Agent extraction and topicalization in Bikol

Cheryl Lim and Michael Yoshitaka Erlewine (mitcho)
National University of Singapore

AFLA 25, Academia Sinica, 10–12 May 2018

Today

• We investigate patterns of A’-dependencies (especially topics) in Bikol, a Central Philippine language closely related to Tagalog.
• Topicalization can target non-subject agents, in contrast to (local) clefting.
  o We propose that topicalization involves a [TOP] probe, which can skip the subject.
  o Clefting involves probing for [D], which cannot skip the subject (Aldridge 2004, 2017).
  o Phase impenetrability explains the impossibility of topicalizing non-subject themes.
• Topicalization can feed long-distance clefting, supporting the view that the Austronesian extraction restriction is not strictly “subject-only,” but rather locality-based.

1. Background: Case, voice, and extraction in Bikol

Bikol exhibits the familiar Philippine-type voice system:

• The choice of voice marker on the verb correlates with the choice of nominative (su)-marked argument. We call the argument cross-referenced by voice the “subject.”
• As in Tagalog, non-subject agents and themes are in genitive case, except for specific non-subject themes which are dative.

(1) Voice alternation in Bikol:
  a. Nag-kaon ning keso su babayi (sa harong). Actor voice (AV)
  AV.PST-eat GEN cheese NOM woman DAT house
  b. K<in>aan su keso kaso babayi (sa harong). Patient voice (PV)
  PV.PST-eat NOM cheese GEN woman DAT house
  c. Pig-kaon-an ning keso kaso babayi su harong. Locative voice (LV)
  LV.PST-eat GEN cheese GEN woman NOM house
  ‘The woman ate (the) cheese (at home).’

• It’s often claimed that only the subject can be A’-extracted in voice system languages.
  At first glance, this is true for Bikol too.

1 We thank Kidjie Saguin and another, anonymous speaker for discussion of judgments. For comments and discussion, we thank the Syntax/Semantics Reading Group at NUS and especially Ted Levin.
2 Data here is from a speaker from Virac, but similar patterns have been confirmed with a speaker from Legazpi City with some lexical differences.
Here, consider topicalization and clefting.

We distinguish internal and external topics (terms from Aissen 1992 on Mayan):

- **Internal topics** have no intonational break and no corresponding pronoun (2).
- **External topics** have an obligatory intonational break and an obligatory pronoun (3).

(“External topicalization” is also called “left dislocation” in some other literature.)

These properties (prosodic break and pronoun) are one-to-one for local topics.

(2) **Subject internal topics:**

a. *Su babayi* nag-kaon ning keso.  
   AV agent subject int. topic  
   NOM woman AV.PST-eat GEN cheese  
   ‘The woman ate (the) cheese.’

b. *Su keso* k<in>aon kaso babayi.  
   PV theme subject int. topic  
   NOM cheese PV.PST-eat GEN woman  
   ‘The woman ate (the) cheese.’

(3) **Subject external topics — with prosodic break and pronoun:**

   AV agent subject ext. topic  
   NOM woman AV.PST-eat NOM.3sg GEN cheese  
   ‘The woman, she ate cheese.’

b. *Su keso*, k<in>aon kaso babayi *(ito).*  
   PV theme subject ext. topic  
   NOM cheese PV.PST-eat GEN woman NOM.DEM  
   ‘The cheese, the woman ate it.’

(4) **Non-subject DPs cannot be topicalized retaining their original case markers:**

a. *Ning keso(,)* nag-kaon su babayi (kaito/kaiyan).  
   GEN cheese AV.PST-eat NOM woman GEN.DEM

b. *Kaso babayi(,)* k<in>aon (=niya) su keso.  
   GEN woman PV.PST-eat GEN.3sg NOM cheese  
   Intended: ‘The woman ate (the) cheese.’

Clefting involves the movement of a focused constituent, followed by a NOM marker, and the gapped clause. Clefting can target the subject but not a non-subject:

(5) **Local subject cleft:**

*Su lalaki su [clause g<in>adan kaso eskwela].*  
NOM man NOM PV.PST-kill GEN student  
‘It’s the man that the student killed.’

