

Conjunctions, clisis, and topicalization in Tagalog

Sergei B. Klimenko

Institute for Linguistic Studies, Russian Academy of Sciences
(St. Petersburg, Russia); sergeybklimenko@gmail.com

Abstract. This paper attempts to provide a more accurate description of three understudied properties of Tagalog conjunctions (the ability to undergo *ay*-topicalization, to host enclitic particles, and to host enclitic pronouns) and to overview issues for future research brought up by this description. The study is based primarily on data from the Tagalog corpus at sketchengine.eu and supplemented by Google searches for a convenience subset of 30 Tagalog conjunctions, consisting mostly of causal conjunctions and several others. The conjunctions in study fall into five categories based on value combinations of the examined properties: group 1 has positive values for all three properties; group 2 conjunctions can host both types of enclitics but does not undergo topicalization; group 3 conjunctions can undergo topicalization and host only particles; group 4 conjunctions can only host particles but cannot undergo topicalization; and group 5 has negative values for all three properties. Most conjunctions examined in this paper belong to group 3. Contrary to the observations made in some published descriptions of Tagalog enclitics, the corpus data show that most of the examined conjunctions can host enclitic particles, and at least three can also host pronouns. The attested value combinations show that the ability to be topicalized and the ability to host pronouns both entail the ability to host particles, which stems from the general properties of topicalized and fronted constituents. The variation found in the behavior of the conjunctions raises a number of questions that need to be studied in further research, including the interaction between conjunction topicalization and information structure, the possibility of the varying conjunction properties pointing to different grammaticalization pathways, the ability to co-occur with the conjunction *at* ‘and’ in place of the topic marker *ay*, and, finally, obligatoriness of being enclitic hosts.

Keywords: Tagalog, conjunctions, enclitics, topicalization, causal constructions.

Acknowledgements. This research was funded by the Russian Science Foundation (project no. 18-18-00472, “Causal Constructions in World Languages (Semantics and Typology)”). I am grateful to Ekaterina A. Baklanova for her valuable comments on a draft of this paper.

Союзы, клитики и топикализация в тагалого

С. Б. Клименко

Институт лингвистических исследований РАН (Санкт-Петербург, Россия);
sergeybklimenko@gmail.com

Аннотация. Цель статьи — более точное описание трех малоизученных свойств союзов в тагалого (способность подвергаться *ау*-топикализации, присоединять энклитические частицы и местоимения). Представленное здесь описание противоречит наблюдениям в публикациях о тагалских союзах в контексте клитизации. Обнаруживаемое варьирование делает возможными будущие исследования по ряду вопросов, включая взаимодействие топикализации союзов и информационной структуры, корреляцию между свойствами союзов и траекториями их грамматикализации и сочетаемость с союзом *at* ‘и’ в функции маркера топика.

Ключевые слова: тагалог, союзы, энклитики, топикализация, каузальные конструкции.

Благодарности. Работа выполнена при поддержке гранта РФФ № 18-18-00472 «Причинные конструкции в языках мира: семантика и типология». Благодарю Е. А. Бакланову за ее ценные комментарии к черновику данной статьи.

1. Introduction

This study offers a description of three understudied properties of Tagalog conjunctions¹: the ability to formally undergo topicalization (see *Sections 3.1 and 3.3*), the ability to host enclitic particles, and the ability

¹ Conjunctions are defined here as connectors of two clauses.

to host enclitic pronouns. To my knowledge, there have been no studies on the use of Tagalog conjunctions in the topicalization construction. As for the ability to host enclitics, the existing studies present conflicting reports on whether Tagalog conjunctions can serve as enclitic hosts. The aim of this paper then is to fill this gap in the description of Tagalog using corpus data and to overview issues for future research brought up by this description.

The paper looks into corpus data for a convenience subset of Tagalog conjunctions consisting of causal conjunctions and several others, including the consequence conjunctions *kaya*, *samakatuwid*, and *anupa*=*ʔ*, as well as a number of others: *para* ‘so that’, *nang* ‘when’, *bago* ‘before’, *kung* ‘if’, *nguni*=*ʔ* ‘but’, *pero* ‘but’, and *at* ‘and’. The causal conjunction list includes 20 items,² all of which mean ‘because’, but differ primarily in their level of formality (e.g. *dahil*, *dahil sa*, *dahilan sa*, *gawa ng*, *bilang* are stylistically neutral, while *palibhasa*, *(ya)yaman*=*g*, *(sa)pagka*(=*t*), *mangyari*,³ *sa dahilan*=*g*, *sa kadahilan(an)*=*g*⁴ are formal, and *(=)kasi*, *(kung) dangan*=*kasi*, *dangkasi*, *(pa)paano*=*kasi*, *porke*(=*t*), *tatal*, *komo*, *kesyo*, *manyapa* ʔ are informal. Also, the conjunctions differ in their frequency (e.g., 41,935 occurrences of *sapagkat* vs 17 occurrences

² Although the causal conjunction list here was intended to be comprehensive, it is possible that more causal conjunctions can be found in the provincial Tagalog dialects. The lists of causal conjunctions presented in other descriptions of Tagalog are less complete: e.g. Schachter, Otanes seven causal conjunctions (*dahil*, *dahil sa*, *dahilan sa*, *gawa ng*, *porke*, *(sa)pagka* ʔ, and *(ya)yaman* + *na/-ng*) are listed in [Schachter, Otanes 1972: 477], while [Malicsi 2013: 153, 157, 170] 13 conjunctions (*kesyo*, *mangyari*, *anupa* ʔ, *kasi*, *dahil*, *sa dahilan*, *porke*, *(sa)pagkat*, *palibhasa* (*ay*), *gawa ng*, *komo*, *tatal*, *(ya)yamang*) and two conjunctive adverbs (*pa(pa)ano kasi* and *dahil dito*) are listed in [Malicsi 2013: 153, 157, 170]. Two of these (*anupa* ʔ and *dahil dito*) actually denote consequence, not reason.

³ *Mangyari* also functions as the verb ‘to happen’. There are also the following combinations, in which *mangyari* has a different meaning: *mangyari*=*pa* ‘of course’ and *mangyari*=*l(am)ang* ‘please/if possible’.

⁴ *Sa dahilan*=*g* and *sa kadahilan(an)*=*g* retain their substantive phrase structure, corresponding to *for a/the reason*, and can be modified by a demonstrative, possessor phrase, and other determiners, without introducing a subordinate clause, like

of *manyapa(')t* and 7 occurrences of *dangkasi* in the corpus). Combined with the fact that some conjunctions, like *dangkasi* and *manyapa't*, are not known to some Manila Tagalog speakers, this might indicate that some of the conjunctions come from Tagalog dialects outside Manila.⁵ *Porke(=t)* [Schachter, Otones 1972: 477] and *kesyo* are different from other causal conjunctions in that their use marks the speaker's skepticism towards the validity of the alleged causal relation between the situations in the two clauses,⁶ as shown in (48) for *porke(=t)*, where this semantics is rendered in the translation with 'just because', and in (56) for *kesyo*.⁷ Such semantics is also available to some other causal conjunctions (e.g. *komo*), but it seems to be obligatory only for *porke(=t)* and *kesyo*.

The data in this study come from the Tagalog (Filipino) Web 2019 (tITenTen19) online corpus of over 230 million tokens [Jakubíček et al. 2013] at sketchengine.eu [Kilgarriff et al. 2004]. The corpus was gathered in June–July 2018 and February 2019 by crawling the web with

in (i). However, since these two phrases also function as connectors of two clauses within a complex sentence, they are considered as conjunctions here.

- (i) *Sa kadahilanan=g ito, marami=ng buhay ang na-sira <...>*
 OBL reason=LK PROX many=LK life NOM PFV.MOD-break[UV]
 'For this reason, many lives were ruined <...>' [6223849]

⁵ For instance, a blog post claims that *manyapa't* is part of the dialect spoken in the municipality of Mendez-Nuñez, Cavite, south off Manila (<https://yellowtrap.wordpress.com/2012/01/27/mga-salitang-mindis/>). Also, reportedly this conjunction is used by speakers in the province of Batangas, south off Cavite (Divine Angeli En-driga, p. c.).

⁶ Most uses of *manyapa(')t* found in the corpus and in Google searches also share the same semantics. However, as shown in example (64) below, it can be used without this meaning as well.

⁷ *Kesyo* also functions as a complementizer with the same semantics of skepticism, like in (i):

- (i) *G<um>awa=nga=ako ng palusot kesyo madami=ako=lagi=ng*
 <AV>do[PFV]=REIT=1SG.NOM GEN excuse COMP much=1SG.NOM=always=LK
ga~gaw-in.
 PROSP~do-UV
 'I made up an excuse that I always have a lot of stuff to do.' [11659446]

Spiderling, a web spider designed for linguistic purposes that looks for text rich websites in a large variety of domains, while avoiding nontextual material, and then removes junk data (navigation links, advertisements, headers and footers), duplicates and near-duplicates [Suchomel, Pomikálek 2012]. The Tagalog corpus contains 449,628 documents from a list of websites that is too long to give here in full.

A number of limitations stem from the use of such a corpus. First, although it is claimed that the corpus has been cleaned of machine-translated content, a sample of 100 sentences containing the oblique nonpersonal case marker *sa* — the most frequent word in the corpus — includes seven sentences from webpages with machine-translated Tagalog texts.⁸ Such sentences, easily identified due to the broken Tagalog syntax and unclear semantic content, were excluded from any counts in this study. Second, Tagalog being the most widely spoken Philippine language, a national Philippine language, and a lingua franca for a large number of nonnative speakers, it is highly likely that the corpus contains a certain amount of data produced by nonnative speakers. The sample, thus, is likely to contain constructions that commonly occur in native speech, as well as those that represent marginal aberrations stemming from a nonnative lack of competence, native idiosyncrasy or lapse of judgment. The quality of such data can be fully assessed only through elicitation of acceptability judgments from native speakers, which was not performed for this study. However, higher frequency of a construction type in the sample probably indicates a greater chance that it is commonly used and accepted within the speech community.

The paper is organized as follows. *Section 2* offers an overview of certain relevant aspects of Tagalog grammar, including the case markers, inversion constructions, and clitics. *Section 3* presents the data found in the corpus, which shows that Tagalog conjunctions fall into five types, regarding the three properties under study here. *Section 4* lists the discrepancies between the data found in the corpus and the descriptions in previous

⁸ The sample was formed using the built-in sketchengine function for pseudorandom extraction of tokens from all parts of the corpus. Users extracting samples of the same size from the same corpus get the same result, which ensures replicability of the sample.

studies, as well as discusses a number of related issues that need to be studied in future research: the correlation of conjunction topicalization with information structure of the utterance, possible pathways of conjunction grammaticalization on the basis of similarities between the five conjunction types and other constituents regarding topicalization and clitic placement, and the ability to be used in combination with the conjunction *at* ‘and’, which seems to be able to replace *ay* ‘TOP’.

2. Tagalog background

2.1. Case markers

Arguments in Tagalog can be realized as one of the four following types with either of the two case systems presented below: nonpersonal substantive⁹ phrases, personal nouns, personal pronouns, or demonstratives. Nonpersonal and personal substantive phrases have syntactically free case markers, while all pronominal and demonstrative forms are inherently marked for case. Two systems of case can be distinguished in Tagalog: one for nonpersonal substantives and demonstratives with nominative, genitive, and oblique cases;¹⁰ and that of personal nouns and personal

⁹ The term *substantive* (cf. [Vanoverbergh 1931; Newell 1993]) is preferred here to *nominal phrase* because in Philippine languages the functions of nominal phrases can be performed by a member of any lexical class [Himmelman 2005: 127]. Substantives are potentially referential phrases that consist of a clause with some lexical item (verb, adjective, noun, numeral, prepositional phrase, or existential predicate) as its predicate, which is introduced with substantivizers (case markers, demonstrative pronouns, existential predicates, some numerals, plural markers, some verbs) if it is originally inherently nonreferential. Apart from the presence of a predicate with its arguments in substantivized phrases, clitic placement [Kroeger 1993: 123] and wh-extraction [Richards 2009: 144] also indicate that such constituents are clauses.

¹⁰ Nominative, genitive, and oblique here are used simply as convenient labels for the notions they refer to. The use of these terms bears no implications for the dominant

pronouns with nominative, actor form, and nonactor form. Nonpersonal case markers (*Table 1*) include only three forms (one for each case), while personal noun case markers (*Table 2*) comprise a paradigm of nine markers with two categories: case and number. The plural forms have two variants. The second one is used only colloquially.

Table 1. Tagalog nonpersonal case markers

NOM	GEN	OBL
ang	ng ¹¹	Sa

Table 2. Tagalog personal noun case markers

number	NOM	ACT	NACT
SG	si	ni	kay
PL	sina / sila	nina / nila	kina / kila

Nominative markers are used to introduce: (i) any type of participants that can be the subject of a verb in a corresponding voice; (ii) the subject in equational and non-verbal sentences (examples (32), (43), (44), (51)); (iii) appositive phrases (examples (19), (49)). Nominative personal case markers are also used when the personal noun is the predicate.¹² The nominative nonpersonal case marker introduces definite independent predicates.¹³ The actor forms of personal markers are used only with actors in the undergoer voice (examples (74), (79)), possessors (examples (2),

syntactic alignment type in Tagalog. This paper is agnostic towards the issue of Tagalog being an ergative, accusative or symmetric language.

¹¹ This is an orthographic abbreviation of /nan/.

¹² (i) ⟨...⟩ *si* *Maymay* *ang* <um>awit ⟨...⟩
 PERS.SG.NOM PN NOM <AV>sing[PFV]
 ‘...The one who sang was Maymay...’ [66920]

¹³ (i) *Ang ahas* *ang pinaka-tuso* *sa lahat ng hayop* ⟨...⟩
 NOM snake NOM SUPER-cunning OBL all GEN animal
 ‘The most cunning among all animals was the snake...’ [2002855]

(46), (58), (67), (79)),¹⁴ and personal nouns in inclusory constructions [Reid 2009].¹⁵ The nonactor forms are used to introduce any other roles in nonsubject positions (not requiring prepositions¹⁶) (examples (16), (43), (57), (75), (97)). In the nonpersonal system, genitive is used to introduce actors in undergoer voice constructions (examples (5), (52), (53), (59), (64), (79), (88)), possessors (examples (38), (40), (42), (53), (58), (62), (64), (96)), included substantives,¹⁷ and some of the roles denoted by the nonactor forms in the personal case marking system (e.g. patient (examples (15), (19), (22), (24), (43), (48), (60), (65), (66), (92), (97)), ground

¹⁴ Strictly speaking, *ng*, *ni*, and *nina/nila* mark not only possession, but any kind of pertinence relation (e.g. *ang doktor ng pamilya* [NOM doctor GEN family] ‘the family doctor’ or *hari ng gubat* [king GEN forest] ‘king of the jungle’ [Malicsi 2012: 78; Gallego, Zubiri 2013: 48].

