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

Emma Keith[1] and I Wayan Arka[1, 2]
[1]Australian National University
[2]Udayana University

Key Points

- The symmetrical voice system which we see in other languages of Western Indonesia (Ross 2002) has disappeared in SM, and in its place has emerged an elaborate system of pronominal indexing
- Concominant with this change, the SUBJ-only constraint on relativisation has been lost, and SUBJ can no longer be gapped/controlled
- Instead, pronominal indexing is used to co-index GFs in different clauses, relying on PERS/NUM features instead of PIVOT and at times creating ambiguity
- We explain the behaviour of this co-indexing system with a modified version of X-bar theory for LFG c-str (Bresnan et al. 2015; Kroeger 1993)

Background

- The Mentawai languages are a group of closely-related languages spoken on the Mentawai islands, an archipelago of four islands near Sumatra, Indonesia
- Sipora is one of these islands, with a population of ~23k; unlike many other Indonesian regional languages SM is still spoken by all age groups in most domains except for formal education (in Indonesian)
- The Mentawai languages are apparently a first-order branch of the Sumatran family within Malayo-Polynesian (Billings and McDonnell 2024), and have had diverged significantly from what we typically expect of Malayo-Polynesian languages of Western Indonesia

Loss of Symmetrical Voice

- Most languages of Western Indonesia have symmetrical voice systems: alternation between an **Actor Voice**, which maps Actor to SUBJ and Undergoer to OBJ; and an **Undergoer Voice**, which does the opposite – neither being more marked (Riesberg 2014)
- SM **lacks** this alternation: there is only one voice morpheme on transitive verbs: *pasi-/masi-*, and it maps SUBJ to A

```
SUBJ   OBJ

Actor  Undergoer    AV

SUBJ OBJ

Actor  Undergoer    UV

C-str determines GFs, e.g.,

IP→NP(↑SUBJ)=↓ VP↑=↓
VP→V’ NP(↑OBJ)=↓
```

Loss of Symmetrical Voice

[cicing-e]$_{SUBJ}$ Ø-sepak [tiang]$_{OBJ}$
A V U
[dog-DEF] UV-kick [1SG]

[tiang]$_{SUBJ}$ ny-epak [cicing-e]$_{OBJ}$
U V U
[1SG] UV-kick [dog-DEF]
'I kicked the dog' (Artawa 1998:8)

Ø- VOICE.PREF (↑SUBJ)σ=↑σU
N- VOICE.PREF (↑SUBJ)σ=↑σA

[Balinese: Symmetrical Voice]

[si Yosep]$_{SUBJ}$ masi-itco' [HP]$_{OBJ}$
A V U
[PRT name] AV.REAL-see mobile
'Yosep saw the mobile phone'

masi- VOICE.PREF (↑SUBJ)σ=↑σACTOR
(↑MOOD)=REAL

[SM: No voice alternation]

Loss of Voice Alternation

- Contra previous analyses (Gil 2015; Lenggang et al. 1978), we do *not* analyse *i-* as a UV marker: although it is formally similar to Nias *ni-*; Enggano *di-* (Brown 2001; Hemmings to appear), A-initial clauses with *i-* on the verb appear
- Instead, *i-* is one of a paradigm of NOM-aligned pronominal indexing morphemes which co-references the SUBJ, wherever it is in the c-str

[Yosep]$_{SUBJ}$ a-i-kukru [jo'jo' nera]$_{OBJ}$
A^1/U^2 V A^2/U^1
[name] REAL-3SG.NOM-chase [dog that]
[1]Yosep chased the dog (preferred)
[2]Yosep was chased by the dog (requires discourse context)

Proof of Referentiality

- *i-* can be shown to be referential where there is no SUBJ NP, wherein *i-* provides pronominal SUBJ – ((↑PRED)='PRO')

ta' [i]$_{SUBJ}$-arep
NEG 3SG.NOM-hear
'S/he doesn't hear (i.e., is deaf).'

- If a DP does **not** have the required PERS/NUM features, including if it is non-referential, co-indexation is disallowed

[kasei]$_{OBJ}$ a-[i]$_{SUBJ}$-kukru [jo'jo' nera]$_{SUBJ}$
[who] REAL-3SG.NOM-chase [dog that]

'Who was chased by the dog/who did the dog chase?'