(6) **Ungrammatical local non-subject agent cleft:**

* *Su / kaso eskwela su [clause g<in>adan su lalaki].*  
NOM / GEN student NOM PV.PST-kill NOM man  
Intended: ‘It’s the student that killed the man.’

Both topicalization (retaining original case markers) and clefting appear to reflect the subject-only extraction restriction.
2. Local agent topics

→ Non-subject agents cannot be topicalized, if they appear in nominative case:

(7) Non-subject agent internal topic:

Su babayi k<in>aon (*=niya) su keso.
NOM woman PV.PST-eat GEN.3SG NOM cheese
‘The woman ate the cheese.’

(8) Non-subject agent external topic:

Su babayi_2 k<in>aon *=(niya) su keso.
NOM woman PV.PST-eat GEN.3SG NOM cheese
‘The woman, she ate the cheese.’

* ‘The woman, the cheese ate it.’

- Notice that there are two nominative DPs in (7–8), but their interpretations are unambiguous: The preverbal NOM is the agent and the postverbal NOM is the theme subject.

In contrast, non-subject themes cannot be topicalized:

(9) Non-subject themes cannot be topicalized, even in nominative case:

* Su keso(,) nag-kaon su babayi (kaito/kaiyan).
NOM cheese AV.PST-eat NOM woman GEN.DEM
Intended: ‘The cheese, the woman ate it.’

We therefore concentrate on PV (Patient Voice) clauses for the rest of this talk.

Q: How are the agent topics in (7–8) possible? Why are they in nominative case?

3. Proposal

3.1. Voice and case in Bikol

(10) Working assumptions for Austronesian voice:

a. Voice morphology is in v. V head-moves to v. vP is a phase. Phrases below the double line are inaccessible due to Phase Impenetrability (Chomsky 2000).

b. The agent is base-generated in Spec,vP.

c. In Non-Actor Voices, the subject DP moves to an outer specifier of vP. This can be thought of as the effect of an EPP feature [uD*] on NAV v (Aldridge 2004, 2008) or of object shift (Rackowski & Richards 2005).

d. Constituents in vP are subject to scrambling. All linearizations of vP with the verb (v+V) as the leftmost constituent can be generated.

Actor Voice: 

Non-Actor Voices:

3
Proposal for morphological case in Bikol:
Morphological case can be **structurally assigned** or realized with **context-sensitive defaults**. See Marantz (1991), Baker (2015: ch. 4).

a. Oblique/prepositional cases are structurally assigned—e.g. DOM dative *sa*.
b. T bears a [uD] probe and assigns nominative to its goal.
c. **Default case in vP is genitive** (Erlewine, Levin, and Van Urk 2018).
d. **Default case in CP is nominative**.

In both AV and NAV, the **highest DP in vP is the subject**, so probing for [uD] by T will necessarily find the subject, giving it nominative case.

**An alternative to structural nominative:** T bears a [uD] probe and Agree with T delays the Spell-Out of the subject DP, making it act as if it is in the higher phase for default case determination. This approach has the advantage of explaining why the case on topics is systematically the same as the case on subjects: Both are default nominative.

### (12)

**a. Actor Voice:**

```
<table>
<thead>
<tr>
<th>TP</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
</tr>
<tr>
<td>vP</td>
</tr>
<tr>
<td>Spell-Out</td>
</tr>
</tbody>
</table>
```

NOM for agent subject
Default GEN for theme (if any)
⇒ e.g. “v+V NOM=S/Ag GEN=Th”

**b. Non-Actor Voices:**

```
<table>
<thead>
<tr>
<th>TP</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
</tr>
<tr>
<td>vP</td>
</tr>
<tr>
<td>Spell-Out</td>
</tr>
</tbody>
</table>
```

NOM for subject (e.g. theme)
Default GEN for agent
⇒ e.g. “v+V NOM=S GEN=Ag”

#### 3.2. Topics

⇒ We propose two heads in the clause periphery:

**Top₂ for external topics > Top₁ for internal topics.**³

```
<table>
<thead>
<tr>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
</tr>
<tr>
<td>Top₂P</td>
</tr>
</tbody>
</table>
```

³ A similar organization is proposed for Mayan languages by Aissen (1992). However, Aissen argues that the equivalent of Top₂ is above C in Mayan, as external topics are unavailable in embedded clauses, unlike in Bikol.
In Non-Actor Voices (NAV), both the subject and agent are at the vP edge and accessible for [TOP] probing. This allows for subject topicalization in all Voices, and agent topicalization in NAV.