¹⁵ (i) *Hiwalay=na=kami ni Leslie.*
 separated=IAM=1PL.NOM PERS.SG.ACT PN
 ‘Leslie and I have already broken up.’ [183159]

¹⁶ Tagalog prepositions are defined here as lexemes that mark the semantic roles of substantives, and which are normally used in front of a case-marked substantive (with a number of exceptions, including unmarked temporal adjuncts and some prepositions that combine with nominative substantives whose nominative marker is dropped if they are indefinite: e.g. *bilang ang taong ito* ‘as this person’ vs. *bilang tao* ‘as a person’). Tagalog has at least four types of case marking frames for prepositions: nouns functioning as prepositions in the phrase *sa ... ng* (e.g. *sa kabila ng* ‘despite’), prepositions with oblique/nonactor marking (*para sa* ‘for’), with genitive/actor marking (e.g. *gawa ng*), and with nominative marking (e.g. *matapos ang* ‘after’). Some prepositions can be used in different frames: for instance, *imbes* ‘instead of’, *matapos* ‘after’, *makalipas* ‘after’, *pagkatapos* ‘after’, *bago* ‘before’ can be used with both genitive/actor or nominative marking (with differing frequencies of each marking), while *puwera* ‘except for’ and *maliban* ‘except for’ (I thank Ekaterina A. Baklanova for directing my attention to the latter preposition) can be used with all three cases. Cf. [Malicsi 2012] for a different classification.

¹⁷ (i) *Nag-ka-tingin-an=kami ng pinsan=ko.*
 PFV.STEM-MOD-look-REC=1PL.NOM GEN cousin=1SG.ACT
 ‘My cousin and I looked at each other.’ [187453]

of directed motion,¹⁸ etc.), while oblique introduces another set of roles, excluding actor and substantives in inclusory constructions, but partially overlapping with that of genitive (e.g. predicative possessor,¹⁹ patient (example (78)), location (examples (8), (18), (23), (39), (45)), ground of directed motion (example (8), (23), (60), (76), (89), (96)), recipient (example (19), (41), (77)), etc.). In addition, all case markers co-occur with certain prepositions and conjunctions (example (37), (39), (41), (43), (47), (53), (66), (67), (69), (71), (72), (73), (74), (91)).

The traditional approach to describing the Tagalog case system distinguishes three cases for all types of arguments (nonpersonal substantives and demonstratives, personal nouns and pronouns): nominative, genitive, and oblique (cf. [Reid, Liao 2004; Gallego 2015: 66]), suggesting that there is a complete symmetry between the four systems. Traditionally, the markers *ni* and *nina / nila* are labeled as genitive, while *kay* and *kina / kila* are labeled as oblique (cf. [Schachter, Otanes 1972: 93]), ignoring the discrepancy between the functions of the genitive and oblique nonpersonal case markers, on the one hand, and ‘genitive’ and ‘oblique’ personal case markers, on the other, in many Philippine languages. Himmelmann draws a similar distinction between the systems of personal and nonpersonal case markers, though in different terms [Himmelmann 2005: 145].

2.2. Inversion constructions

Tagalog has three constructions where certain constituents are moved from a position to the right of the predicate into a position before it: topicalization, fronting, and nonemphatic inversion.

¹⁸ (i) ⟨...⟩ *p<um>unta=ako ng USA.*
 <AV>go=1SG.NOM GEN PLN
 ‘...I went to the USA.’ [22640835]

¹⁹ (i) *Sa kapatid=ko ang mga iyan, ⟨...⟩*
 OBL sibling=1SG.ACT NOM PL MED.NOM
 ‘Those belong to my brother,..’ [189687786]

Topicalization (also referred to as *ay*-inversion and contrastive inversion [Schachter, Otanes 1972: 485–500], *ay*-inversion and topicalization [Kroeger 1993: 123–124], external and internal topicalization [Lee, Billings 2008: 246], sentence initial left-detached position [Nagaya 2007: 353]) moves a constituent into the leftmost position in the sentence detached from the rest of it either by the topic marker *ay/=y* (in formal style) or a pause [Schachter, Otanes 1972: 485]. It functions as the information structure topic [Rachkov 1966; Frolova 1983; Shkarban 1989: 93; Himmelmann 1991: 10], contrastive²⁰ or not [Nagaya 2007: 363], or presupposed element [Nagaya 2007: 361], and never contains focal information [Kroeger 1993: 63–64]. This construction, exemplified in (1), is widespread in Philippine languages [Reid, Liao 2004: 447].

- (1) *Siya ay isa-ng tunay na kaibigan* ⟨...⟩²¹
 3SG.NOM TOP one=LK genuine LK friend
 ‘He is a real friend ⟨...⟩’ [2083]²²

A number of constituents can undergo topicalization. Apart from nominative and oblique / non-actor substantives, prepositional phrases, possessors, temporal, spatial, manner, modal and sentential adverbials, conditional, concessive and temporal clauses, precedence *pag*-forms, phrases with *ni*... ‘not even...’ and *kahit* ‘although’ [Frolova 1983; Schachter, Otanes 1972: 485–495; Shkarban 1989: 100; Himmelmann 1991: 10; Nagaya 2007: 363–365;

²⁰ The examples cited in [Kroeger 1993: 67–68] for which the author claims the *ay*-inverted elements get pragmatic focus seem to be instances of contrastive topic.

²¹ All examples in this paper come from the corpus, unless otherwise specified. Standard Tagalog orthography is followed here. Some corpus examples with nonstandard orthography are respelled (e.g. *lenguwahe* is respelled as *lengguwahe*), which means that in some forms the hyphen is part of the normal spelling of the word, rather than a morpheme boundary marker reflected in glosses. Glottal stops are not represented in standard Tagalog orthography, which means that all roots that are spelled with an initial vowel, in fact, start with a glottal stop.

²² All corpus examples in this paper are provided with their token number, which can be used within the corpus query language notation [#] to search for a token with the given position. For instance, the token number for example (1) can be searched for using ‘[#2083]’ in a CQL query in the concordance.

Frolova 2017], topicalization can also be applied to certain particles [Kaufman 2006: 158] ((=)*halos* ‘almost’, (=)l(*am*)*ang* ‘only’, the optative markers (=) *sana*, *kahimanawari*, etc.; see example (16)), conjunctions, and clauses in complex sentences without conjunctions. The latter case is illustrated in (2), while the data on topicalization of conjunctions are presented in Section 3.

- (2) *Malinis na malinis=na ang mga palad ni*
 clean LK clean=IAM NOM PL palm PERS.SG.ACT
Ore ay kuskos=pa=rin=siya nang kuskos.
 PN TOP rub=CONT=also=3SG.NOM LK rub
 ‘Ore’s palms were already very clean, but he still kept rubbing them.’ [80638988]

In some instances of topicalized nominative constituents (e.g. in phrases with quantifiers *bawat* ‘each’, *lahat* ‘all’, *isa* ‘one’), the nominative case marker is optionally omitted:

- (3) *Isa=ng posible=ng explanation ay dahil*
 one=LK possible=LK TOP because
maaari=ng akusa-han²³ si Jesus <...>
 possible=LK accuse-UV[INF] PERS.SG.NOM PN
 ‘One possible explanation is because Jesus could be accused of <...>’
 [26960814]

Omission of *ang* in non-topicalized position is also common in colloquial Tagalog, like in (7):

- (4) <...> *medyo s<um>akit (ang) ulo=ko.*
 bit <AV>ache[INF] NOM head=1SG.ACT
 ‘<...> my head started aching a bit.’ [17253172 and 24018858]

²³ There is no one-to-one correspondence between the Tagalog voice affixes (<*um*>, *-in*, *-an*, and *i-*) and the semantic roles they require as their subjects, and voice paradigms display a lack of predictability. In addition, in many instances there seems to be very little objective basis for assigning purely semantic labels to Tagalog voice constructions (e.g., patient voice, theme voice, conveyance voice, and other widespread labels). Therefore, I simply label voice affixes as denoting actor voice or undergoer voice, without specifying what particular nonactor role the latter require.

In fronting (also referred to as emphatic inversion [Schachter, Otanes 1972: 496], adjunct fronting [Kroeger 1993: 64], clause-initial precore slot fronting [Nagaya 2007: 353]), a constituent is fronted to the position before the predicate not separated from the rest of the sentence by any marker or pause. Fronting can be applied to oblique / nonactor phrases, temporal adverbials without case markers, prepositional phrases [Schachter, Otanes 1972: 496–498]. Thus, it seems that any constituent that can be fronted can also undergo topicalization. Fronted elements constitute narrow focus [Kroeger 1993: 64; Nagaya 2007: 353]. A construction with a fronted oblique demonstrative denoting a location is given in (5), where the focal element is in italics in the translation:

- (5) *Doon=siya na-aresto*²⁴ *ng mga pulis.*
 DIST.OBL=3SG.NOM PFV.MOD-arrest[UV] GEN PL police
 ‘He was arrested by police *there.*’ [16222717]

Nonemphatic inversion attaches certain manner and temporal adverbials with the linker *na/=ng*²⁵ to the left side of the predicate without any change in information structure [Schachter, Otanes 1972: 498]. The adjective *mabilis* ‘quick’ in nonemphatic inversion modifies the predicate verb in (6).²⁶

²⁴ For ease of presentation, the verbal prefixal forms *mag-*, *maN-*, *maki-*, *ma-*, and *nag-*, *naN-*, *naki-*, *na-* complexes are rendered here as portmanteau morphemes. However, I support the suggestion made in [Wolff 1973: 72] and [De Guzman 1978: 87] to treat each of these prefixal forms as a complex, consisting of the infix <*um*> or <*in*> and a corresponding stem-deriving prefix (*pag-*, *paN-*, etc.), as it allows for a more adequate generalization of the Tagalog morphology (cf. [Kroeger 1998: 8; Nagaya 2007: 347]).

²⁵ According to [Schachter, Otanes 1972: 498], frequency adverbials with *beses* and *ulit* ‘times’ can occur in nonemphatic inversion without a linker. However, such constructions seem to be instances of fronting.

²⁶ Different positions are preferred for different manner adverbs. *Tumakbo nang mabilis* ‘run LK quick’ has 40 tokens in the corpus vs. 141 tokens for *mabilis na tumakbo* ‘quick LK run’. On the other hand, *naglakad nang mabilis* ‘walk LK quick’ has 7 tokens vs. 62 tokens for *mabilis na naglakad* ‘quick LK walk’, *nagsalita nang mabilis* ‘speak LK quick’ has 0 tokens vs. 3 tokens for *mabilis na nagsalita* ‘quick LK speak’,

- (6) *Mabilis na nag-reply si Stephanie.*
 quick LK AV.PFV.stem-reply PERS.SG.NOM PN
 ‘Stephanie quickly replied.’ [6969228]

2.3. Clitics

Apart from clitics that do not change their position (e.g. the sentence-final particles: =*a* (counter-expectation), =*e* (cause, contradiction, or regret), =*o* (polite request, attention attraction), =*ha* (importunity, sarcasm) [Schachter, Otanes 1972: 461–463], and clitic forms of some function words, like the topic marker *ay/=y*, the linker *na/=ng*, etc.), Tagalog has clitics whose position depends on the position of other constituents in the sentence. Hereafter, the discussion is focused only on the latter type.

Tagalog clisis system is described as having two positioning types: enclitic personal pronouns and enclitic particles [Schachter, Otanes 1972: 187].

The personal pronoun paradigm (Table 3) includes 33 forms with four categories: case, number, person, and predicativity. The person category involves four values: 1, 2, 1/2, and 3. 1/2-person pronouns refer to both first and second persons (‘you and me / us’). Such pronouns are often described as dual or inclusive (cf. [Schachter, Otanes 1972: 88]). However, the representation of such forms in this way leads to introducing the category of inclusivity manifested only by the first person pronouns, while creating some gaps in the paradigm, which results in a less economical description of the system.²⁷ The number category involves the opposition of minimal and non-minimal membership, as the 1/2-person pronouns cannot be characterized as singular [Reid 1971; Foley 1997: 111], although the more traditional terms ‘singular’ and ‘plural’ are used throughout this paper for the sake of the readers’ convenience. Predicativity is

while *sumigaw nang malakas* ‘shout LK loud’ has 26 tokens vs. 7 tokens for *malakas na sumigaw* ‘loud LK shout’, *nagsalita nang malakas* has 18 tokens vs. 4 tokens for *malakas na nagsalita* ‘loud LK speak’, and *sumagot nang malakas* ‘answer LK loud’ has 4 tokens vs. 0 tokens for *malakas na sumagot* ‘loud LK answer’.

²⁷ This approach is adopted from [Cunningham, Goetz 1963].

relevant only for actor forms: the predicative forms, apart from being used as a predicate (this function is not available for non-predicative forms), also can occur preposed to a modified constituent with possessor or actor reference.

Table 3. Tagalog personal pronouns

Number	Person	NOM	ACT		NACT
			PRD	NPRD	
Minimal membership	1	(=) ²⁸ (a)ko	Akin	=ko	s(a)akin
	2	ikaw/=ka	Iyo	=mo	sa(i)yo
	1/2	{(=)kata}	{kanita}	{=nita}	{sakanita}
	3	(=)s(i)ya	kan(i)ya	=n(i)ya	sakan(i)ya
Non-minimal membership	1	(=)kami	Amin	=namin	s(a)amin
	2	(=)kayo	Inyo	=(ni)nyo	sainyo
	1/2	(=)tayo	Atin	=natin	s(a)atin
	3	(=)sila	kanila	=nila	sakanila

The second person minimal-membership pronoun has two forms: *=ka* is an enclitic, while *ikaw* is an independent form that is used in the topic and predicate positions, as well as the subject position when there are other constituents between it and the predicate.

There is another minimal-membership dual pronoun, *=kita*, which is difficult to accommodate in any tabular form, since it stands for the non-existing combination *=ko=ka* (i.e. =1SG.ACT=2SG.NOM) [Schachter, Otanes 1972: 184].

Traditionally, actor pronouns are treated as counterparts of the genitive nonpersonal marker, while nonactor pronouns are treated as counterparts

²⁸ Parentheses in *Table 1* indicate two cases: (i) optional sounds, (ii) the fact that some forms are used either as an enclitic, or not depending on their position. Curly brackets mean that the 1/2 minimal membership pronouns are obsolete in Tagalog (although, at present it is unclear if all Tagalog dialects have lost them), while their counterparts are still widely used in many other Philippine languages

of the oblique nonpersonal marker (cf. [Schachter, Otnes 1972: 88]). However, the same reinterpretation of this paradigm can be done as in the case of the personal case markers, since, for example, while nonpersonal Patient or Path can be marked with the genitive marker *ng*, personal Patient or Path has to be realized as a nonactor pronoun.

