Extraction and licensing in Toba Batak

Michael Yoshitaka ERLEWINE (mitcho)

National University of Singapore
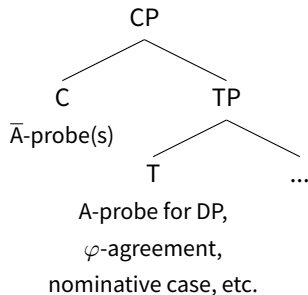mitcho@nus.edu.sg

GLOW 40
March 2017

(1) **Traditional division of labor:** (Chomsky, 1986, a.o.)

    a.   C: $\bar{\text{A}}$-probe(s) for certain XP(s) (focus, *wh*, ...)

    b.   T: A-probe for DP

```
                    CP
                   /  \
                  C    TP
            Ā-probe(s)  / \
                       T   ...
              A-probe for DP,
               φ-agreement,
            nominative case, etc.
```

In many languages, we observe the independent effects of probing by C and probing by T, creating distinct Spec,CP and Spec,TP positions.

(2) **Independent Spec,CP and Spec,TP in English:**

    a.   [$_{TP}$ Stephanie will [be [$_{vP}$ _____ buying the book.

    b.   [$_{CP}$ What will [$_{TP}$ Stephanie [be [$_{vP}$ _____ buying _____?

In contrast, extraction in many Austronesian lgs is regulated by "voice":

(3) **Austronesian voice systems:** (from Erlewine, Levin, and Van Urk, to appear)

    a. <u>A privileged argument:</u> One argument is designated the *subject*, in a particular form and/or structural position.

    b. <u>Articulated voice morphology:</u> Morphology on the verb varies with the choice of subject argument.

    c. **Extraction restriction:** $\bar{\text{A}}$-extraction is limited to the subject argument (Keenan and Comrie, 1977, a.o.).

    d. <u>Marking of non-subject agents:</u> Non-subject agents are restricted in position or morphologically marked.

☞ Voice system languages seem to conflate T-properties (subjecthood) with C-properties (extractability). We don't observe clearly distinct effects of C and T.

- I investigate **patterns of preverbal fronting in Toba Batak**, a predicate-initial Austronesian language.
- Data here is from elicitation with four speakers currently in Singapore.

U. Michigan Museum of Anthropology

C and T in Toba Batak

Toba Batak follows the Austronesian "subject-only" extraction restriction (Silitonga, 1973; Keenan and Comrie, 1977; Cole and Hermon, 2008, a.o.)

☞ I nonetheless argue that **Toba Batak has the distinct heads C and T** with their traditional division of labor (1): C can attract *wh*/focused XPs; T can Case-license and attract DPs.

Patterns of extraction to Spec,CP and Spec,TP are limited, cf English (2):

(4) **Only non-DPs can move to Spec,CP over Spec,TP:**

    a. *[$_{CP}$ Aha [$_{TP}$ si Poltak [man-uhor ___ ___]]]?
          what   PN Poltak ACT-buy
     Intended: '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?'

The role of Case-licensing

Idea 1: The asymmetry in (4) is due to **the lack of Case-licensing** for the DP in Spec,CP. (Toba Batak has no case marking, so this is abstract Case.)

What in English (2b) receives accusative low and then moves to Spec,CP.

(2b) $\checkmark$ [CP What will [TP Stephanie be buying __ACC__ ?

But **Toba Batak has no structural Case-licensor in the lower domain of the clause** (e.g. no accusative). I discuss the Case-licensing of in-situ DPs later.

☞ **The limited means of nominal licensing contributes to the observed Austronesian subject-only extraction asymmetry.**

> **Idea 2:** C and T can be **bundled into a single head, CT**. CT probes for targets that are simultaneously *wh*/focused and nominal, and inherits T's Case-licensing ability.

(5) **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 attract multiple targets:

(6) **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?'

Bundling C and T

CT also inherits from C the ability to attract multiple targets:

(6) **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?'