→ **Topicalization of non-subject themes is not possible** (in AV) because non-subject themes are not at the vP phase edge and thus inaccessible for probing from above, by Phase Impenetrability.

(14) Internal and external topicalizations of subjects and non-subject agents:

a. \([\text{Top}_2 \text{ NOM}=\text{DP}_1], [\text{Top}_1 \text{ TP} ... \text{V}=\text{siya}_i \text{ GEN}=\text{Ag} \text{ S external topic}]\)

b. \([\text{Top}_2 \text{ NOM}=\text{DP}_1], [\text{Top}_1 \text{ TP} ... \text{V}=\text{niya}_i \text{ NOM}=\text{S} \text{ Ag external topic}]\)

c. \([\text{Top}_2 \text{ , } \text{Top}_1 \text{ NOM}=\text{DP}_1] [\text{TP} ... \text{V} = t_i \text{ GEN}=\text{Ag} \text{ S internal topic}]\)

d. \([\text{Top}_2 \text{ , } \text{Top}_1 \text{ NOM}=\text{DP}_1] [\text{TP} ... \text{V} = t_i \text{ NOM}=\text{S} \text{ Ag internal topic}]\)

(See Appendix A for data showing that external topicalization is island-insensitive.)

The strict ordering of \(\text{Top}_2 > \text{Top}_1\) is supported by examples with **multiple preverbal DPs**:

(15) Multiple preverbal DPs: external < internal, *internal < external

a. Si Pedro, , su babayi g<in>adan (=niya).
   NOM Pedro NOM woman PV.PST-kill GEN.3sg
   ‘Pedro, he killed the woman.’
   * ‘Pedro, the woman killed him.’

b. Si Pedro, , su babayi g<in>adan =siya.
   NOM Pedro NOM woman PV.PST-kill NOM.3sg
   ‘Pedro, the woman killed him.’
   * ‘Pedro, he killed the woman.’

(16) No prosodic break possible after the second DP:

* Si Pedro (,) su babayi , g<in>adan (=niya/=siya).
   NOM Pedro NOM woman PV.PST-kill GEN.3sg/NOM.3sg

Note that each example in (15) is unambiguous. These can be derived by our proposal:

(17) a. **Derivation of (15a):**

```
\begin{array}{c}
[\text{vP} \\
| \text{subject} \\
\text{DP[Top]} \\
| \text{woman} \\
| =niya \\
| \text{agent} \\
| =siya \\
| \text{v} \\
| \text{...} & \text{...} \\
\end{array}
```

- \([\text{uTOP}]\) on \(\text{Top}_1\) finds ‘woman,’ fronts it.
- \([\text{uTOP}]\) on \(\text{Top}_2\) finds \(=\text{niya}\), co-indexes it with a base-generated topic DP.

\(\Rightarrow [\text{Top}_2 \text{ NOM}=\text{DP}_1, [\text{Top}_1 \text{ NOM}=\text{S} [\text{V}=\text{niya}_i] \Rightarrow [\text{Top}_2 \text{ NOM}=\text{DP}_1, [\text{Top}_1 \text{ NOM}=\text{Ag} [\text{V}=\text{siya}_i]
```

b. **Derivation of (15b):**

