

Microvariation in the Clitic Doubling Parameter: Comparative Evidence from the Sambalic Cross-Referent Pronouns¹

Leo Gregory A. Fordan

Abstract

This paper investigates the cross-referent pronoun (CRP) through a comparative study of its distribution in three Sambalic languages: Sambal (ISO 639-3 [*xsb*]), Botolan [*sbl*], and Bolinao [*smk*]. The CRP is a result of clitic doubling (CD), a phenomenon which has been extensively studied in the Romance and Balkan languages, but its manifestation in Philippine languages

¹This paper was originally presented at the 16th International Free Linguistics Conference (FLC) at the University of Santo Tomas, Manila. I'd also like to use this space to thank my language consultants: Tito Igo Campos for Sambal, Ma'am Harriet Villanueva for Botolan, and Ma'am Wena Buban for Bolinao.

remains underexplored. The CRP is defined here as a pronominal-like element that co-refers with a nominal argument within the same clause. Data for this study were gathered through a sentence list elicited in real time with native speakers, followed by judgment tests. My analysis shows that the CRP is subject to distinct language-specific licensing conditions: in Sambal, the CRP is largely licensed for a semantic AGENT having competing arguments; in Bolinao, the CRP acts more like an agreement marker; and in Botolan, what was previously described as a CRP (see Antworth, 1979) is actually a resumptive pronoun. This demonstrates that even closely related languages may nevertheless differ substantially in realizing a common feature, as with a parametric approach. This also suggests that although the CRP can serve a pragmatic function in reference tracking, its variation among the three languages is driven solely by language-internal (i.e., structural) factors such as individual phonologies and morphologies. Ultimately, the findings deepen our understanding of the nature of language as they demonstrate how CRP distribution is shaped not by syntax alone but by the interaction of different components of grammar.

Keywords: Sambalic languages, clitic doubling, cross-referent pronoun, syntax, parameters

1 Introduction

Human languages have long been viewed as being built from the same architecture, behaving only in different ways on the surface. This assumption is put to the test most clearly within closely related language groups where we expect to see greater similarity but are instead confronted by more variation. It is then the task of the linguist to uncover such variation and, in doing so, move us closer to understanding the nature of language itself. This paper is one such effort. Here, we examine a phenomenon in the Sambalic languages where a pronominal form appears alongside a DP, both referring to the same entity. Referred to here as the cross-referent pronoun (CRP), it is an element that may seem purely redundant at first but, upon closer inspection, reveals a systematic distribution across constructions and languages. In bringing attention to this phenomenon, the present study contributes to our ongoing effort to explain the morphosyntax of Philippine languages and, ultimately, to refine linguistic theory through the lens of underdocumented languages such as those in the Sambalic group.

This paper focuses on three (3) Sambalic languages: Sambal (ISO 639-3 [*xsbl*]),² Botolan [*sbl*], and Bolinao [*smk*]. These languages belong to the Central Luzon (CL) microgroup (Figure 1) which is further classified into (a) Kapampangan [*pam*], (b) Dumagat Remontado [*agv*], and (c) the Sambalic branch, to which the Sambal, Botolan, and Bolinao

²Usually called “Tina Sambal” in the older literature. Eberhard et al. (2026) claim that the term is pejorative. However, according to my various experiences with native Sambal speakers, it is merely an exonym. Nonetheless, I have decided against it for economy.

languages belong, along with the Ayta languages: Abellen [*abp*], Ambala [*abc*], Mag-antsi [*sgb*], Mag-indi [*blx*], and Magbukun [*ayt*] (Eberhard et al., 2026).

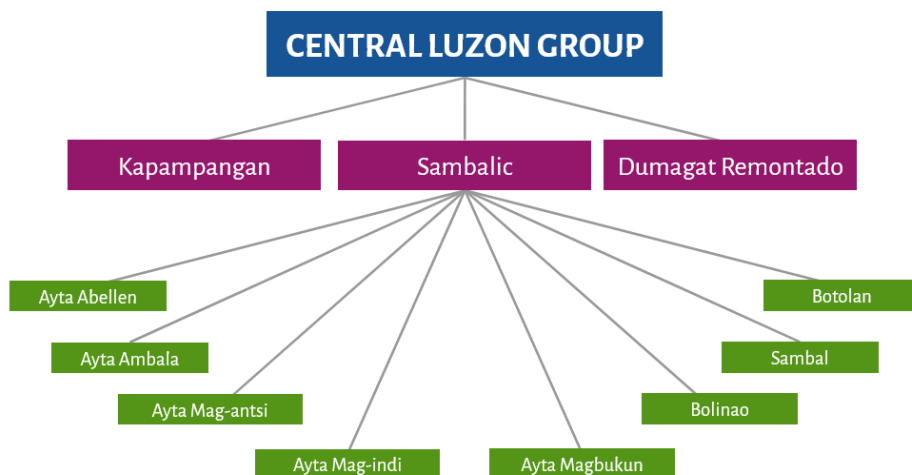


Figure 1. Genetic Relations within the Central Luzon Group³

In the earlier literature, the languages examined in this paper have not always been distinguished and are still usually confused for each other, with previous studies such as Himes (2012), Komisyon sa Wikang Filipino (KWF, 2016), and Viray and Schreerer (1929) having considered all three or at least two as different varieties of a single language. The Ethnologue (Eberhard et al., 2026), however, recognizes them as separate languages and the same is true here.⁴

³Created by the author based on Eberhard et al. (2026).

⁴Despite the high level of mutual intelligibility between Sambal and Botolan, the speakers usually assert that their variety is a different language and this is apparent with the differences even in basic vocabulary. Sambal speakers, for instance, would describe Botolan as a “purer” variety.

The speakers of Sambal are primarily found in five municipalities in northern Zambales: Iba, Palauig, Masinloc, Candelaria, and Sta. Cruz (Elgincolin et al., 1988). Moreover, it is also spoken in the Municipality of Infanta in Pangasinan, the western border of Tarlac Province, and the Panitian village in the Province of Palawan (Eberhard et al., 2026). This paper focuses on its dialect in Palauig. Meanwhile, Bolinao is spoken in the municipalities of Anda and Bolinao in western Pangasinan Province. Lastly, Botolan speakers are found in the municipalities of Botolan and Cabangan in Zambales Province.

Of the three languages, only Botolan is not endangered and is classified as EGIDS Level 5 where some standardized literature is used, although not widespread or sustainable (Eberhard et al., 2026). Meanwhile, Sambal is at Level 7 where it is no longer transmitted by the child-bearing generation, while Bolinao is at Level 6b, where it is losing speakers but still enjoys some use by all generations.

2 Methodology

The data analyzed for this paper were gathered from consultations with one native speaker each of Sambal, Botolan, and Bolinao through an online videoconferencing platform, where all sessions were recorded with consent. Prior to data collection, informed consent was obtained from all consultants through signed consent forms, which explicitly informed them that the data would be used for academic purposes, including conference presentations and potential publication in scientific journals.

