Co-Indexing System and the Changing Face of Austronesian Voice: Insights from Sipora Mentawai

Emma Keith & I Wayan Arka

INTRODUCTION. This paper presents an LFG analysis of the synchronic and diachronic typology of voice and pronominal indexing in Sipora Mentawai (SM) (mwv, ISO639-3, Barrier Islands, Austronesian (AN), Indonesia), drawing on recent fresh fieldwork data. The study situates SM (est. ~25,000 L1 speakers) within western Austronesian, a subgroup of AN characterised by symmetrical voice (i.e., A/U are equally selectable as SUBJ without PASS-like demotion). SM exhibits unusual and intriguing features, including pronominal co-indexing affixes reflecting NOM alignment and the loss of the PIVOT-SUBJ-only constraint typically associated with AN symmetrical voice. These traits distinguish SM from Indonesian-type AN languages but align it with neighbouring Barrier Islands languages like Enggano and, in some respects, with central-eastern AN languages such as Kambera (Klamer 1996). We demonstrate that LFG's modular and parallel architecture provides a robust framework for analyzing how these properties interact and evolve, shaping SM's typological profile.

KEY DATA POINTS. Like Enggano (Hemmings and Dalrymple to appear, Hemmings to appear), SM develops two different sets of bound pronominals on the verb, but unlike Enggano the sets include nominative suffixes in addition to indexing prefixes.[1] Crucially, SM verb stems signify irrealis (IRR) by default (1a), which may be overridden by REAL(is) *a-* (1b) or any of the several *m*-initial voice morphemes (e.g. *masi-*, 2a).

1 a. **ra**-*matei-ake'*   *sikoinan.*   b. *a-**ra**-matei-ake'(-an)*   *sikoinan.*
      3PL.IRR-dead-CAUS   crocodile   REAL-3PL.IRR-dead-CAUS(-PERF)   crocodile
      'They will kill the crocodile.'   'They (have) killed the crocodile.'

Like Enggano, SM exhibits erosion of its AN voice, notably the loss of its AN voice symmetricality property, giving rise to multiple 'active' transitive sentences: those with a pronominal prefix (1), and those with *masi-,* a remnant of Actor Voice (AV) morphology (2a). The co-indexing system in SM allows alternative structures shown in (2a-b), with the same logical meaning and mood but distinct information structure. We argue that (2b) is not a PASS(ive) or Undergoer Voice counterpart of (2a) although it is translatable into English passive free translation.

2 a. Si Yosep **masi**-itco'   [HP]_U   b. HP   nera   a-**i**-itco'   si Yosep
      ART Yosep AV.REAL-see mobile.phone   mobile.phone   that   REAL-3SG.A-see ART Y
      A   Undergoer   Undergoer   Actor
      'Yosep saw a/the mobile phone.'   'The mobile phone was seen by Yosep.'

In our analysis, the prefix *i-* in (2b) is not a PASS (or Undergoer Voice) marker, despite its formal resemblance with the passive markers in other AN languages in western Indonesia, such as *ni-* (Nias) and *di-* (Enggano). At its current stage of evolution, SM *i-* remains pronominal and referential. In LFG's formalism, it carries (($\uparrow$PRED)='pro') as shown in (9e) below. Critical evidence comes from the content question in (3a). This demonstrates that, as a question pronoun without the specified PERS/NUM, features, SUBJ *kasei* 'who' cannot be coindexed; *i-* cannot be questioned. Consequently, only reading (3a.i)—which questions the OBJ—is possible, while reading (3a.ii)—which questions the SUBJ (index *j*)—is unacceptable. Note that in the declarative equivalent structure (3b), where SUBJ and OBJ both possess the required features, the sentence is ambiguous (out of context), as indicated by possible indices *i/j* linked to both co-arguments.

The concomitant loss of AN voice symmetricality entails the disappearance of the privileged PIVOT/SUBJ-only constraint, the hallmark of AN voice symmetricality. We discuss two pieces of evidence coming from complex structure formation. First, consider the embedded clausal ADJUNCT and COMP structures in (4a-

---

[1] The complete indexing sets are as follows. The suffix set is inherited from the PMP NOM2 paradigm (cf. Ross 2006), while the prefix set seems to be a unique areal innovation.

		1	2	3
SG		*ku-/-ku*	*nu-/-nu*	*i-/-na*
PL	INCL	*ta-/-ta*	*nu- kam/-mui*	*ra-/-ra*
	EXCL	*ku- kai/-mai*		

b), which demonstrate that the (prefixed) SUBJ in Sipora Mentawai (SM) does not bear a privileged PIVOT function. That is, unlike in Indonesian-type languages with a robust symmetrical voice system, such as Balinese (example (5)), the prefixed SUBJ in SM cannot be gapped (or controlled), as indicated by *( ). Its properties align with those of SUBJ in the AN co-indexing languages of eastern Indonesia, such as Kambera (Klamer 1996). Notably, in the equivalent Balinese structure in (5), the selected Actor=PIVOT-SUBJ argument must be syntactically controlled (i.e., gapped). The verbal voice in Balinese, **AV ng-** in (5) clearly marks A=SUBJ/PIVOT selection.

