

MULTIPLE EXTRACTION AND VOICE IN TOBA BATAK*

Michael Yoshitaka Erlewine
National University of Singapore
mitcho@nus.edu.sg

I investigate \bar{A} -extraction in Toba Batak. Contrary to the claims of previous work on the language (especially Cole and Hermon 2008), I show that multiple, simultaneous extractions to the left periphery is possible, though only in limited configurations. The pattern of possible and impossible multiple extractions motivates a particular organization of the left periphery: specifically, features associated with C and T begin on a single head, probing together, and then splitting if joint probing yields no matching target. I model this using *CT head-splitting* (Martinović 2015). The distribution of the optional particle *na* further supports this approach. Finally, I discuss lessons for the analysis of Austronesian voice and the role of Case.

1. Introduction

Work on comparative formal syntax has identified two positions in the clause periphery, traditionally labeled C and T,¹ which are commonly associated with two very different sets of properties (Chomsky 1986, a.o.). T is commonly associated with properties of subjects, including ϕ -agreement and nominative case assignment, and often requires a nominal specifier, satisfied through \bar{A} -movement (the EPP property). In contrast, the specifier of C is the landing site of \bar{A} -movement and, accordingly, C is the locus of \bar{A} -probes which attract constituents with certain information-structural status, which may or may not be nominal. This division of labor between C and T is remarkably common across language families of the world.

However, in 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 particular argument is privileged with a particular case form and \bar{A} -extraction is limited to this argument, combining properties traditionally associated with C and T (see Erlewine, Levin, and Van Urk 2015, to appear, for an overview). Some authors have in fact proposed that this privileged argument occupies Spec,TP (Guilfoyle, Hung, and Travis 1992, a.o.) while others associate it with Spec,CP (Richards 2000; Pearson 2001, a.o.).

*This project would not be possible without my Batak teachers, Paris Lubis and Richard Siburian. I also thank Hannah Choi, Mary Dalrymple, Hadas Kotek, Theodore Levin, Martina Martinović, David Pesetsky, Nora Samosir, Yosuke Sato, Coppe van Urk, Michelle Yuan, and audiences at AFLA 23 and MIT. Errors are mine.

¹C refers to *complementizer*, associated with clause-typing semantics, and T refers to *tense*, associated with finiteness and temporal interpretation. I simply refer to these heads as C and T here and concentrate on their functions as heads associated with certain syntactic processes.

In this paper, I investigate patterns of \bar{A} -extraction in Toba Batak, spoken in northern Sumatra. My work here is based primarily on elicitation with two speakers currently living in Singapore. The patterns of \bar{A} -extractions available in the language motivate a particular architecture of C and T, which helps resolve the tension inherent to Austronesian voice systems, summarized briefly above. In particular, I propose that the traditional division of labor between C and T is extant in Toba Batak, but this is not immediately visible due to these functions often being combined onto a single head. This can be modeled through a range of *feature inheritance* theories (Chomsky 2008; Ouali 2008; Fortuny 2008; Legate 2011, a.o.), but is most naturally captured under Martinović’s (2015) theory of *CT head-splitting*, wherein C and T begin the derivation as a single head, CT, which splits under certain circumstances.

2. Toba Batak basics

2.1. Voice and word order

Toba Batak has a two-way “voice” system similar to that of neighboring Malayic languages. Consider the examples in (1) below, which are two ways of saying ‘Poltak read the book.’ The marker *si* precedes proper names (PN).

- (1) a. *Man*-jaha**r** buku *si* Poltak. b. *Di*-jaha**r** si Poltak buku.
 ACT-read book PN Poltak PASS-read PN Poltak book
 ‘Poltak read the book.’

The two sentences in (1) differ in the choice of *pivot*—the one argument that commands a privileged status, italicized here. The prefix on the verb (also italicized) reflects the choice of pivot argument. Following previous literature (van der Tuuk 1864/1971; Schachter 1984a; Cole and Hermon 2008), I refer to *maN*- (1a) as ACTIVE and *di*- (1b) as PASSIVE, though I should warn against conflation with Indo-European active/passive alternations. The non-pivot DPs—the ACTIVE theme *buku* in (1a) and the PASSIVE agent *si Poltak* in (1b)—must be adjacent to the verb; I return to this and related facts in section 6.

The canonical declarative order is verb-initial, but pivot-initial clauses as in (2) are common in elicitation. Cumming 1984 reports on a corpus study where one third of declaratives were found to have such a fronted pivot. She describes this fronting as associated with topichood and reports that such fronted topics are “overwhelmingly definite” or generic; I will therefore describe this as topicalization.

- (2) a. *Si* Poltak [*man*-jaha**r** buku _]. b. *Buku* [*di*-jaha**r** si Poltak _].
 PN Poltak ACT-read book book PASS-read PN Poltak
 ‘Poltak read the book.’

