

A guided tour of Austronesian clitics

Phonology, morphology and syntax

Daniel Kaufman
Queens College, CUNY & ELA
Daniel.Kaufman@qc.cuny.edu

1 Basic geography

- Second-position pronominal and adverbial enclitics are present in every **Philippine language** to some extent. Their distribution has led to suggestions that such patterns be reconstructed to Proto-Malayo-Polynesian, the ancestor of all Austronesian languages outside of Taiwan.
- Outside the Philippines, we find far greater diversity in clitic types and positions.
- In **Sulawesi**, the 2P pattern for nominative clitics and adverbs is generally continued but genitive/ergative clitics become verb-adjacent proclitics. In languages of Southeast Sulawesi, more complex agreement patterns develop where both arguments are indexed by verbal morphology.
- In **Borneo**, the Philippine pattern is continued in northern subgroups like Dusunic but generally lost in other subgroups. Most Bornean languages seem to have lost 2P clitics and many have moved towards a more isolating type of morphology, in contrast to Sulawesi.
- In **Sumatra**, 2P clitics are rare but we find exactly the same development of genitive/ergative becoming verbal proclitics as seen in Sulawesi.
- **Eastern Indonesia/West Papua** is a very mixed bag. There are few generalizations that can be made except that they have all moved away from the Philippine system significantly. Overall, 2P pronominal clitics seem to be very rare here. Pronominal clitics have become agreement markers of different types.

2 Phonology

- What is a clitic? Modern thinking on clitics owes much to the work of Arnold Zwicky (Zwicky 1985, 1977) although an enormous amount of work precedes and follows (see Spencer and Luís 2012 and Anderson 2005 for recent reviews).

- (1) CLITIC DIAGNOSTICS (Halpern 1995:14)
 - a. being lexically stressless/accentsless
 - b. occupying one of a characteristic set of positions (2P, verb-adjacent, etc.).

- The notion that clitics are phonologically dependent led to their common treatment as “prosodic affixes”, e.g.

(2) PROSODIC SUBCATEGORIZATION FRAME (Inkelas 1989, Zec & Inkelas 1990)

]ω __]ω

- Philippine languages show very clearly that syntactic dependency is not contingent on prosodic dependency. Most (but not all) disyllabic clitics in Tagalog typically have homophonous non-clitic counterparts.

“While there may be languages for which a phonological account of the non-occurrence of clitics in phrase-initial position is possible, no well supported analyses of this kind have been presented. And in fact it is extremely unlikely that such a prosodically based account will be adequate in general. That is because some special clitics that must be placed post-initially are not prosodically deficient. Tagalog, for instance, has a huge system of clitics, most of which are prosodically autonomous and bear their own stress. There seems no phonological reason why these could not occur initially, and if they do not, that fact must be due to some other constraint.” (Anderson 2005:141)

- This can be seen with the following disyllabic clitics *kami* and *sána* in Tagalog.

- (3) a. Kami ay nag-lútò ng=ampalaya b. Nag-lútó=kami ng=ampalaya
 1P.NOM TOP AV.PRF-cook GEN=bitter.melon AV.PRF-cook=1P.EX.NOM GEN=bitter.melon
 ‘We cooked bitter melon’ ‘We cooked bitter melon’
- (4) a. Sána ay mag-lútó=sila b. Mag-lútó=sána=sila
 OPT TOP AV-cook=3P.NOM AV-cook=OPT=3P.EX.NOM
 ‘Hopefully, they will cook.’ ‘Hopefully, they will cook.’