For the elicitation tool, I made a 100-sentence list based on the 700+ sentence list of the UP Department of Linguistics, picking

sentences which specifically target possible CRP usage. A translation task using this list was then done in real time with each item followed by a judgment/intuition test whenever deemed necessary. Testing the consultants' judgment of hypothetical utterances—particularly by inserting a potential CRP or moving it around—informed me of the possible combinations and constraints which did not emerge spontaneously during the translation task. The data were transcribed using a practical orthography with the conventions being based on the KWF guidelines, including the use of <ë> to represent a schwa-like vowel.

Every consultant was an adult native speaker of a Sambalic language who acquired it as their first language and continues to use it regularly. While all the consultants are multilingual and also speak Tagalog and English, the data here is supposed to reflect their judgments for their respective languages in isolation, i.e., assuming an idealized speaker-hearer situated in a homogenous speech community.⁵ Therefore, with the focus of this paper being on the grammatical competence underlying CRP distribution, variables such as age, gender, level of education, language contact, etc. were not taken into account in the present study, as they are not directly relevant to the theoretical goals pursued.

⁵I acknowledge, however, that the consultants are multilingual and situated in complex sociolinguistic environments. Nonetheless, their judgments are treated here as reflecting an assumed stable underlying system. Alternative approaches would treat such multilingual and sociolinguistic knowledge as part of linguistic competence itself but the present analysis sets these factors aside in order to focus only on the structural conditions governing CRP distribution.

As for the author: I am a native speaker of Tagalog/Filipino [tɔl/fil] with passive fluency in Sambal (Palauig dialect), which facilitated basic comprehension of the elicited data while still relying on native speaker judgments.

3 Overview of Clitic Doubling

Crystal (2008) defines a clitic as “a form which resembles a word, but which cannot stand on its own as a normal utterance, being phonologically dependent upon a neighbouring word (its host) in a construction” (p. 80). In Tagalog, for example, different kinds of words behave as clitics such as personal pronouns and adverbials (Schachter & Otnes, 1972). In this paper, we suppose that the cross-referent pronoun (CRP) is a species of clitic doubling (CD) observed in other languages of the world. Poletto and Pinzin (2024) define CD as “the double occurrence of the same constituent twice inside a single clausal unit; one of the two is represented by a clitic while the other has the properties of a whole phrase” (p. 1). Meanwhile, it is defined by Fischer et al. (2019) as “the doubling of a verbal argument by means of a clitic pronoun inside the same propositional structure” (p. 52). Both definitions indicate structural proximity between the clitic doublee and the element being doubled (henceforth, the *associate*), with the latter restricting CD to verbal constructions. The associate can either be a (a) full pronoun, (b) DP, or (c) CP (Živojinović, 2018).

CD is best understood from studies on Romance and Balkan languages, in which it is known to typically target particular grammatical functions, such as direct objects in Rioplatense Spanish

[*spa*](Jaeggli, 1982), accusatives in Torlak [*srp*] (Živojinović, 2018), and postnominal possessive strong pronouns in French [*fra*] and Greek [*ell*] (Alexiadou, 2016). However, drawing broad generalizations from these systems is not always straightforward: Spanish doubling clitics in particular resist a uniform analysis, sometimes resembling inflectional morphology or independent pronominal elements depending on the dialect and construction type (Belloro, 2015). Of the common properties exhibited by CD among the Romance languages as noted by Poletto and Pinzin (2024, p. 2), e.g., “[w]henver an XP displays CD inside the clausal domain, it also does with left and right dislocation,” none seem to apply to CD as observed in the Philippine languages. This then necessitates comparing the instances of CD within a smaller group of languages, which in this paper shall be the Sambalic subgroup.

For the Sambalic languages, we expect the clitic doublee to come in the form of a THIRD PERSON PRONOUN appearing alongside a co-referent DP within the same clause.⁶ In various literature on the Philippine languages, this clitic doublee is usually referred to as the *cross-referent pronoun*, *cross-reference pronoun*, *cross-referential pronoun* (CRP) and other variants (see Antworth, 1979; Del Corro, 2019; Fordan, 2025b; Villareal, 2020), or less commonly as the *copy pronoun* (Rowell, 1986). Meanwhile in Reid and Liao (2004, p. 11), this has been described as the *pronominal agreement marking* of GENITIVE and NOMINATIVE arguments.

Given that CD is realized differently across languages, we also assume that the licensing conditions for the CRP will vary even within a small language group.

⁶In Fordan (2026), I posit that the CRP is the head of its own functional projection, the clitic phrase (CIP) which is base-generated in the extended verbal spine and c-commanded by its associate at LF.

4 The Cross-Referent Pronoun

In this paper, the cross-referent pronoun (CRP) is defined as an anaphoric element in the form of a THIRD PERSON pronoun that locally co-refers with a DP; that is, instead of replacing its antecedent, it is found in the same clause along with it (Fordan, 2026). In effect, the CRP serves as an “echo” to an argument (Del Corro, 2019). This phenomenon cannot be subsumed under topicalization or dislocation because no prosodic break can be heard, as in Tagalog topicalization (see Ceña, 2012) or Spanish dislocation (see Jaeggli, 1982).

Furthermore, the CRP in this group is typologically significant as it represents an instance of clitic doubling, reminiscent of the Romance languages, which, Kaufman (2024) notes, is observed in the northern Philippines but is rare in the central-southern zones. Interestingly, the CRP is more often observed in languages found in the northern parts of the Philippines, such as those in the Central Luzon group which includes the Sambalic subgroup and Kapampangan (Del Corro, 2019; Rowsell, 1986), some Northern Luzon languages like Ilokano [*ilo*] (Constantino et al., 1967) and Itneg Inlaud [*iti*] (Gonzales, 2025), and also Ivatan [*ivv*] (Reid & Liao, 2004).

Antworth (1979) takes note of the CRP as appearing optionally in inverted sentences where the fronted *chômeur* can license a CRP in the GENITIVE. In my undergraduate thesis (Fordan, 2023), I only briefly mention the existence of the CRP in Sambal, noting its optionality and that its morphological case marking usually reflects that of its associate. Meanwhile, Villareal (2020) focuses on the THIRD PERSON SINGULAR pronouns *ya* and *na* in his description of the CRP in Bolinao, where he notes their distribution in sentences with negated nominals

and POSSESSIVE markers, while also positing that the NOMINATIVE case marker *yay* is historically related to the CRP *ya*. However, a more detailed analysis of this phenomenon is found only in the works on Kapampangan by Del Corro (2019) and Rowsell (1986). In Kapampangan, CRP usage is obligatory in certain constructions and is used for marking grammatical relations (i.e., PASSIVE ACTOR or SUBJECT in a verbal sentence) in a clause while also following semantic constraints on ANIMACY and NUMBER (Del Corro, 2019; Rowsell, 1986).

Taken together, these accounts point to a gap in our understanding of CRPs in the Sambalic languages, particularly its cross-linguistic obligatoriness (i.e., when it is required), morphological case behavior (i.e., what case maps onto another), and semantic sensitivity (i.e., whether or not ANIMACY and NUMBER play a role). From the perspective of UG, closely related languages can be particularly useful specimens for understanding how a shared underlying grammatical architecture can allow for variation in the realization of certain features, thereby informing us what is (micro-)parametrically variable. Previous work on clitic doubling in non-Philippine languages has shown that superficially similar constructions can vary substantially despite the genetic relatedness of the languages (see Fischer et al., 2019; Tsakali & Anagnostopoulou, 2008).

