

# Multiple extraction and voice in Toba Batak

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

Austronesian Formal Linguistics Association 23, Tokyo University of Foreign Studies, June 2016<sup>1</sup>

## Introduction

Toba Batak has a Malay/Indonesian-type voice system and is thought to only allow extraction of one DP at a time (Cole and Hermon, 2008).

- ① **Multiple, simultaneous extractions** to the left-periphery—including extraction of multiple DPs—is possible, under certain circumstances.
  - When multiple DPs are fronted, voice morphology tracks the DP moved to **immediately pre-verbal** position.
- ② The pattern of possible multiple extractions motivates a **head-splitting view of the C-T connection** (Martinović, 2015; Aldridge, 2015): CT starts as a single head, but sometimes splits.
  - Different probes associated with C and T, but they **first probe together**.

- Often simply *Hata Batak* 'Batak language'
- Spoken in northern Sumatra, around Lake Toba
- Two million speakers, according to Ethnologue
- Data here from elicitation with two speakers in Singapore



U. Michigan Museum of Anthropology

## 1 Background

### 1.1 Voice in Toba Batak

Toba Batak exhibits a two-way voice alternation, similar to Malay/Indonesian languages: (PN = proper name marker)

- (1) Schachter (1984a, p. 123):
  - a. **Mang-ida** si Ria **si** Torus.  
ACT-see PN Ria PN Torus
  - b. **Di-ida** si Torus **si** Ria.  
PASS-see PN Torus PN Ria  
'Torus saw Ria.'

☞ The **voice prefix** tracks the choice of **pivot** argument (here sentence-final). I refer to *maN-* (16a) as ACTIVE and *di-* (16b) as PASSIVE.

Verb-initial order is the canonical declarative order, but more than one third of declaratives in some texts have a fronted pivot (Cumming, 1984):

- (2) a. **Si Torus** [**mang-ida** si Ria \_\_\_\_].  
PN Torus ACT-see PN Ria
- b. **Si Ria** [**di-ida** si Torus \_\_\_\_].  
PN Ria PASS-see PN Torus  
'Torus saw Ria.'

Cumming (1984) describes this fronting as associated with topichood and reports that such fronted topics are "overwhelmingly definite" or generic.

<sup>1</sup>This project would not be possible without my Batak teachers, Paris Lubis and Richard Simalungun. I also thank Hannah Choi, Hadas Kotek, František Kratochvíl, Theodore Levin, David Pesetsky, Nora Samosir, Yosuke Sato, Coppe van Urk, and the Sihombing family. Errors are mine.

In transitive clauses, the DP argument that is *not* the pivot (Schachter's "internal noun phrase") must be strictly verb-adjacent:

- (3) **Adding *nantoari* 'yesterday' to (16a,b):** (Schachter, 1984a, p. 125)
  - a. (✓Nantoari) mang-ida (\*) si Ria (✓) si Torus (✓).  
ACT-see PN Ria PN Torus
  - b. (✓Nantoari) di-ida (\*) si Torus (✓) si Ria (✓).  
PASS-see PN Torus PN Ria  
'Torus saw Ria yesterday.'

Emmorey (1984) shows that this argument always forms a unit together with the verb for the purposes of nuclear stress assignment.

If a DP is fronted, it must be the pivot:

- (4) **Actor *wh*-question:**
  - a. ✓ *Ise* [**mang-allang** pinahan-on \_\_\_\_]?  
who ACT-eat pork-this
  - b. \* *Ise* [**di-allang** \_\_\_\_ pinahan-on]?  
who PASS-eat pork-this  
'Who ate this pork?'
- (5) **Patient *wh*-question:**
  - a. \* *Aha* [**ma-nuhor** \_\_\_\_ **si** Poltak]?  
who ACT-buy PN Poltak
  - b. ✓ *Aha* [**di-tuhor** si Poltak \_\_\_\_]?  
who PASS-buy PN Poltak  
'What did Poltak buy?'

Fronting of non-DPs does not interact with voice; both voices are possible, with corresponding postverbal word order:

- (6) **Oblique *wh*-question:**
  - a. ✓ [*Tu ise*] [**ma-nuhor** buku **si** Poltak]?  
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?'

(4–6) are my examples but Clark (1984, 1985) and Cole and Hermon (2008) describe the same pattern.

### 1.2 Cole and Hermon (2008)

Based on such facts, Cole and Hermon (2008) argue for a **V(oice)P-fronting** analysis for Toba Batak clauses:

- The non-pivot DP argument, if there is one, stays in-situ in VoiceP;
- All other arguments are moved out of VoiceP;
- VoiceP remnant-moves, freezes;

⇒ The non-pivot DP argument will be adjacent to the verb and cannot subsequently move

Related to more general questions about the derivation of verb-initiality; see also discussion in Chung (2008).

