

Copyright

Haowen Jiang

April, 2016

RICE UNIVERSITY

Nominalization and Possession in Formosan Languages

by

Haowen Jiang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

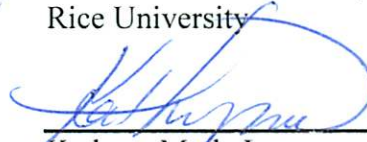
APPROVED, THESIS COMMITTEE



Masayoshi Shibatani, Chair
Professor, Department of Linguistics,
Rice University



Suzanne Kemmer
Associate Professor, Department of Linguistics,
Rice University



Kathryn M. de Luna
Assistant Professor, Department of History,
Georgetown University

HOUSTON, TEXAS

April, 2016

To my grandmother

Abstract

Nominalization and Possession in Formosan Languages

by

Haowen Jiang

This dissertation investigates nominalization and possession in Formosan languages from a functional-typological perspective, where nominalization is a metonymic process of creating denoting expressions. Verbal-based and nominal-based nominalization are each the topic of the two primary parts of this study. Special attention is paid to nominalizations lacking a lexical status, covering constructions traditionally called relativization and possession.

In almost all Formosan languages, the semantic role of an argument nominalization is determined by a small set of affixes on a class of morpho-lexical word forms called Focus-words. Conservative languages demonstrate up to four grammatical categories of Focus-words, marked by two broad sets of affixes (Set I and II) reconstructable in Proto-Austronesian (PAn). Focus-words with Set II affixes predominantly have both predicate and argument functions, which has been explained in terms of a historical reanalysis whereby erstwhile nominalizations were reinterpreted as default verbs, thus marginalizing the use of those with Set I affixes, which are considered verbal throughout the Austronesian history. However, it is argued that Focus-words with both Set I and II affixes can equally constitute argument nominalizations, both subject to the same grammatical restrictions, be it within or across languages. The new analysis suggests PAn was a language employing the gap strategy for argument nominalizations, thus rendering superfluous the question of how or when the nominalization-into-verb reanalysis took place. Consequently, the result casts doubt on any genetic subgrouping based on such a syntactic reanalysis.

The second part explores possessive NPs, and identifies three structural types that are constructionally and paradigmatically defined. The literature shows vigorous interest in the possessor-possessum syntagm, but generally overlooks possessive substantives, or phrases including the possessor but denoting the possessum instead. Possessive substantives in Formosan are important because they expose different syntactic functions of so-called genitive markers across languages even when reflexes of PAn *ni and *nu are involved. Moreover, distributions of these reflexes across the three types have implications on which type PAn might belong to, which in turn helps reveal how modern languages might have diversified from that prototype.

Acknowledgments

I always take pleasure in reading the acknowledgments of scholarly works because they often give away the challenges one has to overcome, some of which require more than just elbow grease and coffee. But things stop being fun when I am the one facing those challenges. First, I underestimated the time needed to finish this work, for both data collection and manuscript writing. Second, it took some missionary-like efforts to gain access to first-hand data, including door-to-door visits in search of competent native speakers as well as follow-up phone calls after meetings that had been agreed on were suddenly canceled. Access to second-hand data was not straightforward either because the publications I was looking for were often buried in a library that was oceans away from me. Third, paradoxically, as more data accumulated, I became less sure of where this dissertation should be heading for. The outcome presented here emerges from several revisions so that the present organization is nothing like the one I proposed in my prospectus. And finally there's also the funding issue, which is critical to almost all research projects, and this present one is no exception.

Despite the challenges that I faced throughout my Ph.D. study, I have had the good fortune of receiving assistance and support from teachers, fellow linguists, language consultants, and funding organizations. First, I would like to express my sincere gratitude to my advisor Masayoshi Shibatani for his continuous guidance and immense knowledge, from which I have benefited enormously, on both the academic and personal levels. I would also like to thank my two other committee members, Suzanne Kemmer and Kathryn M. de Luna, for their constructive comments and hard questions, which pushed me to think of my research from broader perspectives. Heartfelt thanks also go to my other teachers at Rice University, including Michel Achard, Robert Englebretson, Nancy Niedzielski, and Christina M. Willis Oko (in alphabetical order by the last name), for broadening my knowledge in linguistics. I particularly appreciate Laura C. Robinson's continued help and advice even after she left Rice. More generally, I am indebted to Shuanfan Huang, who guided me into the field of Austronesian languages and supervised my M.A. thesis, and Loren Billings, who is more a friend than a research collaborator and who has offered me countless help with research and beyond. And there are direct discussions with or indirect inspirations from numerous Austronesianists, including Edith Aldridge, Robert Blust, Henry Yung-li Chang, Arthur Holmer, Ritsuko Kikusawa, Amy Pei-jung Lee, Lawrence Reid, Malcolm Ross, Laurent Sagart, Li-may Sung, Stacy Fang-Ching Teng, Shigeru Tsuchida, Naomi Tsukida, Joy Jing-lan Wu, Foong Ha Yap, Marie Meili Yeh, and Elizabeth Zeitoun, without whom the present work would not be the way it is now. Second, I am also grateful for the timely favors my fellow linguists have done me, including, but not limited to, Sihwei Chen, Tingchun Chen, Victoria Chen, Yi-ting Chen, Rik de Busser, Huei-ju Huang, Gujing Lin, Naonori Nagaya, Chia-jung Pan, Michael Tanangkingsing, Chunming Wu, most of whom are professors of various ranks by now. I would shamelessly ask them about languages of their expertise or/and request them to introduce me to their language consultants. Additionally, my Ph.D. life was greatly enriched by my colleagues at Rice University, such as Jennifer Hoecker Brindley,

Sarah Seewoester Cain, Akua Asantewaa Campbell, Hussain T. Hijazi, Penelope Howe, Anthony Koth, Carlos Molina-Vital, Ann Marie Olivo-Shaw, Evgeniya Poluektova, and Bethany Townsend, without whom my study would be duller than “colorless green ideas.” Sarah Lee, in particular, deserves special thanks for helping me through thick and thin and always believing in me more than I did myself. Third, I was lucky enough to have met many kind and patient language consultants, whose invaluable insight into their own native languages has furthered my work. Since a complete list of language consultants is provided in the main text, I will not repeat all their names here. Instead, I would love to pay special attribute to three of them. Ms. Atrung from Rikavung in Taitung is a devoted Catholic always ready to help out as much as she can. At the end of elicitation sessions, she would turn down monetary remunerations, and often give me some *avay* (i.e. rice dumpling) that she sold to earn a living. Moreover, Ms. Uzu Paterelau from Laladhengane in Pingtung is a retired teacher always wearing a smiling face and rosy cheeks. She would take my questions very seriously and murmur to herself in Rukai at full speed for some time until she slowed down to answer me in Mandarin. I later found out that both of them are suffering from some illness. I wish them both a quick recovery. The last consultant I would like to mention is Ms. Sayungu ’e Tiaki’ana from Tapang in Chiayi. She was a retired school principal proud of her mother tongue as well as the rooftop garden she created. She had a sharp mind despite her age and was able to recognize a rather long sentence that I asked her two days earlier. Unfortunately, the Tsou dictionary she gave me two years ago has now become my last memory of her. May her soul rest in peace. I thank all the three speakers, as well as many others who I don’t have space here to cover, for sharing their precious knowledge and life experience with me. Finally, I am cordially thankful for the following organizations that have provided me with financial support in various forms: Department of Linguistics at Rice University, Ministry of Education of the Republic of China, Institute of Linguistics at Academia Sinica, Taipei Economic and Cultural Office in Houston, Archilife Research Foundation, and the Chiang Ching-kuo Foundation for International Scholarly Exchange.

Equally, if not more, important is the moral and emotional support I was bestowed upon by friends and family members. Much as I would like to list all the friends who have helped me out in times of need, I am not able to do so because this journey has been so long that I regrettably lose track of them. Alternatively, I want to single out Huiling Yang and Yueting Lee. They not only epitomize “a friend in need is a friend indeed,” but also make fuzzy the boundary between friendship and family. Last but not least, I cannot thank enough Rohan Shen and her family for their long-lasting love and support.

If this work is of any value, I would like to dedicate it to my grandmother, or *ama* as I would call her, who made innumerable sacrifices to raise me until the last few years of her life. If she can feel anything in the world where she is now, I hope I have done her proud by completing this piece of writing.

Table of Contents

Abstract	i
Acknowledgments	ii
Table of Contents	iv
List of Figures	vii
List of Tables	viii
Abbreviations & Symbols	x
Chapter 1	1
Introduction	1
1.1. Research topics and languages	1
1.1.1. Nominalization: What it is and what it is not	2
1.1.2. The connection between possession and nominalization	10
1.1.3. Formosan languages	18
1.2. Methodology and data	19
1.2.1. Theoretical orientations	19
1.2.2. Scope and structure of research	21
1.2.3. Sources of data	23
Chapter 2	26
Linguistic Background of Formosan Languages	26
2.1. Geographical distributions and linguistic vitality	26
2.2. Genetic classifications and subgroupings	29
2.3. Grammatical features	33
2.3.1. Focus morphology	34
2.3.2. Nominal relation markers	39
2.3.3. Person-form clitics	47
2.4. Grammatical relations and glossing principles	55
2.4.1. Focus categories	55
2.4.2. Argument realization and indexing	62
2.4.2.1. Ergativity and beyond	63
2.4.2.2. Argument encoding and indexing	72
2.5. Chapter summary	82
Chapter 3	84
Verbal-based Nominalization I: Synchronic Variations	84
3.1. Preliminaries: Two perspectives	85
3.1.1. Synchronic perspective: The nominal-verbal controversy	86
3.1.2. Diachronic perspective: The nominalization-into-verb hypothesis	93

3.2. Problems in previous studies	97
3.2.1. Confusion of the internal and external syntax of nominalizations	97
3.2.2. Differential treatment of first- and second-generation affixes	107
3.2.3. Differential treatment of AF- and NAF-constructions	120
3.3. Tsou	124
3.3.1. Analytic circumstantial nominalizations	125
3.3.2. Lexical argument nominalizations.....	128
3.4. Rukai	135
3.4.1. Argument nominalizations.....	136
3.4.2. Event/result nominalizations	151
3.5. Puyuma.....	157
3.5.1. AF-NAF asymmetry in nominalizations	158
3.5.2. Nominal-verbal controversy and the Nuclear Austronesian hypothesis.....	175
3.5.2.1. Nanwang Puyuma	176
3.5.2.2. Rikavung Puyuma	181
3.6. Chapter summary	195
Chapter 4.....	197
Verbal-based Nominalization II: Alternative History of Focus Affixes.....	197
4.1. Actor/possessor isomorphism	199
4.2. Lexicalization preference	202
4.3. Cognate constructions of the Mstem.....	208
4.4. Cognate constructions of NAF-words.....	222
4.4.1. Suffixal NAF-words	222
4.4.2. Mixed NAF-words.....	240
4.5. Chapter summary	248
Chapter 5.....	249
Verbal-based Nominalization III: A Case Study of Central Amis	249
5.1. Overview of Amis verb forms.....	251
5.1.1. Classes of AF verbs	251
5.1.2. Verb forms other than AF.....	260
5.1.3. Asymmetries between AF/PF and CF/LF.....	265
5.2. Argument nominalizations	269
5.2.1. Actor nominalizations.....	270
5.2.1.1. The realis Mstem-ay.....	270
5.2.1.2. The irrealis Ca~Mstem	283
5.2.2. Circumstantial nominalizations	291

5.2.2.1. Derivations of forms	293
5.2.2.2. Syntactic functions	313
5.3. Event nominalizations	320
5.4. Generalizations and reevaluations.....	326
5.4.1. External syntax	326
5.4.2. Internal syntax	337
5.5. Chapter summary	344
Chapter 6.....	346
Nominal-based Nominalization I: General Issues	346
6.1. POSS and GEN in the nominalization account.....	348
6.2. Variations of nominalized person indexes in Formosan	353
6.3. Existential vs. equational type of possession	357
6.4. Syntactic functions of possessive substantives	359
6.4.1. Predicate functions.....	361
6.4.2. Argument functions	369
6.4.2.1. NP-use.....	372
6.4.2.2. Modification-use	374
6.5. A typology of possessive substantives	381
6.6. PAN *nu reflexes and their implication on the development across types.....	389
6.7. More arguments against the “headless” analysis	400
6.8. Semantic constraints on possessive substantives	406
6.9. Chapter summary	418
Chapter 7.....	421
Nominal-based Nominalization II: Structural Types.....	421
7.1. Full-nominal possession.....	422
7.1.1. Type A: Unmarked share-NMLZ languages	423
7.1.2. Type B: Marked share-NMLZ languages	425
7.1.3. Type C: Split-NMLZ languages	429
7.2. Person-form possession.....	433
7.2.1. Type I: Unmarked share-person languages	434
7.2.2. Type II: Marked share-person languages	437
7.2.3. Type III: Split-person languages	443
7.3. Chapter summary	456
Chapter 8.....	459
Conclusion	459
References.....	471

List of Figures

Figure 2.1: Austronesian languages of Taiwan: Distributions and vitality assessment (Tsukida & Tsuchida 2007: 286)	27
Figure 2.2: Blust's (1999: 45) subgrouping of Formosan languages.....	31
Figure 6.1: Locations of fifteen Formosan languages	347

List of Tables

Table 1.1: A taxonomy of nominalizations with English examples	20
Table 1.2: Basic information on native speakers consulted in this study	24
Table 2.1: Classes of AF verbs in Tamalakaw Puyuma (after Tsuchida 1980: 211).....	61
Table 2.2: Nominal relation markers in Rikavung Puyuma (after H. Jiang 2013)	78
Table 3.1: Proto-Austronesian verbal morphology (Ross 2002: 33)	94
Table 3.2: Focus-words in non-imperative affirmative sentences in Mayrinax Atayal (after L. 1995, 2000a, 2002)	105
Table 3.3: Morphological verb classes in Budai Rukai (after L. Sung 2011: 530)	138
Table 3.4: Classes of AF verbs in Nanwang Puyuma (after S. Teng 2008: 123)	161
Table 3.5: Imperfective forms in Nanwang Puyuma (after S. Teng 2008: 123)	162
Table 3.6: Focus-words in realis AF and NAF nominalizations in Nanwang Puyuma..	170
Table 3.7: Aspectual forms of AF and NAF in Nanwang Puyuma (as per S. Teng 2008: 140-141).....	171
Table 3.8: Syntagmatic potentials of NAF-words in Rikavung Puyuma.....	194
Table 4.1: Nominalization function of the Mstem in Formosan languages.....	211
Table 4.2: Grammatical functions of Suffixal NAF-words in Formosan languages	239
Table 4.3: Reconstructions of Mixed NAF-words in PAn (as per Ross 2012: 1264)	241
Table 4.4: Nominalization function of reflexes of PF *STEM-en in Formosan languages	243
Table 5.1: AF verb classes in Central Amis	251
Table 5.2: Class I and II verbs in Central Amis.....	254
Table 5.3: External vs. internal negation of Actor nominalizations in Central Amis.....	279
Table 5.4: Actor nominalizations of Class I and II verbs in Central Amis.....	287
Table 5.5: Basic and reduplicated Mstem/Kstem pairs in Central Amis	289
Table 5.6: Morphological correspondence among AF, CF, and LF forms in Central Amis (after J. Wu 2006, 2007)	293
Table 5.7: CF-words deriving from Class I & II AF-words in Central Amis.....	296
Table 5.8: LF-words deriving from Class I & II AF-words in Central Amis (I).....	304
Table 5.9: LF-words deriving from Class I & II AF-words in Central Amis (II).....	309
Table 5.10: CF- and LF-words deriving from Class III AF-words in Central Amis	311
Table 5.11: CF and LF forms in Central Amis	312
Table 5.12: External vs. internal negation of Locative nominalizations in Central Amis	319
Table 5.13: Negation patterns of some verb forms in Central Amis	327
Table 5.14: Negative verb forms for the verbal negation in Central Amis.....	331
Table 6.1: Topic vs. nominalized person forms in Tsou.....	355
Table 6.2: Topic vs. nominalized person forms in Kavalan	355
Table 6.3: Topic vs. nominalized person forms in Tgdaya Seediq.....	356
Table 6.4: Verbal vs. nominal negators in some Formosan languages.....	364
Table 6.5: Possessive NPs in Kaohsiung Isbukun Bunun, Taromake Rukai, and Saisiyat	375
Table 6.6: Three types of possessive NPs in Northern Paiwan	379

Table 6.7: Possessive NPs in two Bunun dialects.....	380
Table 6.8: Three types of Formosan languages that retain PAn *ni.....	383
Table 6.9: Possessive NPs in Type A through C languages	384
Table 6.10: Formosan interrogative words with the PAn *nu (after Reid 2007: 250) ...	390
Table 6.11: Possessive NPs and interrogative words in some Formosan languages	392
Table 6.12: Functions of PAn *nu reflexes in Tsou, Seediq, and Amis	398
Table 6.13: Plain vs. nominalized person forms in Rikavung Puyuma.....	413
Table 6.14: Topic vs. nominalized person forms in Rikavung Puyuma.....	414
Table 7.1: Topic vs. nominalized person forms in Thao	435
Table 7.2: Topic vs. nominalized person forms in Kaohsiung Isbukun Bunun.....	436
Table 7.3: Nominalized person forms in six Type II languages	440
Table 7.4: Topic vs. nominalized person forms in Takibakha Bunun	442
Table 7.5: Topic vs. nominalized person forms in Northern Paiwan	444
Table 7.6: Topic vs. nominalized person forms in Plngawan Atayal	445
Table 7.7: Subject vs. nominalized person forms in Budai Rukai.....	447
Table 7.8: Topic vs. nominalized person forms in Saisiyat.....	448
Table 7.9: Topic vs. nominalized person forms in Tgdaya Seediq.....	449
Table 7.10: Topic vs. nominalized person forms in Central Amis	451
Table 7.11: Free nominalized person forms in Tgdaya Seediq	452
Table 7.12: Free nominalized person forms in Central Amis.....	453
Table 7.13: Functional distributions of three sets of person forms in Central Amis	456
Table 7.14: Full-nominal and person-form possession in Formosan languages.....	457

Abbreviations & Symbols

1 first person	EPIS epistemic
2 second person	ERG ergative
3 third person	EVID evidential
ABLT abilitative	EX existential
ABS absolutive	EXCL exclusive (plural)
ACC accusative	EXCM exclamative
ACT (non-Topic) Actor	EXT Extrovert (in Amis)
ADVZ adverbializer	F feminine
AF Actor Focus	FIL filler
AFF affirmative	FS false start
ALIM alimentary	FUT future
ASP aspect	GEN genitive
ASSO associative (plural)	HAB habitual
AUG augmentative	HORT hortative
AUX auxiliary	HUM human
CA Conveyance applicative	IMP imperative
CAUS causative	IND indicative
CERT (epistemic) certainty	INS instrument
CF Conveyance Focus	INT Introvert (in Amis)
CLF classifier	INTJ interjection
CMN common	INV inverse marker
CMPR comparative	INVIS invisible
COL collective	IPFV imperfective
COM comitative	IRR irrealis
CONJ conjunction	ITER iterative
COP copular	K Kstem marker
DAT dative	LA Locative applicative
DEF definite	LF Locative Focus
DEM demonstrative	LIG ligature
DEP dependent	LNK linking sound
DET determiner	LOC locative
DIM diminutive	M masculine
DIST distal	MED medial
DM discourse marker	N neuter
DSTT distant (past/future) time	NAF non-Actor Focus
DUR durative	NALIM non-alimentary
DYN dynamic	NEG negative
EMPH emphatic	NHUM non-human

NM non-masculine
NMLZ nominalizer
NMRK nominalization marker
NPST non-past
NOM nominative
NTOP non-Topic
OBL oblique
OBV obviative
OPT optative
PASS passive
PF Patient Focus
PFV perfective
PL plural
POSS possessive
PROG progressive
PROX proximal
PRS present
PRT particle

PSN personal
PSSB (epistemic) possibility
PST past
PTOP preposed Topic
QP question particle
REFL reflexive
REP reportative
RLS realis
SBJ subject
SG singular
STAT stative
SUBORD subordinator
SUP superlative
TEMP temporal
TOP Topic
UND (non-Topic) Undergoer
VIS visible

Symbols used in target languages:

- (x) x is an optional element (e.g. a segment, a morpheme, etc.)
- *(x) x's absence would produce a linguistic unit that is otherwise grammatical
- (*x) x's presence would produce a linguistic unit that is otherwise grammatical
- <x> x is an infix
- x-y x and y are at a morphemic boundary
- x~y x is a reduplicated morpheme of y (or the other way around)
- x=y x and y are joined by clisis and constitute a prosodic unit
- [x y] x and y constitute a syntactic unit
- {x/y} x and y are in a paradigmatic relationship (i.e. either x or y applies in a given slot)

|Chapter 1

Introduction

*Whoever rebukes a person will in the end gain favor rather than
one who has a flattering tongue.*
Proverbs 28: 23, New International Version

This dissertation investigates nominalization and possession in Formosan languages of the Austronesian family from a functional-typological perspective. This chapter introduces research topics and languages in this study (§1.1) and then delineates the methodology and data (§1.2).

1.1. Research topics and languages

This section explains the three key terms in the title of this study, namely nominalization (§1.1.1), possession (§1.1.2), and Formosan languages (§1.1.3).

1.1.1. Nominalization: What it is and what it is not

The term “nominalization” has a long history in the linguistic literature (e.g. Lees 1960; Chomsky 1970), but the scope of linguistic phenomena considered deserving such a term often varies from one study to another. It is thus necessary to first clarify what nominalization is all about and what it is not, as the term is applied in this present study.

Perhaps the most widely recognized sense of nominalization is that it is a word-derivation process “turning something into a noun” (Comrie & Thompson 2007: 334). The input of this process can be not only from such word classes as verbs (e.g. *feeling* < *feel*) or adjectives (e.g. *redness* < *red*), but also from nouns (e.g. *womanhood* < *woman*; all examples taken from Bussmann 1996: 804). By extension, word forms resulting from undergoing such a process are called nominalizations, or more specifically, lexical nominalizations.

At the other side of the lexical extreme, there is another sense of nominalization that happens beyond the word level, often called syntactic nominalization, which Trask (1993: 183) defines as “a noun phrase derived from another category which is not a projection of the lexical category Noun, particularly from a verb phrase or sentence.” English examples illustrated by him are given in (1).

- (1) English (Trask 1993: 183)
- a. [*Lisa's going topless*] *upset her father.*
 - b. [*To quit your job*] *would be a mistake.*
 - c. [*That she smokes*] *surprises me.*

In these examples, a group of words as a whole function as the subject just like a lexical noun, despite the fact that no single word form in it denotes the same class of entities as the whole construction.

Some might object that the *that*-construction in (1)c is not a nominalization because no category-specific morphosyntactic changes are made to the “finite” verb *smokes*, unlike *going* in (1)a or *to quite* in (1)b, both of which are “non-finite”.¹ This is precisely why some researchers prefer restricting nominalization only to those cases involving transcategorical markings, and propose a separate term for the more functional and thus form-independent notion of making reference to an entity concept with whatever linguistic means available in a language. For such a functional notion, Spruiell (1990) proposes the term “participantization” and González (2012) “referentialization.” However, transcategorical markings are language-specific (and even construction-specific) mechanisms for signaling whether a linguistic construction is reporting and asserting an event or used as an encapsulated prop to be manipulable in the discourse on a par with typical nouns (see Hopper & Thompson 1984). There are languages where no morphosyntactic changes are necessary for either function, as illustrated by Mandarin Chinese in (2).

(2) Mandarin Chinese (Hopper & Thompson 1984: 737)

- a. **tā** **zài** **tiàowǔ**
 3SG DUR dance
 ‘{She/He} is dancing.’
- b. **tiàowǔ** **hěn** **méiyìsi**
 dance very dull
 ‘Dancing is very dull.’

In the second example, “a state of affairs is being presented as non-challengeable... as an argument...[T]his function is known as a ‘nominalization’, whether or not it occurs with characteristic morphology” (Hopper & Thompson 1984: 737). If we exclude the

¹ For instance, Milsark (2006: 436-437) takes the bracketed phrase in [*Their playing the overture so badly*] *disgusted the audience* to be a nominalization whereas that in [*That the tree was dying*] *saddened us all* a complementation.

Mandarin case from the domain of nominalization simply because of its lack of dedicated morphosyntactic markings, then the very notion of nominalization is essentially dependent on what morphosyntactic markings are available in a language and may thus not be applicable in all languages. Nevertheless, it is generally desirable to use the same term for a cross-linguistically comparable phenomenon, which, in this case, is how languages package a state of affairs into a linguistic unit such that it can be commented on or acted upon just like a typical noun in those languages. It is in this sense that the three examples in (1) are comparable not only within English but also with the Mandarin one in (2)b. Morphosyntactic markings, on the other hand, are hardly comparable across constructions in a language, let alone across languages. So, it is decisively unclear whether (2)b from Mandarin should be analogized to which of the three in (1) from English since the distinct morphosyntactic markings found in the latter are simply non-existent in the former to begin with. Moreover, finiteness may be definable in one language, but a cross-linguistically valid definition is hard to come by because it is often distributed over an array of features (Maas 2004; Aikhenvald 2011: 264). Therefore, both transcategorical markings and the notion of finiteness are better kept apart from the notion of nominalization.

In addition, variations within and across languages with respect to the internal structure of encapsulated expressions with thing-like denotations make transcategorical markings even more trivial than the widespread pattern whereby languages treat lexical nouns and complex structures in a similar manner using whatever grammatical means readily available to them. This point is best illustrated by what Dryer (2004: 47) calls “noun phrases without nouns”, namely, those “lacking a noun which denotes the kind of

thing that the referent of the noun phrase belongs to.” He shows nominal expressions of this nature from diverse languages, including Kutenai in (3).

(3) Kutenai (Dryer 2004: 58)

- a. ʔa ʔ'inax-i [niʔ niʔtahat-nana]
 back go-IND DEF boy-DIM
 ‘The little boy went back.’
- b. ʔa ʔ'inax-i [niʔ k=uqaka]
 back go-IND DEF SUBORD=win
 ‘The one who won went back.’
- c. ʔukat-i [niʔ-s k=aʔxu] xaʔʔin
 take-IND DEF-OBV SUBORD=carry dog
 ‘Dog took what she carried.’

In these examples, NPs consist of a definite marker plus either a noun or a verb marked by something that indicates its embedded status (called subordinator in the Kutenai literature). Crucially, embedded verbs are categorically nothing like nouns because they can potentially have argument structures identical to those of non-embedded matrix verbs. The traditional analysis of NPs like those in (3)b-c is to posit an elliptical or “missing” noun, which is modified by clause-like structures. In other words, they are so-called “headless” or “free relative” clauses. However, Dryer (2004) argues against the elliptical analysis for at least three reasons. First, these NPs are often used in contexts where speakers cannot provide a noun. Second, these NPs may have denotations so generic that the only possible noun that could have been elided would mean something like ‘thing’, but it is doubtful that such an abstract noun exists in all the languages that have NP structures similar to those in Kutenai. Third, in some languages there are simply no NP structures from which their elliptical versions can be arrived at simply by eliding a noun.

Like Dryer (2004), Shibatani (2009) and Shibatani & bin Makhshen (2009) consider the elliptical analysis empirically unmotivated. Furthermore, they have argued

that the elliptical analysis reflects the skewed perspective in the tradition of linguistics to start from endocentric NPs with modifying elements and modified nouns and then consider exocentric ones without modified nouns as derivatives of the former. In this view, attributive clause-like structures are always relative clauses, even when they make up the whole NP without syntactically modifying anything. However, on the basis of relative constructions in various languages, Shibatani (2009) argues that what has long been called relative clauses are nominalized structures, called grammatical nominalizations, which are “neither syntactically nor semantically subordinate to, or dependent on, the nominal head they modify” (ibid.: 163). Grammatical nominalizations denote entity (thing-like) concepts on their own right just like basic nouns such that they can fulfill NP slots and refer to something (the referential function) or modify a noun within NPs and help restrict the reference of that noun (the restricting function).

Likewise, basic nouns can be referential (e.g. *the dog*) or restrictive (e.g. *the dog food*). The equivalent effect of applying the relativization-based (or more generally modification-based) perspective to basic nouns would be similar to saying *the dog* is a headless or elliptical version of *the dog food*, which is simply “wrong-headed” (Shibatani & bin Makhshen 2009: 321). Unlike basic nouns, however, grammatical nominalizations like those in (3)b-c do not have ontologically homogeneous denotations. While *the dog* can only refer to a specific class of animals (as well as other idiosyncratic entities based on metaphorical or metonymic extensions), grammatical nominalizations like *what she carried* may refer to miscellaneous transportable objects from diverse semantic domains (e.g. a dog, an umbrella, a key, etc.), thus lacking a time-stable lexical status. Much as lexical nominalizations may denote either a state of affairs (e.g. *employment* < *employ*) or

a crucial participant in it (e.g. *employer* < *employ*), which can be respectively referred to as event and argument nominalizations, so do grammatical nominalizations. Grammatical event and argument nominalizations are respectively exemplified by the bracketed constructions in (1) from English and those in (3)b-c from Kutenai.

Although the observation between nominalization and relativization is not entirely new, the claim in Shibatani (2009) is stronger than ever and the definition of nominalization is much more functional than recognized by most functional linguists. For instance, Givón (2001: 190) considers nominalization one of the many “strategies” for relativization in many languages. Comrie & Thomson (2007: 379) claim that “in certain languages relativization is indistinct from nominalization.” In these two studies, nominalization is judged on the basis of whether clause-like structures have distinct markings when nominalized as compared with when used for matrix predication, or whether clause-like structures co-occur with demonstratives or case markers. However, from the functional perspective in Shibatani (2009), the morphosyntactic markings previously identified as nominalization “strategies” are merely formal indicators of the more general nominalization function, which is to denote entity concepts. If a language uses clause-like structures to denote something in the same way Kutenai does, but without identifiable morphosyntactic markings (such as the subordinator *k=* in (3), which might as well be called a nominalizer by now), it should have the nominalization function all the same since languages are comparable because of what they do rather than what linguistic forms are used to perform what they do. An analogy from possessive NPs is that simple juxtaposition (e.g. *Dumnab ram* ‘Dumnab’s house’ in Kobon) is just as good an encoding mechanism as nominal declension (e.g. *kniga Ivan-a* ‘Ivan’s book’ in

Russian; both examples taken from Croft 2003: 32-34) despite lack of identifiable morphosyntactic markings in the former. Therefore, the stronger claim in Shibatani (2009: 196) is that “in a large number of languages, if not most, around the globe, nominalization is the basis for relativization.” In other words, the nominalization process creates linguistic structures needed for their restrictive function, which is traditionally called relativization.

Nominalization as outlined by Shibatani (2009) may seem too encompassing, but his approach is in fact harmonious with how nominalization is defined in some functionally oriented traditions. In Functional Grammar, nominalization covers cases where “a group, phrase or clause comes to function as part of, or in place of (i.e. as the whole of), a nominal group” (Halliday & Matthiessen 2004: 358). The referential and restrictive function of grammatical nominalizations respectively serves in place of and as part of a nominal group. In Cognitive Grammar, nominalization is a cognitive process of conceptual reification whereby linguistic expressions of various sizes are grouped and manipulated “as a unitary entity for higher-order purposes” (Langacker 1999: 3). Working in this framework, Heyvaert (2003) takes a similar “radically functional” approach to nominalization in English. Moreover, her cognitive-functional approach and Shibatani’s functional-typological approach both manage to distinguish nominalizations from clauses in functional terms. Grammatical nominalizations, in which no single word form can be reasonably analyzed as the lexical head noun of the whole construction, have been called “clausal nominalizations” or “nominalized clauses” (Comrie & Thomson 2007: 376). The last two terms might give the impression that nominalizations can sometimes be clauses or vice versa, which would downplay the functional definition of

nominalizations. However, Heyvaert (2010: 73) explicitly points out that what nominalizations do is to “reclassify a processual starting point, of which the ‘size’ can vary from a verb stem to a processual expression with complements or even a full clause.” In other words, clauses can be the starting point of nominalizations, but they are not the same animal as nominalizations. Thus, nominalized clauses simply mean nominalizations that have a clausal character (by illustrating the same structural features found in clauses). More importantly, nominalizations and clauses fulfill different pragmatic functions despite their potential structural resemblance. According to Shibatani’s (2011) functional definitions, nominalizations denote entity concepts and presuppose a state of affairs whereas clauses predicate and assert a state of affairs. In this functional sense, so-called relative clauses are not really clauses but nominalizations instead since they presuppose rather than assert a state of affairs. For instance, “She bought me the book” asserts book-buying, but “I am reading the book that she bought me” presupposes rather than asserts book-buying. Importantly, lack of assertion in nominalizations allows them to be embedded into higher-order syntax.

In brief, nominalization is more than just a word-derivation process. More generally, it is “a metonymic process yielding constructions associated with a denotation comprised of entity (thing-like) concepts that are metonymically evoked by the nominalization structures such as events, facts, propositions, and resultant products (event nominalization), as well as event participants (argument nominalization)” (Shibatani 2016). As products of such a process, nominalizations share nominal syntactic properties with nouns by virtue of their association with an entity-concept denotation, which allows nominalizations and nouns alike to function as NP heads (the NP-use) or in

construction with a head noun (the modification-use). A language may have several structural types of denoting expressions with varying degrees of internal complexity, but picking one or some of them as nominalizations while disregarding others is not feasible when it comes to comparing nominalizations across languages.

1.1.2. The connection between possession and nominalization

Comrie & Thompson (2007: 379) discuss several nominalization processes for creating new nouns out of existing ones. Through various morphological means, a source noun can be turned into a derived one denoting the abstract quality of the source noun's referent (e.g. *womanhood* < *woman*) or having attributes that are either quantitatively (e.g. augmentative) or qualitatively (e.g. pejorative) different from those of the source noun's referent. Since these processes all take nouns as the input, they can be referred to as nominal-based nominalization, as opposed to verbal-based nominalization discussed in the previous section. In addition, there is yet another type of nominal-based nominalization that is normally not recognized in the modification-based perspective. It involves what is often called genitive or (adnominal) possessive constructions. This section discusses the connection between the notion of possession with the functional definition of nominalization adopted in this study.

To begin with, possessive NPs are illustrated by English in (4).

(4) English (Dryer 2004: 48)

[*The dog's dishes*] are here and [*the cat's*] are over there.

Unlike the first possessive NP, the second one does not contain a noun that denotes the type of things that the whole NP denotes, and is thus another type of "noun phrases

without nouns” discussed in Dryer (2004). Independent possessive phrases like *the cat’s* are often referred to in the literature as “free genitives” (e.g. Stolz et al. 2008: 390), “headless adnominal” (e.g. Noonan 2008a: 130), “headless possessive” (Koptjevskaja-Tamm 1995; van der Voort 2009), or “genitive phrases without the possessum” (e.g. Lander 2009: 588), all of which reflect the modification-based perspective whereby the construction in question is considered a derivative of the syntagm between the possessor (POR) and the possessum (PUM). However, in the nominalization-based perspective argued for by Shibatani (2009), terms like “free genitives” are just as oxymoronic as “free relatives” because both of them are equally denoting expressions on their own right. Much as the suffix *-hood* is a nominal-based nominalizer that derives a noun denoting the abstract quality of the referent of the base noun (see Hamawand 2011: 159 for more nominalizing suffixes in English), the possessive morpheme *’s* is a nominal-based nominalizer that creates a nominal whose denotation is metonymically associated with the referent of the base nominal. In the words of Taylor (1996: 20), the possessive morpheme *’s* “designates a schematic instance” of the PUM. In Cognitive Grammar parlance (Langacker 2009a), if *the cat* is a conceptual reference point, *the cat’s* evokes a set of target nominals mentally accessible in relation to the reference point. Following Nagaya (2011), I shall refer to the reference point and target as the *plain nominal* and *nominalized nominal* respectively. While nominalized nominals like *N-hood* (where *N* stands for a noun) have rather coherent semantic properties, similar to verbal-based lexical nominalizations, those like *X’s* (where *X* stands for a nominal phrase) denote ontologically heterogeneous entities, akin to verbal-based grammatical argument nominalizations.

More importantly, a unified nominalization-based account of both genitive/possessive and relative constructions is motivated by the symbolic and structural parallelism between the two in diverse languages. Much as relative constructions involve a schematic entity which bears crucial relevance in an event (e.g. the agent, patient, etc.), genitive/possessive constructions do so with a schematic entity that bears crucial relevance with respect to another entity (i.e. the POR). The schematic entity concept is created in a nominalization process that relates to a presupposed state of affairs in the former case (i.e. verbal-based nominalization), and to a cognitively salient entity in the latter (i.e. nominal-based nominalization). In both cases, a new denoting expression is created out of something known, which is “in essence a metonymic process... and... not mechanically constrained” (Spruiell 1990: 119). In this regard, grammatical nominalizations are like exocentric compounds that denote entities in terms of their salient features (e.g. *redbreast* for a type of robins). Assuming there is a missing noun every time a grammatical nominalization takes up an NP is tantamount to saying there is a missing *bird* (or some other nouns) every time we use *redbreast*, which is counterintuitive.

Although genitive/possessive constructions are traditionally discussed under the notion of possession, numerous studies (e.g. Heine 1997; Baron et al. 2001; Stolz et al. 2008; McGregor 2009a; Aikhenvald & Dixon 2013) have come to the conclusion that the range of their semantics is so diverse, both within and across languages, that “linguistic possession implies nothing more than *the existence of some association* or relationship between possessor and possessed” (Langacker 2009b: 81; emphasis mine). In fact, many grammarians even use the term “associative” to describe constructions that express

prototypical possessive relationships and beyond, including Li & Thompson (1981: 113) for Mandarin, Noonan (1992: 156) for Lango, Frajzyngier & Shay (2002: 324) for Hdi, and Kruspe (2004: 214) for Semelai, to name just a few. Thus, the idea that these scholars try to capture with the term “association” matches the functional notion of nominalization as a metonymic process.²

The structural parallelism between nominal-based genitive/possessive and verbal-based relative constructions is reflected in not only the morphemes involved but also their syntactic behaviors. This point can be illustrated by Amami Ryukyuan languages discussed in Shibatani & Shigeno (2013). They distinguish two types of nominalization-related markers. One is called a *nominalizer* (hereafter NMLZ), which is directly responsible for converting a source structure into a nominalization, and the other is a *nominalization marker* (hereafter NMRK), whose major function is not to create a nominalization per se but is instead to indicate that a nominalization is being used as an NP head.³ For instance, Yoron Amami uses *si* and *mun(u)* in both nominal-based (N-based) and verbal-based (V-based) nominalizations when they fulfill a complete NP (the NP-use), but not when they modify a noun (the modification-use), as in (5).

²In this sense, genitive/possessive constructions are quite similar to associative plurals (Daniel & Moravcsik 2013). In both cases, there is a focal salient reference entity (called X) and there is typically a relator (called Y) that derives X into nominalized nominals or associative plurals. They both denote a group of heterogenous entities associated with X, but they differ in one crucial point: the denotation of X-Y includes X for associative plurals but excludes X for nominalized nominals (cf. *tanaka-tachi* ‘Tanaka and his associates’ vs. *tanaka=no* ‘what pertains to Tanaka’ in Japanese). However, associative plurals fall outside the scope of this present study.

³Markers with functions similar to the NMRK are sometimes referred to as “dummy” (Lehmann et al. 2004: 57), “absolute” (Heath 1999: 87), or “prop-word” (Rissanen 1997: 117).

(5) Yoron Amami (Shibatani & Shigeno 2013: 132)

- a. *sinse:=nu hasa*
 teacher=NMLZ umbrella
 ‘teacher’s umbrella’ [Modification-use in N-based nominalization]
- b. *sinse:=nu={si/mun(u)}=ja are*
 teacher=NMLZ=NMRK=TOP DIST
 ‘The teacher’s is that.’ [NP-use in N-based nominalization]
- c. [*aca=ga jumju:-ru sinbun*]
 father=NOM read-NPST.NMLZ newspaper
 ‘the newspaper which Father is reading’ [Modification-use in V-based nominalization]
- d. [*amma:=ga mica={si/mun(u)} ko:sa*]
 mother=NOM cook.PST.NMLZ=NMRK ate
 ‘I ate what Mother cooked.’ [NP-use in V-based nominalization]

Distinct markings of NMLZ and NMRK in N-based nominalizations are widespread in various Japanese and Ryukyuan varieties, although the two types of markers have been converged into one in Modern Standard Japanese (see Shibatani 2013). Based on the patterns of constructions like those in (5) across Amami Ryukyuan languages, Shibatani & Shigeno (2013) propose an overall diachronic development whereby markings spread from the NP-use to modification-use and from N-based to V-based nominalizations. Consequently, if a NMRK that was originally found in the NP-use spread to the modification-use, a new modification pattern would arise. This is precisely the case in Uyama Amami, where the NMRK *mun* is generally found in the NP-use of N-based nominalization but is additionally used in the modification environment as well when singular speech-act participant (SAP) person forms are involved, as shown in (6) below. Differential markings of N-based nominalization across person forms seem to be a common phenomenon, as we will also see in Formosan languages.

(6) Uyama Amami (Shibatani & Shigeno 2013: 124)

- a. **sinsi:=nu** kutsu
 teacher=NMLZ shoes
 ‘teacher’s shoes’ [Modification-use in N-based nominalization]
- b. **sinese:=nu=**mun**** do
 teacher=NMLZ=NMRK PRT
 ‘It’s teacher’s.’ [NP-use in N-based nominalization]
- c. **{wa:/ja:}(=**mun**)** kutsu
 {my/your}=NMRK shoes
 ‘{my/your} shoes’ [Modification-use in N-based nominalization]
- d. **{wa:/ja:}=**mun**** do
 {my/your}=NMRK PRT
 ‘It’s {mine/yours}.’ [NP-use in N-based nominalization]

In addition to Japanese and Ryukyuan languages, the structural parallelism between genitive/possessive and relative constructions is so widespread in Sino-Tibetan languages that the phenomenon is dubbed Standard Sino-Tibetan Nominalization (SSTN) by Bickel (1999). More generally, Aristar (1991) found that identical or similar morphology is used for both constructions in nine separate language families from diverse geographical areas, with a dedicated morpheme even reconstructable for Indo-European. As a result, he concluded that “[t]he similarity can be neither genetic nor areal... there is a *prima facie* case that the *two patterns are generated by the same process*” (ibid.: 10; emphasis mine). That process, according to him, is the “head-referent binding-anaphor strategy,” whereby a pronominal element combines with modifiers to facilitate their incorporations with higher-order syntax. Interestingly, what he called binding anaphors in fact corresponds well with NMLZ or NMRK, and his description of newly arising noun-modifying constructions in Agaw languages outlines a diachronic development similar to Shibatani & Shigeno’s (2013) proposal for Amami Ryukyuan languages (see also the grammaticalization from apposition to attribution proposed for Indo-European languages by Lehmann 2002: 64). Specifically, genitive/possessive and

relative modifiers in most Agaw languages are marked by the same set of bound morphemes that agree with the modified noun with respect to its gender and number (called agreeing modifiers), and only some of them additionally have separate genitive/possessive and relative modifiers that do not agree with the modified noun (called non-agreeing modifiers). Crucially, agreeing modifiers historically arose later than non-agreeing ones, which are lost in some languages. The former are formed on the basis of the latter by adding agreeing morphemes, as illustrated by Bilin Agaw in (7).

(7) Bilin Agaw (Aristar 1991: 13)⁴

- a. *səxant-a* *gərw-əd*
 be.merciful-NMLZ man-DAT
 ‘to the man who is merciful’ [Modification-use]
- b. *səxant-ä-x^w-əd*
 be.merciful-NMLZ-NMRK.M.SG-DAT
 ‘to the one (masculine) who is merciful’ [NP-use]

In the present terminology, the “agreeing modifier morpheme” *-x^w* (for M.SG; cf. *-ri* for F.SG and *-w* for PL) in the second example is a NMRK because it indicates that a denoting expression is being used in the referential function, as opposed to the restricting function in the first example. Strictly speaking, it is not an “agreeing” morpheme because in the second example there is no noun with which it could have agreed. Rather, it designates a schematic entity that is grammatically masculine and singular. Once the NMRK migrates to the modification context, along come new-generation genitive/possessive and relative constructions (i.e. agreeing modifiers), as in (8).

⁴ The morpheme *-a* alternates with *-ä* when in the word-internal position. Both are originally glossed as REL.

- (8) Bilin Agaw (Aristar 1991: 13)⁵
- a. 'aqwa [ja'ag-na-ɣ^w]-əɪ
 water drink-1PL-NMLZ.M.SG-to
 'to water we do not drink' [Modification-use]
- b. ti'idad [adäri-ɣ^w]-əd
 order lord-NMLZ.M.SG-DAT
 'by the order of the lord' [Modification-use]

The “head-referent binding-anaphor strategy”, which Aristar (1991) argues to be the process that gives rise to the structural parallelism between genitive/possessive and relative constructions in diverse language families, is more generally a process of creating grammatical nominalizations, which do not necessarily involve anaphoric constructions, as in Japanese and Ryukyuan languages. More importantly, even when grammatical nominalizations do consist of markers that are historically anaphoric, the use of grammatical nominalizations is by no means restricted to anaphoric contexts. For instance, because referents of grammatical argument nominalizations are unspecified, they are commonly used in contexts where the identity of their referents is provided by a noun they modify (i.e. relative constructions) or one with which they are equated (i.e. cleft-type constructions), or is inferred from the linguistic (i.e. the anaphoric use) or situational context, or where such an identity is the subject of inquiry (i.e. content-word questions), or irrelevant at all because a generic statement is being made to all their potential referents (Shibatani 2009: 193).

Finally, languages do not always use the same formative for both nominal-based and verbal-based nominalization, but even when the isomorphism does not occur, we can still talk about nominal-based nominalization as such because it often involves formatives such as demonstratives, definiteness markers, or nominal classifiers, which play crucial

⁵ Given the data presented in Aristar (1991), -x^w seems to be an allomorph of -ɣ^w, which is originally glossed as REL in the first example and GEN in the second.

roles in verbal-based nominalization across languages (see Yap et al. 2011). Another connection between nominal-based genitive/possessive constructions and verbal-based nominalizations is the crosslinguistic trend to encode a salient participant in the latter as if it were the POR nominal in the former (see Lander 2009). This is also the case in Formosan languages, to which I turn below.

1.1.3. Formosan languages

There are at least two senses of the term “Formosan languages.” In one sense (Tsukida & Tsuchida 2007), it refers to all the indigenous languages of Taiwan (historically known as Formosa due to the epithet *ilha formosa* ‘beautiful island’ given by Portuguese sailors), which predate the Han Chinese immigration in the early 17th century. They all belong to the Austronesian family. In the other sense (Blust 1999), which is narrower, the term designates a collection of languages that are negatively defined in genetic terms, namely, non-Malayo-Polynesian Austronesian. The two senses would have identical reference were it not for Yami/Tao (spoken mostly on Orchid Island of Taiwan), which is included in the first sense but excluded from the second. To disambiguate, I use Austronesian languages of Taiwan for the first sense and save Formosan languages for the second.

The significance of Formosan languages can be stated in at least two aspects. Genetically, despite their small number, Formosan languages cover nine out of ten first-order subgroups of the Austronesian family (with at least over one thousand languages) according to Blust’s (2009) subgrouping account based on phonological evidence. Many of them are morphosyntactically rather conservative and retain the verbal morphology

that has been reconstructed for Proto-Austronesian. Thus, how we understand Formosan languages would have direct import on what Proto-Austronesian would have been like. On the other hand, when compared with other Austronesian languages, Formosan languages are “extremely diverse at all linguistic levels, from phonology to morphology to syntax” (P. Li 2008a: 523). This linguistic diversity makes Formosan languages still a great object of crosslinguistic comparisons even though they are all genetically related.

More details about Formosan languages will be presented in Chapter 2.

1.2. Methodology and data

This section outlines the overall methodology, including theoretical orientations (§1.2.1) and the scope and structure of the dissertation (§1.2.1). It also specifies sources of data (§1.2.3) employed in this study.

1.2.1. Theoretical orientations

The theoretical orientations adopted in this study is the *functional-typological approach*, which “became generally recognized in the 1970s” (Croft 1999: 87). The term “functional” is set against “formal”. While the formal approach defines the domain of inquiry in terms of linguistic forms, the functional approach does so in terms of linguistic functions. Also, unlike the formal approach, the functional approach puts emphasis on the issue of comparability in crosslinguistic comparisons. The functional approach holds that it is problematic to use linguistic forms as the basis of crosslinguistic comparisons since they can vary dramatically from one language to another. On the other hand, the term “typological” implies at least three types of activities when analyzing crosslinguistic data:

classifying a functional domain of research into structural types, making generalizations across types, and finally explaining the generalizations thus established (ibid.). Although it also deals with crosslinguistic data, the formal approach differs crucially from the functional approach in that the former seeks explanations within languages (e.g. abstract constructs defined by constituency structures) whereas the latter does so beyond linguistic knowledge (e.g. social, interactional, and cognitive motivations). Other basic tenets and theoretical principles in the functional-typological approach have been elaborated in such works as Stassen (1985), Comrie (1989), and Croft (2003), and they are followed here.

The domain of research in this study is nominalization, as has been outlined in §1.1.1 and §1.1.2. The types of nominalizations mentioned previously are summarized in Table 1.1 and illustrated with English examples.

Table 1.1: A taxonomy of nominalizations with English examples

Input types	Denotations	Lexical	Grammatical
Verbal-based (PROCESS-based)	Event	<i>employ-ment</i>	<i>that he told me his story</i>
	Argument	<i>employ-er</i>	<i>the one who told me</i>
Nominal-based (THING-based)	Associated Entity	<i>midfield-er</i>	<i>the woman's</i>

The taxonomy here is mostly based on Shibatani (2010, 2011). Verbals and nominals, can be understood as linguistic expressions of any size that respectively designate (or profile) PROCESS and THING as they are defined in Cognitive Grammar. A PROCESS is “a complex relationship that develops through conceived time and is scanned sequentially

along this axis” (Langacker 2008: 112), and a THING is “any product of grouping and reification” (ibid.: 105). These broad definitions allow room for nominalization processes that take complex structures as their input, which often happen when clitics are involved. For instance, the input of nominal-based grammatical nominalizations can not only be lexical nominals, but also grammatical ones, as in [*The person who told me’s*] *name must remain a secret* (Anderson 2005: 92).

1.2.2. Scope and structure of research

This study comprises two primary parts, one on verbal-based nominalizations and the other on nominal-based ones.

Verbal-based nominalization in Formosan languages, or more generally, western Austronesian (i.e. non-Oceanic Austronesian; Himmelmann 2005a), is an especially interesting topic because of the intricate interaction between verbal morphology and the grammaticalized semantic roles assigned to a specific argument. Such a system has been hypothesized to result from a drastic renovation of the verbal system in either the Proto-Austronesian (PAN) period or some interstage after that. According to the nominalization-into-verb hypothesis (Starosta et al. 1982), some modern verbal forms were historically nominalizations while others have remained strictly verbal throughout the history, and grammatical functions of the latter type have been greatly reduced over time due to the expansion of the former type. This hypothesis also becomes the basis of a subgrouping account (Ross 2009, 2012) where three Formosan languages (non-Nuclear) do not form the same subgroup as all the other Austronesian ones (Nuclear) because the ancestor of Nuclear languages innovated the nominalization-into-verb reanalysis, which did not

happen to non-Nuclear ones. One of the goals of this study is to reassess these claims by making fairer comparisons across various verbal forms, both within and across languages. To achieve that, Chapter 3 first identifies conceptual problems in previous studies and focuses on synchronic variations of verbal-based nominalizations across languages, especially those of the three non-Nuclear languages. On the basis of Chapter 3, Chapter 4 attempts to revive an alternative account of the Austronesian verbal morphology without resorting to the nominalization-into-verb hypothesis. The significance of the alternative account is that if it turns out to be correct any subgrouping account relying on the nominalization-into-verb hypothesis would lose its foundation. Chapter 5 is an in-depth case study on verbal-based nominalizations in one particular Formosan language, namely Amis, where many PAn verbal affixes are retained but organized in a way different from her sister languages.