3 a. *kasei_j*   *a-**i**-kukru*   [*jo'jo' nera*]_*i*?
      who   REAL-3SG.A_*i/*j*-chase   dog   that
      i) 'Who was chased by the dog?/Who did the dog chase?'
      ii) ?* 'Who chased the dog.'

   b. *Yosep_i/j*   *a-**i**-kukru*   [*jo'jo' nera*]_*i/j*?
       Yosep   REAL-3SG.A_*i/j*-chase   dog   that
       i) 'Yosep chased the dog.'   (Yosep=A/SUB, 'dog'=pt/OBJ) (preferred)
       ii) 'Yosep was chased by the dog?' (Yosep=P/OBJ, 'dog'=agt/SUBJ)[2]

4 a. a-mei   **aku**   ka   pelabuhan [*(**ku-**)gaba   iba   s(i)=abeu]ADJUNCT   *(ku- is obligatory)*
      REAL-go 1SG   LOC harbour   1SG-look.for   fish   REL-large
      'I went to the harbour to look for big fish.'   [Sipora Mentawai]

   b. **aku**   **masi**-guglu-ake'   toga   nera [*(**i-**)kukru   jo'jo']COMP *(i- is obligatory)*
      1SG   AV.REAL-command-APPL   child   that 3SG.A-chase   dog
      'I made the child chase the dog.'   [Sipora Mentawai]

5   Made Rawi macelep   [_   **ng**-aba/*aba   yeh a   lumbur]XADJUNCT
    name   MID.enter   SUBJ **AV**-bring/**UV.**bring water one   glass
    'Made Rawi entered bringing a glass of water.'   [Balinese, Arka 2003:24]

Second, intriguing evidence comes from relativisation: OBJ in SM can be relativised, even in the active structure with overt AV prefix *masi-* as seen in (6). This is impossible in Indonesian-type languages with symmetrical voice. OBJ relativisation is made possible in SM due to its evolution in allowing IHRCs (internally headed relative clause), which (unlike in Balinese) requires no 'extraction' privileging SUBJ. In our analysis the relativised head in (6) is a zero pronoun in the SUBJ position (i) or in OBJ position (ii); this cannot be made overt. Relativised argument ambiguity is the hallmark of IHRC cross-linguistically (Bonneau 1992). SM also allows a RC marked by the affixal clitic *si=,* example (7). Note the same morpheme (*si*) appears before a proper name (e.g., *Si Tiur*), traditionally glossed as ART(icle); we assign the category of D (cf. the entry in (9d)) for both ART and REL(ativiser). This gives DEF=+ to the DP.

6 [a-*masi-kukru*   *jo'jo' nera*]IHRC *niate' si   Tiur*
    REAL-AV.REAL-chase dog   that   COP   ART   Tiur
    (i) The one who chased the dog is Tiur   (A-SUBJ relativisation)
    (ii) The one who the dog chased is Tiur   (P-OBJ relativisation)

---

[2] Inserting the pronominal copy *nia* immediately after the verb, as shown below, disambiguates the structure; only reading (ii) is acceptable. This is explained by the interaction of discourse pragmatics (anaphoricity/i-str) and syntax in SM: the pronominal copy in the OBJ position must find a pragmatically prominent antecedent, preceding it in a higher left-peripheral position. This DP, bearing contrastive TOP, is analysed as a 'dislocated' NP. Due to the uniqueness condition and its backgrounding, making it pragmatically less prominent, the DP 'dog' cannot serve as the antecedent of OBJ *nia*. Consequently, reading (i) is unacceptable.
    [*Yosep_j*]TOP-C   [[*a-**i**-kukru*   [nia]OBJ*j*]V, [*jojo'   nera*]SUBJ_*i*]VP
    Yosep   REAL-3SG.A_*i*-chase 3SG   dog   that
    (i) * 'Yosep chased the dog.' (Yosep=SUB, 'dog'=OBJ);
    (ii) 'Yosep was chased by the dog?' (Yosep=OBJ, 'dog'=SUBJ).