### 1.3 $\bar{A}$ -movements

Two types of  $\bar{A}$ -movements will be relevant here: *wh*-movement and focus movement.

*Wh*-words prefer to front, but can stay in-situ. *Wh*-in-situ is not an echo question, as diagnosed by question embedding:

- (7) **True optional *wh*-movement:**
- Hu-boto [*ise* [mang-allang pinahan]].  
PASS.1sg-know who ACT-eat pork
  - Hu-boto [mang-allang pinahan *ise*].  
PASS.1sg-know ACT-eat pork who
  - Hu-boto [di-allang *ise* pinahan].  
PASS.1sg-know PASS-eat who pork  
'I know [who ate the pork].'
- (8) ***Wh*-movement optional for adjuncts too:**
- Andigan* ma-nuhor buku si Poltak?  
when ACT-buy book PN Poltak
  - Ma-nuhor buku si Poltak *andigan*?  
ACT-buy book PN Poltak when
  - Ma-nuhor buku *andigan* si Poltak?  
ACT-buy book when PN Poltak  
'When did Poltak buy the book?'

(Passive variants of (8a,b,c) all possible, with positions of Poltak and book reversed.)

Only-phrases are also best when fronted:

- (9) **Focus-fronting preferred but both ok:**
- [Holan si Poltak] [mang-allang indahan \_\_\_\_].  
only PN Poltak ACT-eat rice
  - Mang-allang indahan [holan si Poltak].  
ACT-eat rice only PN Poltak  
'Only POLTAK ate rice.'

## 2 Multiple extractions

**Q1:** Can you front two DPs at the same time?

**A1:** At first glance, no.

- (10) ***Wh*-actor, regular DP patient:** 'Who ate the pork?'
- Ise* [mang-alang pinahan \_\_\_\_]?  
who ACT-eat pork
  - Pinahan-on [di-allang *ise* \_\_\_\_]?  
pork-this PASS-eat who
  - \**Ise* pinahan-on [mang/di-allang \_\_\_\_]?  
who pork-this ACT/PASS-eat
- (11) ***Wh*-patient, regular DP actor:** 'What did Poltak buy?'
- Aha* [di-tuhor si Poltak \_\_\_\_]?  
what PASS-buy PN Poltak
  - Si Poltak [ma-nuhor *aha* \_\_\_\_]?  
PN Poltak ACT-buy what
  - \**Aha* si Poltak [maN/di-tuhor \_\_\_\_]?  
what PN Poltak ACT/PASS-buy

Cole and Hermon (2008, p. 183) discuss data such as (10c, 11c) and say this is predicted by their account.

**Q2:** But what if they're both  $\bar{A}$ -operators that prefer to front?

**A2:** They can both be fronted!

- (12) ***Wh*-actor, only patient:** 'Who ate only rice/pork?'
- Ise* [mang-allang holan indahan \_\_\_\_]?  
who ACT-eat only rice
  - Holan pinahan [di-allang *ise* \_\_\_\_]?  
only pork PASS-eat who
  - Ise* holan pinahan [{"mang/'di}-allang \_\_\_\_]?  
who only pork {'ACT/'PASS}-eat
- (13) ***Wh*-patient, only actor:** 'What did only Poltak eat?'
- Aha* [di-allang holan si Poltak \_\_\_\_]?  
what PASS-eat only PN Poltak
  - Holan si P [mang-allang *aha* \_\_\_\_]?  
only PN P ACT-eat what
  - Aha* holan si P [{"mang/'di}-allang \_\_\_\_]?  
what only PN P {'ACT/'PASS}-eat

**Q3:** What about non-DP *whs*? I remember those don't interact with voice.

**A3:** I'm glad you asked!

(14) **Non-DP *wh*, regular DP:**

- Andigan* buku-i [{"maN/'di}-tuhor ho \_\_\_\_]?  
when book-that {'ACT/'PASS}-buy 2sg  
'When did you buy that book?'
- Andigan* si Poltak [{"maN/\*di}-tuhor buku \_\_\_\_]?  
when PN Poltak {'ACT/\*PASS}-buy book  
'When did Poltak buy the book?'

(15) **Summary:**

- \*DP[*wh*] DP V... (10–11)
- ✓DP[*wh*] DP[*only*] V... (12–13)
- ✓Non-DP[*wh*] DP V... (14)

**Lesson 1:** The non-pivot DP (internal noun phrase) *can* be moved, in certain circumstances, contra Cole and Hermon (2008).

**Lesson 2:** Voice tracks the choice of *immediately preverbal* DP.

## 3 Proposal

### 3.1 Voice

Recall that when multiple DPs are extracted, **voice tracks the *immediately preverbal* DP**.