On the other hand, Chapter 6 and Chapter 7 investigate nominal-based grammatical nominalizations in fifteen Formosan languages/dialects. The Formosan literature shows vigorous interest in their modification-use, that is, the possessor-possessum syntagm, but generally pays little attention to their NP-use, or phrases including or relating to the possessor but denoting the possessum instead, called *possessive substantives* by Ultan (1978). Possessive substantives in Formosan languages are important because they reveal the different syntactic functions of so-called genitive markers across languages even when cognate forms are involved. Moreover, by comparing forms in the modification- and NP-use, we are able to define structural types across languages, and distributions of cognate forms across types, in turn, provide us with some clues as to how the observed types might have evolved. Chapter 6 discusses these

general issues and Chapter 7 presents specific examples from each of the fifteen Formosan languages/dialects.

Finally, Chapter 8 concludes the present study and discusses its implications.

1.2.3. Sources of data

The linguistic data that form the basis of the present study are both first- and second-hand. First-hand data were collected through direct elicitations with native speakers. Since the field research involves human subjects, I have applied for Rice Institutional Review Board (IRB) approval and secured clearance from it (Protocol Number 09-127X). The field research qualifies for exempt status and has been granted the approval (Rice Federal-Wide Assurance Number 00003890). Second-hand data were drawn from various publicly available sources, including academic publications, language teaching materials (9-Level Textbooks and Supplementary Materials)⁶, Online Dictionaries of Indigenous Languages (ODIL)⁷, Corpus of Formosan Languages at National Taiwan University (NTU Corpus)⁸, and Formosan Language Digital Archive at Academia Sinica (FLDA)⁹.

First-hand data are all indicated as such by “Fieldnotes” in parentheses, and native speakers consulted in this study are listed in Table 1.2 below. Sources of all the second-hand data cited here are always specified. Unless otherwise stated, Formosan data are

⁶ <http://web.klokah.tw/>

⁷ <http://e-dictionary.apc.fishweb.com.tw/Index.htm>

⁸ http://corpus.linguistics.ntu.edu.tw/index_en.php

⁹ <http://formosan.sinica.edu.tw/>

Table 1.2: Basic information on native speakers consulted in this study

Languages	Dialects/Varieties	Ethnic Names	Years of Birth
Amis	(Coastal)	Ipay Kukuy	1943
	Central	Ramay Kalitang	1961
Atayal	(Wulai)	Umasu Utaw	1929
	Squliq	Yayu Lasun	1930
	PIngawan	Kumuy Nawi Yumin Nawi	1939 1948
Bunun	Takibakha	Abuc Tasitaluman Tiang Maitangan	1931 1949
	(Kaohsiung) Isbukun	Husung Istanda Hanaivaz Takistaulan	1933 1951
Kavalan	PatRungan	Abas Ungi	1933 1941
Paiwan	(Makazayazaya) Northern	Kaleskes Lataukatu	1949
Puyuma	Nanwang	Waka Raera	1936
	Rikavung	Malada Atrung	1935 1936
	Katripul	Kimti Lisem Kadadepan	1935 1941
Rukai	Budai	Uzu Paterelau Paerec	1948 1951
	Taromake	Lralruy	196?
Saaroa	---	Amalanamalhe Salapuana	1948
Saisiyat	Northern	Oyong a TaheS	1945
Seediq	Tgdaya	Lubi Neyung	1934
		Ape Neyung	1946
		Dakis Pawan	1954
Thao	---	Kilash Lhkatafatu	1923
		Lujang Katamarutaw	1937
Tsou	Tapang	Sayungu 'e Tiaki'ana	1939

transcribed in the conventional orthographies, following the writing systems recommended by the Ministry of Education, Taiwan, but with some minor modifications,

which will be explained in footnotes where relevant.¹⁰ Except for obvious typos and transcription formats, second-hand data are always presented as they are in the original sources. However, their glosses are subject to change since they are “part of the analysis, not part of the data” (Leipzig Glossing Rules), and their free translations may be modified as I see fit. Finally, gloss abbreviations follow the suggestions in Leipzig Glossing Rules wherever possible.

¹⁰ Correspondences between conventional orthographies and IPAs are downloadable at: http://language.moe.gov.tw/result.aspx?classify_sn=42&subclassify_sn=448

|Chapter 2

Linguistic Background of Formosan Languages

To set the scene for subsequent discussions, this chapter provides a linguistic background of Formosan languages. Topics covered include geographical distributions and linguistic vitality (§2.1), genetic classifications and subgroupings (§2.2), grammatical features relevant to nominalization (§2.3), and finally grammatical relations and glossing principles (§2.4).

2.1. Geographical distributions and linguistic vitality

Based on historical records, there were once roughly 26 Austronesian languages spoken in Taiwan, “roughly” because whether a linguistic system counts as a separate language or dialect of a language is always at issue and disputable. Their geographical

distributions over Taiwan are mapped in Figure 2.1, where the degrees of endangerment for each language are also indicated.¹¹

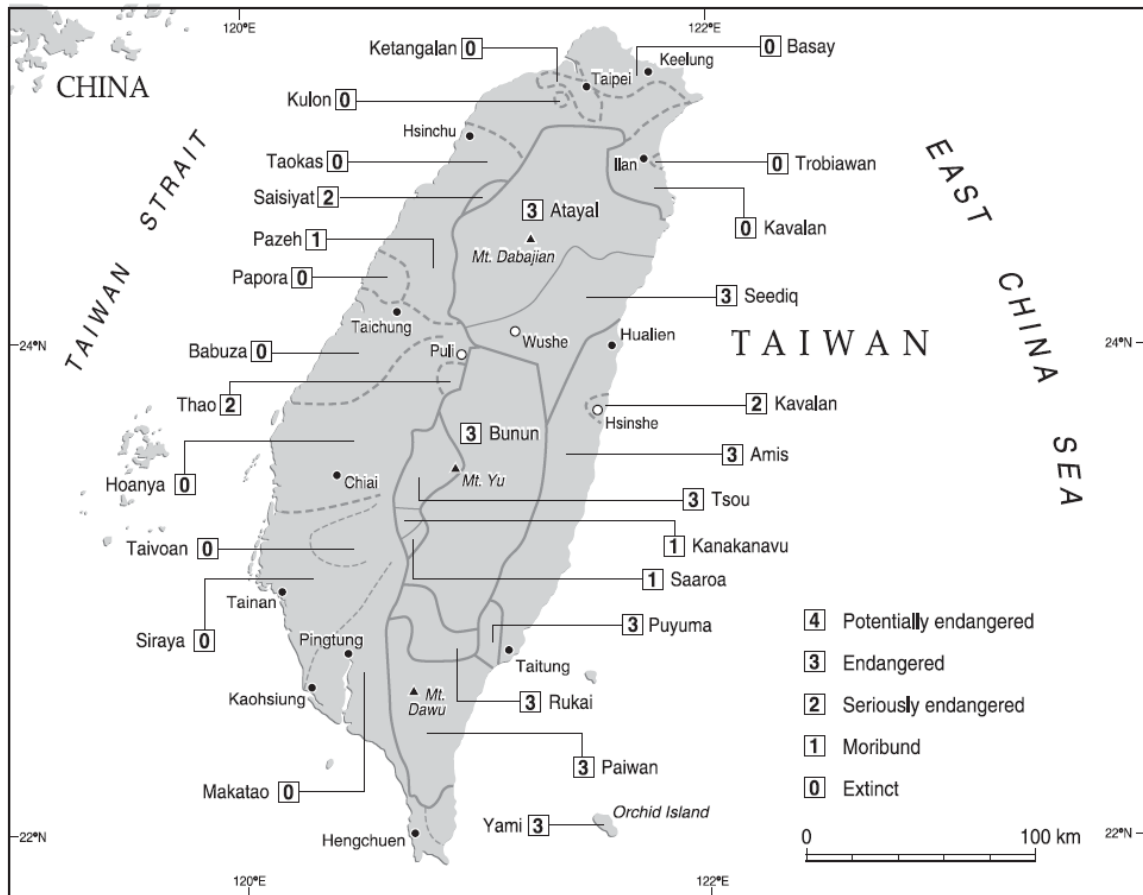


Figure 2.1: Austronesian languages of Taiwan: Distributions and vitality assessment (Tsukida & Tsuchida 2007: 286)

The linguistic vitality assessed by Tsukida & Tsuchida (2007) is generally congruent with a later update from UNESCO Atlas of the World's Languages in Danger (Moseley 2010). Notice that the Kavalan language is indicated twice on the map, once as extinct in Yilan County, and the second time as seriously endangered around Hsinshe area in Hualien County. This is because Ilan was historically the major settlement of the Kavalan people, who, starting from the early 19th century, migrated southwards en masse to Hualien and

¹¹ I would like to thank the authors for granting me the permission to use their map here.

Taitung County due to their conflicts with the Han Chinese (F. Hsieh & S. Huang 2007: 95). Hsinshe is allegedly the primary Kavalan settlement of modern time. However, aside from Hsinshe, there are at least ten other neighboring locations (all within Hualien and Taitung County) where fluent native speakers have been found (see D. Yen 2012: 3).

Languages such as Ketangalan (or Ketagalan), Taokas, Papora, Babuza, Hoanya, and Siraya must have been extinct for at least 70 years. This is inferred from Dyen's (1962b: 262) account that "[they] belong to dead or nearly dead languages. In 1935, they were not spoken as native languages or were only remembered by old people." Unfortunately, Pazeh (or Pazih), moribund as of 2007, followed the same doom as its last speaker died in 2010 (Blust 2013: 51). Excluding extinct languages, we are now left with only half of the original number (as far back as historical records go).

According to the statistics from the Council of Indigenous Peoples (CIP) of Taiwan, ethnic Austronesian peoples account for less than 2% of the total population of Taiwan (about 23.1 million based on the 2010 national census).¹² However, the actual number of fluent speakers (those who can perform all communicative functions in a language) is certainly much smaller than CIP's figures. Depending on languages, the number of speakers ranges from less than 10 (e.g. Kanakanavu) to over hundreds of thousand (e.g. Amis, the one with the greatest number of speakers). Since Taiwan is socio-economically dominated by ethnic Han Chinese, younger generations of ethnic Austronesian have been shifting to lingua francas such as Mandarin and/or Taiwanese Southern Min (and to a lesser degree Hakka, especially among Saisiyat speakers). Accordingly, like many minority languages of the world, Austronesian languages of Taiwan all share the same destiny of gradually losing their vitality in daily

¹² The official English website of CIP is http://www.apc.gov.tw/portal/index.html?lang=en_US.

communications. The number of existing speakers as well as languages will most likely keep decreasing if no drastic measures are implemented in the near future. Fortunately, some efforts have been made to document and revitalize Austronesian languages of Taiwan (see Rau & Florey 2007). Considering their dire situation, studies on Formosan languages appear urgent.

2.2. Genetic classifications and subgroupings

All Formosan languages belong to the Austronesian family, which, among the world's major language families, ranks first in terms of geographical spread and second with regards to the number of languages, according to Blust's (2013: 759) statistics. It is generally agreed that many Formosan languages are high-order within Austronesian. However, the status of Formosan languages as a genetic subgroup (i.e. the Proto-Formosan hypothesis) as well as their internal subgroupings has long been debated. Over the past four decades, various proposals have been offered and the divergence is mostly due to the different types of evidence cited, ranging from lexical to phonological and morphosyntactic. Without repeating too much of what has been said in the literature, only some prominent proposals are highlighted here in the chronological order they occurred. For a survey of subgrouping proposals in greater depth and width, the reader is referred to Ross (2012).

The earliest empirically-based attempt to classify the Austronesian seems to be Dyen (1962a), of which Dyen (1965) is a more widely circulated version. Based on evidence from lexicostatistics, Dyen identified two subgroups of Formosan languages: (i) Atayalic subfamily, which consists of Atayal and Seediq (or Seedik in his original

spelling); (ii) Central Formosan Hesion (changed into East Formosan Hesion in the 1965 revision), which includes Amis (or Ami in his original spelling), Paiwan, Bunun, and Thao. These two groups were assumed to descend from Proto-Formosan (PFm). About the same time, based on phonological evidence, Haudricourt (1965) proposed a tripartite classification of Austronesian: (i) Western Austronesian, which is spread across areas ranging from Madagascar to the Palau Islands and Botel-Tobago; (ii) Northern Austronesian, which is exclusively restricted to the “Highland languages” of Taiwan (Atayal, Paiwan, Puyuma, Ami, and Bunun); (iii) Eastern Austronesian, which includes all the languages in Melanesia and Polynesia as well as most languages in Micronesia. Thus, both Dyen and Haudricourt embraced the Proto-Formosan (PFm) hypothesis.

The doubt on PFm was later expressed by Ferrell (1969), who put forward a tripartite classification of Formosan languages: (i) Tsouic, which includes Tsou, Kanakanabu, and Saaroa; (ii) Atayalic, which is identical to Dyen’s Atayalic subfamily; (iii) Paiwanic, which comprises all the members of Dyen’s Central Formosan Hesion and many others that are neither Tsouic nor Atayalic. Ferrell’s three-way classification had since been the dominant view until Blust’s (1999) ground-breaking work came about.

Within the thirty years between Ferrell (1969) and Blust (1999), new proposals popped up one after another. An especially interesting one is Starosta (1995), who, based on evidence from shared innovations of morphosyntax, suggested a tree-like model with multiple binary splits. Of the many claims made in this model, three are worth mentioning. First, by subsuming Philippine languages like Tagalog and Ilokano under Formosan languages, he equated PFm with Proto-Austronesian (PAN), which was against the general consensus at that time. Second, by scattering Tsou, Saaroa, and Kanakanabu,

he took the initiative to discredit Ferrell's Tsouic hypothesis, a theme later taken up by H. Chang (2006) and Ross (2009, 2012). Third, he was probably the first one who treated Rukai as the first offshoot of the Austronesian, a view later adopted and further supported by P. Li (2008) and Aldridge (2014).

Rejecting Starosta's tree-like model and all of his predecessors, Blust (1999) proposed a multiple rake-like model with ten first-order subgroups of the Austronesian family, nine of which are Formosan. His list of the nine Formosan subgroups is reproduced in Figure 2.2.

- | |
|--|
| <ol style="list-style-type: none"> 1. Atayalic (self-evident). 2. East Formosan: 1) merger of $*t/C$, 2) merger of $*j/n$, 3) shift of $*q > ?$ <ol style="list-style-type: none"> 2.1. Northern branch (Basay-Trobiawan; Kavalan): 1) merger of $*q/\emptyset$, 2) merger of n/N; 3) irregular change in $*susu > /sisu/$ 'breast' <ol style="list-style-type: none"> 2.1.1. Basay-Trobiawan: 1) merger of $*s/l$ 2.1.2. Kavalan 2.2. Central branch (Amis) 2.3. Southwest branch (Siraya) 3. Puyuma 4. Paiwan 5. Rukai 6. Tsouic 7. Bunun 8. Western Plains: 1) merger of $*n/\eta$, 2) merger of $*s$ and $*t$ in non-final position (extended to final position in Thao), 3) merger of $*w/y/\emptyset$ through truncation of the diphthongs $*-ay$ and $*-aw$ (a change that is still in progress in Thao; for details cf. Tsuchida 1982: 9ff, Blust 1996a) <ol style="list-style-type: none"> 8.1. Central Western Plains <ol style="list-style-type: none"> 8.1.1. Taokas-Babuza 8.1.2. Papora-Hoanya 8.2. Thao 9. Northwest Formosan: 1) shift of $*C > *s$, 2) shift of $*q > ?$ <ol style="list-style-type: none"> 9.1. Saisiyat 9.2. Kulon-Pazeh: 1) merger of $*q/\emptyset$, 2) merger of $*C/S$ except in 'fire' and 'wood/tree', where $*S > /h/$ |
|--|

Figure 2.2: Blust's (1999: 45) subgrouping of Formosan languages

His proposal was grounded in the phonological mergers and shifts (some of which are specified in the figure) that he identified using comparative method. In this model, all the

non-Formosan Austronesian languages belong to one single subgroup called Malayo-Polynesian (MP) languages (not listed in Figure 2.2), all of which share innovations attributable to Proto-MP (PMP). For instance, preconsonantal and word-final *S in PAN disappears in PMP (e.g. PAN **kuSkus* > PMP **kuku* ‘nail (of finger/toe)’; Tryon 2006: 25). Among Austronesian languages of Taiwan, Yami/Tao is the only one belonging to the Malayo-Polynesian subgroup, or more specifically the Batanic group within it, on a par with those spoken on the Batanes Islands, north of the Philippines. This means at least one language from each of the ten proposed first-order subgroups of the Austronesian family can be found in Taiwan. Blust (2013: 31) compares this situation of high diversity to “finding representatives of every branch of the Indo-European language family within the borders of the Netherlands.” The nine groups of Formosan languages, however, do not share any innovations not found in PMP, indicating that the Formosan languages do not share a common ancestor aside from PAN. In other words, extra-Formosan is synonymous with Malayo-Polynesian, and conversely non-Malayo-Polynesian is equivalent to Formosan. This ten-way classification of the Austronesian, which explicitly rejects the PFM hypothesis, has ever since become the most widely adopted view among Austronesianists (Ross 2002: 17).

On the other hand, alternative proposals aside from Blust (1999) that opt for more nesting at the higher level are also quite common, including Sagart (2004), P. Li (2008), and Ross (2009, 2012). Among them, the last proposal, which is built on top of Blust (1999), crucially hinges on nominalization and will be reviewed in due course.

To sum up, while there is little doubt regarding the status of Malayo-Polynesian languages as a genetic subgroup, the literature shows divided opinions on whether

Formosan languages also have such a status. Dyen (1962a, 1965), Haudricourt (1965), and Starosta (1995) embraced the PFM hypothesis whereas Ferrell (1969), Blust (1999), and Ross (2009) either doubted or rejected it altogether. Moreover, there is a general consensus among Austronesianists that Formosan languages, despite their small number, are most diverse of the Austronesian family (see P. Li 2008a), a fact that makes typological works based on them all the more significant.

2.3. Grammatical features

Formosan languages are predominantly predicate-initial except for those that are highly Sinicized such as Saisiyat and Thao. The majority of them share various grammatical features with those of the Philippines, so they are often collectively referred to as the Philippine-type languages, which are “found in a geographically delimitable area” covering (from north to south) Taiwan, the Philippines, parts of Kalimantan, and southern Sulawesi (Wolff 2002: 438-439). Himmelmann (2005: 113) characterizes the Philippine-type languages as those that have: “(i) at least two formally and semantically different undergoer voices; (ii) at least one non-local phrase marking clitic for nominal expressions; (iii) pronominal second position clitics.” Because they are particularly relevant to nominalization, these three grammatical properties are introduced in this section under the headings of Focus morphology (§2.3.1), nominal relation markers (§2.3.2), and person-form clitics (§2.3.3) respectively.

Other grammatical aspects of Formosan and Philippine languages can be found in typologically oriented descriptions such as Starosta (1988a), Reid & Liao (2004), Zeitoun (2004), Ross & Teng (2005a), and Himmelmann (2005a).

2.3.1. Focus morphology

One distinctive feature of Philippine-type languages is a small set of affixes on the verb indicating “the semantic role of one of the participants involved in the state of affairs denoted by the predicate” (Himmelmann 2004: 1479). The selected participant is typically specified by a dedicated relation marker for free nominals (see §2.3.2) or/and a bound person form (§2.3.3) on the verb so that its semantic role varies as the verb form changes (to be illustrated below). Perhaps the most prominent example of Philippine-type languages is Tagalog, where verb forms marked by <um>, -in, -an-, and i- were respectively referred to as “active voice,” “direct passive,” “local passive,” and “instrumental passive” in the early years of Philippine linguistics (Blake 1936; Bloomfield 1917; Wolff 1973). Given the one-active and three-passive analysis, the selected NP was then called “subject.”

In subsequent years, the active-passive analysis fell into disfavor, but “[w]hat to call this system of verbal affixation and the selected nominal to which it corresponds has remained controversial” (Quakenbush 2005: 8). A great number of terminologies have been proposed to reflect what each researcher thinks should be the best way to analyze the verbal system in Philippine-type languages. Alternative proposals for “voice” include “focus,” “case,” “trigger,” “recentralization,” and “topicalisation” while the selected “subject” NP also bears such names as “focus,” “pivot,” or “topic.” A more detailed chronology of terminologies for Philippine-type verbal affixes is documented by Blust (2002), included in a volume (Wouk & Ross 2002) dedicated to historical and typological studies of these affixes.

Paradoxically, terminological differences do not really prevent Austronesianists from understanding each other (or linguistic data) since the terms proposed are by and large comparable. Nevertheless, the terms one chooses to use do reflect the theoretical framework in which one works as well as the perspective from which one looks into the Philippine-type verbal system and argument structure. Disagreements among proposals, for the most part, stem from how to reconcile the fact that the verbal affixation in Philippine-type languages bears some functional resemblances to the voice system in Indo-European languages with respect to some properties but not others (Himmelman 2002), and at the same time the fact that properties of the subject NP in Indo-European languages are often split over two separate NPs in Philippine-type languages (Schachter 1976). These two issues as well as others will be discussed later in §2.4, where justifications for terminologies and glossing principles adopted in this study are presented.

Suffice it to say that for the purpose of this study I use Focus to refer to affixes on the verb that delineate the semantic role of a selected argument. Focus, when capitalized, is intended as a proper name for the Philippine-type voice system, to be distinguished from *focus* as is commonly understood in the pragmatics literature. Accordingly, verb forms corresponding to “active voice,” “direct passive,” “local passive,” and “instrumental passive” as those in Tagalog are respectively termed Actor Focus (AF), Patient Focus (PF), Locative Focus (LF), and Conveyance Focus (CV), following the convenient labels in Himmelman (2005a) (except that he opts for “voice” rather than “Focus”).¹³ Moreover, due to the syntactic asymmetry between AF-constructions on the one hand and PF/LF/CF-constructions on the other, the latter three Focus types are often

¹³ On a related note, Blust (2013: 437) points out one advantage of choosing Focus over voice since Focus “uniquely identifies languages that otherwise must be called by the longer and more cumbersome term ‘Philippine-type languages’, while ‘voice’ obviously does not.”

collectively referred to as non-Actor Focus (NAF) (more on this in §2.4.1). When a language formally and semantically distinguishes at least two NAF types, it then meets Himmelmann’s (2005: 113) first criterion of being a Philippine-type language. On the other hand, I adopt the terms used in Shibatani (1988, 1991) for the two NPs that have been shown to demonstrate subject properties in many studies. The selected argument whose semantic role is somewhat specified on the verb is termed *Topic* (glossed TOP) while the other subject-like argument is referred to as *Actor* (glossed ACT), which is typically most agentive and may or may not double as the Topic as well.

The Focus morphology and its interaction with the notion of Topic and Actor NPs are best illustrated with examples having the same verb stem but marked by different Focus affixes. For instance, the Tagalog examples in (1) “are essentially equivalent but differ in the interpretation of definiteness and effectedness of the arguments” (Kaufman 2009b: 3).

(1) Tagalog (Kaufman 2009b: 3)¹⁴

- a. **k<um>áin** ng=dagà sa=pinggan pára sa=áso ang=púsa
 <AF>eat UND=rat OBL=plate for OBL=dog TOP=cat
 ‘The cat ate a rat on the plate for the dog.’
- b. **k<in>áin** ng=púsa ang=dagà sa=pinggan pára sa=áso
 <PF.RLS>eat ACT=cat TOP=rat OBL=plate for OBL=dog
- c. **k<in>áin-an** ng=púsa ng=dagà ang=pinggan pára sa=áso
 <RLS>eat-LF ACT=cat UND=rat TOP=plate for OBL=dog
- d. **i-k<in>áin** ng=púsa ng=dagà sa=pinggan ang=áso
 CF-<RLS>eat ACT=cat UND=rat OBL=plate TOP=dog

¹⁴ Where necessary, Topic NPs are underlined throughout this study. In the source reference, PF is glossed as a zero morpheme after the verb root. Here, however, the infix <in>, when used in PF constructions, is taken to be a portmanteau for both perfectivity/realis and PF marking, following Zeitoun et al. (1996: 29) and Blust (1998a: 347). Non-realistic PF is marked by *-in*. See Himmelmann (2005: 363) for a complete paradigm of verb forms in Tagalog, the very analysis on which the glossing for TAM is based in this and subsequent Tagalog examples. Finally, the morpheme *ng* is transcribed as *nang* [naŋ] in the cited work based on its pronunciation, but has been changed here based on the conventional orthography.

All these examples involve four event participants, each of which can be selected as the *ang*-marked Topic NP, whose semantic role varies from agent to patient, location, and beneficiary as indicated by Focus affixes. By contrast, the Actor NP (i.e. the cat) is invariably the most agentive one among the four and is marked by *ang* in (a), where it is the Topic-cum-Actor, but by *ng* in (b)~(d) instead, where it is the non-Topic Actor. By the same token, affixes assuming similar forms and functions are also attested in Paiwan, a Formosan language, as illustrated in (2).

(2) Northern (Sandimen) Paiwan¹⁵

- a. **na-tekelj** ti=palang ta=vava
 PFV-<AF>drink TOP=P. UND=wine
 ‘Palang drank wine.’ (A. Chang 2000: 97)
- b. **tekelj-{in/en}** azua=a vava ni=palang
 drink-PF TOP.DIST=LIG wine ACT=P.
 ‘Palang drank that wine.’ (S. Wang 2005: 24)
- c. **t<in>ekelj** azua vava ni=palang
 <PFV.PF>drink TOP.DIST wine ACT=P.
 ‘Palang drank that wine.’ (A. Chang 2000: 98)
- d. **k<in>eljem-an** ni=palang tjay=kalau aicu=a gaku
 <PFV>beat-LF ACT=P. UND=K. TOP.PROX=LIG school
 ‘Palang beat Kalau at this school.’ (A. Chang 2000: 99)
- e. **s<in>i-tekelj** ni=palang tua=vava ti=avakaw
 CF-<PFV>drink ACT=P. UND=wine TOP=A.
 ‘Palang drank wine for Avakaw.’ (S. Wang 2005: 25)

Aside from the set of cognate NAF affixes as illustrated in Tagalog and Paiwan, there is another set, which has corresponding Focus functions just like the first set but is used in a non-indicative mood (e.g. imperative or optative) or in syntactically dependent contexts (i.e. after various preverbal elements such as the negator, modals, or discourse

¹⁵ As in Tagalog, non-perfective PF in Paiwan is marked by *-in* (or *-en*) whereas perfective PF by *<in>*, glossed as a portmanteaux for both perfectivity and PF marking (see Footnote 14). The perfective LF form of the root *tekelj* ‘drink’ would be *t<in>ekelj-an*, but it just so happens that no available example with such a word form has been found in A. Chang (2000). Nevertheless, the perfective LF form of the root *keljem* ‘beat’ suffices to illustrate the same point.

connectors) in “almost” all languages that have it (Ross 2002, 2009). For instance, where Paiwan has *-in/-en* or *<in>* for PF, *-an* for LF, and *si-* for CF in affirmative indicatives, as in (2), it uses *-aw* for PF, *-ay* for LF, and *-an* for CF in affirmative non-indicatives that express the speaker’s volition, wish, desire, exhortation, or expectation (variously referred to as subjunctive, projective, hortative, or optative in the literature; the last term is adopted for the purpose of glossing), as in (3).¹⁶

(3) Northern (Sandimen) Paiwan (A. Chang 2006: 188)

- a. **tja=kan-aw=anan=[a** **tja=cengel]**
 1INCL.ACT=eat-PF.OPT=first=TOP 1INCL.GEN=meal.box
 ‘Let’s eat our meal box first!’
- b. **ku=vecik-ay=[a** **tigami]** tanusun
 1SG.ACT=write-LF.OPT=TOP letter 2SG.OBL
 ‘I would like to write you the letter!’
- c. **ku=vecik-an=emun**
 1SG.ACT=write-CF.OPT=2SG.TOP
 ‘I would like to write (something) for you!’

Descriptively speaking, cognate NAF affixes like those in (1) and (2) can be called Mixed NAF affixes because they consist of an assortment of prefixes, infixes, and suffixes whereas those in (3) can be referred to as Suffixal NAF affixes because they are all suffixes. Functional labels instead of structural ones could have been adopted were it not for some exceptions (see §4.4.1), for which the “almost” caveat from above is reserved. For instance, Puyuma uses Suffixal NAF affixes rather than Mixed ones for affirmative indicatives, as illustrated in (4).

¹⁶ In Xrakovskij’s (2001) model, strictly optative, hortative, and imperative meanings are all *hortative* in the broadest sense. They only differ in terms of which speech-act participants are the performer of a prescribed action.

(4) Nanwang Puyuma (S. Teng 2008: 147)

- a. **tu=trakaw-aw** na paisu kan isaw
 3.ACT=steal-PF TOP.DEF money ACT.SG I.
 ‘Isaw stole the money.’
- b. **tu=trakaw-ay=ku** dra paisu kan isaw
 3.ACT=steal-LF=1SG.TOP UND.INDF money ACT.SG I.
 ‘Isaw stole money from me.’
- c. **tu=trakaw-anay** i tina=taw dra paisu
 3.ACT=steal-CF TOP.SG mother=3.GEN UND.INDF money
 ‘{He/She/They} stole money for {his/her/their mother}.’

The existence of Mixed and Suffixal Focus affixes has been given diachronic explanations, a topic to be covered later in §3.1.2 and §4.4.

In short, Focus morphology is a grammatical way of organizing the argument structure, which is regulated through the interaction between Focus affixes on the verb and the Topic argument. All Formosan languages have the Focus system as demonstrated above to varying degrees except for Rukai (see §3.4), where the basic verbal system is drawn between active and passive rather than between AF and NAF (P. Li 1973; Zeitoun 1995, 2007). Maximally four formally distinct Focus types involving two sets of cognate affixes are found in many Philippine-type languages, but the semantic range of the Topic argument in construction with a given Focus form varies across languages. This issue is deferred to §2.4.1 since it is directly related to how Focus affixes are glossed in this study.

2.3.2. Nominal relation markers

As the other side of the coin, intertwined with Focus morphology is the nominal phrase marking system, which consists of a small set of function words (typically monosyllabic) that predominantly occur at the left periphery of a nominal phrase (NP),

such as *ang* and *ng* in Tagalog (see (1) above) as well as *ti* and *ni* in Paiwan (see (2) above). They are collectively referred to as nominal relation markers in this study, which encode a variety of syntactic, semantic, pragmatic, or/and discourse information regarding the referent of the NPs that they mark. Here “relation” is intended as shorthand for “grammatical relation,” which broadly covers syntactic functions and semantic roles (Farrell 2005: 8).

According to Reid’s (2002: 296-297) survey, more than thirty terms have been invented in the literature to designate these phonologically short function words, including “article,” “prepositional particles,” “preposition,” “determiner,” “noun-marking particle,” “noun-phrase marker,” “construction identifier,” “ligature,” “case-marking particles,” or simply “case markers.”¹⁷ The last one seems to be the most commonly adopted term in the literature (Starosta 1993; P. Li 1997b; L. Huang et al. 1998; Ross 2006), although there has been some concern questioning whether it is legitimate to call them case markers in the first place (Shibatani 1988; Reid 2002; M. Chang 2004; M. Wang 2005; W. Shih 2012). The wealth of terms reflects the recalcitrance of these markers for pigeon-holed categorization, which is largely due to the fact that they encode much more information than prototypical case markers do. Glossing principles for nominal relation markers are discussed in §2.4.2.

Depending on languages, nominal relation markers potentially indicate the specificity, definiteness, visibility, evidentiality, semantic roles, or/and syntactic functions of the NPs that they mark. One extreme example is found in Tsou, where the relevant markers encode not only grammatical functions and deictic distance, but also the

¹⁷ A much longer list is provided by Blust (2015: 437-439), including the references where various terms are adopted.

epistemological and perceptual status of the referent of the NPs they collocate with (H. Chang 2011; see the references therein for other analyses). For instance, all the monosyllabic markers in (5) indicate the same Topic relation for an NP, but vary with regards to different pragmatic and discourse factors concerning the referent of that NP.

(5) Tsou (H. Chang 2011: 120)¹⁸

- a. **cuma=[na i=ko ait-i]**
 what=TOP NAF.RLS=2SG.ACT see-LF
 ‘What are you reading?’ (invisible and unwitnessed)
- b. **cuma=[’o i=ko ait-i]**
 what=TOP NAF.RLS=2SG.ACT see-LF
 ‘What are you reading?’ (invisible but witnessed)
- c. **cuma=[co i=ko ait-i]**
 what=TOP NAF.RLS=2SG.ACT see-LF
 ‘What are you reading?’ (invisible but perceptible)
- d. **cuma=[si i=ko ait-i]**
 what=TOP NAF.RLS=2SG.ACT see-LF
 ‘What are you reading?’ (visible and medial)
- e. **cuma=[’e i=ko ait-i]**
 what=TOP NAF.RLS=2SG.ACT see-LF
 ‘What are you reading?’ (visible and proximal)

The semantics of these markers is not readily integrated into English free translations, so only the Topic relation they indicate are shown in glossing. Considering this arbitrary practice, chances are that markers glossed simply as TOP or ACT in fact express more information than is indicated by the labels.

In addition, another piece of information often not shown in the glosses for nominal relation markers is the contrast between personal and common nouns (Reid & Liao 2004: 469). In Paiwan, for instance, TOP and ACT are respectively expressed by *ti* and *ni* if they mark singular personal nouns, but by *a* and *nua/na* instead if they mark common nouns, as contrasted in (6).

¹⁸ Tsou has nominal relation markers that are unaccented and phonologically attach to the word form preceding them. See §2.3.2 for more justifications.

(6) Northern Paiwan (A. Chang 2006: 113-114)

- a. **s<in>angutj** ti=kalalu ni=palang
 <PFV.PF>kiss TOP=K. ACT=P.
 ‘Palang kissed Kalalu.’
- b. **k<in>ac=[a** vatu] n(u)a=’atjuvi
 <PFV.PF>bite=TOP dog ACT=snake
 ‘A snake bit the dog.’

However, much like the distinction between alienable and inalienable nouns (Lichtenberk 2005), the boundary between personal and common nouns varies from one language to another. In languages where such a distinction is made formally, personal nouns include at least personal names, and may additionally cover person forms or/and some kinship terms. To the extent that the categorization between personal and common nouns is lexically determined, such a contrast makes nominal relation markers functionally come close to noun class markers (see H. Chang et al. 1998).

Aside from their functional intricacy, nominal relation markers often demonstrate phonological dependency. According to Reid (1978: 34), they “tend to be single syllable, morphosyntactically free, but phonologically bound, cliticized to the preceding or following stressed word,” hence the term “noun-marking particle” in the literature. In other words, nominal relation markers in many languages are phrasal clitics that have scopes over a nominal phrase whose grammatical relations they mark. Despite the fact that nominal relation markers are phonologically deficient, they are commonly transcribed in the literature as if they were prosodically independent word forms. In this study, I transcribe them as clitics (by using the equal sign) only when there is clear and consistent evidence from prosody indicating their clitichood. Otherwise, they will be presented as independent word forms without equal signs following the common practice. If a language has at least one nominal relation clitic that marks something other than

locative roles, it then meets Himmelmann's (2005: 113) second criterion of being a Philippine-type language.

The positioning of nominal relation clitics can be stated elegantly using Klavans' (1985) three binary parameters, two of which are structural notions and the other a phonological one. In terms of Parameter 1 Dominance, nominal relation clitics attach to the initial (as opposed to final) constituent of an NP, and in terms of Parameter 2 Precedence, they occur before (as opposed after) the selected constituent (i.e. the structural host). In other words, they occur at the left edge of an NP. However, with respect to Parameter 3 Phonological Liaison, these clitics attach to their preceding (hence enclitics) or following (hence proclitics) phonological word (i.e. the phonological host), depending on languages. When the structural host and the phonological host is one and the same constituent, then we see straightforward examples of phrasal proclitics, such as *ang* and *ng* in Tagalog (see (1) above) as well as *ti* and *ni* in Paiwan (see (2) above). However, the structural host and phonological host of a clitic do not always converge, and clitics of this nature are dubbed "clitics with dual citizenship" by Klavans (1985: 104) and termed *ditropic* by Embick & Noyer (1999: 291). For instance, L. Li (2010) argues that ACT/UND and TOP in Isbukun Bunun are both ditropic enclitics, meaning clitics that syntactically mark the grammatical relation of a following NP but phonologically attach to whatever constituent that appears right before them, as illustrated in (7).

- (7) Isbukun Bunun (L. Li 2010: 9, 78, 80)
- a. kaun-un=[**mas** utung]=[**a** bunbun]
 eat-PF=ACT monkey=TOP banana
 ‘The monkey eats/ate a banana.’
- b. mudah-un=in=ta=[**a** tama]
 beat-PF=already=1INCL.ACT=TOP father
 ‘We already beat Father.’
- c. mudah-un=in=ta=[**mas** tama]
 beat-PF=already=1INCL.TOP=ACT father
 ‘Father already beat us.’

These examples show that ACT *mas* and TOP *a* always attach to the preceding word form regardless of whether that word form additionally hosts other clitics. In addition, the phonological liaison between ACT/UND *mas* and its preceding word form is further illustrated by its alternative short form, which is *is* if the phonological host ends with a consonant (e.g. *saipuk=is hangvang* [AF.raise=UND water.buffalo] ‘to raise a water buffalo’) or simply a consonantal segment *s* instead if the host ends with a vowel (e.g. *siza=s tanga* [AF.take=UND hoe] ‘take a hoe’; both examples taken from R. He et al. 1986: 107).¹⁹

Finally, research on ditropic clitics in Formosan languages is almost an unexplored area (except L. Li 2010), and it seems that ditropic enclitics are more common than ditropic proclitics although more research is needed to confirm this. Here I only make a brief note on ditropic enclitics, which can be identified based on three criteria: (i) They form a phonological word with their preceding word form; (ii) they mark the grammatical relations (syntactic or/and semantic) of their following NP, the presence of which is obligatory; and (iii) there is often an audible pause between ditropic enclitics and the first constituent of the following NP. By these criteria, at least two other

¹⁹ Incidentally, Guina-ang Bontok from northern Philippines also has a ditropic enclitic (*a*)*s*, which marks locative and temporal roles (Reid 1970: 23; cited in Himmelmann 2005: 132).

Bunun languages, Takituduh and Takibakha (both Northern Bunun), have nominal relation markers that are ditropic enclitics. Illustrated in (8) is Takituduh, where TOP for personal nouns is either *at* or *kat* and UND for personal nouns either *it* or *t*.

(8) Takituduh Bunun (Y. Su 2008: 17-18)

- a. tu<sa~>sauc=[**at** tama]
 <IPFV~>AF.sing=TOP father
 'Father is singing.'
- b. mas'i=[**kat** bukut]
 AF.cough=TOP B.
 'Bukut coughs.'
- c. ma-ludaq=ak=[**it** ulang]
 AF-beat=1SG.TOP=UND U.
 'I beat Ulang.'
- d. silulu=[**t** ulang]=[**at** talum]
 AF.pull=UND U.=TOP T.
 'Talum pulls Ulang.'

In both cases, the choice between the two allomorphs is conditioned by whether the word form preceding them ends with a consonant or vowel.

On the other hand, when no morphophonemic changes or phonological liaisons are involved, it becomes less straightforward to identify ditropic enclitics. One such language is Tsou, where the stress of content words systemically falls on the penultimate syllable while most grammatical markers, including nominal relation markers, are unaccented. Nevertheless, both phonological and morphosyntactic evidence suggests that nominal relation markers in Tsou are also ditropic enclitics like those in the three Bunun languages mentioned above. In terms of prosody, there are typically noticeable pauses between nominal relation markers and the first constituent of the NPs that they precede and form a syntactic unit with, as shown by OBL *to* and TOP *'e* in an excerpt from a naturally occurring narrative of the Frog story (Mayer 1969) in (9).

(9) Tsou (NTU Corpus|Frog3: IU.06-09)²⁰

- a. ... (1.0) o=he=la X ,_
NAF.RLS=3PL.ACT=DSTT
- b. ... (1.4) poekotv-a=[to] ,\
lock.up-PF=OBL
- c. ... (1.0) pania]=[e] ,_
bottle=TOP
- d. ... fo'kunge]=ho i=he .. aʉt'ʉc-a .\
frog=CONJ NAF.RLS=3PL.ACT raise-PF
'They locked up the frog in a bottle and raised (it there).'

Moreover, the attachment of nominal relation markers to their phonological host does not change the stress of that host. All these content words *poekotva*, *pania*, and *fo'kunge* have their stress on the penultimate syllable, with or without a following unaccented marker. On the other hand, the morphosyntactic evidence comes from the relative order between the GEN marker (that indicates possessive and part-whole relationships) and person-form possessor indexes. For example, the Tsou people usually refer to their older siblings as *ohaeva* and younger ones as *ohaesa* regardless of genders, both of which are directly marked by possessor indexes if possessed (e.g. *ohaeva*= 'u 'my older sibling'). In cases where the gender of a sibling needs to be specified, the GEN marker, together with words for men (i.e. *hahocng#*) and women (i.e. *mamespingi*), is added after either kinship term, as in *ohaesa*=[no *mamespingi*] 'younger sister' (lit. 'younger sibling of women'). Crucially, when the whole phrase is possessed, the possessor index attaches directly to

²⁰ In the NTU Corpus, discourse information is transcribed following Du Bois (1993), where the equal sign symbol indicates lengthening instead of clitic hood as in the Leipzig Glossing Rules. Here the equal sign is meant to mark clitic hood, not lengthening. To assess the prosody of this excerpt, visit the following link for access to the audio recording: http://corpus.linguistics.ntu.edu.tw/read.php?lang=tsou&art=frog_3. Both Zeitoun (2005) and H. Huang (2010) distinguish two *la* morphemes in Tsou. While they both recognize the auxiliary *la*, which attracts person-form indexes and indicates habitual aspect, the former analyzes the other *la* as an experiential marker, "which indicates that an event has been experienced/happened in the past" (Zeitoun 2005: 282), whereas the latter, based on discourse data, treats the other *la* as "an actualized particle... indicating the speaker's recognition that the event occurred in the past, or the speaker's inference that the event is expected to occur in the future." (H. Huang 2010: 157-158) The latter analysis is followed here, and the non-auxiliary *la* is assumed to be an enclitic (judged by prosody) and glossed here as DSTT for distant time in the past or future.

neither of the two content words in that phrase, but to the GEN marker instead, as in *ohaesa*=[*no*='u *mamespingi*] 'my younger sister'.²¹ It would be difficult to account for the order between the GEN *no* and 1SG.GEN='u, both of which are unaccented, without analyzing them both as position-sensitive clitics. In the current clitic analysis, the GEN *no* is a ditropic enclitic, which forms a phonological word with a host that precedes it, whereas possessor indexes like 1SG.GEN='u are Wackernagel enclitics, which attach after a phonological word.²² The two positioning principles then give rise to the interspersing order as observed.

In a nutshell, nominal relation markers are functionally loaded with syntactic, semantic, pragmatic, or/and discourse information regarding the referent of an NP, and are more often than not phonologically deficient, thus clinging onto their preceding or following word form to satisfy the prosodic requirement in each language. Regardless of whether they are proclitics or ditropic enclitics, they predominantly occur at the left periphery of an NP, thus called peripheral clitics by Himmelmann (2005: 131), in contrast to second-position clitics, to which bound person forms in most Philippine-type languages belong and to which I turn in the following section.

2.3.3. Person-form clitics

Most Formosan languages have free and bound person forms, but in this section I primarily discuss bound ones because of their special morphosyntax. Readers are referred

²¹ Another similar example is found in H. Huang (2010: 85). Compare *ongko*=[*no cou*] 'name in Tsou' with *ongko*=[*no=su cou*] 'your name in Tsou', where =*su* '2SG.GEN' attaches directly to neither of the two nouns *ongko* 'name' and *cou* 'Tsou'.

²² Zeitoun (2005) tentatively treats person-form indexes in Tsou as affixes, which is also the most common practice in the literature, except for G. Lin (2010), who adopts the clitic analysis.

to the literature (P. Li 1997a; L. Huang et al. 1999; Ross 2015b) for more encompassing studies of person forms (both bound and free) in Formosan languages.²³

Bound person forms in most Formosan languages are also clitics like nominal relation markers, but the two tend to differ in one crucial aspect. In Himmelmann's (2005: 131) term, nominal relation markers are peripheral clitics while bound person forms are typically second-position, or Wackernagel, clitics. And the existence of person-form Wackernagel clitics satisfies Himmelmann's (2005: 113) third criterion for a Philippine-type language. If stated in Anderson's (2005: 23, 31) typology of clitics, nominal relation markers are phonological clitics, which are phonologically deficient, and bound person forms are morphosyntactic ones, which demonstrate morphosyntax motivated by a set of principles distinct from those that govern free forms in a language.

There have been some debates over whether bound person forms in Formosan and Philippine languages should be analyzed as person agreement markers or pronouns (e.g. H. Chang 1997: 117-124 and P. Li 2006 on Kavalan; Zeitoun 1997 on Mantauran Rukai; H. Chang 1999 on Seediq; Ochiai 2009 on Tgdaya Seediq; H. Liao 2005 on Central Cagayan Agta). Haspelmath (2013) proposes to solve long-held controversy of this nature by distinguishing three types of *indexes* (his term for bound person forms, including affixes and clitics alike), depending on the constraints of their co-occurrence with *conominals* (his term for free nominals coreferential with indexes). If conominals are obligatory, then bound person forms are *gramm-indexes*, equivalent to traditional agreement markers. If conominals are prohibited, on the other hand, then bound person

²³ I follow Haspelmath (2013) in using "person forms" as a cover term for what is traditionally called person agreement markers and pronouns, be it bound or free. Free person forms are "pro-nouns" in the strict sense because they function as NP substitutes. Bound person forms, by contrast, do not have this function, and are thus called indexes for short in his proposal.

forms are *pro-indexes*, comparable to traditional pronouns. Additionally, if bound person forms are allowed to optionally co-occur with conominals, they are said to be *cross-indexes*, a third category beyond the traditional dichotomy between person agreement markers and pronouns. Recognizing cross-indexes not only avoids the attempt to squeeze bound person forms into either of the two traditional categories, but also respects the typological fact that cross-indexes are the most common type of the three across languages. In what follows, I show how this framework of argument indexing helps to do away with the common assumption that person agreement markers are affixes whereas bound pronouns are clitics. The demonstration also suffices to illustrate the special morphosyntax of person-form indexes.

For instance, both Kavalan and Tgdaya Seediq have two morphological paradigms of person-form indexes (called index-sets by Haspelmath 2013), with one set indexing the Actor and the other the Topic. Actor and Topic indexes in Kavalan and Tgdaya Seediq are illustrated in (10) and (11) respectively.

(10) Kavalan

- a. **pukun-an=na=iku**
beat-LF=3.ACT=1SG.TOP
'{She/He/They} beat me.' (A. Lee 1997: 45)
- b. **mai pukun-an=na=iku**
NEG beat-LF=3.ACT=1SG.TOP
'{She/He/They} did not beat me.' (D. Yen 2012: 108)
- c. **mai=iku** **pukun-an=na**
NEG=1SG.TOP beat-LF=3.ACT
'{She/He/They} did not beat me.' (D. Yen 2012: 108)
- d. **pukun-an=na=ti=iku** **ni=buya**
beat-LF=3.ACT=already=1SG.TOP ACT=B.
'Buya already beat me.' (Y. Yeh 2005: 32)

(11) Tgdaya Seediq (H. Chang 1997: 98, 190, 99)

- a. **bube-un=ku=na**
 beat-PF=1SG.TOP=3SG.ACT
 ‘{She/He} will beat me.’
- b. **wada=ku=na** **bube-un** **na=pawan**
 PST=1SG.TOP=3SG.ACT beat-PF ACT=P.
 ‘Pawan beat me.’
- c. **wada=ku=na** **bube-un** **na=pawan** **ka=yaku**
 PST=1SG.TOP=3SG.ACT beat-PF ACT=P. TOP=1SG
 ‘Pawan beat me.’

In both languages, Actor and Topic indexes may cluster with the verb in NAF-constructions, and the internal order is Actor before Topic in Kavalan but Topic before Actor instead in Tgdaya Seediq (except for situations where portmanteau morphemes of ACT and TOP indexes are required; see Holmer & Billings 2014). When there is a preverbal auxiliary (e.g. the negator or tense/aspect/mood markers), in Kavalan only Topic indexes may attach after it while Actor indexes have to stay with the lexical verb. By contrast, both Actor and Topic indexes attach to the preverbal auxiliary in Tgdaya Seediq. H. Chang (1997) argued that Topic indexes in Kavalan are pronominal enclitics while Actor indexes are agreement suffixes considering the mobility and promiscuity of the former. On the other hand, he analyzed both Actor and Topic indexes in Tgdaya Seediq as agreement suffixes despite their mobility and promiscuity. One of his major reasons was that Actor indexes in Kavalan, as well as Actor and Topic indexes in Tgdaya Seediq (i.e. those claimed to be agreement suffixes), co-occur with conominals. While Kavalan only allows non-SAP conominals, Tgdaya Seediq permits both SAP and non-SAP ones, as respectively shown in (10)d and (11)c. However, the co-occurrence constraints on conominals can be independently motivated and have not much to do with whether indexes are affixes or clitics, which should be judged based on factors such as

mobility and promiscuity, among others. In fact, D. Yen (2012) and Holmer & Billings (2014) have demonstrated that the two index-sets in Kavalan and Tgdaya Seediq respectively are best analyzed as clitics to better account for their external and internal ordering. As morphosyntactic clitics, these index-sets function well without conominals, and simply vary with respect to whether or not they may co-occur with conominals. In Haspelmath's (2013) terminology, Topic and SAP Actor indexes in Kavalan are pro-indexes whereas its non-SAP Actor indexes are cross-indexes. Since there are no Topic indexes for non-SAPs in Kavalan, the generalization is that this language has pro-indexes for SAPs but cross-indexes for non-SAPs. By contrast, all index-sets in Tgdaya Seediq, for SAPs and non-SAPs alike, are cross-indexes. These statements are all it takes to capture the differences between indexes in the two languages. There is thus no need to make a hard decision between agreement markers and pronouns. Meanwhile, cliticness of these indexes can still be demonstrated by well-established principles.

Furthermore, Formosan languages predominantly use the same index-sets to index both the non-Topic Actor and the possessor (POR; see §2.4.2 for exceptions). In most cases, POR indexes attach immediately after or before the possessum noun (PUM), depending on languages. However, there are syntactic environments where POR indexes have to attach to something else, thus giving rise to what Nichols & Bickel (2013) calls *floating marking*, “where the marker is positioned with respect not to the head or the dependent of the phrase but relative to the phrase boundaries.” I discuss two such environments below, both involving the negator as a host of POR indexes.

Wackernagel clitics that index the POR may constantly stay with the PUM noun (i.e. head-adjacent clitics) or attach to other prenominal hosts if there is any available (i.e. phrasal clitics), as respectively illustrated by Kavalan and Pngawan Atayal:²⁴

(12) Kavalan (Supplementary Materials; S.P. 5-3)

- a. **patudan=su** **tiyau=ni**
 teacher=2SG.GEN DIST.TOP=QP
 ‘Is that person your teacher?’
- b. **usa, usa patudan=ku** **aizipna**
 NEG NEG teacher=1SG.GEN 3SG.TOP
 ‘No, {she/he} is not my teacher.’