A comparative study is therefore warranted to determine the specific licensing conditions for CRPs in Sambal, Botolan, and Bolinao which shall allow us to move beyond the properties of individual languages and eventually identify which aspects of CRP distribution are stable across the subgroup and which reflect points of variation. By establishing what is shared and what is divergent across these three Sambalic languages,

this study will then deepen our understanding of this underdescribed phenomenon in the Philippine languages and situate these findings within UG.

Take note, however, that while this study is informed by insights from a Chomskyan view of language, I do not assume that all aspects of the Sambalic CRP are reducible to narrow syntax alone. Instead, different components of the phenomenon appear to be best explained at different levels of grammatical organization. Structural constraints on the distribution of the CRP (such as locality restrictions, doubling of particular semantic roles, and preference for verbal constructions) invite a generative treatment, as these properties are mostly regular and could reflect underlying grammatical architecture. At the same time, the optionality of the CRP and its conditioning by referential clarity point to an interaction with pragmatics. Moreover, phonological factors may also play a role in licensing the CRP (as we shall see later), insofar as its realization could be sensitive to euphony considerations at the PF interface. Adopting an eclectic approach, therefore, allows the analysis to remain explanatory without forcing the CRP into a single module of grammar.

Before delving into the CRP behavior within each language, Table 1, containing the THIRD PERSON PRONOUNS in the three languages, is provided here.

Table 1. Third Person Pronouns in Sambal, Botolan, and Bolinao

			Sambal	Botolan	Bolinao
NOM	SG	Full	<i>hiya</i>	<i>hiya</i>	<i>siya</i>
		Minimal	<i>ya</i>	<i>ya</i>	<i>ya</i>
	PL	Full	<i>hila</i>	<i>hila</i>	<i>sara</i>
		Minimal	<i>la</i>	<i>la</i>	<i>ra</i>
GEN	SG		<i>na</i>	<i>na</i>	<i>na</i>
	PL		<i>la</i>	<i>la</i>	<i>ra</i>
OBL	SG		<i>kuna</i>	<i>kuna</i>	<i>kuna</i>
	PL		<i>kunla</i>	<i>kunla</i>	<i>kunra</i>

For the Sambalic languages, as in Kapampangan (see Del Corro, 2019), the CRP is restricted to THIRD PERSON arguments. While the literature does not establish a universal diachronic pathway whereby clitic doubling emerges first in the THIRD PERSON, some studies show that clitic doubling systems are sensitive to the distribution of PERSON features (e.g., Saab, 2024). This could be due to the tendency for THIRD PERSON reference to be more prone to ambiguity and therefore get reinforced by the grammar.

4.1 Sambal

In Sambal, the CRP is best described as AGENT-oriented. It surfaces predominantly in TRANSITIVE⁷ clauses and consistently targets the AGENT, regardless of whether it is the PIVOT (i.e., it is NOMINATIVE-marked) or a non-PIVOT argument as in any voice

⁷An INTRANSITIVE in this paper would refer to syntactic intransitivity as in Ceña and Nolasco (2011) which supposes that there can be another argument apart from the core argument.

configuration. This distribution indicates that the CRP in Sambal is sensitive to semantic role (agentivity) rather than to grammatical function per se, aligning it with the event structure (who did what to whom) rather than to the voice system. The following sentences in the PATIENT voice show that the non-PIVOT, agentive DP *nin anak* is licensed to have a CRP. The examples throughout the paper shall be ordered as follows: (a) cliticized case marker + CRP (most natural) → (b) full case marker + CRP → (c) absence of CRP (shows optionality).

- (1) a. *Kinan nan anak yay mangga.*
 k<in>an na_i=n anak_i yay mangga_j
 <PFV.PAT>eat CRP=GEN child NOM mango
- b. *Kinan na nin anak yay mangga.*
 Kinan na_i nin anak_i yay mangga_j.
 <PFV.PAT>eat CRP GEN child NOM mango
- c. *Kinan nin anak yay mangga.*
 k<in>an nin anak yay mangga
 <PFV.PAT>eat GEN child NOM mango
- d. **Kinan nin anak ya yay mangga.*
 k<in>an nin anak ya_i yay mangga_i
 <PFV.PAT>eat GEN child CRP NOM mango
 ‘The child ate the mango.’
- (2) a. *Kinan lan aw-anak yay mawmangga.*
 k<in>an la_i=n ?<aw>~?anak_i yay
 <PFV.PAT>eat CRP=GEN PL~child NOM
 m<aw>~mangga_j
 PL~mango

b. *Kinan la nin aw-anak yay mawmangga.*

k<in>an la_i nin ?<aw>~?anak_i yay
 <PFV.PAT>eat CRP GEN PL~child NOM
 m<aw>~mangga;
 PL~mango

c. *Kinan nin aw-anak yay mawmangga.*

k<in>an nin ?<aw>~?anak yay m<aw>~mangga
 <PFV.PAT>eat GEN PL~child NOM PL~mango
 ‘The children ate the mangoes.’

As expected, the CRP can also be licensed in the ACTOR voice where the PIVOT is also the AGENT.

(3) a. *Nangan yan mangga yay anak.*

nang-kan ya_i=n mangga_j yay anak_i
 ACR-eat CRP=GEN mango NOM child

b. *Nangan ya nin mangga yay anak.*

nang-kan ya_i nin mangga_j yay anak_i
 ACR-eat CRP GEN mango NOM child

c. *Nangan nin mangga yay anak.*

nang-kan nin mangga yay anak
 ACR-eat GEN mango NOM child
 ‘The child ate a mango.’

However, we do not expect the OBLIQUE-marked argument to be doubled because there is no way for it to be the AGENT in this position.

(4) a. *Dinumani ya yay anak ha dalaga.*

d<in><um>ani ya_i yay anak_i ha dalaga_j
 <PFV><ACR>approach CRP NOM child OBL young lady

b. *Dinumani yay anak ha dalaga.*

d<in><um>ani yay anak ha dalaga
 <PFV><ACR>approach NOM child OBL young lady
 ‘The child approached the young lady.’

Unlike Kapampangan (see Del Corro, 2019), Sambal does not seem to have a strict ANIMACY constraint for the CRP. And so, even non-ANIMATE NPs are able to have a CRP as long as they are agentive in the event structure.

(5) *Napati ya nan bagyo.*

na-pati ya na_i=n bagyo_i
 PFV-kill 3SG.NOM CRP=GEN storm
 ‘He was killed by the storm.’

As expected, NPs which are not agentive in any way are unable to accept a CRP.

(6) **Nakâkit ya nan bag.*

na-kâ-kit ya na_i=n bag_i
 PFV-MOD-see 3SG.NOM CRP=GEN bag
 ‘He saw a bag.’

On the other hand, the CRP is barred in INTRANSITIVE sentences despite the sole argument being agentive. This seems to suggest at this point that the CRP is only licensed for agentive DPs with competing arguments as it does not have so much utility for bare intransitives.