If a single DP is \bar{A} -extracted, it must be the pivot. This is true in the topicalization examples in (2) above and is also explicitly reflected in the *wh*-fronting contrasts in (3–4) below:

- (3) *Agent wh-question* ⇒ ACT: (4) *Patient wh-question* ⇒ PASS:
- a. ✓ *Ise* [*mang-allang pinahan-on* ___]? a. **Aha* [*man-uhor* ___ *si Poltak*]?
- who ACT-eat pork-this who ACT-buy PN Poltak
- b. **Ise* [*di-allang* ___ *pinahan-on*]? b. ✓ *Aha* [*di-tuhor* *si Poltak* ___]?
- who PASS-eat pork-this who PASS-buy PN Poltak
- ‘Who ate this pork?’ ‘What did Poltak buy?’

As noted above, \bar{A} -movement being restricted to the one designated pivot argument is familiar from many other Austronesian languages.

Non-DP constituents do not participate in the voice alternation. In stark contrast to DPs, the fronting of non-DPs is independent of the choice of voice. The PP ‘for who’ can be *wh*-fronted (5) out of both active and passive clauses, with corresponding changes in postverbal DP word order.

- (5) *Extraction of non-DPs does not interact with voice:*
- a. ✓ [Tu *ise*] [*man-uhor buku si Poltak*]? (*maN-tuhor* > *manuhor*)
- DAT who ACT-buy book PN Poltak
- b. ✓ [Tu *ise*] [*di-tuhor si Poltak buku*]?
- DAT who PASS-buy PN Poltak book
- ‘[For who] did Poltak buy the book?’

Examples (3–5) here are from my own elicitation work but these same patterns are described in Clark 1984, 1985 and Cole and Hermon 2008. \bar{A} -extraction of DPs is limited to the pivot 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

I will take a moment here to show that *wh*-movement in Toba Batak is optional but preferred, as is the fronting of exhaustive focus with ‘only.’ 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 (6) below show embedded *wh*-questions with and without fronting.

- (6) *Both wh-movement and wh-in-situ are grammatical:*
- a. Hu-boto [*ise* [*mang-allang pinahan* ___]].
- PASS.1sg-know who ACT-eat pork
- b. Hu-boto [*mang-allang pinahan ise*].
- PASS.1sg-know ACT-eat pork who
- c. Hu-boto [*di-allang ise pinahan*].
- PASS.1sg-know PASS-eat who pork
- ‘I know [who ate the pork].’

The embedding in (6) provides evidence that the language truly allows *wh*-in-situ, rather than allowing *wh*-in-situ only in matrix questions through specialized constructions such as echo questions or so-called “declarative syntax questions” (Bobaljik and Wurmbrand 2015), both of which cannot be embedded.

The availability of both movement and in-situ *wh*-questions also extends to non-DP, adjunct *wh*-words as well, as seen by the embedded ‘when’ questions in (7). (Both linear positions of *andigan* ‘when’ in (7b) are grammatical.) The embedded questions in (7) are all ACTIVE, but PASSIVE variants of (7a,b), with corresponding changes in the order of postverbal DPs, are also all grammatical.

- (7) *Wh-movement is optional for adjuncts too:*
- a. Hu-boto [andigan [man-uhor buku ho]].
PASS.1sg-know when ACT-buy book you
 - b. Hu-boto [man-uhor buku {andigan} ho {andigan}].
PASS.1sg-know ACT-buy book when you when
‘I know [when you bought the book].’

Constituents with the exhaustive focus particle ‘only’ *holan* similarly prefer to be fronted, but can also be in-situ:

- (8) *Focus-fronting preferred but both ok:*
- 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 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 topics (see e.g. (2)), which I called *topicalization* above, following Cumming 1984. 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 pivots and non-DPs.

In the next section, I turn to patterns of multiple extraction in Toba Batak. One lesson will be that we must ultimately recognize *wh*/focus-fronting as a distinct process in the language, independent of the free fronting of pivots as in (2). For convenience, I will refer to both *wh*-phrases and constituents modified by *holan* ‘only’ as “formally focused,” formalized as [+FOC].