(13) Pngawan Atayal (Supplementary Materials; S.P. 5-3)²⁵

- a. **sinsi=su** **xiya**
 teacher=2SG.GEN 3SG.TOP
 ‘Is {she/he} your teacher?’
- b. **a_rat, a_rat=mu** **sinsi** **xiya**
 NEG NEG=1SG.GEN teacher 3SG.TOP
 ‘No, {she/he} is not my teacher.’

The negated nominal predicate in Pngawan Atayal, as in (13)b, then shows the first syntactic environment for floating marking.

The second environment is where members from two index-sets attach to one single noun serving as the nominal predicate. The set closer to the noun indexes the Topic argument while the one further away from it indexes the POR of that noun. As a result,

²⁴ All the four examples have also been confirmed by my Kavalan and Pngawan Atayal consultants.

²⁵ Word-final glottal stops in Atayal dialects are often transcribed in the literature. However, their transcription is often not consistent, both within and across studies (see Y. Tang 2002). In this study, word-final glottal stops in Atayal dialects are not transcribed all together for the following three reasons. First, all lexical words end with either a prolonged vowel or a consonant, including the glottal stop, except for unaccented functional words or some loan words. Second, Y. Tang (2002) found that there is no significant difference in length between words claimed to end with a short vowel without the glottal and those with the same short vowel but with the glottal stop, both of which, however, are significantly shorter than those ending with a long vowel. Third, Lambert (1999) argued that the postvocalic glottal stop at the word-final position is epenthetic in nature, and that its phonetic presence is simply to satisfy the iambic foot. In other words, word-final glottal stops are by and large predictable, and are therefore not represented in phonemic transcriptions. On the other hand, prosodically long vowels [i:] and [u:] will be transcribed as <iy> and <uw> respectively (see Neban 2001; cf. Egerod 1965 and P. Li 1980 for different analyses of these sequences on the phonemic level).

POR indexes are not directly attached to the PUM noun. This seems to be a common feature of Atayalic languages (including Atayal and Seediq), where POR indexes also share the same forms with non-Topic Actor indexes. Interestingly, however, the position of POR indexes does not always agree with that of non-Topic Actor indexes despite their isomorphism. For instance, in Truku and Tgdaya Seediq both non-Actor Topic and non-Topic Actor indexes (in that order) attach to the verb when it is not preceded by any potential host but to the preverbal host instead if that is available:

(14) Truku Seediq (Tsukida 2009: 323, 447)²⁶

- a. **se-pehapuy=ku=na** qesurux ka=kumu
 CF-cook=1SG.TOP=3SG.ACT fish.UND ACT=K.
 ‘Kumu cooked fish for me.’
- b. 'ini=**su=na** baq-i hug
 NEG=2SG.TOP=3SG.ACT give-LF.DEP QP
 ‘Did’nt {she/he} give you (one)?’

(15) Tgdaya Seediq

- a. **bube-un=su=mu**
 beat-PF=2SG.TOP=1SG.ACT
 ‘I will beat you.’ (H. Chang 1997: 108)
- b. ini=**su=mu** bbe-i
 NEG=2SG.TOP=1SG.ACT beat-LF.DEP
 ‘I did not beat you.’ (Ochiai 2009: 36)

The same distribution obtains when Topic and POR indexes (in that order) attach to a single noun serving as the nominal predicate. However, if there is a potential host preceding the nominal predicate, POR indexes stay with its PUM noun whereas Topic indexes go with the initial host in Truku Seediq. But this is not the case in Tgdaya Seediq,

²⁶ The transcriptions here are identical those in the cited work. In Seediq, the verb forms in negative and imperative constructions are identical. In spite of this, a function-based glossing is adopted here, DEP for those after the negator (called connegative in Holmer 2006) and IMP for imperatives. By contrast, verb forms for these two functions may be distinct in Puyuma, to be discussed in §3.5.

where the position of POR indexes matches that of their isomorphic non-Topic Actor indexes. The contrast is shown in (16) and (17).

(16) Truku Seediq (Courtesy of Naomi Tsukida)

- a. **kuyuh=su=na** ka=isu
 wife=2SG.TOP=3SG.GEN TOP=2SG
 ‘You are his wife.’
- b. **adi=su** kuyuh=**na** ka=isu
 NEG=2SG.TOP wife=3SG.GEN TOP=2SG
 ‘You are not his wife.’

(17) Tgdaya Seediq

- a. **tama=su=mu** isu
 father=2SG.TOP=1SG.GEN 2SG.TOP
 ‘You are my father.’ (Holmer & Billings 2014: 119)
- b. **uxe=su=mu** tama
 NEG=2SG.TOP=1SG.GEN father
 ‘You are not my father.’ (Courtesy of Arthur Holmer)

Speakers of Tgdaya Seediq would immediately reject (17)b if the POR index *=mu* is attached to the nominal predicate *tama* ‘father’ instead, following the pattern in (16)b from Truku Seediq (p.c. Arthur Holmer). This is so despite the fact that elsewhere *tama=mu* ‘my father’ is an acceptable phrase in Tgdaya Seediq. Notwithstanding their identical forms, POR and non-Topic Actor indexes in Truku Seediq show disagreeing behaviors, which suggests that identical form does not always guarantee identical morphosyntax. This is also one of the reasons I choose to gloss indexes by functions rather than by forms. More glossing principles are discussed in §2.4.

To conclude this section, person-form clitics abound in Formosan languages. Although the same cognate forms are involved (see Ross 2006, 2015b), their behaviors differ drastically from one language to another in terms of the constraints on co-occurrence with their conominals and with regards to the ordering internal and external to these clitics. Their positioning is motivated by various factors (syntactic and otherwise),

and further complicated by their interactions with other types of clitics (e.g. TAM, evidential/epistemic, or discourse markers). Since many possible orderings have not yet been systemically tested, the exact syntactic scope of these essentially Wackernagel clitics still awaits additional research. Meanwhile, there is a growing number of studies specifically dedicated to this line of research, including L. Li (2010) on Isbukun Bunun, D. Yen & Billings (2011) on Mantaoran Rukai, D. Yen (2012) and D. Yen & Billings (2014) on Kavalan, Holmer & Billings (2014) on Seediq, Y. Chang (2012) and H. Liao (2005b) on Atayal, and finally S. Teng (2015) and H. Jiang & Billings (2015) on Puyuma.

2.4. Grammatical relations and glossing principles

This section provides justifications for glossing principles concerning Philippine-type grammatical relations, which are operated by Focus categories (§2.4.1) on the one hand and argument realization/indexing on the other (§2.4.2). For reasons outlined below, the former is glossed by form and the latter by function. Nevertheless, nothing concluded in this study hinges upon how the data are glossed.

2.4.1. Focus categories

The neat correspondences among four Focus categories in Tagalog, Paiwan, and Puyuma demonstrated above (see §2.3.1) may suggest that the semantic roles indicated by each Focus category are invariable across languages. In fact, despite the seemingly semantic labels like Actor, Patient, Location, and Conveyance, there are considerable variations with respect to the range of semantic roles each Focus type encompasses, both across languages and across verb roots within the same language. On formal grounds,

however, exponents of these categories all involve cognate affixes that are rather stable across languages (see Wouk & Ross 2002). Thus, I choose to gloss Focus categories in terms of their exponent forms.

For instance, A. Chang (2006: 75) lists the following specific semantic roles expressible by each of the four Focus types in Northern Paiwan: agent, experiencer, unintentional stimulator, and unintentionally affected goal for AF; patient, stimulus, and recipient for PF, location, partially affected patient, source, result, and time for LF; finally instrument, recipient, beneficiary, and reason for CF. As expected, not all languages that maintain the four Focus types like Paiwan assign the same range of semantic roles to the Topic NP selected by the verb marked for each Focus type. Thus, despite their semantic-looking denomination, Focus types in each language are better considered language-specific grammatical categories generalized over predicate types (resembling macro-roles) rather than verb-specific participant roles (G. Lin 2010).²⁷

Crosslinguistic variations aside, there seems to be a unifying semantic basis that constrains the categorical assignment of the three NAF types across languages. S. Huang (2005), for example, proposes a localist account by arguing that the Topic argument in PF, LF, and CF encodes a patient object, an abstract location, and a transported theme respectively, from which other verb-specific participant roles can be derived from metaphorical or metonymical extensions. On the other hand, however, despite some unifying principles of Focus selection across languages, stem-specific idiosyncrasies still abound, which then leads to unexpected gaps of word forms that are possible in principle. In Rikavung Puyuma (more on this language in §3.5.2.2), for instance, the three AF forms *kzeng* ‘<AF>pull’, *tengez* ‘<AF>beat’, and *sukun* ‘<AF>push’ all

²⁷ This is also why the labels for Focus types are capitalized.

express bodily action, but each of them selects a different NAF type when the Topic NP is the individual acted upon, as illustrated in (18).

(18) Rikavung Puyuma (Fieldnotes)

- a. **taw=kezeng-aw** ni=misak i=atrung
 INV=pull-PF ACT.SG=M. TOP.SG=A.
 ‘Misak pulled Atrung.’
- b. **taw=tengez-ay** ni=misak i=atrung
 INV=beat-LF ACT.SG=M. TOP.SG=A.
 ‘Misak beat Atrung.’
- c. **taw=sukun-anay** ni=misak i=atrung
 INV=push-CF ACT.SG=M. TOP.SG=A.
 ‘Misak pushed Atrung.’

With everything else being held equal, mixing the three NAF affixes across the three verb stems other than the combinations shown above would lead to unacceptable sentences. This situation cannot be easily accounted for by semantics alone, especially considering the contrast between a pulling and pushing event, which only differ in terms of the direction of the patient’s movement with respect to the agent. In both the (a) and (c) example, the referent of the Topic NP is equally a patient object (since it is acted upon) and a transported theme (since it moves over space). To some extent, the variations of stem-affix combinations of Focus-words across Formosan languages are quite comparable to those of verb-particle constructions across Germanic languages, where originally spatial and compositional verb-particle combinations may either acquire quite disparate meanings in different languages or end up having lexical gaps in some languages but not in others, even when cognate forms are involved. One example given by Thim (2012: 4) is *auf-machen* ‘up-make’ in German, which means “to open”, as compared with its cognate form *make up* in English, which means “to resolve a quarrel” (among others). Thus, it remains an open question why some verb stems are inclusive

enough to be marked by all Focus types available while others are rather selective, and more importantly what might motivate the lexical gaps. Before more comprehensive studies are available, it is necessary to respect the stem-specific selection of Focus categories, which makes it difficult to gloss them by semantics alone. At the very least, Focus selection is determined not only by how a language systemically extends general Focus semantics via metaphor or/and metonymy (e.g. whether to allow temporal terms to be the Topic of LF verbs), but also by how each verb stem in a language inclusively or exclusively selects certain Focus type(s).

Another area where Focus categories show variations across languages is the permutation between PF and LF, which is rather common among Formosan languages (S. Huang 2005: 784). Specifically, this phenomenon in Truku Seediq has been shown to be conditioned sometimes by default tense/aspect interpretations, as in (19), and other times by whether the Topic argument is fully or partially affected, as in (20).

(19) Truku Seediq (Tsukida 2000: 57)

- a. **tepaq-un**=nami ka=hini
eat-PF=1INCL.ACT TOP=PROX.LOC
'We will swim here.'
- a. **tepaq-an**=nami ka=hini
eat-LF=1INCL.ACT TOP=PROX.LOC
'We (usually) swim here.'

(20) Truku Seediq (Tsukida 2005: 318)

- a. wada=mu **hepuy-un** ka=seqemu
PST=1SG.ACT cook-PF TOP=corn
'I cooked (all) the corn.'
- b. wada=mu **hepuy-an** ka=seqemu
PST=1SG.ACT cook-LF TOP=corn
'I cooked (some of) the corn.'

What (19) shows is a contrast between PF *-un* marking future events and LF *-an* indicating habitual ones.²⁸ On the other hand, (20) illustrates the difference between a fully affected patient in construction with a verb marked by PF *-un* and a partially affected one in construction with a verb marked by LF *-an*. In Hopper & Thompson's (1980) framework of semantic transitivity, a lower degree of affectedness of the patient is attributed with a lower degree of transitivity (called Affectedness of O), which suggests that LF is associated with a lower degree of transitivity. This idea is consistent with another parameter of transitivity (called Affirmation), according to which negative polarity is said to show a lower degree of transitivity than its affirmative counterpart. In Nanwang Puyuma, for instance, PF marking in affirmative indicatives is permuted to LF in their corresponding negatives, as shown in (21).

(21) Nanwang Puyuma (S. Teng 2008: 207)²⁹

- a. **tu=pa-ka-lradram-aw=ku**
 3.ACT=CAUS-K-know-PF=1SG.TOP
 ‘{She/he/they} let me know.’
- b. **adri tu=pa-ka-lradram-i=ku**
 NEG 3.ACT=CAUS-K-know-LF.DEP=1SG.TOP
 ‘{She/he/they} didn't let me know.’
- c. **tu=beray-ay=ku dra paisu kan nana=lri**
 3.ACT=give-LF=1SG.TOP UND.INDF money ACT.SG mother=1SG.GEN
 ‘My mother gave me (some) money.’
- d. **adri tu=beray-i=ku dra paisu**
 NEG 3.ACT=give-LF.DEP=1SG.TOP UND.INDF money
 ‘{He/She/They} didn't give me money.’

²⁸ Similarly, the PF STEM-*un* and the LF STEM-*an* in both Wulai and Sliq Sqliq Atayal respectively describe future and past events by default (S. Chen 2007: 15). See Zeitoun & L. Huang (1997) for the default TAM interpretations of Focuc categories across five Formosan languages.

²⁹ When the possessor is first person singular, the word for “mother” is *nana*, the suppletive form of *tina*, which is used elsewhere. The cited work (S. Teng 1008: 12) uses <l> for the alveolar lateral /l/ and <lr> for the retroflex lateral /l̥/, which is linguistically more logical than the conventional orthography, where the other way around is the norm. Nevertheless, conventional orthographies have their own unique history, and are “illogical” in one way or another. This present study respects and adopts the conventional Nanwang orthography and transcribes the data accordingly.

In affirmative indicatives, PF is marked by *-aw* and LF by *-ay*. However, regardless of whether a verb stem is marked by either of them in affirmatives, it can only be marked by *-i* in negatives, which is regularly derivable from the offglide of LF *-ay*. In other words, the PF/LF distinction observable in affirmatives is neutralized into LF marking under negation, the exact context where transitivity is expected to be low. Similar permutations of PF into LF under negation are also documented in Mayrinax Atayal (L. Huang 2001: 55) and Truku Seediq (Tsukida 2009). Thus, both Affectedness of O and Affirmation Parameter suggest that LF demonstrates lower transitivity than PF. However, as Tsukida (2005) points out, there is one contrast between PF and LF that is not predicted by Hopper & Thompson's parameters. According to their Mode Parameter, realis events are supposed to have higher transitivity than irrealis ones. However, whenever there is a contrast between realis and irrealis events, LF is used to encode realis and PF irrealis (see (19) above), which is contrary to the prediction of Mode Parameter. Nevertheless, regardless of whether they are motivated by degrees of transitivity or something else, permutations across Focus categories, of which the Topic argument assumes essentially the same semantic role, make it unrealistic to gloss Focus affixes purely by function.

Finally, I follow Ross (2015c) in referring to two highly schematic morphological stems as the Mstem and Kstem, the two "principal parts" from which various possible verb forms can be predictably derived. By introducing the notion of Mstem and Kstem, he was able to avoid conventional labels like "finite" or "non-finite," to define verb classes in each language, and subsequently to reconstruct verb classes in Proto-Austronesian. The Mstem can be defined as the AF-word serving as the matrix predicate in the affirmative realis-indicative, and its exponents may be the root infixes by /m/ (as

well as its cognate forms; hence the term Mstem), the /m/-initial root, the root prefixed by /ma/, or simply the unaffixed root, depending on verb classes and languages. On the other hand, the Kstem is “purely morphological, that is morphomic (Aronoff 1994), in that there is no morphosyntactic feature shared in common by the morphological structures of which each is a part” (Ross 2015c: 283). One important generalization about the Kstem is that it is often the base to which NAF or causative affixes are attached. Exponents of the Kstem may include the unaffixed root, the /p/-initial root, or the root prefixed by /ka/ (as well as its cognate forms; hence the term Kstem), with the latter typically marking stativity in numerous Formosan languages (Zeitoun & L. Huang 2000). Because they are useful, efficient, and theory-neutral labels, the Mstem and Kstem will recur constantly in subsequent discussions.

For the purpose of glossing, morphological makeups of the Mstem and Kstem determine how they are glossed in this study. This can be illustrated by Tsuchida’s (1980: 211) data in Tamalakaw Puyuma, where he identified five verb classes, as summarized in Table 2.1. The Mstem is used in AF affirmative indicatives whereas the Kstem in AF imperatives (among others).

Table 2.1: Classes of AF verbs in Tamalakaw Puyuma (after Tsuchida 1980: 211)

	I	II	III	IV	V
Mstem	<i>m-iturus</i> ‘AF-follow’	<i>ma-tikezir</i> ‘AF-stand.up’	<i>may-kavang</i> ‘AF.wear-clothes’	<i>riked</i> ‘AF.laugh’	<i>veray</i> ‘AF.give’
Kstem	<i>iturus</i> ‘follow’	<i>ka-tikezir</i> ‘K-stand.up’	<i>pay-kavang</i> ‘K.wear-clothes’	<i>ka-riked</i> ‘K-laugh’	<i>veray</i> ‘give’

These five classes are defined by the morphological alternations between the Mstem and Kstem: *m*-ROOT vs. ROOT in Class I, *ma*-ROOT vs. *ka*-ROOT in Class II, *may*-ROOT vs. *pay*-ROOT in Class III, ROOT vs. *ka*-ROOT in Class IV, and finally ROOT vs. ROOT in Class V. The present glossing principles are as follows. First, AF is always indicated in glossing for the Mstem even when there is no dedicated morphology other than the root (as in Class IV and V). Second, the gloss K, an abstract label for the Kstem, specifies the affix of the Kstem that either alternates with an affix in the Mstem (as in Class II and III) or is added on top of the Mstem (as in Class IV). Finally, in cases where the Kstem is morphologically less marked than the Mstem (as in Class I) or simply identical to the Mstem (as in Class V), the gloss K is not used, which would imply a given form is the Kstem since the Mstem is always marked with the gloss AF.

2.4.2. Argument realization and indexing

While the previous section explains why Focus affixes are glossed based on form, this current one presents some justifications for why arguments realized as free nominals or/and indexed by bound person forms are glossed according to semantico-syntactically defined functions that they serve.

Just like the Focus system, the argument alignment pattern in Philippine-type languages has long been a contentious topic, and will most likely “[continue] to be a matter of controversy” (Himmelmann 2005a: 112). Since this topic is only tangential to the present work and a great amount of literature has been devoted to it, I do not intend to add anything new in this section but to briefly highlight some major proposals as well as explain the one adopted in this study (§2.4.2.1), which provides the basis for glossing

(§2.4.2.2). The glossing labels used here are definitely not standard because there is simply none (Himmelmann 2005a: 146). Nevertheless, for a comparative study like the present one, principled glossing is more important than which proposal is adopted as the basis for glossing. The reader is referred to H. Liao (2004: 144), G. Lin (2010: 22), and references therein for more detailed surveys of various approaches.

2.4.2.1. Ergativity and beyond

There are two primary approaches to argument structure in the Formosan and Philippine literature. One is the ergative hypothesis, and the other the symmetrical-voice hypothesis, with perhaps the former much more popular and widely adopted in various theoretical orientations (Schachter & Reid 2009). It is not uncommon that the same language is judged to be ergative by some researchers, but symmetrical-voice by others. Not surprisingly, the different judgments are often due to the different types of criteria (semantics, syntax, or discourse) either analysis is based on.

Under the ergative hypothesis (e.g. L. Huang 1994; Starosta 1997, 1999; H. Liao 2002, 2004; S. Wang 2004; Aldridge 2004a, 2012) NAF-constructions are transitive and the AF-construction is intransitive. Since the patientive argument in transitive NAF-constructions is marked in the same manner as the Actor-cum-Topic in the intransitive AF-construction to the exclusion of the agentive argument in transitive NAF-constructions, the alignment pattern is said to be ergative-absolutive (i.e. S/P pivot to the exclusion of A), although Austronesianists often choose to use glosses such as GEN and NOM instead of ERG and ABS for the A and P/S argument respectively (see H. Liao & Reid 2004 and Ross & S. Teng 2005). One of the consequences of the ergative hypothesis is

that since NAF-constructions are ergative, the AF-construction with patientive arguments would be essentially antipassive, where these arguments are significantly less definite than their counterparts in ergative constructions. Accordingly, the patientive argument in ergative NAF-constructions and that in the antipassive AF-construction are expected to be definite and indefinite respectively, at least significantly so. However, this expectation is not always born out among Formosan languages. For instance, while arguing for the ergative analysis in Puyuma, Ross & S. Teng (2005: 771) admit that the same analysis “cannot be applied” to languages like Kavalan, Seediq, and Squliq Atayal “without considerable modification” because discourse studies in these languages do not support a significantly indefinite patientive argument in the AF-construction. To illustrate by Kavalan, under the premise that the AF-construction is intransitive and NAF-constructions are transitive, the alignment pattern in terms of nominal relation markers would be ergative-absolutive, as in (22), where S and P are marked by *ya* whereas A by *na*.

(22) Kavalan (H. Chang 2004: 104)

- a. p<m>ukun tu=sunis **ya=baqi**
 <AF>beat UND=child TOP=old.man
 ‘The old man is beating a child.’
- b. pukun-an=na **na=baqi** **ya=sunis**
 beat-LF=3.ACT ACT=old.man TOP=child
 ‘The old man beat the child.’

While it is less questionable that the NAF-constructions like the second example are transitive, it has been a major concern regarding whether the AF-construction like the first example is grammatically intransitive and hence antipassive, especially considering its patientive argument can be as specific or even definite as those in (23).

(23) Kavalan

- a. p<m>ukun tu=**wasu unay ya=sunis=su**
 <AF>beat UND=dog MED TOP=child=2SG.GEN
 ‘Your child beat that dog.’ (H. Jiang 2006: 70)
- b. t<m>ayta=**iku timaimu**
 <AF>see=1SG.TOP 2PL.UND
 ‘I saw you (guys).’ (P. Li 1978: 353)

Although it is true that the AF-construction patientive argument is typically interpreted as indefinite unless modified by something else (see (22)a), discourse data, as well as examples like (23), show that this particular argument is not significantly less definite than the NAF-construction patientive Topic argument. It may very well be the case that the default indefinite interpretation of the AF-construction patientive argument is “a matter of pragmatic inference, not of grammaticisation” (Ross 2002: 31).

Unsatisfied with the antipassive analysis for the AF-construction, some researchers (e.g. Kroeger 1993; Ross 2002; Himmelmann 2005a) support the symmetrical-voice hypothesis instead, whereby AF- and NAF-constructions are symmetrical in terms of transitivity. The symmetry often comes in two aspects. Morphologically, both AF and NAF verbs are equally marked by affixes (see (22) above), with neither of them being the morphological default. This is in sharp contrast to languages with an ergative-antipassive system, where ergative verbs are morphologically less marked than antipassive ones. Syntactically, the agentive argument in both AF- and NAF-constructions are equally marked by nominal relation markers (see (22) again), which is unlike typical ergative-antipassive languages where A is more marked than S/P.

Similar in spirits (though not necessarily in argumentation) to the symmetrical-voice hypothesis is Shibatani’s (1988) view that the Philippine-type voice is not straightforwardly ergative or accusative, written at the time when the accusative

hypothesis was still on the table, according to which the AF-construction is active whereas NAF-constructions are passive (i.e. S/A pivot to the exclusion of P).³⁰ He contends that the Philippine-type voice should be better recognized as a distinct Topic construction of its own. The main thrust of his argumentation hinges on Schachter's (1976) pioneering observations in Tagalog, where subject properties (e.g. reflexive binding, equi-NP deletion, imperative addressee, quantifier floating, and relativization) are split over two NPs. However, unlike Schachter, who concluded that the notion of subject is not applicable to Philippine-type languages because of the split phenomenon, Shibatani argues the opposite by adopting a prototype approach to the notion of subject. This approach does not presuppose subjects across languages should always possess the same concatenation of properties or converge to one single NP. In fact, the split phenomenon is precisely what makes the notion of subject equally applicable to Philippine-type languages since the Topic-cum-Actor in the AF-construction exhibits the maximal number of subject properties whereas the non-Topic Actor and the Topic in NAF-constructions, but not any other random NPs, consistently demonstrate their respective shares of those same properties possessed by the Topic-cum-Actor. In other words, the two grammatically prominent NPs in NAF-constructions are also realizations of subjecthood, though they are less prototypical ones. Moreover, in Cebuano, from which Shibatani's (1988) data are mostly drawn, some subject properties (e.g. control behavior) are even simultaneously observable in both of the two prominent NPs in NAF-constructions rather than simply attributed to either of them, thus bolstering the idea of two arguments slicing up subjecthood (see *ibid.*: 125 for a complete list of subject

³⁰ The statement here is strictly speaking anachronic since Shibatani (1988) predated the symmetrical-voice hypothesis.

properties and their realizations between the Topic and Actor). This phenomenon is then likened to so-called “Dative Subject Construction” in languages like Japanese and Russian, where the two arguments of certain mental predicates (with the human experiencer marked in dative and the stimulus in nominative) both share subject properties that elsewhere converge to one single argument.

Since the accusative hypothesis has generally been abandoned, only some of Shibatani’s (1988) arguments against the ergative hypothesis are summarized here. First, unlike syntactically ergative languages like Dyrbal, where the controller of the gap in the second coordinated clause is always coreferential with the absolutive argument in the first coordinated clause (i.e. encoding S/P), the controller in Cebuano is the Actor nominal regardless of whether it is also the Topic or not (i.e. encoding S/A), as in (24).

(24) Cebuano (Shibatani 1988: 107)

- a. ni-bunal **si** **juan** ni **pedro** ug ni-lakaw
 AF-hit TOP J. UND P. and AF-leave
 ‘Juan hit Pedro and left.’ (Juan left.)
- b. gi-bunal-an **ni** **juan** si **pedro** ug ni-lakaw
 PFV-hit-LF ACT J. TOP P. and AF-leave
 ‘Juan hit Pedro and left.’ (Juan left.)

In terms of nominal relation markers alone, Cebuano does illustrate the ergative-absolutive pattern (*si* for S/P and *ni* for A) like Dyrbal. But the control behavior shows the nominative-accusative pattern instead like English, with S/A being the potential gap-controller to the exclusion of P. When the second coordinated clause is a NAF-construction with the Actor and Topic realized by distinct arguments, the control-gap configurations are much more complicated, but even there the syntax does not follow the ergative-absolutive pattern. Second, AF and NAF verb forms are equally complex, with neither of them being the morphological default, which is also one piece of the evidence

advanced in the symmetrical-voice hypothesis. Third, antipassive is essentially a valency-decreasing mechanism that marginalizes (or “demotes”) the patientive argument, but the non-Topic patientive argument in the AF-construction “is an integral element of the clause, and omission of it results in an elliptical expression, which an antipassive form is not” (ibid.: 113). Last but not least, like passive forms, antipassive ones are rather marked and thus show extremely low text frequency according to many studies (4.9% in Eskimo, citing Kalmár 1979; 11% in Australian languages, citing Tsunoda 1988). By contrast, the AF-construction with the patientive non-Topic argument is much more frequent in texts, taking up 24% in Tagalog (citing Cooreman et al. 1984) and amounting to as high as 52% in Cebuano. Therefore, Shibatani (1988: 113) concludes that “the differences between the [Actor-construction] and the antipassive outweigh the similarity between them. The same conclusion can be drawn with regard to the proposal that the [NAF-constructions] be considered as an ergative construction.”

Finally, H. Chang (2004) evaluates the ergative and symmetrical-voice hypothesis against data from six Formosan languages and concludes that the ergative hypothesis is “too strong a claim” (ibid.: 117) because the AF-construction in some of these languages is in no way close to antipassive/intransitive, both semantically and morphosyntactically. One of his crucial arguments is that the AF-construction is neither semantically nor syntactically homogeneous, and shows varying degrees of transitivity among them. For instance, in Saisiyat, Paiwan, Seediq, and Mayrinax Atayal, stative or resultative events with only the experiencer are expressed by AF prefixes like *m-* or *ma-* (depending on languages), but causative events with both the causer and causee by AF infixes like $\langle m \rangle$,

<um>, <en>, or <om> (depending on languages).³¹ The minimal pair in (25) from Mayrinax Atayal illustrates the contrast.

(25) Mayrinax Atayal (H. Chang 2004: 108; citing L. Huang 2000)³²

- a. **ma-taqu**=[ku 'ulaqi]
 <AF>fall=TOP child
 'The child fell down.' (The child is on the ground now.)
- b. **t<um>aqu**=[cku nabakis]=[ku 'ulaqi]
 <AF>fall=UND old.man=TOP child
 'The child made the old man fall down (e.g. by pushing him).'

On the other hand, he also argues that languages like Tsou and Kavalan do not conform to the symmetrical-voice hypothesis, unlike Tagalog, which was concluded to be a symmetrical-voice language by both Kroeger (1993) and Ross (2002). One crucial question is whether the non-Topic patient in the AF-construction plays a core or peripheral syntactic role. According to Kroeger (1993: 47-48), the *ng*-marked non-Topic patient in Tagalog assumes a core role whereas the one marked by *sa* has only a peripheral one because the former, not the latter, can be the controller of the gap in an adverbial construction introduced by *nang*, as contrasted in (26).

³¹ See M. Yeh (2003: 64) for a summary table of high- vs. low-transitivity AF affixes in five Formosan languages.

³² Nominal relation markers in Mayrinax Atayal are consistently ditropic. The mismatch between syntax and prosody is also noticed by T. Chen (2010: 21). Word-final glottal stops in Mayrinax Atayal are not transcribed here, which differs from the practice in the original source. See Footnote 25 for the rationale.

(26) Tagalog (Kroeger 1993)³³

- a. nang-huli **ng=magnanakaw** ang=polis
 AF.RLS-catch UND1=thief TOP=police
 [nang p<um>a~pasok sa=banko]
 ADVZ <AF>IPFV~enter UND2=bank
 ‘The police caught {a/the} thief when entering the bank.’
 (Either the police or {a/the} thief was entering the bank.)
- b. b<um>isita si=juan **sa=hari** [nang nag-i~isa]
 <AF>visit TOP=J. UND2=king ADVZ AF.RLS-IPFV~alone
 ‘Juan visited the king alone.’ (Juan is alone.)

Since there are non-Topic patients that can be a potential controller just like the Topic, whose core syntactic function is never controversial, it follows that the AF-construction in Tagalog are morphosyntactically transitive with two core arguments (cf. Ross 2002). Building on the control behavior in Tagalog, H. Chang (2004) uses a similar test to distinguish languages where the non-Topic patient in the AF-construction can be a controller, such as Truku Seediq, from those where this is not possible, such as Tsou and Kavalan, both of which, he argues, demonstrate asymmetrical transitivity between AF- and NAF-constructions, both semantically and grammatically. The two types of languages are illustrated in (27) and (28) respectively.

(27) Truku Seediq (H. Chang 2004: 106)

pskiyux **knan** **m-usa** ka=hiya
 AF.force 1SG.UND AF-go TOP=3SG
 ‘{She/He} forces me to go.’

(28) Kavalan (H. Chang 2004: 111)

- a.* pawRat=iku **tu=sunis** **ma-ynep**
 AF.force=1SG.TOP UND=child AF-sleep
- b. pawRat=iku **tu=sunis** **pa-qa-ynep**
 AF.force=1SG.TOP UND=child CAUS-K-sleep
 ‘I force the child to sleep.’

³³ See Footnote 14 for the way Tagalog data are represented and glossed here.

Unlike in Truku Seediq, the non-Topic patient in Kavalan cannot be the controller of a complement predicate. To remedy (28)a, the complement predicate has to be causativized, thus making the Topic argument the controller. Thus the control phenomenon is a common way to find out whether a controversial argument has a core or peripheral syntactic role, and it is generally agreed that syntactic processes like control behavior are more reliable than the marking of nominal relational markers or person forms (see Ross 2002: 28; Himmelmann 2005a: 147; more on this point below). Taken together, the two parts of H. Chang's (2004) study suggest that neither the ergative nor symmetrical-voice hypothesis neatly applies to all the six languages examined.

Language variations aside, the generalization from the studies reviewed above seems to be that ergativity in Philippine-type languages does not go much beyond patterns of argument marking because their verbal morphology, syntactic processes, and discourse behaviors do not resemble those of prototypical ergative languages. More importantly, even if we take argument marking seriously, the way syntactic-semantic primitives like S/A/P is determined in Philippine-type languages has been based on two types of syntactic constructions, that is, taking S from the AF-construction but A/P from NAF-constructions. However, the way S/A/P is defined in Dixon's (1994) typology is to identify S/A/P from the same syntactic construction, such as the active-voice construction in English. If we were to take S from the active-voice construction (e.g. *The master came.*), but A/P from the passive-voice construction (e.g. *The master was heard by the slaves.*), we would come to the conclusion that English is an ergative language based on argument marking (i.e. S/P pivot), which is a very distorted view.³⁴

³⁴ The English examples are modelled on those in Dixon (1994).

Finally, even if we restrict the definition of ergativity to the alignment pattern of argument marking alone, “nothing else necessarily accompanies this,” (Dixon 1994: 219), a point cited by both Himmelmann (2005: 158) and Blust (2013: 457). Since no typological properties have been found to correlate with ergativity as defined by argument alignment, not much information is lost even if we do not gloss Philippine-type languages using case labels such as ERG and ABS (or GEN and NOM, depending on one’s preference), as is commonly done in the Austronesian literature. More importantly, even if we focus only on argument alignment patterns alone, ergativity is not generalizable to all argument types or/and all noun classes of free nominals. The last point is to be demonstrated in the next section.

2.4.2.2. Argument encoding and indexing

Following the spirits of Shibatani (1988, 1991), this study uses syntactic-semantic categories (Topic, Actor, Undergoer, and Oblique) to gloss nominal relational markers for full nominals and bound person forms that index those nominals. I first define these categories and then demonstrate the diverse syncretic patterns within and across languages, which is the very fact that makes it difficult to make generalizations over forms and to gloss the data entirely based on form. Rukai is perhaps the only Formosan language where these categories are not necessary.

First, the Topic argument is defined as free nominals or/and person-form indexes whose semantic role is somewhat constrained by Focus affixes, which grammatically categorize Topic into subtypes (maximally Actor-Topic, Patient-Topic, Location-Topic, and Conveyance-Topic). Second, the Actor argument is the one that is either the Actor-

Topic or the most agentive argument when the Topic is not also the Actor-Topic. As restatements of previous studies, the Topic has referential prominence and is thus highly definite across languages whereas the Actor has agentive prominence and thus tends to be human or at least animate. The two are most syntactically privileged categories due to their prominence. The Actor-Topic is no doubt the grammatical subject since it realizes the intersection of two types of prominence. When the two relations do not converge on the same NP, languages vary with respect to how either of the two slices up the pie of subjecthood. Third, the Undergoer is defined negatively as the non-Topic and non-Actor patientive argument. Finally, the Oblique is a miscellaneous category for anything other than the three just defined, covering such semantic roles as location, goal/source of motion, recipient, time, instrument, beneficiary, and more unless a language has a dedicated marker for a specific type of roles (e.g. LOC for locative or/and temporal roles). The Topic is always glossed as TOP regardless of its semantic roles. As a result, when something is glossed as ACT, UND, or OBL, it is implied that it is a non-Topic.³⁵ This is due to the fact that the marking for the Topic, as it were, overrides that for others.

It should be pointed out that given their definitions these categories are used here as *comparative concepts*, rather than *descriptive categories* (Haspelmath 2010). Under the assumption of categorical particularism, no categories in two linguistic systems would be completely identical even though they do demonstrate a great number of overlapping functions. Accordingly, categories in a language are the result of functional compartmentalization within that language as a system and are thus language-specific

³⁵ In other words, as a functional category, the Actor can be the Topic or a non-Topic. But based on the present glossing principles, ACT is understood to be a non-Topic since the Actor-Topic is simply glossed TOP. The Undergoer and Oblique are always non-Topics by definition. The present glosses TOP, ACT, UND, and OBL roughly correspond to SPEC (for specific), GEN, NPIV (for non-pivot and non-agent), and LOC respectively in Ross (2002: 22).

descriptive categories, which are not readily comparable across languages. However, for the sake of crosslinguistic comparison, typologists designate comparative concepts that are formulated based on conceptual-semantic notions. Importantly, comparative concepts and descriptive categories, though they often bear the same labels, “stand in a many-to-many relationship” (ibid.: 665).

The difference outlined above can be demonstrated by comparing how the non-Topic Actor is marked between two closely related linguistic systems as well as how the Undergoer nominal is marked differentially within the same language. The first part is illustrated by Nanwang and Rikavung Puyuma as in (29).

(29) Nanwang Puyuma (S. Teng 2008: 147, 159)

- a. tu=trakaw-aw na paisu **kan** **isaw**
 3.ACT=steal-PF TOP.DEF money ACT.SG I.
 ‘Isaw stole the money.’
- b. masupeng=ku **kan** **nana=lri**
 AF.miss=1SG.TOP UND.SG mother=1SG.GEN
 ‘I miss my mother.’

(30) Rikavung Puyuma (Fieldnotes)

- a. {tu/taw}=kezeng-aw na=takulris **ni=misak**
 INV=pull-PF TOP.DEF=goat ACT.SG=M.
 ‘Misak pulled the goat.’
- b. m-na’u~na’u=ku **kani=misak**
 AF-IPFV~see=1SG.TOP UND.SG=M.
 ‘I keep looking at Misak.’

In Nanwang Puyuma, singular personal free nominals (including personal names and some kinship terms) that function as the non-Topic Actor are marked by *kan*, and so are those fulfilling the non-Topic Undergoer. Since the same marker is also used to introduce various non-Undergoer peripheral roles, all uses of the marker *kan* could be analyzed as one single language-specific descriptive category Oblique, as is done by S. Teng (2008).

In Rikavung Puyuma, however, the non-Topic Actor and the Undergoer consisting of

equivalent nominals are marked by *ni* and *kani* respectively, and only the latter is also used to mark various non-Undergoer peripheral roles.³⁶ Thus, from a language-internal perspective, *ni* and *kani* could be analyzed as two distinct descriptive categories (often called Genitive and Oblique respectively). But descriptive categories are meaningful only in the overall nominal marking system of each language. On the other hand, from a comparative point of view, the non-Topic Actor in both Nanwang and Rikavung demonstrates similar syntactic and discourse functions (e.g. controlling anaphor binding) regardless of whether they are marked the same as or differently from the Undergoer. Thus, consistent glossing based on semantico-syntactically defined categories such as Actor makes the present comparative task transparent and relieves the reader of the memory load necessary for keeping track of which language-specific categories in which languages are used to express which functions.

The second part has to do with the marking of the Undergoer nominal, which can often be marked in more than one way, a phenomenon comparable to differential object marking (DOM). Common semantic correlates responsible for DOM include definiteness and animacy. When this happens, multiple markers for the Undergoer in the same language are arbitrarily numbered as UND1 and UND2, such as *ng* and *sa* in Tagalog, both having been demonstrated in (26) above. Interestingly, it is often the case that one of them is also used to mark other non-Topic and non-Actor peripheral roles, in which case the gloss UND would not be used. This applies to Tagalog *sa*, as in (31).

³⁶ In this regard, Rikavung is similar to two other non-Nanwang varieties of Puyuma. See S. Teng (2009) for details.

(31) Tagalog (Nagaya 2007: 356)³⁷

nag-meryenda=lang=siya **sa=bahay**

AF.RLS-have.a.snack=only=3SG.TOP LOC=house

‘{She/He} only had a snack in the house.’

As a result, Tagalog *sa* would have two different glosses, depending on whether or not it marks the non-Topic patientive argument (i.e. the Undergoer). This practice may not be desirable since *sa* could simply be given one consistent label based on form. However, by singling out the patientive Undergoer from other roles we would be able to compare the syncretic patterns across markers in the same language. Taking the two markers under discussion for instance, *ng* and *sa* in Tagalog share the function of marking the Undergoer, but they each have another function to the exclusion of the other. While *sa* also marks non-Undergoer locative roles, as in (31), *ng* also introduces the non-Topic Actor, as in (1)b~d. To put it schematically, *ng* shows the syncretic pattern of ACT=UND to the exclusion of LOC and *sa* of UND=LOC to the exclusion of ACT. Thus, on a continuum going from (non-Topic) Actor to Undergoer and then Locative, the two markers find their niches on either side of it, showing a shared function in the midpoint (i.e. the Undergoer).³⁸

The four comparative concepts defined above are in response to variations of syncretic patterns across and within languages. The variations within languages are correlated with whether a given function is realized by free nominals or bound person forms, and sometimes even by noun class types among free nominals. This is

³⁷ The glossing for aspect and mood follows the analysis in Himmelmann (2005 363), which slightly differs from the original glosses in the cited reference. When marking non-Undergoer roles, the marker *sa* is commonly glossed as Dative (as in the cited work) or Locative (as in Reid 2005). Here the latter label is chosen.

³⁸ See Shibatani (1988) for detailed syncretic patterns in four languages of the Philippines. Regardless of crosslinguistic variations, there seems to be a general tendency for a marker to specify continuous roles on the continuum (either all three or two of them on either side of it).

demonstrated by six linguistic systems below, which, despite the small number, are sufficient to make the point.

In terms of maximal distinctions of free nominals assuming various functions, two common syncretic patterns are found, as in (32).

(32) Syncretic patterns of nominal relation markers in six Formosan languages³⁹

a. TOP|ACT=UND=OBL: Nanwang Puyuma, Isbukun Bunun, and Tsou

b. TOP|ACT|UND=OBL: Rikavung Puyuma, Northern Paiwan, and Kavalan

The first pattern in (32)a means that the (non-Topic) Actor, Undergoer, and Oblique all receive the same marking to the exclusion of the Topic. On the other hand, the second in (32)b indicates that Undergoer and Oblique both receive the same marking, distinct from the marking for the (non-Topic) Actor, which is again distinct from the marking for the Topic. Thus, the two patterns vary in terms of how the (non-Topic) Actor is treated, which has been demonstrated by Nanwang and Rikavung Puyuma above.

The reason to emphasize maximal distinctions is that in some languages not all free nominals follow the same syncretic pattern. For instance, Rikavung Puyuma shows three syncretic patterns across nominal types, including both of the patterns in (32), as summarized in Table 2.2.

³⁹ The equal sign indicates identical marking and the vertical bar distinct marking.

Table 2.2: Nominal relation markers in Rikavung Puyuma (after H. Jiang 2013)

Nominal types	TOP	ACT	UND	OBL
Personal singular (TOP ACT UND=OBL)	<i>i</i>	<i>ni</i>	<i>kani</i>	<i>kani</i>
Personal associative plural & Definite common (TOP=ACT UND=OBL)	<i>na</i>	<i>na</i>	<i>kana</i>	<i>kana</i>
Indefinite common (TOP ACT=UND=OBL)	<i>a</i>	<i>za</i>	<i>za</i>	<i>za</i>

Notice that the pattern shown by personal associative plural nouns and definite common ones is where the ergative hypothesis fails. They show a *neutral alignment* (Siewierska 2004: 51) rather than the ergative-absolutive because S/A/P are all marked the same (i.e. TOP=ACT). It seems that personal nouns tend to make more formal distinctions than non-personal ones and it is the latter that often defies a unified analysis. As Himmelmann (2005: 146) puts it, “[f]rom a cross-linguistic point of view, the distribution of the non-personal markers in particular is somewhat unusual. To date, there is no standard analysis and terminology in use for these forms in the literature on western Austronesian languages.”⁴⁰

When it comes to bound person forms, the same languages in (32) demonstrate four syncretic patterns as in (33).

⁴⁰ The term “western Austronesian languages” in the quote is intended as shorthand for non-Oceanic Austronesian languages.

(33) Syncretic patterns of bound person forms in six Formosan languages

- a. TOP|ACT: Nanwang Puyuma, Northern Paiwan, and Kavalan
- b. TOP|ACT=UND: Isbukun Bunun
- c. Topic-Actor=non-Topic Actor: Tsou
- d. Topic-Actor|non-Topic Actor=non-Actor-Topic: Rikavung Puyuma

These languages have no more than two sets of person-form clitics indexing the Topic, (non-Topic) Actor, and Undergoer. (The Oblique is never indexed, so is excluded from the discussion here.) The first and most dominant pattern is (33)a, where two index sets are available, with one set indexing the Topic and the other indexing the (non-Topic) Actor. The two are structurally distinguished by the forms or/and positions of indexes with respect to the verb (or preverbal auxiliaries). Illustrative examples from Northern Paiwan are shown in (34), where the Undergoer is not indexed, but is expressed by free forms instead.

(34) Northern (Sandimen) Paiwan⁴¹

- a. **na-pacun=aken** **tjaymadju**
 PFV-AF.see=1SG.TOP 3SG.UND
 ‘I saw {him/her}.’ (A. Chang 2006: 72)
- b. **ku=k<in>eljem=esun** **katiaw**
 1SG.ACT=<PFV.PF>beat=2SG.TOP yesterday
 ‘I beat you yesterday.’ (A. Chang 2006: 65)
- c. **ku=si-k<in>eljem=esun** **tjaymadju**
 1SG.ACT=CF-<PFV>beat=2SG.TOP 3SG.UND
 ‘I beat {him/her} for you.’ (A. Chang 2000: 98)

The second pattern in (33)b is found in Isbukun Bunun, where both the (non-Topic) Actor and the Undergoer are indexed by the same person forms to the exclusion of the Topic.

⁴¹ Like Rikavung Puyuma, some Paiwan varieties also have two person-form clitics clustering together instead of sandwiching the verb on both sides. However, the clustering pattern in Paiwan is apparently rather limited, so is not discussed here. For historical developments of the two patterns, see Ross (2015a)

(35) Isbukun Bunun (L. Li 2010: 61-62)⁴²

- a. na=ma-ludah=**im=su**
 FUT=AF-beat=1EXCL.TOP=2SG.UND
 ‘We will beat you.’
- b. na=ludah-un=**su=im** auba
 FUT=beat-PF=2SG.ACT=1EXCL.TOP QP
 ‘Will you beat us?’

The Topic index precedes the Undergoer index in the AF-construction, but follows the (non-Topic) Actor index in NAF-constructions. In either case, the agentive argument always precedes the patientive one.

Unlike the first two, the third pattern in Tsou and the fourth one in Rikavung Puyuma do not draw a line between the Topic and (non-Topic) Actor in terms of argument indexing, and both languages only have one index set in terms of *forms*. Given such a structural limitation, they have to distinguish the agentive role from the patientive one by some other means. Tsou solves this problem by only indexing the agentive role, namely the Actor, whether it also doubles as the Topic or not, as shown in (36).

(36) Tsou (G. Lin 2010: 92)

- a. mi='o c<m>uhu=[to moatʰ'nʌ]
 AF.RLS=1SG.TOP <AF>butcher=UND goat
 ‘I butchered a goat.’
- b. i='o chu-a=['o moatʰ'nʌ]
 NAF.RLS=1SG.ACT butcher-PF=TOP goat
 ‘I butchered the goat.’

Assuming that the AF-construction is intransitive and the NAF-constructions are transitive, the alignment pattern of bound person forms is clearly nominative-accusative (i.e. S/A pivot) rather than ergative-absolutive, which is only found among free nominals (see (32)a above). Similarly, Rikavung Puyuma also indexes the Topic Actor and non-Topic Actor, but on different sides of the verb. Additionally, the non-Topic Actor and the

⁴² Post-vocalic glottal stops are transcribed in the cited work but not here.

non-Actor Topic are indexed on the same side of the verb, thus creating a clitic cluster.

Relevant examples are shown in (37).

(37) Rikavung Puyuma⁴³

- a. **m-na'u~na'u=ku** kannu
 AF-IPFV~see=1SG.TOP 2SG.UND
 'I keep looking at you.' (Fieldnotes)
- b. **ku=nu=sukun-anay**
 1SG.ACT=2SG.TOP=push-CF
 'I pushed you.' (H. Jiang & Billings 2015: 90)
- c. **taw=ku=nu=sukun-anay**
 INV=1SG.TOP=2SG.ACT=push-CF
 'Did you push me?' (H. Jiang & Billings 2015: 100)

Interestingly, Rikavung Puyuma solves the semantic problem of who acts upon whom by resorting to a person-based hierarchy whereby the first person must precede the second person and by using an inverse marker, as has been argued by H. Jiang & Billings (2014, 2015). Since the order between the first- and second-person form is fixed, the presence or absence of an inverse marker helps to determine the flow of action. Assuming once again the AF-construction is intransitive and the NAF-constructions are transitive, the S index is not treated in the same manner as the A or P index, which would render the nominative-accusative or ergative-absolutive alignment respectively. Instead, argument indexing in Rikavung Puyuma demonstrates a *horizontal alignment* (Comrie 2013: 23), whereby S is treated differently from *both* A and P. More specifically, A and P in Rikavung Puyuma shows a *hierarchical alignment*, defined by Siewierska (2004: 51) as one where "[e]ither one or the other is singled out for special treatment depending on which is higher on the person and/or animacy hierarchies." This is in contrast to the

⁴³ The fact that (37)b is translated into a statement but (37)c into a question has nothing to do with the *taw* marker, but is due to a pragmatic concern instead. As has been argued in H. Jiang & Billings (2015), formal variations of person-form indexes in Rikavung Puyuma are conditioned by positional allomorphy rather than semantic roles or syntactic functions. Nevertheless, person-form indexes are still given functional glosses to be consistent with the glossing principles adopted in the present study.

patterns found among free nominals in the same language, with personal singular nouns and indefinite common ones having an ergative-absolutive alignment whereas personal associative plural nouns and definite common ones a neutral alignment (see Table 2.2).

When commenting on western Austronesian languages in general, Himmelmann (2005: 158) observes that “clear-cut cases of ergative alignment are restricted to person marking systems, with little or no evidence of an ergative distribution of noun phrase markers.” Regardless of the extent to which this statement is true of Formosan languages, one thing for sure is that person forms and free nominals in the same language tend to show different alignment patterns. Among the six Formosan languages demonstrated above, we see ergative, accusative, and horizontal alignments with respect to bound person forms. And of free nominals, the ergative alignment is most dominant, though sometimes it is not generalizable to all nominal classes, as in Rikavung Puyuma. Due to the diverse patterns of argument realization and indexing, I choose syntactic-semantic categories like Topic/Actor/Undergoer/Oblique over case labels like GEN/NOM or ERG/ABS for the purpose of glossing.

Finally, notwithstanding the aforementioned principles, glosses should always be understood with the premise that they only highlight certain functions of a particular morpheme and that the same glosses across languages only capture partial functional similarities.