- (7) a. **Nirong ya yay anak.*
n-irong ya_i yay anak_i
PFV-stand CRP NOM child
'The child stood up.'
- b. *Nirong yay anak.*
n-irong yay anak
PFV-stand NOM child

Meanwhile, although we might expect CD to only occur in verbal sentences as is typical of Romance languages, as we have seen earlier in the definition by Fischer et al. (2019), it also appears in EXISTENTIAL sentences and adjectival predicates for Sambal.

- (8) a. *Main yan bulaklak a|yay⁸ dalaga.*
main ya_i=n bulaklak a|yay dalaga_i
EXIST CRP=GEN flower NOM young.lady
- b. *Main bulaklak a|yay dalaga.*
main Ø bulaklak a|yay dalaga
EXIST GEN flower NOM young.lady
'The young lady has a flower.'
- (9) a. *Humin yan bulaklak a|yay dalaga.*
humin ya_i=n bulaklak a|yay dalaga_i
NEG.EXIST CRP=GEN flower NOM young.lady

⁸The NOM marker *yay* can be optionally reduced to the form *a* when found lower in the sentence (see Fordan, 2025a)

- b. *Humin bulaklak a|yay dalaga.*
 humin Ø bulaklak a|yay dalaga
 NEG.EXIST GEN flower NOM young.lady
 ‘The young lady does not have a flower.’

EXISTENTIAL words in the Philippine languages have been analyzed variously and into different categories. For instance, Ceña (2012) considers Tagalog *may*⁹ a determiner and does not distinguish it from the longer form *mayroon*. In Sambal, the EXISTENTIAL *main* appears to be the base form of an aspectless verb which becomes *nagkamain* in the PERFECTIVE and *magkamain* in the PROSPECTIVE. Given this, I suppose that *main* in Sambal is a pseudo-verb, comparable to Tagalog verbs such as *gusto*, *puwede*, and *ayaw* which behave similarly. Under this analysis, the earlier assumption that the CRP is licensed only in verbal constructions poses no theoretical difficulty to us. Moreover, even in the absence of the usual AGENT-PATIENT relations, we still see an “active” role undertaken here by the PIVOT, as in ‘the young lady possessing a flower.’

We only run into problems with the fact that the CRP is also licensed in Sambal with adjectival predicates. In the following examples, *maganda* ‘good’ is clearly an adjective and is negated by *ambo*. In Sambal, *kai/kay* is used when negating verbs (Fordan, 2023).

Here, we see that the Sambal CRP is not only used to mark the semantic AGENT but also to foreground the TOPIC (which I assume to be different from the PIVOT). Although *mako* ‘to go’ is morphologically a

⁹Since *may* cannot take a DP complement (as expected for true verbs), as in **May ng tao*, this analysis seems to be correct for Tagalog. In Sambal, *Main nin (GEN) tawo* is actually acceptable and is natural.

verb, it is essentially a gerund in this position and therefore, deverbalized. This allows it to be an argument to the adjective which then licenses the CRP due to a competing argument. As with the EXISTENTIAL constructions, there is no clear agentive DP here, although there is some “active” quality here contained in the act of ‘going.’

- (10) a. *Ambo yan¹⁰ maganda nin mako ha kabaret.*
ambo ya_i=n ma-ganda nin mako_i ha kabaret
NEG CRP=LNK ADJ-good COMP INF.go OBL cabaret
- b. *Ambo maganda nin mako ha kabaret.*
ambo ma-ganda nin mako ha kabaret
NEG ADJ-good COMP INF.go OBL cabaret
- c. *Ambo yan maganda yay mako ha kabaret.*¹¹
ambo ya_i=n ma-ganda yay mako_i ha kabaret
NEG CRP=LNK ADJ-good NOM INF.go OBL cabaret
‘It is not nice to go to a cabaret.’

¹⁰ This seems anomalous because the associate is not a DP on the surface but simply a deverbalized predicate that occupies an argument position (complement to the adjective). It could be that the complementizer (in Sambal, *nin* is polysemous as both GEN and COMP) acts as the anchor, allowing it to license a CRP and be treated as a nominal argument. Thus, that the CRP in (10a) has NOMINATIVE morphology need not be a problem and instead, it supports the analysis that the construction is structurally nominalized.

¹¹ Examples (10a) and (10c) have a slight difference in meaning due to their differing forms. In (10a), where *nin* is analyzed as COMP, the clause is interpreted descriptively, i.e., it denotes the general act of going to a cabaret and thus, the judgment ‘not nice’ applies to the activity in a generic sense. In contrast, (10c), which involves NOMINATIVE marking, results in a more referential interpretation where it picks out a specific instance of going to a cabaret from among other possible activities, and that is ‘not nice.’

Furthermore, Sambal also consistently licenses only one CRP, even in DITRANSITIVE constructions where we have three arguments. An additional CRP for the INDIRECT OBJECT is barred.

- (II) a. *Impanaliw na nin anakabayo yay dalaga nin kendi.*
 ing-pang-haliw= na_i nin anakabayo $_i$ yay dalaga $_k$
 PAT.PFV-MOD-buy=CRP GEN young.man NOM young.lady
 nin kendi $_j$
 GEN candy
- b. **Impanaliw na ya nin anakabayo yay dalaga nin kendi.*
 ing-pang-haliw= $na_i=ya_k$ nin anakabayo $_i$ yay
 PAT.PFV-MOD-buy=CRP=CRP GEN young.man NOM
 dalaga $_k$ nin kendi $_j$
 young.lady GEN candy
 ‘The young man bought candy for the young lady.’

However, a second CRP may be licensed when an agentive DP’s competing argument is not found locally (i.e., it is in another CP) where we see both arguments having a CRP, including the one in the embedded clause.

- (I2) a. *Pinastang nan dalaga nu nakaali yayna duman yay anakabayo.*
 p<in>astang $na_i=n$ dalaga $_i$ nu na-ka-ali
 <PFV.PAT>question CRP=GEN young.lady C PFV-MOD-leave
 ya $_j$ -yna duman yay anakabayo $_j$
 CRP-already Q NOM young.man

b. *Pinastang nan dalaga nu nakaaliyna duman yay anakabayo.*

p<in>astang na_i=n dalaga_i nu
<PFV.PAT>question CRP=GEN young.lady C
na-ka-ali-yna duman yay nakabayo_j
PFV-MOD-leave-already Q NOM young.man

‘The young lady asked if the young man has already left..’

This example supports the view that CRP usage is also driven by pragmatic considerations, despite the licensing conditions being based on structure. In the host clause, the DP *nin dalaga* ‘young lady’ has no other CRP-eligible argument within its own clause that would normally motivate CRP realization, and thus, under a strictly structural account, the CRP here would be unexpected. Nevertheless, the CRP is still licensed, because the embedded clause introduces another salient argument, the DP *yay anakabayo* ‘young man,’ creating an utterance wherein marking the AGENT or TOPIC becomes relevant.

Up to this point, we have only seen that the CRP is licensed when there are competing arguments to the agentive DP. However, the CRP is also unexpectedly licensed for intransitives in the RECENT PERFECTIVE, where there is no other argument, while still agreeing in CASE and NUMBER features with its associate. Since this construction is PIVOT-less (because there is no VOICE), we must assume here that the CRP is instead used to mark either the agency of its non-PIVOT argument or the non-pivoothood of its agentive argument. As we have seen earlier that bare intransitives cannot license a CRP despite the sole argument being agentive, there must be another factor at play.