Note that (6) contrasts minimally from the earlier ungrammatical (4a):

(4a) *[_{CP} Aha   [_{TP} si Poltak [man-uhor ____ ____]]]?
      what   PN Poltak   ACT-buy

   Intended: 'What did Poltak buy?'

Bundling C and T

The availability of multiple DP fronting as in (6) has, to my knowledge, never before been documented in any Austronesian voice system language. It is unpredicted by all previous accounts for Toba Batak clause structure (Clark, 1992; Baldridge, 2002; Cole and Hermon, 2008; Hermon, 2009).

☞ **Joint probing by CT derives this privileged status of being simultaneously *wh*/focused and nominal.**

Roadmap

Voice in Toba Batak

A two-way voice alternation, similar to nearby Malayic languages.

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

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

The **voice prefix** tracks the choice of **subject** argument (bold, here sentence-final). I refer to *maN-* (7a) as ACTIVE and *di-* (7b) as PASSIVE. Note that the "PASSIVE" agent is not demoted or oblique.

Voice in Toba Batak

Predicate-initial order is the canonical declarative order, but subject-initial order (8) is also common.

(8) a. **Si Poltak** [**man**-jaha buku ____].
         PN Poltak ACT-read book

     b. **Buku** [**di**-jaha si Poltak ____].
         book PASS-read PN Poltak
         'Poltak read a book.'

Postverbal word order

☞ Postverbal word order is free with one exception: **the non-subject DP argument must be immediately postverbal** (if any).

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

   a. Manga-lehon <u>buku</u> tu si Uli **si Poltak**.
       ACT-give   book to PN Uli PN Poltak
       'Poltak gave a book to Uli.'

   b. Manga-lehon <u>buku</u> **si Poltak** tu si Uli.

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

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

Postverbal word order

☞ Postverbal word order is free with one exception: **the non-subject DP argument must be immediately postverbal** (if any).

(10) **Adding *nantoari* 'yesterday' to (7a,b):**

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

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

       'Poltak read a book yesterday.'

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

(11) **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?'

☞ This is the famed "subject-only" Austronesian extraction asymmetry
    (Silitonga, 1973; Keenan and Comrie, 1977, a.o.).

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

(12) **Theme *wh*-question** $\Rightarrow$ 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?'

☞ This is the famed "subject-only" Austronesian extraction asymmetry (Silitonga, 1973; Keenan and Comrie, 1977, a.o.).

18

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

(13) **Extraction of non-DPs does not interact with voice:**

    a. $^{\checkmark}$[_PP_ Tu ise] [**man**-uhor buku **si Poltak** _____]?
          for who ACT-buy book PN Poltak

    b. $^{\checkmark}$[_PP_ Tu ise] [**di**-tuhor si Poltak **buku** _____]?
          for who PASS-buy PN Poltak book

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

(7–13) are my examples but the same patterns have been described by Silitonga (1973), Clark (1984, 1985), Schachter (1984a), and Cole and Hermon (2008).

Wh/**focus-movement**

Wh-words prefer to front, but can stay in-situ.

(14) **Optional *wh*-movement: *Ise* 'who'**

    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?'

Adjunct *wh*-movement

Wh-words prefer to front, but can stay in-situ.

(15) **Optional *wh*-movement: *Andigan* 'when'**

    a. Andigan [man-uhor buku ho ____]?
       when   ACT-buy   book you

    b. Man-uhor buku andigan ho?
       ACT-buy   book when   you

    c. Man-uhor buku ho andigan?
       ACT-buy   book you when

       'When did you buy the book?'

Passive variants all possible, with positions of *buku* and *ho* reversed.

Focus-movement

Phrases with *holan* 'only' and *pe* 'even' similarly can be in-situ but prefer to move. (*holan* on handout)

(16) **Optional focus-fronting: *Pe* 'even'**

    a.   [Naniura pe]   [di-allang ahu ___].
          naniura even   PASS-eat 1sg

    b.   Di-allang ahu [naniura pe].
          PASS-eat   1sg   naniura even

    c.   Mang-allang [naniura pe]   ahu.
          ACT-eat   naniura even 1sg

        'I eat even *NANIURA*.'