```
\begin{array}{c}
[\text{vP} \\
| \text{subject} \\
\text{DP[Top]} \\
| \text{agent} \\
| =siya \\
| \text{v} \\
| \text{...} & \text{...} \\
\end{array}
```

- \([\text{uTOP}]\) on \(\text{Top}_1\) finds \(=\text{siya}\), but it’s a clitic so can’t be fronted.
- Probing further, \(\text{Top}_1\) finds the agent ‘woman’ and fronts it.
- \([\text{uTOP}]\) on \(\text{Top}_2\) finds \(=\text{siya}\), co-indexes it with a base-generated topic DP.
Open question: What are the conditions under which bound pronouns can be null? Notice that \(=niya\) is optional in (15a), but they are generally optional when long-distance; see Appendix A.

Summary

- **Subjects and non-subject agents are at the vP phase edge** and can be targeted by [TOP] probes for (internal and external) topicalization. Non-subject themes are never accessible, due to Phase Impenetrability.
- Topics receive **default nominative** by being in the higher phase (CP).

4. **Clefts and the nature of the extraction restriction**

Recall that, unlike local topicalization, local clefting is limited to the subject:

(18) **Grammatical local subject cleft:**

\[
\text{Su lalaki su } \left[ \text{clause } \text{g<in>adan kaso eskwela.} \right.
\text{NOM man NOM PV.PST-kill GEN student}
\text{‘It’s the man that the student killed.’}
\]

(19) **Ungrammatical local non-subject agent cleft:**

\[
\ast \text{Su eskwela su } \left[ \text{clause } \text{g<in>adan su lalaki.} \right.
\text{NOM student NOM PV.PST-kill NOM man}
\text{Intended: ‘It’s the student that killed the man.’}
\]

\[\Rightarrow\] We propose that (a) **the cleft clause is reduced** (TP) and (b) **clefting involves probing for [D]**. (Long-distance clefting is *island-sensitive*; see Appendix B.)
- With [D] probing, it’s not possible to skip the subject which is the highest DP in the clause.
- Without Top\(_1\) or Top\(_2\), it’s not possible to put a non-subject above the subject and feed clefting.

Alternatively, we might imagine that clefts specifically probe for the “subject.”

Support for our approach comes from the fact that **long distance clefts can target non-subject agents**:

(20) **Long-distance cleft of embedded non-subject agent:**

\[
\text{Su lalaki su } \left[ \text{TP pig-balita ning radyo } \left[ \text{CP na } \text{g<in>adan } (=niya) \text{ su eskwela.} \right. \right.
\text{NOM man NOM PV.PST-report GEN radio that PV.PST-kill GEN.3sg NOM student}
\text{‘It’s the man that the radio reported that \(t\) killed the student.’}
\]

(21) **Also possible: Long-distance cleft of embedded subject**

\[
\text{Su lalaki su } \left[ \text{TP pig-balita ning radyo } \left[ \text{CP na } \text{g<in>adan } (=siya) \text{ ning eskwela.} \right. \right.
\text{NOM man NOM PV.PST-report GEN radio that PV.PST-kill NOM.3sg GEN student}
\text{‘It’s the man that the radio reported that the student killed \(t\).’}
\]

\[\Rightarrow\] We propose that in (20), **embedded topicalization feeds long-distance clefting**.

- Unlike the edge of the cleft clause itself, which is reduced (TP), the **embedded clause is a full CP**, including Top\(_1\) and Top\(_2\) projections.
- Topicalization (internal or external) of a non-subject agent within the embedded CP makes it the highest DP in the embedded clause.
- The cleft probes for [D], **finding the highest DP** in the embedded clause, **which is the topic**.
Q: How can clefting probe into the embedded clause in (20–21)?
A: Notice that the matrix clause in (20–21) is Patient Voice, so the embedded CP is formally the subject. **Movement of the CP to Spec,vP allows for subsequent probing into the CP** (Van Urk and Richards, 2015). This correctly predicts that long-distance clefting is ungrammatical with an AV matrix verb (22).

(22) Long-distance clefting is ungrammatical across a AV matrix clause:

\[ * \text{Si A. su [TP nag-balita si Pedro [CP na g<in>adan (=niya) su lalaki. NOM A. NOM AV.PST-report NOM Pedro that PV.PST-kill GEN.3sg NOM man Intended: 'It's Andrew that Pedro reported that t killed the man.'} ] } \]