- (13) a. *Kaiirong nan anak.*
 ka-i~irong na_i=n anak_i
 REC~stand CRP=GEN child
- b. *Kaiirong na nin anak.*
 ka-i~irong na_i nin anak_i
 REC~stand CRP GEN child
- c. *Kaiirong nin anak.*
 ka-i~irong nin anak
 REC~stand GEN child
 ‘The child just stood up.’

4.2 Botolan

Meanwhile in Botolan, the use of a CRP-like element is relatively more restricted in that it is only licensed to appear in inverted sentences where it is also still optional. According to the current data, Botolan is also the only Sambalic language of the three that allows focusing (the PIVOT), i.e., using the INVERSION marker *ay*, which is also found in some Philippine languages, such as Tagalog (Schachter & Otanes, 1972). In Antworth (1979), it is claimed that the CRP is licensed as long as it is an *ay*-inversion, regardless of voice. However, my data reveal that a CRP-like element is only licensed in Botolan when the sentence is in the ACTOR voice. Take note, however, that some of the judgments, as reported by my consultant in elicitation, are only marginally acceptable. At this point, as we are still unsure whether Botolan has a genuine CRP, a “?” will be used to gloss the CRP-like element.

- (14) *Hinumakay ha kabayo hi Pedro.*
<in><um>hakay ha kabayo hi Pedro
<PFV><ACR>ride OBL horse NOM Peter
- (15) a. *Hi Pedro ay hinumakay ya ha kabayo.*
hi Pedro ay <in><um>hakay ya ha kabayo
NOM Peter PM <PFV><ACR>ride ? OBL horse
- b. *?Hi Pedro ay hinumakay ha kabayo.*
hi Pedro ay <in><um>hakay ha kabayo
NOM Peter PM <PFV><ACR>ride OBL horse
'Peter rode a horse.'

Contrary to previous research, the following sentences (16) and (17a)-(17b) based on Antworth (1979)¹² are no longer acceptable.¹³ This is best explained by the case mismatch that results from the movement because only the PIVOT (*ya kabayo*) is expected to move with the sentences being in the PATIENT voice. In Tagalog, undergoing \bar{A} -extraction for a non-PIVOT argument is only of marginal acceptance.¹⁴

- (16) *Hinakyan ni Pedro ya kabayo.*
h<in>akay-an ni Pedro ya kabayo
<PAT>saddle-LOC GEN Peter NOM horse

¹²The original sentence in Antworth (1979) was *Hiniyawan ni Pedro ya kabayo*, which also means 'Peter saddled the horse' but my consultant was not familiar with the verb.

¹³We can safely attribute this to language change.

¹⁴"Among DPs, only the PIVOT argument of a clause can be targeted for the formation of \bar{A} -dependencies such as relative clauses, focus constructions, and wh-questions." (Hsieh, 2023, p. 520)

- (17) a. ?*Hi Pedro ay hinakyan na ya kabayo.*
 hi Pedro ay h<in>akay-an na ya kabayo
 NOM Peter PM <PAT>saddle-LOC ? NOM horse
- b. **Hi Pedro ay hinakyan ya kabayo.*
 hi Pedro ay h<in>akay-an ya kabayo
 NOM Peter PM <PAT>saddle-LOC NOM horse
 ‘Peter saddled the horse.’

The following are more examples taken directly from Antworth (1979) with the acceptability judgments that I was able to gather. From (19)-(21b), we still see instances of *ay*-inversion focusing a non-PIVOT argument to which Antworth (1979) claims that a “CRP” can be licensed.

- (18) *Nilabi ya nin masyado nin tawtawo.*
 <in>labi ya nin masyado nin tawtawo
 <PAT>love 3SG.NOM LNK very.much GEN people
- (19) a. ?*Hay tawtawo ay nilabi la ya nin masyado.*
 hay tawtawo ay <in>labi la ya nin masyado
 NOM people PM <PAT>love ? 3SG.NOM LNK very.much
- b. **Hay tawtawo ay nilabi ya nin masyado.*
 hay tawtawo ay <in>labi la ya nin masyado
 NOM people PM <PAT>love ? 3SG.NOM LNK very.much
 ‘The people loved him very much.’
- (20) *Inlako ni Juan ya baboy na.*
 <in>lako ni Juan ya baboy=na
 <PAT>sell GEN John NOM pig=3SG.GEN

- (21) a. ?*Hi Juan ay inlako na ya baboy na.*
 hi Juan ay inlako na ya baboy=na
 NOM John PM <PAT>lako ? NOM pig=3SG.GEN
- b. **Hi Juan ay inlako ya baboy na.*
 hi Juan ay <in>lako ya baboy=na
 NOM John PM <PAT>sell NOM pig=3SG.GEN
 ‘John sold his pig.’

Furthermore, unlike Sambal, Botolan does not license the CRP in both TRANSITIVE and INTRANSITIVE sentences, even when the AGENT is non-PIVOT.

- (22) *Nirëng (*ya) ya anak.*
 ni-irëng (*ya) ya anak
 PFV.ACR-stand.up CRP NOM children
 ‘The child stood up.’
- (23) *Nipanpirëng *hila=y|ya aw-anak.*
 ni-pan-pi-irëng *hila=y|ya aw-anak
 PFV.ACR-MOD-MOD-stand.up 3PL.NOM=NOM|NOM children
 ‘The children stood up.’
- (24) *Kapipi-irëng nin|*na=n anak.*
 kapipi-irëng nin|*na=n anak
 REC.PFV-stand.up GEN|CRP=GEN child
 ‘The child just stood up.’
- (25) *Kinan nin|*na=n anak ya mangga.*
 k<in>an nin|na=n anak ya mangga
 <PAT>eat GEN|CRP=GEN child NOM mango
 ‘The child ate the mango.’

Due to this significant restriction, we have good reason to suggest that, although Antworth (1979, p. 53) calls this element a “cross-referent pronoun,” it is better described as a kind of resumptive pronoun¹⁵ like *her* in *Mary, I like her* (Crystal, 2008, p. 415). Focusing and topicalization both render the sentence discontinuous, i.e., they create a greater structural distance between the supposed doublee and its associate, where there is an INVERSION marker or a prosodic break, respectively. Moreover, since the supposed associate has already moved to an \bar{A} -position, it is no longer found in the same verbal domain as its doublee. As described earlier, we assume in this paper that locality is an important feature of the CRP. Meanwhile, the element at hand serves to maintain dependency across \bar{A} -movement and so, it does not reflect the typical redundancy of the CRP. This is analogous to how Fischer et al. (2019) differentiate CD from clitic left and right dislocation: “We assume that [CD] is different from clitic left and right dislocation... [which] necessarily involves prosodic cues ... whereas this need not be the case in [CD] structures” (p. 1).

4.3 Bolinao

Villareal’s (2020) work on Bolinao only gives examples in which the CRP is allowed, noting also that it has undergone grammaticalization into a DEFINITE article, but it does not describe its licensing conditions particularly. The distribution of the CRP in Bolinao is similar to that of Sambal in that it does not require movement in order to be licensed.