3. Multiple extractions in Toba Batak

I now investigate the possibility of \bar{A} -extracting multiple constituents simultaneously to the left periphery in Toba Batak. Very little previous work has attempted to investigate such multiple simultaneous fronting. When it comes to DP arguments, the characterization given above and in all previous work on Toba Batak—that only the pivot DP can be fronted—immediately predicts that the fronting of multiple DPs should be impossible. And at first glance, this appears to be correct:²

- | | |
|---|--|
| <p>(9) <i>Wh agent, regular DP patient:</i></p> <p>a. <i>Ise</i> [mang-alang pinahan]?
 who ACT-eat pork</p> <p>b. <i>Pinahan-on</i> [di-allang ise]?
 pork-this PASS-eat who</p> <p>c. *<i>Ise pinahan-on</i> [mang/di-allang]?
 who pork-this ACT/PASS-eat
 ‘Who ate the pork?’</p> | <p>(10) <i>Wh patient, regular DP agent:</i></p> <p>a. <i>Aha</i> [di-tuhor si Poltak]?
 what PASS-buy PN Poltak</p> <p>b. <i>Si Poltak</i> [ma-nuhor aha]?
 PN Poltak ACT-buy what</p> <p>c. *<i>Aha si Poltak</i> [maN/di-tuhor]?
 what PN Poltak ACT/PASS-buy
 ‘What did Poltak buy?’</p> |
|---|--|

Examples (9a,b) are two grammatical forms of the matrix question ‘Who ate the pork?’ As noted above, Toba Batak allows for fronting of the *wh*-word, which must be the pivot (9a), and also allows *wh*-in-situ and free topicalization of definite pivot DPs, resulting in (9b). This topicalization and *wh*-movement cannot cooccur to yield a *wh* DP followed by a topic DP, as observed in (9c).³ The contrast in (10) is completely parallel, but with a referential agent and *wh* patient. Cole and Hermon (2008, p. 183) discuss data such as (9c, 10c) as further support for their view that non-pivot DPs are frozen and cannot move, to be discussed in section 6.

This situation changes, however, if the two DPs in question are a *wh* DP and a DP with the exhaustive focus particle *holan*; in other words, if both are formally

²I do not indicate postverbal gap positions here.

³The opposite order—a topic DP followed by a *wh* DP—is grammatical as a matrix *wh*-question (i). However, there are reasons to believe that (i) is a hanging topic construction that should be distinguished from a true multiple extraction. First, this topic requires a following prosodic break, indicated by # in (i), unlike other preverbal constituents that are studied here. Second, this word order is disallowed in embedded clauses (ii).

- (i) Buku-i *(#) *ise* [man-jahar ___]?
 book-that who ACT-read
 ≈ ‘That book, who read?’
- (ii) *Hu-boto [buku-i (#) *ise* [man-jahar ___]].
 PASS.1sg-know book-that who ACT-read
 Intended: ‘I know [who read that book].’

In what follows, I will disregard such hanging topic constructions.

focused. The examples in (11–12) below show that it is possible to front the *wh* DP followed by the DP with ‘only’ (c), in addition to fronting just the *wh*-word (a) as the pivot or just the DP with ‘only’ (b) as the pivot.⁴ Here too I italicize the pivot DP on grammatical examples, as determined by the choice of voice morphology on the verb.

- (11) *Wh agent, ‘only’ patient:*
- a. *Ise* [mang-allang holan indahan ___]?
who ACT-eat only rice
 - b. *Holan pinahan* [di-allang ise ___]?
only pork PASS-eat who
 - c. *Ise holan pinahan* [{*mang/√di}-allang ___ ___]?
who only pork {*ACT/√PASS}-eat
‘Who ate only rice/pork?’
- (12) *Wh patient, ‘only’ agent:*
- a. *Aha* [di-allang holan si Poltak ___]?
what PASS-eat only PN Poltak
 - b. *Holan si Poltak* [mang-allang aha ___]?
only PN Poltak ACT-eat what
 - c. *Aha holan si Poltak* [√mang/*di]-allang ___ ___]?
what only PN Poltak {√ACT/*PASS}-eat
‘What did only Poltak eat?’

Examples of the form of (11–12) have never before been described. The availability of these multiple extraction variants in (11c) and (12c) has a number of implications for our understanding of Toba Batak syntax. First, contrary to all previous descriptions of Toba Batak, we learn that it is possible to front multiple constituents to the left periphery. Second, the contrast between examples (11–12) where multiple extraction is possible and the earlier examples in (9–10) above shows us that the grammar must distinguish *wh*/focus-fronting from the free fronting of topical, referential constituents, e.g. topicalization as in (2) above.⁵ Third, when multiple DPs are fronted, voice morphology tracks track the choice of DP fronted to *immediately preverbal* position: PASSIVE in (11c) and ACTIVE in (12c). Fourth and finally, non-pivot DPs *can* be moved, contrary to the explicit claims and predictions of Cole and Hermon 2008, which will be discussed further in section 6.

⁴The opposite order, with the *only* DP above the *wh*-phrase, is also ungrammatical, which I take to be for reasons of semantic interpretation; see e.g. Beck 2006. I have thus far not been able to elicit any multiple *wh*-questions.

⁵In particular, this shows that approaches such as in the recent Aldridge to appear, where all extraction is taken to be driven by a [uφ] probe (equivalent to [uD] here), are not rich enough to capture the full pattern of Toba Batak extraction.