	CLITIC	FREE
aspect	= <i>na</i> ‘already’	∅
	= <i>pa</i> ‘still’	∅
focus	= <i>din</i> ‘also’	∅
	= <i>man</i> ‘even’	∅
	= <i>naman</i> ‘switch topic’	(<i>naman</i>)
	= <i>ngà</i> ‘emphasis’	∅
	= <i>lang</i> ‘only’	∅
	= <i>lámang</i>	<i>lámang</i>
	= <i>talaga</i> ‘emphasis’	<i>talaga</i>
politeness	= <i>pò</i> , = <i>hò</i> ‘politeness’	∅
mood	= <i>pala</i> ‘surprise’	∅
	= <i>yàtà</i> ‘perhaps’	∅
	= <i>sána</i> ‘hopefully’	<i>sána</i>
	= <i>náwa</i> ‘hopefully’	<i>náwa</i>
	= <i>ba</i> ‘question marker’ (= <i>baga</i>)	∅ (<i>baga</i>)
	= <i>daw</i> reported speech	∅

Trad. labels	Gloss	Features	NOM	GEN	NOM	GEN	OBL
			CLITIC		FREE		
1 st sing.	1S	[1]	=ako	=ko	ako	ákin	sa ákin
2 nd sing.	2S	[2]	=ka	=mo	ikaw	iyo	sa iyo
3 rd sing.	3S	[∅]	=siya	=niya	siya	kaniya	sa kaniya
1 st excl. pl.	1+3	[1, p]	=kami	=námin	kami	ámin	sa ámin
(1 st dual)	1+2	[1,2]	=kata/kita	=ta	kata/kita	kanita	sa kanita
1 st incl. pl.	1+2P	[1,2,p]	=táyo	=nátin	táyo	átin	sa átin
2 nd pl.	2P	[2,p]	=kayo	=ninyo	kayo	inyo	sa inyo
3 rd pl.	3P	[∅,p]	=sila	=nila	sila	kanila	sa kanila
			Portmanteau forms: [1.GEN+2.NOM] =kita, kita				

- Prosodic weakness is often equated with prosodic dependency but this need not be the case. I argue that clitics *are* prosodic words, but *still deficient* compared with lexical words. Much like Wackernagel's (1892) original treatment of the Germanic verb: a phonological/prosodic word but one that had “lost its accent”.

2.1 Prosodization of clitics

- A number of phonological phenomena (vowel lowering, glottal deletion, tapping) support a prosodic structure as in Figure 1.
- Monosyllabic and disyllabic clitics are adjuncts to prosodic words and prosodic phrases.
- Enclitics form a single prosodic phrase with their host but proclitics adjoin to the left edge of the prosodic phrase.
- There thus exists a prosodic phrase boundary between proclitics and their hosts.

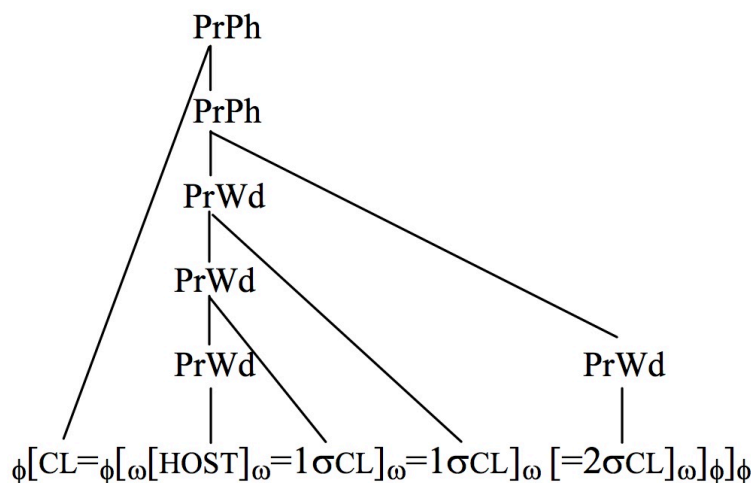


Figure 1: Prosodization of clitics in Tagalog

2.2 Intonation

- How does the prosodic structure posited earlier effect intonation?
- Initial edge tones are anchored to the left edge of either a maximal or embedded prosodic phrase. Final edge tones are anchored to the right edge of the maximal prosodic phrase. When a phrase begins with functional heads (case markers, determiners, plural markers) there is some flexibility as to where the initial edge tones dock.
- The following example shows how an initial PPh edge tone docks to the plural marker *maŋa* rather than the following lexical word. (This seems to be a significant difference with at least English.)