*'Who chased the dog?'

Co-Referencing Requirements and Behaviour

- The LFG architecture provides a neat way to understand the behaviour of *i-* and other pronominal indexing morphemes:
- Since there is no UV or PASS, the mapping between syntactic and macro-semantic roles always produces the same result: SUBJ=A, OBJ=U
- *i-* provides referential features to SUBJ, whether the co-indexing free SUBJ DP is Spec of IP (a) or of CP (b); if this indexing SUBJ DP is absent, then 'PRO' is provided (c). It is a pronoun, semantically referential, and therefore cannot be co-indexed with a variable operator, such as *kasei* 'who' (d).

No voice morphology

'PRED ... <A, (U)>'

[-o] [-r]

[-r]

<SUBJ, (OBJ)>

Voice morphology

'PRED ... <A, (U)>'

[-o] [-r]

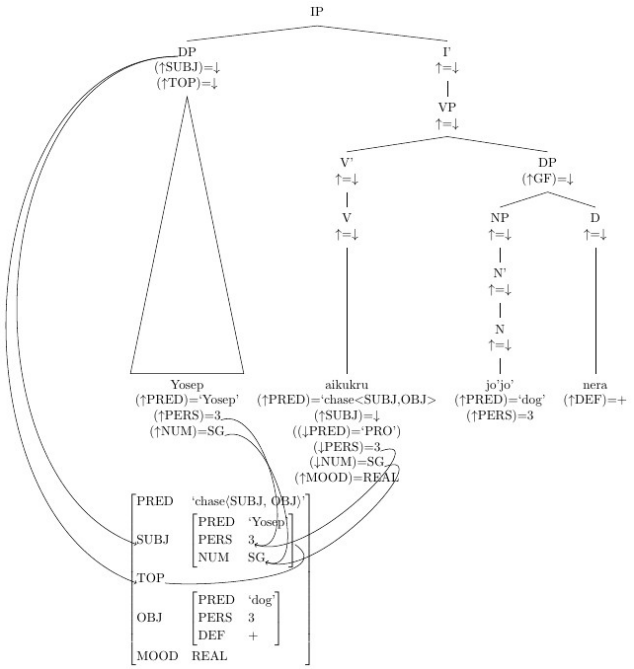
[-r]

(↑SUBJ)σ=↑ σACTOR

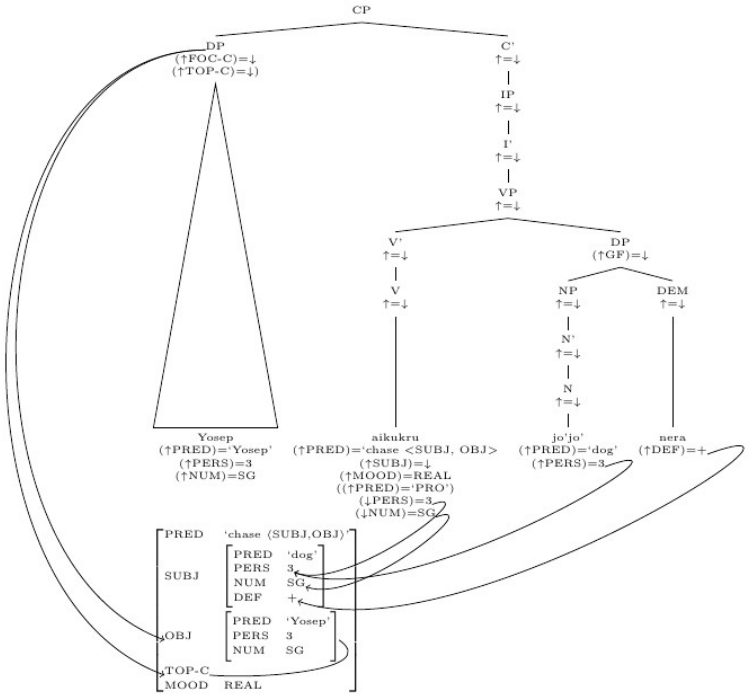
<SUBJ, (OBJ)>

QW Non-Referentiality Constraint:
A question word (QW) functioning as an operator (Q-Op) cannot be cross-referenced or indxed by a pronominal marker with referential value in argument indexing.