The situation is different still with a DP and non-DP. The examples in (13) below show that the simultaneous extraction of a non-DP *wh*-phrase and a non-focused, referential DP, in that order, is grammatical.⁶

- (13) *Simultaneous fronting of non-DP wh and topic DP is grammatical:*
- a. Andigan *buku-i* [{"*maN/√di}-tuhor si Poltak __ __]?
 when book-that {*ACT/√PASS}-buy PN Poltak
 - b. Andigan *si Poltak* [{√maN/*di}-tuhor buku __ __]?
 when PN Poltak {√ACT/*PASS}-buy book
 ‘When did Poltak buy the book?’ (maN-tuhor > manuhor)

The availability of the multiple extractions in (13) is perhaps unsurprising, given that the fronting of non-DPs does not interact with voice, as we saw in (5). However, it’s important to note that it is not simply the case that the simultaneous extraction of any DP and non-DP is grammatical. Examples in (14) show that the combination of a *wh* DP and a referential non-DP is ungrammatical in either order:

- (14) *Simultaneous fronting of wh DP and referential non-DP is ungrammatical:*
- a. *Ise* [man-angko buku [pp sian toko buku] __]?
 who ACT-steal book from store book
 - b. **Ise* [pp sian toko buku] [man-angko buku __ __]?
 who from store book ACT-steal book
 - c. **[pp Sian toko buku] ise* [man-angko buku __ __]?
 from store book who ACT-steal book
 ‘Who stole books from the book store?’

The evidence presented here shows that multiple extractions are possible in Toba Batak but only in a particular, limited set of configurations, summarized in (15) 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,’ [±FOC]).

- (15)a. * [+FOC, +D] [−FOC, +D] V... (9–10)
 b. √ [+FOC, +D] [+FOC, +D] V... (11–12)
 c. √ [+FOC, −D] [−FOC, +D] V... (13)
 d. * [+FOC, +D] [−FOC, −D] V... (14b)
 e. * [−FOC, −D] [+FOC, +D] V... (14c)

Specifically, we observe that *wh*/focus-fronting—a traditional function of C—and the attraction of nominals—traditionally a function of T, the EPP—interact in a nontrivial fashion in Toba Batak. In the next section, I present my proposal which derives this distribution in (15) from a particular understanding of the C-T connection.

⁶The opposite order is grammatical but involves a hanging topic; see footnote 3 above.

4. Proposal

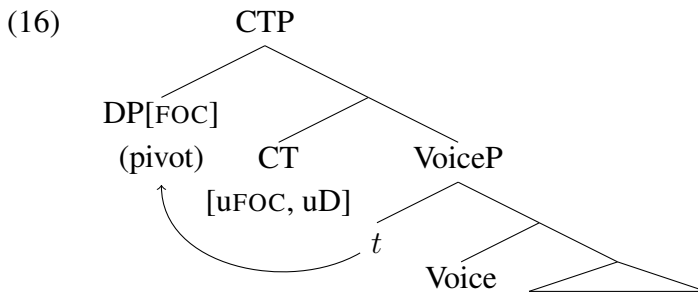
The pattern of grammatical multiple extractions in Toba Batak prompts us to reconsider the relationship between C and T. I propose that the key to the Toba Batak extraction patterns observed is to take the probes associated with C and T—[uFOC] and [uD], respectively—and allow them to first *probe jointly* for a target that simultaneously satisfies both probes (featurally, [+FOC, +D]). If this probing fails to find a target, the probes then probe separately. I assume that such joint probing presupposes that the probes [uFOC] and [uD] originate on the same head.

My work here is not the first to propose that there is a nontrivial relationship between the features and functions of C and T. Work on topics such as subject extraction asymmetries (Pesetsky and Torrego 2001, a.o.) and the morphosyntax of C and T (see e.g. Fortuny 2008 for a review) have all converged on the idea that there must be a tight connection between C and T. One prominent approach to the C-T relationship is the *feature inheritance* hypothesis of Chomsky 2008 which proposes that the features of T such as ϕ -agreement and Case-licensing probes all originate on C and are passed down to T. See also Ouali 2008; Fortuny 2008; Legate 2011 for additional discussion of feature inheritance.

Here I will adopt a recent, alternative conception of the C-T connection which I think most naturally derives the Toba Batak facts. This is the *CT head-splitting* hypothesis of Martinović 2015, which states that the traditional heads C and T start their life as a single head, CT,⁷ but “splitting occurs in cases where a feature cannot be checked” (Martinović 2015, p. 64). This approach is motivated by Martinović’s study of Wolof clause structure and extraction asymmetries.