- (11) [maŋa ba:ta: ŋa: pala sila]
 /maŋa=ba:taʔ=ŋaʔ=pala=sila/
 PL=child=EMPH=MIRA=3P.NOM
 ‘They are really children!’

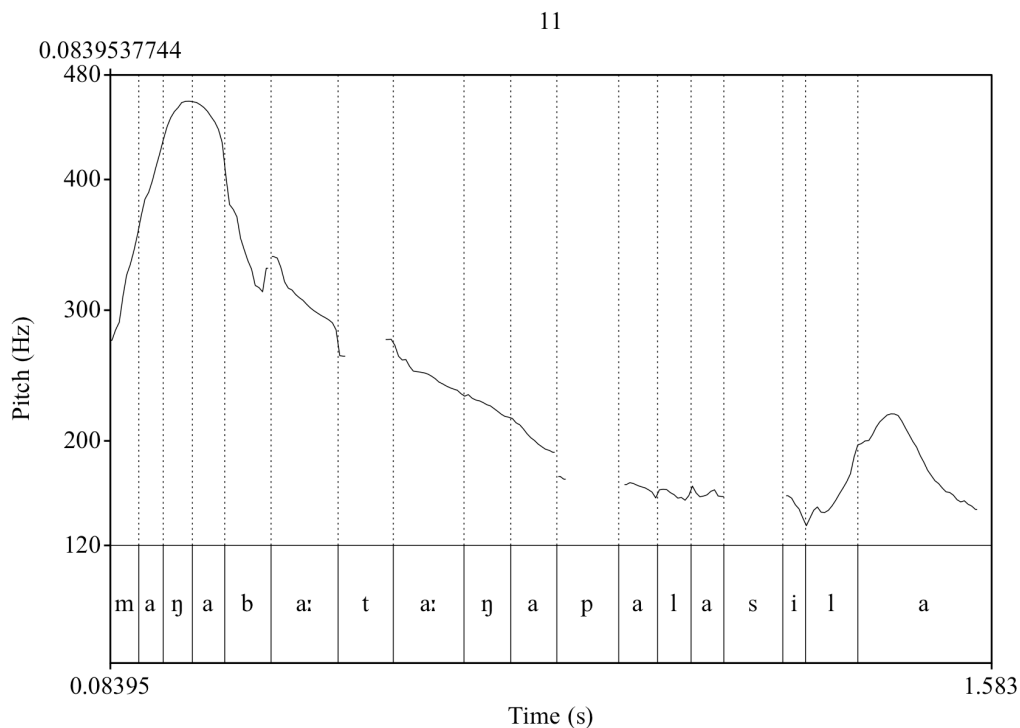


Figure 2: Enclitics and intonation

3 Morphology: cluster-internal ordering

- All Philippine languages allow clustering of pronominal and adverbial clitics (although almost all languages impose some restrictions on what pronominal clitics can co-occur).

- Every imaginable factor can play a role in determining the order of clitics within a clitic cluster.
- Syllable count is the primary factor in Tagalog. Within the 1σ and 2σ domains, case and scope determine ordering.
- Figure 3 shows the ordering of adverbial clitics in the clitic cluster.