Co-Referencing Requirements and Behaviour



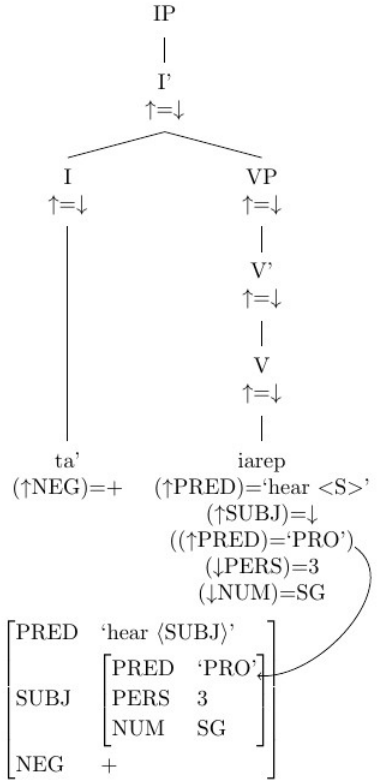
a: Yosep=A/SUBJ



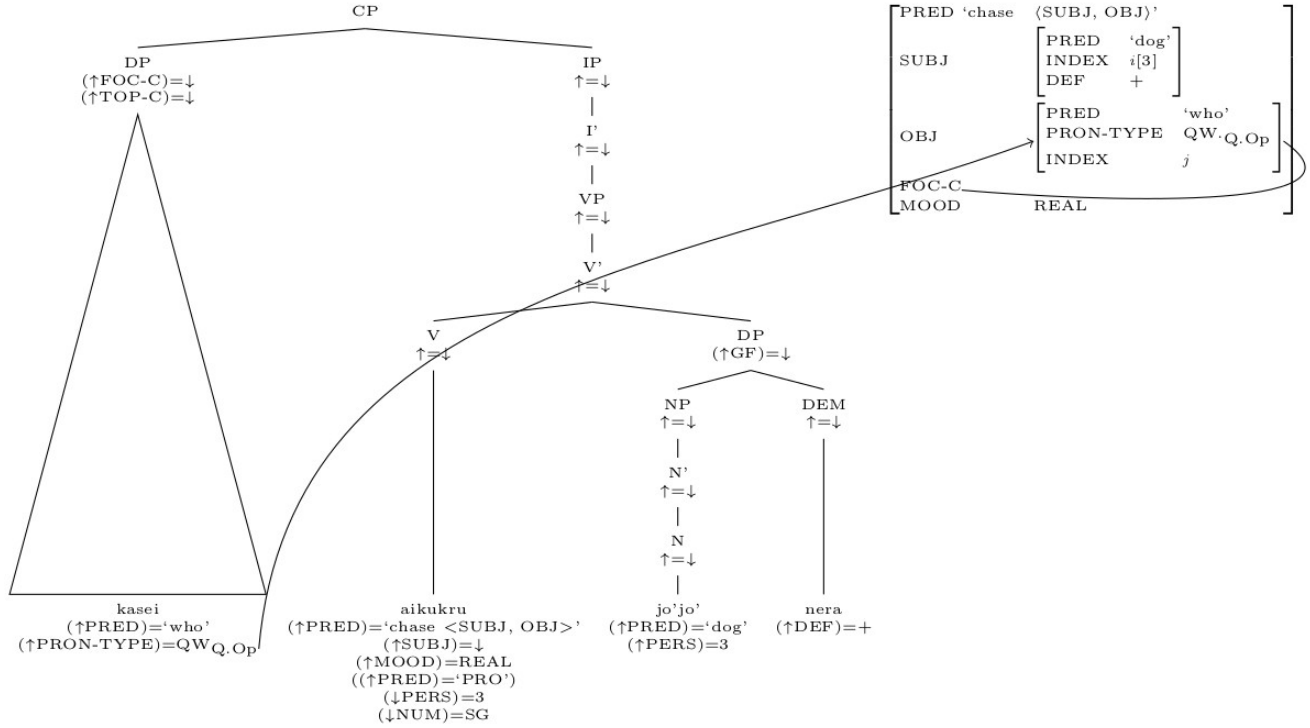
b: Yosep=P/OBJ

Co-Referencing Requirements and Behaviour

c: Providing own PRED value



d: QW Non-Referentiality Constraint



Paragidms: Pronominal Indexing Affixes

	1			2		3	
	SG	PL		SG	PL	SG	PL
		INCL	EXCL				
PREFIX	ku-	ta-	ku- kai	nu-	nu- kam	i-	ra-~da-
SUFFIX	-ku/C_ -kku/V_	-ta/C_ -tta/V_	-mai	-nu/C_ -m/V_	-mui	-na	-ra~-da/C_ -dda/V_