"*Wh*/focus-movement"?

NB: At this point, there may be no processes of "*wh*/focus-fronting" *per se*. These fronting examples could simply be due to a general process of optional fronting to preverbal position, with the language *wh*/focus-in-situ at its core.

However, I will show in the next section that *wh* and *holan*/*pe*-focused phrases command a privileged status. I refer to them as **"formally focused"** ([+FOC]).

Roadmap

The extraction restriction

As we have seen, Toba Batak exhibits the famed "subject-only" Austronesian extraction restriction: if a DP is extracted, it must be the subject (Silitonga, 1973; Keenan and Comrie, 1977; Cole and Hermon, 2008, a.o.).

☞ **This predicts that multiple DPs cannot be simultaneously fronted.**

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

A1: At first glance, no.

(18)   *Wh* **agent, referential DP theme:**

    a. Ise [mang-allang babi _____]?
        who ACT-eat   pork

    b. Babi [di-allang ise   _____]?
        pork PASS-eat who

    c. *Ise   babi [mang/di-allang ___ ___]?
        who pork ACT/PASS-eat
        'Who ate pork?'

(Cole and Hermon (2008) discuss data such as (18c, 19c) and say this is predicted by their account.)

Multiple extractions: [+FOC, +D] [−FOC, +D]

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

A1: At first glance, no.

(19) *Wh* theme, referential DP agent:

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

    b. Si Poltak [man-uhor aha _____]?
       PN Poltak ACT-buy what

    c. *Aha si Poltak [maN/di-tuhor ___ ___]?   (=4a)
       what PN Poltak ACT/PASS-buy
       'What did Poltak buy?'

(Cole and Hermon (2008) discuss data such as (18c, 19c) as support for their account.)

Multiple extractions: [+FOC, +D] [+FOC, +D]

Q2: But what if they're both formally focused and prefer to front?

A2: They can both be fronted!

(20) ***Wh* agent, *only* theme:**

Ise   [holan babi] [{*mang/$^\checkmark$di}-allang ___ ___]?
who   only   pork   {*ACT/$^\checkmark$PASS}-eat

'Who ate only pork?'   cf (18c)

☞ When multiple DPs are fronted, **the subject must be immediately preverbal**.

Multiple extractions: [+FOC, +D] [+FOC, +D]

Q2: But what if they're both formally focused and prefer to front?

A2: They can both be fronted!

(21) *Wh* **theme,** *only* **agent:**

Aha   [holan si Poltak] [{$^\checkmark$mang/*di}-allang ___ ___]?

what only   PN Poltak {$^\checkmark$ACT/*PASS}-eat

'What did only Poltak eat?'   (=6), cf (19c)

☞ When multiple DPs are fronted, **the subject must be immediately preverbal**.

Q2: But what if they're both formally focused?

A2: They can both be fronted!

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

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

'Only infants only drink milk.'

☞ When multiple DPs are fronted, **the subject must be immediately preverbal**.

Multiple extractions: [+FOC, −D] [±FOC, +D]

Q3: Recall that non-DP extraction doesn't interact with voice. Can a non-DP be moved with a DP?

A3: Good question. It depends…

(23) ***Wh*-non-DP, (focused) subject DP:**

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

    b. Andigan [(holan) si Poltak] [{✓maN/*di}-tuhor buku ____ ____]?
        when   only   PN Poltak   {✓ACT/*PASS}-buy   book
        'When did (only) Poltak buy the book?'   (=4b)

Multiple extractions: [+FOC, +D] [−FOC, −D]

But it's not simply that any DP and non-DP can be simultaneously fronted:

(24) **A *wh*-DP and referential non-DP:**

    a.   Ise   [man-angko buku [$_{PP}$ sian toko buku] _____]?
          who   ACT-steal   book   from store book
          'Who stole books from the book store?'