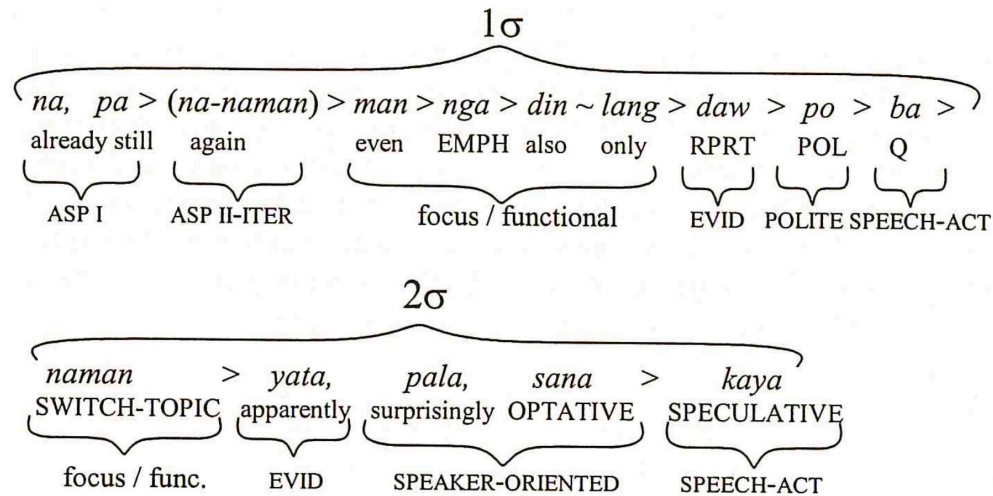


Figure 3: Cluster-internal clitic ordering in Tagalog

- The mere role of phonology in determining clitic order contravenes the fundamental premise of “late-insertion” models of morphology.
- That this is an active constraint in the grammar is clear from the positioning of new adverbial clitics, which are also positioned according to syllable count, e.g. (12).

(12) hindi:=na:=sila=sigu:ro d<um>atiŋ
 NEG=EMPH=3P.NOM=maybe <AV>arrive
 ‘Maybe they really didn’t arrive.’

- A major question with regard to scope: Do all 2P clitic clusters respect scope relations in a left-to-right manner?
- Garifuna, a VSO Arawak language, shows identical ordering of clitic functions, e.g.
 TENSE<EVIDENTIAL<SPEECH ACT

(13) Ka=**ba**=**funa**=**san** a-yanu-ha Garifuna n-uma?
 who=FUT=SPEC=QM VRB-speak-VRB Garifuna 1s.G-with
 ‘(I wonder) who will speak Garifuna with me?’ (Andy Palacio, *Amunegu*)

- When case comes into play in clitic ordering, it always positions genitive clitics before nominative ones, reflecting the unmarked order of arguments in conservative MP languages.
- This case-based ordering principle can also be interpreted as Agent>Patient, even though it applies more widely (experiencers, possessors, etc.). In certain Formosan languages, thematic

(20) *Old Malay*

- a. ni-galar-ku
PV-title-1SG.GEN
'I titled (him).' (Karang Brahi r.9, 14-15, Kota Kapur r. 4,8)
- b. ni-minuŋ-ña
PV-drink-3SG.GEN
'He drank (it).' (Talang Tuwo r. 5)

- In an interesting case of degrammaticalization, the proclitic agent was reanalyzed as an independent syntactic position.
- Kaufman 2014 argues that the “wedge” in this case was the use of kin titles in place of pronouns for politeness. Eventually, these titles were modified to further increase deference (e.g. ‘slave’s master’ for 2nd person) and at this point there was little evidence that the titles occupied a verb-adjacent clitic position.

(21) telah **hamba**^[1]=ampun-i=lah dosa dan ke-salah-an meréka itu
already slave=forgive-APPL=EMPH sin and NMLZ-wrong-NMLZ 3PL that
'I (slave) have already forgiven their sins and errors.'
(Bayan Budiman 214:27)

(22) Hamba^[1]=lah bayan yang **tuan hamba**^[2]=pelihara-kan dahulu itu
slave=EMPH parrot RELT master slave=care.for-APPL earlier that
'It was I (slave) who was the parrot you (slave’s master) cared for at the time.'
(Bayan Budiman 14:26)

- In more formal registers of Indonesian, there is still a robust 1/2 vs. 3 distinction in the use of titles. These can only refer to 1st and 2nd persons.

(23) “IMPOSTER” PROCLISIS PARADIGM

- a. Mana yang akan bapak^[2]=pilih?
which RELT FUTURE father=choose
'What will sir/father (you) choose?'
- b. Mana yang akan bapak^[1]=pilih?
which RELT FUTURE father=choose
'What will sir/father (I) choose?'
- c. Mana yang akan di-pilih bapak?
which RELT FUTURE PV-choose father
'What will sir/father (he) choose?'

- More exotic clitic positions are attested in eastern Indonesia. One of the most bizarre is found in Manggarai, an SVO language with clause-final pronominal clitics that double the subject! (Possibly the grammaticalization of a right-dislocated subject.)