- **All** have NOMinative alignment
- Suffixes occur in RELative clauses and on perfective (but not realis)-marked verbs
- Prefixes occur in main clauses, irrespective of TAM marking
- The suffixes are apparently older, and reflect the PMP NOM2 set (Ross 2006), while the prefixes seem to be an areal innovation, having cognate forms in Nias (Brown 2001) and Enggano (Hemmings and Dalrymple to appear)

Implications of the loss of voice alternation

1) Loss of PIVOT
2) Loss of SUBJ-only restriction in relativisation
3) Flexibility of constituent ordering, and resultant DF correlates in c-str

Loss of PIVOT

- In symmetrical-voice languages (e.g., Balinese, Arka 2003), only the SUBJ has access to PIVOT, which provides privileges such as the ability to be gapped/controlled within a subordinate clause from the main clause
- In SM as in the pronominal co-indexing languages of Eastern Indonesia (e.g., Kambera – Klamer 1996), the pronominal SUBJ in subordinate clauses can never be gapped, since the SUBJ does **not** have exclusive access to relativisation – no PIVOT in these languages

Loss of PIVOT

- In Balinese, since PIVOT is an exclusive property of the SUBJ, a gapped arument in an XADJUNT or XCOMP will always be the SUBJ; there is no ambiguity

```
Made Rawi macelep    [_SUBJ ng-aba/*Ø-aba     yeh    a  lumbur]XADJ
NAME      MIDV.enter [    AV-bring/*UV-bring water one glass]
```

‘Made Rawi entered bringing a glass of water.’ (Arka 2003:24)

- In SM, in the absence of a PIVOT, pronominal co-indexing is required to express coreferentiality in ADJUNCT and COMP clauses. Crucially, this co-indexing is anaphoric and may introduce ambiguity, as it can refer to either the matrix SUBJ or the matrix OBJ

Loss of PIVOT

a-mei aku ka pelabuhan [**ku-**gaba iba s(i)=abeu]$_{ADJ}$
REAL-go 1SG LOC harbour [1SG.NOM-look.for fish REL=large]

‘I went to the harbour to look for big fish.’

aku masi-guglu-ake’ toga nera [**i-**kukru jo’jo’]$_{COMP}$
1SG AV.REAL-command-APPL child that [3SG.NOM-chase dog]

‘I made the child chase the dog.’

- In both sentences above, the pronominal coindexing from inside ADJ/COMP is compulsory

Loss of PIVOT

- The loss of PIVOT status can be explained functionally: there must always be a way for ACTOR to be the SUBJ of the ADJ/COMP; if the method for detaching the matrix SUBJ from the ACTOR disappears then so does the SUBJ-only constraint

Balinese: (↑XADJ SUBJ) = (↑SUBJ)