    b.   *Ise   [$_{PP}$ sian toko buku] [man-angko buku ___ ___]?
          who   from store book   ACT-steal   book

Finally, multiple [+FOC] non-DPs can be simultaneously fronted:

(25)   **Multiple fronting of two [+FOC, −D] targets:**
    Boasa [holan [PP sian   toko   buku]] [man-angko buku ho ____ ____]]?
    why   only   from store book   ACT-steal   book 2sg
    'Why do you only steal books from the BOOK STORE?'
    (I.e.: Why don't you steal books from other places?)

Summary

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

    a. *[+FOC, +D] [−FOC, +D] V...   (18–19)

    b. $^\checkmark$[+FOC, +D] [+FOC, +D] V...   (20–22)

    c. $^\checkmark$[+FOC, −D] [±FOC, +D] V...   (23)

    d. *[+FOC, +D] [−FOC, −D] V...   (24b)

    e. $^\checkmark$[+FOC, −D] [+FOC, −D] V...   (25)

☞ **The non-subject DP *can* be moved**, contra Cole and Hermon (2008) a.o., but only in multiple focused DP extraction (26b).

- Various multiple extractions with non-DPs are possible (26c,e), but still not entirely free (26d).

Roadmap

Clause structure and voice

(27) **Working assumptions:**

    a.   ACTIVE/PASSIVE is in Voice, above vP; the verb head-moves to Voice

    b.   Voice correlates with/determines the **highest DP** in VoiceP:

        • ACTIVE $\iff$ agent highest

        • PASSIVE $\iff$ patient highest

```
        AspP
        /  \
     Asp    VoiceP
    (AUX)   /    \
        Voice     vP
     ACT/PASS-V   /\
```
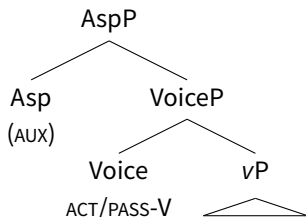
Property (27b) is (roughly) shared with many previous works on Austronesian voice, including Guilfoyle et al. (1992); Aldridge (2004); Rackowski and Richards (2005).

(27)   **Working assumptions:**

   c.   Constituents in VoiceP are subject to **scrambling**. All
        linearizations with the verbal complex leftmost can be
        generated.

Scrambling of postverbal constituents (27c) has been independently
proposed for many Austronesian languages, including Malagasy (Paul, 2000;
Pearson, 2000), Tagalog (Kroeger, 1991/1993; Richards, 1993; Wegmüller, 1998;
Rackowski, 2002), Tongan (Otsuka, 2002, 2005), and Niuean (Clemens, 2014).

Nominal licensing

Toba Batak lacks case morphology, but nonetheless **nominals must be licensed** by PF.

☞ There is **no structural Case licensor within VoiceP**.

- The subject is (Case-)licensed by T via Agree.
- One DP can be **licensed by adjacency**, which allows for postverbal PASSIVE agents and ACTIVE themes.

Licensing by adjacency

The idea of licensing by adjacency goes back to Baker (1988, et seq).

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

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

    b. Am-**seuan**-wan-ban   hliawrade.
        AGR$_{3sg}$-man-come-PAST lady
        'The men came to the lady.'

Baker (1988, 106ff): 'Come' is unaccusative, so there is only one Case-licensor: T. T Case-licenses with the 'men' in (28a). The 'lady' instead must be an oblique. In (28b), the 'man' incorporates into the verb, freeing T to license the 'lady.'

Massam (2001) observes a similar alternation with NPs:

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

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

    b.   Kua tō   **huli   talo** e   magafaoa he māla.
        PERF plant shoot taro ABS family   LOC farm
        'The family planted taro shoots at the farm.'

Massam (2001): Limited to NPs (not DPs): no pronouns or proper names; no possessors, number markers, or indicative relative clauses. Limited to direct objects and instrumentals.