- (24) MANGGARAI
- a. Hia pa'u eta mai bubuŋ mbaru hitu=i
 3s fall above from top.roof house that=3s.NOM
 '(S)he fell down from the top roof of that house.' (Arka & Kosmas 2005:90)
- b. Latuŋ hitu cero l=aku=i
 corn that fry by=1s=3s.NOM
 'The corn is (being) fried by me' (Arka & Kosmas 2005:95)

5 Theoretical considerations

- Three interesting theoretical topics:
 - Clitic coercion: Full phrases that are dragged into clitic positions (and vice versa).
 - Defining 2P: What counts as the host? A prosodic word? A syntactic phrase?
 - Motivating 2P and its typology: Phonology? Morphology? Syntax? Semantics? Discourse factors?
- The idea that clitics and phrasal syntax belong to two different worlds (e.g. morphology and syntax) cannot be correct, as there is plenty evidence for significant interaction.
- In the following types of sentences, a clitic appears in an unambiguous clitic position (in this case, between negation and the verb) but is coordinated or modified by a phrase. Not all such examples are judged as perfectly grammatical but I have collected many naturally occurring examples, many of which are judged so.

- (25) hindi[=ko at naŋ=maŋa=kasamahan=ko=ŋ Filipina] kailanman s<in>irà
 NEG=1S.GEN and GEN=PL=colleague=1S.GEN=LNK Filipina ever <PV.PRF>destroy
 aŋ=tiŋin naŋ=iba sa=ámin
 NOM=view GEN=other OBL=1P.GEN
 'Neither I nor my Filipina colleagues ever destroyed the view of others towards us.' (Kaufman 2010b:26)

- (26) Saan[=kayo=ŋ lima] nag-tuloy matápos kayo=ŋ maka-babà
 where=2P.NOM=LNK five AV.BEG-continue after 2P.NOM=LNK AV.NVL-descend
 naŋ=Bus?
 GEN=bus
 'Where did you five continue to after getting off the bus?' (Kaufman 2010b:34)

- As has now been discussed by several authors (Spencer and Luís 2012; Anderson 2005), different types of clitics may define 2P differently within a single language:

“The abiding impression is that Tagalog clitics are sensitive to very specific aspects of specific grammatical constructions. It may well be, of course, that with sufficient ingenuity we might unearth a small set of simple, abstract principles governing such behaviour, but it is rather more likely that systems such as this are irreducibly idiosyncratic and require

direct reference to specific constructions or to a relatively unstructured list of component grammatical properties.” (Spencer and Luís 2012:176)

- (27) [Bukas=**ba** naŋ=gabi] ay [sa~sayaw=**sila** naŋ=pandango]?
 tomorrow=QM GEN=night TOP AV-INCM~dance=3P.NOM GEN=fandango
 ‘Tomorrow night, will they dance a fandango?’ (Schachter and Otnes 1982:429)

- In fact, it’s not as terrible as Spencer and Luís (2012) make it out to be. The scope of the clitic (e.g. phrase, clause, sentence) determines its placement and...
- the clitic must be visible to the constituent in which its positioned. That is, it can’t be embedded in a sub-constituent, as in (28-b), where =ka is embedded within a fronted oblique phrase.
- This may not be too different from how adverbs with clausal scope are positioned in languages like English, as shown in (29).

- (28) DETERMINING 2P IN TAGALOG: PREDICATES VS. FRONTED FOCUS PHRASES

- a. kaníno=**ka**=ŋ anak?
 OBL.who=2S.NOM=LNK child
 ‘Whose child are you?’
- b. kaníno[*=**ka**]=ŋ anak[=**ka**] nag-bigay naŋ=pérà?
 OBL.who=2S.NOM=LNK child=2S.NOM AV.BEG-give GEN=money
 ‘To whose child did you give money?’

- (29) _{IP}[(Perhaps) we (✓) will (✓) buy (✓) _{DP}[another (*) fancy (*) pair (*) of (*) shoes]] (✓).