2.5. Chapter summary

This chapter has introduced some background information on Formosan languages, which are genetically non-Malayo-Polynesian Austronesian and typologically

quite similar to Philippine languages. Shared grammatical features include Focus morphology, nominal relation markers that are typically phonological clitics, and Wackernagel person-form clitics. Moreover, some issues concerning the glossing principles adopted in this study are addressed, such as the grammatical nature of Focus categories and the diverse argument alignment patterns across languages.

|Chapter 3

Verbal-based Nominalization I: Synchronic Variations

The best candidate word forms to be reasonably called verbs as a distinct lexical class in Formosan languages are those marked by Focus affixes (reconstructable in Proto-Austronesian; see Table 3.1 below), and Focus-marked word forms are indeed generally referred to as verbs by convention. I will follow this tradition, but in cases where I hope to be neutral, I speak of Focus-marked word forms as a morpho-lexical class, or Focus-words for short, and refer to constructions of which Focus-words are constituent elements as Focus-constructions, regardless of what syntactic functions they might assume.⁴⁴ By extension, Focus-words from a specific Focus category can be called AF/PF/LF/CF-words, and constructions where they are the category-determining word form of a specific Focus category can be referred to simply as AF/PF/LF/CF-constructions.

⁴⁴ This practice is borrowed from Himmelmann (2008: 285), who refers to Focus-marked word forms as “V-words”, where V stands for “voice-marked.” Since I have been using the term Focus instead of voice (which is more or less a terminological choice), an equivalent shorthand would create an awkward effect, which is why it is avoided in the first place.

This chapter investigates the synchronic variations of entity-denoting expressions consisting of Focus-words in Formosan languages, or Focus nominalizations for short. As preliminaries, §3.1 outlines both the synchronic and diachronic perspective to the syntactic functions of Focus-words in Philippine-type languages. Then §3.2 presents three conceptual problems in the previous studies of Formosan nominalization. It is argued that steering away from these problems would help us obtain a broader and clearer picture of what Focus-words do across languages (this chapter) and even how they might evolve over time (Chapter 4). Finally, §3.3, §3.4, and §3.5 respectively concentrate on Tsou, Rukai, and Puyuma because they each differ from all the other Formosan languages in such a unique way that they play an important role in the Nuclear Austronesian hypothesis (see §3.1.2).

3.1. Preliminaries: Two perspectives

An ostensible feature of Philippine-type languages is that many Focus-words, which show grammatical distinctions for tense/aspect/mood (TAM for short) and often attract argument-indexing clitics, can function as both matrix predicates (occurring in the sentence-initial position) and arguments (occurring in the post-predicate position introduced by nominal relation markers just like underived nouns). Syntactically versatile Focus-words in Philippine-type languages have not only spawned the nominal-verbal controversy from a synchronic perspective (§3.1.1), but also encouraged the nominalization-into-verb hypothesis from a diachronic perspective (§3.1.2). Both are discussed in this section.

3.1.1. Synchronic perspective: The nominal-verbal controversy

The predicate-argument isomorphism has long been observed in Tagalog by Bloomfield (1933: 173), who refers to Focus-words as “transient” words, which “are by no means confined, like our verbs, to predicative position; they can figure equally well, for instance, in equational sentences.” Example (1) illustrates his point, where the Focus-word is in bold.

- (1) Tagalog (Kaufman 2009b: 19)⁴⁵
- a. **nag-íngay** ang=áso
 AF.RLS-make.noise TOP=dog
 ‘The dog made noise.’ [Predicate function]
- b. áso ang=**nag-íngay**
 dog TOP=AF.RLS-make.noise
 ‘The one that made noise was a dog.’ [Argument function]

The phenomenon is indisputable, but how to reconcile both the predicate and argument functions of Focus-words has presented an analytical problem, thus resulting in a long-standing point of contention and controversy.⁴⁶ Ever since Bloomfield, at least three types of analyses have been proposed, with the nominal-verbal distinction being the critical culprit, as summarized in (2).

- (2) Three analyses of Focus-words in Philippine-type languages
- a. Both the predicate and argument function are verbal.
 b. Both the predicate and argument function are nominal.
 c. The nominal-verbal controversy is a moot question unless morphological and syntactic levels of categories are distinguished.

For researchers who vouch for (2)a (e.g. Kroeger 1993), Focus-words are verbs and always verbal and thus serve as matrix predicates, and their argument function is

⁴⁵ See Footnote 14 for the way Tagalog data are represented and glossed here.

⁴⁶ Schachter (1985) discussed a similar debate in Nootka, which, like many Philippine-type languages, allows verb-like forms to occur as predicates as well as arguments. See Haspelmath (2012: 116) for a summary of similar debates on word-class distinctions in other language families.

accounted for in terms of “headless relative clauses.” That is, for every instance of Focus-words occurring in the argument position and denoting an event participant, it is a relative clause (RC) “in disguise,” with the head of that RC being “elided” or “deleted” (i.e. a null nominal head). Accordingly, under this analysis Focus-words in the argument position always involve relativization (headed or not) since they are verbal, and there is thus no nominalization involved in (1)b.⁴⁷

Reid (2002: 299) held another twist of the view in (2)a by rejecting a null nominal head because “there’s no such thing as a headless relative clause.” In his analysis, Focus-words are always verbs, and the nominal head of denoting phrases consisting of Focus-words is not an invisible noun that somehow gets deleted, but nominal relation markers like *ang*, which are “specifying-nouns meaning ‘the one’” (ibid.: 285).⁴⁸ As a result, the Topic phrase in (1)b is said to be an NP headed by *ang* with a RC as its dependent. The consequence of this analysis is that even the Topic phrase in (1)a is an NP headed by *ang*, with the lexical noun *áso* ‘dog’ being its dependent (meaning “the one being a dog”). However, unlike *the one* in English, these “specifying-nouns” do not make up legitimate NPs without a following lexical item, be it a Focus-word or a lexical noun like *áso* ‘dog’ (ibid.: 306), but lexical nouns like *áso* ‘dog’ can dispense with “specifying-nouns”. The uneven dependency relationship between lexical nouns and “specifying-nouns” shows that the latter is syntactically more deficient, thus questioning its status as the syntactic head of NPs. Moreover, Reid’s analysis of Philippine languages cannot be easily applied to Formosan languages where nominal relational markers have been all but lost. For

⁴⁷ For (1)b Richards (2009: 140) posited not only a null nominal head, but also a null copula.

⁴⁸ See Starosta (2009: 510) for a similar proposal in Tsou, where similar markers were called “relator nouns, nouns with minimal semantic content which function as the syntactic head of a construction and carry syntactic or semantic features characterizing the noun phrase as a whole.”

instance, compared with her sister languages, Squliq Atayal has an impoverished system of nominal relation markers, which are often not used in daily conversations. Its Topic nominal is introduced by *qu*, which is optional whether immediately followed by a lexical noun or Focus-word, as in (3).

(3) (Jiangshi) Squliq Atayal (L. Huang & Hayung 2011: 6, 35)⁴⁹

- a. m-qwas **qu** **tali** ki ciwas krryax
 AF-sing TOP T. COM C. often
 ‘Tali often sings with Ciwas.’
- b. t<m>apih tapih **limuy** ru m-qwas sayun
 <AF>fan fan L. CONJ AF-sing S.
 ‘(When) Limuy is fanning a fan, Sayun is singing.’
- c. sayun ru **tali (qu)** m-usa mngka
 S. CONJ T. TOP AF-go Taipei
 ‘Those who will go to Taipei are Sayun and Tali.’

Crucially, with or without the marker *qu* the AF construction *m-usa mngka* in (3)c is just as denoting a phrase as the personal name *tali* in (3)a. If the marker *qu* were the head of an NP, we would again end up having headless RCs when *qu* is absent, the very idea rejected by Reid.

On the other hand, Reid also rejected the claim that Topic phrases like those in (1)b are “zero-derived deverbal nominalizations” (ibid.: 299) due to the wide range of word classes that are allowed to follow the marker *ang* in Tagalog, with the negator and a preposition illustrated in (4).

(4) Tagalog (Reid 2002: 299, 301)

- a. **ang=hindí** ma-ta~talíno
 TOP=NEG AF-IPFV~smart
 ‘those who are not smart’ (citing Schachter & Otnes 1972: 518)
- b. **ang=pára** sa=bátà
 TOP=for OBL=child
 ‘the one that is for the child’ (citing Lemaréchal 1982: 21)

⁴⁹ Word-final glottal stops in Squliq Atayal are not transcribed here. See Fotenote 25 for the rationale.

While he conceded the possibility of Focus-words being nominalizations, such as the Topic phrase in (1)b, “it is unlikely that negative, modal, directional and similar auxiliary verbs are nominalizations. To my knowledge, there are no languages that have nominalized auxiliary verbs” (ibid.: 300). This might be true if one takes the lexical view of nominalization meaning a word-derivation process which turns something into a lexical noun. However, if one allows for the grammatical view of nominalization, the phrases in (4) can still be nominalizations because there is ample cross-linguistic evidence for “nominalizations with no lexically derived noun” (see Comrie & Thompson 2007: 376), that is, exocentric constructions whose denotations do not come from one specific member of their components.

In sharp contrast, those who hold (2)b (e.g. Kaufman 2009a, 2009b) argue that Focus-words in conservative Philippine-type languages like Tagalog are inherently nominal and hence harmonious with their argument function. Even when serving as matrix predicates, they are still nominals, i.e. nominal predicates. Two pieces of evidence are often cited to support the nominalist hypothesis. First, different Focus-words denote entities with different thematic roles in a state of affair, depending on the Focus affixes involved, as in (5).

(5) Tagalog (Kaufman 2009b: 5)

- a. ang=b<um>ili
TOP=<AF>buy
‘the {buyer/one who bought}’
- b. ang=b<in>ili
TOP=<PF.RLS>buy
‘the (thing) bought’

- c. **ang=b<in>ili-han**
 TOP=<RLS>buy-LF
 ‘the (place) bought at’
- d. **ang=i-b<in>ili**
 TOP=CF-<RLS>buy
 ‘the one bought for’

Thus, Focus affixes are compared to argument nominalizers such as *-er* and *-ee* suffixes in English. Second, the basic structure in Tagalog is bipartite with two coreferential NPs, with the nominal predicate (including Focus-words) preceding the sole argument, and this gives a straightforward explanation for the “subjects-only” (or Topic-only in the present terminology) restrictions on question formation, relativization, and Topic fronting, all observed by numerous Philippine-type languages, as in (6).

(6) Tagalog (Kaufman 2009b: 4)

- a. **sino ang=b<um>ili ng=téla**
 who TOP=<AF>buy UND=cloth
 ‘Who bought the cloth?’ (cf. (1)b)
- b. **ang=babáe=ng b<um>ili ng=téla**
 TOP=woman=LIG <AF>buy UND=cloth
 ‘the woman who bought the cloth’
- c. **ang=babáe ay b<um>ili ng=téla**
 TOP=woman PTOPTOP <AF>buy UND=cloth
 ‘The woman, (she) bought cloth?’

All the three operations target at the Topic argument of the AF-word *b<um>ili* ‘<AF>buy’ (i.e. the buyer) because (6) would be unacceptable when constituents other than the buyer are questioned, modified, or fronted. Moreover, content-word interrogative sentences like (6)a are commonly analyzed as cleft-like structures much like (1)b above (e.g. Aldridge 2013). Accordingly, Focus-words are treated on a par with underived nouns. Not surprisingly, Austronesian nominalism, as Kaufman (2009b) calls it, has spurred much debate among Austronesianists and beyond. In fact, a complete issue of *Theoretical Linguistics* (Vol. 35.1) is dedicated to this topic. In a strong version of the nominalist

hypothesis Focus-words are always nominal, and the same nominal structure is used as matrix predicates, referential arguments, or restrictive modifiers of a nominal. If the best candidate to claim verbhood turns out to be constantly nominal, it follows that a system like this essentially “makes no use of the verbal category,” (Kaufman 2009b: 43), at least as far as the evidence goes in Tagalog.⁵⁰

The analyses mentioned above vary with respect to where to locate the source of nominality of denoting phrases consisting of Focus-words (e.g. *ang=nag-íngay* ‘TOP=AF.RLS-make.noise’ in Tagalog), which are in a paradigmatic relationship with uncontroversial nouns (e.g. *ang=áso* ‘TOP=dog’ in Tagalog). The answer ranges from an invisible but restorable noun, on which Focus-constructions are syntactically dependent, to nominal relation markers that introduce the whole phrase, and finally to Focus-words that are analyzed to be “inherently nominal.” An immediate challenge of the last analysis, which equates Focus-words with underived nouns on the lexical level, is how to account for those examples in (4) as raised by Reid (2002), where there seems to be no good candidate to attribute the nominality to.

The third type of analysis not only rejects the unnecessary distinction between nominal-like and verbal-like arguments (and hence between underived nouns and Focus-words analyzed as “headless relative clauses”) which only exists in the syntax of Philippine-type languages from the Indo-European point of view, as maintained in (2)a, but also refuses to accept total nominalism as advocated by (2)b, whereby there are no categorical distinctions whatsoever between nouns and verbs on the lexical level, a claim

⁵⁰ Nevertheless, Kaufman (2009b: 43) does end with a mitigated version of the nominalist hypothesis by saying “[t]he strength of the evidence presented, however, argues for at least some nominal component within the predicate phrase whether or not voice/aspect forms must be treated exactly on par with unambiguous nouns.”

that often becomes questionable when all relevant phenomena are considered (see Schachter & Shopen 2007: 5). As a compromise, supporters of (2)c (e.g. Himmelmann 2008) propose to solve the conundrum by distinguishing morphological word classes (nouns vs. verbs) from syntactic functions (predicates vs. arguments). What sets Philippine-type languages apart from Indo-European ones is then the fact that morphological and syntactic categories generally align with each other in the latter group but not in the former, where a general distinction can be made between verbs and nouns on the lexical level but syntactic functions are typically insensitive to lexical categories.⁵¹ In other words, the fact that a word form has the same argument function as an underived noun, as in (5), does not necessarily make it lexically a noun; otherwise one would have to say the negator and the preposition in (4) are both nouns, which no one ever seems to claim. In a similar vein, by distinguishing the morphological, syntactic, and semantic component of a word form, Spencer (2005) identifies six logically possible types of mixed categories (with deverbal nominals being a prominent member) across languages. Those particularly relevant to Focus morphology are what he calls “m-inert derivation” or “m-inert transposition” (where “m-inert” stands for “morphologically inert”), referring to cases where a morphologically defined category (i.e. Focus-words in Austronesian languages) maintains its morphology even when treated syntactically like another category (e.g. nouns marked by nominal relation markers). Once we suspend the assumption that morphological categories always have to line up with syntactic functions, which is more accurate in some languages than in others, the nominal-verbal controversy in Philippine-type languages becomes pointless and Philippine-type languages are no

⁵¹ On a related note, in his descriptive oriented work on nominalizations in Awetí, Drude (2011: 190) “implicitly argues that it is important to keep formal (morphological) properties apart from syntactic roles the forms may play, and both from functional uses.”

longer “unique” in this regard since many other languages also allow realis-indicative verbal forms to be treated like underived nouns on the phrasal level (e.g. Inuktitut languages; see Johns 1992).

In this present study, I adopt this last approach to the nominal-verbal controversy not only because it is typologically more informative, but also because in languages where verbal and nominal properties can be consistently distinguished, Focus-words, or Focus-constructions in general, typically demonstrate both, and thus deviates from underived (and uncontroversial) nouns.

3.1.2. Diachronic perspective: The nominalization-into-verb hypothesis

This section gives a brief summary of the diachronic perspective to Focus-words, most notably the nominalization-into-verb hypothesis (Starosta et al. 1982; widely known as the SPR hypothesis), on which the Nuclear Austronesian hypothesis (Ross 2009, 2012) is built.

One of the earliest reconstructions of Proto-Austronesian (PAN) verbs is Wolff (1973), who identified verbal paradigms such as Independent Non-past, Independent Past, Dependent, and Subjunctive, with each paradigm consisting of four affixes corresponding to the four-way distinction of Focus categories. Although there were some missing and mistaken forms in his reconstructions, his paradigms remain to be the basis on which subsequent works made additions or revisions (e.g. Dahl 1976; Starosta et al. 1982; Starosta 1992; Ross 1995, 2002). For instance, reproduced in Table 3.1 is the version according to Ross (2002), who renames Wolff’s paradigms.⁵²

⁵² There is one paradigm not included here, called Independent Non-past in Wolff (1973) and Indicative

Table 3.1: Proto-Austronesian verbal morphology (Ross 2002: 33)

	AF	PF	LF	CF
Indicative Neutral	*<um>STEM	*STEM-en	*STEM-an	*Si-STEM
Indicative Perfective	*<um><in>STEM	*<in>STEM	*<in>STEM-an	*Si-<in>STEM
Non-indicative Atemporal	*STEM	*STEM-u/-a	*STEM-i	*an-i + STEM
Non-indicative Projective	*<um>STEM-a	*STEM-aw	*STEM-ay	*an-ay + STEM

Among them, the two Non-indicative paradigms require further explanations. Atemporal forms are used in either imperatives or in contexts where the verb is syntactically dependent on “preverbs”, such as negators, TAM auxiliaries, or coordinate sequential markers (called co-subordinate forms in Paiwan by C. Wu 2013). Projective forms, on the other hand, express “irrealis events and states, i.e. intention, possibility and exhortation” (Ross 2002: 37). Interestingly, the two paradigms in both Indicative and Non-indicative are morphologically related. While Perfective forms are derivable from Neutral ones through the infixation of *<in> (except for the irregularity in PF; see Blust 1998b), Projective forms are derivable from Atemporal ones through the suffixation of *-a.

Durative in Ross (2002) because the forms in it are regularly derivable from Indicative Neutral undergoing *Ca-* reduplication. Unlike all the other Focus affixes, CF affixes in Non-indicatives were reconstructed as preverbal auxiliaries rather than suffixes, and this was mostly based on the fact that in Squaliq Atayal *anay* precedes a CF verb (marked by *s-*), which Ross considers to be a PAn situation. This detail should not concern us here. What is important, though, is that reflexes of **anay* are realized as suffixes in all languages that have them, including even Squaliq Atayal, where *s-STEM-anay* is an alternative form of *anay s-STEM* (see M. Yeh 2013: 62).

Despite the conventional term, reflexes of Perfective *<in> in modern languages often indicate not only perfective aspect but also past tense relative to a reference time.

One influential hypothesis was proposed by Starosta et al. (1982) to account for, among others, the morphological hodgepodge of the Indicative affixes in Table 3.1 (i.e. two infixes, two suffixes, and one prefix). They argued that these Indicative affixes were once “noun-deriving affixes in PAn” (ibid.: 148), and the derived nominals, with their thematic roles specified by these affixes, were used as nominal predicates in equational/identificational sentences. Over time, these nominalized predicates were reinterpreted as the default verbal forms, thus marginalizing the syntactic contexts applicable for predicates marked by Non-indicative affixes, which were presumably the erstwhile verbal forms. The nominalization-into-verb hypothesis helps to explain the widespread pattern of Indicative forms serving both predicate and argument functions. This process, should it happen in English, can be analogically summarized as going from ‘John is the beater of my child’ to ‘John beat my child’. The SPR hypothesis, in its core spirits, is then comparable to similar proposed developments of “verbal nouns” or “participles” having restructured the verbal system in diverse language families (see Sasse 2009).⁵³ Over the past three decades, the SPR hypothesis has been received with various degrees of favor by different scholars, from disbelief (e.g. Blust 1998) and skepticism (e.g. Ross 2002) to acceptance (e.g. Ross 2009, 2012) and total support (e.g. Kaufman 2009a, 2009b).

The organization in Table 3.1 had long been attributed to PAn until Ross (2009), who made a brand new claim on genetic subgroupings based on the SPR hypothesis as

⁵³ For instance, Gildea (1998) suggests that the indicative verbal forms with various tenses/aspects in many Cariban languages have evolved from argument or event nominalizations.

well as some new observations. Because the functional labels in Table 3.1 do not apply to all daughter languages, he referred to affixes in the lower two rows as first-generation verbal affixes and those in the upper two as second-generation ones.⁵⁴ Importantly, he argued that only second-generation affixes underwent the nominalization-into-verb reanalysis (hence the term “second-generation”), which took place not at the PAn period as previously believed, but was innovated at a later stage by a subgroup dubbed *Proto Nuclear Austronesian* (PNAn), which is ancestral to all modern Austronesian languages except for Tsou, Puyuma, and Rukai, collectively called *non-Nuclear Austronesian*.⁵⁵ His major evidence was that the verbal systems in the three non-Nuclear languages are more readily derived from the reconstructed first-generation affixes than from second-generation ones, and that second-generation Focus-words in non-Nuclear languages are only limited to nominal functions, in line with the idea of the SPR hypothesis that second-generation affixes were once nominalizers. The observed correlation between first-generation affixes with verbal functions and second-generation affixes with nominal ones in non-Nuclear languages suggests to Ross that the nominalization-into-verb reanalysis did not happen in PAn yet, but in PNAn instead, where the reanalysis was innovated. An immediate consequence of this new analysis, as Ross also argued, is that the number of first-order subgroups of PAn would be reduced from Blust’s (1999) ten (based on phonological evidence) to four (based on morphosyntactic evidence): Tsou,

⁵⁴ As a result, *⟨um⟩ is both a first- and second-generation verbal affix.

⁵⁵ The restatement here is a bit simplified because reconstructed verbal paradigms of PAn in Ross (2009) are much more sophisticated than those in Ross (2002). For instance, Atemporal forms were further split into Imperative and Dependent ones, which share identical forms in LF and CF, but not in AF and PF. And *Sa- is reconstructed for second-generation CF in addition to *Si-. In spite of these additional revisions, the restatement should be truthful to the major claim in Ross (2009).

Puyuma, Rukai, and Nuclear Austronesian, with the last group including all the other Formosan as well as all the Malayo-Polynesian languages.

Specific phenomena that uniquely distinguish Tsou, Rukai, and Puyuma from all the others will be addressed in §3.2.2/§3.3, §3.4, and §3.5 respectively.

3.2. Problems in previous studies

This section discusses three conceptual problems that are commonly found in previous studies of Formosan nominalization, including confusion of the internal and external syntax of nominalizations (§3.2.1), differential treatment of first- and second-generation Focus affixes (§3.2.2), and that of AF- and NAF-constructions (§3.2.3).

3.2.1. Confusion of the internal and external syntax of nominalizations

As mentioned in §3.1.1, the nominal-verbal debate concerning the predicate and argument function of Focus-words in Philippine-type languages often results from not distinguishing lexical categories that are morphologically defined from syntactic functions. One immediate consequence of equating syntactic functions with lexical categories is that properties of a construction are attributed to one syntactically dominant constituent of that construction, thus confusing the external syntax of a nominal structure, definable by its paradigmatic relationship with an uncontroversial noun, with its internal syntax, which can vary in complexity both within and across languages.

For instance, P. Li (2002: 234), a dedicated work on nominalization in Pazeh, concluded that “[m]orphology alone, therefore, does not tell whether a certain form is a

noun or a verb. We have to resort to syntactic evidence.” The type of syntactic evidence resorted to was illustrated by the contrast in (7).

(7) Pazeh (P. Li 2002: 234-235)

- a. **m-angit** lia ki rakihan
 AF-cry already TOP child
 ‘The child started to cry.’
- b. rakihan ka, sasay ki **m-angit**
 child PTOP why TOP AF-cry
 ‘As for the child, why did (he) cry?’

As P. Li (2002: 235) put it, “[i]nterestingly enough, when a verb follows a case marker and occurs in the position of [a] noun, it functions as a noun.” Thus, the Mstem *m-angit* was said to be a verb in (7)a and a noun in (7)b. However, as the syntactic complexity of the Topic argument increases, similar Mstems somehow become “verb phrases,” which are linked with a head noun they modify by the ligature *a* in the same manner as two underived nouns would, shown in (8).

(8) Pazeh (P. Li 2002: 235)

- a. nahani ki [**m-angid**] a [rakihan]
 AF.come TOP AF-cry LIG child
 ‘The child who cried came.’
- b. duila lia ki [**ma-siatu** rubahing] a [mamais]
 AF.leave already TOP AF.wear-clothes red LIG woman
 ‘The woman who wears red (clothes) has left.’

Both the ligature *a* and the nominal relational marker *ki* were taken to be “evidence that verb forms are nominalized” (ibid.: 236) in examples like (8). On the other hand, however, he hesitated to commit himself to the view that relative constructions in (8) are really nominalized simply because “there is no morphological evidence for... nominalization or that subordinate structures are nominalized in [Pazeh]” (ibid.: 237).

Some inconsistencies arise from the analysis above. First, it is dubious that the Mstem introduced by *ki* in (7)b is a noun whereas essentially the same form in (8)a is a

verb instead.⁵⁶ Second, if what is nominalized in (8) is the verb forms following the marker *ki* as claimed, where do the the “verb phrases” that modify a head noun come from? Third, it seems contradictory to maintain that only syntactic but not morphological evidence can tell the difference between a verb and a noun, but at the same time that syntactic evidence (such as the ligature *a* and the nominal relational marker *ki*) is not good enough for something being nominalized due to lack of morphological evidence. All the inconsistencies can be avoided if one distinguishes the external and internal syntax of complex denoting expressions like those in (8). Their external syntax is nominal because they are marked by *ki* and serve as the Topic argument like uncontroversial nouns. That is, they are in a paradigmatic relationship with underived nouns. Their internal syntax, however, varies from a single verb in (7)b to a verb phrase with an argument of its own in (8)b. Thus, the nominal property attributed to the Mstem *m-angit* in (7)b is better considered belonging to the construction the word form is part of rather than to the word form itself.⁵⁷ This may not be obvious in (7)b, where the denoting expression consists of only one single word form. In (8)b, however, it is clear what is nominalized is not the Mstem *ma-siatu* but the verb phrase it is part of. Otherwise, it would not be able to collocate with a patientive argument. P. Li’s (2002) reluctance to treat relative constructions in (8) as nominalized expressions in spite of the syntactic evidence that he identified also reflects the common lexical view of nominalization as a word derivation process. The lexical view inevitably seeks evidence of nominalization in

⁵⁶ The Mstem *mangit*, as in (7), becomes *mangid* in (8)a because of intervocalic voicing in the latter case (see Tsuchida 1993b: 302). Considering the sound change, the ligature *a* in Pazeh is most likely an enclitic like that in Paiwan. However, not having access to audio recordings of Pazeh, I present it here as a phonological free form, as was done in the cited work.

⁵⁷ See also P. Li’s (1973: 171) analysis of Taromake Rukai, where syntactic distributions of word forms were taken to be evidence for the lexical categories between nouns and verbs.

morphology, so it is not encompassing enough to subsume crosslinguistic phenomena constrained by similar principles under the same rubric.

Moreover, even when nominalization is taken beyond the lexical level, the boundary between lexical and higher-level nominalization is often mistaken because the external property of a denoting expression consisting of one single word form is attributed to that particular word form rather than the construction it is part of. For example, L. Huang (2002) distinguished lexical from clausal nominalizations in Mayrinax Atayal. Based on her analysis, clausal nominalizations occur in equational and relative constructions, as illustrated by the PF form in (9), where it was said to be a “clause” (ibid.: 207) serving as the Topic argument in the first example and the modifier of the Topic in the second.

(9) Mayrinax Atayal (L. Huang 2002: 207)

- a. pa-ka-shahiya cubalay=[ku **ta~tahq-un=nia**
 AF.IRR-K-delicious very=TOP.DEF IRR~cook.dish-PF=3SG.ACT
 ‘What {he/she} is going to cook will be very delicious.’
- b. pa-ka-shahiya cubalay=[ku raramat=ka **ta~tahq-un=nia**
 AF.IRR-K-delicious very=TOP.DEF dish=LIG IRR~cook.dish-PF=3SG.ACT
 ‘The dish that {she/he} is going to cook will be very delicious.’

However, it is not immediately clear why structures with the same external and internal syntax suddenly become “lexical nominalizations” in (10) below, aside from its free translations taken from the cited work.⁵⁸

⁵⁸ A glossing principle in the cited work seems to be that clausal nominalizations are glossed morpheme by morpheme whereas lexical ones are not. Thus, the original glosses for *nanubu'un* and *paptwaun* are “drink” and “work” respectively.

(10) Mayrinax Atayal (L. Huang 2002: 215)

- a. **balaiq=[ku na~nubu'-un=mu]**
 AF.good=TOP.DEF IRR~drink-PF=1SG.ACT
 ‘My drink is good.’
- b. **kia=[ku pa~ptwa-un=nia]**
 EX=TOP.DEF IRR~work-PF=3SG.ACT
 ‘{She/He} has work to do.’

In fact, one could translate (10)a in the same manner as (9)a, namely, “What I am going to drink is good.” There are just no principled ways to distinguish the proposed clausal nominalizations in (9) from lexical ones in (10) for at least two reasons. For one thing, both cases involve the same morphological processes, namely, *Ca-* reduplication, which marks irrealis, and PF *-un* suffixation, or schematically *Ca~Kstem-un*. Its realis perfective counterpart *<in>Kstem*, where *<in>* indicates past situation time relative to a reference time, is likewise used in both so-called “clausal” nominalizations, as in (11)a, and “lexical” ones, as in (11)b.

(11) Mayrinax Atayal (L. Huang 2002: 205, 216)⁵⁹

- a. **situinq=[ku b<in>ainay=[ni yaya]=[i ikuing]]**
 clothes=TOP.DEF <PF.PFV>buy=ACT mother=LOC 1SG
 ‘What Mother bought me is clothes.’
- b. **ma-hnuq=[ku b<in>ainay=[nku nabakis]]**
 AF-cheap=TOP.DEF <PF.PFV>buy=ACT.DEF elder
 ‘What the old {woman/man} bought was cheap.’

For the other, instances judged to be “lexical” nominalizations are semantically no less compositional than those considered “clausal” nominalizations, and both denote time-bound entities (cf. (9) through (11) above). In other words, putative lexical nominalizations do not have specialized semantics distinct from putative clausal ones.

⁵⁹ The original gloss for *binainay* in (11)b is “things.being.bought”, and the original free translation of (11)b has been revised. Nevertheless, what is at issue here is not so much how the target language should be glossed or translated into English as what rationale there is for the proposed distinction between lexical and clausal nominalizations given the grammar of Mayrinax Atayal.

This can also be seen in the wide range of denotations of the irrealis LF form *ga~ghapuy-an* ‘IRR-cook-LF’ (which was illustrated as a “lexical” nominalization; L. Huang 2002: 221), potentially referring to the micro-location for where food is cooked (“cooker”), the macro-location for the cooking event (“kitchen”), and any provisional place where someone will cook. Thus, lack of semantic constancy indicates the extemporized nature of putative lexical nominalizations.

Contradictory statements similar to those in P. Li (2002) can also be found in L. Huang (2002) because properties of the whole are attributed to parts of the whole. On the one hand, it was concluded that “it is legitimate to postulate, regardless of whether an element in Mayrinax Atayal is a verbal category or a nominal, that it is syntactically determined, instead of morphologically” (ibid.: 204). In other words, the claim was that syntax rather than morphology determines whether something is verbal or nominal. On the other hand, however, it was also stated that “verbal categories which are generally identifiable by their being affixed with [F]ocus or/and tense/aspect/mood markers can sometimes present the aforementioned nominal properties” (ibid.: 203). If the nominal-verbal distinction is made not by morphology as claimed, why can verbal categories be generally identifiable by certain affixes? The nominal properties in the quote refer to the possibility of being marked by nominal relation markers and taking up the argument position. The question is what these nominal properties belong to, a particular word form within a denoting expression or the denoting expression as a holistic construction? By dividing up lexical and clausal nominalizations without a vigorous basis, one would have to say the observed nominal properties belong to the PF-word in (11)b since it is a “lexical” nominalization, but to the construction of which the same word form is a part in

(11)a because that construction is a “clausal” nominalization. However, if nominal properties are consistently attributed to the external syntax of a denoting expression rather than to a specific member of its internal constituents, then the assumed distinction between lexical and clausal nominalizations would evaporate. The two delineated types are nothing but instantiations of the same grammatical process that creates “clausal nominalizations” in Comrie & Thompson’s sense (2007: 376), meaning “nominalizations with no lexically derived nouns.”

One could potentially argue that the nominalization in (11)a is more clausal than that in (11)b because the former has one extra argument than the latter (notice that this would not work for the putative distinction between (9)a and (10)a), but the internal syntax of a nominalization can be as complex as is permitted in an indicative sentence, as shown in (12) and (13).

(12) Mayrinax Atayal (C. Wu 2013: 112, 114)

- a. *ki'i*=[*i* **ba~bu'-un**]=[*ni* *watan*]=[*ku* *bauwak*]]
 PSSB=LIG IRR~shoot-PF=ACT W.=TOP.DEF pig
 ‘Watan will possibly shoot the pig.’
- b. *nanuwan*=[*ku* *ki'i*=[*i* **ba~bu'-un**]=[*ni* *watan*]]]
 what=TOP.DEF PSSB=LIG IRR~shoot-PF=ACT W.
 ‘What is it that Watan will possibly shoot?’

(13) Mayrinax Atayal (C. Wu 2013: 112, 115)

- a. *asi*=[*ki* **m<in>-usa**]=[*i* *bari*]=[*i* *watan*]]
 CERT=LIG AF<PFV>-go=LOC Miaoli=TOP W.
 ‘Watan must have been to Miaoli.’
- b. *ima*=[*ku* *asi*=[*ki* **m<in>-usa**]=[*i* *bari*]]]
 who=TOP.DEF CERT=LIG AF<PFV>-go=LOC Miaoli
 ‘Who is it that must have been to Miaoli?’

If the phrases introduced by *ku* in (9) and (10) are nominalizations, there is no good reason why those in (12)b and (13)b should not be. The nominalizations in (12)b and (13)c both have modals that qualify the validity of a state of affairs. Does that mean they

are even more clausal than the nominalizations in (9), both of which are already considered clausal? If so, we would end up having an infinite number of degrees of clausal nominalizations, depending on how many clausal elements (e.g. patientive arguments, modals, or spatio-temporal adverbials, etc.) happen to co-occur with a Focus-word. Maintaining that nominalizations in (10) are more lexical than those in (9) or (12)b and (13)b would be like claiming that the English noun *book* as in *a book* is more lexical than that as in *a very interesting book*, which is untenable. Nevertheless, different degrees of nominalization do exist in a single language, but they should be definable by restrictions on concomitant elements (e.g. the English nouniness *squish* by Ross 1973), and yet no similar restrictions can be found among argument nominalizations in Mayrinax Atayal (see §3.5.1 for a similar point in Puyuma).

Therefore, a simpler solution is to treat nominalizations in (9) through (13) all as exocentric constructions, which externally demonstrate nominal syntax due to their denoting power shared with underived nouns but internally no single word form is directly responsible for the nominal properties of the whole. Exocentric constructions of this nature are termed *grammatical nominalizations* by Shibatani (2009: 187), the denotations of which are often spatio-temporally anchored in an event. When a Focus-word is used within an argument nominalization, the nominalized argument is always the Topic NP that is gapped (see (12) and (13) above), and the same constraint is carried over when an argument nominalization modifies another nominal (i.e. the relative construction; cf. the two examples in (9) above). Thus, there is no need for postulating separate constraints in relative constructions (Shibatani 2009: 192). An overall analysis of grammatical nominalizations in Mayrinax Atayal not only accounts for the complexities

of their nominal-internal syntax but also finds good support in L. Huang’s (2002: 197) observation that “[e]xcept for [F]ocus and tense/aspect/mood markers...there are no productive morphological devices to produce lexical nominals; nominalized elements and verb forms are...identical.” That is to say, any Focus-word serving as the matrix predicate in non-imperative affirmative sentences, which allows TAM modifications and subcategorizes its co-occurring constituents, can potentially be put into use as an entity-denoting expression, which is then embedded in contexts where underived nouns are expected, serving as a referential NP or a restrictive modifier within an NP. Relevant word forms are schematically summarized in Table 3.2, and the predicate and argument functions of AF forms, which involve both the Mstem and Kstem, are illustrated in (14) through (17).

Table 3.2: Focus-words in non-imperative affirmative sentences in Mayrinax Atayal (after L. 1995, 2000a, 2002)

	AF	PF	LF	CF
Realis Neutral	Mstem	Kstem- <i>un</i>	Kstem- <i>an</i>	<i>si</i> -Kstem
Realis Perfective	< <i>in</i> >Mstem	< <i>in</i> >Kstem	< <i>in</i> >Kstem- <i>an</i>	< <i>in</i> >Kstem
Irrealis Future	<i>pa</i> -Kstem	<i>Ca</i> ~Kstem- <i>un</i>	<i>Ca</i> ~Kstem- <i>an</i>	<i>Ca</i> ~Kstem
Irrealis Optative	Mstem- <i>ay</i>	Kstem- <i>aw</i>	Kstem- <i>ay</i>	Kstem- <i>anay</i>

(14) Mayrinax Atayal: AF Realis Neutral (L. Huang 2002: 212)

- a. **m-aquwas**=[ku irawing=mu]
 AF-sing=TOP.DEF friend=1SG.GEN
 ‘My friend is singing.’
- b. **baq-un=mu**=[ku **m-aquwas**=ka haca]
 know-PF=1SG.ACT=TOP.DEF AF-sing=LIG DIST
 ‘I know that one over there who is singing.’

(15) Mayrinax Atayal: AF Realis Perfective

- a. **m-<in>aniq**=[cu bunga]=[’i sayun]
 AF-<PFV>eat=UND.INDF sweet.potato=TOP S.
 ‘Sayun ate sweet potatoes.’ (L. Huang 2002: 202)
- b. **ta~tuting-un=mi**=[ku xuil=ka **m-<in>aniq**=[cu siyam]]
 IRR~beat-PF=1SG.ACT=TOP.DEF dog=LIG AF-<PFV>eat=UND.INDF pork
 ‘I will beat the dog that ate pork.’ (L. Huang 1995: 211)

(16) Mayrinax Atayal: AF Irrealis Future (L. Huang 2002: 211)

- a. **pa-p-aquwas**=[ku irawing=mu]
 IRR-K-sing=TOP.DEF friend=1SG.GEN
 ‘My friend will sing.’
- b. **baq-un=mu**=[ku **pa-p-aquwas**=[ka haca]
 know-PF=1SG.ACT=TOP.DEF IRR-K-sing=TOP.DEF DIST
 ‘I know that one over there who will sing.’

(17) Mayrinax Atayal: AF Irrealis Optative (L. Huang 2002: 201)

- m-aras-ay**=ci=[cu guqiluh]
 AF-bring-OPT=1SG.TOP=UND.INDF banana
 ‘I’d like to bring bananas.’

The only exception is Irrealis Optative forms in the last row of Table 3.2, all marked by first-generation affixes such as *-aw*, *-ay*, and *-anay*. L. Huang (2002) pointed this out, but offered no explanation as to why they have predicate but no argument functions. I will speculate on a possible motivation for such a restriction on Irrealis Optative forms later in §4.4.1, where cognate word forms in other Formosan languages are also taken into consideration.

Notwithstanding the arguments against L. Huang's analysis of lexical nominalizations, grammatical nominalizations (or grammatical constructions in general) do undergo lexicalization if they are entrenched enough and have acquired specialized semantics, much as the lexicalization of verb-applicative-marker combinations (see Peterson 2007: 169) or that of provisional epithets into conventional terms (e.g. *redbreast* in English). Lexicalization, however, has to be supported by evidence such as phonological coalition or semantic idiosyncrasy (see M. Yeh 2011: 572 for evidence of this nature in Saisiyat), rather than judged by whether something can be readily translated into an English noun or what clause-like elements can or cannot be found in a nominalization. Presumably, grammatical constructions consisting of one single word form are especially prone to lexicalization because “[t]he stripping away of argument structure... is a major conduit for the reanalysis of nominals derived from verbs into basic nouns.” (Foley 2014: 5) However, a construction is not necessarily lexicalized simply because it happens to be comprised of one single word form. Conversely, lexicalization can happen irrespective of the internal complexity of a construction, such as *methinks* or *whachamacallit* in English. Therefore, instead of treating all nominalizations made up of one single Focus-word as lexical, it is suggested that their lexical status should be more constrained and be investigated case by case.

3.2.2. Differential treatment of first- and second-generation affixes

Collapsing the internal and external syntax of nominalizations also leads to differential and unfair treatments of what constructions should or should not count as nominalizations in the first place. This section first discusses the contrast between first-

and second-generation Focus affixes, and the next one (§3.2.3) deals with the asymmetry between AF- and NAF-constructions.

Among the reconstructed Focus affixes in PAn (see Table 3.1), the AF *<um> (or its allomorphs) is the only one reflected in all Formosan and Philippine languages, and also the only one that is both first- and second-generation affixes in the Nuclear Austronesian hypothesis. Due to its non-distinguishing nature, reflexes of AF *<um> are not discussed in this section, but will be included in the next one, where the AF-NAF asymmetry is discussed.

Many conservative Philippine-type languages have both first- and second-generation NAF affixes, but there are at least two extreme examples where only either set is attested. In (standard) Tagalog, only second-generation NAF affixes are found.⁶⁰ Word forms with these affixes in Tagalog have both predicate and argument functions, and specialists of Tagalog are often happy to include them in studies on nominalizations. Kaufman (2009a, 2009b) even made the radical claim that all apparent verbs in Tagalog are in fact nominals and that the language makes no use of the verbal category whatsoever (see §3.1.1). Relevant examples have been shown and are repeated in (18) and (19).

⁶⁰ Kaufman (2009a) mentions that second-generation NAF affixes are preserved in some provincial dialects of Tagalog.

(18) Tagalog (Kaufman 2009b: 3) [= (1) in Chapter 2]

- a. k<in>áin ng=púsa ang=dagà sa=pinggan pára sa=áso
 <PF.RLS>eat ACT=cat TOP=rat OBL=plate for OBL=dog
 ‘The cat ate the rat on the plate for the dog.’
- b. k<in>áin-an ng=púsa ng=dagà ang=pinggan pára sa=áso
 <RLS>eat-LF ACT=cat UND=rat TOP=plate for OBL=dog
- c. i-k<in>áin ng=púsa ng=dagà sa=pinggan ang=áso
 CF-<RLS>eat ACT=cat UND=rat OBL=plate TOP=dog

(19) Tagalog (Kaufman 2009b: 5) [= (5)]

- a. ang=b<in>ili
 TOP=<PF.RLS>buy
 ‘the (thing) bought’
- b. ang=b<in>ili-han
 TOP=<RLS>buy-LF
 ‘the (place) bought at’
- c. ang=i-b<in>ili
 TOP=CF-<RLS>buy
 ‘the one bought for’

By contrast, in Tsou, only first-generation NAF affixes are productive (more specifically those deriving from the Dependent set in PAN as reconstructed by Ross 2012: 1264), and specialists of Tsou (e.g. M. Chang 2002; Zeitoun 2005) typically exclude word forms with these affixes from the domain of nominalizations because they are “verbal” and thus involve only relativization. However, Focus-words in Tsou are used in exactly the same constructions where those in Tagalog are called for, including at least “relative clauses, cleft equational sentences, and content interrogatives” (Starosta 1985/2009: 507; examples ensue). Moreover, all these constructions in both Tsou and Tagalog are subject to the same constraints whereby only the Topic argument can be relativized, clefted, or questioned (or more commonly known as the “subjects-only constraint”). It is then worth asking why Tsou deserves a special treatment and what makes something verbal in Tsou in the first place.

Zeitoun (2005) identified one morphological and three syntactic criteria for verbs as a word class distinct from nouns in Tsou. Morphologically, verbs in Tsou are marked by Focus affixes, but this does not make Tsou unique since it is equally true of all the other Philippine-type languages. Syntactically, Tsou verbs have three properties: (i) they are obligatorily preceded by auxiliary verbs (called “beginners” by T. Tung 1964) that grammatically encode AF/NAF distinction and TAM information; (ii) they never host person-form indexes whereas nouns do; and finally (iii) nouns “are always preceded by a case marker... verbs never are, unless they modify a noun.” (ibid.: 264) The first property of Tsou verbs is illustrated in (20), where they function as matrix predicates, similar to Tagalog Focus-words in (18).

(20) Tsou (Zeitoun 2005: 284)

- a. $i=si$ **teaph-a**=[ta skayʉ]=[to ino]=[to oko]
 NAF.RLS=3SG.ACT put-PF=OBL cradle=ACT mother=TOP child
 ‘Mother put the child into a cradle.’
- b. $i=si$ **teaph-i**=[to oko]=[ta ino]=[ta skayʉ]
 NAF.RLS=3SG.ACT put-LF=UND child=ACT mother=TOP cradle
 ‘Mother put the child into the cradle.’
- c. $i=si$ **teaph-neni**=[to tacʉmʉ]=[to ino]=[’e oko]
 NAF.RLS=3SG.ACT put-CF=UND banana=ACT mother=TOP child
 ‘Mother put bananas (into something) for the child.’

The contrast between verbs and nouns as stated in (ii) is shown in (21), where the noun *amo* attracts person-form indexes but not the verb *su’nov-a*.

(21) Tsou (G. Lin 2010: 183)

- $i=si$ **su’nov-a**=[to naau]=[’o amo=**si**]
 NAF.RLS=3SG.ACT angry-PF=ACT N.=TOP father=3SG.GEN
 ‘Naau is angry at her father.’

However, the verbal properties in (i) and (ii) are in fact correlated. It is precisely because preverbal auxiliaries are obligatory that person-form indexes, which are essentially Wackernagel second-position clitics (see §2.3.3), attach to them rather than to lexical

verbs. In other Philippine-type languages, person-form indexes commonly attach to lexical verbs or preverbal elements, depending on whether the latter are present or not.⁶¹ Finally, the two-part claim as stated in (iii) is not entirely correct. For one thing, while it is true that nouns serving as arguments are always preceded by nominal relation markers (or case markers as in the quote above), those that function as predicates are not.⁶² Compare, for instance, the possessive NP as an argument in (21) with the one as the predicate in (22).

(22) Tsou (Zeitoun 2005: 266)
 zou 'oko=su=[na a'o]
 EMPH child=2SG.GEN=TOP 1SG
 'I am your child.'

For the other, which is more important, Focus-words in Tsou are just as free to be marked by nominal relation markers as those in other Philippine-type languages including Tagalog, even if they *do not* modify a noun. A verbal phrase marked by a nominal relation marker does modify a noun in (23), but definitely not in (24), which is contrary to the statement in (iii).

⁶¹ Even person-form indexes in Tsou show some degrees of mobility by attaching to preverbal elements other than those auxiliary verbs identified by Zeitoun (2005: 268). Moreover, the mobility of person-form indexes is also found in cases where they either attach to an auxiliary that precedes a special word form (which would be verbal in this context; schematically AUX=PRO X) or to that very same word form when auxiliaries are not available (which would be nominal in this context; schematically X=PRO). See §3.3 for details.

⁶² In fact, according to M. Chang (2004: 257), nominal relation markers are prohibited from occurring before nominal predicates. Another situation where nouns are not marked by nominal relation markers is when they are incorporated with generic action verbs such as “make” and “take, pick” (e.g. *moeoeai yongku* ‘AF.make basket’; H. Huang 2010: 241). Incorporated nouns denote non-specific entities, and it is an open question whether they should be analyzed as syntactic arguments or part of the verbal complex.

(23) Tsou (Zeitoun 2005: 273)⁶³

os='o eobak-a=['o i=si pasunaenv-neni=[to
NAF.RLS=1SG.ACT beat-PF=TOP NAF.RLS=3SG.ACT sing-CF=ACT

paicɿ]]=ci cou
P.=LIG man

'I beat the man that Paicɿ sang for.'

(24) Tsou (G. Lin 2010: 254)

i=ko haf-a=['o i=si toa-i=[to
NAF.RLS=2SG.ACT take-PF=TOP NAF.RLS=3SG.ACT pick-LF=ACT

ino] **nehucma]**
mother yesterday

'Did you bring what Mom picked yesterday?'

Examples like (24) do not just come from elicitations, but are widely found in naturally occurring texts (see Huang et al. 2001: 77-81 as well as the appendix therein). One interesting example taken from those texts is (25), where neither of the two verbal phrases that make up NPs modifies a noun.

(25) Tsou (S. Huang et al. 2001: 99; also cited in H. Huang 2010: 52)

... 'a=['i'o mo
AFF=TOP AF.RLS

... **mɿm'ɿ]=[to hana]]=['o i=si a'ɿmt-a ɿmɿɿ-a]**
grow=UND flower=TOP NAF.RLS=3SG.ACT really-PF like-PF

'The one he really loves is the one who grows flowers.'

In other words, the argument function of Focus-words in Tsou is no more syntactically dependent on a so-called head noun than the argument of Focus-words in her sister languages. Given the right context, all the verb forms in (20) can assume argument functions, as in (26), similar to (19) in Tagalog.

⁶³ The CF form for "sing" in the cited work is *pasunaenoveni*, which is most likely a typo. Its AF form is *pasunaeno*, which due to the regular sound change of /o/ to /v/ alternates with *pasunaenv-*, the stem for all NAF forms (or the Kstem in Ross's 2015 term), including PF *pasunaenv-a* and CF *pasunaenv-neni* (G. Lin 2010: 76, 207).

(26) Tsou (after Zeitoun 2005: 284)

- a. 'o=i=si **teaph-a**
 TOP=NAF.RLS=3SG.ACT put-PF
 'the (thing) put into.'
- b. 'o=i=si **teaph-i**
 TOP=NAF.RLS=3SG.ACT put-PF
 'the (place something) is put into.'
- c. 'o=i=si **teaph-neni**
 TOP=NAF.RLS=3SG.ACT put-CF
 'the one (something) is put into for.'

Therefore, there is really no good reason why Tagalog examples in (19) are nominalizations but Tsou ones in (26) are not, considering the relevant constructions in both languages semantically denote event participants and syntactically serve as arguments.

Nevertheless, one could still argue that Tsou Focus-words are verbal and that no single word form in (26) is nominalized. Both are true in terms of the internal syntax of (26). However, when interacting with higher-order constituents, verbal phrases as a whole are in a paradigmatic relationship with uncontroversial nouns, and it is in this sense that the former are nominalizations. Verbal phrases in Tsou are just as readily to take up any argument position as nouns, as compared in (27).

(27) Tsou (Fieldnotes)

- a. mo yonghu=he=[ta ceopngu=su]=[e ceopngu='u]
 AF.RLS AF.pretty=CMPR=OBL hat=2SG.GEN=TOP hat=1SG.GEN
 'My hat is prettier than your hat.'
- b. mo yonghu=he=[ta os=ko **phin-i]**=[e
 AF.RLS AF.pretty=CMPR=OBL NAF.RLS=2SG.ACT buy-LF=TOP
 os='o **phin-i]**
 NAF.RLS=1SG.ACT buy-LF
 'What I bought is prettier than what you bought.'

As a result, what makes Tsou unique comes down to the fact that its Focus-words are obligatorily preceded by auxiliary verbs, which then attract person-form and other types

of clitics. In fact, bound person forms on nouns and those on auxiliaries, which index the possessor and the (Topic or non-Topic) Actor as in (27)a and (27)b respectively, are identical in form with only one exception:⁶⁴ the former use =*taini* and the latter =*ta* for visible non-SAPs (see Zeitoun 2005: 278). In languages like Tagalog, person-form indexes in situations like (27)b would attach to Focus-words, as they do to underived nouns. Within the grammar of Tsou, it may be reasonable to say that its Focus-words are more verbal than its nouns. However, the crucial question is whether it is sensible to maintain that Focus-words in Tsou are somehow more “verbal” than those in Tagalog such that the former involve relativization but the latter comprise nominalization when they are both used to form denoting expressions? The answer should be a clear no since what uniquely distinguishes Focus-words in Tsou from those in Tagalog is the obligatory presence of auxiliaries, which is simply not a grammatical requirement in Tagalog. A crosslinguistic comparison like this is just as unfair as saying English verbs are more “verbal” than Chinese ones because the former inflect for tense but the latter do not. Therefore, the internal structure of denoting expressions in a language may resemble noun or verb phrases as defined by the grammar of that language, but across languages those defining criteria are not readily comparable. What is comparable, however, is how languages construct linguistic expressions that denote various event participants as characterized by a state of affairs (or argument nominalizations for short). More importantly, argument nominalizations with first-generation affixes in Tsou hold an appositional or equational relationship with the nominal they modify (i.e. relative constructions) or with the nominal that is predicated of them (i.e. cleft-like and content-

⁶⁴ The different forms for 1SG and 2SG across the two examples in (27) are conditioned by phonology and have nothing to do with whether they attach to nouns or auxiliary verbs.

word interrogative constructions), thus obeying what is commonly called “subject-only” constraint in the same manner as nominalizations with second-generation affixes in Tagalog and many others. For instance, the interrogative pronoun *sia* ‘who’ is predicated of forms as simple as a pronoun or as complex as an argument nominalization denoting the Topic nominal, whose semantic role is indicated by an LF-word, as in (28).