Firstly, like Sambal, verbs in the ACTOR (as in example (26) and PATIENT voices (as in examples (27) will normally have a CRP where

¹⁵It is coreferential but not due to clitic doubling and therefore, not a true CRP.

the associate is also the semantic AGENT. Take note that the absence of a CRP here evokes only marginal acceptance, conveying a feeling that something is missing.

- (26) a. *Nangan yan mangga a anak.*
 nang-kan ya_i=n mangga a anak_i
 ACR-eat CRP=GEN mango NOM child
- b. *Nangan ya nin mangga a anak.*
 nang-kan ya_i nin mangga a anak_i
 ACR-eat CRP GEN mango NOM child
- c. *?Nangan nin mangga a anak.*
 nang-kan nin mangga a anak
 ACR-eat GEN mango NOM child
 ‘The child ate a mango.’
- (27) a. *Kinan nan anak a mangga.*
 k<in>an na_i=n anak_i a mangga
 <PAT>eat CRP=GEN child NOM mango
- b. *Kinan na nin anak a mangga.*
 k<in>an na_i nin anak_i a mangga
 <PAT>eat CRP GEN child NOM mango
- c. *?Kinan nin anak a mangga.*
 Kinan nin anak a mangga
 <PAT>eat GEN child NOM mango
 ‘The child ate the mango.’

EXISTENTIAL constructions such as example (28a) show a parallel with Sambal. In Sambal, the EXISTENTIAL predicates in (8)–(9) license the CRP despite lacking an obvious TRANSITIVE structure which, under

my analysis, is possible because the EXISTENTIAL word in Sambal seems to behave like a pseudo-verb that still introduces relations between arguments. A similar pattern is observed in Bolinao, however, contrary to what is expected, the CRP here is entirely optional.

- (28) a. *Kasa ya sa bali a balasang.*
 kasa ya_i sa bali a balasang_i
 NEG.EXIST CRP OBL house NOM young.lady
- b. *Kasa sa bali a balasang.*
 kasa sa bali a balasang
 NEG.EXIST OBL house NOM young.lady
 ‘The young lady is not at home.’

Moreover, the CRP is also licensed for the DP complement to a noun predicated by an adjective and is thus a non-verbal sentence. Nonetheless, the DP *nin anak* still has some form of agency despite not being a verbal argument such that it is described as performing an action. This reveals that, at least for the CRP in both Sambal and Bolinao (and perhaps for the Sambalic group), predication type may not be relevant to CRP licensing. In light of this, we must subscribe to Poletto and Pinzin’s (2024, p. 1) definition of clitic doubling, where it is simply “the double occurrence of the same constituent twice inside a single clausal unit,” rather than to a definition which requires verbal status, as in Fischer et al. (2019, p. 52), in order to better reflect the cross-linguistic data.

- (29) a. *Paspas yay palayo nan anak.*
 paspas yay palayo na_i=n anak_i
 fast NOM run CRP=GEN child

- b. *Paspas yay palayo na nin anak.*
paspas yay palayo na_i nin anak_i
fast NOM run CRP GEN child
'The child runs fast.'

Meanwhile, unlike Sambal, DITRANSITIVE constructions in Bolinao are able to license two CRPs. As expected, CRP use is also most preferred.

- (30) a. *Ipinanaliw na yan tubabayo a balasang nin kendi.*
i-p<in>ang-haliw=na ya_i=n tubabayo_k a
BEN-<PFV>MOD-buy=CRP CRP=GEN young.man NOM
balasang_i nin kendi_j
young.lady GEN candy
- b. *?Ipinanaliw nin tubabayo a balasang nin kendi.*
i-p<in>ang-haliw nin tubabayo a balasang nin
BEN-<PFV>MOD-buy GEN young.man NOM young.lady GEN
kendi
candy
'The young man bought candy for the young lady.'

However, like Sambal, a second CRP is also licensed when a competing argument to the agentive DP is found in an embedded clause where both arguments have a CRP.

- (31) a. *Nibarità nan anak nin nagawid yaynay tubobâyo.*
 ni-barità na_i=n anak_i nin nag-awid
 PFV-tell CRP=GEN child C PFV-go.home
 ya_j-yna-y tubobâyo_j
 CRP=already=NOM young.man
 ‘The child said that the young man had gone home.’
- b. ?*Nibarità nan anak nin nagawid anay tubobâyo.*
 ni-barità na_i=n anak_i nin nag-awid ana=y
 PFV-tell CRP=GEN child C PFV-go.home already=NOM
 tubobâyo_j
 young.man
 ‘The child said that the young man had gone home.’

Like Sambal, the CRP is also licensed in the RECENT PERFECTIVE as it is also an agentive non-PIVOT.

- (32) a. *Kâdëng nan anak.*
 kâ-dëng na_i=n anak_i
 REC-stand CRP=GEN child
- b. *Kâdëng na nin anak.*
 kâ-dëng na_i nin anak_i
 REC-stand CRP GEN child
- c. **Kâdëng nin anak.*
 kâ-dëng nin anak
 REC-stand GEN child
 ‘The child just stood up.’

With this, we see that the Bolinao CRP is almost obligatory for all its structurally licensed positions across construction types. This

could have a morphological explanation. Because both the GENITIVE CM and the LINKER (for modification) in Bolinao are *nin*, the CRP provides a distinct marker separating argumenthood from modification, preventing any potential ambiguity. Unlike Sambal and Botolan whose GENITIVE CM and LINKER are morphologically distinct, Bolinao gives near-obligatory status to the CRP as a kind of differential argument marking.

Apart from depending on valency and theta roles, the Bolinao CRP is also obligatory based on NUMBER: any DP argument with a plural noun requires the CRP *sara*.

- (33) a. *Nagsiudëngan saray a-anak.*
 nag-si-udëng-an sara_i-y a~anak_i
 PFV-MOD-stand-MOD CRP=NOM PL~child
- b. **Nagsiudëngan yay a-anak.*
 nag-si-udëng-an yay a~anak
 PFV-MOD-stand-MOD NOM PL~child
 ‘The children stood up together.’
- (34) a. *Niaway sara sina Pedro tan Juan.*
 ni-away sara_i sina [Pedro tan Juan]_i
 PFV-fight CRP NOM Peter and John
- b. **Niaway sina Pedro tan Juan.*
 ni-away sina Pedro tan Juan
 PFV-fight NOM Peter and John
 ‘Peter and John fought each other.’

Interestingly, the singular counterpart of this construction does not allow a CRP. In these examples, it is apparent that the Bolinao CRP is

also connected to plurality, besides agentivity. In this way, the CRP also acts as an agreement marker for plural DPs in Bolinao.

- (35) a. *Inumdëng yay anak.*
in-m-udëng yay anak
PFV-ACR-stand NOM child
- b. **Inumdëng ya yay anak.*
in-m-udëng ya_i yay anak_i
PFV-ACR-stand CRP NOM child
‘The child stood up.’

Similarly, the CRP is also not licensed for INTRANSITIVE sentences in the ACTOR voice with an OBLIQUE argument.

- (36) a. *Kimadani yay anak sa balasang.*
ki-ma-dani yay anak sa balasang
PFV-ACR-go NOM child OBL young.lady
- b. **Kimadani ya yay anak sa balasang.*
ki-ma-dani ya_i yay anak_i sa balasang
PFV-ACR-go CRP NOM child OBL young.lady
‘The child approached the young lady.’