- The domain of a clitic is determined both by its semantic scope and syntax. In most Philippine-type languages, genitive clitics are bound within DP, that is, they cannot cross the edge of DP. But in those languages where there is no morphological signal for the DP edge, as in Chavacano, the boundary itself disappears.

- | | | | |
|------|--|---|--|
| (30) | CEBUANO | TAGALOG | CHAVACANO |
| a. | Unsa’y gi-buhat= nila ?
what=INDEF.NOM PV.PRF-do=3P.GEN
‘What did they do?’ | b. Ano ang g<in>awa= nila ?
what NOM <PV.PRF>do=3P.GEN
‘What did they do?’ | c. Kósa= silá ya-asé?
what=3.P PST-do
‘What did they do?’ |

- Much of clitic typology can be summed up in three principles:

- (31) PROSODIC ASYMMETRY GENERALIZATION

Rightwards prosodic attachment is coerced by direct morphosyntactic constituency.

- (32) EDGE ASYMMETRY GENERALIZATION

Syntactic displacement of a clitic from the edge of its host only occurs on the host’s left boundary

- (33) SYNTACTIC DISPLACEMENT CONSTRAINT (Kaufman 2010:128)

Unambiguous (branching) heads of phrases are never displaced to 2P

Phrase	WP]	[XP]						[YP
Host position	A=1	2=B=3	4=...=5	6=C=7	8=D			
<i>sister</i>	✓	✓	✗	✗	✗	✗	✓	✗
<i>non-sister</i>	(✓)	(✓)	✓	✗	✗	✗	✓	✗

Figure 4: Clitic typology (Kaufman 2010)

- “Ditropic clitics” (Type 1 in Fig. 4), or enclitic prepositions, are borne of a need to begin a prosodic phrase with a prosodic word head (PrWd_{Hd}), which is typically associated to a lexical word rather than a functional word.

(34) LIMOS KALINGA

- a. Nan-anup dadit tagu=t bolok
 AV.PRF-hunt PL.NOM person=OBL pig
 ‘The people hunted pig.’ (Ferreirinho 1993:12, Kaufman 2010b:154)

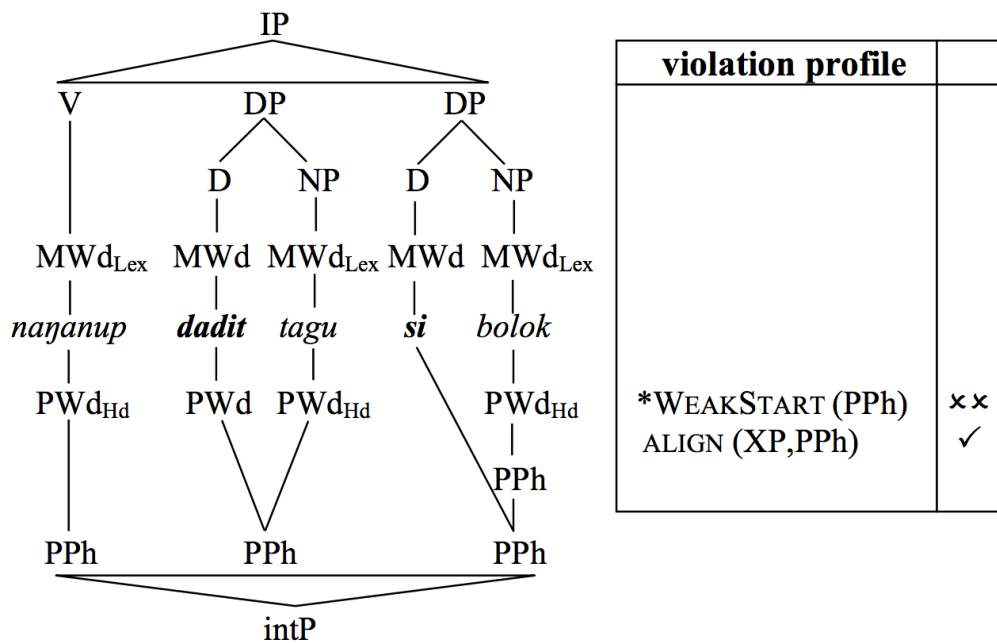


Figure 5: Well aligned clitics (Kaufman 2010b:156)