(28) Tsou

- a. *zou sia=[na suu]*
 EMPH who=TOP 2SG
 ‘Who are you?’ (Zeitoun 2005: 283)
- b. *zou sia=[na i=si ait-i=[ta pasuya]]*
 EMPH who=TOP NAF.RLS=3SG.ACT see-LF=ACT P.
 ‘Who is it that Pasuya saw?’ (M. Chang 2004: 169)

Hence, first-generation NAF affixes in Tsou serve the same function as second-generation ones in languages that have them, and there is no good basis on which to exclude the former from the study of nominalization.

There is yet one last concern about analyzing those Tsou argument-denoting constructions in (24) through (26) as nominalizations. Starosta (1988a) identified two types of relative constructions among Formosan and Philippine languages. In his “equational” type, (second-generation) Focus nominalizations (which are lexical ones in his view) are either juxtaposed or linked to the head noun with a ligature, as in all Philippine languages as well as all Formosan ones but Tsou. Relevant examples from Tagalog are illustrated in (29), where Focus nominalizations and head nouns can swap positions around the ligature.

(29) Tagalog (Aldridge 2004b: 100)

- a. libro=ng b<in>ili ni=maria
 book=LIG <PF.RLS>buy ACT=M.
 ‘the book Maria bought’
- b. b<in>ili ni=maria=ng libro
 <PF.RLS>buy ACT=M.=LIG book
 ‘the book Maria bought’

The second type, called “verbal”, was exclusively attributed to Tsou. However, aside from the “verbal” issues discussed earlier, there is no good reason why relative constructions in Tsou are not “equational” like those in Tagalog. The verb phrases in Tsou either precede or follow the head noun they modify, with the marker *ci* occurring in between in either case, as shown in (30) (see (49) below for another example from a spontaneous narrative).

(30) Tsou (Zeitoun 2005: 273-274; citing H. Chang 1998: 69)

- a. o='u=cu ait-i=['o o=si tpos-i=[to
 NAF.RLS=1SG.ACT=already see-LF=TOP NAF.RLS=3SG.ACT write-LF=ACT
 pasuya]=ci tposɤ]
 P.=LIG book
 ‘I have read the book that Pasuya wrote.’
- b. o='u=cu ait-i=['o tposɤ=ci o=si
 NAF.RLS=1SG.ACT=already see-LF=TOP book=LIG NAF.RLS=3SG.ACT
 tpos-i=[to pasuya]]
 write-LF=ACT P.
 ‘I have read the book, which Pasuya wrote.’

Like Tagalog *ng* (or its allomorph *na*), the Tsou marker *ci* is not permitted unless it connects two constituents, with one modifying the other. However, some might still argue that the Tsou marker *ci* is a “relativizer”, which differs from the “ligature” in Tagalog. The difference is that the Tagalog ligature freely connects all kinds of modifiers with their modifyees, but the Tsou marker *ci* is only restricted to verbal modification, distinct from nominal modification (Szakos 1994: 78; Zeitoun 2005: 273). Since the Tsou

marker *ci* does not link two nouns, it follows that the construction in (30) cannot be equational and that the formal distinction between verbal and nominal modification in Tsou is a legitimate reason for distinguishing relativization (headed or not) from nominalization. Many examples have been tested to confirm this in previous studies, but it only takes a few right ones to refute it. It turns out that the marker *ci* not only links two (lexical) nouns (by the criteria in Zeitoun 2005), but also shows the type of semantic restrictions on the two nominals it links which make them “equational” in a stricter sense. Examples in (31) are illustrative.

(31) Tsou (Fieldnotes)

- a. 'a **oko=ci pasuya**=[’e t<m>oycɛ=[ta evi]]
 AFF child=LIG P.=TOP <AF>chop=UND tree
 ‘The one who chopped trees is the child Pasuya.’
- b. na’no yonghu=[’e **lema’coho=ci mamespingi**]
 AF.very AF.pretty=TOP teacher=LIG woman
 ‘The female teacher is very pretty.’

The two nominals linked by *ci* have to be appositional, as in (31). Likewise, the grammatical nominalizations consisting of verb phrases are appositional with the nouns they modify, as in (30). Due to this semantic restriction, linking a place name and a person, for instance, is not allowed in Tsou, but is permitted in languages like Amis (e.g. *takaw a tamdaw* (Kaohsiung LIG person) ‘person born and raised in Kaohsiung’; J. Wu 2012). In this sense, relative constructions in Tsou should be just as “equational” as those in her sister languages, if not more so. More importantly, just as it is nonsensical to maintain that the modifying noun in (31) is “headless” elsewhere when there is no accompanying modified noun, so is the belief that the verb phrases serving as arguments in (24) through (26) are “headless relative clauses” in disguise.

This section has contrasted a differential treatment of first- and second-generation NAF affixes across two extreme languages. On the one hand, Tagalog NAF Focus-words only reflect second-generation affixes and have been claimed to be highly “nominal.” On the other, Tsou NAF Focus-words only reflect first-generation affixes and have been assumed to be “clearly verbal.” This situation is rather odd considering the overlapping functions of NAF Focus-words in both languages, which in fact can be as “verbal” as they are allowed by either language when they serve as arguments or modifiers of arguments. More importantly, the differential treatment is unwarranted because the unique defining property of something being verbal in Tsou is simply not a grammatical requirement in Tagalog. From a comparative point of view, Focus-words in Tsou constitute constructions that are externally as nominal as underived nouns, and as such these constructions are no less nominalized than those in Tagalog or other Philippine-type languages.

Moreover, excluding first-generation Focus-words in Tsou from the study of nominalization simply because they are preceded by auxiliaries appears even more unfair considering similar constructions with second-generation Focus-words are also found in Atayalic languages. For instance, Focus-words in Squaliq Atayal are optionally preceded by aspectual auxiliaries, which attract person-form indexes just like auxiliary verbs in Tsou, as in (32).

(32) (Jianshi) Squaliq Atayal (M. Yeh 2013: 249)

- a. **hluy-un=mu** qu ruma kira
 pull-PF=1SG.ACT TOP bamboo later
 ‘I will pull bamboos later.’
- b. **wal=mu** **hluy-un** qu ruma la
 PFV=1SG.ACT pull-PF TOP bamboo PRT
 ‘I pulled bamboos.’

Crucially, with or without preverbal auxiliaries, Focus-words in Squliq Atayal can as readily be part of a denoting phrase as those in Tsou. This is demonstrated in (33), where the denoting phrase fulfills an entire NP in the first example and modifies another nominal within an NP in the second.

(33) (Jianshi) Squliq Atayal⁶⁵

- a. *blaq truq-an qu [pmzy-an=mu ramat]*
 AF.good dig.with.a.hoe-LF TOP plant-LF=1SG.ACT vegetable
 ‘(The field) where I planted vegetables is easy to to plow.’ (M. Yeh 2013: 192)
- b. [*wal bhy-an na yumin*] *ka kneril ga, cyux m-’uyay*
 PFV beat-LF ACT Y. LIG woman PTOP PROG AF-hungry
 ‘As for the woman that Yumin beat, (she) is hungry.’ (A. Liu 2005: 100)

However, specialists of Atayal do not seem to have issues with granting Focus-words like those in (33) nominalization functions,⁶⁶ which is in sharp contrast with the commonly held view that Focus-words in Tsou have nothing to do with nominalization.

Finally, there is yet one more reason to include Focus-words in Tsou in the present study of nominalization. This is because under certain circumstances they do derive word forms that are syntactically indistinguishable from underived nouns in the language, a process that resembles grammatical nominalizations having undergone lexicalization in other Philippine-type languages. This is to be demonstrated in §3.3.

The unjust differential treatment of first- and second-generation NAF affixes is found not only across but also within languages. This section has contrasted the former type, specifically between Tagalog and Tsou. An illustration of the latter type is illustrated by Rikavung Puyuma in §3.5.2.2.

⁶⁵ If the Actor in (33)b were SAPs, person-form indexes would obligatorily attach to the aspectual auxiliary as in (33)a.

⁶⁶ In fact, the original gloss for *-an* in (33)a is specifically *LOCNMZ*, which stands for Locative nominalizer.

3.2.3. Differential treatment of AF- and NAF-constructions

The second type of differential treatment in Philippine-type languages is between AF- and NAF-constructions due to their conspicuous morphosyntactic asymmetry. The contrast was probably first pointed out by Lopez (1941). On the basis of Tagalog and to a lesser extent Sangir, Capell (1964) later argued that NAF-words, but not their AF counterparts, are “verbal nouns” because the (non-Topic) Actor of the former is encoded in the same way as the possessor of underived nouns, which is formally distinct from the Topic-cum-Actor of the latter. Over the past half century, this observed Actor-possessor isomorphism, together with the notion of finiteness (defined as compared with “ordinary” verbs in affirmative indicative sentences), has continued to be the deciding factor based on which later researchers determine whether Focus-words are nominal or verbal, which then in turn has consequences for whether one analyzes a Focus-construction as nominalization or relativization.

Some central claims taken from the literature on Formosan nominalization are quoted in (34) below.

(34) Previous views on nominalization vs. relativization in Formosan languages

- a. D. Liu (1999: 151; emphasis mine) on Central Amis⁶⁷: “the Amis cleft clauses and the Amis relative clauses...obligatorily undergo the process of nominalization. This linguistic fact is not common [...]cross-linguistically. Syntactically speaking, both cleft clauses and relative clauses are clauses with missing gaps inside in nature. Compared to other Formosan languages, *this obligatory nominalization for the “incomplete” clauses, the cleft clauses and the relative clauses, is language-specific.*”
- b. H. Chang & A. Lee (2002: 364; emphasis mine) on Kavalan: “[T]he nominalization-relativization distinction is not trivial... In Kavalan, *nominalization and headless relativization are grammatically and semantically distinct from each other.* Nominalization can be marked with the suffix *-an* and relativization is marked with the enclitic *=ay*; the affected verbs in nominalization cannot take accusative noun phrases as their complements, but verbs in relativization can; *nominalization* turns its hosts into *arguments*, while *relativization* turns them into *modifiers*. The Kavalan evidence also indicates that *a functional definition of nominalization is inadequate.*”
- c. S. Teng (2008: 105; emphasis mine) on Nanwang Puyuma: “In Puyuma, NP_{rel} is always a gap, but two different RC strategies are utilised according to whether the NP_{rel} is an actor or not. *If the NP_{rel} has the role of actor, then the RC is manifested as a finite clause; if not, then the RC is a nominalised clause.*”⁶⁸
- d. L. Sung (2011: 531-533; emphasis mine) on Budai Rukai: “Nominalized clauses with *-anə* or *-∅* are considerably pervasive in Budai; furthermore, they correlate to relativization in interesting ways...*When the relativized argument is the subject... the relative clause is essentially identical to the ordinary predicative clause counterpart... and no nominalization is involved...When it is the non-subject argument...that gets relativized, this is where -∅ (zero) nominalization comes into play... The same observation is attested in the -anə type of nominalization as well.*”⁶⁹

First, according to (34)a, Central Amis is unique among Formosan languages in that its relative and cleft constructions with missing gaps ought to undergo “obligatory nominalization”, which is “not common [...]cross-linguistically”, but similar constructions in other Formosan languages do not involve nominalization. However, the claim results from, again, confusing the internal and external syntax of an argument-

⁶⁷ There are two instances of p.151 in the cited work. The quote comes from the one in Chapter 5.

⁶⁸ Exactly the same claim was also made in Saaroa by J. Pan (2012: 289-290).

⁶⁹ Since Rukai has only an impoverished Focus system, AF- and NAF-words in it should be understood as word forms that contain reflexes of the PAn AF *<um> and LF *-an respectively. Here I am following the recent reconstructions in Ross (2015c), where PAn *<um> is reflected as *u- in Proto-Rukai. The morpheme is found in the realis verbal form *wa*-STEM in Budai/Taromake/Labuan/Tona (where *wa*- comes from *<um> plus *a-), and *u*-STEM in Maga as well as *o*-STEM in Mantauran (where *u*- or *o*- comes from *<um>).

denoting construction. What is nominal about Amis relative and cleft constructions is their external distributions within higher-order syntax, in terms of which similar constructions in most other Formosan languages are just as nominal as underived nouns are allowed in each language. More arguments against (34)a as well as more details on Amis nominalization will be presented in Chapter 5.

Although a view like (34)a, which singles out Amis from other Formosan languages, is not commonly shared among Formosanists, its underlying assumption surely is, namely, that there are valid crosslinguistic criteria by which one can distinguish nominalized relative or cleft constructions from those that are not. Even more prevalent is the view in (34)b that nominalization and headless relativization are distinct operations with the former creating arguments and the latter modifiers of arguments. However, the basic distribution facts as described by H. Chang & A. Lee (2002) are not even a fair characterization of the Kavalan language. As shown in the quote, three points were made to support the claim that word forms marked by *=ay* are relativized whereas those by *-an* are nominalized. The first is the different marking between the two, which does not explain much because the functions of these morphemes are precisely what is at issue to begin with. The second point has to do with whether word forms marked by these two formatives can take an “object noun phrase” (i.e. the non-Topic and non-Actor patientive Undergoer; yes for *=ay* but no for *-an*), and the third with the syntactic functions of these word forms (modifiers for *=ay* and arguments for *-an*). However, the latter two points have been demonstrated by F. Hsieh (2011) to be non-differential with respect to the two markers in question, thus invalidating the conclusion in (34)b.

There is, however, a robust asymmetry between Kavalan *=ay* and *-an*. That is, the former is used in AF-constructions whereas the latter in NAF ones, which brings us to the shared claim in (34)c through (34)d. Specifically, attributive phrases involving Focus-words are either “finite”, that is, “ordinary predicative clauses”, or “nominalized clauses,” depending on the grammatical relation of the relativized argument. In Nanwang Puyuma, the RC is said to be a “finite clause” when the relativized argument is the Actor, but a “nominalized clause” instead when the relativized argument is a non-Actor. Likewise, in Budai Rukai the RC is claimed to involve no nominalization when the relativized argument is the subject, but is nominalized instead when the relativized argument is anything other than the subject. In this view, finiteness is a feature exclusively reserved for relativization, and nominalization only happens when finiteness is somehow lost. The distinction between nominalization and relativization then hinges upon the notion of finiteness, which is notoriously difficult to define in a principled manner across languages (Maas 2004; Aikhenvald 2011: 264). Moreover, it is somehow believed that AF-words are essentially more “finite” than NAF ones regardless of their syntactic functions and thus the former can only be relativized if they are to modify a noun, which can even be absent, thus giving rise to so-called headless relative constructions. At the core of this belief is the implicit notion that “headless relative construction” and “nominalization” can be, and should be, distinguished in Formosan languages.

To examine the validity of the claim in (34)c through (34)d, I shall compare AF- and NAF-constructions side by side in terms of their referential and restricting functions and examine whether it is motivated to draw the line between “finite” and “nominalized”

clauses, and consequently between relativization and nominalization. It should be pointed out that the question is not so much about whether a given structure should be called headless relative clauses or nominalizations (which sometimes turns into a terminological debate).⁷⁰ Rather, the more crucial question is what language-particular and cross-linguistic criteria there are to sustain a meaningful distinction between the two terms.

Himmelmann (2005a: 127) points out that in western Austronesian languages “[s]ometimes a distributional distinction between nouns and verbs pertains only to one or two fairly specific syntactic contexts.” One such context is negation, which will play an important role in the following discussions.

3.3. Tsou

The Tsou language is mostly spoken around the Mt. Ali area (or *psoseongana* in the vernacular, literally meaning “pine forest”) across seven major villages in Chiayi County, and one in Nantou County. The language is usually classified into three distinct varieties in the literature, including TapangꞰ, Tfuya, and Luhtu, which vary mostly in lexicon and phonology, but not much in morphosyntax (T. Tung 1964; P. Li 1979). Both of my Tsou consultants spent most of their childhood in TapangꞰ Village, so my own Tsou data specifically belong to the TapangꞰ variety unless otherwise indicated, which also seems to be the variety on which most previous studies are based.

⁷⁰ For instance, Martínez Fabián & Lagendoen (1996) analyzed suffixes *-me* and *-’u* in Yaqui (Uto-Aztecan) as a marker for nominalization and relativization respectively. However, González (2012) treated both as marking nominalization. See §4.2 for relevant examples in Yaqui.

This section discusses Tsou nominalizations not covered in §3.2.2, including analytic circumstantial nominalizations (§3.3.1) and lexical argument nominalizations (§3.3.2).

3.3.1. Analytic circumstantial nominalizations

As has been shown in §3.2.2, grammatical argument nominalizations in Tsou consist of auxiliary verbs followed by verb phrases with their arguments, and they denote the Topic argument as indicated by Focus affixes on the verb. In addition, Tsou has such analytic constructions as those in (35) for circumstantial nominalizations denoting manner/means/degree and reason/cause, for which other Formosan languages would use synthetic word forms instead (e.g. Amis *sa-ka-ulah* (CA-K.INT-like) ‘reason for liking’; see Chapter 5 for details).

(35) Tsou⁷¹

- a. **te=ko=la** **aezuh-a=[’o** **la=ko** **hia** **aomotʰʰ]**
 NAF.IRR=2SG.ACT=DSTT change-PF=TOP AF.HAB=2SG.TOP how AF.talk
 ‘You should change the way you (normally) talk.’
 (ODIL; under the entry *aezuha*)
- b. **cuma=[na** **mi=ko** **kua** **mongoi]**
 what=TOP AF.RLS=2SG.TOP why AF.leave
 ‘What is the reason you left?’ (Fieldnotes)

These constructions have been analyzed by M. Chang (1998, 2002) as “internal-head relative clause,” as if word forms like *hia* and *kua* were the head nouns of relative constructions. However, other than the impression one would get from free translations, they are anything but head nouns in the grammar of Tsou for at least three reasons.

⁷¹ The two morphemes, *hia* and *kua*, have only the indefinite function, but no interrogative function.

First, both the linear position between preverbal auxiliaries and lexical verbs (a point also made by M. Chang 2002, but for a different purpose) and the adverbial semantics suggest that *hia* and *kua* were historically verbs, or more specifically, “adverbial verbs” (H. Huang 2010: 175) such as *aha’o* and *aha’va* in (36).

(36) Tsou (H. Huang 2010: 181)

- a. **mo** **aha’o** mieboct=[na cou
 AF.RLS AF.suddenly AF.fart=TOP person
 ‘The person suddenly farted.’
- b. **o=si=cu** **aha’va** lehtothom-neni=[’e eatatiskova
 NAF.RLS=3SG.ACT=already NAF.suddenly wrestle-CF=TOP living.being
 ‘(She, i.e. the mother bear,) suddenly wrestled with the man.’

H. Huang (2010: 178) identifies two properties of adverbial verbs that are not shared by lexical verbs. Semantically, adverbial verbs “encode information more peripheral to events, such as manner, aspect, frequency, time, attitude, etc.” Syntactically, adverbial verbs do not determine the argument structure whereas lexical verbs do. In (36)b, for instance, it is the lexical CF verb *lehtothom-neni* rather than the adverbial verb *aha’va* that determines the Conveyance Topic. This is essentially due to the fact that adverbial verbs make a grammatical distinction only between AF and NAF as preverbal auxiliaries do. Although neither *hia* nor *kua* seems to show Focus alternations, it seems to be no coincidence that they both end with the vowel /a/, which might be relatable to PF -a.

Second, as in the case with adverbial verbs, it is the lexical verb after *hia* or *kua* that determines the argument structure. In (37), for instance, the AF verb *cocvo* and the LF verb *cocvi* after *hia* respectively select Actor (i.e. the one laughing) and Location (i.e. the one laughed at) as the Topic argument.

(37) Tsou (M. Chang 2002: 344)

- a. mo o'ha ʌmnʌ=[o mi=ta hia cocvo=[to
 AF.RLS NEG AF.good=TOP AF.RLS=3SG.TOP how AF.laugh=UND

yangui]=[e pasuya]]
 Y.=TOP P.

‘The way Pasuya smiles at Yangui is not good.’

- a. mo o'ha ʌmnʌ=[o i=ta hia cocv-i=[to
 AF.RLS NEG AF.good=TOP NAF.RLS=3SG.ACT how laugh-LF=ACT

pasuya]=[e yangui]]
 P.=TOP Y.

‘The way Pasuya laughs at (i.e. mocks) Yangui is not good.’

Third, synchronically speaking, forms like *hia* are highly grammaticalized such that they attract person-form indexes, a property of preverbal auxiliaries but not of lexical verbs. When preverbal auxiliaries are absent, person-form indexes attach to *hia* instead, as contrasted in (38) and (39).

(38) Tsou

- a. te=ko=la aezuh-a=[o la=ko hia aomotʌ'ʌ]
 NAF.IRR=2SG.ACT=DSTT change-PF=TOP AF.HAB=2SG.TOP how AF.talk

‘You should change the way you (normally) talk.’ [= (35)a]
 (ODIL; under the entry *aezuha*)

- b. te=ko=la aezuh-a=[o hia=su aomotʌ'ʌ]
 NAF.IRR=2SG.ACT=DSTT change-PF=TOP how=2SG.TOP AF.talk

‘You should change the way you talk.’ (Fieldnotes)

(39) Tsou

- a. mo o'ha ʌmnʌ=[o mi=ta hia cocvo=[to
 AF.RLS NEG AF.good=TOP AF.RLS=3SG.TOP how AF.laugh=UND

yangui]=[e pasuya]]
 Y.=TOP P.

‘The way Pasuya smiles at Yangui is not good.’ (M. Chang 2002: 344) [= (37)a]

- b. taunona'vʌ=[e hia=si ti'usnu=[to mo'o]=[e pasuya]]
 AF.surpring=TOP how=3SG.TOP AF.whip=UND M.=TOP P.

‘The way Pasuya whips Mo'o is surprising.’ (Fieldnotes)

Although regular nouns also attract person-form indexes, they never occur immediately before lexical verbs, unlike *hia* in the (b) examples of (38) and (39). That is to say, by

hosting person-form indexes *hia* helps to create the same effect of having obligatory preverbal auxiliaries, that is, to prevent person-form indexes from attaching to lexical verbs. When there is nothing available in the preverbal position to attract person-form indexes, they inevitably attach to the next word form available, thus giving rise to lexical nominalizations to be discussed in §3.3.2.

Therefore, circumstantial nominalizations involving *hia* for manner and *kua* for reason are as exocentric as argument nominalizations that denote the Topic argument, where no single word form is solely responsible for the nominality of the whole construction.

3.3.2. Lexical argument nominalizations

Aside from grammatical argument nominalizations, Tsou also has lexical ones, where denoting expressions made up by verbs are treated syntactically like underived nouns. Compared with that in other Formosan languages, the boundary between grammatical and lexical nominalizations in Tsou is relatively easy to discern due to the distinct syntactic properties of lexical verbs and nouns (recall §3.2.2 for how either is defined in Tsou).

Three types of lexical nominalizations are worth mentioning here. The first one is where a stative verb is used to denote entities associated with the attribute expressed by that verb, as discussed by M. Chang (2002), from which relevant examples in (40) are drawn.

(40) Tsou (M. Chang 2002: 336-337)

- a. mo **cohũmũ**=[’o suika]
 AF.RLS AF.sweet=TOP watermelon
 ‘The watermelon is sweet.’
- b. mo notaico=[’e **cohũmũ**=si=[ta suika]]
 AF.RLS AF.central=TOP AF.sweet=3SG.GEN=GEN watermelon
 ‘The sweet part of the watermelon is in the center.’

The word form *cohũmũ* in the first example is clearly a verb because it serves as the matrix predicate preceded by an auxiliary, but the same form in the second can be argued to be a noun because it serves as the Topic argument without being preceded by any auxiliary and more importantly it attracts person-form indexes, which regular verbs in Tsou are not allowed to do. The nominal use in (40)b is only restricted to a small subset of stative verbs (e.g. **kaebũ=si* ‘AF.happy/like=3SG.GEN’; *ibid.*: 342), much as only a small subset of English adjectives can be preceded by *the* to denote a group of entities characterizable by attributes expressed by those adjectives. The lexical nominalization in (40)b is in sharp contrast with grammatical nominalizations involving the same word form followed by an auxiliary, as in (41).

(41) Tsou

- la *kaebũ*=[to mo **cohũmũ**]=[’o sosea]
 AF.HAB AF.like=UND AF.RLS AF.sweet=TOP ant
 ‘Ants like anything sweet.’ (ODIL; under the entry *cohũmũ*)

While the lexical nominalization process is rather restrictive and creates word forms that can be argued to be nouns by the grammatical criteria in Tsou, the grammatical nominalization process is highly productive and does not create any lexical nouns.

The other two types of lexical nominalizations denote entities that are specifically designated to perform a certain action or to be regularly used for the purpose of that action without making reference to the spatio-temporal aspect of an event.

Nominalizations of this semantic nature can be called *dispositional*, as opposed to *episodic* ones, whose denotations are spatio-temporally bound to an event (terms borrowed from Alexiadou & Schäfer 2010).

Dispositional Actor nominalizations denote someone who specializes in doing something or does it by profession, and they take the form of *le*-Mstem. They contrast with episodic Actor nominalizations, which are expressed by grammatical nominalizations in the form of Mstem preceded by an obligatory auxiliary as discussed in §3.2.2. In what follows, I contrast the syntactic differences between *le*-Mstem and the same Mstem preceded by the auxiliary *la*, which indicates habitual aspect.

At least three syntactic clues show that *le*-Mstem behaves like a regular noun whereas the same Mstem preceded by the habitual *la* is treated like a regular verb. First, *le*-Mstem can be a nominal predicate by itself whereas *la* Mstem cannot unless marked by a nominal relation marker, as in (42).

(42) Tsou (Fieldnotes)

- a. 'a eno **le-aotothomɯ**=[e pasuya]
 AFF indeed specialize-AF.cure=TOP P.
 'Pasuya is indeed a doctor.'
- b.* 'a eno **la aotothomɯ**=[e pasuya]
 AFF indeed AF.HAB AF.cure=TOP P.
- c. 'a eno=[o **la aotothomɯ a'o**]=[e pasuya]
 AFF indeed=TOP AF.HAB AF.cure 1SG.UND=TOP P.
 'Pasuya is indeed the one who regularly cures me.'

Second, the Mstem marked by *le*- attracts person-form indexes as regular nouns do whereas the same Mstem after the habitual *la* cannot, as in (43).

(43) Tsou (Fieldnotes)

- a. 'a zou **le-aotothomɯ='u=[e** pasuya]
 AFF EMPH specialize-AF.cure=1SG.GEN=TOP P.
 'Pasuya is my doctor.'
- b.* **la** **aotothomɯ='u=[e** pasuya]
 AF.HAB AF.cure=1SG.GEN=TOP P.

Third, the Mstem preceded by the habitual *la* takes the patientive Undergoer argument as regular verbs do whereas the same Mstem marked by *le-* cannot, as in (44).

(44) Tsou (Fieldnotes)

- a.* 'a zou **le-aotothomɯ** **a'o=[e** pasuya]
 AFF EMPH specialize-AF.cure 1SG.UND=TOP P.
- b. **la** **aotothomɯ** **a'o=[e** pasuya]
 AF.HAB AF.cure 1SG.UND=TOP P.
 'Pasuya cures me regularly.'

Thus, the three tests indicate that *le-*affixation turns the verbal Mstem into a lexically nominal form that is syntactically treated like underived nouns. In this sense, *le-* qualifies as a nominalizer.

On the other hand, there is evidence suggesting that the nominal function of *le-*Mstem may have resulted from lexicalization of the same mechanism as grammatical argument nominalizations discussed in §3.2.2. Specifically, aside from its nominal use, *le-*Mstem also functions like the regular Mstem and assume all the verbal properties. First, *le-*Mstem is preceded by the same preverbal auxiliaries as those that co-occur with a typical Mstem, as compared in (45) (cf. *le-yaezoi* 'farmer').

(45) Tsou

- a. **ta='u=la** **le-yaezoi**
 AF.IRR=1SG.TOP=DSTT specialize-AF.till.the.soil
 ‘I want to (make a living by) farming (in the future).’
 (ODIL; under the entry *leyaezoi*)
- b. **ho** **ta='u=la** **ahueu** **m-eeecunu=[no emoo=su]**
 CONJ AF.IRR=1SG.TOP=DSTT AF.should AF.go.over=OBL house=2SG.GEN
 ‘Oh, I should go (with you) to your house.’ (G. Lin 2010: 364; citing T. Tung 1964)

Second, like the regular Mstem, *le*-Mstem is followed by a verb to form serial verb constructions (SVCs; see G. Lin 2010: 351), as in (46) (cf. *le-ma'coho* ‘teacher’).

(46) Tsou

- a. **la=ta** **le-ma'coho** **pasunaeno=[e pasuya]**
 AF.HAB=3SG.TOP specialize-AF.teach AF.sing=TOP P.
 ‘Pasuya (makes a living by) teaching singing.’
 (ODIL; under the entry *lema'coho*)
- b. **mi='o** **aothom#** **m-aavo=[to phingi]**
 AF.RLS=1SG.TOP AF.try AF-open=UND door
 ‘I tried to open the door.’ (G. Lin 2010: 368)

Third, which is more important, *le-* is prefixable not only to AF-words, as has been shown above, but also to NAF-words, as in (47), indicating the prefix is part of the productive verbal system.

(47) Tsou (Fieldnotes)

- la=ta** **le-teoc-a=[ta** **pasuya]=[o evi='u]**
 NAF.HAB=3SG.ACT specialize-chop-PF=ACT P.=TOP tree=1SG.GEN
 ‘Pasuya specializes in cutting down my trees.’

All the examples in (45) through (47) show that adding *le-* to a verb stem does not create a lexical noun and has no syntactic consequences. Instead, the prefix only contributes additional semantics to the verb stem it is attached to, which makes it functionally equivalent to what is called lexical prefixes in the literature (Tsuchida 2000). For instance, like *le-*, the lexical prefix *o-* is compatible with both AF- and NAF-words, as in (48), where it adds the meaning of “eating” to the reduplicated stem *so~son#* for AF and

so~sona for PF (called “portmanteau verbs” by H. Huang 2010), a function that its semantically equivalent free forms (i.e. AF *bonu* and PF *ana*) cannot accomplish.

(48) Tsou (H. Chang 2009: 452)

- a. *mi=ta* **o-so~sonu** *bonu*=[to *yosku*]
 AF.RLS=3SG.TOP eat-IPFV~AF.easy AF.eat=UND fish
 ‘{She/He} eats fish without any difficulty.’
- b. *i=ta* **o-so~son-a** *an-a*=[’o *yosku*]
 NAF.RLS=3SG.ACT eat-IPFV~easy-PF eat-PF=TOP fish
 ‘{She/He} ate the fish without any difficulty.’

Therefore, considering all the distribution facts regarding the prefix *le-*, it is highly plausible that it starts out as a verbal prefix that modifies the semantics of the verb stem it is attached to like lexical prefixes. The nominal use of *le*-Mstem in (42) through (44) is then the consequence of a lexicalization process whereby the verbal use in (45) through (47) is gradually stripped of auxiliaries, which encode event-specific information such as Focus and modality. This is similar to some sporadic amphibious word forms that have both the verbal and nominal use, such as *pasunaeno* in (49), taken from a spontaneous narrative about a love story.

(49) Tsou (S. Huang et al. 2001: Appendix 90-91)

- mi=cu* **pasunaeno**=[to *mo* *con=ci* *i=si* *na’n-a*
 AF.RLS=already AF.sing=UND AF.RLS one=LIG NAF.RLS=3SG.ACT very-PF
- uṃnu-a=ci* **pasunaeno**]=ho *toveucu*=[to *buvnu=ci* *i=si*
 like-PF=LIG song=CONJ AF.pick=UND flower=LIG NAF.RLS=3SG.ACT
- na’n-a* *uṃnu-a*=[to *oko=no* *mamespingi*]]
 very-PF like-PF=ACT child=GEN woman
- ‘(The boy) sang one much-loved song, and picked flowers that the girl liked very much.’

Where dispositional Actor nominalizations make use of the prefix *le-*, dispositional non-Actor ones involve *’o-*, which is prefixed to a PF verb. For instance, the result form *’o-yon-a* ‘regularly-reside-PF’ (cf. AF *yon*, LF *yon-i*, and CF *yon-eni*), which denotes the place where one resides or the habitat of animals, is syntactically treated like

an underived noun. It can also be followed by an AF verb to form a compound nominal, which denotes designated places for some activities. Examples in (50) are illustrative.

(50) Tsou (ODIL; under the entry *'oyona* and *'oyonapei'i*)

- a. *la* *noepe=he=[to* *toevosʉ]=[ʼo* *'o-yon-a]=[to*
 AF.HAB AF.high=CMPR=OBL Swinhoe's.pheasant=TOP regularly-reside-PF=GEN
ftʉftʉ]]
 Mikado.pheasant
 'The habitat of the Mikado pheasant is higher (in altitude) than that of the Swinhoe's pheasant.'
- b. *mi=ta* *yon=[ta* *'o-yon-a* *pei'i]=[ʼe* *ino]*
 AF.RLS=3SG.TOP reside=OBL regularly-reside-PF AF.cook=TOP mother
 'Mother is in the kitchen (i.e. where one regularly stays while cooking).'

In addition to the nominal use in (50), the non-Actor *'o*-Kstem-*a* form also has the verbal use, where it is preceded by an auxiliary that attracts person-form indexes and optionally followed by an AF verb, just like a typical verb. This is shown by *'o-eon-a* 'regularly-stay-PF' (cf. AF *eon*, LF *eon-i*, and CF *eon-eni*) in (51), to be compared with a typical PF verb without the prefix *'o-* in non-harmonizing SVCs (wherein the two verbs belong to two different Focus categories; see G. Lin 2010: 375), as shown in (52).

(51) Tsou (Fieldnotes)

'a *Zou* *taipahu=[ʼo* *la=ʼu* *'o-eon-a* *{bonʉ/t<m>opsʉ}]*
 AFF EMPH Taipei=TOP NAF.HAB=1SG.ACT regularly-stay-PF {AF.eat/<AF>study}
 '(The place) where I regularly {eat/study} is Taipei.'

(52) Tsou (G. Lin 2010: 376)

i=ʼo *haf-a* *uh=[to* *taipahu]=[ʼo* *naau*
 NAF.RLS=1SG.ACT take-PF AF.go=OBL Taipei=TOP N.
 'I took Naau to Taipei.'

Therefore, the prefix *'o-* does not necessarily make the morphological verb to which it is attached become a lexical noun. Rather, it simply adds the adverbial meaning "regularly" to a verb. As in the case of the *le*-Mstem, the nominal use of the *'o*-Kstem-*a* form, as in (50), can again be seen as resulting from depriving a grammatical nominalization, as in

(51), of its preverbal auxiliary, which encodes time-bound information. Such a process is presumably motivated by the fact that both AF *le*-Mstem and NAF *'o*-Kstem-*a* express time-stable events, thus rendering time-bound auxiliaries dispensable. As a result, the same SVCs in (51) can host person-form indexes as regular nouns do, as in *'o-eon-a* {*bon#t<m>ops#*}=*'u* 'my usual {dining/studying} place'. The fact that person-form indexes attach to the second AF verb instead of the first PF verb agrees well with G. Lin's (2010: 410) conclusion that non-harmonizing SVCs are both phonologically and morphologically a well-integrated unit like a typical verb.

3.4. Rukai

The Rukai language is rather diverse, and six distinctive dialects are often identified, with Maga and Tona closely related to each other on the one hand (called Northern dialects in Zeitoun 1995: 169) and Budai, Labuan, and Taromake on the other (called South-eastern dialects; *ibid.*: 175).⁷² Mutual intelligibility between the two groups is reported to be low (P. Li 1977). The genetic relationship between Mantauran, the sixth dialect, with the other five, is less obvious although it is geographically adjacent to Northern dialects. While P. Li (1977, 2001) suggests Mantauran is more closely related to Northern dialects, Zeitoun (2003) proposes that its closer relative should be South-eastern dialects instead.

In this section, I focus on Budai Rukai, particularly with respect to some claims made in L. Sung (2011).

⁷² Taromake is more commonly known as Tanan due to P. Li (1973), which comes from the erstwhile Chinese name of the village where this variety is primarily spoken. However, Taromake is the self-denominational term preferred by the local community.

3.4.1. Argument nominalizations

I first introduce lexical word forms that can reasonably be called nouns and verbs in Budai Rukai, and then assess L. Sung's (2011) claim (as quoted in §3.2.3) that there are non-nominalized and nominalized relative constructions in this language, depending on whether the relativized argument is the subject or not.

Underived nouns serving as arguments are marked by nominal relation markers *ka*, *ku*, or *ki*, which have been given various case labels in the literature due to the complicated selection principles determined by both syntactic and semantic factors. Descriptively speaking, these principles can be summarized as follows (based mostly on C. Chen 2008): (i) if an argument is the subject, use *ka* or *ku* for all semantic types of nouns, and *ki* for only personal names;⁷³ (ii) if an argument is a non-subject, use *ka* or *ku* for non-human non-generic nouns, and *ki* for non-human generic nouns, location nouns, or human nouns; (iii) the selection between *ka* and *ku* for subjects and non-subjects alike is often correlated with the visibility or definiteness of an NP referent; and (iv) only *ki* is used to mark the full-nominal possessor regardless of its semantic properties. In addition, there are also some lexicalized or construction-specific restrictions on which marker should be selected, including at least the following (based on data from the literature and my fieldnotes): (v) predicate nominals can only be marked by *ka* (which is optional in affirmatives but obligatory in negatives); (vi) some compounds consisting of two nouns can only be linked by *ka*, irrespective of their grammatical relations; (vii) bound person forms combine with only *ku* to become free forms (e.g. *ku=naku* 'DET=1SG.NOM'); and finally (viii) when a noun is modified by

⁷³ However, no instance of *ki*-marked subjects is found in W. Shih's (2012) discourse-based study.

demonstratives the marker *ki* is predominantly favored over the other two if any of the three is used at all between demonstratives and nouns.⁷⁴ These selection principles present evident glossing challenges. In the present study, I adopt L. Sung's (2011) case labels in scenarios (i) through (iv), whereby the subject, non-subject, and possessor are respectively glossed as NOM, OBL, and GEN regardless of which markers are used. For scenarios (v) through (viii), on the other hand, I simply use DET as a cover label for lack of better choices. Despite the somewhat arbitrary division between the two groups of scenarios, presumably the same markers are involved, but they have become so specialized that it is almost impossible to make generalizations over the various functions of a given marker.

Both nouns and verbs can take up the sentence-initial position to serve as the matrix predicate without additional morphosyntactic changes. When they are negated, however, nouns require the *ka* marker in addition to the negator whereas verbs do not.

(53) Budai Rukai (L. Sung 2011: 526)⁷⁵

- a. ngudradrekay {ka/ku} salrabu
 Rukai NOM S.
 'Salrabu is a Rukai.'
- b. kai=[ka ngudradrekay] {ka/ku} salrabu
 NEG=DET Rukai NOM S.
 'Salrabu is not a Rukai.'

Verbs come in at least four morphological classes, definable by the morphological alternations between the Mstem and Kstem, as summarized in Table 3.3.

⁷⁴ The last point is based on the natural discourse data in W. Shih (2012).

⁷⁵ Based on its prosody and its distributions as described in Y. Tang (2008), the negator *kai=*, as well as the future marker *lri=*, is tentatively analyzed as a proclitic. One indicator is the flexible order between the two morphemes (e.g. *lri=kai=thingale* 'FUT=NEG=know' vs. *kai=lri=thingale* 'NEG=FUT=know').

Table 3.3: Morphological verb classes in Budai Rukai (after L. Sung 2011: 530)

	I	II	III	IV
Mstem	<i>wa-lrumay</i> 'RLS-beat'	<i>ma-dalame</i> 'RLS-like'	<i>ki-a-kamadha</i> 'get-RLS-mango'	<i>duduli</i> 'red'
Kstem	<i>lrumay</i> 'beat'	<i>ka-dalame</i> 'K-like'	<i>ki-kamadha</i> 'get-mango'	<i>duduli</i> 'red'

The Mstem functions as the matrix predicate in realis-indicative (or called non-future in C. Chen 2008) sentences, be it affirmative or negative, as illustrated in (54) through (57).

(54) Budai Rukai (L. Sung 2011: 527)

- a. **wa-lrumay** {ki/*ka/*ku} lrailrai {ka/ku} salrabu
 RLS-beat OBL L. NOM S.
 'Salrabu beat Lrailrai.'
- b. kai=**wa-lrumay** {ki/*ka/*ku} lrailrai {ka/ku} salrabu
 NEG=RLS-beat OBL L. NOM S.
 'Salrabu did not beat Lrailrai.'

(55) Budai Rukai (L. Sung 2011: 527)

- a. **ma-dalame** {ki/*ka/*ku} lrailrai {ka/ku} salrabu
 RLS-like OBL L. NOM S.
 'Salrabu likes Lrailrai.'
- b. kai=**ma-dalame** {ki/*ka/*ku} lrailrai {ka/ku} salrabu
 NEG=RLS-like OBL L. NOM S.
 'Salrabu does not like Lrailrai.'

(56) Budai Rukai (Y. Tang 2008: 55)

- a. **ki-a-kamadha** ku basakalrane
 get-RLS-mango NOM B.
 'Basakalrane picked mangos.'
- b. kai=**ki-a-kamadha** ku basakalrane
 NEG=get-RLS-mango NOM B.
 'Basakalrane did not pick mangos.'

(57) Budai Rukai (L. Sung 2011: 526)

- a. **duduli** ka laimay
 red NOM clothes
 ‘The clothes are red.’
- b. **kai=duduli** ka laimay
 NEG=red NOM clothes
 ‘The clothes are not red.’

On the other hand, the Kstem is used in a variety of syntactic environments and is also the required base when certain morphemes are added, including at least *-a* ‘IMP’, *-ane* ‘NMLZ’, *sa=* ‘when.PST’, *lu=* ‘when.RLS’, and *lri=* ‘FUT’ (Y. Tang 2008: 36).⁷⁶

Class I and II are often said to encode dynamic and stative events respectively. Although the semantic characterization generally matches up with the morphology (e.g. *wa-kane* ‘RLS-eat’ vs. *ma-barengere* ‘RLS-miss’), there are also apparent counterexamples (e.g. *wa-thingale* ‘RLS-know/miss’ vs. *ma-tuas* ‘RLS-leave’). Thus, C. Chen (2008: 145) argues the choice between *wa-* and *ma-* to be more a root-specific morphological requirement than an absolute reflection of semantic contrast. Nevertheless, there are some roots that are equally amenable to both *wa-* and *ma-* prefixation, with the latter expressing middle or reflexive situation types (e.g. *wa-cuake* ‘(of a person) to break something’ vs. *ma-cuake* ‘(of something) to break into pieces’; Y. Tang 2008: 38). Class III is to a large extent denominal, consisting of a lexical prefix (with meanings such as “to wear” or “to get”, “to go”, or “be at somewhere”, etc.) and a nominal root denoting the theme object or location. In addition, a prefix in Class III also combines with the Kstem, which leads to a passive construction that has been grammaticalized from the

⁷⁶ The Mstem and Kstem in Budai Rukai are often called finite and non-finite forms respectively. However, the diverse contexts where the Kstem are required do not support a coherent concept of non-finiteness regardless of how it is defined. Ross (2015c) points out that the Kstem is more a morphological requirement than a syntactically or semantically generalizable notion. One good way to show this in Budai Rukai is that *pa*-causativization is only possible with the Kstem of Class I (e.g. *wa-kane* ‘RLS-eat’ vs. *pa-kane* ‘CAUS-eat’) and II verbs (e.g. *ma-cuake* ‘RLS-break’ vs. *pa-ka-cuake* ‘CAUS-K-break’), but applies equally to the Mstem and Kstem of Class III verbs (e.g. *pa-ngu-(a)-cilri* ‘CAUS-go-(RLS)-lose’).

denominal type in (56) (cf. *get*-passive in English), attested in all Rukai varieties (Elizabeth & S. Teng 2009). The active and passive of Class I and II are illustrated in (58) and (59).

(58) Budai Rukai (L. Sung 2011: 529)

a. **wa-lrumay** {ki/*ka} lrailrai {ka/*ki} salrabu
 RLS-beat OBL L. NOM S.

‘Salrabu beat Lrailrai.’ (cf. (54)a)

b. **ki-a-lrumay** {ki/*ka} salrabu {ka/*ki} lrailrai
 PASS-RLS-beat OBL S. NOM L.

‘Lrailrai was beaten by Salrabu.’

(59) Budai Rukai (C. Chen 2008: 77)⁷⁷

a. **ma-barenger=aku** ki ina
 RLS-miss=1SG.NOM OBL Mom

‘I miss Mom.’

b. **ki-a-ka-barengere** nakuane ka ina
 PASS-RLS-K-miss 1SG.OBL NOM Mom

‘Mom was missed by me.’

In the active, the agent and patient are marked in NOM (i.e. *ka/ku* for human nouns or SAP indexes on the verb) and OBL (i.e. *ki* for human nouns or person forms ending with *-ane*) respectively whereas the marking is reversed in the passive. Thus, the agent in active and the patient in passive are generalizable into the grammatical category subject in Rukai.

In addition to matrix predicates, verbs can also be part of the phrase in a typical argument position, which brings us to the contrast between relativization and nominalization as maintained in L. Sung (2011). It was claimed that Budai Rukai has “nominalized clauses” involving the Kstem-*ane* (where *-ane* is a nominalizer reflecting PAN LF *-an) or just the Kstem on the one hand (called zero nominalization), and

⁷⁷ C. Chen (2008) chose to keep the free variations between *ako* and *aku*, or more generally between [o] and [u], from his consultants. I decide to replace all his instances of <o> into <u> since the transcription system I adopt is by and large phonemic.

“ordinary relative clauses” with the Mstem like those in (54) through (57) (called finite verbs) on the other. In what follows, I compare the nominalized type involving the Kstem-*ane* and the “relativized” type concerning the Mstem and argue that such a dichotomy is unmotivated in Budai Rukai.

The internal syntax of the Kstem-*ane* has been shown by L. Sung (2011) to share many common properties with the Mstem, such as allowing non-subject arguments (marked in OBL), tense/aspect marking, temporal and frequency adverbials, passivization, reciprocalization, reflexivization, and causativization. In short, “the internal structures resemble in nearly all respects that of a sentence.” (ibid.: 525) These internal properties will be not repeated here. Instead, I focus on how the Kstem-*ane* and Mstem interact with higher-order external syntax. According to L. Sung (2011: 538), the Kstem-*ane* is nominal because it demonstrates two “hallmarks of nominality, i.e. embeddedness under a case marker/demonstrative and the presence of genitive subject”, which are discussed below in that order.

The nominalized and the “relativized” type are illustrated in (60) and (61) respectively, where the constructions in question are indicated by brackets.⁷⁸

(60) Budai Rukai (L. Sung 2011: 534)

ma-lras=aku kane ka lacenge [ka **ta-aga-ane**
 RLS-hate=1SG.NOM eat OBL vegetable OBL RLS-cook-NMLZ

ki lrailrai]

GEN L.

‘I hate to eat the vegetables that Lrailrai cooked.’

⁷⁸ It is irrelevant that the modifying phrase follows the modified noun in (60) but precedes it in (61). Either order is possible for either type.

(61) Budai Rukai (L. Sung 2011: 532) (cf. (58) above)

- a. ngu-a-bere [ku **wa-lrumay** ki lrailrai] ku lasu
 go-RLS-escape NOM RLS-beat OBL L. NOM man
 ‘The man that beat Lrailrai ran away.’
- b. ngu-a-bere [ku **ki-a-lrumay** ki salrabu] ku lasu
 go-RLS-escape NOM PASS-RLS-beat OBL S. NOM man
 ‘The man that was beaten by Salrabu ran away.’

The marker between a modifying phrase and a modified noun (i.e. the second instance of *ka* and *ku* in (60) and (61) respectively) was glossed REL (for relativizer) in the cited work. This practice is not followed here due to its problems. For one thing, it leads to an odd situation where the so-called relativizer sometimes precedes but sometimes follows what it presumably relativizes. For another, when an underived noun modifies another, the same pattern is found, with both nouns marked by the same marker, as in (62).

(62) Budai Rukai (Fieldnotes)

wa-kane=aku ku lrubu [ku lapanay]
 RLS-eat=1SG.NOM OBL porridge OBL corn
 ‘I ate corn porridge.’

Thus, a simpler analysis is that both the modifier and the modifiee in (60) and (61) are equally independent nominal expressions, much like the two nouns in (62). This analysis is supported by several parallel distributions of the two types of modifier phrases in (60) and (61) on a par with underived nouns.

First, both types can make up a complete NP and be pluralized, showing that both are nominal expressions independent of a modified noun.

(63) Budai Rukai⁷⁹

- a. *wa-malra* [ku **la=hana**] ka balenge
 RLS-get OBL PL=flower NOM B.
 ‘Balenge plucked flowers.’ (C. Chen 2008: 19)
- b. *mu-a-bere* [ku **ta-lrumadh-ane** ki camake]
 go-RLS-escape NOM RLS-beat-NMLZ GEN C.
 ‘The one Camake beat ran away.’ (Fieldnotes; based on L. Sung 2011: 532)
- c. *kai lapanay*, [**la=sa-pa-kan-ane** ki beke]
 PROX corn PL=INS-CAUS-eat-NMLZ OBL pig
 ‘The corn (here) is (meant to be) used for feeding pigs.’ (Fieldnotes)
- d. *mu-a-bere* [ku **(la=)wa-lrumay** ki camake]
 go-RLS-escape NOM PL=RLS-beat OBL C.
 ‘{The one/Those} who beat Camake ran away.’
 (Fieldnotes; based on L. Sung 2011: 532)
- e. *mu-a-bere* [ku **(la=)ki-a-lrumay** ki salrabu]
 go-RLS-escape NOM PL=PASS-RLS-beat OBL S.
 ‘{The one/Those} who {was/were} beaten by Salrabu ran away.’
 (Fieldnotes; based on L. Sung 2011: 532)

Second, both types can be modified by demonstratives, as in (64).

(64) Budai Rukai⁸⁰

- a. *ma-tu-mane kai Iribange*
 RLS-do-what PROX window
 ‘What happened to this window?’ (L. Sung 2011: 534)
- b. *masamali=aku kai ta-lrumadh-ane* ki salrabu ki lrailrai
 surprised=1SG.NOM PROX RLS-beat-NMLZ GEN S. OBL L.
 ‘I am surprised at (the consequence resulting from) Salrabu’s beating Lrailrai.’
 (L. Sung 2011: 537)
- c. *ka cegaw ma-dalame kawai ki ababay, kuini si-a-belrengay*
 NOM C. RLS-like MED DET woman PROX wear-RLS-flower
 ‘Cegaw likes the woman there, this one that wears a flower.’ (C. Chen 2008: 118)

⁷⁹ There is a morphophonemic alternation between the postvocalic glide as in {ay} (phonetically [aj]) and {dh} (phonetically [ð]) (cf. *wa-lrumay* and *ta-lrumadh-ane*). In addition, the verb meaning to “escape” is *ngu-a-bere* in L. Sung (2011) (see (61) above), but this form was not recognized by my consultants, who provided me with *mu-a-bere* instead (in the year of 2013). This is congruent with a later footnote in L. Sung (2015: 295) stating that older speakers prefer *mu-a-bere*. See below for more discussions on the semantic differences between the matrix argument in (63)b and (63)e, both of which denote a patientive participant.

