier in the paper she suggests that Toba Batak does not have argument *wh*-movement, only pseudoclefts: "The five VP-raising languages reviewed by P [=Potsdam 2009] (Malagasy, Maori, Niuean, Seediq, and Toba Batak) indeed do not allow argument Wh-movement. They stick to Wh-in-situ or clefts" (p. 779).

adopted for apparent *wh* fronting in many Austronesian languages; see Potsdam 2009 for discussion and references.

I argue that Toba Batak does have true fronting. In general, the examples that I have described in this paper cannot be reanalyzed as instances of pseudoclefting. First, note that in all of the fronting examples that I discuss, the particle *na* is optional wherever it is possible; it is never obligatory. This is also true in the presence of the focus enclitic *do* as in (61), which is itself optional. 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 with the use of *na* in relative clauses which is obligatory, casting doubt on a possible reanalysis of these examples as involving relativization.

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

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

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

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

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

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 1972:181–182 and Keenan and Comrie 1977:68–69. 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 Case-licensing — above a lower domain (*v*P) with generally free word order.

I proposed here that nominal licensing is the key to explaining the verb adjacency requirement of postverbal non-subject DPs and the limited patterns of extraction in Toba Batak. 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?* in (2a): the non-subject in Toba Batak will not be licensed, unlike the English accusative *what*.

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

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

References

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

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

Aldridge, Edith. 2008. Phase-based account of extraction in Indonesian. *Lingua* 118:1440–1469.

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

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

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

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

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

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

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

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

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

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

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

Bobaljik, Jonathan David. 2017. Distributed morphology. In *Oxford research encyclopedia of linguistics*, ed. Mark Aronoff. Oxford University Press. URL `http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-131`.

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

Bonet, Eulàlia, and Daniel Harbour. 2012. Contextual allomorphy. In *The morphology and phonology of exponence*, ed. Jochen Trommer, 195–235. Oxford University Press. Manuscript.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Erlewine, Michael Yoshitaka, Theodore Levin, and Coppe van Urk. 2017. Ergativity and Austronesian-type voice systems. In *Oxford Handbook of Ergativity*, ed. Jessica Coon, Diane Mas-

sam, and Lisa deMena Travis, 373–396. Oxford University Press.

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

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

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

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

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

Gallego, Ángel. 2017. Remark on the EPP in labeling theory: Evidence from Romance. *Syntax* 20:384–399.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

Martinović, Martina. 2017. Head-splitting at the Wolof clausal periphery. In *Proceedings of WCCFL 34*, ed. Aaron Kaplan, Abby Kaplan, Miranda K. McCarvel, and Edward J. Rubin, 364–371.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Schachter, Paul. 1984. 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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