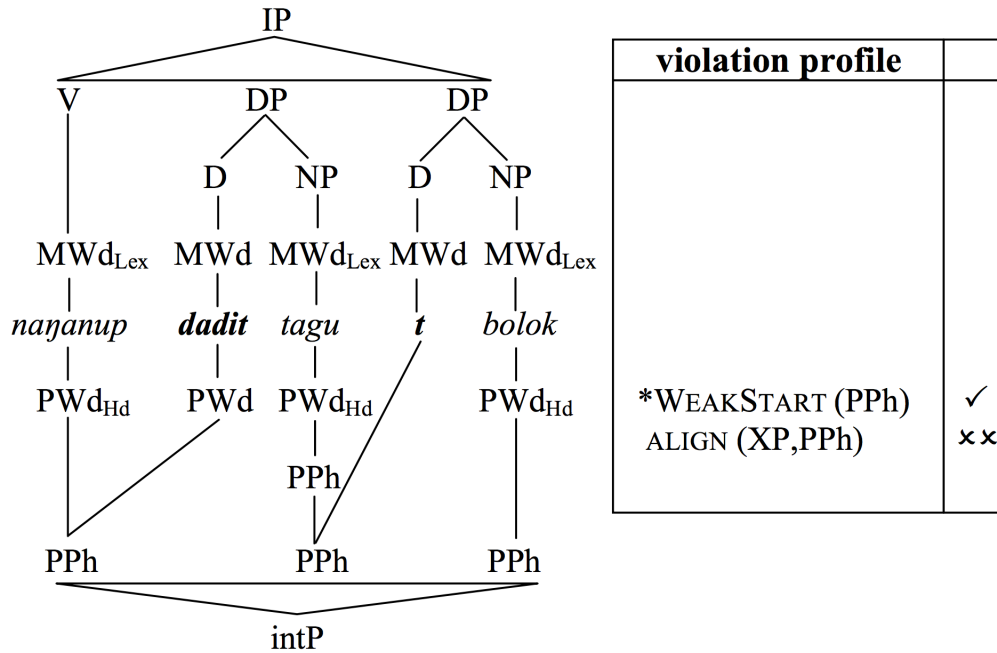


Figure 6: Misaligned (ditropic) clitics in Limos Kalinga (Kaufman 2010b:156)

- Himmelmann (2014) argues convincingly that ditropic clitic behavior stems from:
 - Processing factors:** function words are easily activated and may precede lexical choices, and
 - Discourse factors:** certain types of function words can act as floor-keeping devices.
- The items that become ditropic clitics are delimited by Himmelmann thus:

“The basic idea is that the relevant set of function words has to (i) be preposed and project a specific construction with a specific target that signals the possible syntactic completion of the unit thus projected. Such constructions could be called target-specific constructions.” (Himmelmann 2014:931)
- Where does this leave “pure” prosodic factors?
- It seems to me that pronominal and adverbial clitics could act as excellent floor-keeping devices but they never behave as ditropic 1P clitics. Ditropic clitics are overwhelmingly case markers, prepositions, coordinators and subordinators.

“Although a change in position may occur in grammaticization processes (e.g. a preposed function word becomes a postposed one), it is generally the case that grammatical elements do not change position relative to their hosts once they have reached the clitic stage. Thus, proclitics usually become prefixes and enclitics usually become suffixes (see also Bybee et al. 1990:3, passim).” (Himmelmann 2014:931)
- This is true for *branching heads* but not true for non-branching heads (pronominal clitics, adverbial clitics), even though pronominal clitics give an excellent preview of a following predication (by telling us how many arguments there are and what features they have).
- We still need a division between branching morphosyntactic heads and non-heads. Branching heads are fixed in their position and may become ditropic for the reasons posited by Himmel-

mann (2014) as well as prosodic reasons. But 2P clitics avoid phrase-initial position solely for prosodic reasons and they do this in a different way.