⇒ The pivot DP is fronted first.<sup>2</sup>

☞ The pivot DP is in a designated position (Guilfoyle, Hung, and Travis, 1992, a.o.) at the edge of the lower phase. DP probing from above will find the pivot first.

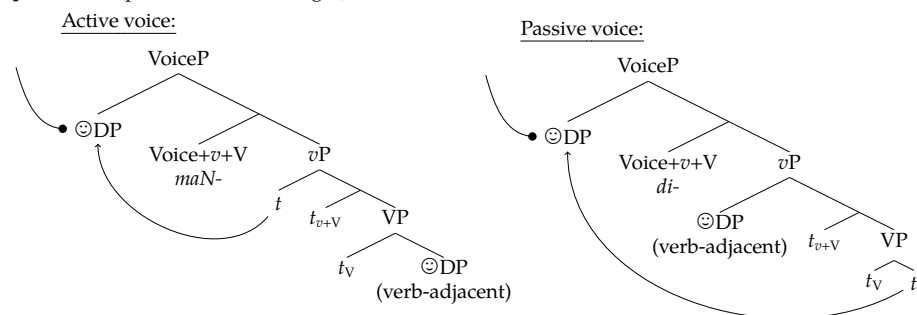
(16) **Working assumptions for voice (Erlwine, Levin, and Van Urk, 2015, to appear, in progress):**

- One DP (the pivot) is attracted to a designated position (but pronounced low or to the right)
  - Voice morphology tracks this choice of pivot.
  - DPs need licensing (abstract Case):
    - the pivot DP must be licensed from above (nominative)
    - one DP (the non-pivot) can be licensed by PF adjacency with the verb (Levin, 2015, and references there)
- ⇒ this is the source of strict verb-adjacency for the non-pivot argument (when postverbal)

The voice details in (16) could conceivably be swapped out for different approaches to voice morphology.

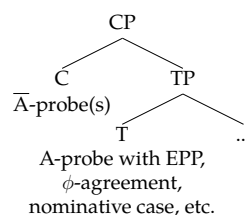
<sup>2</sup>David Pesetsky and Hadas Kotek independently suggest that this generalization could be taken instead to show that the pivot is extracted *last* in multiple extractions; specifically, the lower phase could keep track of each DP extracted out of it, and voice reflects the identity of the *last* DP moved out. I have not (yet) worked out how such an account could be fleshed out for the Batak facts here.

VoiceP is the lower phase; actors are generated in Spec,vP below Voice (pace Legate, 2014). **The pivot is Spec,VoiceP** (pronounced to the right).



### 3.2 Theoretical background: C and T

- (17) **Traditional division of labor:** (Chomsky, 1986, a.o.)
- C:  $\bar{A}$ -movement probe(s)
  - T: A-movement probe, fills Spec,TP with one DP (EPP)



Many languages exhibit an interdependence between C and T (see e.g. Fortuny, 2008 for a review), motivating a tighter connection:

- Feature inheritance:** T features originate on C (Chomsky, 2008; Ouali, 2008; Fortuny, 2008; Legate, 2011, a.o.)
- CT splitting:** C and T begin as a single head, with the option of splitting (Martinović, 2015; Aldridge, 2015)

“the splitting occurs in cases where a feature cannot be checked... or because there is no available position for its goal to move into.”  
Martinović (2015, p. 64)

### 3.3 Proposal

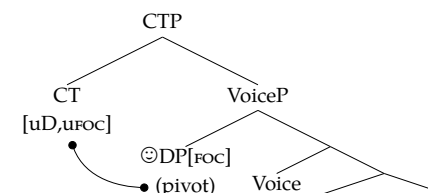
- (18) **Proposal:**
- I adopt CT splitting: **CT starts as one head**
  - C is associated with a probe for *wh*- and *only*-phrases: [uFoc]
  - T is associated with a probe for a DP: [uD]
  - These probes can (Case-)license their agreement targets; subsequent movement is optional
  - CT will first probe to satisfy [uD,uFoc] together<sup>3</sup>; C and T split if no [D,Foc] target is found.**

(Partially matching targets will trigger defective intervention.)

<sup>3</sup>The idea that satisfying multiple probes simultaneously is preferred and must be attempted first is present in, for example, the economy condition of Pesetsky and Torrego (2001).