The nonactor pronouns and predicative forms of actor pronouns are nonenclitic, while the nominative forms are nonenclitics when topicalized or function as the predicate [Schachter, Otnes 1972: 184].

Some other lexical units can optionally behave like enclitic personal pronouns, including personal nouns and demonstratives [Schachter, Otnes 1972: 184; Billings 2005: 317].

Similar to other Philippine languages [Billings, Kaufman 2004: 18–20], in Tagalog enclitic pronouns are obligatorily hosted by a predicate that does not have any fronted constituent (7), a fronted constituent in emphatic inversion (8), the negation marker (*hindi*) (9) [Schachter, Otnes 1972: 183–193], and the prohibitive marker (*huwag*) (10) [Malicsi 2013: 75].

- (7) *Masaya=siya ngayon.*
happy=3SG.NOM today

‘He is happy today.’ [Schachter, Otnes 1972: 183]

- (8) *Sa maliit na bahay sa probinsya=siya pu~punta.*
OBL small LK house OBL province=3SG.NOM PROSP~go[AV]

‘It’s the little house in the province that he’s going to.’ [Schachter, Otnes 1972: 188]

- (9) *Hindi=siya masaya ngayon.*
NEG=3SG.NOM today

‘He isn’t happy today.’ [Schachter, Otnes 1972: 183]

- (10) *Huwag=siya=ng mayabang.*
PROH=3SG.NOM=LK boastful

‘Let him not be boastful.’ [9521392]

There is also a number of optional hosts [Schachter, Otnes 189–193]. For instance, in a construction with the interrogative *bakit* ‘why’ enclitic pronouns can be hosted by the predicate (11) or by the interrogative (12).

- (11) **Bakit** *t<in>anong=mo=siya.*
 why <PFV>ask[UV]=2SG.ACT=3SG.NOM
 ‘Why did you ask him?’ [Schachter, Otnes 1972: 191]
- (12) **Bakit=mo=siya** *t<in>anong.*
 why=2SG.ACT=3SG.NOM <PFV>ask[UV]
 ‘Why did you ask him?’ [Schachter, Otnes 1972: 191]

Unlike enclitic particles, enclitic pronouns that are arguments of the sentence predicate cannot be hosted by topicalized constituents [Schachter, Otnes 1972: 189; Billings, Kaufman 2004: 22] (example (13)).

- (13) **Bukas**(*=*siya*) *ay a~alis*(*=*siya*).
 tomorrow=3SG.NOM TOP PROSP~leave[AV]=3SG.NOM
 ‘Tomorrow, he’ll leave.’ [Schachter, Otnes 1972: 189]

Enclitic particles — often referred to as adverbial in works on Philippine languages [Rubino 1997: 320–353; McFarland 2001; Billings, Kaufman 2004: 22; Kaufman 2006: 158; Robinson 2008: 195–210], as they are said to encode “aspectual, mood, evidential and functional” meanings [Kaufman 2006: 158]²⁹ — include the following well-known items: the question markers =*ba* and =*baga*, (=)*kasi* ‘because’, the speculative or emphatic marker =*kaya*, the reportative marker =*daw* / =*raw*, =*din* / =*rin* ‘also’, the honorific markers (=)*ho* and (=)*po*, (=)*l(am)ang* ‘only’, the concessive marker =*man*, =*muna* ‘at first’, the iamitive and imperative marker =*na*, the continuative marker =*pa*, the marker of emphasis, information structure topic change, reluctance, and polite imperative (=) *naman*, the reiterative marker =*nga*,³⁰ the counter-expectation marker =*pala*, the optative marker (=)*sana*, =*tuloy* ‘as a result’, (=)(*y*)*ata* ‘perhaps’ [Schachter, Otnes 1972: 411], (=)*muli* / (=)*juli* / (=)*julit* ‘again’, (=)*agad* ‘right away’, (=)*siguro* ‘perhaps’, (=)*talaga* ‘indeed’ [McFarland 2001],

²⁹ Other generalizations can also be found, e.g. “aspectual clitics and the question marker *ba*” [Spencer, Luis 2012: 58].

³⁰ =*nga* is usually translated as ‘indeed’, however in many instances this translation is not applicable, as illustrated by its use in a variety of constructions, like the speaker’s agreement (i), insistence (ii), or doubt (iii). It seems that the more generalized

‘the optative marker (=)*nawa*, and (=)*marahil* ‘probably’ [Malicsi 2013: 74]. There are also at least the following enclitic particles, not listed in the cited sources: (=)*halos* ‘almost’, (=)*dapat* ‘must’, (=)*lagi* ‘always’, (=)*lahat* ‘all’, (=)*malamang* ‘likely’, (=)*umano*, (=)*dumano*, (=)*mismo* ‘self/very’, (=)*kamo* ‘as you said’. Examples (14)–(15) illustrate the enclitic use of (=)*halos* and (=)*dapat*.

(14) *Hindi=ko=halos=siya ma-kilala...*
 NEG=1SG.ACT=almost=3SG.NOM MOD-recognize[UV]
 ‘I almost couldn’t recognize him...’ [138920884]

(15) ⟨...⟩ *mag-hanap=ka=na=dapat ng trabaho.*
 AV.STEM-search[INF]=2SG.NOM=IAM=must GEN work
 ‘⟨...⟩ you should look for a job now.’ [9534004]

As noted in [Kaufman 2006: 158], and as indicated with the use of parentheses in the list above, some enclitic particles can also function as non-enclitics when used in the topic position. Kaufman presents an example with the optative marker (=)*sana* in the topicalized position. Example (16) illustrates the same type of construction with (=)*l(am)ang* ‘only’.³¹

(16) *Lamang ay k<in>a~kailangan=ninyo=ng iwan*
 only TOP <IPFV>IPFV~need[UV]=2PL.ACT=LK leave[UV.INF]

meaning of =*nga* is that of reiteration of previously mentioned or presupposed information.

- (i) *Siguro=nga.*
 perhaps=REIT
 ‘Yes, it seems so.’ [6124258]
- (ii) *Bakit=nga=ba ga-nito?!*
 why=REIT=Q like-PROX.GEN
 ‘But why is it like this?!’ [7564994]
- (iii) *Talaga... di=nga?*
 indeed NEG=REIT
 ‘Really... is it so?’ [705980]

³¹ The enclitic use of (=)*l(am)ang* ‘only’ is shown in examples (19), (33), (37), (53), (58), (61), (70), (72), (79), (87), (95).

sa akin ang inyo=ng pangalan at <...>
 1SG.NACT NOM 2PL.ACT=LK name and

‘The only thing is you need to give me your name and <...>
 [7646251]

Some enclitics, though, occur as nonenclitics only when used as a standalone utterance, for instance, the honorific marker *ho* in (17), which is used as a polite request to repeat or further elaborate on what was said by the other party in the conversation.³²

- (17) *“Inum-in=mo yun=g una.”*
 drink-UV[NEUT]=2SG.ACT DIST.NOM=LK first
“Ho? Bakit=ako i~inom nito!”
 HON why=1SG.NOM PROSP~drink[AV] PROX.GEN
 “Drink the first one.”
 “What? Why would I drink this!” [3900241]

This also pertains to some combinations of enclitic particles whose meaning differs from the combined meaning of its constituent parts, like the combination (=)*na=naman*, which stands for the emphatic ‘again’, as shown in (18) [Malicsi 2013: 73].³³

- (18) *Nag-ka-problema=po=kasi kanina sa*
 PFV.STEM-ACQ-problem[AV]=HON=because earlier OBL
trak kaya=po na-tagal-an=na=naman. Na=naman?!
 truck therefore=HON PFV.MOD-take_long-UV=IAM=EMPH IAM=EMPH
 ‘Because I’ve got a problem with the truck, so it took me long again.
 Again?!’ [51530534]

In addition to the hosts available to enclitic pronouns, enclitic particles can be hosted by a topicalized constituent, like in (19) and (20).

³² Some enclitics have nonenclitic homonyms, like *tuloy* ‘stay/continue’ vs. the enclitic =*tuloy* ‘as a result’.

³³ There are also some other enclitic combinations of this kind, e.g. (=)*nga=pala* (‘(=)REIT=CE’) ‘by the way’. There are also combinations whose meaning does not equal the combined meaning of its constituents that do not occur as nonenclitics, like =*na=l(am)ang* (‘=IAM=only’) ‘instead’.

- (19) *Nito=ng Pebrero=lang ay nag-propose=na*
 PROX.GEN=LK February=only TOP PFV.STEM-propose[AV]=IAM
si Arthur Solinap ng kasal sa kanya=ng
 PERS.SG.NOM PN PN GEN wedding OBL 3SG.ACT=LK
long-time girlfriend na si Rochelle Pangilinan.
 LK PERS.SG.NOM PN

‘Just this February, Arthur Solinap proposed to his long-time girlfriend Rochelle Pangilinan.’ [172952]

- (20) *Anak=ko, ako='y maligaya*
 offspring=1SG.ACT 1SG.NOM=TOP happy
sapagkat ikaw=na ay d<um>ating.
 because 2SG.NOM=IAM TOP <AV>arrive
 ‘My child, I am happy because you have arrived’ [49083218]

According to [Schachter, Otnes 1972: 429–435], Tagalog enclitic particles are split into four groups with varying positional properties: (i) *=ba*, (=)*kasi*, *=kaya*, *=man* are hosted by a sentence-initial predicate, and a topicalized argument or adverbial; (ii) *=daw* / *=raw*, *=din* / *=rin*, (=)*ho*, (=)*po*, (=)*naman*, *=nga*, *=pala*, (=)*sana*, *=tuloy*, (=)(*y*)*ata* are hosted by an initial or non-initial predicate, and a topicalized argument or adverbial; (iii) (=)*l(am)ang* and *=muna* are hosted by an initial or non-initial predicate, and a topicalized argument; (iv) *=na* and *=pa* are hosted by an initial or non-initial predicate only.³⁴ However, counter-evidence to these observations can be found in corpus data. For instance, in (19) *=lang* is hosted by the topicalized adverbial *nito=ng Pebrero*, while in (20) *=na* is hosted by the topicalized argument *ikaw*. Other discrepancies with corpus data can be

³⁴ As follows from Schachter, Otnes’ observations regarding the positional properties of Tagalog enclitic particles, some enclitic particles can occupy different positions in the sentence without any apparent semantic change (examples (i)–(ii)). Other enclitic particles are hosted within the clause that they semantically modify, like *=l(am)ang* ‘only’ in example (iii) where it modifies *chismis* ‘gossip’ (see Footnote 9 for arguments in support of the clausehood status of substantive phrases).

- (i) *Sila=muna ay sa-sayaw ng pandanggo*
 3PL.NOM=at_first TOP PROSP~dance[AV] GEN fandango

found in Schachter, Otones' description of some of the properties of Tagalog enclitic particles. For example, contrary to their claim [Schachter, Otones 1972: 433], =*na* and =*pa* can be hosted by the interrogative *bakit* 'why', like in (21):

- (21) *Bakit=pa=kami ma-ta~takot?*
 why=CONT=1PL.NOM AV.STEM-PROSP~get_scared
 'Why would we get scared?' [1950056]

Unlike enclitic pronouns, enclitic particles can also be hosted by the initial constituent of a topic [Schachter, Otones 1972: 429], like in (22), and a fronted constituent [Schachter, Otones 1972: 434; Kaufman 2008: 33; Kaufman 2010b: 192], like in (23).

- (22) *Bukas=ba ng gabi ay*
 tomorrow=Q GEN evening TOP
sa~sayaw=sila ng pandanggo?
 PROSP~dance[AV]=3PL.NOM GEN fandango
 'Tomorrow night, will they dance a fandango?' [Schachter, Otones 1972: 429]
- (23) *Sa bahay=ba sa probinsya=siya*
 OBL house=Q OBL province=3SG.NOM

bukas ng gabi.
 tomorrow GEN evening

'They will first dance a fandango tomorrow night.' [Schachter, Otones 1972: 431]

- (ii) *Sila='y sa~sayaw=muna ng pandanggo*
 3PL.NOM=TOP PROSP~dance[AV]=at_first GEN fandango

bukas ng gabi.
 tomorrow GEN evening

'They will first dance a fandango tomorrow night.' [Schachter, Otones 1972: 431]

- (iii) *Wag=kayo=ng man-i~niwala sa chismis=lang.*
 PROH=2PL.NOM=LK AV.STEM-PROSP~believe OBL gossip=only

'Don't believe in something that is just a gossip.' [152755600]

This paper is not concerned with the questions of what individual enclitics can be hosted by different conjunctions and why. Instead, it aims to find out what conjunctions can generally host any of the enclitics.

pu~punta?
 PROSP~go[AV]

‘Is it the house in the province that he’s going to?’ [Schachter, Otanes 1972: 434]

Lee, Billings label the enclitic pronoun ordering type as verb-adjacent ordering, while that of the enclitic particles as Wackernagel ordering, and suggest to use constructions that contain both an inverted adjunct and a negation marker at the same time as a test for the ordering type of a clitic [Lee, Billings 2005: 241–242].

In addition, some Philippine languages allow enclitic pronouns to be hosted by some conjunctions.³⁵ According to [Billings, Kaufman 2004: 19–20] and [Lee, Billings 2005: 243, 248] claim that Tagalog conjunctions cannot host enclitic pronouns (see example (24)),³⁶ while some conjunctions in Mapun (example (25)) [Billings, Kaufman 2004: 19–20], Tausug (example (26)) [Lee, Billings 2005: 243; Lee, Billings 2008: 193–194], Masbatenyo (example (27)) [Lee 2008], Binukid (example (28)) [Peng, Billings 2008: 187–190, 200], and Maranao (example (29)) [Kaufman 2010a: 184] can.

TAGALOG:

- (24) ***Kung hindi=ka nag-luto ng itlog, <...>***
 if NEG=2SG.NOM AV.PRF-cook GEN egg

‘If you didn’t cook eggs, <...>’ [Billings, Kaufman 2004: 20; Lee, Billings 2005: 251]

MAPUN:

- (25) ***Bong=ko ya’ pa-tagong nangis, <...>***
 if=2SG.NOM NEG AV-stop AV.cry

‘If you don’t stop crying, <...>’ [Billings, Kaufman 2004: 20; Lee, Billings 2005: 251]

³⁵ Complementizers in the terminology used in [Billings, Kaufman 2004] and [Lee, Billings 2005].

³⁶ This is also said to be true for most other Central Philippine languages, including Tagakaulo, Kaagan, Mansaka, Davawenyo, and Mamanwa [Lee, Billings 2005: 248, 251].