- Finally, how is the class of clitics determined if they are not phonologically *dependent*? In Philippine languages, they are a functional class defined by their “weak” semantics. This can be seen clearly in mixed languages like Kolehiyala English (Bautista 1996) and Chavacano (Forman 1972).

- (35) *Send=you=naman=ako ng picture nina Jodi at Cholo nun*
 send=you=EMPH=1S.NOM GEN picture GEN.PR.PL Jodi and Cholo when
 bata=pa=sila.
 young=still=3S.NOM
 ‘Send me a picture of Jodi and Cholo when they were young.’
- (36) *I-pasa=you=naman sa akin yung novel summary...*
 CV.INF-pass=you=EMPH OBL 1S that.NOM.LNK novel summary
 ‘Pass to me that novel summary...’
- (37) *Call=you=naman=ako*
 call=2SG=EMPH=1S.NOM
 ‘You call me!’
- (38) *Add=mo=me*
 add=2S.GEN=me
 ‘Add me!’

6 Summing up

- Austronesian clitics offer a wealth fascinating data with implications for every corner of grammar. Some of the more burning questions involve acquisition, language contact and details of positioning and prosodization, as well as the historical development of the clitic and agreement systems of Indonesia and beyond.
- The variation, which I have not really been able to convey in this talk, is truly massive.

References

- Anderson, Stephen R. 2005. *Aspects of the Theory of Clitics*. New York: Oxford University.
- Bautista, Ma. Lourdes S. 1996. Notes on three sub-varieties of philippine english. In *Readings in Philippine Sociolinguistics*, ed. Ma. Lourdes Bautista, 94–101. Manila: DLSU Press.
- Billings, Loren, and Daniel Kaufman. 2004. Towards a typology of Austronesian pronominal clisis. In *Proceedings of AFLA 11*, ed. Paul Law, volume 34 of *ZAS Papers in Linguistics*, 15–29. Berlin: Zentrum für Allgemeine Sprachwissenschaft.

- Billings, Loren A. 2005. Ordering clitics and postverbal R-expressions in Tagalog: A unified analysis? In *Verb first: On the syntax of verb-initial languages*, ed. Andrew Carnie, Heidi Harley, and Sheila Ann Dooley, volume 73 of *Linguistik Aktuell*, 303–339. Amsterdam: John Benjamins.
- Forman, Michael. 1972. Zamboangueno texts with grammatical analysis: A study of philippine creole spanish. Doctoral Dissertation, Cornell University.
- Halpern, Aaron. 1995. *On The Placement And Morphology Of Clitics*. Stanford, CA: Center for the Study of Language (CSLI).
- Himmelmann, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90:927–960.
- Kaufman, Daniel. 2009. Austronesian typology and the nominalist hypothesis. In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, ed. Alexander Adelaar and Andrew Pawley, Pacific Linguistics. Canberra: ANU Press.
- Kaufman, Daniel. 2010a. The grammar of clitics in Maranao. In *Piakandatu ami Dr. Howard P. McKaughan*, ed. L. Billings and Nellke Goudswaard, 132–157. Manila: Linguistic Society of the Philippines and SIL.
- Kaufman, Daniel. 2010b. The morphosyntax of Tagalog clitics: A typological approach. Doctoral Dissertation, Cornell University, Ithaca, NY.
- Kaufman, Daniel. 2014. The syntax of Indonesian imposters. In *Cross-linguistic Studies of Imposters and Pronominal Agreement*, ed. Chris Collins, Oxford Studies in Comparative Syntax, 89–120. New York: Oxford University Press.
- Schachter, Paul, and Fe T Otones. 1982. *Tagalog Reference Grammar*. University of California Press.
- Spencer, Andrew, and Ana R. Luís. 2012. *Clitics: An Introduction*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Wackernagel, Jacob. 1892. über ein Gesetz der indo-germanischen Wortstellung. *Indogermanische Forschungen* 1:333–436.
- Zwicky, Arnold. 1977. *On Clitics*. Bloomington: Indiana University Linguistics Club.
- Zwicky, Arnold. 1985. Clitics and particles. *Language* 61:283–305.