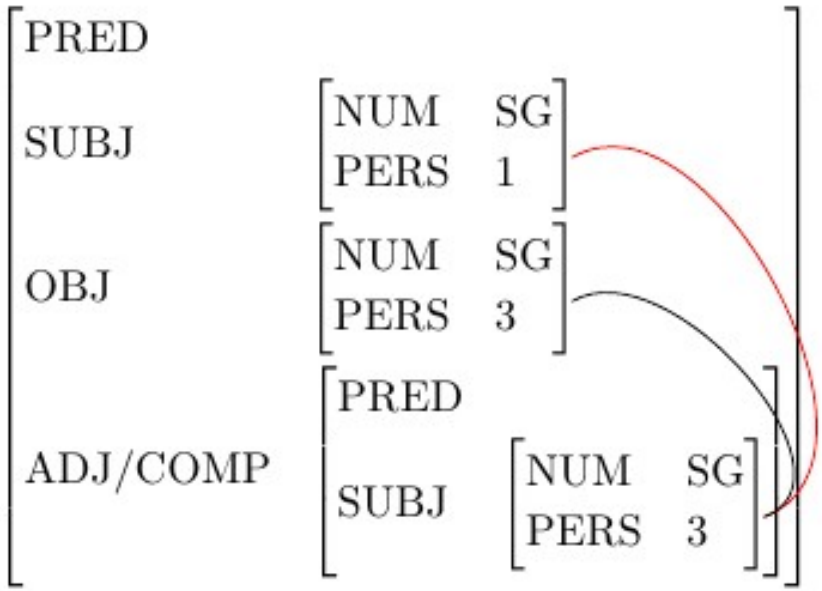
$$\begin{bmatrix} \text{PRED}_1 & \langle \ldots \rangle \\ \text{SUBJ} & \\ \text{OBJ} & \\ \text{XADJUNCT/XCOMP} & \begin{bmatrix} \text{PRED}_2 & \langle \ldots \rangle \\ \text{SUBJ} & \\ \text{OBJ} & \end{bmatrix} \end{bmatrix}$$

Sipora Mentawai: No restriction

$$\begin{bmatrix} \text{PRED}_1 & \langle \ldots \rangle \\ \text{SUBJ} & \begin{bmatrix} \text{INDEX} & i \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{INDEX} & i \end{bmatrix} \\ \text{ADJ/COMP} & \begin{bmatrix} \text{PRED}_2 & \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & i \end{bmatrix} \\ \text{OBJ} & \end{bmatrix} \langle \ldots \rangle \end{bmatrix}$$

Loss of PIVOT

- SM resolves ambiguities in a different way, namely through **anaphoric agreement**: the PERS/NUM features of the embedded SUBJ must match those of the relevant Grammatical Function (GF)



Loss of SUBJ-Only Resctriction in Relativisation

- Unlike symmetrical voice languages, SM allows the relativisation of OBJ – like with COMP/ADJ, no constraining equation exists
- This is because SM permits Internally-Headed Relative Clauses (IHRCs), where no ‘extraction’ with a SUBJ-only requirement exists
- This is what typically occurs with IHRCs: relativised arguments are ambiguous (Bonneau 1992)

Loss of SUBJ-Only Constraint in Relativisation

[a-masi-kukru _$_{i}$ [jo'jo' nera]$_{j}$]$_{IHRC}$ niate' si Tiur
[REAL-AV.REAL-chase SUBJ dog that] COP ART NAME

'The one who chased the dog is Tiur' (SUBJ is target of relativisation)

[a-masi-kukru _$_{i}$ [jo'jo' nera]$_{j}$]$_{IHRC}$ niate' si Tiur
[REAL-AV.REAL-chase OBJ dog that COP ART NAME

'The one who the dog chased is Tiur' (OBJ is target of relativisation)

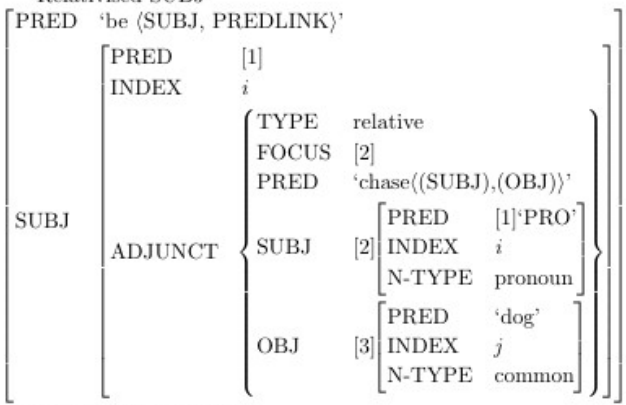
[nganga [si=buru' [si=kau-[ra]$_{SUBJ}$ [tai kebbuk-at-ta]$_{SUBJ}$]$_{REL2}$]$_{REL1}$
[language [REL=old [REL=give-[3PL.NOM] [PL.PR older.sibling-NMLZ-1PL.INCL.POSS

'The old language that our ancestors gave.'