TAUSUG:

- (26) **Bang=siya** *di'* *m-agad* ⟨...⟩
 if=3SG.NOM NEG AV-go_with

‘If he won’t go with us, ⟨...⟩’ [Lee, Billings 2005: 251]

MASBATENYO:

- (27) **Didi=ka=lang** *tindog* **agod=ka** *dili* *ma-init-an*.
 here=2SG.NOM=just stand[AV] so.that=2SG.NOM NEG ABIL-hot-RV

‘Stand here so that you don’t get sunstroke.’ [Lee 2008: 128]

BINUKID:

- (28) **ta daw=ka=kandin** *mabutawanan*.
 so.that=NOM.2PL=[OBL]/GEN/.3SG release

‘[Poke the crocodile in the eye] so that it will let go of you.’ [Peng, Billings 2008: 188]

MARANAO:

- (29) **Kaan=ta** *di'* *pe-ketey*.
 so.that=1,2S.NOM NEG FUT-delay

‘So that we shall not be delayed.’ [Kaufman 2010a: 184]

Conjunctions can function as enclitic hosts in some other Austronesian languages as well, cf. Seediq [Holmer, Billings 2014] in example (30) and South Sulawesi languages [Kaufman 2008] in example (31).

SEEDIQ:

- (30) ⟨...⟩ **ado=ku** *m-beyax* *t<m>alang* *yaku*.
 because=1SG.NOM AV-strong <AV>run 1SG

‘⟨...⟩ because I am good at running.’ [Holmer, Billings 2014: 113]

SEKO PADANG:

- (31) **I=koi:=ka'** *mang-keki'* *tu-lino* ⟨...⟩
 COND=1S.F=1S.A AV-bite person-world

‘If I bite a person ⟨...⟩’ [Kaufman 2008: 39]

On the other hand, some sources claim that some Tagalog conjunctions can host enclitics as well. According to [Kaufman 2010a: 184] only one Tagalog conjunction, *kaya* ‘therefore’, can host enclitics. According

to [McFarland 2001: 11–13], the situation is somewhat more complicated, as *kaya* ‘therefore’ can host both enclitic particles and pronouns when it is not topicalized, like in (32), and only enclitic particles when it is topicalized, like in (33).³⁷ In addition, *kasi* ‘because’ can host enclitic particles, like in (34), *kundi* ‘if not’ (see example (35)) and *para* ‘in order to’ (see example (36)) can host both enclitic particles and pronouns.

- (32) *Mahal=ko=po si Darwin kaya=ako*
 beloved=1SG.ACT=HON PERS.SG.NOM PN therefore=1SG.NOM
na-ri-rito.
 PRED-PRED~PROX.OBL

‘I love Darwin, that’s why I’m here.’ [McFarland 2001: 11]

- (33) *Kaya=lang ay masyado=ng personal.*
 therefore=only TOP excessive=LK personal

‘It’s just that it’s too personal.’ [McFarland 2001: 11]

- (34) *Kasi=naman ikaw.*
 because=EMPH 2SG.NOM

‘Because you (are like that).’ [McFarland 2001: 11]

- (35) *Kundi=ka d<um>ating, wala=sana=ng problema.*
 if_not=2SG.NOM <AV>arrive[PFV] NEG.EXIST=OPT=LK problem

‘If you hadn’t come, there wouldn’t have been a problem.’ [McFarland 2001: 13]

- (36) *Wala=ng t<um>awag para=siya k<um>ain.*
 NEG.EXIST=LK <AV>call[PFV] so_that=3SG.NOM <AV>eat[INF]

‘No one called for him to eat.’ [McFarland 2001: 13]

According to [Schachter, Otones 1972: 188, 191, 433–434, 463–483], Tagalog conjunctions differ along three parameters: the ability to host enclitic particles, the ability to host enclitic pronouns, and obligatoriness of being a host. Four groups of conjunctions are thus formed: (i) nonhosts, (ii) optional hosts for particles, (iii) optional hosts for both

³⁷ McFarland’s example in (33) actually contains the combination *kaya=lang* ‘but the thing is’, rather than the conjunction *kaya* ‘therefore’.

particles and pronouns, and (iv) obligatory hosts for both particles and pronouns (*Table 4*).

Table 4. Tagalog conjunction properties (from [Schachter, Otanes 1972])

Conjunctions	Particles	Pronouns
<i>at</i> ‘and’, <i>kung</i> ‘if’, <i>hanggang</i> ‘until’, <i>hangga’t</i> ‘while’, <i>bagama’t</i> ‘although’, <i>nang</i> ‘when’, <i>o (kaya)</i> ‘or (else)’, <i>pero</i> ‘but’, <i>dahil sa</i> ‘because’, <i>dahilan sa</i> ‘because’, <i>gawa ng</i> ‘because’, <i>(sa)pagka’t</i> ‘because’, <i>upang</i> ‘so that’	–	–
<i>(kung) sakali</i> ‘in case if’, <i>noon</i> ‘when’, <i>sa sandali</i> ‘the moment that’, <i>(mag)buhat</i> ‘since’, <i>(mag)mula</i> ‘since’, <i>bagaman</i> ‘although’, <i>maliban</i> ‘unless’, <i>bukod</i> ‘besides’, <i>dangan</i> ‘were it not that’, <i>dahil</i> ‘because’, <i>(ya)yaman + na/-ng</i> ‘because’	(+)	–
<i>(ka)pag(ka)</i> ‘if/when’, <i>tuwi</i> ‘whenever’, <i>matapos</i> ‘after’, <i>pagkatapos</i> ‘after’, <i>porke</i> ‘because’, <i>maski</i> ‘even if’, <i>imbis</i> ‘instead’, <i>(sa) lugar</i> ‘instead’, <i>(sa) halip</i> ‘instead’, <i>miyentras</i> ‘while’, <i>para</i> ‘so that’, <i>(at) saka</i> ‘and then’, <i>haba</i> ‘while’, <i>samantala</i> ‘while’	(+)	(+)
<i>bago</i> ‘before’, <i>kundi</i> ‘if not’	+	+

These observations are, however, contradicted by the corpus data, as shown in *Section 3*.

Importantly for any discussion concerning how the data presented below compare to what has been reported in the earlier works mentioned above, in most cases the latter do not provide any explicit indications of where their data comes from. It is unclear what data the authors of [Billings, Kaufman 2004: 19–20], [Lee, Billings 2005: 243, 248], and [Kaufman 2010a: 184] used to make their generalizations regarding the alleged (in) ability of Tagalog conjunctions to host enclitics, as they do not provide any information on the source of their Tagalog data. Schachter, Otanes do not provide such information either, although they state elsewhere that their description of Tagalog pronunciation is based on the Manila dialect spoken by “recent college graduates” [Schachter, Otanes 1972: 1]. Also, there are multiple references to “speakers” without any further elaboration

throughout their work, from which one can assume that each of their data points comes from an undefined number of native speakers. McFarland [McFarland 2001: 11–13] is the only one who explicitly states the source of his data on Tagalog conjunctions, which is a 1 million-word corpus of 50 Filipino novelettes published in the 1990s [McFarland 2001: 1].

3. Data presentation

The data presented below show that there is significant variation in the distribution of the three properties discussed above — ability to undergo topicalization, to host enclitic particles, and to host enclitic pronouns — in Tagalog conjunctions. The conjunctions under study here are split into five groups based on the combinations of these properties. As shown in *Table 5*, the ability to be topicalized and the ability to host pronouns both entail the ability to host particles.

Table 5. Conjunction properties

	group 1	group 2	group 3	group 4	group 5
topicalization	+	–	+	–	–
particles	+	+	+	+	–
pronouns	+	+	–	–	–
conjunctions	<i>kaya</i> ‘therefore’ <i>para</i> ‘so that’	<i>bago</i> ‘before’	<i>palibhasa</i> ‘because’ (=) <i>kasi</i> ‘because’ <i>tatal</i> ‘because’ (<i>kung</i>) <i>dangan=kasi</i> ‘because’ (<i>pa</i>) <i>paano=kasi</i> ‘because’ <i>porke</i> (=’ <i>t</i>) ‘because’ <i>komo</i> ‘because’ <i>kesyo</i> ‘because’ <i>dahil</i> ‘because’ (<i>sa</i>) <i>pagka</i> (=’ <i>t</i>) ‘because’ <i>dangkasi</i> ‘because’	(<i>ya</i>) <i>yaman=g</i> ‘because’ <i>dahil sa</i> ‘because’ <i>dahilan sa</i> ‘because’ <i>gawa ng</i> ‘because’ <i>bilang</i> ‘because’	<i>anupa=t</i> ‘therefore’ <i>at</i> ‘and’ <i>nang</i> ‘when’

	group 1	group 2	group 3	group 4	group 5
			<i>mangyari</i> ‘because’ <i>sa dahilan=g</i> ‘because’ <i>sa kadahilan(an)=g</i> ‘because’ <i>manyapa(’)t</i> ‘because’ <i>samakatuwid</i> ‘therefore’ <i>nguni(=’)</i> ‘but’ <i>pero</i> ‘but’ <i>kung</i> ‘if’		

The following procedure was used to establish the properties of each conjunction under study. The ability to undergo topicalization was established through searches for combinations of a conjunction with the topic marker *ay/=’y* with and without an intervening unspecified word.³⁸ The ability to host enclitic particles was checked through searches for combinations of a conjunction with the following particles: *=ba* ‘Q’, *=baga* ‘Q’, *=daw/ =raw* ‘REP’, *=din/ =rin* ‘also’, *(=)ho* ‘HON’, *(=)po* ‘HON’, *(=)l(am)ang* ‘only’, *=man* ‘CONC’, *=muna* ‘at first’, *=na* ‘IAM’, *=pa* ‘CONT’, *(=)naman* ‘EMPH’, *=nga* ‘REIT’, *=pala* ‘CE’. Other particles have high-frequency non-enclitic counterparts (e.g., *(=)kasi* ‘because’, *(=)sana* ‘OPT’). The ability to host enclitic pronouns was checked through searches for combinations of a conjunction with the pronouns that function only as enclitics, including the following: *=ka* ‘2SG.NOM’, *=ko* ‘1SG.ACT’, *=mo* ‘2SG.ACT’, *=(ni)nyo* ‘2PL.ACT’, *=namin* ‘1PL.ACT’, *=natin* ‘1/2PL.ACT’, *=n(i)ya* ‘3SG.ACT’, *=nila* ‘3PL.ACT’. The pronoun *=kita* ‘1SG.ACT+2SG.NOM’ was not included in this list, since it is a homograph of the adjective *kita* ‘seen’, while the pronouns *(=)ako* ‘1SG.NOM’, *(=)kayo* ‘2PL.NOM’, *(=)kami* ‘1PL.NOM’, *(=)tayo* ‘1/2PL.NOM’, *(=)s(i)ya* ‘3SG.NOM’, *(=)sila* ‘3PL.NOM’ frequently occur in non-enclitic positions, and searches for combinations with these forms yield too many results that have to be checked manually, as the corpus is untagged, which would make the procedure too time-consuming.

As mentioned in the Introduction, it is impossible to ascertain just on the basis of corpus data if a certain construction is grammatical

³⁸ This is a wild card indicated with a standalone asterisk (*) in sketchengine.

or represents a deviation from the norms existing in the language community. Due to this, in an attempt to exclude at least some such constructions, the following principle was used as a criterion of whether a combination of a conjunction with the topic marker *ay* or an enclitic is considered as existing in the language. If more than five examples of a combination of a conjunction with the topic marker, particles, or pronouns were found in the corpus, the conjunction was considered to be able to combine with the said elements. If less than five examples of such combinations were found, the possible combinations were further searched for in Google. If such searches found additional instances of such combinations on webpages of Philippine online newspapers, magazines, blogs, or other social media, and the combined number of instances in Google searches and the corpus was at least five, the conjunction was considered as having the property in question. Otherwise, the conjunction was given negative value for the property in question. For instance, the conjunction *dahil* occurs in combination with an enclitic pronoun twice in the corpus and at least one time in Google searches, hence the negative value for this property in *Table 5*.³⁹

When a conjunction was found to pass the requirements for both the topicalization property and the enclitic hosting property, I attempted

³⁹ Apart from legitimate instances of combinations of *dahil* with an enclitic pronoun, like in (i), which comes from a publication by the Philippine tabloid *Abante Tonite*, the corpus also contains a number of sentences with such combinations, which turn out to be missing a certain element, like the negative marker (*hin*)*di* in (ii), have a misspelled element, like in (iii), where *dahil* should be the verb *dalhin*, or represent a nonstandard use of the actor form of the pronoun instead of the nonactor form with the preposition *dahil* (*sa*), like in (iv), where *dahil nya* is supposed to be *dahil sa kanya*:

- (i) *Na-enjoy=nila* *ang Kenya dahil=nila* *na-kita*
 PFV.MOD-enjoy[UV]=3PL.ACT NOM Kenya because=3PL.ACT PFV.MOD-see[UV]
roon *ang marami=ng hayop.*
 DIST.OBL NOM many=LK animal
 ‘They enjoyed Kenya because they saw a lot of animals there.’ [180301944]
- (ii) ⟨...⟩ *at huwag=ka=ng* *ma-ta-takot* *sa akin.*
 and PROH=2SG.NOM=LK AV.STEM-PROSP~fear 1SG.NACT

to find examples where the conjunction hosted an enclitic while used in the topicalization construction, in order to illustrate both properties at once for the sake of saving space.

This criterion does not eliminate the possibility that the combinations found in the corpus and Google searches might be unacceptable in native speakers' grammar. In particular, all five attested examples might come from the same author or even the same text (cf. Footnote 47). However, they provide evidence that such combinations exist in the language at least marginally as part of dialectal or idiolectal variation, or as a result of other language interference. For example, the corpus contains 23 instances of the combination *pero ay* 'but TOP' (with 13 instances coming from the same blog). However, four Tagalog native speakers (all having higher education, aged 20, 29, and 31; one has been exposed to Cebuano, another one to Ilokano) whom I asked to evaluate acceptability of the construction in (37), which comes from a blog, found it unacceptable.