Licensing by adjacency

There are also DPs (full nominals) which are licensed by adjacency: non-subject agents in Malagasy and Balinese (see Levin, 2015).

(30) **Malagasy non-subject agents must be immediately postverbal; can be proper names, definite (Pearson, 2005, 391ff):**

    a. Vonoin-dRamatoa amin'ny antsy ny akoho.
        PASS.kill-Ramatoa with-Det knife Det chicken
        'Ramatoa is killing the chickens with the knife.'

    b. Nohanin' {*haingana} ny gidro {haingana} ny voankazo.
        PASS.eat   {*quickly}   Det lemur {quickly}   Det fruit
        'The lemur ate the fruit quickly.'

Erlewine et al. (2015, to appear): In many Austronesian voice languages, non-subject agents lack structural Case licensing. In some languages, these arguments are licensed through PF adjacency with the verb.

Licensing by adjacency

☞ **Nominals can be licensed (or waived of the licensing requirement) by PF adjacency with the verb.**

I abstract away from the precise mechanism underlying licensing by adjacency here, but see Levin (2015) for one proposal. Languages may then vary in the size of adjacency-licensed nominals (N, NP, DP) and possible base positions.

The limited means of nominal licensing explains the word order restrictions observed: The subject will be licensed by T. The non-subject DP will be licensed by adjacency...if immediately postverbal.

(31) **Nominal licensing in Toba Batak:**

Man-jaha 😘 [DP buku] 😀 [DP si Poltak].
ACT-read   book   PN Poltak

'Poltak read a book.'

Licensing by adjacency in Toba Batak

The limited means of nominal licensing explains the word order restrictions observed: The subject will be licensed by T. The non-subject DP will be licensed by adjacency…if immediately postverbal.

(31)   **Nominal licensing in Toba Batak:**

Man-jaha (*nantoari) 😢[DP buku] 😀[DP si Poltak].
ACT-read   *yesterday   book   PN Poltak

'Poltak read a book.'

Licensing by adjacency in Toba Batak

We also explain an asymmetry between *holan* 'only' and *pe* 'even':

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

    a. Mang-allang [DP naniura] pe] [DP ahu].
        ACT-eat   naniura even   1sg

        'I eat even *NANIURA*.'   (=16c)

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

        'Only POLTAK ate rice.'   (=17c)

☞ **C and T exist in Toba Batak** with C triggering *wh*/focus-fronting and T associated with subject Case-licensing/fronting.

(Assume C and T are unpronounced for now.)

(33) **The content of C and T:**

	C = [uFOC]	T = [uD]
Case-licenses target?	no	yes
Must front target?	yes	no
Can probe multiply?	yes	no

Invocations of these probes are optional.

Assume first that the subject is [−FOC]:

(34) **[uD] on T probes, Case-licenses subject:**

The subject in Spec,VoiceP is necessarily the closest DP for [uD] probing from above. [uD] optionally fronts the subject.

Assume first that the subject is [−FOC]:

(34) **[uD] on T probes, Case-licenses subject:**
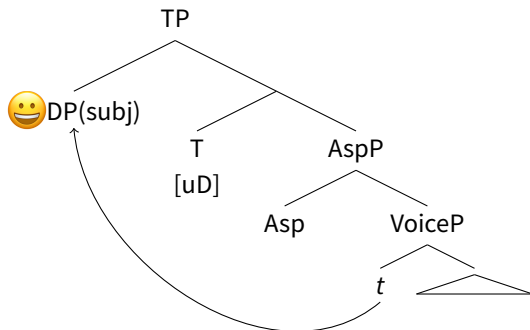
The subject in Spec,VoiceP is necessarily the closest DP for [uD] probing from above. [uD] optionally fronts the subject.
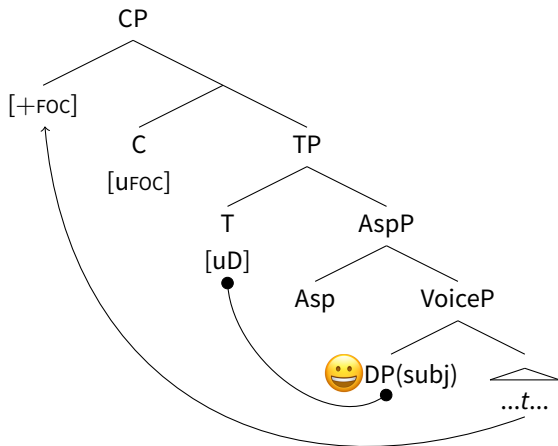
(35) [uꜰᴏᴄ] **on C probes, attracts target:**