⁸⁰ The original transcriptions involving word-final <au> and <ai> have been changed into <aw> and <ay> respectively in some cases, but remain intact in others. This follows the phonological analysis in C. Liu (2008), where the stress assignment and other phonological changes (e.g. the one mentioned in Footnote 79) are cited as evidence for the phonemic difference between vowel-glide sequences and two consecutive vowels (cf. *pagay* [pájaj] ‘rice’ vs. *kikai* [kíkáj] ‘PROX’; *ibid.*: 46).

When showing the external distributions of the Kstem-*ane*, L. Sung (2011: 538) commented that “the presence of case marker/demonstrative clinches the identity as a nominal.” By this very same criterion, the Mstem marked by case markers/demonstratives in both (63) and (64) should be equally nominal.

Third, both types can serve as the sole argument of the affirmative existential predicate *yakai* (< *i-a-kai* ‘LOC-RLS-PROX’; see also Zeitoun et al. 1999) or its negative counterpart *kadrúa*. The latter construction is illustrated in (65).

(65) Budai Rukai (Y. Tang 2008: 122, 120, 119)

- a. *kadrúa* *ku* **alralregele** *i-kai* *ki* *cukuy*
 NEG.EX NOM fly LOC-PROX OBL table
 ‘There no flies on the table.’
- b. *kadrúa* *ku* [**a-kitubi-ane=li** *numiane*]
 NEG.EX NOM FUT-ask.for-NMLZ=1SG.GEN 2PL.OBL
 ‘There is nothing that I will ask you for.’
- c. *kadrúa* *ku* [**wa-lra<bua~>buale** *i-gaku*]
 NEG.EX NOM RLS-<IPFV~>run LOC-school
 ‘This is nobody running in the school.’

Fourth, both types can be a fronted topic marked by *yai*, which is subsequently commented on, as in (66).

(66) Budai Rukai (Y. Tang 2008: 75, 112, 15)⁸¹

- a. *ka* **ama** *yai* *ngudradrekay*
 NOM father PTOPTOP Rukai
 ‘(My) father, (he is) a Rukai person.’
- b. *kikai* **sa-ka-ua~ung-ane** *yai* *laububulu* *ki* *pulaludhane*
 PROX INS-K-IPFV~work-NMLZ PTOPTOP stuff GEN P.
 ‘This (tool for) working, (it is) Pulaludhane’s stuff.’
- c. *(*ka*) **ki-a-kace** *yai* *agi=li*
 NOM PASS-RLS-bite PTOPTOP younger.sibling=1SG.GEN
 ‘The one that got bitten, ({she/he}) is my younger {sister/brother}.’

⁸¹ The original morphological decomposition for *sa-ka-ua~ung-ane* in the cited reference is *sa-kaua-ung-ane* ‘INS-work-RED-NMLZ’. However, the Mstem is *ma-ungu* and its Kstem *ka-ungu*. Thus, the reduplicated syllable should be /ua/ instead of /ung/.

The fronted topics in the first two examples are expected since they are nominal expressions, but the Mstem *ki-a-kace* in the third example can also denote someone provided that it is marked by a nominal relation marker, without which the third example would be unacceptable and the same verb would assert a predication instead of denoting a participant of that predication, as in (67).

(67) Budai Rukai (Y. Tang 2008: 13)

ki-a-kace ki sulraw ka kaka
 PASS-RLS-bite OBL snake NOM older.sibling
 ‘(My) older {sister/brother} got bitten by a snake.’

Fifth, as pointed out by L. Sung (2010: 548), the Kstem-*ane* can only be negated by the nominal pattern (as in (53) above) whereas the Mstem is negated by the verbal pattern (as in (54) through (57) above). However, the Mstem also has the option to be negated by the nominal pattern when the predication is presupposed instead of asserted. The two negation types for which the Mstem is equally suitable are contrasted in (68).

(68) Budai Rukai (Y. Tang 2008: 56, 90)

- a. **kai=wa-ta<maku~>maku kai=wa-ung~ungulu** ku bava
 NEG=RLS-<IPFV~>smoke NEG=RLS-IPFV~drink OBL alcohol
 ‘{She/He} does not smoke (and) does not drink.’
- b. **wa-gelre~gelres kai=[ka wa-sena~senay]**
 RLS-IPFV~cry NEG=DET RLS-IPFV~cry
 ‘{She/He} is crying rather than singing.’

In the first example, what is negated is an assertion (that someone drinks), but in the second the negation targets at a presupposition (that someone is crying) (see Y. Tang 2008: 91). The second example suggests that the Mstem, with appropriate marking, can function as a nominal predicate. This is confirmed in (69), where a regular noun, a temporal nominal expressed by the Kstem-*ane* (additionally prefixed by *kala-* to indicate

temporality), and the Mstem are in a paradigmatic relationship serving as the negated nominal predicate, which requires the *ka* marker.

(69) Budai Rukai (Y. Tang 2008: 86, 96, 90)⁸²

- a. **kai=su=[ka situ=li]**
 NEG=2SG.NOM=DET student=1SG.GEN
 ‘You are not my student.’
- b. **kai=[ka kala-udal-ane] kayasasane**
 NEG=DET TEMP-rain-NMLZ now
 ‘Now is not the raining (season).’
- c. **kai=naku=[ka ma-bi<tu~>tulru=nga]**
 NEG=1SG.NOM=DET RLS-<IPFV~>fat=SUP
 ‘I am not the fattest one.’

As in (68)b, the negation in (69)c is external to the nominalization consisting of the Mstem. Thus, what is negated in (69)c is not the predicate itself, but the identification relationship between an individual and someone characterized by that predicate. On the other hand, negation can also happen within a nominalization, with the negated Mstem occurring in an argument phrase just like its affirmative counterpart, as in (70) (cf. (65)c).

(70) Budai Rukai (Y. Tang 2008: 168)

kadrua ku [kai=wa-thingale iniane]
 NEG.EX NOM NEG=RLS-know 3SG.OBL
 ‘This is no one who does not know {her/him}.’ (Lit. ‘(He who) doesn’t know {her/him} does not exist.’)

Now turning to the other hallmark of nominality demonstrated by the Kstem-*ane*, which was said to have a “genitive subject”. What it means is that that the agentive argument of the active Kstem-*ane* and the patientive one of the passive Kstem-*ane* (additionally prefixed by *ki-*) are encoded in the same manner as the possessor of underived nouns, which is marked by *ki* for full nominals or/and by bound person

⁸² It is of interest to note that in addition to cliticizing right after the negator *kai*, bound person forms indexing the subject in Taromake Rukai have the option to go between /ka/ and /i/, with the latter part being optional in some cases (P. Li 1973: 228). In addition, the negator in Mantaoran Rukai is simply *ka* (Zeitoun 2007). This suggests that the negator *kai* was historically bimorphemic.

forms.⁸³ As a result, the genitive subject and the non-subject would receive the same marking *ki* when they are both human nouns (but not when the non-subject is a non-human noun), as in (71), but have different realizations when at least one of them is a person form, as in (72).

(71) Budai Rukai (L. Sung 2011: 547)

- a. *masamali=aku ku [ta-lrumadh-ane ki salrabu ki lrailrai]*
 surprised=1SG.NOM OBL RLS-beat-NMLZ GEN S. OBL L.
 ‘I am surprised at Salrabu’s beating Lrailrai.’ [Active: GEN=agent]
- b. *masamali=aku ku [ta-ki-lrumadh-ane ki lrailrai ki salrabu]*
 surprised=1SG.NOM OBL RLS-PASS-beat-NMLZ GEN L. OBL S.
 ‘I am surprised at Lrailrai’s being beaten by Salrabu.’ [Passive: GEN=patient]

(72) Budai Rukai (L. Sung 2011: 547, 549, 456)

- a. *masamali=aku ku [ta-lrumadh-ane ki salrabu iniane]*
 surprised=1SG.NOM OBL RLS-beat-NMLZ GEN S. 3SG.OBL
 ‘I am surprised at Salrabu’s beating {him/her}.’ [Active: GEN=agent]
- b. *malisi ka salrabu ka [ta-kan-ane=nga=li]*
 Angry NOM S. OBL RLS-eat-NMLZ=already=1SG.GEN
 ‘Salrabu is angry at my having eaten already.’ [Active: GEN=agent]
- c. *masamali ka salrabu ku [ta-ki-draedrangelr-ane=li]*
 surprised NOM S. OBL RLS-PASS-cheat-NMLZ=1SG.GEN
 ‘Salrabu is surprised at my being cheated.’ [Passive: GEN=patient]

The Mstem, on the other hand, is not allowed to collocate with the genitive subject, which is presumably what kept L. Sung (2010) from analyzing the argument-denoting Mstem as nominalized in spite of its external nominal distributions parallel to the Kstem-*ane*. Since the marker *ki* as GEN is also used as OBL, which introduces non-subjects of both the Kstem-*ane* and Mstem, the genitive subject factor really comes down to the fact that the Kstem-*ane* collocates with possessor indexes whereas the Mstem does not.

However, the question is, should compatibility with possessor indexes be the only criterion for nominalizations? The answer is negative from both a language-internal and

⁸³ The two types of marking co-occur only when the possessor is non-SAP, as in *senate=(ini) ki cegaw* (book=3SG.GEN GEN C.) ‘Cegaw’s book’.

language-external perspective. In Budai Rukai, there are clearly nominal forms that cannot host possessor indexes, such as **kala-udal-ane=li* ‘TEMP-rain-NMLZ=1SG.GEN’ (cf. (69)b). Conversely, there are construction-specific forms that host possessor indexes, but no one seems to have ever analyzed them as nominal. One such example is a cleft-like “focus-presupposition articulation” (Andrews 2007: 150), where a grammatical argument nominalization constitutes the presupposition part and a bound person form indexing the (pragmatic) focus, which is cliticized to *amani* ‘COP’, as in (73).

(73) Budai Rukai (Y. Tang 2008: 90, 77)⁸⁴

- a. *amani=aku* [ku wa-kupa]
 COP=1SG.NOM NOM RLS-steal
 ‘It is me who stole (something).’
- b. *amani=ini* kuiya [ku wa-angeale ku angatu]
 COP=3SG.GEN yesterday NOM RLS-carry.on.the.shoulder OBL wood
 ‘It is {she/he} who carried wood on the shoulder yesterday.’

I double-tested all the bound person forms in (74) using the construction in (73).

(74) Budai Rukai: Bound person forms indexing NOM and GEN (cf. P. Li 1996; Zeitoun 1997: 316)⁸⁵

	1SG	1EXCL	1INCL	2SG	2PL	3SG	3PL
NOM	= <i>aku</i> / <i>=naku</i>	= <i>nai</i>	= <i>ta</i>	= <i>su</i>	= <i>numi</i>	--	--
GEN	= <i>li</i>	= <i>nai</i>	= <i>ta</i>	= <i>su</i>	= <i>numi</i>	= <i>ini</i>	= <i>lini</i>

As far as syncretic forms are concerned, it is unclear whether NOM or GEN is involved.

However, among non-syncretic ones, when the (pragmatic) focus is 1SG, only NOM is possible, but when it is a non-SAP (be it singular or plural), GEN is required instead.

Despite its gloss GEN in (73)b, the form /*ini*/ does not index any possessor in that example.

It only means the construction requires a form that elsewhere indexes a non-SAP

⁸⁴ In Budai Rukai, one of two non-identical vowel sequences is subject to gliding and two identical vowel sequences undergo coalescence (cf. H. Huang 2006), so forms like /*amani=aku*/ and /*amani=ini*/ are phonetically realized as [a.ma.njá.ku] and [a.ma.ní.ni] respectively.

⁸⁵ These forms are identical to those in the two cited works except for =*lini* ‘3PL.GEN’, which was not included in either of the references. However, this particular form was listed in Y. Tang (2008: 16), found in C. Chen’s (2008: 272) naturally occurring texts, and also recognized by my consultants.

possessor (e.g. *paisu=ini* ‘{her/his} money’). The point is to show that it is not well-founded to determine the nominal status of a particular form simply based on whether or not that form can be the host of possessor clitics, which tend to be morphosyntactically diverse entities (e.g. regular nouns, negators, TAM, evidential/epistemic, or discourse markers, etc.) across Formosan languages (see §2.3.3).

Therefore, the (so-called finite) Mstem marked by nominal relation markers in Budai Rukai is better analyzed as grammatical nominalizations like those involving the Kstem-*ane* because both types illustrate various verbal properties internally, as L. Sung (2010) has shown, but show a robust nominal syntax externally. Functionally speaking, they both denote entity concepts that can be further acted or predicated upon, thus rendering arbitrary the distinction between “headless relativization” and nominalization as maintained in L. Sung (2010). Admittedly, the Mstem has some properties not shared by the Kstem-*ane*, but so is the case between the Kstem-*ane* and Kstem (so-called zero) nominalizations (see *ibid.*: 549). This should not pose a problem since they are all different constructions to begin with. The situation in Budai Rukai is similar to those in conservative Philippine-type languages like Tagalog (see §3.1.1) and Mayrinax Atayal (see §3.2.1), where researchers have no problems calling realis-indicative verbs nominalizations when they are marked by nominal relation markers and fulfill the argument function.

More importantly, based on the data in Zeitoun (1995), it is the norm rather than the exception among Rukai dialects to use the same realis-indicative Mstem to assert a predication and to denote the subject argument of a presupposed predication (i.e. grammatical nominalizations), and the functional distinction is achieved with the help of

nominal relation markers.⁸⁶ The only exception is Mantauran Rukai, where realis-indicative verbs take the Mstem (in the form of *o*-STEM, among others) whereas subject nominalizations (called subjective nominalizations in Zeitoun 2002) assume the *ta*-Kstem instead. Even in this exceptional language, the form *ta*-Kstem is not lexically a noun because several verbal properties are observed within the denoting phrase it is part of (ibid.: 276). To illustrate the point, Budai and Mantauran Rukai are contrasted in (75) and (76) respectively.

(75) Budai Rukai [Mstem *wa-bai* | Kstem *bai*]

- a. **wa-bai** ku cingpi musuane
 RLS-give OBL pencil 2SG.OBL
 ‘({She/He}) gave the pencil to you.’ (C. Chen 2008: 42)
- b. aneane [ku **wa-bai** musuane ku paisu]
 who NOM RLS-give 2SG.OBL OBL money
 ‘Who was it that gave you the money?’ (Zeitoun 1995: 360)

(76) Mantauran Rukai⁸⁷ [Mstem *o-va'ai* | Kstem *va'ai*]

- a. **o-va'ai-nga=na=inome** dhona'i vekenelre
 DYN-give-already=1EXCL.NOM=2PL.OBL that land
 ‘We already gave you {that/those} land(s).’ (Zeitoun 2007: 56)
- b. aanga=i [ta-va'a_=**im**ia'e paiso]
 who=3SG.GEN.VIS SBJ.NMLZ-give=2SG.OBL money
 ‘Who was it that gave you the money?’ (Zeitoun 1995: 360)

It seems to be no coincidence that Mantauran is also the only Rukai dialect where (prenominal) relation markers are extremely impoverished, and lack of such markers, which help to distinguish an assertion from an argument associated with that assertion in other Rukai dialects, is adequately compensated by the use of two distinct forms for the

⁸⁶ The Rukai Mstem takes the form of *wa*-STEM in Budai/Taromake/Labuan/Tona, *u*-STEM in Maga, and *o*-STEM in Mantauran, all containing the PAN *<um> based on Ross's (2015c) recent reconstructions.

⁸⁷ Following the notation in D. Yen & Billings (2011), I use the underscore “_” to indicate instances of vowel deletion. The morpheme =*i* in the (b) example alternates with =*ni*, and the conditioning factor is unclear (Zeitoun 2007). Despite its gloss GEN, the morpheme =*i* does not index the possessor in the present example, or more generally, when combined with interrogative words. The gloss GEN is used, following the practice in the Mantauran literature, simply because bound person forms like =(n)*i* do index the possessor elsewhere (e.g. *kapadha'anenga=i* ‘all his/her houses’; ibid.).

two functions, as in (76). In addition, like subject nominalizations in Budai (as in (75)b), the *ta*-Kstem in Mantauran cannot host possessor clitics as regular nouns do (Zeitoun 2002: 276, 2007: 359), which is, however, never cited as evidence against its nominalization function.

The comparison between Budai and Mantauran Rukai shows that whether languages use the same verb form to assert a state of affairs and to denote a participant in it is a language-specific issue and should have no direct bearing on how comparable linguistic expressions that we call nominalizations are delineated. More specifically, argument nominalizations are comparable across languages not because of the presence of certain markers with similar functional load (i.e. those that we can comfortably call nominalizers) but because of their shared function to denote an argument by means of a non-assertive predication expressed by whatever resources at the disposal of a language. The cross-dialectal comparison in Rukai also corroborates the idea that “whether or not a form in question has a finite verb form is to a large extent irrelevant” to nominalization (Shibatani 2009: 187).

3.4.2. Event/result nominalizations

Having argued that the Mstem in their referential and restricting functions are no less nominalized than the Kstem-*ane* based on their external distributions, I move on to investigate another claim in L. Sung (2010), namely, that the Mstem only nominalizes (or relativizes in the original wording) the subject whereas the Kstem-*ane* only non-subjects when they serve as argument nominalizations. While the first part of the claim is correct, the second part is not entirely true. Example (77) shows that the semantic role of the

subject of the Mstem internal to a grammatical nominalization is identical to that of the matrix nominalized argument (marked in NOM) as characterized by that very same Mstem.

(77) Budai Rukai (Fieldnotes) (revised from (63) above)⁸⁸

a. mu-a-bere [ku wa-lrumay ki camake]
 go-RLS-escape NOM RLS-beat OBL C.

‘The one who beat Camake ran away.’ (Active: SBJ=agent; NMLZ=agent)

b. mu-a-bere [ku ki-a-lrumay ki salrabu]
 go-RLS-escape NOM PASS-RLS-beat OBL S.

‘The one who was beaten by Salrabu ran away.’ (Passive: SBJ=patient; NMLZ=patient)

This means that the nominalized argument targets at the subject of the Mstem. By contrast, it is not true that the nominalized argument targets at non-subjects of the Kstem-*ane*. In this section, I argue the fact that the Kstem-*ane* (more specifically, *ta*-Kstem-*ane* for realis and *a*-Kstem-*ane* for irrealis) nominalizes its non-subject argument is just an epiphenomenon of its fundamental function to nominalize contingent results as invoked in an event nominalization like those in (71) and (72) above. In other words, the Kstem-*ane* denotes either an event or what pertains as a result of that event. This view is also congruent with C. Chen’s (2008: 126) semantic analysis of the nominalizer *-ane*, which “associates a consequent state to... events.” Given the same Kstem-*ane* form, whether the matrix predicate is event-oriented or participant-oriented plays an important role in whether it denotes an event as a whole or a specific participant in that event, as contrasted in (78).

⁸⁸ As an aid to facilitate comprehension of the data, SBJ (for subject) specifies the semantic role of the subject of a verb internal to a grammatical nominalization whereas NMLZ indicates the semantic role of the matrix nominalized argument (marked in NOM) as characterized by that verb.

(78) Budai Rukai (L. Sung 2011: 547, 534)⁸⁹

- a. *masamali=aku* [ku **ta-lrumadh-ane** ki salrabu ki lrailrai]
 surprised=1SG.NOM OBL RLS-beat-NMLZ GEN S. OBL L.
 ‘I am surprised at (the consequence resulting from) Salrabu’s beating Lrailrai.’
 [= (71)a]
- b. *ngi-a-caeme* [ka **ta-lrumadh-ane** ki salrabu ki muni]
 REFL-RLS-sick NOM RLS-beat-NMLZ GEN S. OBL M.
 ‘As a result of Salrabu’s beating her (=Muni), Muni (somehow) got sick.’

The semantic link is presumably motivated by a metonymic extension such that the event/result isomorphism (e.g. *painting* as a process of using paint and a result of that process) is widespread in diverse languages (Koptjevskaja-Tamm 2006).⁹⁰

The following demonstration supports the present analysis by showing (i) what argument the *Kstem-ane* seems to nominalize is subject to pragmatic inferences rather than determined by syntactic relations, and (ii) there are not only counterexamples but also ambiguous cases where both the subject and non-subject argument are potential targets of the *Kstem-ane* nominalization. I start with the active *Kstem-ane* and then move on to its passive counterpart, where there is more room for ambiguity.

Assuming for now with L. Sung (2010) that GEN marks the subject of the *Kstem-ane* (called genitive subject), OBL would mark a non-subject. As a result, the agent and patient are marked by GEN and OBL respectively in the active, but by OBL and GEN instead in the passive. Some initial examples I obtained are similar to (79) (see also (63)b), where

⁸⁹ The morpheme *ngi-* was given no functional gloss in the cited work. Compared with *ma-caeme* ‘RLS-sick’, which expresses the state of being sick, *ngi-a-caeme* in (78)b has an anticausative sense implying getting sick without an obvious cause. Since *ngi-* also has the reflexive function in the strict sense (i.e. someone acting on oneself) and that (strict) reflexive and anticausative functions are typically expressed by the same form across languages, *ngi-* is thus glossed REFL here.

⁹⁰ Metonymy also explains why the same “internally headed relative clauses” typically denote both an event and an argument in that event, depending on the matrix predicates they co-occur with. See Langacker (2009a) for such an analysis regarding Japanese data. As he points out, an analogy of the event/argument indeterminacy on the lexical level is to use *piano* to refer to the sound of a piano, as in *I can hear a piano*.

the active *Kstem-ane* does nominalize the non-subject patient argument, which is consistent with L. Sung's (2010) claim.

(79) Budai Rukai (Fieldnotes) (cf. (63)b)

mu-a-bere [ku **ta-sulav-ane** ki salrabu]

go-RLS-escape NOM RLS-cure-NMLZ GEN S.

'The one that Salrabu cured ran away.' (Active: SBJ=agent; NMLZ=patient)

[As a result of Salrabu's curing someone, that person ran away.]

However, as I changed the matrix predicate in (79), a different result emerged, as in (80), where the active *Kstem-ane* denotes the subject agent rather than the non-subject patient.

(80) Budai Rukai (Fieldnotes)

kuali [ku **ta-sulav-ane** ki salrabu]

tired NOM RLS-cure-NMLZ OBL S.

'The one that cured Salrabu is tired.' (Active: SBJ=agent; NMLZ=agent)

[As a result of someone's curing Salrabu, that person is tired.]

Here, the matrix predicate *kuali* 'tired' is attributed to the agent curer, which is most likely inferred from world knowledge about the process of a curing event. Aside from the matrix predicate, another crucial difference between (79) and (80) is whether the marker *ki* is interpreted as marking the subject agent, which is glossed GEN, or the non-subject patient, which is glossed OBL. By contrast, there is no such room for multiple interpretations when it comes to argument nominalizations involving the *Mstem*, as in (77), where the nominalized argument can only be the agent subject in active or patient subject in passive irrespective of pragmatic inferences.

To avoid the potential ambiguity caused by the marker *ki*, person names are replaced with person forms, which display distinct markings between GEN and OBL. It is found again that the active *Kstem-ane* can nominalize not only the non-subject patient, but also the subject agent, as in (81).

(81) Budai Rukai (Fieldnotes)

- a. mu-a-bere [ku **ta-sulav-ane=li**
 go-RLS-escape NOM RLS-cure-NMLZ=1SG.GEN
 ‘The one that I cured ran away.’ (Active: SBJ=agent; NMLZ=patient)
 [As a result of my curing somebody, (that person) ran away.]
- b. mu-a-bere [ku **ta-sulav-ane** **nakuane**]
 go-RLS-escape NOM RLS-cure-NMLZ 1SG.OBL
 ‘The one that cured me ran away.’ (Active: SBJ=agent; NMLZ=agent)
 [As a result of somebody’s curing me, (that person) ran away.]

Similar results pertain to the passive *Kstem-ane*. With person-form arguments, the interpretations are rather consistent across speakers, as in (82), where the passive *Kstem-ane* nominalizes not only the non-subject agent, but also the subject patient.

(82) Budai Rukai (Fieldnotes)

- a. mu-a-bere [ku **ta-ki-sulav-ane=li**
 go-RLS-escape NOM RLS-PASS-cure-NMLZ=1SG.GEN
 ‘The one I got cured by ran away.’ (Passive: SBJ=patient; NMLZ=agent)
 [As a result of my being cured by someone, that person ran away.]
- b. mu-a-bere [ku **ta-ki-sulav-ane** **nakuane**]
 go-RLS-escape NOM RLS-PASS-cure-NMLZ 1SG.OBL
 ‘The one that got cured by me ran away.’ (Passive: SBJ=patient; NMLZ=patient)
 [As a result of somebody’s being cured by me, that person ran away.]

With full-nominal arguments, however, there are some cross-speaker variations, most likely due to the dual functions of the marker *ki* for both GEN and OBL. Upon hearing (83), one speaker volunteered two interpretations, which differ in terms of whether the matrix argument is the non-subject agent or subject patient of the passive *Kstem-ane*.

(83) Budai Rukai (Fieldnotes; Speaker A) (cf. (82)a)

- mu-a-bere [ku **ta-ki-sulav-ane** ki salrabu]
 go-RLS-escape NOM RLS-PASS-cure-NMLZ=1SG.GEN GEN S.
 ‘The one that Salrabu got cured by ran away.’ (Passive: SBJ=patient; NMLZ=agent)
 [As a result of Salrabu’s being cured by someone, that person ran away.]
 ‘Salrabu ran away after being cured.’ (Passive: SBJ=patient; NMLZ=patient)
 [As a result of her (i.e. Salrabu’s) being cured by someone, Salrabu ran away.]

By contrast, another speaker interpreted the same sentence as in (84), where the marker *ki* was understood as introducing the non-subject agent (glossed as OBL) instead of the subject patient (glossed as GEN).

(84) Budai Rukai (Fieldnotes; Speaker B) (cf. (82)b)

mu-a-bere [ku **ta-ki-sulav-ane** ki salrabu]
 go-RLS-escape NOM RLS-PASS-cure-NMLZ OBL S.
 ‘The one that got cured by Salrabu ran away.’ (Passive: SBJ=patient; NMLZ=patient)
 [As a result of somebody’s being cured by Salrabu, that person ran away.]

One of the interpretations in (83) and the one in (84) are both counterexamples to the claim that the *Kstem-ane* nominalizes only non-subjects.

Therefore, examples (79) through (84) show that both the subject (marked in GEN) and the non-subject (marked in OBL) of the *Kstem-ane* nominalization, in both the active and passive construction, can potentially be the argument invoked as the result of an event, contrary to the claim in L. Sung (2010). It is of special interest to note that precisely because of this flexibility the passive construction in (84) ends up having the same propositional meaning as its active counterpart in (79) and they differ only in terms of the perspective in which an event is construed. Both examples are repeated in (85) for ease of comparison.

(85) Budai Rukai (Fieldnotes)

a. mu-a-bere [ku **ta-sulav-ane** ki salrabu]
 go-RLS-escape NOM RLS-cure-NMLZ GEN S.
 ‘The one that Salrabu cured ran away.’ (Active: SBJ=agent; NMLZ=patient)
 [As a result of Salrabu’s curing someone, that person ran away.] [= (79)]

b. mu-a-bere [ku **ta-ki-sulav-ane** ki salrabu]
 go-RLS-escape NOM RLS-PASS-cure-NMLZ OBL S.
 ‘The one that got cured by Salrabu ran away.’ (Passive: SBJ=patient; NMLZ=patient)
 [As a result of somebody’s being cured by Salrabu, that person ran away.] [= (84)]

The matrix nominalized argument in the first example is construed with respect to the non-subject patient of the active *Kstem-ane*, but that in the second example in terms of

the subject patient of the passive *Kstem-ane*. The present analysis also explains C. Kuo's (1979) data, cited in L. Sung (2010) and reproduced in (86).

(86) Budai Rukai (L. Sung 2010: 535; citing C. Kuo 1979: 19)

ni-(ki)-kupa-ane kai adhadhame
 PFV-PASS-steal-NMLZ PROX bird
 'This bird is a stolen one.'

If it were the case that the *Kstem-ane* only nominalizes non-subjects, the optionality of the passive marking in (86) would be just an accidental qualm because the patient (i.e. the bird) is the non-subject in the active (i.e. *ni-kupa-ane*) but the subject in the passive (i.e. *ni-ki-kupa-ane*). In light of the data in (85), however, (86) can be accounted for by different event construals. Without the passive marker, the nominalized argument invoked by a stealing event in (86) is the non-subject patient, but with the passive marker it becomes the subject patient. This is again in sharp contrast with argument nominalizations with the *Mstem*, where the passive marking always has import on the semantic role of the nominalized argument.

Finally, the conclusion that the *Kstem-ane* form may denote an event or a participant invoked by an event agrees well with the fact that when the *Kstem* is replaced with an underived noun *X*, the result form *X-ane* may denote what is metonymically associated with *X* (see §6.4.1).

3.5. Puyuma

Among Puyuma dialects, there is a general noticeable divide between Nanwang (or *puyuma* in the vernacular) and non-Nanwang. Those of the latter group often delineated in the literature include at least (from north to south) Ulivelivek, Pinaski,

Tamalakaw, Rikavung, Kasavakan, and Katripul.⁹¹ Nanwang has been extensively investigated in S. Teng (2007, 2008), from which many Nanwang examples demonstrated in this section are drawn, and of non-Nanwang dialects Tamalakaw is the one with most detailed grammatical descriptions (Tsuchida 1980, 1992a), a variety most closely related with and geographically adjacent to Rikavung, which I will turn to for resolving a nominal-verbal debate in Puyuma. In terms of phonology, Nanwang is more conservative whereas with respect to morphosyntax non-Nanwang dialects are more conservative (S. Teng 2009).

In this section, I first examine S. Teng's (2008) claim (as quoted in §3.2.3) that there are two types of relative constructions in Nanwang Puyuma, with one involving "finite clauses" and the other "nominalized clauses" (§3.5.1). Then I move on to discuss a nominal-verbal debate in Puyuma (§3.5.2) since it has far-reaching implications for the Nuclear Austronesian hypothesis, which will be reevaluated against data from both Nanwang and non-Nanwang dialects.

3.5.1. AF-NAF asymmetry in nominalizations

Before looking into nominalizations in Nanwang Puyuma, it is necessary to first lay out the grammatical differences between nouns and verbs at the word level as well as the verb forms used as matrix predicates. According to S. Teng (2008: 49), lexical verbs

⁹¹ S. Teng (2009) mentioned yet another non-Nanwang village, Alipay, which is located to the north of Pinaski. The reason Alipay is not listed here is that the village is extremely Sinicized compared with the other seven. Loren Billings and I visited all eight villages in 2015 and were able to find some fluent speakers in all except Alipay, where the only local person we could find possessing some conversational fluency in Puyuma is an ethnic Chinese man (born in 1938). He claimed to start learning Puyuma at the age of 52. Although he is a late learner, we were impressed by his pronunciation and his ability to produce correct sentences with the expected Focus-words in all the various syntactic contexts that we tested. However, the data elicited from him were not included in our joint research.

and nouns in Nanwang can be distinguished by the two syntactic tests in (87), which are respectively illustrated in (88) and (89).

(87) Verb-noun distinctions at the word level in Nanwang Puyuma (S. Teng 2008: 49)⁹²:

- a. Verbs are negated by *adri* whereas nouns by *amelri*.
- b. Both NAF verbs and nouns can host person-form proclitics, but only nouns collocate with free possessive person forms.

(88) Nanwang Puyuma (S. Teng 2008: 49)

- a. {*adri*/**amelri*} **saygu**
 NEG AF.capable
 ‘(She/He) cannot (do it).’
- b. {**adri*/*amelri*} **a suan**
 NEG TOP.INDF dog
 ‘(It is) not a dog.’

(89) Nanwang Puyuma (S. Teng 2009: 826)⁹³

- a. **ku=rungas-aw ku=kiruwan**
 1SG.ACT=take.off-PF 1SG.GEN=clothes
 ‘I took off my clothes.’

⁹² The cited work (p.57) also touches upon verbal-nominal distinctions at the root level, which are not discussed here. To summarize, verbal and nominal roots are not precatégorial in most cases because they are identifiable by morphological markedness. For instance, verbal roots are those that can be used in imperative constructions without additional morphology whereas nominal roots are those that can be used as arguments without being additionally affixed. By this criterion, only a small number of roots are both verbal and nominal (e.g. *senay* meaning both “sing” as a command and “song” as a piece of music).

⁹³ The original transcription for “clothes” is *kiruan*, which is changed here to *kiruwan* [ki.ru.wán]. Two other changes made here have to do with glossing. In the cited work, “Gen” (for genitive) and “PRS” (for possessor) are used to gloss the person-form non-Topic Actor and the possessor respectively. To be consistent with the glossing principles adopted here, Gen is replaced with ACT and PRS with GEN, although bound person forms in Nanwang for both functions are identical in form across all persons/numbers (see S. Teng 2015: 410). In addition, free possessive person forms are further decomposed into a nominal relation marker, the linking sound /n/, and a person-form clitic, following the analysis in S. Teng (2009, 2015). The morphemic breakdown will help us understand later in §3.5.2 why some Focus-words collocate with free possessive person forms while others do not. The linking sound /n/ is homorganic with the following consonant. This is comparable to so-called interfixes found within a compound in Germanic languages (e.g. German *Schwan-en-gesang* ‘swansong’; see Haspelmath 2002: 86 for details). Even more relevant is the fact that, as Halpern (2001) points out, idiosyncratic phonological alternations are often found between special clitics and their host or between two special clitics. An example he gave (citing Simpson & Withgott 1986: 167) is from a French dialect, where /z/ is inserted between two special clitics, as in *Donnez=moi=z=en* ‘Give me some’. This clitic-internal linking sound is glossed as LNK, to be distinguished from LIG for the attributive ligature that connects the modifier and the modifiee (Foley 1976).

- b. **ku=rungas-aw** **na=n=ku** kiruwan
 1SG.ACT=take.off-PF TOP.DEF=LNK=1SG.GEN clothes
 ‘I took off my clothes.’
- c.* **na=n=ku** **rungas-aw** **ku=kiruwan**
 TOP.DEF=LNK=1SG.ACT take.off-PF 1SG.GEN=clothes

Next, as has been mentioned in §2.3.1, Puyuma is unique among Philippine-type languages in that it uses first-generation NAF affixes like *-aw/-ay/-anay* in the affirmative realis-indicative, unlike all the other languages that have these cognate forms (except for *-ai* in Kanakanavu, reflecting PAn LF *-ay; see §4.4.1). By contrast, the AF construction in the realis-indicative, be it affirmative or negative, makes use of the Mstem as in other languages. Relevant examples are repeated in (90).

(90) Nanwang Puyuma (S. Teng 2008: 147) [= (4) in Chapter 2]

- a. **trakaw** dra paisu i isaw
 <AF>steal UND.INDF money TOP.SG I.
 ‘Isaw stole money.’
- b. **tu=trakaw-aw** na paisu kan isaw
 3.ACT=steal-PF TOP.DEF money ACT.SG I.
 ‘Isaw stole the money.’
- c. **tu=trakaw-ay=ku** dra paisu kan isaw
 3.ACT=steal-LF=1SG.TOP UND.INDF money ACT.SG I.
 ‘Isaw stole money from me.’
- d. **tu=trakaw-anay** i tina=taw dra paisu
 3.ACT=steal-CF TOP.SG mother=3.GEN UND.INDF money
 ‘{He/She/They} stole money for {his/her/their mother}.’

Moreover, the Mstem in Nanwang Puyuma can be categorized into at least six morphological classes depending on how it alternates with its corresponding Kstem, which is the base of NAF affixes *-aw/-ay/-anay* as well as the form for AF imperatives (among others). The six classes are each illustrated with one Mstem/Kstem pair in Table 3.4 (cf. Table 2.1 for a similar table in Tamalakaw Puyuma).

Table 3.4: Classes of AF verbs in Nanwang Puyuma (after S. Teng 2008: 123)

	I	II	III	IV	V	VI
Mstem	<i>kasu</i> '<AF>bring'	<i>ma-dreki</i> 'AF-scold'	<i>aremeng</i> 'AF.dark'	<i>mi-kiping</i> 'AF.have-clothes'	<i>ma-rengay</i> 'AF-tell'	<i>beray</i> 'AF.give'
Kstem	<i>kasu</i> 'bring'	<i>ka-dreki</i> 'K-scold'	<i>k-aremeng</i> 'K-dark'	<i>pi-kiping</i> 'K.have-clothes'	<i>rengay</i> 'tell'	<i>beray</i> 'give'

Both stems can undergo *Ca-* reduplication or *a-* affixation (depending on stems) to produce the Imperfective form (or schematically α -Mstem/ α -Kstem; following Ross 2015a), as shown in Table 3.5 below. While the AF α -Mstem expresses ongoing or habitual events (among others), the AF α -Kstem describes future ones (among others), as in (91).

(91) Nanwang Puyuma (S. Teng 2008: 135, 208)

- a. *sa~salrpit=ku*
 <AF>IPFV~flog=1SG.TOP
 'I am flogging (somebody with a twig).' [Mstem: *salrpit* | Kstem: *salrpit*]
- b. *pa~pulrang=ku*
 IPFV~help=1SG.TOP
 'I will help.' [Mstem: *p<en>ulrang* | Kstem: *pulrang*]

In what follows, I first present S. Teng's (2008) analysis regarding nominalization, and then point out its problems, some of which are similar to those that have been outlined in §3.2.

To begin with, a distinction between lexicalized and gerundive nominalization was made in S. Teng (2008). Specifically, "[t]he difference between these two processes is evident in that gerundive nominals are productive, may have an argument NP licensed by the valency of the stem, and are negated like a verbal construction, whereas lexicalised

Table 3.5: Imperfective forms in Nanwang Puyuma (after S. Teng 2008: 123)⁹⁴

	I	II	III
α -Mstem	<i>ka~kasu</i> '<AF>IPFV~bring'	<i>ma-dra~dreki</i> 'AF-IPFV~scold'	<i>a<ra~>remeng</i> '<IPFV~>AF.dark'
α -Kstem	<i>ka~kasu</i> 'IPFV~bring'	<i>ka-dra~dreki</i> 'K-IPFV~scold'	<i>k-a<ra~>remeng</i> 'K-<IPFV~>dark'
	IV	V	VI
α -Mstem	<i>mi-a-kiping</i> 'AF.have-IPFV-clothes'	<i>ma-ra-rengay</i> 'AF-IPFV~tell'	<i>ba~beray</i> 'IPFV~AF.give'
α -Kstem	<i>pi-a-kiping</i> 'K.have-IPFV-clothes'	<i>ra~rengay</i> 'IPFV~tell'	<i>ba~beray</i> 'IPFV~give'

nominalizations are not productive and are negated like a nominal construction” (ibid.: 129-130). By these very same criteria (i.e. negation, productivity, and argument-taking), I will compare so-called lexicalized and gerundive nominalizations consisting of NAF-words alongside with constructions made up of AF-words that serve similar syntactic functions.

When NAF-words syntactically occur in the argument position and semantically denote the non-Actor, they do not assume forms like the Kstem-*aw/-ay/-anay* in (90), but undertake <in>Kstem-(*an*) instead (among others), where <in> expresses perfective

⁹⁴ The Mstem, α -Mstem, Kstem, and α -Kstem are respectively referred to as Realis Unmarked, Realis Progressive, Imperative, and Irrealis in the cited work. While these functional labels do describe their prominent functions, these four types of forms have much more functions than the labels would suggest (see S. Teng 2008 for details). Thus, they are mentioned here by their formal features. For ease of reference, the α -Mstem and α -Kstem are collectively called the Imperfective form due to the same morphological process they both involve. Finally, a seventh class is included in the original reference, but it has identical forms for both the Mstem and Kstem just like Class VI, from which the seventh class is only distinguishable in terms of how the Mstem/Kstem pair is changed into the Imperfective form. For instance, the Imperfective form of the isomorphic Mstem/Kstem *ki-lrengaw* ‘(AF.)get-sound’ is *ki-a-lrengaw* ‘(AF.)get-IPFV-sound’.

aspect by default as well as relative past time and *-an* indicates nominalization.⁹⁵ In terms of negation, the form *<in>Kstem-(an)* was said to be either a lexicalized or gerundive nominalization (S. Teng 2008: 131) depending on whether it is negated by the nominal or verbal pattern, as illustrated in (92) (cf. (88) above).⁹⁶

(92) Nanwang Puyuma (S. Teng 2008: 131)

- a. **amelri** [na=n=tu **ni-lradra~lradram-an**] ta=ngai
 NEG TOP.DEF=LNK=3.ACT NAF.RLS-ITER~know-NMLZ 1INCL.GEN=language
 ‘Our language is not what they have been learning.’
- b. uwa alak [dra pa-trungtrung-an] [dra **adri=driya**
 go.IMP AF.take UND.INDF CAUS-sound-NMLZ UND.INDF NEG=still
 b<in>arekep-an dra kulitr]
 <NAF.RLS>assemble-NMLZ OBL.INDF skin
 ‘Go get a drum (i.e. the thing that makes sounds) that has not yet been assembled with skin.’

However, this analysis suffers from confusing scope of negation with the lexicalized/gerundive distinction it aims to illustrate. In the present analysis, the form *<in>Kstem-(an)* in the first example forms a referential NP, which is then negated by the nominal pattern, so what is negated is the identifying/classifying relationship between two nominals. In the second example, however, the negation has scope over a verbal predicate, which is externally nominal (as evidenced by the marking of *dra*) and denotes something characterized by a negative predication, which is further used as a restricting

⁹⁵ The form *<in>Kstem-(an)* is meant to be schematic, covering word forms in the shape of the Kstem infixed by *<in>* (or its allomorphs *in-* and *ni-*) and optionally suffixed by *-an*. Interestingly, the phonological constraints that determine the selection of the three allomorphs *<in>*, *in-*, and *ni-* are also applicable to the three AF allomorphs **, *m-*, and *me-* (see S. Teng 2008: 26-27). Unless the stem undergoes additional morphological processes (such as *Ca-* or CVCV- reduplications; see Table 3.6) to convey imperfectivity, the infix *<in>* expresses perfective aspect as well as relative past time by default. Because *<in>* is found in word forms with not only perfective but also imperfective meanings, it is glossed as RLS instead of PFV, which deviates from the practice in the literature (see also the analysis of Tagalog *<in>* in Himmelmann 2005b). Moreover, unlike other Puyuma dialects, Nanwang Puyuma does not distinguish realis nominalizations for different NAF categories (see S. Teng 2012 and §4.5.2). Thus, the gloss NAF is used for *<in>* instead of Focus-specific labels such as PF/LF/CF.

⁹⁶ When discussing both lexicalized and gerundive nominalizations, S. Teng (2008) commented that the suffix *-an* is optional when the infix *<in>* is present. Thus, the presence or absence of the suffix *-an* does not play a role in distinguishing the two proposed types of nominalizations.

phrase to modify another nominal (i.e. a drum). In other words, negation is external to a nominalization in (92)a, but internal to the one in (92)b, and the contrast has nothing to do with the lexicalized/gerundive nature of a nominalization. On the other hand, when Focus-words syntactically occur in the argument position and semantically denote the Actor, they assume the same Mstem as when they function as the matrix predicate. Probably because of this, the Mstem in its referential and restricting function was not analyzed as nominalized, but constantly “finite” or “relativized” instead (S. Teng 2008: 105). However, like *<in>Kstem-(an)* in (92), the Mstem, regardless of its morphological classes (see Table 3.4), can not only be negated by the nominal and verbal pattern, but also occur in a referential or restricting phrase, as in (93).

(93) Nanwang Puyuma (S. Teng 2008: 191, 69)⁹⁷

- a. **amelri** [a **seneng**] na unan
 NEG TOP.INDF <AF>special TOP.DEF snake
 ‘The snake was not a special one.’
- b. **driyama** idrini i, taturu [kana trau]
 so TOP.PROX PTOP <AF>warn UND.DEF person
 [kana **adri** **iraman** kantu walrak]
 UND.DEF NEG AF.commiserate UND.DEF.3.GEN child
 ‘So, this (story), (it is meant to) warn those people who do not commiserate with their children.’

If the *<in>Kstem-(an)* form in (92)a were really lexicalized as claimed, so should the Mstem in (93)a be based on the same negation factor, but such an analysis is not found in S. Teng (2008). The present analysis, however, gives a consistent treatment to the Mstem in (93) and *<in>Kstem-(an)* in (92) as grammatical nominalizations, which are internally verbal, based on the negation pattern and their ability to take arguments subcategorized by the Mstem/Kstem, and at the same time externally nominal, as evidenced by their

⁹⁷ The original transcription for the nominal relation marker *na* in (93)a is *ina*, which is very likely a typo judging from the original gloss for it as well as the fact that *ina* means ‘mother/aunt’.

collocation with nominal relation markers and their syntactic functions as arguments, modifiers of arguments, or (negated) nominal predicates.⁹⁸ More importantly, a denoting phrase consisting of the Mstem or <in>Kstem-(*an*) is syntactically on an equal footing with a typical NP comprised of underived nouns, which is supported by two distribution facts: (i) the order between the former (i.e. the modifying phrase) and the latter (i.e. the modified phrase) is rather flexible, and (ii) the former can syntactically dispense with the latter, thus serving as an independent NP on its own (see *ibid.*: 104-105). Therefore, the modifying phrase with the Mstem in (93)b should be no less nominalized than that with <in>Kstem-(*an*) in (92)b, which is in sharp contrast to the claim that there are nominalized and finite (thus non-nominalized) relative constructions in Nanwang Puyuma (*ibid.*: 105). In addition, just as the Mstem in (93)a is no more lexicalized than that in (93)b simply because of the different negation patterns they illustrate, so too the <in>Kstem-(*an*) in (92)a is no more lexicalized than that in (92)b.

Second, while it is generally true that lexicalized nominalizations are less productive than gerundive ones, the productivity criterion is not helpful in determining whether NAF nominalizations are lexicalized or gerundive because there seem to be as many, if not more, instances of the <in>Kstem-(*an*) form that can be negated by the nominal pattern, which was assumed to imply lexicalized nominalizations, as those that can be negated by the verbal pattern, which was taken to be an indication of gerundive nominalizations. Just as instances of the Mstem applicable to the nominal and verbal negation are equally productive (see (93) above), so are those of the <in>Kstem-(*an*) form (see (92) above).

⁹⁸ In S. Teng's (2008) tree diagrams, however, the terminal node for the Mstem in its referential/restricting functions is lexically a verb but that for the <in>Kstem-(*an*) in the same functions is lexically a noun.

The last criterion for the proposed distinction between lexicalized and gerundive nominalizations in S. Teng (2008) is whether they can take patientive arguments. “[S]ome nominals can take arguments and are therefore evidently gerundive” (ibid.: 142). Thus, the NAF nominalization in (94) would be gerundive because it takes a non-Topic patientive argument (i.e. *walrak* ‘child’).

(94) Nanwang Puyuma (S. Teng 2008: 105)

tu=lrasedr-aw=dar i tralu~tralun [na=n=tu **in-abak-an**
3.ACT=hide-PF=often LOC AUG~grass TOP.DEF=LNK=3.ACT NAF.RLS-pack-NMLZ

kana walrak] [na padrekan]
UND.DEF child TOP.DEF backpack

‘She often hid in the grass the backpack where she packed the child.’

However, the very same nominalization can in fact be negated by the nominal pattern, which would suggest a lexicalized nominalization based on S. Teng’s (2008) negation criterion, as shown in (95) (cf. (92) above).

(95) Nanwang Puyuma (Fieldnotes; based on S. Teng 2008: 105)

amelri [na=n=ku **in-abak-an** **kana** belrbelr]
NEG TOP.DEF=LNK=3.ACT NAF.RLS-pack-NMLZ UND.DEF banana

[idrini na padrekan]
TOP.PROX LIG backpack

‘This backpack is not where I packed the bananas.’

In other words, (95) would produce conflicting results according to two of the criteria that S. Teng (2008) proposed to differentiate lexicalized from gerundive nominalizations. In the present analysis, however, (95) is not only unproblematic but also expected since argument-taking is a verbal property subcategorized by the <in>Kstem-(an) internal to the nominalization whereas nominal negation applies to the nominalization as a whole without necessarily implying that any constituent member of it is lexically a noun. Similarly, AF nominalizations may take non-Topic patientive arguments, and whether

argument-taking is permissible or not is determined by the valency of the Mstem, as illustrated in (96).

(96) Nanwang Puyuma

- a. amau kuiku [na **sa~senay**]
 NEG 1SG TOP.DEF <AF>IPFV~sing
 ‘The one who was singing is me.’ (S. Teng 2008: 193)
- a. amau kuiku [na **sermud kana tugi**]
 NEG 1SG TOP.DEF <AF>push UND.ASSO.PL T.
 ‘The one who pushed Tugi and his associates is me.’ (Fieldnotes)

Just as it is unfeasible to maintain that the Mstem with a patientive argument in (96)b is more gerundive than the one without any in (96)a, it is equally problematic to hold that the <in>Kstem-(an) form with a patientive argument, as in (94), is more gerundive than the one without any. In fact, S. Teng (2008: 142-143) admitted that “when there is no argument present, we often cannot tell whether a given construction is a lexical nominal or a gerundive nominal.” A less deceptive conclusion to draw is that the argument-or-not criterion simply does not contribute to the proposed lexical/gerundive distinction, which is merely a matter of contingent realizations of potential arguments subcategorized by the Mstem/Kstem.

So far, I have shown that the proposed criteria for distinguishing lexicalized from gerundive nominalizations in the shape of <in>Kstem-(an) are not reliable. In fact, the putative distinction would disappear once the internal and external syntax of nominalizations are kept apart. In addition to morphosyntax, the present analysis also finds support in semantics. For instance, *k<in>a-bekas-an* ‘<NAF.RLS>K-run-NMLZ’ and *k<in>-iedreng-an* ‘<NAF.RLS>K-sleep-NMLZ’ respectively denote “place having been traversed” and “place having been slept at”, and both were given by S. Teng (2008: 138) as examples of lexical nominalizations that denote locations. However, these nominals

“do not refer to places typically associated with the named event; they can refer to any place where the named action has taken place” (ibid.). The semantically extemporized nature thus casts doubt on their status as lexical nominalizations. Moreover, she (ibid.: 139) concluded that locative nominals “with <in> are in fact nominalised RCs without a PIBU”, where PIBU (for primary information-bearing unit) is a term taken from Croft (2001: 259) meaning “the most contentful item that most closely profiles the same kind of thing that the whole constituent profiles.” Crucially, it seems conflicting to claim that something is lexically a locative noun but at the same time a nominalized RC without a PIBU. This inconsistency can be avoided if the observed nominal properties are not attributed to the <in>Kstem-(an) word form at the lexical level (, which would give rise to putative lexical nominalizations), but to the overall construction it is part of, together with its potential syntagmatic constituents such as the verbal negator.

In addition, the same analysis can also be applied to AF nominalizations (consisting of the Mstem as well as its morphological derivatives; see below), and it renders superfluous the attempt to make a distinction that is not motivated by the grammar of Puyuma. For instance, a construction containing the Imperfective form of the Mstem was analyzed by S. Teng (2008: 135) as a “relative clause” in (97)a, where it modifies a nominal, but as indeterminate “between a noun and a relative clause” in (97)b, where it constitutes a referential NP.