- (37) *May batas=nga=tayo pero ay para=lang sa*
 EXIST law=REIT=1/2PL.NOM but TOP for=only OBL
mayroon=g pera.
 EXIST=LK money

'We have the law, but only for those who have money.' [<http://arvin95.blogspot.com/2010/03/tama-ba.html>]

Dahil=mo=dapat=ako ka-takut-an.

because=2SG.ACT=must=1SG.NOM STEM=fear-UV[INF]

'{...} and don't be afraid of me. Because you don't need to be afraid of me.'
 [170082797]

- (iii) *I-uwi=mo o=kaya dahil=mo sa simbahan,*
 UV-take_home[INF]=2SG.ACT or=SPEC because=2SG.ACT OBL church
i-lagay=mo sa altar.
 UV-put[INF]=2SG.ACT OBL altar

'Take it home, or else bring it to the church and put it on the altar.' [186528837]

- (iv) *Masaya=ako sa ngayon dahil=nya.*
 happy=1SG.NOM OBL now because_of=3SG.ACT
 'I am happy now because of her.' [5046125]

3.1. Group 1

The first group of conjunctions consists of those that have all three features, including only two conjunctions: *kaya* ‘therefore’ and *para* ‘so that’. Examples (38) and (40) show constructions with the conjunctions in the topicalized position followed by enclitic particles, while examples (39) and (41) illustrate their ability to host pronominal enclitics.

- (38) *Kaya=naman=po ay nais=ko=po=ng bahagi ng*
 therefore=EMPH=HON TOP desired=1SG.ACT=HON=LK part GEN
akin=g pagsusulat ang akin=g mga mahal
 1SG.ACT=LK writing NOM 1SG.ACT=LK PL dear
na mambabasa.
 LK reader

‘So, I want my dear readers to be part of my writing.’ [7224305]

- (39) *Just remember na kaya=ka an-dito sa*
 LK therefore=2SG.NOM PRED-PROX.OBL OBL
Saudi ay para mag-work para sa family <...>
 PLN TOP so_that AV.STEM-work[INF] for OBL family

‘Just remember that this is why you are here in Saudi — in order to work for the family <...>’ [145527]

- (40) <...> *para=po ay ma-iwas-an ang casualties,*
 so_that=HON TOP MOD-avoid-UV[NEUT] NOM casualties
para ma-iwas-an=po ang maging biktima
 so_that MOD-avoid-UV[INF]=HON NOM AV.INF.become victim
ng disgrasya.
 GEN accident

‘<...> in order to avoid the casualties, in order to avoid becoming an accident victim.’ [83249578]

- (41) *Sige hintay-in=ko ang code para sa blog=mo*
 alright wait-UV[INF]=1SG.ACT NOM code for OBL blog=2SG.ACT
para=ko ma-i-lagay sa side bar<...>
 so_that=1SG.ACT MOD-UV-put[INF] OBL side_bar

‘Alright, I will wait for the code for your blog so that I can put it to the side bar <...>’ [34655628]

3.2. Group 2

The second group includes only one conjunction, *bago* ‘before’, which can host both particles and pronouns, as shown in (42), but cannot undergo topicalization.

- (42) *Bago=mo=ba ma-kilala si*
 before=2SG.ACT=Q MOD-get_acquainted[UV.INF] PERS.SG.NOM
James Younghusband fan=ka=na=talaga ng football?
 PN PN fan=2SG.NOM=IAM=indeed GEN football
 ‘Were you already a football fan before you got to know James Younghusband?’ [144611007]

3.3. Group 3

Group 3 is the biggest one and consists of conjunctions that can be topicalized and host particles, but not pronouns. The following are examples illustrating the use of the conjunctions with an enclitic particle in the topicalized position: *palibhasa* ‘because’ (see example (43)), *(=)kasi* ‘because’ (see example (44)), *tutal* ‘because’ (see example (45)), *(kung) danggan=kasi* ‘because’ (see example (46)), *(pa)paano=kasi* ‘because’ (see example (47)), *porke(=t)* ‘because’ (see example (48)), *dahil* ‘because’ (see example (49)), *(sa)pagka(=t)* ‘because’ (see example (50)), *dangkasi* ‘because’ (see example (51)), *mangyari* ‘because’ (see example (52)), *samakatuwid* ‘therefore’ (see example (53)).

- (43) *May bf si Janice at may*
 EXIST boyfriend PERS.SG.NOM PN and EXIST
nang-ya~yari=na sa kanila,
 IPFV.STEM-IPFV~happen[AV]=IAM 3PL.NACT

palibhasa=nga ay close=kami kaya
 because=REIT TOP close=1PL.NOM therefore

na-ka~ka-pag-kwentu-han=kami ng tungkol sa sex.
 MOD.IPFV-IPFV~AV-STEM-tell-RECP=1PL.NOM GEN about OBL sex

‘Janice has a boyfriend and something is already happening between them, because we are close, that’s why we can tell each other about sex.’ [177278423]

- (44) *Tu~tulong-an=daw=nya=ako dito sa gawain=g*
 PROSP~help-UV=REP=3SG.ACT=1SG.NOM PROX.OBL OBL chore=LK

bahay kasi=daw ay may sakit=nga ito=ng
 house because=REP TOP EXIST sickness=REIT PROX.NOM=LK

si Jhino.
 PERS.SG.NOM PN

‘He said that he would help me here with housework because Jhino is sick.’ [201937955]

- (45) *Na-isip-an=nya=ng ma-ligo sa ilog*
 MOD.PFV-think-UV=3SG.ACT=LK AV.STEM-bathe[AV] OBL river

tatal=naman ay wala=na=ng tao sa paligid.
 because=EMPH TOP EXIST.NEG=IAM=LK person OBL surroundings

‘She decided to bathe in the river, because there were no people around anyway.’ [112541291]

- (46) *At dangan=din=kasi=y hindi lengguwahe ni*
 and because=also=because=TOP NEG language PERS.SG.ACT

Tina ang Tagalog
 PN NOM Tagalog.

‘And also because Tagalog is not Tina’s language.’ [228574191]

- (47) *Bukod sa hindi=na=niya kaya ang gastos,*
 apart OBL NEG=IAM=3SG.ACT able NOM expenses

napaka-hirap mag-aral sa kanya=ng
 INTENS~difficult AV.STEM-study[INF] OBL 3SG.ACT=LK

<in>upa-han=g silid.
 <PFV>rent-UV=LK room

Paano=naman=kasi ay napaka-ingay.

how=EMPH=because TOP INTENS-loud

‘Apart from that he can’t afford the expenses, it is very difficult to study in the room he rented. Because it is very loud.’ [27312099]

- (48) ***Porket=ba ay s<in>uwerte=ka=na at***
 because=Q TOP <PFV>get_lucky[UV]=2SG.NOM=IAM and
nag-ka-visa=ka=na kahit pulubi ay halos
 PFV.STEM-ACQ-visa[AV]=2SG.NOM=IAM although beggar TOP almost
hindi=muna ma-bigy-an ng barya,<...>⁴⁰
 NEG=at_first MOD-give-UV[INF] GEN change

‘Just because you got lucky and you’ve already got a visa, you won’t even give any change to a beggar, <...>’ [https://issuu.com/ofwakomagazine/docs/ofw_ako_magazine_online_edition_issue_012]

- (49) ***May matindi=ng dahilan=pala kung bakit***
 EXIST intense=LK reason=CE if why
nam-ahinga=ulit ang singer-actress na si
 PFV.STEM-rest=again NOM singer-actress LK PERS.SG.NOM
Selena Gomez sa social media, dahil=pala=y may
 PN PN OBL social media because=CE=TOP EXIST
p<in>ag-da~daan-an ito=ng sakit.
 <IPFV>STEM-IPFV-go_through-UV PROX.NOM=LK sickness

‘It turns out that there is a serious reason why the singer-actress Selena Gomez took a break from social media, because, it turns out, she is going through a sickness.’ [133786840]

- (50) ***Sapagkat=po ay kung ano=po ang mga***
 because=HON TOP if what=HON NOM PL
mag-ka~ka-lapit na streets ay yun=na=po ang
 RECP-PL~SHR-near LK streets TOP DIST.NOM=IAM=HON NOM
d<in>i~diretso=nila
 <IPFV>IPFV~go_straight_to[UV]=3PL.ACT

⁴⁰ No instances of the combinations of *porke ay*, *porke’y*, *porket ay*, *porke’t ay* are found in the corpus, but some are found in Google searches.

‘Because whatever streets are near each other, they go straight to those.’ [166629291]

- (51) *Dangkasi=raw ay iba=na si Jinkee* <...>
 because=REP TOP different=IAM PERS.SG.NOM PN

‘They say, because Jinkee is different now <...>’⁴¹ [<http://globalbalita.com/?s=jinkee+pacquiao>]

- (52) *Mangyari=po=y ni-linlang=ako ng ahas,*
 because=HON=TOP PFV-deceive-UV=1SG.ACT GEN snake
kaya=ako na-tukso=ng k<um>ain <...>
 therefore=1SG.NOM PFV.MOD-tempt[UV]=LK <AV>eat[INF]

‘Because the snake deceived me, that’s why I got tempted to eat it <...>’ [2003222]

- (53) *Samakatuwid=pala ay maaari=ka=ng ma-bautismu-han*
 therefore=CE TOP possible=2SG.NOM=LK MOD-baptise-UV[INF]

sa pamamatigan=lamang ng pag-tanggap ng salita
 OBL help=only GEN NMLZ-receive GEN word
ng Diyos!
 GEN god

‘Therefore, as it turns out, you can get baptized just through accepting the word of God!’ [221398457]

The following pairs of examples illustrate the use of the rest of the conjunctions in this group in the topicalized position and with enclitic particles: *komo* ‘because’ ((54)–(55)), *kesyo* ‘because’ ((56)–(57)), *kung* ‘if’ ((58)–(59)), *pero* ‘but’ ((60)–(61)), *nguni* (=‘i’) ‘but’ ((62)–(63)), *manypa* (‘)t ‘because’ in ((64)–(65)), *sa dahilan*=g ‘because’ ((66)–(67)), and *sa kadahilan*(an)=g ‘because’ ((68)–(69)).

- (54) *Komo kami ay extranjero at sila=naman*
 because 1PL.NOM TOP foreigner and 3PL.NOM=EMPH
komo ay taga-rito, bakit=naman kami
 because TOP from-PROX.OBL why=EMPH 1PL.NOM

⁴¹ Only three instances of *dangkasi* in the topicalized position are found in the corpus, while several instances are found in Google searches.

*p<in>a-baya-an?*⁴²

<PFV>abandon-UV

‘Just because we are foreigners and because they are from here, why are we abandoned?’ [<http://testnews.abs-cbn.com/global-filipino/08/13/15/pinoys-see-justice-filipina-imprisoned-spain>]

- (55) *Komo=ba may langib ang tupa,*
because=Q EXIST scab NOM sheep

la~layo=na ang pastol?
PROSP~draw_away[AV]=IAM NOM shepherd

‘Just because a sheep has a scab, will the shepherd draw away?’
[151349857]

- (56) *Kesyo ay wala=raw=ako=ng lovelife kaya*
because TOP EXIST.NEG=REP=1SG.NOM=LK love_life therefore

*nag-pa~pa-pansin=daw=ako.*⁴³
IPFV.STEM-IPFV~CAUS-pay_attention=REP=1SG.NOM

‘They say, because I don’t have any love life, I seek attention.’
[<https://www.wattpad.com/299456157-the-nerd%27s-future-husband-wattys2017-5/page/2>]

- (57) *Mayroon=din=naman=g na-pi~pikon sa kanya,*
EXIST=also=EMPH=LK IPFV.STEM-IPFV~get_offended[AV] 3SG.NACT

kesyo=raw siya ay hipokrito at inconsistent ang
because=REP 3SG.NOM TOP hypocrite and inconsistent NOM
views=niya bilang tao.
views=3SG.ACT as person

‘There are those who get offended by him, because, they say, he is a hypocrite and his views as a person are inconsistent.’ [8986966]

- (58) *Sa tatlo=ng sala ng Muntinlupa RTC,*
OBL three=LK hall GEN PLN regional_trial_court
ang i-s<in>ampa=ng motion to quash ng kampo
NOM UV-<PFV>file=LK motion_to_quash GEN camp

⁴² Only one example in the corpus and several in Google searches.

⁴³ No examples in the corpus but several in Google searches.

ni De Lima ay bukas=pa=lamang=ito
 PERS.SG.ACT PN TOP tomorrow=CONT=only=PROX.NOM

di-dingg-in kung ay i-ba~basura ang
 PROSP~hear-UV if TOP UV-PROSP-junk NOM

k<in>a~ka-harap=nito=ngkaso <...>
 <IPFV>IPFV~SHR-face[UV]=PROX.GEN=LK case

‘In the three halls of Muntinlupa regional trial court, the motion to quash filed by De Lima’s camp, it will be heard only tomorrow if the case she is facing will be junked <...>’ [216297206]

(59) *May balak=pa=pala=siya=ng t<um>akbo bilang Presidente*
 EXIST plan=CONT=CE=3SG.NOM=LK <AV>run[INF] as president

kung=daw siya ang ‘i-s<in>ugo ng Diyos’?
 if=REP 3SG.NOM NOM UV-<IPFV> send GEN god

‘Does he also have a plant, as it turns out, to run for president if, as he said, he was sent by God?’ [213061009]

(60) *Noon=g dati ang PAL ay*
 DIST.GEN=LK before NOM Philippine_airlines TOP

nag-o~opera ng mga lipad na hindi
 IPFV.STEM-IPFV-operate[AV] GEN PL flight LK NEG

nag-mu~mula sa hub, pati=na=rin mga non-stop
 IPFV.STEM-IPFV~originate OBL hub even=IAM=also PL non_stop

na mga lipad sa Europa at mga Domestiko=ng
 LK PL flight OBL Europe and PL domestic=LK

mga operasyon pero=y hindi=na i-p<in>a-tuloy
 PL operation but=TOP NEG=IAM UV-<PFV>CAUS-continue

pag-dating ng Krisis Pinansiyal sa Asya
 PREC-arrive GEN crisis financial OBL Asia

noon=g 1999.
 DIST.GEN=LK

‘Before the Philippine Airlines used to operate flights not originating from the hub, including also non-stop flights to Europe and domestic operations, but these were discontinued with the start of the Financial Crisis in Asia in 1999.’ [159756688]

- (61) *Doon, s<in>abi=ko na mahalaga ang*
 DIST.OBL <PFV>say[UV]=1SG.ACT LK important NOM
compatibility hindi=lamang sa sex, pero=rin sa
 compatibility NEG=only OBL sex but=also OBL
katalinuhan, sa personalidad sa pag-tingin paano
 intelligence OBL personality OBL NMLZ-look how
ga~gasta-hin ang pera, atbp.
 PROSP~spend-UV NOM money etc.