Evidence comes from patterns of long-distance clefting across embedded topics:

(23) Long-distance clefts across embedded topics:

a. NOM=Ag [TP ... [CP NOM=S V=niya]]
Su lalaki su [TP pig-balita ning radyo [CP na su eskwela g<in>adan (=niya). NOM man NOM PV.PST-report GEN radio that NOM student PV.PST-kill GEN.3sg
i. ‘It’s the man that the radio reported that the student killed t.’
ii. ‘It’s the man that the radio reported that t killed the student.’

b. NOM=S [TP ... [CP NOM=Ag V=siya]]
Su lalaki su [TP pig-balita ning radyo [CP na su eskwela g<in>adan =siya. NOM man NOM PV.PST-report GEN radio that NOM student PV.PST-kill NOM.3sg
i. ‘It’s the man that the radio reported that the student killed t.’
ii. ‘It’s the man that the radio reported that t killed the student.’

Notice that the cleft focus corresponds thematically to the clitic pronoun in the embedded clause:
- ‘The man’ is the embedded agent (=niya) in (23a).
- ‘The man’ is the embedded subject theme (=siya) in (23b).

The embedded clauses in (23a,b) each started with both an external and internal topic. **Cleft formation necessarily attracts the highest DP among the two topics**, which is the embedded clause’s external topic. (For the derivation of simultaneous external and internal topics, see (17).)

(24) Long-distance cleft from an embedded clause with two topics:

c. DP[TOPIC, FOCUS] NOM [TP ... [CP na [Top2 t [Top1 DP [V... d. *DP[TOPIC, FOCUS] NOM [TP ... [CP na [Top2 DP [Top1 t [V... This supports the view that cleft-formation involves probing for [D] (equivalent to Aldridge’s (2004, 2017) [φ] probe), which unambiguously targets the subject in cases without embedded topics. In particular, there is no preference for attraction of the “subject.”

---

4 See also Rackowski & Richards (2005) and Halpert (2012, 2016, to appear) for a similar approach, where Agree with a phase “unlocks” it for subsequent probing.
5. Conclusion

- Internal and external topicalization in Bikol can target **non-subject agents** as well as subjects, but not **non-subject themes**.
  
  ➔ **Subjects and non-subject agents are the only DPs at the vP phase edge** (see Erlewine & Levin ms). (See also Appendix A on long-distance external topics.)

- **Local clefts can only target subjects**, which is compatible with (a) probing for the highest DP (with a restriction against forming topics in cleft clauses) or (b) probing for the “subject.”

- The behavior of long-distance clefts shows that **clefts are just attracting the highest DP**. In particular, **embedded topicalization (probing for [TOP]) can feed clefting (probing for [DI]), resulting in the grammaticality of long-distance clefting of non-subject agents**.
  
  ➔ The “subject-only” extraction restriction is best thought of as a **locality-based effect** (as in Aldridge 2004, 2017, Rackowski & Richards 2005, a.o.), rather than thematic role-based.

References
Appendix A: Long-distance topics are external topics; island-insensitive
Both subjects and non-subject agents can be topicalized long distance. In either case, (a) there is an 
obligatory prosodic break, but (b) the corresponding pronoun is optional.

(25) Long-distance subject topic:  
Su lalaki *(s) pig-balita ning radyo [CP na g<in>adan (=siya) ning eskwela. 
NOM man PV.PST-report GEN radio that PV.PST-kill NOM.3sg GEN student 
‘The man, the radio reported that the student killed (him).’

(26) Long-distance non-subject agent topic:  
Su eskwela *(s) pig-balita ning radyo [CP na g<in>adan (=niya) su lalaki. 
NOM student PV.PST-report GEN radio that PV.PST-kill GEN.3sg NOM man 
‘The student, the radio reported that (he) killed the man.’

Given the obligatory prosodic break and optionality of the pronoun, it is unclear whether these are 
internal or external topics — i.e. moved or base-generated high.

We argue that long-distance topics as in (25–26) are always external topics, with optional 
pronunciation of the pronoun. This dependency is island-insensitive, with or without the pronoun. 