In order to concentrate here on the left periphery of Toba Batak, I will abstract away from the details of the derivation of basic, verb-initial clauses in Toba Batak. I will, however, assume that the pivot DP occupies a designated position—the specifier of VoiceP—with the Voice head tracking this choice of DP in Spec, VoiceP. I will, however, briefly return to related questions of clause structure in section 6.

I begin by discussing the simple case where we will front a single *wh* or focused DP pivot. The CT head probes jointly for [uFOC, uD] and finds a matching target: the focused pivot DP at the edge of VoiceP. Attracting this DP to Spec, CTP results in the fronting of a single DP, the formally focused pivot.

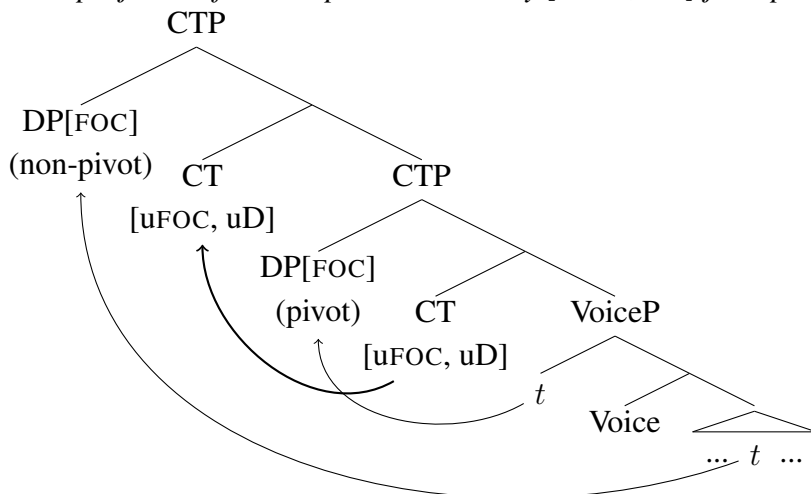


⁷Aldridge 2015 also discusses the application of a joint CT head for Austronesian languages.

I note that unfronted pivot DPs in Spec,VoiceP are postverbal. The tree in (16) is meant to simply illustrate that the pivot argument is hierarchically highest in VoiceP.

Having found a matching target for joint probing by [uFOC, uD] in (16), the CT head has no motivation to split. Following movement of the pivot, CT may probe again for [uFOC, uD]. If it finds another [+FOC, +D] target past the pivot position, it can move it.⁸ In such a case, I propose that CT remerges with its projection and *reprojects* (thick arrow in (17)),⁹ in order to host an additional specifier. I will present evidence for this reprojection of the CT head in section 5.

(17) *CT reprojection for multiple extraction by [uFOC, uD] joint probing:*



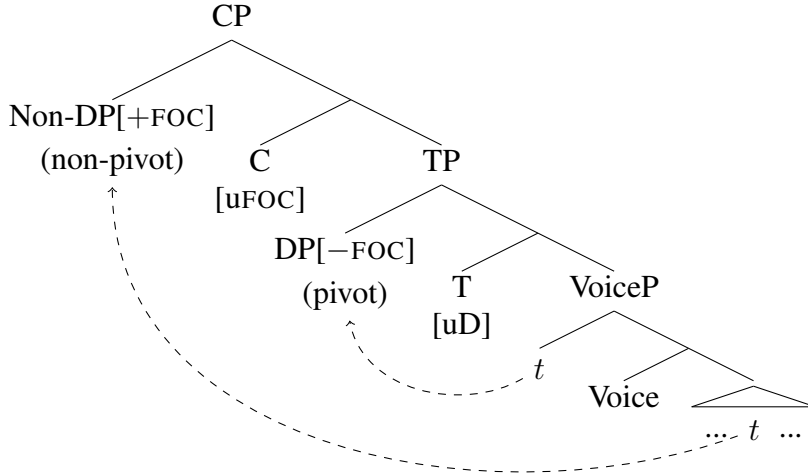
This approach derives the fact that, when two formally focused DPs are fronted as in (11–12), the immediately preverbal DP will be the pivot. The pivot is highest in the VoiceP and therefore will necessarily be the first target moved by CT.

Now consider a case where no DPs in the clause are formally focused. First, CT will probe for [uFOC, uD], but will find no target. It will therefore split into C and T with the traditional division of labor: C is the host of the [uFOC] probe and T is the host of [uD]. This is illustrated in (18) below. Probing by [uD] on T allows for the free fronting of the [–FOC] pivot DP to Spec,TP—what I called topicalization above in (2). Probing by [uFOC] on C can front any [+FOC] constituent to Spec,CP, which in this case will necessarily be a non-DP, as we are considering the case where no DP in the clause is formally focused. This movement to Spec,CP alone yields the fronting of focused non-DPs as in (5).