‘There, I said that compatibility is important not only in sex, but also in intelligence, in personality, in the views on how to spend money, etc.’ [201067532]

- (62) *T<um>ayo=pa=rin=siya sa harap ng pintuan,*
 <AV>stand_up[PFV]=CONT=also=3SG.NOM OBL front GEN door
t<um>ango, na-gustu-han ang mga malulungkot
 <AV>nod[PFV] PFV.MOD-like-UV NOM PL sad
ngunit ay matatamis na mga kataga=ng <in>ukit
 but TOP sweet LK PL word=LK <PFV>carve[UV]
sa tula.
 OBL poem

‘She still stood up in front of the door, nodded, enjoying the sad but sweet words carved in the poem.’ [5937254]

- (63) *Na-i~intindi-han=ko=po na bulag ang*
 IPFV.MOD-IPFV-understand-UV=1SG.ACT=HON LK blind NOM
hustisya, ngunit=po tila nag-iba=na ang
 justice but=HON as_if PFV.STEM-change[AV]=IAM NOM
ibig sabihin.
 meaning

‘I understand that justice is blind, but it’s like its meaning has changed.’ [52308424]

- (64) <...> *may kakaiba=ng ngiti=ng nam-u~mutawi*
 EXIST unusual=LK smile=LK IPFV.STEM-IPFV~be_uttered
sa mga labi ng bawat tao=ng
 OBL PL lip GEN each person=LK

ma-ka~ka-salubong=mo sa kalye, sa eskuwela o
 MOD-PROSP~SHR-meet[UV]=2SG.ACT OBL street OBL school or

di=kaya='y sa opisina. Manyapa='y sadya=ng
 NEG=SPEC=TOP OBL office because=TOP special=LK

kakaiba=ng sigla ang na-i-du~dulot ng
 unusual=LK liveliness NOM IPFV.MOD-UV-IPFV~bring GEN

simoy ng Pasko sa atin=g mga
 breeze GEN Christmas OBL 1/2PL.ACT=LK PL

Pilipino <...>⁴⁴
 Filipino

'<...> there are unusual smiles on the lips of every person that you meet in the street, at school or at the office. Because the breeze of Christmas brings us Filipinos some especially unusual liveliness <...>' [https://issuu.com/angdiaryonatin/docs/adn_507/4]

(65) *So manyapa't=ba wala=ako,*
 so because=Q EXIST=1SG.NOM

ha~hanap=ka ng iba?⁴⁵
 PROSP~search[AV]=2SG.NOM GEN different

'So, just because I'm not there, you will look for someone else?'
 [94441082]

(66) *At h<um>i~hingi=ito ng kapatawaran sa hindi*
 and <AV>IPFV~request=PROX.NOM GEN forgiveness OBL NEG

pag-sabi ng tunay=nito=ng karamdaan sa dahilan
 NMLZ-say GEN real=PROX.GEN=LK sickness OBL reason

ay ayaw=nito na mag-alala=sila=ng
 TOP disliked=PROX.GEN LK AV.STEM-WORRY[INF]=3PL.NOM=LK

⁴⁴ No occurrences of *manyapa'y*, *manyapa't ay*, *manyapat ay*, *manyapa ay*, or *manyapa at ay* are found in the corpus, while *manyapa't* occurs in Google searches five times. These numbers indicate that there is a possibility that, in fact, in all these instances *manyapa'y* is actually the misspelled *manyapa't*.

⁴⁵ One occurrence of *manyapa't ba* and one occurrence of *manyapat ba* in the corpus, and another five occurrences of *manyapat ba* in Google searches.

mag-ina.⁴⁶

NMLZ.DU-mother

‘And he was asking forgiveness for not telling telling about his real sickness because he did not want his wife and child to worry.’
[96202070]

- (67) <...> *sa dahilan=po=ng masama ang lagay*
 OBL reason=HON=LK bad NOM condition
ni Kapitan Tiago at ako=’y
 PERS.SG.ACT captain PN and 1SG.NOM=TOP
kailangan=g <...>
 needed=LK

‘<...> because Captain Tiago’s condition was bad and I was needed <...>’ [12303008]

- (68) <...> *at ako ay hindi=na b<um>aba sa*
 and 1SG.NOM TOP NEG=IAM <AV>descend[PFV] OBL
kadahilanan ay malakas ang ulan <...>
 reason TOP strong NOM rain

‘<...> and I didn’t get out because it was raining heavily <...>’⁴⁷
 [https://lawphil.net/judjuris/juri1999/apr1999/gr_126303_1999.html]

- (69) <...> *at na-rule out=po ang RA 7610*
 and PFV.MOD-rule_out[UV]=HON NOM Republic_Act_7610
sa kadahilanan=daw=po=ng di=naman=daw=po
 OBL reason=REP=HON=LK NEG=EMPH=REP=HON

nag-karoon ng psychological effect sa bata.
 PFV.STEM-acquire[AV] GEN psychological effect OBL child

‘<...> and Republic Act 7610 was ruled out because, as they said, there were no psychological effect on the child.’ [2476341]

⁴⁶ There are only five occurrences of *sa dahilan ay* in the corpus, four of which come from the same blog.

⁴⁷ No occurrences of this combination in the corpus, but a number of them is found in Google searches.

Note that in (66) and (68) *ay* replaces the linker =g in *sa dahilan=g* and *sa kadahilan(an)=g* ‘because’.

3.4. Group 4

Conjunctions in this group cannot be topicalized or host pronouns, but can host particles. The following are examples for the causal conjunctions comprising this group: (*ya*)*yaman=g* (70), *dahil sa* (71), *dahilan sa* (72), *gawa ng* (73), and *bilang* (74).

- (70) *Yaman=din=lamang na etiko ang paksa,*
 because=also=only LK ethics NOM topic
na-isip=ko bigla ang mga
 PFV.MOD-think[UV]=1SG.ACT suddenly NOM PL
na-ga~ganap na mga krimen <...>
 IPFV.STEM-IPFV~happen[AV] LK PL crime

‘Just because the topic was ethics, I suddenly thought of the happening crimes <...>’ [49514019]

- (71) *Na-move ang pasukan=namin sa June 15*
 PFV.STEM-move[AV] NOM start_of_classes=1PL.ACT OBL June
imbis na June 8 siguro dahil=din sa konti ang
 instead LK June perhaps because=also OBL few NOM
nag-e~enrol=nga <...>
 IPFV.STEM-IPFV~enrol=REIT

‘The start of our classes was moved to June 15 instead of June 8, probably also because few people were enrolling <...>’ [33295849]

- (72) *Mayroon=po=ng masama=ng nang-ya-yari*
 EXIST=HON=LK bad=LK IPFV.STEM-IPFV~happen[AV]
sa kanya dyan subalit wala=sya=ng ma-gawa
 3SG.ACT MED.OBL however EXIST.NEG=3SG.NOM=LK MOD-do[UV.INF]
dahilan=po sa bago=pa=lang=po=sya dyan <...>
 because=HON OBL new=CONT=only=HON=3SG.NOM MED.OBL

‘Something bad is happening to him there, however he can’t do anything because he is just new there <...>’ [26651700]

- (73) *Ah oo=nga masarap ang siopao dito,*
 ah yes=REIT delicious NOM siopao PROX.OBL
gawa=daw ng pusa ang palaman ⟨...⟩
 because=REP GEN cat NOM filling
 ‘Ah yes indeed, the siopao here is delicious, because, they say, it is made with cats ⟨...⟩’ [20972282]
- (74) *Hindi=na malayo=ng gaw-in=naman ni Thea*
 NEG=IAM far=LK do-UV[INF]=EMPH PERS.SG.ACT PN
iyon para sa kahit kanino=niya=ng mga
 DIST.NOM for OBL although who.OBL=3SG.ACT=LK PL
kapatid, bilang=din=naman panganay=siya ⟨...⟩
 sibling because=also=EMPH first_child=3SG.NOM
 ‘It was not unlikely then that Thea would do it for any of her siblings, because, of course, she was the eldest one ⟨...⟩’ [118476413]

3.5. Group 5

The last group of conjunctions includes *anupa’t* ‘therefore’, *at* ‘and’, and *nang* ‘when’, which cannot be topicalized and cannot host any enclitics. Examples (75) with *anupat* and (76) with *at* and *nang* show that enclitics have to be hosted elsewhere.

- (75) *Napaka-liit na bagay=ba para sa inyo na pagur-in*
 INTENS~small LK thing=Q so_that 2PL.NACT LK tire-UV[INF]
ang mga tao, anupat pa~pagur-in=din=ninyo
 NOM PL person therefore PROSP~tire-UV[INF]=also=2PL.ACT
ang akin=g Diyos?
 NOM 1SG.ACT=LK god
 ‘Is it a small thing for you to weary men, so you will weary my God also?’ [96981072]
- (76) *At nang na-buwag=na=siya sa grupo*
 and when PFV.MOD-unbind[UV]=IAM=3SG.NOM OBL group

ay d<um>iretso=nga=ito sa
 TOP <AV>go_straight_to[PFV]=REIT=PROX.NOM OBL

k<in>a-ro~roon-an=ko.
 <IPFV>STEM-IPFV~DIST.OBL-UV=1SG.ACT

‘And when he left the group, he went straight to where I was.’
 [3523223]

4. Discussion

4.1. Discrepancies with the previous studies

The data presented above contradict some of the previous descriptions of the behavior of Tagalog conjunctions. The generalization that Tagalog conjunctions cannot host enclitic pronouns, made in [Billings, Kaufman 2004: 19–20; Lee, Billings 2005: 243, 248], does not hold up against the corpus data, as well as the observation in [Kaufman 2010a: 184] that only *kaya* ‘therefore’ can host enclitics. At least three conjunctions among those covered in this paper (*kaya* ‘therefore’, *para* ‘so that’, and *bago* ‘before’) can host both enclitic particles and pronouns, confirming the observations made in [McFarland 2001: 11–13; Schachter, Otones 1972: 191]. On the other hand, the observations in [Schachter, Otones 1972] regarding the conjunctions present in this paper and in their description (see *Table 4*) contradict the corpus data in the following: (i) *kung* ‘if’, *pero* ‘but’, and the causal conjunctions *dahil sa*, *dahilan sa*, *gawa ng*, (*sa*)*pagka=’i* can host enclitic particles, in spite of Schachter, Otones’ claim that they cannot; (ii) *porke(=t)* ‘because’ can host only enclitic particles, in spite of their claim that it can host both types of enclitics.

4.2. Topicalization correlation with information structure

The possibility to place conjunctions into the topicalized position raises the question if such a structure actually reflects any changes in the information structure of the utterance. It is difficult to study this matter

using only written sources, since topicalization is not always marked with the topic boundary marker *ay*, but also with a pause in speech, which might or might not be reflected with a comma in writing, as mentioned in *Section 2.2*.⁴⁸ Hence, this question requires a further experimental study involving a number of native speakers.

As a first approximation to the issue, it is safe to claim that at least in some instances conjunction topicalization does reflect information structure changes. In such instances, topicalization of a conjunction indicates that the information focus is contained in the rest of the clause, following *ay*. Nontopicalization of a conjunction, on the other hand, indicates that the rest of the clause constitutes the information topic, while the conjunction itself is part of the focus.

The immediate left context of example (38) above (with the topicalized conjunction *kaya* ‘therefore’) given in (77) shows that the construction in (38) answers the question ‘What is the consequence of this situation regarding my attitude to my writing and my readers?’.

- (77) *Hindi=naman=po=ako nag-kulang sa pagsasabi*
 NEG=EMPH=HON=1SG.NOM PFV.STEM-lack[AV] OBL saying
na ang mga akda=ko ay h<in>a~handog=ko
 LK NOM PL work=1SG.ACT TOP <IPFV>IPFV~offer[UV]=1SG.ACT
sa akin=g mga mabubuti=ng mambabasa.
 OBL 1SG.ACT=LK PL good=LK reader

‘I haven’t got tired of saying that my works are my gift to my good readers.’ [7224284]

In example (39) with the nontopicalized *kaya* ‘therefore’, on the other hand, the context (‘Just remember that...’) indicates that the clause

⁴⁸ The topic marker *ay* is generally considered to be characteristic of formal style (cf. [Schachter, Otones 1972: 485]). Hence, it could be argued that nonoccurrence of the informal style conjunctions among those examined here with *ay* in the corpus stems from their stylistic incompatibility, while in fact they might be able to undergo topicalization marked with intonation. Although this is certainly a possibility, the corpus data does not support this objection, as, for instance, it contains 50 occurrences of the informal conjunction *kasi* ‘because’ and *ay* at the beginning of the sentence (*Kasi ay* and *Kasi’y*).

following *kaya* constitutes presupposition, while the clause following *ay* is the information focus, answering the question ‘Why are you here in Saudi?’.

In case of the topicalized conjunction *palibhasa* ‘because’ in (43), the information in the clause following the conjunction *kaya* ‘therefore’ (‘... that’s why we can tell each other about sex’) is presupposed from the information given in the first clause in the sentence (‘Janice has a boyfriend and something is already happening between them, ⟨...⟩’). The clause after the topicalized *palibhasa* ‘because’ introduces new information in the situation, answering the question “Why can we tell each other about sex?”.

Example (78) shows a construction with a clause introduced with a non-topicalized *palibhasa* ‘because’, containing presupposed information, while the consequence clause answers the question ‘What happens because they are a child?’.

- (78) *Nag-la-laro ang nakababata=mo=ng kapatid.*
 IPFV.STEM-IPFV~play NOM younger=2SG.ACT=LK sibling
Naki-usap=ka=ng tulung-an=niya ngunit,
 PFV.STEM-ask=2SG.NOM=LK help-UV[INF]=3SG.ACT but
palibhasa *bata=pa ay hindi na-ka-u-unawa*
 because child=CONT TOP NEG IPFV.MOD-AV-IPFV~understand
sa iyo=ng pangangailangan.
 OBL 2SG.ACT=LK need

‘Your younger sibling is playing. You asked them to help you but, since they are a child, they do not comprehend what you need.’
 [91323945]

On the other hand, there are examples in the corpus where topicalization of *palibhasa* ‘because’ does not indicate that the following clause is the information focus. This is the case in (79), where the clause with *palibhasa* ‘because’ contains the information presupposed from the left context, while the consequence clause answers the question ‘What happened because she was looking at the young man?’.

- (79) *Ng<in>iti-an=lamang=sya ng dalaga. Palibhasa='y*
 <PFV>smile-UV=only=3SG.NOM GEN girl because=TOP

naka-tingin sa binata ay hindi na-malay-an
 STAT-look OBL youth TOP NEG PFV.MOD-realize-UV

ni Shiela na papalapit=pala ang ama
 PERS.SG.ACT PN LK approaching=CE NOM father

ni Jeffrey sa k<in>a-ta~tayu-an=nya.
 PERS.SG.ACT PN OBL <IPFV>STEM-IPFV-stand-UV=3SG.ACT

‘The girl just smiled at him. Because she was looking at the young man, Shiela did not realize that Jeffrey’s father was coming closer to where she was standing.’ [92775013]

Another possibility that needs to be studied in further research is that some of the conjunctions might not occur with clauses containing the information topic at all. In that case, topicalization of the conjunction might be optional, without marking any semantic distinction.