(35) [ᴜꜰᴏᴄ] **on C probes, attracts target:**
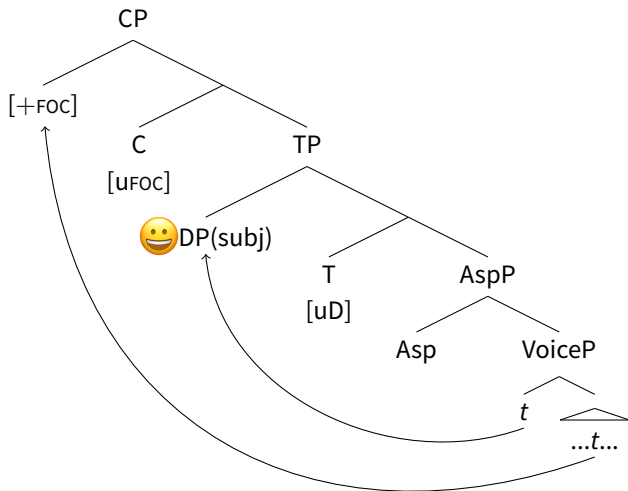
☞ **Anything fronted by [uFOC] to Spec,CP must be [−D].**
C does not Case-license. A non-subject DP could be licensed
postverbally by adjacency to the verb, but fronting to Spec,CP would
break this adjacency.

This explains the contrast betwen DPs and non-DPs in Spec,CP (4), repeated here:

(4) **Only non-DPs can move to Spec,CP over Spec,TP:**

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

       Intended: 'What did Poltak buy?'   (=19c); pattern (26a)

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

       'When did Poltak buy a book?'   pattern (26c)

Bundled CT

☞ **C and T can be bundled.**

- Bundling of heads accounts for variation in 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), causatives (Pylkkänen, 2002, 2008), V2 requirements (Hsu, 2016, to appear)...
- The CT bundling here is most similar to Legate's (2011) and Martinović's (2015) proposals for subject/non-subject extraction asymmetries.

Their probes get bundled too: [**uFOC+D**] **probes for targets that are simultaneously** [+**FOC**, +**D**].

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

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

(37) **Fronting a [+FOC] subject with CT:**
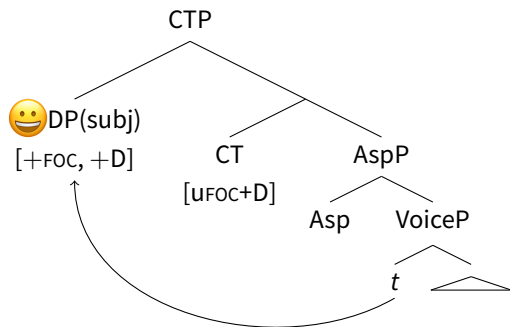
CT inherits Case-licensing ability of [uD]. (Probing by bundled CT is incompatible with clauses with [−FOC] subjects: the intervening [+D] subject will trigger defective intervention for [uFOC+D].)