The ungrammatical examples can be explained by phonological reasons, instead of purely through syntax. Using the CRP in those sentences would require the CRP *ya* to be immediately followed by the NOMINATIVE marker *yay*, resulting in a repetitive *ya-ya* sequence. In *Morpolohya ng Filipino*, Ceña (2021) proposes the *Principle of Minimal Distinction* for explaining why the PERFECTIVE PATIENT form of *tawag* is *tinawag* instead of **tinawagin*. He goes on to speculate that perhaps the phenomenon goes beyond affixation, as in the case of the LOCATIVE

verb *iwanan*, which becomes *iwanan*. I assume that the same forces are at work here due to the similar sounding *ya* and *yay* despite their having different functions. In the plural version, the NP *a-anak*, although still marked by *yay*, receives a different CRP, *sara*, with no repetitive segment. Lastly, like Botolan, topicalized¹⁶ INTRANSITIVE sentences in Bolinao also obligatorily include a CRP-like element referring back to the fronted noun. This is also analyzed here as a resumptive pronoun.

- (37) *Si anak, inumdëng ya.*
si anak Ø <in>umdëng ya
NOM child PM <PFV>stand RSM
'The child stood up.'

5 Summary and Conclusion

This paper has shown that the CRP is attested in two of the three selected Sambalic languages, while Botolan does not fully realize a typical CRP, but rather a CRP-like element that only appears in inverted sentences. The relevant differences of their licensing conditions are summarized in Table 2.

¹⁶Sambal also allows for topicalization, but I was not able to elicit a CRP-like element:

- (i) *Yay anak, nîrong.*
yay anak Ø <ni>irong
NOM child PM <PFV>stand
'The child stood up.'

Microvariation in the Clitic Doubling Parameter

Table 2. Comparison of the CRP in Sambal, Botolan, and Bolinao

Property	Sambal	Botolan	Bolinao
Optionality	Optional	N/A	Almost obligatory
Role of the Associate	AGENT	N/A	AGENT
Structure conditions	(a) Licensed in transitives, intransitives with an OBLIQUE argument, any RECENT PERFECTIVE verb. (b) Barred for bare intransitives.	N/A	(a) Licensed in transitives, intransitives with an OBLIQUE argument, any RECENT PERFECTIVE verb. (b) Barred for bare intransitives.
ANIMACY sensitivity	No strict ANIMACY constraint	N/A	No strict ANIMACY constraint
Number of licensed CRPs for ditransitives	One	N/A	Two
Function(s)	(a) Disambiguates who is acting when PIVOT ≠ AGENT, reinforcing reference	N/A	(a) Disambiguates who is acting when PIVOT ≠ AGENT, reinforcing reference (b) Agreement marker for plurality

Property	Sambal	Botolan	Bolinao
Others	N/A	Not a CRP, but a resumptive pronoun	Has phonological constraints for <i>ya yay</i> sequence

Given the above discussion, this paper affirms that the CRP is best analyzed as a doubled clitic, in line with its analogous phenomena in other languages. Although a CRP-like element is also used in Botolan, it clearly serves a repair function for extraction gaps in inverted sentences, instead of showing local co-reference, and therefore I assume this to be a resumptive pronoun. Thus, this paper shows how genuine CRPs, as in my original definition, are anaphoric elements in the form of a THIRD PERSON pronoun that locally co-refers with an active DP, regardless of predicate type or verbal status.

6 Sambalic CRP as Parametric Variation

With all this, one might ask: *Why does the CRP exist at all in the Sambalic languages?* Some questions about why languages are the way they are can be answered by pointing to the structure or a feature of the language itself. For instance, *Why is Tagalog word order freer than that of English?* It is due to Tagalog having a voice system which marks the role of arguments in a sentence, so movement is allowed because the voice-marking affixes maintain clarity. However, the CRP question is not quite like that, in that it does not entail a straightforward grammar-internal explanation. In other words, no such rule seems to exist that forces a language to have an element that doubles a core argument. Therefore, the CRP question

is almost like asking why a particular language has grammatical gender or prepositions instead of postpositions. Because we can deduce that, fundamentally, it is a question of parametric choice, not necessity; i.e., the existence of a CRP is just one of the many but limited ways that our minds can configure language. Once that option became available, however, we can explain why it persisted and became useful. From a broader perspective, this also supports the idea that languages from unrelated groups, such as the Philippine and Romance languages, can have similar characteristics.

In the same way, with regard to the variation in CRP licensing conditions, no extra-grammatical explanation seems to be necessary. The variation is not attributable to sociocultural factors, but to how each language has grammaticized a clitic doublee. We have no good reason, for instance, to believe that the speakers of Bolinao (where the CRP is almost obligatory) have certain values or attitudes that Sambal and Botolan speakers do not. The same holds for the fact that there are CD languages and non-CD languages.

Thus, ultimately, the case of the Sambalic CRP presents empirical support for the structural variation among languages being mostly arbitrary and being best explained by parameter setting, rather than by culture. More specifically, the existence of the CRP could be tied to a microparameter with the values of allowing the doubling of either a full DP or other possible associates. Taking one step back, this possible configuration of a language is activated by the macroparameter for clitic doubling (see Baker, 2008 for *microparameter* vs. *macroparameter*).¹⁷

¹⁷Baker (2008): “This [microparametric] approach uses the same reasoning as classical parameter theory, but applies it at a higher level of magnification. It tends to

Nonetheless, the CRP may, at the same time, play a pragmatic role. By echoing the agentive argument, the CRP helps the listener to keep track of event structure, especially given the characteristics of the Philippine-type voice system in which agentivity and pivothood are not so precisely linked morphologically. In Philippine languages, the PIVOT is not always the semantic AGENT, and THIRD PERSON arguments in particular are prone to ambiguity. Moreover, some sentences can even be PIVOT-less like the RECENT PERFECTIVE. The Sambalic CRP alleviates this by providing an index for a particular role or grammatical function.

All in all, the present study has attempted to show that a small feature like the Sambalic CRP could be a genuinely multi-component phenomenon whose properties emerge from the interaction of syntactic licensing, interpretive requirements, and constraints at PF. Moreover, it exemplifies how closely related languages can diverge in syntactic details, while at the same time proving how human languages are alike.

Finally, I recommend future work on the CRP in Sambalic languages to situate it within Baker's (2001) parameter hierarchy: researchers could ask which higher-level parameter allows for or blocks CRP usage, which would further clarify the differences between CD and non-CD languages. Also, expanding the coverage to other Sambalic languages (particularly the Ayta languages) would help deepen our understanding of the microparametric variation in CRP licensing across the subgroup.

look for very localized differences with small but nontrivial effects in the grammars of very closely related languages (alias dialects)" (p. 1).