(97) Nanwang Puyuma (S. Teng 2008: 135)

- a. indang=ku kana trau [kana **sa~salrpit**]
 AF.afraid=1SG.TOP OBL.DEF person OBL.DEF <AF>IPFV~flog
 ‘I am afraid of the person who is flogging (somebody with a twig).’
- b. indang=ku [kana **sa~salrpit**]
 AF.afraid=1SG.TOP OBL.DEF <AF>IPFV~flog
 ‘I am afraid of the flogger (, who does so by using a twig).’
 ‘I am afraid of the one who is flogging (somebody with a twig).’

However, the bracketed construction in both examples can be given a unified analysis as grammatical nominalizations without a lexical noun. That is, the Mstem (as well as its morphological derivatives) is lexically a verb ready to take on any verbal properties to the exclusion of underived nouns, but at the construction level has syntactic functions parallel to underived nouns once marked by nominal relation markers, thus illustrating the modification-use in (97)a and the NP-use in (97)b. Moreover, the habitual and progressive readings of (97)b do not provide any good basis for a distinction “between a noun and a relative clause”, which is only as real as the free translations provided, because the two readings are equally available in matrix predicates, as in (98), and thus independent of the nominalization process.

(98) Nanwang Puyuma (S. Teng 2008: 116)

- m-a-ekan** dra kuraw
 AF-IPFV-eat UND.INDF fish
 ‘{She/He} has the habit of eating fish.’
 ‘{She/He} {is/was} eating fish.’

Among various forms involving the Mstem, only the Imperfective α -Mstem was singled out by S. Teng (2008: 133-134) to be “person-denoting nouns”, that is, “nouns denoting the persons that carry out the action denoted by the verb.” However, the α -Mstem is merely one of the many morphological derivatives of the Mstem, itself included, that can denote the executor of an action (or the Actor in general). The Mstem

denotes the Actor in realis-perfective events and is the base on which various morphological processes are operated to give rise to the Imperfective (via *Ca*-reduplication or *a*- affixation, depending on stems) and Iterative forms (via CVCV-reduplication).⁹⁹ Likewise, the *<in>Kstem-(an)* form denotes non-Actors in realis-perfective events and derives its corresponding Imperfective and Iterative forms in the same manner as the Mstem does. The Focus-words in AF and NAF nominalizations are summarized in Table 3.6 and illustrated in (99).¹⁰⁰

Table 3.6: Focus-words in realis AF and NAF nominalizations in Nanwang Puyuma

	Perfective	Imperfective	Iterative
AF forms in nominalizations	Mstem	<i>α</i> -Mstem	CVCV~Mstem
NAF forms in nominalizations	<i><in>Kstem-(an)</i>	<i><in>α-Kstem-(an)</i>	<i><in>CVCV~Kstem-(an)</i>

(99) Nanwang Puyuma (Fieldnotes)

- a. amau i tugi [na {trakaw/tra~trakaw
 COP TOP.SG T. TOP.DEF {<AF>steal/<AF>IPFV~steal
 traka~trakaw} kana bunga
 <AF>ITER~steal} UND.DEF sweet.potato
 ‘The one who {stole/is stealing/repeatedly steals} the sweet potatoes is Tugi.’
- b. na=n=tu {tr<in>akaw/tr<in>a~trakaw/tr<in>aka~trakaw}
 TOP.DEF=LNK=3.ACT {<NAF.RLS>steal/<NAF.RLS>IPFV~steal/<NAF.RLS>ITER~steal}
 kan tugi i, amau a bunga
 ACT.SG T. PTOP COP TOP.INDF sweet.potato
 ‘What Tugi {stole/is stealing/repeatedly steals} is sweet potatos.’

⁹⁹ The term Iterative refers to a specific case of imperfective meanings, one that involves repetitive action. See Zeitoun & C. Wu (2006) for an overview of reduplication in Formosan languages, which typically expresses imperfective meanings of various types.

¹⁰⁰ The form *<in>CVCV~Kstem-(an)* was not listed in S. Teng’s (2008: 141) table, but its instances are found in one of her examples (see (92)a above) as well as in (99)b.

Despite the same morphological processes equally applicable to the AF Mstem and NAF $\langle in \rangle$ Kstem-(*an*), S. Teng (2008: 140) concluded that instances of the $\langle in \rangle$ Kstem-(*an*) are nominalizations, which “express aspectual and modal categories by morphology alone” (i.e. *Ca-* reduplication or *a-* affixation), whereas those of the Mstem are verbal constructions, where “some categories are expressed by morphology and some by clitics.” However, the conclusion was based on the invalid assumption that the Mstem marked by those clitics alluded to in the quote, namely, *=lra* ‘already’, *=driya* ‘still’, and *=dar* ‘often’, expresses “the same aspectual categories” as NAF-words like $\langle in \rangle$ Kstem-(*an*), $\langle in \rangle\alpha$ -Kstem-(*an*), and $\langle in \rangle\alpha\text{-}\alpha$ -Kstem-(*an*) respectively, as summarized in Table 3.7.¹⁰¹

Table 3.7: Aspectual forms of AF and NAF in Nanwang Puyuma (as per S. Teng 2008: 140-141)

	Perfective	Imperfective	Frequentative
AF	Mstem= <i>lra</i>	Mstem= <i>driya</i>	Mstem= <i>dar</i>
NAF	$\langle in \rangle$ Kstem-(<i>an</i>)	$\langle in \rangle\alpha$ -Kstem-(<i>an</i>)	$\langle in \rangle\alpha\text{-}\alpha$ -Kstem-(<i>an</i>)

It is immediately questionable that temporal modifications made by clitics like *=lra*, *=driya*, and *=dar* are functionally equivalent to those achieved through morphological operations like reduplication/affixation. First, at least for the first two clitics, they

¹⁰¹ The word form $\langle in \rangle\alpha\text{-}\alpha$ -Kstem-(*an*), not shown in Table 3.6, results from the NAF form $\langle in \rangle$ Kstem-(*an*) having undergone serial *Ca-* reduplication or *a-* affixation. This morphological process is also applicable to the Mstem, and the result form $\alpha\text{-}\alpha$ -Mstem (e.g. *tra~tra~trekel* ‘<AF>IPFV~IPFV~drink’) was called durative in S. Teng (2008: 117). However, both AF $\alpha\text{-}\alpha$ -Mstem and NAF $\langle in \rangle\alpha\text{-}\alpha$ -Kstem-(*an*) can express prolongation of actions or intensification of states, and thus deserve a shared aspectual label, whatever that should be. Similarly, as a result of the aspectual equivalence assumed in Table 3.7, the NAF $\langle in \rangle\alpha$ -Kstem-(*an*) was called imperfective but the AF α -Mstem progressive instead. However, both forms can be used to describe an on-going event, as shown in (99).

collocate with not only the Mstem but also its morphological derivatives such as Imperfective α -Mstem and Iterative CVCV~Mstem, as in (100).

(100) Nanwang Puyuma (S. Teng 2008: 118)

a. *m-a-ekan=lra* ‘AF-IPFV-eat=already’

b. *ki-lrenga~lrengaw=ta=lra* ‘get-ITER~sound=1INCL.TOP=already’

This suggests that the two types of operations independent of each other. Moreover, attributing perfectivity to the clitic =*lra*, as was done in S. Teng (2008), would lead to semantically incongruous cases where a putative perfective marker attaches to verb forms that clearly express imperfective meanings (i.e. the Imperfective and Iterative form). Second, the aforementioned temporal clitics also attach to underived nouns and have scopes over the entire phrase of which the host noun is a constituent, as in (101).

(101) Nanwang Puyuma (S. Teng 2008: 62, 294)

a. a *lralrak=ku=driya*
TOP.INDF child=1SG.TOP=still
‘I was still a child.’

b. *adri atelr-an* i *drenan=lra* ka
NEG abandon-CF.IMP LOC mountain=already <AF>say
“‘Don’t abandon (her) in the mountains,” (he) said.’

Thus, temporal clitics are more unselective as to both the target and scope of their applications than the reduplication/affixation processes that indicate aspectual distinctions. Finally, precisely because of the promiscuous nature of temporal clitics, it is not surprising to find that the clitic =*lra* attaches to not only the AF Perfective Mstem (e.g. *m-uberek=lra* ‘AF-return=already’; *ibid.*: 181), but also the NAF Perfective <in>Kstem-(*an*) (e.g. *dr<in>ua-an=lra* ‘<NAF.RLS>come-NMLZ=already’; *ibid.*: 267).¹⁰²

¹⁰² Sagart (2013: 487) points out that Puyuma =*lra* (or <la> in his rendition) is more a marker signaling new or unexpected situations than a perfective marker in the strict sense, although its effect does include creating perfective meanings. Markers with similar functional load are widely attested in Formosan languages, and they are sometimes glossed as COS for ‘change of state’, or simply as ‘already’. I choose the latter option in this study. See also Adelaar’s (2011: 124) analysis of =*ato* in Siraya.

Crucially, the latter case presents a problem to the analysis in Table 3.7, where cliticization of temporal markers is considered a property exclusively of the AF Mstem but not of the NAF <in>Kstem-(an). Under the present analysis in Table 3.6, however, the same problem does not arise because cliticization of temporal markers is treated as an operation on top of the reduplication/affixation processes.

One the other hand, I have also demonstrated that the nominal/verbal distinction at the word level based on negation patterns (see (87)a above) applies equally to denoting constructions consisting of the AF Mstem and those made up of the NAF <in>Kstem-(an). Both forms can be negated by the verbal and nominal pattern, depending on the scope of negation. Thus, it is not well-founded to treat the NAF <in>Kstem-(an) as nominalized on the one hand, but the AF Mstem as relativized on the other. What has not been discussed is the other criterion for the nominal/verbal distinction at the word level, namely, that uncontroversial nouns, but not verbs, collocate with free possessive person forms (see (87)b above). This second criterion is addressed below.

One potential argument against the Mstem within NPs being nominalized is that it generally does not combine with free possessive person forms, which is unlike uncontroversial nouns (e.g. *na=n=ku paisu* ‘TOP.DEF=LNK=1SG.GEN money’ for “my money”) or the NAF <in>Kstem-(an) (e.g. *na=n=ku tr<in>ima-(an)* ‘TOP.DEF=LNK=1SG.ACT <NAF.RLS>buy-NMLZ for “what I bought”; S. Teng 2008: 130-131). Hence, the NAF <in>Kstem-(an) is nominalized, but the Mstem is not. However, the logic behind this is flawed because the potential to collocate with free possessive person forms is a sufficient condition for something being nominal rather than a necessary one. The sufficient condition can be stated as: if, say, the phrase *nanku X* is

acceptable, then X is nominal. On the other hand, the necessary condition stipulates this instead: X cannot be nominal unless the phrase *nanku* X is acceptable. Crucially, the necessary condition is obviously not true because there are uncontroversial nouns in Nanwang Puyuma that are prohibited from co-occurring with free possessive person forms, such as personal names (e.g. **nanku senayan* intended for ‘my Senayan’; *ibid.*: 52), some orientational terms (e.g. **nanku nguayan* intended for ‘my front’; *ibid.*: 53), and some kinship terms (e.g. **nanku taina* intended for ‘my mother’; *ibid.*: 422). Therefore, it is unjustifiable to argue against the nominality of the Mstem solely because of its inability to collocate with free possessive person forms, and this particular piece of negative evidence appears even much weaker in light of the robust positive evidence for the Mstem being treated like the *<in>Kstem-(an)* on the phrasal level, as has been demonstrated throughout this section. In fact, the paradigmatic relationship between the AF Mstem and the NAF *<in>Kstem-(an)* in a typical argument position is also taken by Ross (2009: 308) to be evidence for both constructions being nominalizations. Nevertheless, while the verbal and nominalization function of the Mstem are in a homophonous relationship in his analysis, the present study argues for a meronymic (i.e. part-whole) relationship between the two, with the nominalization function being the whole consisting of the verbal Mstem as its part. In other words, Actor nominalizations in Puyuma are grammatical in nature (aside from those that have been lexicalized based on semantic idiosyncrasy or/and irregular sound changes), internally consisting of the Mstem and other potential constituents that a typical verb would subcategorize (e.g. the verbal negator *adri*, its arguments, spatio-temporal adverbials, etc.) while externally demonstrating syntax parallel to typical nouns. The seemingly homophonous relationship

is simply a special case of the part resembling the whole when an Actor nominalization comprises nothing but the Mstem.

3.5.2. Nominal-verbal controversy and the Nuclear Austronesian hypothesis

As a synthesis of both the synchronic and diachronic perspective, the Nuclear Austronesian hypothesis (Ross 2009) is essentially a claim of diachronic changes based on synchronic data, so it might fall apart if synchronic data do not hold up. The external evidence for Nuclear Austronesian comes from Tsou, Rukai, and Puyuma. However, excluding AF affixes, which are considered both first- and second-generation in the Nuclear hypothesis, Tsou has only robust reflexes of first-generation affixes (see §3.2.2 and §3.3) whereas Rukai shows none because its verbal system is drastically different from all the other Formosan languages (see §3.4). Consequently, the hypothesis hinges heavily on Puyuma,¹⁰³ the only non-Nuclear language that has reflexes of both first- and second-generation affixes in all four Focus categories. Thus, Puyuma has since become the center of attention in subsequent works that challenge the hypothesis (Sagart 2010, 2013; Foley 2014) and those that defend it (S. Teng & Ross 2010; Ross 2012). In what follows, I first summarize the arguments from both the challengers and defenders regarding a nominal-verbal debate in Nanwang Puyuma (§3.5.2.1). After the review, I will add some additional data from Rikavung Puyuma (§3.5.2.2) to weigh in on this issue.

¹⁰³ In fact, Ross (2012: 1297) admitted that his “PAn [Proto-Austronesian] reconstruction... relies heavily on a comparison of Puyuma and PNAn [Proto-Nuclear Austronesian]”, although he did claim that “Puyuma is the only Austronesian language that closely reflects the PAn system of verbal morphology.”

3.5.2.1. Nanwang Puyuma

One of the claims in the Nuclear Austronesian hypothesis is that second-generation NAF-words in Puyuma are always nominal because it is a non-Nuclear language, which is exempt from the nominalization-into-verb innovation involving second-generation affixes that were once nominalizers. One point of contention in the debate of the four aforementioned studies concerns whether Puyuma word forms in the shape of *<in>STEM*,¹⁰⁴ where *<in>* is a second-generation affix in the Nuclear hypothesis, are always nominal as claimed. If they can be verbal, then Puyuma would be just like all the Nuclear languages where the *<in>STEM* is both nominal and verbal, which would falsify the Nuclear hypothesis, at least regarding its claim in Puyuma. For ease of reference, the former is called the absolute nominal view whereas the latter the amphibious (between nominal and verbal) view.

According to S. Teng (2008: 130) “[t]he morpheme *<in>* (or its allomorph *ni-*) is infixes or prefixed to verbs to indicate perfective aspect in many Formosan languages. However, in Puyuma, although *<in>* still retains its function of marking perfective aspect, words formed with *<in>* are nominal.” She then showed three properties to demonstrate the nominal nature of the *<in>STEM* in Nanwang Puyuma, as summarized in (102).

¹⁰⁴ I use the schematic form *<in>Kstem-(an)* in §3.5.1, but *<in>STEM* instead in this section. This is because §3.5.1 focuses comparisons of Actor nominalizations with non-Actor ones, which specifically involve the Kstem. This section, however, deals with functions of *<in>* in general, regardless of whether it has to do with the Kstem or nominalization.

(102) Nominal properties of the <in>STEM in Nanwang Puyuma (S. Teng 2008: 130-131)

- a. A verb attracts enclitic Topic indexes, which are obligatory when SAPs are indexed, whereas the <in>STEM cannot do so.
- b. A verb never collocates with free possessive person forms whereas both the <in>STEM and underived nouns have this option.
- c. A verb is negated by *adri* whereas both the <in>STEM and underived nouns are negated by *amelri*.

Regarding the first property, Sagart (2010) found some counterexamples and argued for the amphibious view, so S. Teng & Ross (2010) admitted that (102)a is too general a statement but maintained counterexamples are very rare and reaffirmed the absolute nominal view. One controversial example is illustrated in (103), with two versions of its free translation taken from the aforementioned studies.

(103) Nanwang Puyuma¹⁰⁵

tu=**p<in>auka**=ku kana ragan m-uka i balangaw
 3.ACT=<NAF.RLS>send=1SG.TOP ACT.DEF priest AF-go LOC Taitung
 ‘the priest’s sending me to Taitung’ (S. Teng & Ross 2010: 551)
 ‘The priest sent me to Taitung.’ (Sagart 2010: 197; 2013: 485)

For S. Teng & Ross (2010), Example (103) is a “noun phrase” headed by *p<in>auka*. They supported the absolute nominal view by the second property in (102), as shown in (104)a, and held that (103) is only a nominalized version of the sentence in (104)b, where the indicative verb *pauka*-[y]aw can never collocate with free possessive person forms.

¹⁰⁵ All of the three cited works transcribed the word for “Taitung” and that for “child” (not shown in this example) using the same grapheme <l>, but the laterals in the two words are contrastive, one alveolar and the other retroflex. Thus, either instance of <l> must be a typo. The transcriptions here follow the conventional Nanwang orthography (see Footnote 29), according to which “Taitung” is *balangaw* /baɭaŋaw/ and “child” *walrak* /walak/. Moreover, while S. Teng & Ross (2010) glossed <in> with NMLZ, Sagart (2010, 2013) did not. The gloss NMLZ is not indicated here since the nominalization function is precisely what is at issue. Finally, *pauka* ‘send’ can be further analyzed into *pa-uka* ‘CAUS-go’, but this is not shown here since causatives are trivial to the current controversy.

(104) Nanwang Puyuma (S. Teng & Ross 2010: 551)¹⁰⁶

- a. *na=n=tu* ***p<in>auka=ku*** *kana* *ragan* *m-uka*
 TOP.DEF=LNK=3.ACT <NAF.RLS>send=1SG.TOP ACT.DEF priest AF-go
 i *balangaw*
 LOC Taitung
 ‘the priest’s sending me to Taitung’ (Original translation from the cited reference)
- b. *tu=pauka-[y]aw=ku=lra* *kana* *ragan* *m-uka* *i* *balangaw*
 3.ACT=send-PF=1SG.TOP=already ACT.DEF priest AF-go LOC Taitung
 ‘The priest sent me to Taitung.’

For Sagart (2010, 2013), however, (103) is a “complete sentence” with *p<in>auka* being the indicative verb. Also, Sagart (2013) dismissed (104)a as spurious evidence since what is really at issue is (103), not (104)a, so he questioned the validity to treat the two examples as illustrating the same construction.¹⁰⁷ He then supported his analysis by the question-answer pair in (105).

¹⁰⁶ In fact, even if the nominal use of *p<in>auka* ‘<NAF.RLS>send’ is clear in (104)a due to its collocation with a free possessive person form, it is dubious that (104)a as a whole is a nominal phrase without illocutionary force meaning ‘the priest’s sending me to Taitung’, as translated in the cited work. Judging from the presence of the 1SG Topic index =*ku*, (104)a is better analyzed as an assertive nominal predicate construction meaning ‘I am the one the priest sent to Taitung’, comparable to (i).

(i) Nanwang Puyuma (S. Teng 2008: 191)
a *tipulr=ku*
 TOP.INDF T.=1SG.TOP
 ‘I am a Tipulr.’ (I am from Tipulr.)

This also explains the assertive force in (105)c. Another reason for this analysis is that the patientive role within a true nominal phrase is not coded by the Topic index, but by free person forms for non-Actor and non-Topic arguments, such as *kantaw* in (ii).

(ii) Nanwang Puyuma (L. Huang 2000b: 183)
pa-ka-nquayan *dra* [*ku=ni-na'u-an*] ***kantaw*** *i,* *m-u<a>arak=ku*
 CAUS-K-front OBL.INDF 1SG.ACT=NAF.RLS-see-NMLZ 3.UND PTOP AF-<IPFV>dance=1SG.TOP
 ‘Before (my) seeing {him/her/them}, I was dancing.’

¹⁰⁷ The crucial difference between (103) and (104)a is whether the predicate *p<in>auka* ‘<NAF.RLS>send’ is preceded by a nominal relation marker such as *na*. With such a marker, there is good evidence for *p<in>auka* in (104)a being a nominal predicate (parallel to underived nouns), but without it, the evidence for *p<in>auka* in (103) being a nominal predicate becomes weaker.

(105) Nanwang Puyuma (Sagart 2013: 485)¹⁰⁸

- a. **tu=pauka-[y]aw=yu** kana ragan m-ukua i balangaw
 3.ACT=send-PF=2SG.TOP ACT.DEF priest AF-go LOC Taitung
 ‘Does the priest send you to Taitung?’
- b. **tu=p<in>auka=yu** kana ragan m-ukua i balangaw
 3.ACT=<NAF.RLS>send=2SG.TOP ACT.DEF priest AF-go LOC Taitung
 ‘Did the priest send you to Taitung?’ (Original translation from the cited reference)
- c. **aywa, tu=p<in>auka=ku** kana ragan
 yes 3.ACT=<NAF.RLS>send=1SG.TOP ACT.DEF priest
 m-ukua i balangaw
 AF-go LOC Taitung
 ‘Yes, the priest sent me to Taitung.’ (Original translation from the cited reference)

According to his consultant, (105)c is an appropriate answer to both (105)a and (105)b. If (105)a is a complete sentence, which both sides would agree, then so should be (105)b, which is the perfective (past) rather than nominalized counterpart of (105)a in the amphibious view. Although it is indisputable that the default realis-indicative PF verb is *Kstem-aw* in modern Puyuma (as in (105)a above), the question at issue is whether Puyuma has some residual instances of the *<in>STEM* used as indicative verbs. The answer is negative for S. Teng & Ross (2010) but positive for Sagart (2010, 2013).

Perhaps stronger evidence against S. Teng & Ross’s (2010: 551) absolute nominal analysis of the *<in>STEM* comes from the fact that there are clear cases of the *<in>STEM* in AF- rather than NAF-constructions and that the Topic index is associated with the agentive rather than patientive role, unlike those in (103) through (105). Examples in (106) are illustrative of the *<in>STEM* in AF-constructions.

¹⁰⁸ Sagart’s consultant used *m-uka* and *m-ukua* interchangeably, as in (104) and (105).

(106) Nanwang Puyuma (Sagart 2013: 487)¹⁰⁹

- a. **in-umu=ku** dra eraw
 PFV-AF.hold.in.mouth=1SG.TOP UND.INDF wine
 ‘I tasted some wine.’
- b. **k<in>i-lrengaw=ku** dra piano
 <PFV>AF.get-sound=1SG.TOP UND.INDF piano
 ‘I heard a piano.’

To maintain that the <in>STEM in (106) is the nominal predicate denoting a patientive argument, as the absolute nominal view would do to the <in>STEM (103), would run against the semantics. Among languages that have it the infix <in> most commonly occurs in NAF-constructions rather than AF-constructions, but AF-words with perfective <in> like those in (106) are not an anomaly in Puyuma since they are also found in many Formosan languages (including at least Thao, Pazeh, Saisiyat, Siraya, Kanakanavu, Saaroa, Mayrinax Atayal, Seediq, and Isbukun Bunun; see Ross 2009: 317-320 and the references therein).

As for the third property in (102), Foley (2014) points out a counterexample from S. Teng (2008), and questions the absolute nominal view. The negator for uncontroversial nouns and verbs in Puyuma is respectively *amelri* and *adri*, both of which have been illustrated in (88) above. However, the <in>STEM can be negated not only by the nominal negator *amelri*, which is in accord with the absolute nominal view, but also by the verbal negator *adri*, which goes against such a view, which has been illustrated in (92) above. Thus, Foley (2014: 38) concludes that the word form *b<in>arekep-an* ‘<NAF.RLS>assemble-NMLZ’ in (92)b “is synchronically a verb in Puyuma, whatever its

¹⁰⁹ The word-initial glottal stop in *eraw* was transcribed in the cited work but is dropped here because it is not phonemic here and that there are variations across speakers (see S. Teng 2008). The form *ki-lrengaw* in the second example can potentially be the Mstem or Kstem because it belongs to Class VI verbs (see Table 3.4). It is judged to be the Mstem here based on the argument realization pattern parallel to that in a typical AF-construction in Puyuma.

earlier status.” The same word form in the same example was analyzed by S. Teng (2008: 143) as an instance of “gerundive nominals”, which even she judged to be “verbal constructions” due to their collocations with the verbal negator. Another property of gerundive nominals she pointed out is that they “can be followed by an intransitive verb [i.e. the Mstem] to form a serial verb construction” (ibid.). This is precisely the case in (105), where both the form with a first-generation affix (i.e. *pauka*-[y]aw ‘send-PF’) and the one with a second-generation affix (i.e. *p<in>auka* ‘<NAF.RLS>send’) are followed by the Mstem *m-ukua* ‘AF-go’. Thus, even by her criteria, the controversial form <in>STEM does share some verbal properties with the indisputable PF verbal form *Kstem-aw* to the exclusion of underived nouns.¹¹⁰

3.5.2.2. Rikavung Puyuma

In addition to the arguments made by Sagart (2010, 2013) and Foley (2014), additional data from Rikavung Puyuma, a non-Nanwang variety, suggest that the absolute nominal view is too strong. There are at least four points that can be made here.

First, Rikavung helps to shed light on the controversial Nanwang example in (103) thanks to one crucial grammatical property almost not found in Nanwang. Specifically, uncontroversial verbs in Nanwang NAF-constructions (i.e. first-generation

¹¹⁰ In fact, the <in>STEM is not the only form with a second-generation affix that has been shown to possess both verbal and nominal properties in Puyuma. In Katripul Puyuma, for instance, the form *i*-Kstem, which contains the second-generation CF *i*- (< PAn *Si-), is analyzed by S. Teng (2012) as both nominal and verbal. They are nominal because they collocate with free possessive person forms like nouns do, and verbal because (i) they can be negated by the verbal negator, (ii) they acquire repetitive or intensified meanings (depending on stems) when undergoing CVCV- reduplication, which gives rise to plurality instead on nouns, and finally (iii) there are more restrictions on their occurrences in typical argument positions than on typical nouns. Therefore, like the <in>STEM in Nanwang Puyuma, the CF-word *i*-Kstem in Katripul Puyuma can be as verbal as first-generation Focus-words and at the same time has external nominal distributions comparable to those of underived nouns, thus speaking against the absolute nominal view of second-generation affixes (i.e. NAF <in> in Nanwang and CF *i*- in Katripul).

NAF-words) are sandwiched by person-form indexes for the (non-Topic) Actor on the left and those for the Topic on the right, or schematically ACT=V.NAF=TOP (called the sandwiching pattern). On the other hand, in equational/identificational constructions Topic indexes for the sole argument occur on the right of uncontroversial noun phrases acting as the predicate, or schematically NP=TOP. The gist of the debate regarding the *<in>*STEM (with a second-generation affix) in (103) is whether this form should be analyzed on a par with the verbal Kstem-*aw* in (104)b or with the nominal possessive NP in (104)a. Forms alone are indecisive in this matter due to the complete isomorphism between (non-Topic) Actor indexes on uncontroversial verbs and possessor indexes on uncontroversial nouns in Nanwang (see S. Teng 2015). In Rikavung, however, both the Topic and (non-Topic) Actor indexes for SAPs cluster on the left of uncontroversial verbs of NAF-constructions (i.e. first-generation NAF-words), with the first person always preceding the second person irrespective of their grammatical relations (called the clustering pattern; see (37) on p.81).¹¹¹ In addition, an inverse marker is obligatorily used to determine who is acting on whom when the patientive Topic is at least as person-prominent as the agentive (non-Topic) Actor. Person-prominence is ranked along the hierarchy of the first, second, and then third person from high to low. This design leads to two scenarios where the inverse marker is required: (i) when the second person acts upon the first person, and (ii) when the third person, the lowest-ranked one, assumes the agentive role (see H. Jiang & Billings 2015 for details). On the other hand, as in Nanwang, Topic indexes for the sole argument in Rikavung equational/identificational constructions occur on the right of nominal predicates consisting of uncontroversial

¹¹¹ Based on the ongoing collaboration between me and Loren Billings, this is the predominant pattern in not only Rikavung but also Ulivelivek and Tamalakaw Puyuma. It just so happens that we did more work on Rikavung, so it is demonstrated here.

nouns. As a result, the Topic index of equational/identificational constructions in Rikavung is never confused with that of NAF-constructions. Importantly, it is found that the second-generation <in>STEM can participate in the same NAF-construction as the first-generation Kstem-aw, where the Topic index occurs on different sides of the predicate as compared with that in equational/identificational constructions. This is contrasted in (107).

(107) Rikavung Puyuma (Fieldnotes)

- a. **a=sinsi=ku**
TOP.INDF=teacher=1SG.TOP
'I am a teacher.'
- b. **taw=ku=pa-ua-[y]aw** ni=misak
INV=1SG.TOP=CAUS-go-PF ACT.SG=M.
'Misak made me go.'
- c. **taw=ku=p<in>a-ua** ni=misak
INV=1SG.TOP=<PF.RLS>CAUS-go ACT.SG=M.
'Misak (previously) made me go.'

The clitic clustering pattern in Rikavung is generally considered to be more conservative than the sandwiching pattern in Nanwang (see Ross 2015a), so the example in (107)c might as well reflect an earlier function of <in> as the realis-perfective PF marker on matrix verbs, as is the norm in numerous Nuclear languages (e.g. Tagalog and Paiwan, demonstrated in §2.3.1 above).

So far it has been demonstrated that in both Nanwang and non-Nanwang dialects of Puyuma there are at least some constructions where word forms with second-generation affixes cannot be maintained to be absolutely nominal. But what about the flip side of the claim in the Nuclear hypothesis, namely, that Puyuma word forms with first-generation affixes are always verbal? This then begs the question of what criteria were taken to be nominal or verbal in the hypothesis. Ross (2009) considered *M- (his abstract

label for **<um>*, **ma-* and no affix, all serving the AF function) to be both first- and second-generation affix in PAn because *M-STEM “served both as a realis verbal form and a realis nominalisation...[O]nly the syntactic context determined whether a form in *M- was being used in a noun phrase or a verb phrase. This situation continues in Puyuma.” (ibid.: 307). Then he illustrated his point with (108) from Nanwang Puyuma.

(108) Nanwang Puyuma (Ross 2009: 308)

- a. **sa~senay** i walregan
 <AF>IPFV~sing TOP.SG W.
 ‘Walregan {is/was} singing.’ (citing S. Teng 2008)
- b. amau kuiku na **sa~senay**
 COP 1SG.TOP TOP.DEF <AF>IPFV~sing
 ‘The one who was singing is me.’ (citing Ross & S. Teng 2005b)

The AF-word *sa~senay* ‘<AF>IPFV~sing’, where ** is a reflex of PAn **<um>*, is verbal because it serves as the matrix predicate, as in the first example, and at the same time nominal because it is preceded by the nominal relation marker *na* and functions as the Topic argument, as in the second example. Thus, only the marker *na* “tells us that *na sa~senay* is a nominal.” (ibid.: 308). Therefore, the crucial criterion is whether a form can be marked by nominal relation markers, hence being in a paradigmatic relationship with underived nouns. And it is by this very criterion that NAF-words with second-generation affixes were judged to be always nominal and those with first-generation affixes always verbal. The first part of the claim has been questioned above, and now I move on to show that even the second part of the claim is not free from problems.

Second, it is true that NAF-words with first-generation affixes like PF *-aw*, LF *-ay*, and CF *-anay* (called the *a*-grade series by Ross 2009) are generally not preceded by nominal relation markers (hence verbal), unlike their counterparts with second-generation

affixes (hence nominal). However, there is at least one exception to this general rule, one that involves exclamations. In Rikavung Puyuma, one major function of the marker *kana* is to introduce non-Actor and non-Topic arguments consisting of personal associative plural nouns or common definite ones (see Table 2.2 for the complete nominal marking system), as in (109).

(109) Rikavung Puyuma (Fieldnotes)

- b. m-na'u~na'u=ku **kana=misak**
 AF-IPFV~see=1SG.TOP UND.ASSO.PL=M.
 'I keep looking at Misak and her associates.'
- b. m-na'u~na'u=ku **kana=alrak**
 AF-IPFV~see=1SG.TOP UND.DEF=child
 'I keep looking at the {child/children}.'

Interestingly, the marker is also used to introduce exclamative remarks, as in (110) through (112), where an unaffixed AF-word, an affixed AF-word, and a first-generation NAF-word are illustrated respectively. Corresponding indicative sentences are presented side by side for comparative purposes.

(110) Rikavung Puyuma (Fieldnotes)

- a. tihasar i=masaw
 AF.tall TOP.SG=M.
 'Masaw is tall.'
- b. wa, **kana=tihasar** i=masaw
 INTJ EXCM=AF.tall TOP.SG=M.
 'How tall Masaw is!'

(111) Rikavung Puyuma (Fieldnotes)¹¹²

- a. *m-atras* *i=masaw*
 AF-arrogant TOP.SG=M.
 ‘Masaw is arrogant.’
- b. *wa*, ***kana=m-atras*** *i=masaw*
 INTJ EXCM=AF-arrogant TOP.SG=M.
 ‘How arrogant Masaw is!’

(112) Rikavung Puyuma (Fieldnotes)

- a. *taw=kuang-aw* *na=vavuy* *ni=masaw*
 INV=shoot.with.a.gun-PF TOP.DEF=boar ACT.SG=M.
 ‘Masaw shot the boar with a gun.’
- b. *wa*, ***kana=taw=kuang-aw*** *na=vavuy* *ni=masaw*
 INTJ EXCM=INV=shoot.with.a.gun-PF TOP.DEF=boar ACT.SG=M.
 ‘How (well) Masaw shot the boar with a gun!’

Aside from the interjection, adding *kana* to an indicative sentence essentially turns a statement into an exclamation.¹¹³ The argument-introducing and exclamatory function of *kana* is most likely not a coincidence since similar situations are also attested in many Austronesian languages, where the exclamative function has been argued by Kaufmann (2011) to derive from depriving a predication of its assertive force by marking it in the same way as a typical nominal, thus packaging it as a presupposition instead of an assertion. The exclamative constructions in (110) through (112) resemble what Evans (2007: 367) terms *insubordination*, that is, “the conventionalized main clause use of what, on prima facie grounds, appear to be formally subordinate clauses.” Like Puyuma exclamative constructions, many insubordinated constructions he investigated are high in presuppositionality. Crucially, by Ross’s (2009) criterion the PF-word with the first-

¹¹² While the affixed *m-atras* means “arrogant”, its unaffixed counterpart *atras* means “high (of buildings)” (cf. *haughty* in English).

¹¹³ Caution has to be exercised here that not every indicative can be readily changed into an exclamative in this way. A similar exclamative use of the marker *kana* is also found in Tsuchida’s (1980: 300) spontaneous texts of Tamalakaw Puyuma.

generation PF *-aw* in (112)b would be nominal because of its collocation with the nominal marker *kana*.

Third, one might discard exclamation marks as “peripheral” to the grammar, but there is yet another construction where denoting expressions are created out of first-generation NAF-words in the same way as second-generation ones. Specifically, the construction involves NAF-words with first-generation affixes like PF/LF *-i* and CF *-an* (called the zero-grade series by Ross 2009), to be demonstrated below.

Focus-words serving as matrix predicates in affirmative and negative sentences are shown in (113) and (114) respectively.¹¹⁴

(113) Rikavung Puyuma (Fieldnotes)

- a. {**kezeng/tengez/sukun**}=ku kani=misak
 {<AF>pull/<AF>beat/<AF>push}=1SG.TOP UND.SG=M.
 ‘I’m {pulling/beating/pushing} Misak.’
- b. ku=**kezeng-aw** i=misak
 1SG.ACT=pull-PF TOP.SG=M.
 ‘I pulled Misak.’
- c. ku=**tengez-ay** i=misak
 1SG.ACT=beat-LF TOP.SG=M.
 ‘I beat Misak.’
- d. ku=**sukun-anay** i=misak
 1SG.ACT=push-CF TOP.SG=M.
 ‘I pushed Misak.’

¹¹⁴ Recall from §2.4.1 that unlike AF affixes, NAF affixes in Puyuma are rather stem-specific. Thus, mixing the three stems with the three NAF affixes would create unacceptable sentences if nothing else is changed. Also, under negation PF is permutated to LF, a phenomenon also found in Atayalic languages.

(114) Rikavung Puyuma (Fieldnotes)

- a. 'azi=ku **{kezeng/tengez/sukun}** kani=misak
 NEG=1SG.TOP {<AF>pull/<AF>beat/<AF>push} UND.SG=M.
 'I'm not {pulling/beating/pushing} Misak.'
- b. 'azi=ku **kezeng-i** i=misak
 NEG=1SG.ACT pull-LF.DEP TOP.SG=M.
 'I didn't pull Misak.'
- c. 'azi=ku **tengez-i** i=misak
 NEG=1SG.ACT beat-LF.DEP TOP.SG=M.
 'I didn't beat Misak.'
- d. 'azi=ku **sukun-an** i=misak
 NEG=1SG.ACT push-CF.DEP TOP.SG=M.
 'I didn't push Misak.'

As far as NAF-words are concerned, those in (113) and (114) all involve first-generation affixes (respectively the *a*-grade and zero-grade series), as opposed to those with second-generation affixes, which are embeddable under nominal relation markers in typical argument positions as in (115).

(115) Rikavung Puyuma (Fieldnotes)

- a. mau inku [na={kezeng/tengez/sukun}
 COP 1SG.TOP TOP.DEF={<AF>pull/<AF>beat/<AF>push}
 kani=misak]
 UND.SG=M.
 'The one who is {pulling/beating/pushing} Misak is me.'
- b. mau i=misak [na=n=ku **k<in>ezeng]**
 COP TOP.SG=M. TOP.DEF=LNK=1SG.ACT <PF.RLS>pull
 'The one I pulled is Misak.'
- c. mau i=misak [na=n=ku **t<in>engez-an]**
 COP TOP.SG=M. TOP.DEF=LNK=1SG.ACT <RLS>beat-LF
 'The one I beat is Misak.'
- d. mau i=misak [na=n=ku **i-sukun]**
 COP TOP.SG=M. TOP.DEF=LNK=1SG.ACT CF.RLS-push
 'The one I pushed is Misak.'

Due to the complementary distributions of matrix predicates with first-generation affixes and arguments with second-generation ones, Ross (2009: 317) analyzed Focus-words in (113) and (114) as verbal (respectively called Realis and Negative verbal forms) and

those in (115) as nominal (called Realis nominal forms). As a result, since the AF Mstem with AF affixes functions as both the predicate and argument function, AF affixes have to be both first- and second-generation in the Nuclear hypothesis (see also (108) from Nanwang Puyuma).

However, it is often ignored that negated first-generation Focus-words in (114) can also have external nominal distributions and occur in the same argument positions as second-generation ones in (115). This is demonstrated in (116).

(116) Rikavung Puyuma (Fieldnotes)

- a. mau~mau inku [na='azi {**kezeng/tengez/**
 EMPH~COP 1SG.TOP TOP.DEF=NEG {<AF>pull/<AF>beat/
sukun} kani=misak]
 <AF>push} UND.SG=M.
 'The (only) one who is not {pulling/beating/pushing} Misak is me.'
- b. mau~mau i=misak [na='azi=ku **kezeng-i]**
 EMPH~COP TOP.SG=M. TOP.DEF=NEG=1SG.ACT pull-LF.DEP
 'The (only) one I didn't pull is Misak.'
- c. mau~mau i=misak [na='azi=ku **tengez-i]**
 EMPH~COP TOP.SG=M. TOP.DEF=NEG=1SG.ACT beat-LF.DEP
 'The (only) one I didn't beat is Misak.'
- d. mau~mau i=misak [na='azi=ku **sukun-an]**
 EMPH~COP TOP.SG=M. TOP.DEF=NEG=1SG.ACT push-CF.DEP
 'The (only) one I didn't push is Misak.'

By the very same criterion the argument phrases with second-generation Focus-words in (115) are nominalizations, those with first-generation zero-grade ones in (116) should be as well because they all denote the Topic NP of whatever Focus-words involved. While it is tempting to attribute the nominality of the argument phrases in (115) solely to Focus-words, the same solution does not work well with those in (116), where Focus-words are negated by the verbal negator 'azi (as opposed to the nominal one *melri*; cf. (87)a in Nanwang). Negated Focus-words in Rikavung not only structurally resemble Focus-words in Tsou (negated or not; see §3.2.2), where person-form indexes attach to

preverbal auxiliaries, but also involve the same PAn etyma as Focus-words in Tsou, that is, PAn **<um>* for AF-words and the first-generation zero-grade series (PAn **-a/i/ani*) for NAF-words. In both languages, Focus-words with these cognate affixes can be embedded under nominal relation markers to form grammatical argument nominalizations, where no single word form is solely responsible for the nominality of the whole construction.

Setting aside AF-words, which contain AF affixes considered both first- and second-generation in the Nuclear hypothesis, it is clear that the nominality of argument phrases with verbally negated NAF-words containing first-generation zero-grade affixes, as in (116), does not come from NAF-words per se, but from the construction they are part of. What is less clear, however, is where the nominality of argument phrases with second-generation NAF-words, as in (115), should be sought. Does it reside in second-generation NAF-words lexically or in the constructions they are part of? In what follows, I argue for the latter.

One could potentially argue that second-generation NAF-words are lexically nominal because they, unlike negated NAF-words, collocate with free possessive person forms as underived nouns do (cf. (115)b-d and *na=n=ku turus* ‘TOP.DEF=LNK=1SG.GEN sibling’ for “my sibling”).¹¹⁵ However, such a collocational discrepancy can be given a positional account independent of the lexical status of a given Focus-word. It has to do with the composition of so-called free possessive person forms, which are prosodic words consisting of a nominal relation marker, a person-form clitic, and a homorganic nasal

¹¹⁵ For instance, S. Teng & Ross (2010: 551) took this collocation to be “a crucial test of whether the following head is nominal or verbal.”

sound linking the two (see Footnote 93). When only one of the first two elements is present, the nasal linking sound does not arise, as shown in (117).

(117) Rikavung Puyuma (Fieldnotes)

- a. mau i=misak **na=k<in>ezeng** ni=masaw
 COP TOP.SG=M. TOP.DEF=<PF.RLS>pull ACT=M.
 ‘The one Masaw pulled is Misak.’ (cf. (115)b)
- a. i=ini na=valray mu, **ku=tr<in>ima’-an**
 TOP.SG=PROX LNK=book PTOP 1SG.ACT=<RLS>buy-LF
 ‘As for this book, (it) is what I bought.’ (cf. (115)c)

In other words, for there to be a free possessive form in the first place, a nominal relation marker and a person-form clitic have to be contiguous so that the linking nasal would appear.¹¹⁶ The contiguity requirement is precisely what cannot be met in (116), where the negator *'azi*, itself being prosodically independent, intervenes between a nominal relation marker and a person-form clitic, followed by the negated NAF-words. As far as the argument function is concerned, negated first-generation NAF-words collocate with both nominal relation markers and Actor-indexing clitics just as second-generation NAF-words do. Thus, it is reasonable to infer that the restriction against negated Focus-words collocating with free possessive person forms is only secondary to the fact that negated Focus-words are preceded by the negator, which obligatorily attracts person-form clitics (as is the norm in other Formosan languages), thus preventing person-form clitics from combining with nominal relation markers into prosodically independent person forms.¹¹⁷

¹¹⁶ It is worthy of note that the linking nasal can be optional when the indefinite nominal relation marker *za* combines with person-form clitics, as in the following example:

Rikavung Puyuma (Fieldnotes)
me-na'u {*za=ku=suan/za=n=ku* *suan*} *i=misak*
 AF-see {INDF.UND=1SG.GEN=dog/INDF.UND=LNK=1SG.GEN} dog TOP.SG=M.
 ‘Misak saw a dog of mine.’

¹¹⁷ Variations in person-form clitics abound in Rikavung Puyuma. However, the variations are conditioned not by semantic roles, syntactic functions, or morphological cases, but by positioning with respect to their

On the other hand, within the argument phrase of which second-generation NAF-words are a constituent, they share properties with first-generation NAF-words or verbally negated AF-words to the exclusion of underived nouns. One such property has been shown in (107)b-c above, where NAF-words, of both the first- or second-generation type, serve as the matrix predicate procliticized by the inverse marker and a Topic index for SAPs. Moreover, second-generation NAF-words can be followed by the AF Mstem so as to complement the predication, a property they share with first-generation NAF-words but not with underived nouns, as shown in (118).

(118) Rikavung Puyuma (Fieldnotes)

- a. **na=vu'ir** **mu,** **ku=i-kerutr** **salem**
 TOP.DEF=taro PTOP 1SG.ACT=CF.RLS-dig <AF>plant
 ‘As for the taros, (they) are what I planted by digging (holes on the ground).’
- b. **na=vu'ir** **mu,** **ku=kerutr-anay** **salem**
 TOP.DEF=taro PTOP 1SG.ACT=dig-CF <AF>plant
 ‘As for the taros, I planted them by digging (holes on the ground).’

This shows that although they can have external nominal distributions, second-generation NAF-words (i.e. *i-kerut* in (118)a) also have the subcategorizing properties of highly verbal forms like first-generation NAF-words (i.e. *kerut-anay* in (118)b). Finally, second-generation NAF-words mark their patientive arguments in the same way as verbally negated AF-words, as in (119).

(119) Rikavung Puyuma (Fieldnotes)

- a. **pakirev** **ku=i-sukun** {kannu/kani=misak}
 AF.forcefull. 1SG.ACT=CF-push {2SG.UND/UND.SG=M.}
 ‘(The way) I pushed {you/Misak} is strong. (i.e. I pushed {you/Misak} hard.)’
- b. **'azi=ku** **sahar** {kannu/kani=misak}
 NEG=1SG.TOP AF.like {2SG.UND/UND.SG=M.}
 ‘I don’t like {you/Misak}.’

Therefore, argument phrases with second-generation NAF-words are externally nominal because they can be marked by nominal relation markers to denote the Topic argument, and at the same time they are internally as verbal as other uncontroversial verbal forms in the language. By these criteria, argument phrases with second-generation NAF-words are grammatical nominalizations just like those with first-generation zero-grade NAF-words.

Table 3.8 below summarizes the syntagmatic potentials of all the NAF-words in (113) through (117), and every acceptable combination is cross-referenced with examples. It shows that all NAF-words, be it first- or second-generation, collocate with Actor-indexing clitics and serve as matrix predicates (Cell A through C). However, only second-generation and first-generation zero-grade NAF-words (Cell E and F), but not first-generation *a*-grade NAF-words (Cell D), can be additionally embedded under nominal relation markers to denote the Topic NP.¹¹⁸ Therefore, the potential to be embedded under nominal relation markers is a crucial property that sets second-generation and first-generation zero-grade NAF-words apart from first-generation *a*-grade ones. More importantly, the embedding criterion is sufficient to characterize the syntactic differences among NAF-words without having to rely on the nominal-verbal contrast between first- and second-generation NAF-words under the Nuclear hypothesis. A question worth asking then is why only first-generation *a*-grade NAF-words are restricted from embedding for the purpose of argument nominalizations. This topic is taken up later in §4.4.1, which explores *a*-grade NAF-words in multiple Formosan languages.

¹¹⁸ However, *a*-grade first-generation NAF-words are marked by the nominal relation marker *kana* to express exclamations, as has been shown in (112) above.

Table 3.8: Syntagmatic potentials of NAF-words in Rikavung Puyuma

Polarity	Forms of Focus-words	With Actor-indexes only	With both Actor-indexes and NRMs
AFF	PF: Kstem- <i>aw</i> LF: Kstem- <i>ay</i> CF: Kstem- <i>anay</i> [1st-generation <i>a</i> -grade]	A: YES (113)b-d, (107)b, (118)b	D: NO
	PF: < <i>in</i> >Kstem LF: < <i>in</i> >Kstem- <i>an</i> CF: <i>i</i> -Kstem [2nd-generation]	B: YES (107)c and (118)a	E: YES with the linking nasal (115)b-d
NEG	PF/LF: Kstem- <i>i</i> CF: Kstem- <i>an</i> ¹¹⁹ [1st-generation zero-grade]	C: YES (114)b-d	F: YES without the linking nasal (116)b-d

Note: NRM stands for nominal relation markers.

¹¹⁹ Given the *a*-grade CF suffix *-anay*, its expected zero-grade counterpart is *-ani* (via deletion of the /a/ vowel before the offglide; Ross 2009). In Rikavung Puyuma, the zero-grade CF suffix turns out to be *-an* (as in Kstem-*an*), which is then homophonous with the second-generation LF suffix *-an* (as in <*in*>Kstem-*an*). However, in Katripul Puyuma, both *-ani* and *-an* are attested for the zero-grade CF suffix. Compare the affirmative CF form with its negative counterpart:

Katripul Puyuma (Fieldnotes; provided by Lisem Kadadepan, born in 1941)

- (i) *ku=mu=seksek-anay*
1SG.ACT=2PL.TOP=push-CF
'I pushed you.'
- (ii) *'azi=ku=mu* *seksek-an(i)*
NEG=1SG.ACT=2PL.TOP push-CF.DEP
'I did not push you.'

3.6. Chapter summary

This chapter has investigated argument-denoting expressions comprised of Focus-words in several Formosan languages, including Tsou, Rukai, and Puyuma in particular. They all share the function of denoting the Topic NP of a Focus-word, on which easily identifiable cognate Focus affixes are marked to restrict the grammaticalized semantic roles of that NP. Conservative languages grammatically distinguish up to four such roles, conveniently labeled as Actor, Patient, Locative, and Conveyance. Some of such constructions have been considered *prima facie* instances of nominalization while others do not seem to be worthy of being studied under the same rubric or are deemed headless relativization at best. However, it has been argued above that such a treatment is unfair because it hinges upon whether Focus-words, a morpho-lexical class, are analyzed as lexically nominal or verbal based on language-specific criteria, which are often not comparable across languages. Moreover, it has been shown that by the criteria in specific languages Focus-words are as verbal as they can be in terms of their syntagmatic subcategorizations (except for the restrictions on polarity; e.g. Table 3.8) within the Focus-constructions they are part of, and that the nominality of argument nominalizations containing Focus-words is better attributed to the entire constructions rather than to Focus-words alone. Focus-constructions of this nature are then termed grammatical nominalizations, whose syntactic functions include at least taking up a complete NP or modifying another nominal, be it a lexical noun or yet another grammatical nominalization.

Now suppose PAn resembled conservative Philippine-type languages, its Focus nominalizations would have been exocentric as well, with all Focus-words being

component parts of the composite nominalization structure. If so, this conclusion would lead to an understanding of the history of Focus-words that differs from the account in the SPR or Nuclear Austronesian hypothesis but aligns well with the alternative hypothesis proposed by Ross (2002) (or even earlier, Ross 1995). For that alternative hypothesis to work, some issues have to be addressed so as to account for the transition from PAn to the variations found in modern Formosan languages, which is the task in the next chapter.

|Chapter 4

Verbal-based Nominalization II: Alternative History of Focus Affixes

For the purpose of discussions in this chapter, I stop referring to Focus affixes as the first- or second-generation ones, which assumes among them different diachronic trajectories into the verbal paradigm of proto-languages (either PAn or the proposed PNA). Instead, I rely on their structural features, and refer to Focus-words as either the Mstem (containing reflexes of PAn **<um>* or **ma-*) or NAF-words, with the latter containing Suffixal or Mixed Focus affixes, corresponding to Ross's (2009) first- and second-generation affixes. Suffixal NAF affixes come in two series, referred to as *a*-grade (PAn **-aw/ay/anay*) and zero-grade (PAn **-u/a/i/ani*) affixes, following Ross (2009).