4.3. Similarities with other constituents with differing properties regarding topicalization and clitic placement

Some parallels in the topicalization and clitic hosting behavior of the five groups of conjunctions can be found in constituents of other types.

For instance, similarly to the first group of conjunctions, any phrase that can be fronted can be topicalized, like in (80), and host enclitic particles, like in ((80)–(81)), as well as host enclitic pronouns when fronted (81).

(80) *Dito sa community=naman ay wala=ng aso.*
 PROX.OBL OBL community=EMPH TOP EXIST.NEG=LK dog

‘Here in the community, there are no dogs.’ [162555425]

(81) *Dito sa Canada=naman=daw=niya=talaga*
 PROX.OBL OBL Canada=EMPH=REP=3SG.ACT=indeed

gusto=ng gaw-in ang kanya=ng
 wanted=LK do-UV[INF] NOM 3SG.ACT=LK

medical residency.

‘He says, he really wants to do his medical residency here in Canada. (not anywhere else)’ [134380779]

Bago ‘before’ of group 2 behaves similarly to some *wh*-words that do not occur as the predicate of an equational construction,⁴⁹ like *saan* ‘where’, *paano* ‘how’, and *bakit* ‘why’, in it does not occur topicalized, but can host enclitics of both types, like in (82).

- (82) *Saan=daw=siya pu~punta?*
 where=REP=3SG.NOM PROSP~go[AV]
 ‘Where did he say he was going?’ [12248400]

In constructions with a topicalized substantivized verb or pseudo-verb and a clausal predicate the topicalized constituent cannot occur in any other position [Schachter, Otanes 1972: 487]. Such topicalized constituents behave like any other topicalized units in that they can host enclitic particles, like in ((83)–(84)), and, thus, are similar to group 3 conjunctions:

- (83) *Ang na-rinig=ko=nga,*
 NOM PFV.MOD-hear[UV]=1SG.ACT=REIT
di gaano=ng happy ang mga taga-Star cinema...
 NEG how_much=LK happy NOM PL from-Star_cinema
 ‘What I did hear is that the people from the Star cinema are not so happy...’ [109938607]
- (84) *Ang dapat=po ay mag-tulong-an.*
 NOM must=HON TOP AV.STEM-help-RECP[INF]
 ‘What should be done is helping each other.’ [128137664]

Some constituents of this type, like *dapat* ‘must’, can drop the nominative marker *ang*.

- (85) *Dapat=po ay nag-ka~ka-isa=tayo.*
 must=HON TOP IPFV.STEM-IPFV~SHR-one=1/2PL.NOM
 ‘We should be uniting.’ [44267111]

At least one group 3 conjunction, (=) *kasi* ‘because’, can also be used as an enclitic, which makes it similar to some other lexical units that can

⁴⁹ Equational constructions (i.e., with a substantive predicate) with a substantivized verb as the subject are also referred to as *cleft* (cf. [Kroeger 1993: 62; Nagaya 2007: 348]) or *pseudocleft constructions* [Nagaya, Hwang 2018].

behave both as enclitic and nonenclitic, like the optative markers (=)*sana* and (=)*nawa*, (=)*siguro* ‘perhaps’, (=)*talaga* ‘indeed’, and others, like in ((86)–(87)).

- (86) ***Sana=po*** *ay na-gustu-han=ninyo=ito.*
 OPT=HON TOP PFV.MOD-like-UV=2PL.ACT=PROX.NOM
 ‘I hope you liked it.’ [290116]

- (87) ***P<in>alit-an=na=lang=sana=sila*** *ng iba=ng*
 <PFV>replace-UV=IAM=only=OPT=3PL.NOM GEN different=LK
players ⟨...⟩
 ‘I wish they were replaced with other players instead ⟨...⟩’ [53339]

Group 4 conjunctions are similar to heads of fronted constituents. The head of any fronted constituent can host only enclitic particles, but not pronouns. Also, the head of any constituent that can be fronted cannot undergo topicalization separately from the rest of the phrase. An example of this is given in (88), where the enclitic particle =*na* ‘IAM’ is hosted by the head of the fronted phrase *bukas ng umaga* ‘tomorrow morning’, while the enclitic pronoun =*kayo* ‘2SG.NOM’ can be hosted only by the whole phrase.

- (88) ***Bukas=na*** *ng umaga=kayo mag-pa-tuloy*
 tomorrow=IAM GEN morning=2SG.NOM AV.STEM-CAUS-continue[INF]
ng paglalakbay.
 GEN travel
 ‘Continue your travel tomorrow morning (rather than now).’
 [2012752]

Another example is (89), where the head of the fronted phrase *sa flat=nya* ‘to his flat’, introducing the goal of directed motion, hosts the particle =*daw* ‘REP’, while the pronoun =*kami* ‘1PL.NOM’ has to be hosted at the right edge of the phrase.

- (89) ⟨...⟩ *dahil sa flat=daw=nya=kami pu-punta.*
 because OBL flat=REP=3SG.ACT=1PL.NOM PROSP-go[AV]
 ‘⟨...⟩ because he said we were going to his flat.’ [217143199]

Example (90) includes a fronted prepositional phrase (*para sa experiment...* ‘for the experiment...’), with the particle =*daw* ‘REP’ hosted by the head of the phrase.⁵⁰

(90) ⟨...⟩ *para=daw sa experiment ni*
 for=REP OBL experiment PERS.SG.ACT

Carla sa school=nila.
 PN OBL school=3PL.ACT

‘⟨...⟩ as she said, for Carla’s experiment at their school.’ [2492340]

Example (91) illustrates the only available position for the pronoun =*ka* — following the whole prepositional phrase *para sa iba* ‘for others’.⁵¹

(91) ⟨...⟩ *kahit para sa iba=ka=na k<um>i~kintab.*
 although for OBL different=2SG.NOM-IAM <AV>IPFV~shine

‘⟨...⟩ although you shine for others now.’ [197835777]

Example (92) illustrates the ability of prepositions to host particles (= *daw* ‘REP’) when topicalized, while the pronoun (=) *siya* ‘3SG.NOM’, as expected, has to be hosted by the sentence predicate.

(92) *Dahil=daw dito ay na-ka-kuha=siya ng*
 because=REP PROX.OBL TOP IPFV.MOD-AV-take=3SG.NOM GEN
mababa=ng grades sa school.
 low=LK grades OBL school

‘They said, because of this she got low grades at school.’ [175460215]

In Schachter, Otones’ description there seems to be only two lexical units that resemble group 5 conjunctions in their inability to be topicalized or host enclitics, despite their position at the beginning of the clause:

⁵⁰ The rest of the utterance is ellipted.

⁵¹ I have not been able to find any examples in the corpus where both a particle and a pronoun would be present but hosted by different constituents: the particle by the preposition itself, and the pronoun by the prepositional phrase. Although it does not mean that such constructions do not exist, it seems that the preferred way is to place both types of enclitics together after the fronted constituent, like in (91).

the sentence-initial “rhetorical linker” *e*, marking “a sequential or logical connection”, like in (93), [Schachter, Otanes 1972: 459] and the “adverb” *di* ‘in that case’, which frequently occurs in combination with *e*, like in (94), [Schachter, Otanes 1972: 461].

- (93) *E, isip=ko='y na-sa iyo ang mga ticket.*
 LOG thought=1SG.ACT=TOP PRED-2SG.NACT NOM PL
 ‘Why, I thought you had the tickets.’ [Schachter, Otanes 1972: 459]
- (94) *E, di ga~gaw-in=ko=agad.*
 LOG in_that_case PROSP~do-UV=1SG.ACT=right_away
 ‘Well then, I’ll do it right away.’ [Schachter, Otanes 1972: 461]

These parallels pave the way to speculation that Tagalog conjunctions have been grammaticalized from lexical units of different classes. This seems to be true at least for group 4 conjunctions, as the causal conjunctions *dahil sa*, *dahilan sa*, *gawa ng*, and *bilang* primarily function as prepositions, like in (92), while *(ya)yaman=g* ‘because’ retains the linker, similar to some “initial adverbs”, like *baka sakali* ‘just in case’ [Schachter, Otanes 1972: 460]. Also, cf. the early-20th century use of *dahil* ‘because’ as the noun ‘reason’ in example (i) in Footnote 56.

4.4. *At* ‘and’ with conjunctions

A number of the Tagalog conjunctions discussed above end with the enclitic form =*t* of the conjunction *at* ‘and’⁵²: *nguni(=t)* ‘but’, *(sa)pag-ka(=t)* ‘because’, *porke(=t)* ‘because’, *anupa=t* ‘therefore’, *manyapa=t* ‘because’. Some of these conjunctions can also be used with the full form

⁵² Although *at* ‘and’ somewhat resembles additive markers in that it can be part of conjunctive phrases, it cannot be classified as such, as it lacks the core function of additive markers, i.e. pointing to the existence of an alternative to the associate of the additive [Forker 2016]. In other words, constructions like *Johnathan also came* cannot be used with *at* ‘and’ in Tagalog.

of *at* in the modern language, like *sapagka at* in (95), *porke at* in (96), and *anupa at* in (97).

- (95) *Ga-nyan=talaga ang buhay sa daigdig*
 like-MED.GEN=indeed NOM life OBL world

sapagka at tayo ay tao=lamang.
 because and 1/2PL.NOM TOP person=only

‘Life in this world is indeed like that, because we are only human.’
 [https://www.youtube.com/watch?v=WbnLWuNX408]

- (96) *Kaya huwag=mo=ng isip-in na*
 therefore PROH=2SG.ACT=LK think-UV[INF] LK

porke at ako ay isa=ng huwes ay
 because and 1SG.NOM TOP one=LK judge TOP

sa kurakot nang-galing ang lahat
 OBL corrupt PFV.STEM-come_from[AV] NOM all

ng na-ki-kita=mo dito.
 GEN IPFV.MOD-IPFV~see[UV]=2SG.ACT PROX.OBL

‘That’s why you shouldn’t think that just because I’m a judge, everything you see here came from corrupt officials.’ [165187968]

- (97) *Isa sa mga naging crush=ko ang*
 one OBL PL PFV.become[AV] crush=1SG.ACT NOM

nag-pa-rinig=pa sa akin na bago=sana=ako
 PFL.STEM-CAUS-hear[AV]=CONT 1SG.NACT LK before=OPT=1SG.NOM

man-ligaw sa kanya, h<um>anap=muna=ako
 AV.STEM-court[NEUT] 3SG.NACT <AV>search=at_first=1SG.NOM

ng elevator shoes. Anupa at mula noon,
 GEN elevator_shoes therefore and since DIST.GEN

<um>i~iwas=na=ako=ng maki-sabay
 <AV>IPFV~avoid=IAM=1SG.NOM=LK AV.COM-do_simultaneously[INF]

sa mga kaklase=ng matatangkad.
 OBL PL classmate=LK tall

‘One of those who I had a crush on also mentioned to me that I should look for some elevator shoes first. So, since then I avoid being with those of my classmates who are tall.’ [https://www.

philstar.com/test-microsite-clone/2013/08/15/1090611/kulang-sa-sukat]

Example (95) is one of very few instances of *sapagka at* not in an archaic text. *Nguni at*, on the other hand, seems to occur only in archaic texts, like in example (98), which comes from José Rizal's⁵³ late 19th century translation of Hans Christian's Andersen's "The Ugly Duckling".

- (98) *Nguni at totoo=ng mainam ang*
 but and true=LK good NOM
*l<um>angoi sa tubig <...>*⁵⁴
 <AV>swim[INF] OBL water

'But swimming in water is really good <...>' [219403853]

These conjunctions seem to have undergone the same reinterpretation process as in the case of the interrogative *bakit* 'why', which is historically derived from the combination *bakin at* 'why and', where *at* is used as a complementizer, as shown in (99) [Kaufman 2010b: 189].⁵⁵

- (99) <...> *baquin at i-pa~patay reina=ng*
 why CONJ CV-INCM~kill queen=LNK
uala=ng casalanan.
 N.EXT=LNK sin

'<...> why kill the guiltless queen?' [Kaufman 2010b: 189]

The use of *at* 'and' as a complementizer also happens in the modern language in so-called "explanatory causes", with the main predicate

⁵³ A Filipino writer and national hero.

⁵⁴ In examples (98), (99), (101), and (i) in Footnote 56, the original spelling following the orthography used during the late Spanish colonial period is preserved.

⁵⁵ The same also applies to a number of other conjunctions, including *bagama()* 'although' (also *bagaman*), *kahi()* 'although', *datapwa()* 'but', *subali()* 'however', *danga()* 'were it not that' (also *dangan*), and *hangga()* 'until' (also *hanggang*), and other lexemes, like *diyata()* 'is it true that', *bawa()* 'each', *kailanma()* 'when-ever' (also *kailanman*), *basta()* 'no matter what', *hindi ba()* 'isn't it that', *lalo pa()* 'especially'.

(e.g. *mabuti* ‘good’, *salamat* ‘thank you’, *nagalalak* ‘to be glad’) expressing the speaker’s attitude towards the situation in the complement clause, like in (100) [Schachter, Otanes 1972: 545].

- (100) *Mabuti at pu~punta=ka.*
 good and PROSP~go[AV]=2SG.NOM

‘It’s good (that) you’re going.’ [Schachter, Otanes 1972: 545]

In addition to the conjunctions listed above, a number of other conjunctions under study here are also found in combination with *at* / = *t*, including *kaya* ‘therefore’, and the causal conjunctions *dahil*, *kesyo*, *kasi*, *palibhasa*, (*kung*) *dangan=kasi*, (*pa*)*paano=kasi*, *komo*, *tatal*, (*ya*)*yaman=g*. It is, however, impossible to speak of any properties of such constructions in any detail here, without conducting a study involving native speakers, since there is a strong possibility of typos in written sources in this case, as ‘t’ is next to ‘y’ on the QWERTY layout of computer keyboards. Yet, at least in some instances, the mentioned conjunctions do occur with the conjunction *at*. For instance, example (101) comes from José Rizal’s translation of “The Ugly Duckling”, which contains another four instances of the combination *dahil at* in the same use.⁵⁶

- (101) *Ya=’i nang-ya~yari dahil at*
 MED.NOM=TOP IPFV.STEM-IPFV~happen because and
ako=’i totoo=ng pangit!
 1SG.NOM=TOP true=LK ugly

‘That is happening because I am truly ugly!’ [219402875]

It seems that *at* ‘and’ in such constructions performs the same function as *ay* ‘TOP’, i.e. marks the topicalized position of the conjunction. There

⁵⁶ Pascual H. Poblete’s 1906 translation of José Rizal’s novel “Noli Me Tangere”, written in Spanish, into Tagalog contains a number of constructions where *dahil* functions as a noun, meaning ‘reason’, followed by *at* used as a complementizer, similar to the use of *baquin at* in (99).