(38) **Probing multiply with CT:**
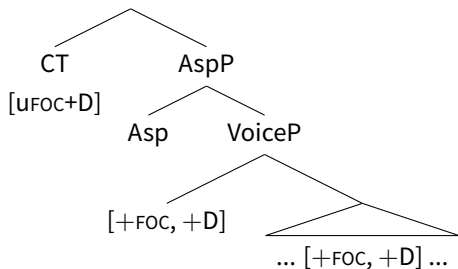
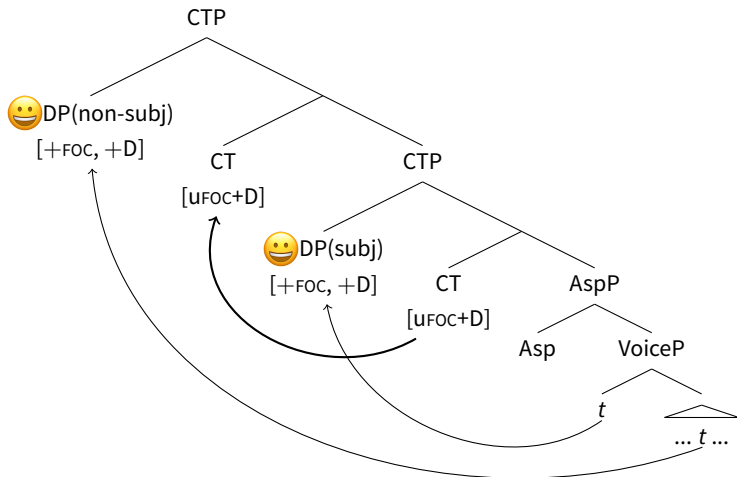When attracting multiple targets, **CT reprojects**.

(38) **Probing multiply with CT:**
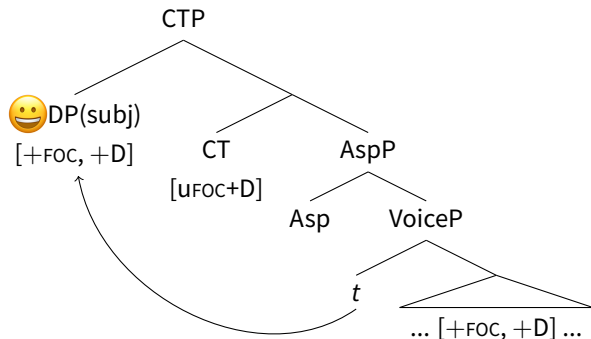
When attracting multiple targets, **CT reprojects**.

Bundled CT

(38) **Probing multiply with CT:**

When attracting multiple targets, **CT reprojects**.

Summary

☞ The traditional organization of C and T — together with a consideration of nominal licensing and the ability to bundle C and T — derives the patterns of grammatical extractions in the language.

(39) **Summary, based on (26):**

a.   *[+FOC, +D] [−FOC, +D] V...   C doesn't Case-license;
                                           CT would hit defective intervention

b.   ✓[+FOC, +D] [+FOC, +D] V...   multiple attraction by CT

c.   ✓[+FOC, −D] [±FOC, +D] V...   attraction by C and T

d.   *[+FOC, +D] [−FOC, −D] V...   C and T are ordered;
                                                T can't attract [−D]

e.   ✓[+FOC, −D] [+FOC, −D] V...   multiple attraction by C

We also derive that, when multiple DPs are fronted, the subject must be in immediately preverbal position.

Roadmap

Supporting evidence comes from **a ubiquitous, optional particle *na***.

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

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

'I know that Poltak is sleeping.'

Due to examples such as (40), *na* has been called a complementizer (see e.g. Silitonga, 1973). *Na* is also used obligatorily for introducing relative clauses, which I do not discuss here. See Erlewine (2016).

Two patterns of *na*

(41) [+FOC, +D] (✓**na**) V...

  Ise   (✓**na**) modom?
  who   NA   sleep
  'Who is sleeping?'

(42) [+FOC, +D] (✓**na**) [+FOC, +D] (✓**na**) V...