⁸I assume that all arguments are generated within VoiceP, with VoiceP properly containing the traditional vP. Here I presume no Phase Impenetrability effects arising from the possible phasehood of VoiceP or vP. See Cole, Hermon, and Yanti 2008 §9 for a similar conclusion.

⁹See 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.

(18) *CT split into C and T:*



Each of these movements can apply optionally and independently (optionality indicated by dashed arrows in (18)). When both apply simultaneously, we yield the configuration where a formally focused non-DP precedes a non-focused pivot DP as in (13), the second of our two¹⁰ grammatical multiple extraction configurations (15).

Finally, I consider the case where the pivot DP is [-FOC] but there is a lower [+FOC, +D] constituent in the clause. We begin with CT probing jointly for [uFOC, uD]. Although a matching [+FOC, +D] target is present in the structure, the higher, intervening [+D] pivot will trigger *defective intervention* (originally Chomsky 2000), causing joint probing by [uFOC, uD] on CT to fail. CT will then split, resulting in the same configuration in (18) above.

There is one remaining problem with the approach just outlined. If a lower, non-pivot DP is formally focused, the [uFOC] probe on the split C head (18) could attract the focused DP. When combined with the optional fronting of the non-focused pivot DP to Spec,TP, this alone would predict the multiple extraction of a focused DP followed by a non-focused pivot DP to be grammatical, contrary to fact (9–10). However, this fronting of a non-DP to Spec,CP will fail for principled reasons of Case-licensing, as will be discussed in section 6.

The proposal here yields the correct pattern of grammatical and ungrammatical multiple extractions in Toba Batak, summarized in (15) above. This pattern reflects a sensitivity to both the features [\pm FOC] and [\pm D], with [+FOC, +D] constituents having more extraction possibilities than those that bear [+FOC] or [+D] but not both. This proposal also hints at a new understanding of the relationship between Austronesian voice systems, where \bar{A} -extraction is often limited to the pivot DP, and more familiar systems with distinct C and T functions. In Toba Batak, we see that the default is for CT to jointly probe and attract a formally focused pivot DP, while the split C-T configuration reflects the traditional division of labor between the heads C and T, familiar from the syntax of many other language families.

¹⁰Not counting those involving hanging topics; see footnote 3.

5. Evidence from the particle *na*¹¹

One aspect of \bar{A} -extractions in Toba Batak that I have not yet discussed is the optional particle *na*. This particle often appears preverbally in examples with a single \bar{A} -extraction as in (19a). In cases of long-distance extraction, *na* can appear by the final landing site of movement as well as at the embedded clause edge (19b). This particle *na* is generally included in translations but is never judged to be obligatory.¹²

(19) *The particle na:*

a. *Ise (na) modom?*

who NA sleep

‘Who is sleeping?’

b. *Aha (na) di-dok si Uli [(na) di-allang si Poltak]?*

who NA PASS-say PN Uli NA PASS-eat PN Poltak

‘What did Uli say that Poltak ate?’

(*na...na* ok too)

Both of my speakers agree on the availability of *na* in the cases of single *wh* or focused DP extractions as in (19). However, there are other configurations where judgments systematically diverge. The symbol % in (20) indicates grammaticality for Speaker A but not Speaker B. There is no position where Speaker B accepts *na* but Speaker A does not. Both speakers’ judgment patterns are robust across sessions.

(20) *Configurations with systematic variation:*

a. *Andigan (%na) di-tuhor ho buku-i?*

when NA PASS-buy you book-that

‘When did you buy that book?’

b. *Andigan (*na) buku-i (%na) di-tuhor ho?*

when NA book-that NA PASS-buy you

‘When did you buy that book?’

The consistent pattern of idiolectal variation here can be straightforwardly captured under my proposal. I propose that Speaker A employs *na* as the realization of the feature bundle [T] whereas Speaker B uses *na* to spell out the more specific feature bundle [C, T]. As the unsplit CT head has the categorial features [C, T], both speakers allow *na* in the examples in (19) where the pivot is formally focused and thus CT remains unsplit. In the examples in (20), the pivot is [–FOC] so CT will necessarily split; we then predict that Speaker B will use no *na* whereas Speaker A will allow *na* in the position of the T head. Neither speaker allows for *na* in between

¹¹I thank Martina Martinović for a stimulating conversation which prompted me to revisit my notes on the particle *na*, which led to the discovery presented in this section.

¹²The particle *na* also introduces relative clauses, in which case its presence *is* obligatory. Here I do not discuss relative clauses and leave their detailed investigation for future research.

the *wh* non-DP and the unfocused DP in (20b) because this is the position of the split C head, which matches neither [T] nor [C, T].