7 References

- Alexiadou, Artemis. (2016). DP internal (clitic) doubling. In Susann Fischer & Mario Navarro (Eds.), *Proceedings of the VII Nereus International Workshop: Clitic doubling and other issues of the syntax/semantic interface in Romance DPs* (Arbeitspapier No. 128). Universität Konstanz.
- Antworth, Evan L. (1979). *A grammatical sketch of Botolan Sambal (LSP Special Monograph Issue)* (8th ed.). Linguistic Society of the Philippines.
- Baker, Mark C. (2001). *The atoms of language: The mind's hidden rules of grammar*. Basic Books.
- Baker, Mark C. (2008). The macroparameter in a microparametric world. In Theresa Biberauer (Ed.), *The limits of syntactic variation* (pp. 351–373). John Benjamins Publishing Company. <https://doi.org/10.1075/la.132.16bak>
- Belloro, Valeria A. (2015). *To the right of the verb: An investigation of clitic doubling and right dislocation in three Spanish dialects*. Cambridge Scholars Publishing.
- Ceña, Resty M. (2012). *Sintaks ng Filipino*. National Commission on Culture and the Arts.
- Ceña, Resty M. (2021). *Morpholohya ng Filipino*. Books Bisooogo.
- Ceña, Resty M., & Nolasco, Ricardo M. (2011). *Gramatikang Filipino: Balangkas*. University of the Philippines Press.
- Constantino, Ernesto, Paz, Consuelo J., & Posoncu, Marietta N. (1967). The personal pronouns of Tagalog, Ilukano, Isinai, and Kapampangan. In Mario D. Zamora (Ed.), *Studies in Philippine*

- anthropology (In honor of H. Otley Beyer)* (pp. 567–591). Alemar Phoenix.
- Crystal, David. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Blackwell Publishing.
- Del Corro, Anicia. (2019). Tampok na katangian ng gramatika ng wikang Kapampangan. *Pandiwa: Lathalaan para sa Wika at Kultura*, 56–72.
- Eberhard, David M., Simons, Gary F., & Robinson, Alison J. (Eds.). (2026). *Ethnologue: Languages of the world* (29th ed.) [Online version]. SIL Global. <https://www.ethnologue.com/>
- Elgincolin, Sotero B., Goschnick, Hella E., & Elgincolin, Priscilla R. (1988). *English-Tina Sambal-Pilipino dictionary*. Summer Institute of Linguistics.
- Fischer, Susann, Navarro, Mario, & Jorge, Vega Vilanova. (2019). The clitic doubling parameter: Development and distribution of a cyclic change. In Miriam Bouzouita, Anne Breitbarth, Lieven Danckaert, & Elisabeth Witzhausen (Eds.), *Cycles in language change* (pp. 52–70). Oxford University Press. <https://doi.org/10.1093/oso/9780198824961.003.0004>
- Fordan, Leo Gregory A. (2023). *Isang gramatikal-iskets ng Sambal Tina* [Unpublished undergraduate thesis]. University of the Philippines Baguio.
- Fordan, Leo Gregory A. (2025a, April). *Linking and embedding functions of case markers in the Sambalic languages* [Paper presentation]. 7th Linguistic Society of the Philippines International Conference (LSPIC), Pangasinan State University, Pangasinan, Philippines.

- Fordan, Leo Gregory A. (2025b, October). *Syntactic properties of the cross-referent pronoun in the Sambalic languages* [Paper presentation]. 16th International Free Linguistics Conference (FLC), University of Santo Tomas, Manila, Philippines.
- Fordan, Leo Gregory A. (2026, April). *Revisiting binding theory: Evidence from clitic doubling in Sambal* [Paper presentation]. 8th Linguistic Society of the Philippines International Conference (LSPIC), West Visayas State University, Iloilo, Philippines.
- Gonzales, Ryn Jean. (2025). *A grammar of Itneg Inlaud* [Unpublished PhD dissertation]. University of the Philippines Diliman.
- Himes, Ronald S. (2012). The Central Luzon group of languages. *Oceanic Linguistics*, 51(2), 490–537. <https://www.jstor.org/stable/23321866>
- Hsieh, Henrison. (2023). Locality in exceptional Tagalog \bar{A} -extraction. *Linguistic Inquiry*, 56(3), 519–560. https://doi.org/10.1162/ling_a_00505
- Jaeggli, Osvaldo. (1982). *Topics in Romance syntax (Studies in generative grammar, Vol. 12)*. Foris.
- Kaufman, Daniel. (2024). The languages of the central and southern Philippines. In Alexander Adelaar & Antoinette Schapper (Eds.), *The Oxford guide to the Malayo-Polynesian languages of Southeast Asia* (pp. 347–374). Oxford University Press. <https://doi.org/10.1093/oso/9780198807353.003.0025>
- Komisyon sa Wikang Filipino. (2016). *Atlas ng mga wika ng Filipinas* [Map].
- Poletto, Cecilia, & Pinzin, Francesco. (2024, June 18). Clitic doubling in the Romance languages. In *Oxford research encyclopedia of*

- linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.643>
- Reid, Lawrence A., & Liao, Hsiu-chuan. (2004). A brief syntactic typology of Philippine languages. *Language and Linguistics*, 5(2), 43–90.
- Rowsell, Lorna. (1986). The function of the copy pronouns in Kapampangan. *Calgary Working Papers in Linguistics*, 12(Summer), 12.
- Saab, Andrés. (2024). Neither agreement nor pronouns: Clitic doubling and weak crossover in Rioplatense Spanish. *Isogloss*, 10(5), 1–31. <https://doi.org/10.5565/rev/isogloss.379>
- Schachter, Paul, & Otnes, Fe T. (1972). *A Tagalog reference grammar*. University of California Press.
- Tsakali, Vina, & Anagnostopoulou, Elena. (2008). Rethinking the clitic doubling parameter: The inverse correlation between clitic doubling and participle agreement. In Dalina Kallulli & Liliane Tasmowski (Eds.), *Clitic doubling in the Balkan languages* (pp. 321–360). John Benjamins Publishing Company. <https://doi.org/10.1075/la.130.11tsa>
- Villareal, John Venson. (2020). *A grammar of Bolinao* [Unpublished master's thesis]. University of the Philippines Diliman.
- Viray, Felizberto B., & Schreerer, Otto. (1929). The Sambali dialect of Bolinao. *The Archive: A Collection of Papers Pertaining to Philippine Linguistics (Paper No. 7)*.
- Živojinović, Jelena. (2018). Torlak clitic doubling: A cross-linguistic comparison. In Andreas Blümel, Jovana Gajić, Ljudmila Geist,

Uwe Junghanns, & Hagen Pitsch (Eds.), *Advances in formal Slavic linguistics 2018* (Ch. 16). Language Science Press.

8 List of Abbreviations

~	reduplication	LF	logical form
?	marginal acceptance	LOC	locative voice
3	third person	LNK	linker
ACR	actor voice	MOD	modal affix
ADJ	adjectivalizer	NEG	negative
BEN	benefactive voice	NOM	nominative case marker
C	complementizer	OBL	oblique case marker
CD	clitic doubling	PAT	patient voice
CL	Central Luzon	PF	phonological form
CP	complementizer phrase	PFV	perfective aspect
CM	case marker	PL	plural
COMP	complementizer	PM	predicate marker
CRP	cross-referent pronoun	REC	recent perfective aspect
DP	determiner phrase	RSM	resumptive pronoun
EXIST	existential	SG	singular
GEN	genitive case marker	Q	question particle
INF	infinitive	UG	Universal Grammar
KWF	Komisyon sa Wikang Filipino	XP	general phrase category