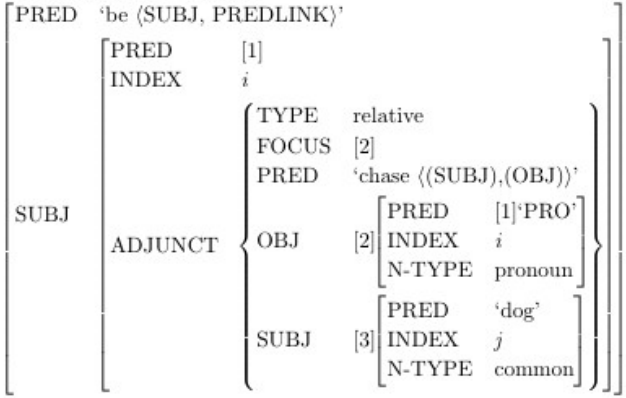
Loss of SUBJ-Only Constraint in Relativisation

- The IHRC is an IP (finite, with REAL mood), which is the SUBJ argument to a PREDLINK enabled by the copula *niate'*
- The head of the REL clause is a zero form
- Because the head of the clause has no PERS/NUM values, PREDLINK can just easily link to the zero form whether it is SUBJ or whether it is OBJ, and c-str does not constrain this

Relativised SUBJ

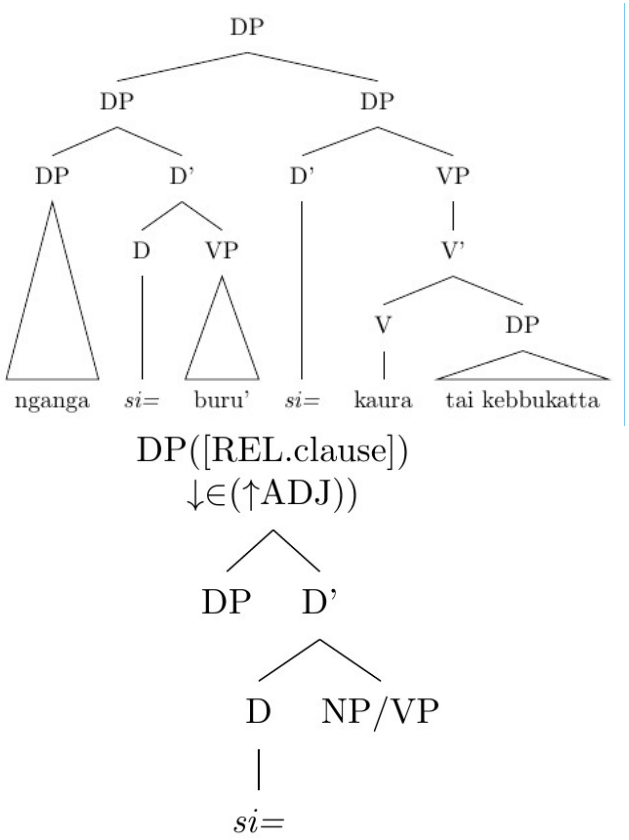


Relativised OBJ - VOS RC



Loss of SUBJ-Only Constraint in Relativisation

- In relative clauses beginning with *si=*, the complement is a VP, which may itself contain pronominal indexing linking outside the IHRC
- Thus, we can generalise the c-str of REL clauses, giving *si=* the label D



Evidence for VP

- Although the NP within a VP may have any GF, as determined by the markedness hierarchy, sentential adverbials cannot intervene in the [V NP] sequence -

*aku masi-gaba [sokat]$_{ADV}$ iba s(i)=abeu
1SGNOM AV.REALlook.for yesterday fish REL=large
aku masi-gaba iba s(i)=abeu [sokat]$_{ADV}$
1SG.NOM AV.REAL-look.for fish REL=large yesterday

‘I looked for fish yesterday.’