Now consider the case of the multiple extraction of two formally focused DPs. Here both speakers allow for the pronunciation of *na* after each DP and in particular allow *na* to be pronounced in both positions simultaneously:

- (21) *The particle na with two wh/focus-fronted DPs:*
 Ise (✓*na*) holan pinahan (✓*na*) di-allang?
 who NA only pork NA PASS-eat
 ‘Who eats only pork?’ (na...na ok too)

This configuration is precisely where my proposal predicts that the CT head will reproject as in (17). The availability of the particle *na* in both positions simultaneously in (21) supports the CT reprojection view presented above, and is not predicted under alternative proposals such as the simple use of multiple specifiers on CTP.

6. The role of Case in Toba Batak

Finally, I turn to the role of abstract Case in Toba Batak. Although nominals in Toba Batak do not bear any case morphology, I argue that there is nonetheless a system of nominal licensing (abstract Case) which plays a crucial role in governing Toba Batak clause structure and word order.

The first motivation for a system of nominal licensing in Toba Batak comes from the following word order restriction. Although postverbal constituents can generally be in any order, the non-pivot DP in a transitive clause must be adjacent to the verb. Example (22) below is reproduced from Schachter 1984a, p. 125.¹³ We see that the adverb *nantoari* ‘yesterday’ can be placed freely, with the exception of the position between the verb and the non-pivot DP:

- (22) *Adding nantoari ‘yesterday’ to (1a,b):*
 a. {Nantoari} mang-ida {*nantoari} si Ria {nantoari} si Torus {nantoari}.
 yesterday ACT-see PN Ria PN Torus
 b. {Nantoari} di-ida {*nantoari} si Torus {nantoari} si Ria {nantoari}.
 yesterday PASS-see PN Torus PN Ria
 ‘Torus saw Ria yesterday.’

Cole and Hermon 2008—the only contemporary syntactic analysis of Toba Batak clause structure—derives this adjacency effect as follows. They propose that all other arguments necessarily move out of the VoiceP constituent, followed by fronting and freezing of the VoiceP. Their analysis is explicitly designed to yield

¹³For what it’s worth, my speakers do not recognize the name *Torus*, suggesting instead that these sentences are about *Sitorus*, with the proper name marker *si* dropped; *si Sitorus* is possible here.

two effects. First, it explains the adjacency requirement observed in (22), assuming that adjuncts such as *nantoari* are necessarily generated outside of VoiceP. Second, it predicts that non-pivot DPs cannot be fronted. Cole and Hermon present data akin to (9–10) above, where a *wh* DP and a referential DP cannot be simultaneously fronted, as support for the latter prediction: extraction of DPs is limited to the pivot, they say, and therefore multiple extraction of two DPs is predicted to be impossible.

I have however shown above that the simultaneous extraction of two co-argument DPs is in fact possible, provided that both DPs are [+FOC]; see (11c) and (12c) above. This teaches us that it is false that non-pivot DPs cannot be fronted and it is also false that non-pivot DPs must necessarily stay verb-adjacent. At the same time, the rejection of the Cole and Hermon approach to Toba Batak syntax means that an alternative explanation for the adjacency facts in (22) must be proposed.

I take the adjacency effect in (22) to be a consequence of a need to Case-license the non-pivot DP through adjacency with the verb. I follow Erlewine, Levin, and Van Urk (2015, to appear) in taking a core property of Austronesian voice systems to be that pivot DPs are Case-licensed by virtue of becoming the pivot. I propose that, in Toba Batak, there is no Case-licensor for DPs internal to the VoiceP. DPs can be licensed by Agreement with the [uD] probe of (C)T or under adjacency with the verb, which can be analyzed as a form of morphological merger or akin to pseudounoun-incorporation.¹⁴ Evidence for this licensing-by-adjacency comes from the fact that the postverbal non-pivot DP forms a phonological unit together with the verb for the purposes of stress assignment, as observed and discussed in Emmorey 1984.

Recall that in the grammatical multiple extractions of DPs in (11c) and (12c), CT never splits into the separate C and T heads. Both fronted DPs have therefore been Agreed with by the [uD] probe on the CT head, in the process of joint probing by [uFOC, uD]. This Agreement with [uD] Case-licenses both DPs. This explains the grammaticality of the multiple extractions in (11c) and (12c), even though the non-pivot DP is not adjacent to the verb.

In contrast, consider the ungrammatical multiple fronting of a formally focused DP followed by a referential pivot DP in (9c) and (10c). In these cases, I claim that CT splits into the traditional C and T heads, with T attracting the referential pivot DP with its [uD] probe and C attracting the non-pivot *wh* DP with its [uFOC] probe. The problem is as follows. The non-pivot in (9c) and (10c) are not Case-licensed in their base position—as they are not adjacent to the verb at PF—nor are they Agreed with by a [uD] probe, as their fronting is due to the [uFOC] probe alone. In this case, the fronted non-pivot DP cannot be Case-licensed, leading to ungrammaticality.