Foc DP pivot

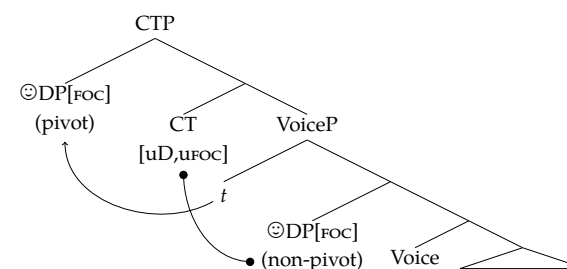
CT probes for [uD,uFoc] together:



Agree; license the pivot; optionally move to preverbal position

Two foc DPs at the edge<sup>4</sup>

CT probes for [uD,uFoc] together *again*:

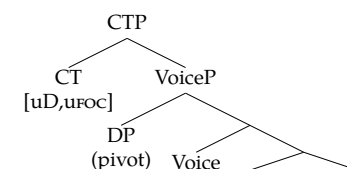


Agree; **license the non-pivot**; move to preverbal position

- ☞ Postverbal non-pivot DPs need verb-adjacency for licensing, but multiple fronting (agreeing with CT) satisfies licensing.

Non-foc DP pivot

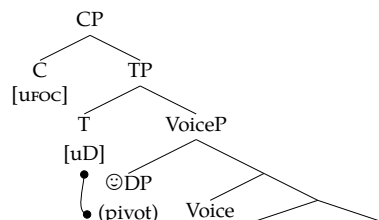
CT probes for [uD,uFoc] together:



- ☞ **If the pivot is not [Foc], CT will not find any [D,Foc] target at the lower phase edge, and must split into C and T.**

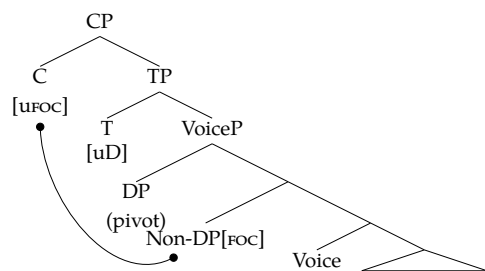
<sup>4</sup>The precise mechanism for optionally fronting additional material must be due to general mechanisms of intermediate movement, such as edge features. Just like other intermediate movement, there must be a restriction that material cannot stay in such positions: by the end of the derivation, VoiceP may only have one unmoved specifier, the pivot DP.

C and T splits; T probes for [uD]:



Agree; license the pivot; optionally move

C probes for [uFOC]:



Agree; move the FOC non-DP

(19) **Summary, based on (15):**

- |                         |  |
|-------------------------|--|
| a. DP V...              | CT splits; T attracts pivot                                    |
| b. DP[FOC] V...         | CT attracts pivot  |
| c. *DP[wh] DP V...      | CT sees non-FOC pivot; CT must split; (defective intervention) |
| d. DP[wh] DP[only] V... | CT attracts pivot; probes again                                |
| e. Non-DP[wh] DP V...   | CT splits; T attracts pivot; C probes                          |

## 4 Conclusion

- ① **Multiple DPs can be simultaneously extracted**, but only if both are formally focused (*wh* or *only*).
  - Motivates *initial joint probing* by [uD] and [uFOC], then separate probing;
  - ☞ In turn motivates a **CT-splitting approach** as in Martinović (2015); Aldridge (2015): **[uD] and [uFOC] must start on the same head**.
- ② **The non-pivot DP can move**, contra Cole and Hermon (2008)
  - Takes away the primary motivation for V(oice)P-fronting;
  - Adjacency facts are better explained by a need for **licensing by adjacency** (Levin, 2015);
  - Voice tracks the pivot, which will be the **first DP attracted** (if any).

Some further directions for study:

- A- and  $\bar{A}$ -properties of these movements
- Multiple non-DP extractions
- Left-dislocated topics, as in Cumming (1984)

## References

- Aldridge, Edith. 2015. Origin of the extraction restriction. LSA 2015 summer institute lecture notes.
- Chomsky, Noam. 1986. *Barriers*. 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. 2008. Indonesian clause structure from an Austronesian perspective. *Lingua* 118:1554–1582.
- Clark, Robin. 1984. The syntactic nature of Logical Form: Evidence from Toba Batak. In 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.
- 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 *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.
- 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. Oxford University Press. URL <http://ling.auf.net/lingbuzz/002629/current.pdf>.
- Fortuny, Jordi. 2008. *The emergence of order in syntax*. John Benjamins.
- 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.
- 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.
- Oualij, Hamid. 2008. On C-to-T feature transfer: the nature of agreement and anti-agreement in Berber. In *Agreement restrictions*, ed. Roberta D'Alessandro.
- Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*, ed. Michael Kenstowicz. MIT Press.
- Schachter, Paul. 1984a. Semantic-role-based syntax in Toba Batak. In Schachter (1984b).
- Schachter, Paul, ed. 1984b. *Studies in the structure of Toba Batak*. Number 5 in UCLA Occasional Papers in Linguistics.