Mood as it relates to Voice

- Verbs in SM are IRRealis by default – optional `((↑MOOD)=IRR)`
- This can be overridden by the feature `MOOD=REAL`, contributed either by `ASP.PREFIX` *a-* or any of three *m*-initial voice morphemes, derived from *p*-initial forms (seemingly an affix process reflecting fossilised Proto-Austronesian active voice *-*um*- plus a dispreference for dissimilar labials deleting the initial syllable (Blust 2013, 2022)
- All stems on which pronominal indexing occurs carry the non-realis-marked *p*- forms, creating an association between co-indexing and IRR in the derived verbs. But pronominal indices do not themselves carry mood features

	STAT	AV.ITR	DIST	AV
REAL	*ma-*	*mu-*	*maN-*	*masi-*
IRR		*pu-*	*paN-*	*pasi-*

Mood as it relates to Voice

ra-matei-ake’ sikoinan
3PL.NOM-dead-CAUS crocodile

‘They will kill the crocodile.’ No TAM marking: IRR

a-ra-matei-ake’(-an) sikoinan
REAL-3PL.NOM-dead-CAUS(-PERF) crocodile

‘They (have) killed the crocodile.’ *a-*: REAL

Firah i-pu-urai sabbat saraina-nia
NAME 3SG.NOM-AV.ITR-sing with sibling-3SG.POSS

‘Firah will sing with her sibling.’ No TAM marking: IRR

ai nia mu-urai
AUX 3SG AV.ITR.REAL-sing

‘S/he singing.’ *m*-initial voice form: REAL

Conclusions

	Co-Indexing Affixes	**Voice Affixes**
Properties	Show both agreement and referential properties, prohibiting argument-GF alternations.	Function solely as argument-SUBJ selector markers, allowing alternations
Position/Function	In SM, may bear DF by functioning as an argument anaphorically linked to DF. May bear SUBJ where the DP provides PRED value and index is therefore agreement affix.	Never display pronominal properties.
Actor Regulation	SM pronominal indices such as **i-** require ACTOR to be SUBJ, like AV **pasi-/masi-**	Never serve as arguments themselves.
Comparison with Balinese	Balinese **a/=a** allows ACTOR to be OBJ, while SM pronominal indexing prefixes forbid this	AV morphology In SM no longer alternates the GF of Actor as in symmetrical-voice languages, e.g. Balinese

Acknowledgments

This research has been supported by the Australian Research Council grant *Languages of Barrier Islands, Sumatra: Description, Typology, and History* (DP230102019). Emma Keith was supported by an Australian Government Research and Training Program scholarship. Thanks are due to Sumario Tatubeket and Arkelaus Saogo for sentences and grammaticality judgments in Sipora Mentawai.

Australian National University

Australian Government
Australian Research Council

Bibliography

Arka, I W. 2003. *Balinese Morphosyntax: A lexical-functional approach.* Canberra: Pacific Linguistics. **Artawa, I K.** 1998. “Ergativity and Balinese Syntax.” *NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia*, 42. **Billings, B. and McDonnell, B.** 2024. “Sumatran.” *Oceanic Linguistics* 63.1:112-74. **Blust, R.** 2013. *The Austronesian Languages.* Canberra: Asia-Pacific Linguistics. **Blust R.** 2023. “Repair and drift in Austronesian languages: Avoidance of dissimilar labials as the onsets of successive syllables.” *Oceanig Linguistics* 62.2:304-23. **Bresnan, J. Asudeh, A., Toivonen, I., Wechsler, S.,** 2015. *Lexical Functional* Syntax. 2nd edition. Oxford: Wiley-Blackwell. **Brown, L.** 2001. *A Grammar of Nias Selatan.* PhD thesis, University of Sydney. **Bonneau, J.** 1992. *The Structure of Internally-Headed Relative Clauses: Implications for Configurationality*. PhD thesis, McGill University. **Gil, D.** 2015. “The Mekong-Mamberamo linguistic area” In Enfield, N. J. and Comrie, B. *Languages of Mainland Southeast Asia: The state of the art*, 266-355. Canberra: Pacific Linguistics. **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. **Lenggang, Z. H.R, Nio, B. K. H., Zainil, M. A., and Adam, S.** 1978. *Bahasa Mentawai*. Jakarta: Badan Pembinaan dan Pengembangan Bahasa. **Riesberg, S.** 2014. *Symmetrical Voice and Linking in Western Austronesian Languages*. Canberra: Pacific Linguistics. **Ross, M.** 2002. “The history and transitivity of Western Austronesian voice and voice-marking” In *The history and typology of Western Austronesian voice systems,* edited by F Wouk and M Ross, 17-62. Canberra: Pacific Linguistics. **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.