  Ise   (✓**na**) holan babi (✓**na**) di-allang?
  who   NA   only   pork   NA   PASS-eat
  'Who eats only pork?'   (*na...na* ok too)

But there are also cases where my speakers' judgments split, consistently:

(43) **Pattern A (three speakers):**

    a.   [+FOC, −D] (✓**na**) V...

        Andigan (✓**na**) di-tuhor ho buku-i?

        when   NA PASS-buy you book-that

        'When did you buy that book?'

    b.   [+FOC, −D] (***na**) [−FOC, +D] (✓**na**) V...

        Andigan (***na**) buku i   (✓**na**) di-tuhor ho?

        when   NA book that NA PASS-buy you

        'When did you buy that book?'

But there are also cases where my speakers' judgments split, consistently:

(44)   **Pattern B (one speaker):**

    a. [+FOC, −D] (***na**) V...

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

        when   NA   PASS-buy you book-that

        'When did you buy that book?'

    b. [+FOC, −D] (***na**) [−FOC, +D] (***na**) V...

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

        when   NA book that   NA PASS-buy you

        'When did you buy that book?'

Q: What's the difference between (41–42) and (43–44)?

A: The examples in (41–42) are exactly where I predict C and T to be bundled as CT. In (43–44), C and T must be separate heads.

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

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

(42) [$_{CTP}$ Ise **na** [$_{CTP}$ holan babi **na** [$_{VoiceP}$ di-allang __ __]]]?
who NA only pork NA PASS-eat

'Who eats only pork?'

The availability of simultaneous *na...na* in (42) supports CT head reprojection.

Roadmap

Conclusion

Austronesian voice system languages are an interesting testing ground for our understanding of the left periphery, subjecthood, and extraction:

(3c) **The Austronesian extraction restriction:**

$\overline{\text{A}}$-extraction (*wh*-movement, relativization, etc.) is limited to the subject argument (Keenan and Comrie, 1977, a.o.).

Toba Batak exhibits this extraction restriction (Silitonga, 1973; Keenan and Comrie, 1977, a.o.): If one DP is extracted, it must be the subject, and it generally seems impossible to extract two DPs simultaneously.

☞ **But multiple DPs can be extracted *if they are both formally focused*.**

Conclusion

The extraction restriction (3c) has been a central topic in the study of Austronesian syntax...*and its traditional description might be wrong*.

☞ **Nominal licensing** may play a key role in this extraction asymmetry.

- Multiple DPs can be extracted if attracted by the **bundled CT probe**, which is both licensing (from T) and can attract multiply (from C).

Conclusion

Possible multiple extractions of DPs with non-DPs motivate the availability of **split C and T**, with their traditional division of labor: C responsible for *wh*/focus-movement and T responsible for subject licensing and movement; C > T.

☞ **The organization of the clause periphery in Austronesian languages might be much more familiar than previously thought.**

Thank you!

Paper: "Extraction and licensing in Toba Batak" LingBuzz 3261

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

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

Baker, Mark 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.

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

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

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

References II

Chomsky, Noam. 1986. *Barriers*. 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.

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.

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

Erlewine, Michael Yoshitaka. 2016. Extraction and licensing in Toba Batak. URL `http://lingbuzz.auf.net/lingbuzz/003261/current.pdf`, manuscript, National University of Singapore.

68

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

Erlewine, Michael Yoshitaka, Theodore Levin, and Coppe van Urk. to appear. Ergativity and Austronesian-type voice systems. In *Oxford Handbook of Ergativity*, ed. Jessica Coon, Diane Massam, and Lisa deMena Travis, 373–396. Oxford University Press. URL `http://ling.auf.net/lingbuzz/002629/current.pdf`.

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

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

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

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

Hsu, Brian. to appear. Verb second and its deviations: An argument for feature scattering in the left periphery. *Glossa* .

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

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

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

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

References V

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

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

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.

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. 2005. The Malagasy subject/topic as an A′-element. *Natural Language & Linguistic Theory* 23:381–457.

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

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

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

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

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

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.

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

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.

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