To summarize, even though Toba Batak does not exhibit overt case alternations, nominals must be (abstract Case) licensed. This licensing helps explain the verb-adjacency of post-verbal non-pivot DPs, discussed in both Schachter 1984a and Cole and Hermon 2008, while also allowing for the limited possibility of fronting the non-pivot DP in multiple extractions. The proposal of Cole and Hermon 2008, in contrast, predicts that non-pivot DP extractions can simply never occur.

¹⁴See Levin 2015 and references there on licensing by adjacency.

7. Conclusion

At first glance, Toba Batak exhibits the familiar Austronesian extraction restriction, where \bar{A} -extraction is limited to the pivot DP, whose choice is cross-referenced on the verb. A closer look shows that multiple extractions—and in particular the simultaneous fronting of two DPs—are possible in certain limited configurations. The observed pattern motivates the view that (a) both [uFOC] and [uD] probes exist in Toba Batak and are associated with C and T, respectively, as is common in many non-Austronesian languages, but (b) these two probes prefer to *probe jointly* from a single head. I model this interaction using the CT head-splitting hypothesis of Martinović 2015, together with head reprojection where necessary, and show that the distribution of the particle *na* in two consistent idiolects offers overt morphological evidence for this proposal.

References

- Aldridge, Edith. 2015. Origin of the extraction restriction. LSA 2015 summer institute lecture notes.
- Aldridge, Edith. to appear. Φ -feature competition: A unified approach to the Austronesian extraction restriction. In *Proceedings of CLS 52*.
- Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14:1–56.
- Bobaljik, Jonathan David, and Susi Wurmbrand. 2015. Questions with declarative syntax tell us what about selection? In *50 years later: Reflections on Chomsky's aspects*, 13–31. MIT Working Papers in Linguistics.
- Browning, Margaret A. 1996. CP recursion and *that-t* effects. *Linguistic Inquiry* 27:237–255.
- Chomsky, Noam. 1986. *Barriers*. 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. 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.
- Clark, Robin. 1984. The syntactic nature of Logical Form: Evidence from Toba Batak. In Schachter (1984b), 9–16.
- Clark, Robin. 1985. The syntactic nature of Logical Form: Evidence from Toba Batak. *Linguistic Inquiry* 16:663–669.
- 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.
- Cumming, Susanna. 1984. The syntax and semantics of prepredicate word order in Toba Batak. In Schachter (1984b), 17–36.

- Emmorey, Karen. 1984. The intonation system of Toba Batak. In Schachter (1984b), 37–58.
- Erlewine, Michael Yoshitaka, Theodore Levin, and Coppe van Urk. 2015. What makes a voice system? On the relationship between voice marking and case. In *Proceedings of AFLA 21*, ed. Amber Camp, Yuko Otsuka, Claire Stabile, and Nozomi Tanaka, 51–68. Asia-Pacific Linguistics.
- Erlewine, Michael Yoshitaka, Theodore Levin, and Coppe van Urk. to appear. Ergativity and Austronesian-type voice systems. In *Oxford Handbook of Ergativity*, ed. Jessica Coon, Diane Massam, and Lisa deMena Travis.
- Fortuny, Jordi. 2008. *The emergence of order in syntax*. John Benjamins.
- Georgi, Doreen, and Gereon Müller. 2010. Noun-phrase structure by reprojection. *Syntax* 13:1–36.
- 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.
- 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.
- Legate, Julie Anne. 2011. Under-inheritance. Presented at NELS 42.
- Levin, Theodore. 2015. Licensing without case. Doctoral Dissertation, MIT.
- Martinović, Martina. 2015. Feature geometry and head-splitting: Evidence from the morphosyntax of the Wolof clausal periphery. Doctoral Dissertation, University of Chicago.
- 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.
- Pearson, Matthew. 2001. The clause structure of Malagasy: A minimalist approach. Doctoral Dissertation, University of California at Los Angeles.
- 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.
- Richards, Norvin. 2000. Another look at Tagalog subjects. In *Formal issues in Austronesian linguistics*, ed. Ileana Paul, Vivianne Phillips, and Lisa Travis, 105–116. Springer.
- Schachter, Paul. 1984a. Semantic-role-based syntax in Toba Batak. In Schachter (1984b), 122–149.
- Schachter, Paul, ed. 1984b. *Studies in the structure of Toba Batak*. Number 5 in UCLA Occasional Papers in Linguistics.
- Surányi, Balázs. 2005. Head movement and reprojection. *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae, Sectio Linguistica* 26:313–342.
- van der Tuuk, Herman Neubronner. 1864/1971. *A grammar of Toba Batak*. Springer.
- Watanabe, Akira. 1992. Wh-in-situ, subjacency, and chain formation. *MIT Occasional Papers in Linguistics* 2.