7 [nganga **si**=buru' [**si**=kau-ra_*i* [tai kebbuk-at-ta]_*i*]REL.CLAUSE
   language REL=old REL=give-3PL.A PL.PERS older.sibling-NMLZ-1PL.INCL.POSS
   'The old language that our ancestors gave.' [Sipora Mentawai]

LFG ANALYSIS. The proposed LFG analysis for SM consists of information specification in lexical entries, c-str and m-str formulation, and related constraints. We adopt a traditional morpheme-based morphology with the m-str generated by the word-formation rule informally shown in (8). Sample entries are given in (9). In terms of c-str, we adopt an LFG-version of X-bar syntax (cf., Kroeger 1993, Bresnan et al. 2015) to account for SM configurational syntax. Strong evidence for a VP structure includes the fact that its NP must be post-verbally adjacent to its V head when it is not given pragmatic focus, and sentential adverbials like *sokat* 'yesterday' cannot intervene in the [V NP] sequence. The core clause (IP) with its extended maximal structure (CP) shows contrastive DF in [Spec, CP], with SUBJ as the default TOP in [Spec, IP]. The post-verbal free NP, co-indexing the pronominal prefix *i-* (cf. examples (2b), (3a)), functions either as an ADJ within VP/outside IP, or the default TOP/SUBJ within IP (example (3b, reading i).

8 V → (MOOD.PREF) + (PRON.PREF) + (VOICE.PREF)+ V.STEM + (PRON.SUFF) + (ASP.SUFF)

9   Sample lexical entries:
  a. *a-* MOOD.PREF (↑MOOD)=REAL b. *-an* ASP.SUFF (↑ASP)=PERF
  c. *masi-* VOICE.PREF (↑SUBJ)σ = ↑σ ACTOR (↑MOOD)=REAL
  d. *si* D ((↑PRED)= 'pro') (↑DEF)= +
  e. *kukru* V.STEM (↑PRED)= 'chase<ACTOR, UNDERGOER>' ((↑MOOD)=IRR)
  e. *i-* PRON.PREF (↑SUBJ)=↓ (↑SUBJ)σ = ↑σ ACTOR
       ((↑PRED)='pro') (↓PERS)=3 (↓NUM)=SG.

10 a.
```
            CP
    DP              C'
(↑FOC-C)=↓  C        IP
(↑TOP-C)=↓      DP   ,    I'
         (↑SUBJ) =↓  I       VP/V'
         (↑TOP) =↓  (AUX)   V'        DP
                          V    DP   ↓∈ (↑ADJ)
```

b.
```
        DP([Rel.Clause])
         ↓∈ (↑ADJ)

    DP          D'

    D              NP/VP
    |
    si=
```

PRED   'BE<SUBJ, PREDLINK>'
PREDLINK [PRED 'TIUR']

          PRED   [1]
SUBJ   ADJUNCT TYPE   *relative*
                   FOCUS [2]
                   PRED   'chase<SUBJ OBJ>'
                   SUBJ   PRED 'dog'
                          DEF +
                   OBJ   [2][PRED [1]['pro']]
                   MOOD   REAL

The IHRC (without *si=*) in (6) is analysed as an embedded clausal unit (i.e., a finite IP containing REAL MOOD) that functions as SUBJ of the copula *niate'*. Its partial f-str is shown in (0). The RC with *si=* in (7) will have a different c-str, forming a DP (as it is headed by D *si=,* cf. (10b). However, its f-str is similar to (0), differing only in that it functions as an ADJ of the matrix PRED, whose value is 'language' instead of 'pro' (tag [1]).

CONCLUSION. This paper makes an empirical contribution to language typology and AN studies by presenting new, salient data on grammatical relations and the co-indexing system in SM. The discussion is typologically framed within the diachrony of AN voice symmetricality, arguing that the emergence of the pronominal indexing system and nominalisation via IHRC has led to the decline of AN voice symmetricality. We have demonstrated how LFG provides a framework for capturing the complexities of the morphosyntax-pragmatics interface in SM. The full paper will present additional data and further develop the LFG-based analysis, contributing theoretically to the diachronic typology of grammatical relations, nominalisation, and voice-alternating systems in AN and beyond.

REFERENCES. Bresnan, J., et al. 2015. *Lexical-Functional Syntax (second edition)*. Oxford: Wiley-Blackwell. **Bonneau, J.** 1992. *The Structure of Internally-Headed Relative Clauses: Implications for Configurationality*. PhD thesis, McGill University. **Hemmings**, C. to appear. "Verbal morphosyntax in Enggano " In *Enggano: Historical and Contemporary Perspectives*, edited by I W Arka, et al. Canberra: Asia-Pacific linguistics. **Hemmings**, **C, and M. Dalrymple**. to appear. "Relative clauses in Enggano." In *Enggano: Historical and Contemporary Perspectives*, edited by I W Arka, et al. Canberra: Asia-Pacific linguistics. **Klamer, M**. 1996.

"Kambera has no passive." *NUSA* 39:12-30. **Kroeger, P.** 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford: CSLI Publications. **Ross, M.** 2006. "Reconstructing the case-marking and personal pronoun systems of Proto-Austronesian" In *Streams Converging into an Ocean: Festschrift in Honor of Professor Jen-kuei Li on his 70th Birthday*, edited by H Y Chang et al. Taipei: Academia Sinica.