

## Three ways to “steal” an element from a CP: Evidence from Formosan

Victoria Chen & Shin Fukuda  
University of Hawai'i at Mānoa

**Introduction:** Three Formosan languages, Puyuma, Amis and Seediq, all exhibit what appear to be instances of Raising-to-object-out-of-CP (ROC) (1)-(3). ROCs in these languages all involve the “raised” phrase (i.e. the XP) thematically belonging to an embedded finite CP that *optionally* appears in the matrix domain. Further, in all three languages, ROCs unrestrictedly apply to all knowledge/perception verbs that select a finite CP.

- (1) aparu=ku i Arasip<sub>i</sub> [dra m-uka ec<sub>i</sub> i Senden]. PUYUMA  
AV.forget=1SG.ABS OBL.LOC Arasip<sub>i</sub> [ C AV-go ec<sub>i</sub> ABS Senden]  
'I forget that Senden has been to Arasip.'
- (2) ma-fana' kaku tuna wacu<sub>i</sub> [Ø ma-palu ni wawa (cingra)<sub>i</sub>]. AMIS  
AV-know 1SG.ABS OBL.that dog<sub>i</sub> [ C PV-beat ERG child (it.ABS)<sub>i</sub>]  
'I know that the child beat that dog.'
- (3) spi-an Hana ka Watan<sub>i</sub> [Ø s<m><n>ipaq laqi=na<sub>i</sub> ]. SEEDIQ  
dream-LV Hana.ERG ABS Watan<sub>i</sub> [ C <AV><PST>beat child=3SG.POSS.(OBL)]  
'Hana dreamt that Watan beat his child.'

In this paper, we first show that, ROCs in these languages impose distinct restrictions on how the XP is associated with the finite CP. We then propose that the observed microvariation is best accounted for by analyzing the ROCs as cases of embedded left dislocations with three independently motivated strategies: (i) concatenation of XP and a propositional CP (4a), (ii) coindexation between XP and an operator base-generated inside a predicative CP (4b), and (iii) coindexation between XP and an operator that undergoes A'-movement inside a predicative CP (4c) (e.g., Aissen 1992; Culicover & Jackendoff 1999; Landau 2011).

- (4) a. V .... XP [CP ..... ]: PUYUMA  
b. V .... XP<sub>i</sub> [CP Op<sub>i</sub> C ... pronoun<sub>ABS</sub><sub>i</sub>.... ]: AMIS  
c. V .... XP<sub>i</sub> [CP Op<sub>i</sub> C ... t<sub>i</sub>.... ]: SEEDIQ

**Similarities:** ROCs in all languages involve a finite CP and the “raised” XP that is in the matrix domain. The finite CP analysis is motivated by the unrestricted aspect marking (5) and embedded voice type ((2), (3), (6), (13)). The status of the XP as a matrix element is suggested by (i) word order (5)-(6), in which the XP precedes both overt C and matrix elements, (ii) unambiguous reflexive binding (6), and (iii) the Case licensing of the XP ((2),(5),(10)), which is inconsistent with the XP being an embedded element.

- (5) ma-ladram kan Isaw<sub>i</sub> i Senden [\*(dra) adri d<em>a-deru ec<sub>i</sub> dra bujir]. PUYUMA  
AV-know OBL Isaw<sub>i</sub> ABS Senden [ C NEG <AV>PROG-cook ec<sub>i</sub> OBL taro ]  
'Senden knows that Isaw is not cooking taros now.'
- (6) ma-lemed ni Kulas<sub>i</sub> cingra<sub>i</sub> inacila [Ø ma-palu ni Mayaw<sub>k</sub> ec<sub>i</sub>]. AMIS  
PV-dream ERG Kulas<sub>i</sub> himself.ABS<sub>i</sub> yesterday [ C PV-beat ERG Mayaw<sub>k</sub> ec<sub>i</sub>]  
'Yesterday Kulas<sub>i</sub> dreamt that Mayaw<sub>k</sub> beat him<sub>i/\*k</sub>'

**How they differ:** First, ROCs in Amis and Seediq require that the XP be identified with an embedded ABS phrase (the *ABS-only* constraint) ((7), (8)) while such restriction is absent in Puyuma ((1),(9)), where the XP can be identified with any embedded element from ABS (5), ERG, OBL (1), and adverbial (9).

- (7) \*ma-tawal aku i Kalingku<sub>i</sub> [Ø tayra ec<sub>i</sub> ci Lisin]. AMIS  
PV.forget 1SG.ERG ABS.LOC Kalingku<sub>i</sub> [ C go.AV ec<sub>i</sub> ABS Lisin]  
'(I forget that Lisin has been to Kalingku.) (cf. (1))
- (8) \*s<um>hong<sub>i</sub>=ku Skangki<sub>i</sub> [Ø me-n-sa ec<sub>i</sub> ka Ikung]. SEEDIQ  
AV.forget=1SG.ABS Skangki.OBL<sub>i</sub> [ C AV-PFV-go ec<sub>i</sub> ABS Ikung]  
'(I forget that Ikung has been to Skangki.) (cf. (1))
- (9) ma-ladram=ku an miranang na birā<sub>i</sub> [\*(dra) wa-ruma=yu ec<sub>i</sub>] PUYUMA  
AV-know=1SG.ABS when be.yellow.AV DF.ABS leaf<sub>i</sub> [ C IRR.AV-go.back=2SG.ABS ec<sub>i</sub>]  
'I know that you will be back when the leafs turn yellow.'