- (i) *¿Anó=ng dahil at h<um>into=ca?*
 what=LK reason and <AV>stop[PFV]=2SG.NOM
 ‘What is the reason that you stopped?’ [224996901]

are no instances of *at* and *ay* following each other in one construction. It remains to be seen in future research if the presence of *at*, like in (101), influences the information status of the following clause.

5. Conclusion

This research focused on three understudied properties — the ability to undergo topicalization, to host enclitic particles and enclitic pronouns — in a subset of Tagalog conjunctions (20 causal conjunctions and 10 others). The data, taken from the Tagalog corpus at sketchengine.eu and supplemented by Google searches, reveal that the above-mentioned conjunctions fall into five groups based on the value combinations for the three properties in question: group 1 has positive values for all three properties; group 2 conjunctions can host both types of enclitics, but does not undergo topicalization; group 3 conjunctions can undergo topicalization and host only particles; group 4 conjunctions can only host particles, but cannot undergo topicalization; and group 5 has negative values for all three properties. Most conjunctions examined in this paper belong to group 3.

Contrary to the observations made in some published descriptions of Tagalog enclitics, the corpus data show that most of the examined conjunctions can host enclitic particles, and at least three can also host pronouns. Schachter, Otones' (1972) relatively detailed discussion of compatibility of conjunctions with the two types of enclitics contradicts the corpus data regarding at least seven conjunctions.

The attested value combinations show that the ability to be topicalized and the ability to host pronouns both entail the ability to host particles, which stems from the general properties of topicalized and fronted constituents.

A number of limitations related to the use of corpus data in this study are mentioned above. These include the impossibility to filter out ungrammatical uses of language constructions characteristic of nonnative speech and idiolectal idiosyncrasies that might not have wider acceptance in the

language community. An attempt to somewhat circumvent this limitation was made by setting a threshold of five occurrences for each construction type that was searched for in the corpus, augmented with Google searches in webpages of Philippine online newspapers, magazines, blogs, or other social media. Also, as mentioned above, according to Schachter, Otones, Tagalog conjunctions also vary in obligatoriness of being enclitic hosts. This matter was not addressed in this study.

The variation in the three properties raises a number of questions that also need to be explored in future research. First, it remains to be seen to what extent placement of a conjunction into the topicalized position triggers changes in information structure. Second, the conjunctions manifest behavior characteristic of some other types of lexical units, which might point to differences in the type of units from which the conjunctions were grammaticalized. Third, some conjunctions can occur with the conjunction *at* ‘and’, which seems to substitute the topic marker *ay* in such combinations. The extent of compatibility of *at* with different conjunctions and the function of such structures remain unclear.

Abbreviations

1 — first person; 2 — second person; 1/2 — first and second persons; 3 — third person; A — case A; ABIL — abilitative; ACQ — acquisitional; ACT — actor form; AV — actor voice; CAUS — causative; CE — counter-expectation; COMP — complementizer; COND — conditional; CONJ — conjunction; CONT — continuative; CV — causal voice; DIST — distal; DU — dual; EMPH — emphatic; EXIST — existential; EXT — existential; F — free pronominal; FUT — future; GEN — genitive; HON — honorific; IAM — iamitive; INCM — incremental; INF — infinitive; INTENS — intensive; IPFV — imperfective; LK — linker; LNK — linker; LOG — logical connection; MED — medial; MOD — modal; N — negative; NACT — nonactor form; NEG — negative; NMLZ — nominalization; NOM — nominative; OBL — oblique; OPT — optative; PERS — personal; PFV — perfective; PL — plural; PLN — place name; PN — personal name; PREC — precedence form; PRED — predicative; PRF — perfective; PROH — prohibitive; PROSP — prospective; PROX — proximal; Q — interrogative marker; RECP — reciprocal; REIT — reiterative; REP — reportative; RV — referent voice; s — singular; SG — singular; SHR — shared notion; SPEC — speculative; STAT — stative; STEM — stem-deriving prefix; TOP — topic marker; UV — undergoer voice.

References

- Billings 2005 — L. Billings. Ordering clitics and and postverbal R-expressions in Tagalog: A unified analysis. A. Carnie, H. Harley, S. A. Dooley (eds.). *Verb first: On the syntax of verb-initial languages*. (Linguistik Aktuell 73). Amsterdam: Benjamins, 2005. P. 303–339.
- Billings, Kaufman 2004 — L. Billings, D. Kaufman. Towards a typology of Austronesian pronominal clisis. P. Law (ed.). *Proceedings of the Eleventh meeting of the Austronesian Formal Linguistics Association*. (ZAS Papers in Linguistics 34). Berlin: Zentrum für Allgemeine Sprachwissenschaft, 2004. P. 15–29.
- Cunningham, Goetz 1963 — M. C. Cunningham, J. E. Goetz. *Pronoun formatives in Amganad Ifugao*. Manila: SIL, Linguistic Society of the Philippines, 1963.
- De Guzman 1978 — V. P. De Guzman. *Syntactic derivation of Tagalog verbs*. (Oceanic Linguistics Special Publications 16). Honolulu: University Press of Hawai'i, 1978.
- Foley 1997 — W. A. Foley. *Anthropological linguistics: An introduction*. Oxford: Basil Blackwell, 1997.
- Forker 2016 — D. Forker. Toward a typology of additive markers. *Lingua*. 2016. Vol. 180. P. 69–100.
- Frolova 1983 — E. G. Frolova. K voprosu o sluzhebnykh sloвах ang i ay v tagalskom yazyke [On function words ang and ay in Tagalog]. V. M. Alpatov (ed.). *Voprosy vostochnogo yazykoznanija* [Topics in the study of oriental languages]. Moscow: Nauka, 1983. P. 240–246.
- Frolova 2017 — E. G. Frolova. Ob upotreblenii sintaksicheskikh pokazately v tagalskom yazyke [On syntactic indicators in the Tagalog language]. *Filologicheskiye nauki. Voprosy teorii i praktiki* [Philology. Theory and Practice]. 2017. Vol. 4. Pt. 2. P. 170–173.
- Gallego 2015 — M. K. S. Gallego. Ang mga nominal marker ng Filipino at Ivatan [Nominal markers of Filipino and Ivatan]. *Daluyan: Journal ng Wikang Filipino*. 2015. Vol. 21. P. 65–95.
- Gallego, Zubiri 2013 — M. K. S. Gallego, L. A. M. Zubiri. Metonymy of NANG. *Social Science Diliman*. 2013. Vol. 9. Pt. 1. P. 39–63.
- Grant 2012 — A. P. Grant. Contact, convergence, and conjunctions: a cross-linguistic study of borrowing correlations among certain kinds of discourse, phasal adverbial, and dependent clause markers. C. Chamoreau, I. Léglise (eds.). *Dynamics of Contact-Induced Language Change*. Berlin: Walter de Gruyter, 2012. P. 311–358.
- Himmelman 1991 — N. P. Himmelman. *The Philippine challenge to universal grammar*. (Arbeitspapier Nr. 15 Neue Folge). Köln: Institute für Sprachwissenschaft, Universität zu Köln, 1991.

- Himmelman 2005 — N. P. Himmelman. The Austronesian languages of Asia and Madagascar: Typological characteristics. A. Adelaar, N. P. Himmelman (eds.). *The Austronesian languages of Asia and Madagascar*. London; New York: Routledge, 2005. P. 110–181.
- Holmer, Billings 2014 — A. Holmer, L. Billings. Clitic pronouns in Seediq. I. W. Arka, N. L. K. M. Indrawati (eds.). *Argument realisations and related constructions in Austronesian languages. Papers from 12-ICAL. Vol. 2*. Canberra: Asia-Pacific Linguistics, 2014. P. 111–139.
- Jakubiček et al. 2013 — M. Jakubiček, A. Kilgarriff, V. Kovář, P. Rychlý, V. Suchomel. The TenTen corpus family. A. Hardie, R. Love (eds.). *Proceedings of the 7th International Corpus Linguistics Conference (CL2013)*. Lancaster: UCREL, 2013. P. 125–127.
- Kaufman 2006 — D. Kaufman. Rigidity versus relativity in adverbial syntax: Evidence from Tagalog. H. Gärtner, P. Law, J. Sabel (eds.). *Clause structure and adjuncts in Austronesian languages*. Berlin: Mouton de Gruyter, 2006. P. 151–194.
- Kaufman 2008 — D. Kaufman. South Sulawesi pronominal clitics: Form, function and position. *Studies in Philippine Languages and Cultures*. 2008. Vol. 17. P. 13–65.
- Kaufman 2010a — D. Kaufman. The grammar of clitics in Maranao. L. Billings, N. Goudswaard (eds.). *Piakandatu ami Dr. Howard P. McKaughan*. Manila: Linguistic Society of the Philippines and SIL, 2010. P. 179–204.
- Kaufman 2010b — D. Kaufman. *The Morphosyntax of Tagalog clitics: A typologically driven approach*. PhD thesis. Ithaca: Cornell University, 2010.
- Kilgarriff et al. 2004 — A. Kilgarriff, P. Rychlý, P. Smrz, D. Tugwell. The Sketch Engine. G. Williams, S. Vessier (eds.). *Proceedings of the 11th EURALEX International Congress*. Lorient: Université de Bretagne Sud, 2004. P. 105–116.
- Kroeger 1993 — P. R. Kroeger. *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications, 1993.
- Kroeger 1998 — P. R. Kroeger. *Nouns and verbs in Tagalog: A reply to Foley*. Paper presented at the 3rd Lexical-Functional Grammar Conference, Asia SIL School, Brisbane, 30 June — 3 July 1998.
- Lee 2008 — C. Lee. Clitic Pronouns in Masbatenyo. *Studies in Philippine Languages and Cultures*. 2008. Vol. 17. P. 121–136.
- Lee, Billings 2005 — C. Lee, L. Billings. Wackernagel and verb-adjacent clisis in Central Philippines. J. Heinz, D. Ntelitheos (eds.). *Proceedings of the 12th meeting of the Austronesian Formal Linguistics Association*. (UCLA Working Papers in Linguistics 12). Los Angeles: University of California, 2005. P. 241–254.
- Lee, Billings 2008 — C. Lee, L. Billings. Clitic-pronoun clusters in Central Philippine. W. Khanittanan, P. Sidwell (eds.). *Papers from the 14th Annual Meeting of the Southeast Asian Linguistics Society 2004. SEALS XIV. Vol. 1. Pacific*

- Linguistics E-5*. Canberra: Pacific Linguistics; The Australian National University, 2008. P. 193–203.
- Malicsi 2012 — J. C. Malicsi. Pang-ukol sa Filipino. *Daluyan*. 2012. Vol. 18. Pt. 1–2. P. 39–82.
- Malicsi 2013 — J. C. Malicsi. *Gramar ng Filipino* [Filipino Grammar]. Quezon City: Sentro ng Wikang Filipino, 2013.
- McFarland 2001 — C. D. McFarland. Filipino enclitics. *Philippine Journal of Linguistics*. 2001. Vol. 32. Pt. 1. P. 1–16.
- Nagaya 2007 — N. Nagaya. Information structure and constituent order in Tagalog. *Language and Linguistics*. 2007. Vol. 8. Pt. 1. P. 343–372.
- Nagaya, Hwang 2018 — N. Nagaya, H. K. Hwang. Focus and prosody in Tagalog. S. Riesberg, A. Shiohara, A. Utsumi (eds.). *Perspectives on information structure in Austronesian languages*. Berlin: Language Science Press, 2018. P. 375–388.
- Peng, Billings 2008 — A. Peng, L. Billings. Binukid pronominal clisis. *Studies in Philippine Languages and Cultures*. 2008. Vol. 17. P. 179–212.
- Rachkov 1966 — G. E. Rachkov. Sluzhebnoe slovo ay v tagalskom yazyke [Function word ay in Tagalog]. S. N. Ivanov (ed.). *Issledovaniya po filologii stran Azii i Afriki* [Studies in philology of Asian and African countries]. Leningrad: Leningrad State University Press, 1966. P. 89–94.
- Reid 1971 — L. A. Reid. *Philippine minor languages: Word lists and phonologies*. (Oceanic Linguistics Special Publications 8). Honolulu: University of Hawai'i, 1971.
- Reid 2009 — L. A. Reid. Inclusory constructions and their development in Philippine languages. A. Adelaar, A. Pawley (eds.). *Austronesian historical linguistics and culture history: A Festschrift for Robert Blust*. (Pacific Linguistics 601). Canberra: Pacific Linguistics; The Australian National University, 2009. P. 267–294.
- Reid, Liao 2004 — L. A. Reid, H. Liao. A brief syntactic typology of Philippine languages. *Language and Linguistics*. 2004. Vol. 5. Pt. 2. P. 433–490.
- Richards 2009 — N. Richards. Nouns, verbs, and hidden structure in Tagalog. *Theoretical Linguistics*. 2009. Vol. 35. Pt. 1. P. 139–152.
- Robinson 2008 — L. C. Robinson. *Dupaningan Agta: Grammar, vocabulary, and texts*. PhD thesis. Manoa: University of Hawai'i, 2008.
- Rubino 1997 — C. R. G. Rubino. *A reference grammar of Ilocano*. PhD thesis. Santa Barbara: University of California, 1997.
- Schachter, Otnes 1972 — P. Schachter, F. T. Otnes. *Tagalog reference grammar*. Berkeley; Los Angeles; London: University of California Press, 1972.
- Shkarban 1989 — L. I. Shkarban. Poryadok slov v tagal'skom yazyke [Word order in Tagalog]. I. F. Vardul' (ed.). *Ocherki tipologii poryadka slov* [Sketches on word order typology]. Moscow: Nauka, 1989. P. 75–108.

- Spencer, Luís 2012 — A. Spencer, A. R. Luís. *Clitics: An Introduction*. New York: Cambridge University Press, 2012.
- Suchomel, Pomikálek 2012 — V. Suchomel, J. Pomikálek. Efficient web crawling for large text corpora. A. Kilgarriff, S. Sharoff (eds.). *Proceedings of the 7th Web as Corpus Workshop*. Lyon: ACL, 2012. P. 39–43.
- Vanoverbergh 1931 — M. Vanoverbergh. Iloko Substantives and Adjectives. *Anthropos*. 1931. Vol. 26. Pt. 3/4. P. 469–488.
- Wolff 1973 — J. U. Wolff. Verbal inflection in Proto-Austronesian. A. B. Gonzalez (ed.). *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*. (LSP Special Monograph Iss. No. 4). Quezon City: Linguistic Society of the Philippines, 1973. P. 71–91.