Open question: Why are there no long-distance internal topics? (Topics can be moved; see clefts.)

(27) Long-distance topics are island-insensitive:  
a. Su babayi, pig-uran [2 bagu pig-hiling (=siya) ni Andrew. 
NOM woman PV.PST-rain before PV.PST-see NOM.3sg GEN Andrew 
‘The woman, it rained [2 before Andrew saw (her)].’
b. Si Andrew, pig-uran [2 bagu pig-hiling (=niya) su babayi. 
NOM Andrew PV.PST-rain before PV.PST-see GEN.3sg NOM woman 
‘Andrew, it rained [2 before (he) saw the woman].’

The behaviour of long-distance topics supports the proposal (a) that external topics involve [TOP] 
probing by Top; to link the base-generated topic to a corresponding pronoun and (b) that probing is 
sensitive to Phase Impenetrability.

- Recall that long-distance clefting require a PV matrix verb (20–22).
- The same requirement is observed with long-distance topics: the matrix verb must be PV as 
in (25–26) rather than AV (28). This cannot be reduced to movement, as long-distance topics 
are island-insensitive (27). Instead, this reflects the requirement [TOP] probing by Top2.
- In the PV matrix clause, the embedded CP is the subject and moves to Spec,vP, allowing for 
subsequent probing into the CP (Van Urk and Richards, 2015).

(28) Base-generated (external) topics can bind into CPs in Spec,vP, but not into VPs:  

a. [CP TOPICi ... [vP pro,  v^θ [vP ... (3), (8) 
b. *[CP TOPICi ... [vP  v^θ [vP ... pro,] (9) 
c. [CP TOPICi ... [vP [CP ... pro, ... ] v^θ [vP ... (25, 26) 
d. *[CP TOPICi ... [vP  v^θ [vP ... [CP ... pro, ... ] (27)
Appendix B: Long-distance clefts are island-sensitive
As we saw in section 4, both subjects and non-subject agents can be clefted long-distance.

Unlike long-distance (external) topics, clefting is island-sensitive:

(30) Clefting is island-sensitive:\(^5\)
a. *Su babayi su pig-uran \(\wedge\) bagu pig-hiling (=siya) ni Andrew.\n   NOM woman NOM PV.PST-rain before PV.PST-see NOM.3sg GEN Andrew\n   Intended: ‘It’s the woman that it rained \(\wedge\) before Andrew saw at (her)].’
b. *Si Andrew su pig-uran \(\wedge\) bagu pig-hiling (=siya) su babayi.\n   NOM Andrew NOM PV.PST-rain before PV.PST-see GEN.3sg NOM woman\n   Intended: ‘It’s Andrew that it rained \(\wedge\) before (he) saw at the woman].’

Appendix C: Non-subject agent topics in other Philippine languages
Although generally understudied, some examples of non-subject agent topics in other Philippine languages can be found in previous literature. See also Pizarro-Guevara & Wagers (last talk).

a. Ang nanay, lu~lutu~in (=niya) ang isda.\n   NOM mother IPFV~cook-PV GEN.3sg NOM fish\n   ‘The mother, (she) will cook the fish.’
b. Ang isda, mag~lu~luto (nito) ang nanay. \(^7\)\n   NOM fish AV~IPFV~cook GEN.DEM NOM mother\n   Intended: ‘The fish, mother will cook (it).’

(32) Kapampangan (Sells 2000: 124):
Ing lalaki e =na =ya seli ing mangga. \(^8\)\n   NOM man not GEN.3sg NOM.3sg bought NOM mango\n   ‘The man did not buy the mango.’

Appendix D: Clefting vs relativization
Recall that local non-subject agents cannot be clefted; see (6/19).

In contrast, local non-subject agents can be relativized:\(^6\)

(33) Local non-subject agent relative:
   Su eskwela [\(\_\_\_\_\_\) na g<in>adan su lalaki] na-takdag.\n   NOM student that PV.PST-kill NOM man PV.PST-fall\n   ‘The student [\(\_\_\_\_\_\) that killed the man] fell.’

It is commonly thought that (pseudo)clefts as in (34) involve (headless) relativization.

Bikol (pseudo)clefts cannot be built from (regular) relative clauses.
Relativization must instead be built on a clause larger than TP or use a selective [REL] probe, not [D].

---

\(^5\) Both examples have alternative parses where the extraction is local, and not related to the content inside the island: ‘It’s the woman/Andrew that it rained on [\(\wedge\) before …].’ What’s important is that the intended readings are unavailable.

\(^6\) Cf “To our knowledge, there is no Philippine language which unambiguously allows relativization of either the Genitive Agent of a transitive sentence, or…” — Reid & Liao 2004: 482