Second, Seediq ROCs obey islands (10), while Puyuma and Amis ROCs do not (11)-(12).

- (10) \*q<um>pahang=ku Imin<sub>i</sub> [Ø kela-un=su [ka kari shelisun [Ø q<em>iyut babuy ec<sub>i</sub>]]].  
 AV-hear=1SG.ABS Imin.OBL<sub>i</sub> [C know-PV=2SG.ERG [ABS anecdote C AV-bite pig.OBL ec<sub>i</sub>]]  
 ‘I heard that you know the anecdote that Imin bit pigs.’ [SEEDIQ] [Complex NP island]
- (11) kilengaw=ku kan Isaw<sub>i</sub> [dra ma-ladram=yu [kana kasaerueru [dra tr<em>ima kana le’u ec<sub>i</sub>]]].  
 AV-hear=1SG.ABS DF.OBL Isaw<sub>i</sub> [C AV-know=2SG.ABS [DF.OBL anecdote C AV-buy DF.OBL owl ec<sub>i</sub>]]  
 ‘I heard that you know the anecdote that Isaw bought the owl.’ [PUYUMA] [Complex NP island]
- (12) ma-fana’ ci Kulas tuna wacu<sub>i</sub> [Ø t<um>angic kaku, [anu ma-patay (cingra)<sub>i</sub>]].  
 AV-know ABS Kulas OBL.that dog<sub>i</sub> [C AV-cry 1SG.ABS [if AV-die (it.ABS)<sub>i</sub>]]  
 ‘Kulas knows that I will cry if that dog dies.’ [AMIS][Adjunct island]

**Proposal:** We analyze ROCs in the three languages as cases of embedded left dislocations that involve the three strategies identified in (4). We argue that the left-dislocated phrase (XP) is base-generated in all languages. Puyuma and Amis ROCs’ immunity to islands (11)-(12) follows from this analysis. The lack of reconstruction effects in Seediq (13) also shows that the XP has not undergone a syntactic movement.

- (13) a. kela-un=mu [Ø qelu-un bi de-bubu ka laqi=teha ]. SEEDIQ  
 know-PV=1SG.ERG [C love-PV very all-mother ABS child=3PL.POSS]  
 ‘I know that all mothers<sub>i</sub> love their children<sub>i/k</sub>.’ (the bound variable reading available).
- b. kela-un=mu ka laqi=teha<sub>i</sub> [Ø qelu-un bi de-bubu ec<sub>i</sub>].  
 know-PV=1SG.ERG ABS child=3PL.POSS<sub>i</sub> [C love-PV very all-mother ec<sub>i</sub>]  
 ‘I know that all mothers<sub>i</sub> love their children\*<sub>i/k</sub>.’ (the bound variable reading unavailable)

The unrestricted relationship between the XP and the CP in Puyuma is captured under the analysis that the embedded CPs are *positional CPs concatenated with the XP* (4a). For Amis and Seediq, we propose that *the XP is coindexed with a null operator (Op) inside a predicative CP*. The *ABS-only* condition on the XP comes from the fact that *Op* must be identified with an ABS phrase in both languages. The divergence in island-sensitivity between the two is further accounted for as following: In Amis, *Op* is base-generated as an embedded topic in [Spec CP] and unselectively binds an embedded ABS DP (4b). Crucially, topics in Amis are observed to unselectively bind any ABS DP regardless of syntactic locality. Postulating *Op* as a base-generated topic inside the CP thus accounts for the lack of locality constraint (12) while maintaining the *ABS-only* constraint. This analysis is supported by the optional overt embedded pronominal copy in ROCs ((2), (12)) as well as in topicalization, while traces in A-/A’-operations are never spell-out as pronouns in the same dialect (see also Wu’s 2000 description of Central Amis). In Seediq, *Op A’-moves to [Spec CP]* (4c) and hence obeys island constraints (10) (no resumptive pronouns are permitted). Together, these languages demonstrate how an embedded CP and a left dislocated XP are associated via three distinct strategies (4). Last, we argue that the matrix behavior of the XP is due to its status as a *base-generated adjunct to the CP*. Based on novel data from the same languages, we argue with Chung (1991, 1994, 1998) and Rackowski & Richards (2005) that *CPs receive structural Case*, and present evidence that the XP inherits the Case assigned to the CP in ROC constructions. The word order in which an XP and a CP can be separated by a matrix element under certain circumstances (e.g., (5)) is accounted for based on independently motivated assumptions that (i) CPs in Formosan languages extrapose (nearly) obligatorily, and (ii) ABS-arguments and TPs undergo phrasal movement (Aldridge 2004).

**Implications:** The proposed analysis adds ROCs in the three Formosan languages to the growing list of languages with ROCs without true “raising” out of CP, and provides support for the notions that restrict the domains in which grammatical operations operates (PIC) and how they proceed (the ban on improper movement). The macrovariation in ROCs among these three Formosan languages also illustrates how closely related languages utilize different strategies to “steal” an embedded element from a finite CP. Last, we discuss further implications of the findings from ROCs in these languages cross-linguistically with previous findings from the relevant literature, especially Massam (1985).

**Selected References:** AISSÉN, J. 1992. Topic and focus in Mayan. *Language* 68:43-80. MASSAM, D. 1985. Case theory and the projection principle. Ph.D. dissertation, MIT. LANDAU, I. 2011. Predication vs. aboutness in copy raising. *NLLT* 29: 779-813.