Under the SPR hypothesis (see §3.1.2), Mixed NAF affixes (and to a lesser degree AF affixes of the Mstem) were all “noun-deriving affixes in PAn” (Starosta et al. 1982: 148), and the nominal Mixed NAF-words were reanalyzed at some point into verbs. The

nominalization-into-verb reanalysis helps to explain why Mixed NAF-words in conservative languages are widely used as both matrix predicates and arguments. Meanwhile, Suffixal NAF-words were believed to be the ancient verbal forms that predated the time when Mixed NAF-words took over their territory, since which Suffixal NAF-words have been marginalized to limited syntactic contexts. Similarly, the Nuclear hypothesis also vouches for the nominalization-into-verb reanalysis, but argues that three languages (Tsou, Rukai, Puyuma) did not undergo such a reanalysis, which only happened once in Proto Nuclear Austronesian (PNAn).

Alternatively, instead of saying Mixed NAF affixes were category-changing nominalizers in PAn and that Suffixal ones were the cornerstone of verbal forms, we could hypothesize that NAF-words marked by both sets were equally verbal, thus going back to the proposal in Ross (2002), where functional distinctions between the two sets lie more in grammatical moods (see Table 3.1) than in the putative nominal-verbal contrast. Then argument-denoting expressions consisting of Focus-words (or Focus nominalizations for short) would be just a matter of PAn making use of the gap strategy, whereby the nominalized Topic argument is gapped. Perhaps not incidentally, the gap strategy is also the most frequent one in East and Southeast Asia (Comrie 1998: 78). If so, PAn would be typologically similar to many Native American languages, as Ross (2002) also pointed out (see §4.2 and Comrie & Estrada-Fernández 2012). Moreover, so-called relative constructions would be the result of tapping into Focus nominalizations juxtaposed with another nominal via the parataxis-to-syntaxis diachronic pathway proposed by Givón (2012), whereby two denoting expressions packed under separate intonation contours (e.g. *John the Butcher*) are condensed into one over time. This would

account for not only the predicate and argument function of Mixed and Suffixal Focus-words, but also the syntagmatic potentials shared by both types of Focus-words across the two syntactic functions, as has been demonstrated in Chapter 3. In the words of Ross (2002: 44), Focus-words in PAN then would be “a single class... used in both predicate and arguments slots.” If so, it is just a matter of terminology whether that single class should be called verbs or something else. More importantly, if this Alternative hypothesis, as it shall be referred to hereafter, turns out to be correct, then questions of how or when the nominalization-into-verb reanalysis took place would become superfluous.

Nevertheless, reviving the Alternative hypothesis also opens up other questions that need to be addressed. This chapter attempts to do so by dealing with issues such as the Actor/possessor isomorphism (§4.1), lexicalization preference of some Focus-words over others (§4.2), cognate constructions of the Mstem (§4.3), and those of NAF-words (§4.4).

4.1. Actor/possessor isomorphism

One celebrated piece of evidence for the nominalization-into-verb reanalysis is the Actor/possessor isomorphism, whereby the Actor of NAF-words is coded in the same or similar manner as the possessor of underived nouns. However, the isomorphism is found among both Suffixal and Mixed NAF-words, in both Nuclear and non-Nuclear languages, as illustrated by Mayrinax Atayal and Rikavung Puyuma in (1) and (2) respectively.

(1) Mayrinax Atayal (M. Huang 2002: 201, 199)

- a. **ras-un=mu**=[ku guqiluh]
bring-PF=1SG.ACT=TOP.DEF banana
'I brought the banana.' [Mixed PF *-un* < PAn **-en*]
- b. **ras-aw=mu**=[ku guqiluh]
bring-PF.OPT=1SG.ACT=TOP.DEF banana
'I want to bring the banana.' [Suffixal PF *-aw* < PAn **-aw*]
- c. **ulaqi=mu**
child=1SG.GEN
'my child'

(2) Rikavung Puyuma (Fieldnotes)

- a. **na=valray mu, ku=tr<in>ima'-an**
TOP.DEF=book PTOP 1SG.ACT=<RLS>buy-LF
'As for this book, (it) is what I bought.' [Mixed LF *-an* < PAn **-an*]
- b. **na=valray mu, ku=trima'-ay=Ira**
TOP.DEF=book PTOP 1SG.ACT=buy-LF=already
'As for this book, I already bought it.' [Suffixal LF *-ay* < PAn **-ay*]
- c. **ku=turus**
1SG.GEN=sibling
'my sibling'

In both languages, the non-Topic Actor indexes share the same forms with possessor indexes.¹²⁰ If the Actor/possessor isomorphism is any good for indicating that Mixed NAF affixes were historically category-changing nominalizers, the same conclusion should apply to Suffixal NAF affixes as well. The latter claim, however, is hardly ever made.¹²¹

Ross (2009: 314) inferred that the Actor/possessor isomorphism found among Suffixal NAF-words was due to a later generalization from Mixed NAF-words at the PNA period. The idea is that speakers of PNA generalized the Actor marking for Suffixal NAF-words, which were verbal, from that for Mixed NAF-words, which were

¹²⁰ Monosyllabic person-form indexes in Mayrinax alternate between /u/ and /i/ vowels (e.g. =mi vs. =mu for 1SG). However, the alternation is phonologically and positionally conditioned (see L. Huang 2000a: 88), and has nothing to do with what roles/relations they index.

¹²¹ K. Mei (1982: 224) also questioned the validity of using the Actor/possessor isomorphism as evidence for something being nominal.

originally nominal and thus collocated with the possessor indexes, after Mixed NAF-words had been reanalyzed as verbal (i.e. nominalization-into-verb). However, Suffixal and Mixed NAF-words share the same syntagmatic subcategorizations not only in terms of Actor marking, but also with respect to other event participants, spatial-temporal adverbials, and serializations of event predication. This is so in both Nuclear languages and Puyuma, the only non-Nuclear language where systematic comparisons between Suffixal and Mixed NAF-words can be made (see §3.5). Since Puyuma did not undergo the nominalization-into-verb innovation under the Nuclear hypothesis, the generalization account cannot explain why Suffixal and Mixed NAF-words in Puyuma share the same syntagmatic subcategorizations, including the Actor/possessor isomorphism, as in Nuclear languages.

Moreover, the Actor/possessor isomorphism is also used to explain the syntactic correlations between the two roles, thus reinforcing the idea that Mixed NAF-words were historically nominal like underived nouns. For instance, conservative Austronesian languages typically restrict the non-Topic Actor phrase from being extracted to the sentence-initial position. Kaufman (2009b: 29) claimed that the basis for the Austronesian extraction restrictions is a “general ban on genitive predicates,” which he observed in Philippine languages. What the ban means is that the genitive phrase that modifies a nominal is prohibited from acting as the predicate of that nominal, as shown in (3) below.

(3) Tagalog (Kaufman 2009b: 28)

- a. ang=koponan **ni=Juan**
 TOP=team GEN=J.
 ‘Juan’s team’
- b.* **ni=Juan** ang=koponan
 GEN=J. TOP=team
 Intended: ‘The team is Juan’s.’

However, there is no such a ban in many Formosan languages (see §7.1.1), some of which even use cognate forms of Tagalog *ni*, but the extraction restrictions on the non-Topic Actor still apply in those languages. In other words, identical forms do not necessarily imply identical syntactic behaviors (see also §2.3.3 for examples from person-form clitics). Since NAF-constructions in general are geared towards the non-Actor, marking the Actor as the possessor could have been motivated by the advantage of encoding the cognitively salient entity (Langacker 1993) in the same manner (i.e. the possessor as opposed to the possessum vs. the Actor as opposed to a non-Actor-oriented event), without necessarily bringing lexical categories into play. If so, the Actor/possessor isomorphism might have been with Suffixal NAF-words as long as it has been with Mixed ones since the PAn period.

4.2. Lexicalization preference

If both AF and NAF Mixed affixes did not start out as category-changing nominalizers, why is it that modern languages demonstrate ample examples of the Mstem and Mixed NAF-words being lexicalized as nominals with semantic constancy and idiosyncrasy whereas similar examples are few and far between among Suffixal NAF-words (see §3.2.2 for some such examples in Tsou)? This lexicalization preference might relate to word order and the original functions of different Focus-words in PAn.

From a series of comparative studies by Ross (1995, 2002, 2009, 2012), we know that there are much more syntactic restrictions on Suffixal NAF-words than on the Mstem or Mixed NAF-words in modern languages that have them. The *a*-grade Suffixal NAF-words “predominantly” (because there are exceptions; see §4.4.1) express the speaker’s desire, volition, wish, or urge (Optative/Hortative forms), and the zero-grade Suffixal NAF-words are used either as a command by themselves (Imperative forms) or immediately after various elements (e.g. negators, TAM auxiliaries, or temporal-sequential markers, depending on languages) that require them (Dependent forms; all terms from Ross 2009, 2012). On the other hand, both the Mstem and Mixed NAF-words convey statements (Indicative forms). Suppose for now this was also the situation in PAn, Optative/Hortative and Imperative forms would share the semantic property of expressing speaker-oriented modality, “which represent[s] speech acts through which a speaker attempts to move an addressee to action” (Bybee & Fleischman 1995: 6). Accordingly, Optative/Hortative and Imperative forms would produce illocutionary forces pragmatically at odds with argument nominalizations, which are meant to characterize entities in terms of highly presupposed state of affairs (see also §4.4.1).

By contrast, Indicative and Dependent forms would be congruent with argument nominalizations due to their assertive nature. Indicative forms are precisely those touted as erstwhile nominalizations, their nominalization function in PAn has never been doubted. However, it is not as clear in the case of Dependent forms. At the very least, grammatical argument nominalizations containing Dependent forms are the norm in Tsou (see §3.2.2), and are restricted to those with internal negative predication in Rikavung

Puyuma (see §3.5.2.2). The latter situation is also generalizable to other languages where Dependent forms are required after the negator, as shown by Squliq Atayal in (4)

(4) Squliq Atayal (Courtesy of Tingchun Chen)

- a. laqi=maku qu [ini **ptas-i** qasa]
 child=1SG.GEN TOP NEG tattoo-LF.DEP DIST
 ‘That one who was not tattooed is my child.’
- b. buqoh nanak qu [ini=nya **niq-i**]
 banana only TOP NEG=3SG.ACT eat-LF.DEP
 ‘The only (thing) {she/he} did not eat is bananas.’

In Truku Seediq, however, Dependent forms can denote the Topic argument without there being any auxiliary-like element, just as their Indicative counterparts do. The Indicative CF *se*-STEM and the Dependent CF STEM-*ani* are respectively shown in (5) and (6), where the predicate function precedes the argument function.

(5) Truku Seediq (Tsukida 2009: 555, 550)¹²²

- a. **se-pehapuy**=na qesurux ka=[baki rubiq]
 CF-cook=3SG.ACT fish.UND TOP=old.man R.
 ‘{She/he} {cooks/cooked} fish for Rubiq’s grandfather.’
- b. ’ima ka=[**se-pehapuy**=na qesurux]
 who TOP=CF-cook=3SG.ACT fish.UND
 ‘Who is it that {she/he} {cooks/cooked} fish for?’

(6) Truku Seediq (Tsukida 2009: 447, 236)

- a. ’ini=na **keret-ani** qesurux ka=[yayu niyi]
 NEG=3SG.ACT cut-CF.DEP fish.UND TOP=knife PROX
 ‘{She/he} {doesn’t/didn’t} cut fish with this knife.’
- b. yayu niyi ka=[**keret-ani** sagas]
 knife PROX TOP=cut-CF.DEP watermelon.UND
 ‘What one should cut the watermelon with is this knife.’

If the Topic argument in (5)b is a nominalization, there is no good reason why that in (6)b should not be. In fact, reflexes of PAn Dependent forms in Truku Seediq (i.e. STEM-*i* for LF and STEM-*ani* for CF) all have the argument function (Tsukida 2009: 235). The

¹²² The transcriptions in (5) and (6) are identical to those in the cited work. The original gloss for *qesurux* in (5)b is “chicken”, which is an error.

parallel comparison here shows once again that it is unwarranted to exclude Dependent forms from the province of nominalization (as also argued in §3.2.2). Since Dependent forms are found within argument nominalizations in Tsou, Puyuma, and Atayalic (including Atayal and Seediq), coming from three first-order subgroups based on both Blust 1999 and Ross 2009, it seems reasonable to surmise that this was also the situation in PAN. This conclusion would dispel one of Ross's (2002) concerns for maintaining the Alternative hypothesis.¹²³

Since PAN was undoubtedly a predicate-initial language and that the constituent order in argument nominalizations followed that in matrix predications, Indicative but not Dependent forms would be the first immediate constituent of an argument nominalization right after a nominal relation marker (NRM), as schematically represented in (7), where X stands for the element that dictates Dependent forms.

(7) Schematic representations of argument phrases in PAN:

- a. NRM + Underived nouns
- b. NRM + Indicative forms + Other argument-structure materials
- c. NRM + X + Dependent forms + Other argument-structure materials

Focus affixes on both Indicative and Dependent forms helped to recover the grammaticalized semantic roles of the nominalized argument that was gapped. As argument phrases of the type in (7)b acquired conventionalized meanings and other argument-structure materials were stripped away, Indicative forms would then be reanalyzed into nominals on a par with basic nouns denoting physical objects (see also Foley 2014: 5) due to their similar phrasal structure.

¹²³ Based on the data available back then, Ross (2002: 46) concluded that Dependent forms “occurred only in the predicate slot”, but not in the argument slot, unlike Indicative forms, which occurred in both. The uneven syntactic distributions of the two forms would then speak against the Alternative hypothesis.

By contrast, Dependent forms would be in a disadvantaged position for lexicalization due to the X element that got in the way. In fact, even Indicative forms would have the same obstacle when they are preceded by TAM markers (e.g. (12) and (13) in Mayrinax Atayal on p.103). If PAn had been a predicate-final language, Indicative forms would have fallen under the same unfavorable circumstances as Dependent forms did in predicate-initial PAn because other argument-structure materials would then precede rather than follow Indicative forms. A good language to demonstrate this hypothetical scenario is Yaqui, a Uto-Aztecan language spoken in Mexico.¹²⁴ It is a predicate-final language that shares many similar features with conservative Philippine-type languages in terms of argument nominalizations. First, the constituent order in nominalizations follows that in matrix predications. Second, the gap strategy is used in subject, direct object, and locative nominalizations, where the verb is marked by role-indicating affixes that help to recover which of the three arguments is nominalized (roughly corresponding to Austronesian AF, PF, and LF affixes). Third, although nominalizations are externally in a paradigmatic relationship with basic nouns (by collocating with case markers, number agreement markers, demonstratives, and determiners), there is no hard evidence that the role-indicating affixes on the verb are category-changing nominalizers because all sorts of verbal features are found within nominalizations. Fourth, the agent in non-subject nominalizations shares the same marking as the possessor of basic nouns. Fifth, nominalizations immediately follow basic nouns to form restrictive relative constructions, and either of them is syntactically independent of the other.

¹²⁴ Yaqui is also spoken in Arizona, USA. Grammatical information about Yaqui presented here is based on González (2012), which investigates the Mexican variety of Yaqui.

To illustrate, compare the direct object phrases (marked in ACC) consisting of basic nouns, as in (8)a, with those comprising basic nouns modified by subject or direct object nominalizations, as in (8)b and (8)c respectively.

(8) Yaqui (González 2012: 71, 72, 88)

- a. U yoeme [uka kari-ta] jinu-k
 DET man.NOM DET.ACC house-ACC buy-PFV
 ‘The man bought the house.’
- b. Joan [uka chu’u-ta] [Maria-ta ke’e-ka-m-ta] me’a-k
 J. DET.ACC dog-ACC M.-ACC bite-PFV-NMLZ-ACC kill-PFV
 ‘John killed the dog that bit Mary.’
- c. Inepo [uka karta-ta] [em kaa tu’i-si
 1SG.NOM DET.ACC letter-ACC 2SG.GEN NEG good-ADVZ
 ji’ote-ka-’u] taya-k
 write-PFV-NMLZ burn-PFV
 ‘I burned the letter that you wrote wrong.’

Like Austronesian Focus nominalizations, Yaqui nominalizations are used in cleft-like constructions and content-word questions, as in (9).

(9) Yaqui (González 2012: 86, 89)

- a. [Wa-me yabe-m tea-ka-me] Joan
 DEM-PL key-PL find-PFV-NMLZ J.
 ‘The one who found those keys is John.’
- a. Jabesa [wa jiosam noktua-me]
 who DEM book read-NMLZ
 ‘Who is that one that is reading the book?’

Crucially, all the event-specific materials preceding the verb within a nominalization are hardly conducive to its reanalysis into a basic noun. However, when the verb expresses time-constant events and is dispensable with its argument-structure materials, reanalysis becomes promising. For instance, the subject nominalizer *-me* can create profession nouns, action/result nouns, or idiosyncratic animal names, as in (10).

- (10) Yaqui (González 2012: 83)
- a. *majta-wa-me* (teach-PASS-NMLZ) ‘student’
 - b. *etbwa-wa-me* (steal-PASS-NMLZ) ‘theft’
 - c. *totte-me* (fold-NMLZ) ‘snail’

Therefore, the lexicalization preference of Indicative forms over Dependent ones is most likely correlated with word order and the crucial syntactic difference between the two forms, namely, the X element that syntactically dictates Dependent forms. It is also because of the X element that zero-grade NAF-words in Tsou (see §3.2.2) and Rikavung Puyuma (see §3.5.2.2) are unfairly excluded from argument nominalizations, despite the fact they do occur within argument phrases, denoting the Topic NP of the Focus-word in question.

4.3. Cognate constructions of the Mstem

The Mstem consists of a stem and the abstract morpheme M-, whose exponents are reconstructable in PAn as *⟨um⟩, *ma-, or no affix at all (Ross 2009: 297). Since the zero exponent of M- is not easy to identify in modern languages if no additional information is available, I thus restrict the current discussion of the Mstem only to modern forms that contain reflexes of PAn *⟨um⟩ or *ma-.

The Mstem is found in all Formosan languages where linguistic data are available.¹²⁵ It is clear that they all use it as the matrix predicate in realis-indicative sentences, but it is less obvious how many of them also use it in the argument slot as a nominalization. The Alternative hypothesis would expect descendant forms of PAn Mstem in modern languages (at least conservative ones) to demonstrate both predicate

¹²⁵ This includes even Rukai, which has an impoverished Focus system. See Footnote 69 for the historical connection between the Mstem in Rukai and that in other Formosan languages.

and argument functions rather evenly. However, only two out of eleven Formosan languages covered in Ross (1995: 757), two out of thirteen in Ross (2002: 41), five out of fifteen in Ross (2009: 317-320), and five out of thirteen in Zeitoun & L. Huang (2006) were judged to have nominalizations made up of the Mstem, with Paiwan and Puyuma being the common denominators in all these studies. The results then beg the question of what should count as positive evidence.

Languages may have more than one way to encode nominals with agentive roles, and when that happens a common semantic contrast is made between dispositional and episodic nominals. For instance, dispositional Actor nominals in Kavalan take the form of *pa*-Kstem-*an* (e.g. *pa-taqsi-an* ‘student’) whereas episodic ones are expressed as Mstem=*ay* (e.g. *taqsi=ay* ‘{one/those} who {studies/study/studied}’). Instances of *pa*-Kstem-*an* are often translated into English profession nouns and those of Mstem=*ay* into English relative constructions, so one is very much tempted to include the dispositional *pa*-Kstem-*an* as a nominalization but sweep the episodic Mstem=*ay* into the realm of relativization, as was done in H. Chang & A. Lee (2002). However, translational equivalence is never good evidence for lexical categories or syntactic distinctions in the target language, let alone a legitimate basis for whether something is a nominalization. More importantly, the two forms have been shown by F. Hsieh (2011) to demonstrate the same syntactic potentials in subcategorizing (non-Topic) patientive arguments. Thus, the dispositional/episodic semantic contrast is not necessarily correlated with syntactic behaviors. Similarly, the two common denominators, Paiwan and Puyuma, also have different structural mechanisms for expressing the dispositional/episodic semantic contrast (see §3.3.2 for this in Tsou). In Central (Pucunug) Paiwan, for instance, episodic

agentive nominals take the Mstem (e.g. *qunu* ‘{one/those} who {cut(s) hair}’) while dispositional ones are expressed by the Mstem additionally prefixed by *'isan-* (e.g. *'isan-qunu* ‘barber’; K. Chen & R. Ma 1986: 32). Again, despite the translations, the dispositional *'isan-*Mstem and the episodic Mstem in fact share the same syntagmatic potentials (e.g. taking the same non-Topic patientive argument phrase), so either both or neither should qualify as nominalizations. However, previous studies picked only the dispositional *pa-*Kstem-*an* in Kavalan but the episodic Mstem in Paiwan as Actor nominalizations, which makes the standard of comparison questionable. The task of choosing an Actor nominalization to be compared with the Mstem used as the matrix predicate is further complicated by the fact that sometimes different affixes are required to derive an Actor nominalization, depending on the semantics of the stem. In (Jianshi) Squaliq Atayal, for instance, *t-* is prefixed to a stative stem to create derogatory or teasing terms for someone with the attribute expressed by that stem (e.g. *t-busuk* ‘drunkard; alcoholic’; cf. Mstem *m-busuk* ‘drunk’) whereas *p-* to a dynamic one to denote those who perform the action expressed by that stem routinely or by profession (e.g. *p-qalup* ‘hunter’; cf. Mstem *q<m>alup* ‘hunt’).¹²⁶

Given the aforementioned problems, a different methodology is taken here, which is based on cognate constructions that involve both syntactic similarity and cognate morphology (see Ross 2015a). Specifically, if the Mstem in a language can be used in cleft-like constructions, content-word questions, or relative constructions, where it occurs as part of a phrase that syntactically takes the argument slot and semantically denotes the

¹²⁶ More examples of these two types of nominals can be found in Hayung (2008: 31, 97), from which the two examples are drawn.

gapped argument, then the Mstem is judged to have a nominalization function.¹²⁷ If not, the forms required in those constructions are indicated separately. The outcome of this methodology shows that the nominalization function of the Mstem is the norm rather than exceptions, as summarized in Table 4.1.

Table 4.1: Nominalization function of the Mstem in Formosan languages

Nominalization function	Languages
YES	Budai & Taromake Rukai, Nanwang & Rikavung Puyuma, Northern Paiwan, Plngawan & Squliq Atayal, Tgdaya Seediq, Isbukun & Takibakha Bunun, Thao, Tsou, Saaroa, Kanakanavu, Pazeh, Siraya
NO	Mantauran Rukai: <i>ta</i> -Kstem Saisiyat: { <i>'ima/kama/AUX</i> }=Mstem/Kstem Kavalan: Mstem(= <i>ay</i>) Central Amis: Mstem- <i>ay</i>

One language from either of the two groups in Table 4.1 is illustrated in (11) and (12) respectively.

¹²⁷ In fact, applicability in one construction entails that in the others. Shibatani (2008: 911) points out that these constructions “share the property of requiring a specific type of gap in the nominalized clause employed.”

(11) (Jiangshi) Squliq Atayal

- a. sayun ru tali [qu **m-usa** mngka]
 S. CONJ T. TOP AF-go Taipei
 ‘It is Sayun and Tali who will go to Taipei.’ (L. Huang & Hayung 2011: 35)
- b. ima [qu **m-ihiy** tali]
 who TOP AF-beat T.
 ‘Who is it that beat Tali?’ (C. Lin 2005: 38)
- c. musa blaq [qu **m-nbu** qa]
 IRR AF.good TOP AF-sick DEM
 ‘The patient (i.e. sick one) would get well.’ (M. Yeh 2013: 82)
- d. [wal **m-anig** mqu] yaba=mu ga, cyux pnep-qulih
 PFV AF-eat snake father=1SG.GEN PTOPTOP PROG AF.angle-fish
 ‘My father, who ate snakes, is fishing.’ (A. Liu 2005: 94)

(12) Mantauran Rukai (Zeitoun 2007: 330, 72, 309)¹²⁸

- a. taotao [**ta-tiptip=ine** ana lroolai]
 T. SBJ.NMLZ-beat=3SG.OBL DEM.VIS child
 ‘It is Taotao that beat that child.’ [cf. Mstem: *o-tiptip*]
- b. aanga=i [**ta-tiptip** ana lroolai]
 who=3SG.GEN.VIS SBJ.NMLZ-beat DEM.VIS child
 ‘Who is it that beat that child?’ [cf. Mstem: *o-tiptip*]
- c. ki-kane=lrhao [dhona’i **ta-ka-ecelrange**]
 NEG-eat=1SG.NOM DEM.INVIS SBJ.NMLZ-K-black
 ‘I won’t eat that black one.’ [cf. Mstem: *ma-ecelrange*]
- d. ma-dhalame=lrhao [dhona’i **ta-ka-ecelrange**] molrae
 STAT-like=1SG.NOM DEM.INVIS SBJ.NMLZ-K-black fabric
 ‘I like the black fabric.’ [cf. Mstem: *ma-ecelrange*]

If either the *t*-STEM or *p*-STEM form in Squliq Atayal (see above) had been chosen as the Actor nominal to be compared with the Mstem, we would come to the conclusion that the Mstem has only predicate but no argument function, which is apparently not so given the examples in (11). Like *t*-STEM and *p*-STEM, the Mstem in Squliq Atayal is just one of the many forms that can be used to denote the Actor.¹²⁹ By contrast, Mantauran Rukai does not have as many forms at its disposal, and consistently uses the *ta*-Kstem rather than the

¹²⁸ See Footnote 87 for what the gloss GEN means for the morpheme =i in the (b) example.

¹²⁹ Similarly, Adelaar (2011: 149-155) distinguishes two types of nominalization in Siraya, one of “verbal bases” and the other of “derived verbs.” Mstem nominalizations belong to the latter type.

Mstem for the nominalization function whether its referent is an attribute-bearer, a routine-doer, or an episodic agent.¹³⁰

There are some pieces of evidence suggesting that using the Mstem for both predicate and argument functions was the original pattern whereas the four minority languages in Table 4.1 were innovators. First, Ross (2012: 1269) reconstructed PAn *ta-STEM as the form for Actor nominalizations. The stative *t*-STEM in Squiliq Atayal expresses Actor nominalizations that are semantically more dispositional than those expressed by the Mstem. Similarly, *ta*-STEM in Pazeh (e.g. *ta-turazak* ‘someone talkative’; P. Li 2002: 232) denotes dispositional agentive nominals whereas its Mstem nominalizations are episodic and situationally bound. Hence, it seems reasonable to hypothesize that the original meaning of PAn *ta-STEM was dispositional in nature. Unlike all the other Rukai dialects, Mantauran uses the *ta*-STEM rather than the Mstem for subject nominalizations in constructions like those in (12), which might have resulted from generalizing the dispositional *ta-STEM to include more episodic situations.¹³¹ In other words, even within Rukai dialects, Mantauran is an outlier due to its restriction against using the Mstem for the purpose of subject nominalizations. As suggested earlier in §3.4.1, this morphosyntactic change seems to be correlated with the fact that Mantauran has the most impoverished inventory of (prenominal) relation markers among Rukai dialects, which are crucial elements for maintaining a structural distinction between expressions denoting a predication and those denoting an argument of that predication. The loss of (prenominal) relation markers in Mantauran may thus be

¹³⁰ Strictly speaking, the *ta*-Kstem also denotes the patient when the Kstem itself is marked by the passive marker *'i-*, as in *ta-'i-alroho* (SBJ.NMLZ-PASS-lift) ‘the one who is lifted’. However, the Mstem of passive forms (identical to the Kstem; e.g. *'i-alroho*) does not involve reflexes of PAn *<um> or *ma-, so passive forms are not taken into consideration here. See Zeitoun (2007: 143) for more on Mantauran passive.

¹³¹ See Footnote 86 for the realizations of the Mstem in Rukai dialects.

compensated to a nicety by maintaining a formal difference between the Mstem for the predicate function and the *ta*-STEM for the argument function.

Second, the two markers *kama* and *'ima* in Saisiyat were singled out as relativizers in M. Yeh (2003) and as Actor nominalizers in Zeitoun et al. (2011). However, H. Huang (2003: 101) shows that the marker *'ima* indicates the speaker's positive epistemic belief (hence glossed EPIS here) and is used in contexts not necessarily involving argument nominalizations, as in (13).

(13) Saisiyat (H. Huang 2003: 103)

- a. 'obay isahini' **'ima**=sa'sa'ih ka=hahila: m-atawaw
 O. now EPIS=every LIG=day AF-work
 'Obay works everyday.'
- b. 'aehoe' **'ima**=Sepat ka=tatakay
 dog EPIS=four LIG=leg
 'Dogs have four legs.'

Similarly, the marker *kama* expresses habitual or dispositional meanings and is used with both matrix predicates and those within argument nominalizations. Accordingly, T. Lee (2010) argues that Epistemic *'ima* and Habitual *kama* are simply two of the many TAM auxiliaries available in the language, along with Experiential *'ina*, Future *'am*, and Progressive *mam*, that are required within Actor nominalizations, where these TAM markers show complementary distributions, as in (14).¹³²

¹³² Nevertheless, probably because of its dispositional nature, the Habitual *kama* is more likely to create lexicalized nominals with non-compositional semantics than other TAM markers, such as *kama=manra:an* 'man, husband' (cf. Mstem *manra:an* 'walk') and *kama=maS'abo* 'woman, wife' (cf. Mstem *maS'abo* 'burn something into ashes'). Once *kama*=Mstem is lexicalized, the result nominal may then be marked by other TAM markers, just like underived nouns.

(14) Saisiyat (T. Lee 2010: 76)

- a. [ʔima=S<om>bet ka=korkoring] ka=hiza' ma'iaeh
 EPIS=<AF>beat UND=child LIG=DIST person
 'that person who beats the child'
- b. [mam=S<om>bet ka=korkoring] ka=hiza' ma'iaeh
 PROG=<AF>beat UND=child LIG=DIST person
 'that person who is beating the child'
- c. [ʔam=S<om>bet ka=korkoring] ka=hiza' ma'iaeh
 FUT=<AF>beat UND=child LIG=DIST person
 'that person who will beat the child'

Unlike the Mstem used for Actor nominalizations, the Mstem functioning as matrix predicates may dispense with TAM markers, as in (15).

(15) Saisiyat (M. Yeh 2003: 14)

yaba' **S<om>bet** ka=korkoring
 father <AF>beat UND=child
 'Father beat the child.'

The comparison between (14) and (15) shows that Actor nominalization in Saisiyat makes use of the same gap strategy as in other Formosan languages (i.e. the first group in Table 4.1). What makes Saisiyat different is that it requires some TAM markers to help substantiate a referent. Since these markers do not seem to have cognate forms in other languages, this TAM requirement is most likely an innovation in Saisiyat.

Finally, both Kavalan and Central Amis add the marker /ay/ to the Mstem to create Actor nominalizations. The requirement of this additional marking is rather strict in Central Amis (see §5.2.1.1), but seems to be flexible in Kavalan (hence the parenthesis in Table 4.1) because, based on examples from the early literature (Tsuchida 1993a), Actor nominalizations in Kavalan can be comprised of the Mstem without the marker /ay/. For instance, compare:

(16) Central Amis (D. Liu 1999: 48)

- a. **mi-kalat** ku=ra oner tu=wacu
 AF.EXT-bite TOP.CMN=DIST snake UND.CMN=dog
 ‘That snake is biting a dog.’
- b. u=ra oner ku=[{**mi-kalat-ay**/*mi-kalat}] tu=wacu]
 CMN=DIST snake TOP.CMN=AF.EXT-bite-NMLZ UND.CMN=dog
 ‘What is biting a dog is that snake.’

(17) Kavalan (Tsuchida 1993a: 94)

- a. **s<m>aqut=ti=iku** tu=paRin
 <AF>move=already=1SG.TOP UND.CMN=tree
 ‘I already moved {the wood/trees}.’
- b. tiana ya=[**s<m>aqut** tu=paRin=a yau]
 who TOP=<AF>move UND.CMN=tree=LIG MED
 ‘Who is it that moved {that piece of wood/those trees}?’

Since the two languages have been in close contact over a hundred years, it is highly possible that one has borrowed the marker /ay/ from the other. Sagart (2013: 491) assumed that the Amis /ay/ (as in the Mstem-ay) is a reflex of the PAN LF *-ay. Nevertheless, distributions of the Amis -ay do not seem to support such a historical connection. In languages that clearly retain the PAN LF *-ay (e.g. Atayal, Seediq, Paiwan, Puyuma, and Pazeh), the cognate marker is at least used in NAF-constructions, although (Mayrinax) Atayal (e.g. *m-aras-ay* [AF-bring-OPT] ‘want to bring’; L. Huang 2002: 201) and Pazeh (e.g. *m-apa’-ay* [AF-carry-IRR] ‘will carry’; P. Li 2000: 91) additionally combine it with the AF Mstem. By contrast, in both Amis and Kavalan the /ay/ marker is only compatible with the AF Mstem, thus making suspicious its historical connection with the PAN LF *-ay. In addition, clear cognates of the PAN LF *-ay are only suffixable to the Mstem/Kstem, to which other Focus affixes may alternatively attach. However, the marker /ay/ in both Amis and Kavalan combines with not only the Mstem, but also forms that generally do not take Focus affixes, such as numerals, possessive substantives, and even underived nouns (H. Jiang 2011). These distributions of /ay/ in both Amis and

Kavalan suggest that its use in verbal-based nominalization (i.e. collocating with the Mstem/Kstem) might have derived from its initial function in nominal-based nominalization (i.e. collocating with underived nouns), a development proposed by Shibatani & Shigeno (2013) for a similar phenomenon in Ryukyuan languages. In what follows, I illustrate this point with Kavalan data from spontaneous speech. Amis nominalization will be treated in more detail later in Chapter 5.

The Kavalan example in (18) shows that when an underived noun is marked by =*ay*, the result form semantically denotes what is metonymically associated with that noun and syntactically makes up a complete argument NP.

(18) Kavalan (NTU Corpus|TeachWeaving.Abas.Ipay: IU 88)

...(0.7) mai qa=mai syuqay tu=sikawman si
 FS EPIS=NEG introduce UND.CMN=speech if
 taRni=na qaya na=[**taypaq=ay**]
 how.would.one.know=3.ACT also ACT.CMN=Taipei=NMLZ
 ‘If (someone) doesn’t explain (the Kavalan) language, how could those Taipei
 (people) possibly understand it?’

The NP *taypaq=ay* in (18) specifically refers to inhabitants of Taipei, much as *New Yorker* may refer to inhabitants of New York. Crucially, this function of the Kavalan =*ay* in deriving a new nominal out of an existing one has not been reported in languages that have unquestionable reflexes of PAn LF *-*ay*.

In addition, the Kavalan =*ay* attaches to attribute words to create nominals characterized by those attributes. For instance, the speaker of (19) was talking with another woman about earthquakes that had happened in her village.

(19) Kavalan (NTU Corpus|Earthquake.Abas.Raciang)

- a. *yau=pama=imi ta=tebalun-an yu 'nay Raya utuz*
 EX=still=1EXCL.TOP LOC=T.-LOC INT MED big earthquake
 'We were still in Tebalun (at the time of) that big earthquake.' [IU 31]
- b. *wama 'nay Raw tasaw Raya=ay*
 only MED INT year big=NMLZ
 'Only in that year (did we have) a big (one, i.e. earthquake).' [IU 203]

After introducing *Raya utuz* 'big earthquake' into the discourse in the first example (IU 31), the speaker anaphorically referred to it as *Raya=ay* 'big one' in the second example after a long stretch of talk (IU 203), where the attribute word *Raya* 'big' is not modifying any noun. A similar situation is also observed in (20), which is taken from one of the eight Frog narratives studied by H. Jiang (2006).

(20) Kavalan (NTU Corpus|Frog.Ungi)

- a. *m-zuqat=ti u-zusa XX 'nay suyki Raya*
 AF-exit=already NHUM-two MED frog big
 'Those two big frogs came out.' [IU 72]
- b. *m-zuqat mazmun=ti ya=sunis=na na=suyki unay Raya=ay*
 AF-exit AF.many=already TOP=offspring=3.GEN GEN=frog MED big=NMLZ
 'The offspring of those frogs, the big ones, are numerous. (They all) came out.'
 [IU 76]

The same attribute word *Raya* 'big' (postnominally) modifies *suyki* 'frog' (a loan word from Taiwanese Southern Min) in the first example (IU 72), and later is nominalized by =*ay* (IU 76), thus referring to big frogs previously introduced into the discourse. However, unlike that in (19), the nominalized phrase in (20) is juxtaposed to a complete nominal phrase (i.e. *suyki unay* 'those frogs') as an afterthought. As this juxtaposition becomes entrenched enough, the marker =*ay* would be generalized into an all-purpose marker in a typical attributively modified phrase. For instance, the prenominal modifier is a temporal adverbial in (21) and an Mstem in (22), both marked by =*ay*.¹³³

¹³³ Despite its initial /m/ sound, the word *masang* 'before, in the past' is not an Mstem, nor does it attract

(21) Kavalan (NTU Corpus|Earthquake.Abas.Raciang: IU.04)

- a. ngid=iku ipil tu=[**masang=ay** utuz] zin=na nani
 AF.want=1SG.TOP hear UND.CMN=before=NMLZ earthquake say=3.ACT DM
 ‘He said, “I want to hear (something about) earthquakes in the past.”’

(22) Kavalan (NTU Corpus|Ancestors.Buya: IU 19-21)

- a. XXX spaw tu=[**m-patay=ay** wasu]
 put UND.CMN=AF.die=NMLZ dog
- b. u [**m-patay=ay** taquq]
 or AF.die=NMLZ chicken
- c. spaw-an=na ta=zenna na=qaniyau
 put-LF=3.ACT LOC=water.field ACT=3PL
 ‘They left dead dogs or dead chickens in the water field.’

Thus, the Mstem is simply one of the various word forms that may host the marker =*ay*, and its distributions speak against the view that it is a reflex of the PAn LF *-*ay*. The same conclusion can also be made about the Amis -*ay*, although it is strictly word-final, unlike the Kavalan =*ay*, which can be phrase-final (see F. Hsieh 2011). In the face of its unclear historical source, the /*ay*/ marker found in Kavalan and Amis Actor nominalizations is most likely a later development.¹³⁴

Finally, there is an implicational relationship between syntactic functions of the Mstem and its perfective counterpart marked by reflexes of PAn *<in> (schematically represented as <IN>Mstem). In languages where the latter form is also available, if the

person-form clitics.

¹³⁴ On a related note, Bunun has a nominalizing suffix -*az* [aǎ], which is marked on underived nouns to create idiosyncratic nominals or on stative Kstems to derive dispositional nominals, although this is not a productive morphological process. Two Bunun varieties are illustrated below:

(i) Isbukun Bunun (R. He et al. 1986: 33)

- a. *halup-az* (peach-NMLZ) ‘persimmon’
 b. *taimang-az* (stupid-NMLZ) ‘idiot’ [cf. Mstem: *ma-taimang* ‘stupid’]

(ii) Takibakha Bunun (I. Chen 2009: 75)

- a. *’vut-az* (snake-NMLZ) ‘worm’
 b. *kulkul-az* (jelly-NMLZ) ‘white fungus’
 c. *sanglav-az* (green-NMLZ) ‘green bamboo viper’ [cf. Mstem: *ma-sanglav* ‘green’]
 d. *duqlaz-az* (white-NMLZ) ‘flying squirrel with a white nose’ [cf. Mstem: *ma-duqlaz* ‘white’]

Mstem has the nominalization function, it follows that the <IN>Mstem is also applicable for the same function. Like the Mstem, the <IN>Mstem can make up an NP by itself or together with other argument-structure materials that it subcategories, as respectively illustrated by Saaroa in (23) (where PAn *<in> is reflected as *hli-*) and Truku Seediq in (24) (where PAn *<in> is reflected as <*en*>). In both examples, the <IN>Mstem occurs in the same argument position as whatever Mixed NAF-words available in the languages.

(23) Saaroa (C. Pan 2012: 320, 318)

- a. ngahla=isa [hli-makari]
 who=3.TOP PFV-AF.call
 ‘Who is it that have called?’
- b. ngahla=isa [a-tama~tamahleng-a=u aari~aari]
 what=3.TOP IRR-IPFV~make-PF=2SG.ACT PL~day
 ‘What is it that you do every day?’

(24) Truku Seediq (Tsukida 2009: 572, 573)

- a. masaw ka [s<en>ipaq laqi=mu] hug
 M. TOP <AF><PFV>hit child=1SG.GEN QP
 ‘Is it Masaw that hit my child?’
- c. manu ka [’uq-un=su merebu]
 what TOP eat-PF=2SG.ACT morning
 ‘What is it that you ate in the morning?’

With or without argument-structure materials, the nominality of Actor nominalizations is better attributed to the construction as a whole rather than to the Mstem or <IN>Mstem alone. This point is best illustrated by Isbukun Bunun, where there are two grammatical sets of demonstrative clitics, one for Topics and the other for non-Topics. Either set has three members, which are deictically contrastive in terms of distance, as in (25).

(25) Isbukun Bunun (H. Lin 1997: 16)

kalat=[a asu={**in/an/a**}]=[mas ngaung={**cin/tan/cia**}]
 AF.bite=TOP dog={PROX/MED/DIST}.TOP=UND cat={PROX/MED/DIST}.NTOP
 ‘{This/that/that} dog bit {this/that/that} cat.’

Importantly, the two sets of demonstrative clitics illustrate different scopes of modification. As argued by H. Jiang (2012), demonstrative clitics for Topics are phrasal whereas those for non-Topics are head-adjacent. Crucially, this difference helps to demarcate the boundary of Actor nominalizations. Compare, for instance, the three examples in (26).

(26) Isbukun Bunun (H. Jiang 2012)

- a. minsum=in=[a **m<in>inbuhbuh** **sia** [**libus=cin**]]
 AF.appear=already=TOP <PFV>AF.get.lost LOC forest=PROX.NTOP
 ‘The (one who) got lost in this forest has returned.’ [The speaker is in the forest.]
- b. minsum=in=[a **m<in>inbuhbuh** **sia** **libus=in**]
 AF.appear=already=TOP <PFV>AF.get.lost LOC forest=PROX.TOP
 ‘This (one who) got lost in the forest has returned.’
 [The speaker is close to the returner.]
- c. minsum=in=[a **m<in>inbuhbuh** **sia** [**libus=cin**]=in]
 AF.appear=already=TOP <PFV>AF.get.lost LOC forest=PROX.NTOP=PROX.TOP
 ‘This (one who) got lost in this forest has returned.’
 [The speaker is in the forest and close to the returner at the same time.]

In (26)a, the non-Topic proximal =*cin* attaches to the locative noun *libus* ‘forest’, which is expected since it is a non-Topic argument. What is less expected, however, is that the Topic proximal =*in* can replace its non-Topic counterpart, giving rise to (26)b, which only makes sense if the Topic proximal =*in* has its limiting scope over the entire Actor nominalization that serves as the Topic argument of the matrix predicate *minsum=in* ‘AF.appear=already’ (a coalescenced form of *minsuma=in* ‘AF.appear=already’). This analysis is confirmed by the different deictic interpretations of *libus* ‘forest’ in (26)a and (26)b. Further evidence comes from (26)c, where both the Topic and non-Topic proximal demonstrative attach to the same host, with each having different scopes of modification.

The nominal function of the <IN>Mstem is generally not recognized in the literature due to its “clausal” nature. However, if the locative phrase *sia libus=cin* ‘in this forest’ is removed from (26)c, the Topic proximal =*in* would then attach to the <*in*>Mstem instead, thus making it syntactically paradigmatic with underived nouns in (25). Since argument-structure materials of a Focus-word are contingent, they should not be a deciding factor for whether or not a Focus-word has the nominalization function. This is precisely why a construal approach is taken in the present study.

To sum up, Formosan languages predominantly use the Mstem (as well as its morphological derivative <IN>Mstem) for both predicate and argument functions, as is expected under the Alternative hypothesis. Only a small number of them use forms other than the Mstem alone for Actor nominalizations, and there are good reasons to believe that these languages are innovators.

4.4. Cognate constructions of NAF-words

Unlike the Mstem, NAF-words in modern languages demonstrate syntactic behaviors that present more challenges to the Alternative hypothesis, primarily because of the uneven distributions of Suffixal and Mixed NAF-words across their predicate and argument functions. This section discusses cognate constructions of Suffixal NAF-words in §4.4.1 and those of Mixed ones in §4.4.2.

4.4.1. Suffixal NAF-words

Since the predicate function of Suffixal NAF-words in modern languages is rarely an issue, this section focuses on their argument function.

As has been mentioned in §4.2, Tsou, Puyuma, Atayal, and Seediq use zero-grade NAF-words for argument nominalizations. Importantly, argument nominalizations consisting of zero-grade NAF-words obey the same Topic-only constraint as those comprised of Mixed NAF-words. For instance, the two argument nominalizations in (27) from Truku Seediq, one headed by the Mixed *se*-STEM and the other by the Suffixal STEM-*ani*, can only denote the Conveyance Topic NP, which is the beneficiary of an action in both cases.

(27) Truku Seediq (Tsukida 2009: 550, 236)

- a. 'ima ka=[**se-pehapuy**=na qesurux]
 who TOP=CF-cook=3SG.ACT fish.UND
 'Who is it that {she/he} {cooks/cooked} fish for?'
- b. baki=su ka=[**keret-ani** sagas]
 grandfather=2SG.GEN TOP=cut-CF.DEP watermelon.UND
 '(The person who) one should cut the watermelon for is your grandfather.'

Since languages that use zero-grade NAF-words for argument nominalizations come from at least three first-order subgroups of the Austronesian family according to both Blust's (1999) and Ross's (2009) subgrouping accounts, it can be inferred that this was also the situation in PAN, which is expected under the Alternative hypothesis.

By contrast, *a*-grade NAF-words are prohibited from argument nominalizations in "almost" (more on this below) all the Formosan languages that retain these forms. This is so irrespective of whether these forms express optative/hortative meanings or not. The syntactic restriction on *a*-grade NAF-words requires some explanation if the Alternative hypothesis is to be maintained.

The restriction against *a*-grade NAF-words functioning as argument nominalizations is specifically mentioned in grammatical descriptions, including at least those of Mayrinax Atayal (L. Huang 2002: 207), Truku Seediq (Tsukida 2009: 236), and

Central Paiwan (W. Huang 2012: 118). As speculated in §4.2, such a restriction might have to do with pragmatic incongruence. The *a*-grade NAF-words in these languages expresses optative/hortative meanings and are used to declare the speaker's volition or urge whereas argument nominalizations are meant to denote someone or something in terms of a presupposed state of affairs.

For instance, Mayrinax Atayal has two classes of Focus-words for unrealized events, the Irrealis Optative (expressed by *a*-grade Suffixal NAF-words) and the Irrealis Future (expressed by Mixed NAF-words; see also Table 3.2). While the former performs the illocutionary act of declaring the speaker's desired state of affairs (see Y. Cheng 2013), the latter asserts that a state of affairs will likely pertain in the future, as contrasted by the Optative PF Kstem-*ay* and Future PF *Ca~Kstem-un* in (28).

(28) Mayrinax Atayal (L. Huang 2001: 64)

- a. **niq-aw**=mu=[ku siyam]
 eat-PF.OPT=1SG.ACT=TOP pork
 'I want to eat the pork (now)!' (i.e. 'If only I could eat the pork!')
- b. **na~niq-un**=mu=[ku siyam]
 IRR~eat-PF=1SG.ACT=TOP pork
 'I will eat the pork (but not immediately).'

Perhaps due to this pragmatic difference, the Optative PF Kstem-*ay* does not have the nominalization function whereas the Future PF *Ca~Kstem-un* does (e.g. *na~niq-un* 'what will be eaten'). The same restriction is generalizable to other Irrealis Optative forms in Mayrinax, all suffixed by the *a*-grade series.

Based on the present speculation, what prevents *a*-grade NAF-words from serving in argument nominalizations is their incompatible semantics rather than their lexical category (presumably as verbs). More weight will be added to this idea if we can find a language where *a*-grade NAF-words are indeed used in argument nominalizations

because they express meanings with illocutionary forces compatible with argument nominalizations.¹³⁵ Such a language is (or *was*, to be more precise) indeed attested, and it is because of this language that the aforementioned caveat “almost” is needed. It is the now extinct (or “dormant” as preferred by language revitalizationists) Siraya language, to which I turn below.

According to Adelaar’s (2011) analysis of 17th-century liturgical texts, Siraya had reflexes of all the three PAn *a*-grade NAF-suffixes (PF *-aw*, LF *-ey*, and CF *-aney*). Focus-words with these suffixes (or Subjunctive forms in his term) “[express] a wish as well as future tense” (ibid.: 113), as illustrated by the two examples in (29) respectively.

(29) Siraya (Adelaar 2011: 205, 96)

- a. **ka-vāngo-aw**=mhu ta ra~ruma=hu
 K-love-PF.IRR=2SG.ACT TOP PL~other=2SG.GEN
 ‘You shall love your neighbors (lit. others).’ [cf. Mstem *mä-vāngo* ‘love’]
- b. **pa-i-alak-aw** lava tumang ta ti Jesus
 K-LOC-offspring-PF.IRR perhaps where TOP PSN Jesus
 ‘Where will this Jesus be born?’ [cf. Mstem *ma-i-alak* ‘give birth’]

While the first example clearly involves the speaker’s will, the second has more to do with non-subjective future. And it is this latter future sense that is found in argument nominalizations consisting of *a*-grade NAF-words, as in (30), where the PF Kstem-*aw* is syntactically embedded under a nominal relation marker just like a typical noun phrase.

¹³⁵ Ross (2009, 2012) reconstructed two types of NAF-words with zero-grade suffixes in PAn, the Imperative form *STEM-u/i/ani and the Dependent form *STEM-a/i/ani. While reflexes of *STEM-a/i/ani have been found to serve in argument nominalizations (in Tsou, Puyuma, and Atayalic at least), those of *STEM-u have not. It seems non-coincidental that *STEM-u is a strictly imperative form, unlike *STEM-i/ani. To go with the idea of pragmatic incongruence, the strictly imperative form *STEM-u would create the same kind of illocutionary forces incompatible with argument nominalizations as hortative/optative NAF-words containing *a*-grade suffixes.

(30) Siraya (Adelaar 2011: 96, 91)

- a. timang ta [Rarenan=au]
 who TOP Mother=1SG.GEN
 ‘Who is my mother?’
- b. mang ta [kan-aw=miän]? mang ta [it-aw=miän]?
 what TOP eat-PF.IRR=1EXCL.ACT what TOP drink-PF.IRR=1EXCL.ACT
 ‘What are we going to eat? What are we going to drink?’

More generally speaking, when embedded under nominal relation markers, all *a*-grade NAF-words in Siraya can denote the gapped argument just as Mixed NAF-words do, except that the former do so with respect to a (relative) future state of affairs. For instance, both the *a*-grade PF-word *ukukua-aw* (marry-PF.IRR) ‘someone to be married; wife-to-be’ (p.153; cf. Mstem *m-ukukua* ‘AF-marry’) and the Mixed PF-word *pa-titil-ən* (K-quarrel-PF) ‘what is quarreled about; issue’ (p.151; cf. Mstem *ma-titil* ‘AF-quarrel’) may denote a patientive nominal. Similarly, both the *a*-grade LF-word *pä-i-hahab’-ey* (K-LOC-rest-LF.IRR) ‘place where one will rest’ (p.154) and the Mixed LF-word *pa-i-sasu-an* (K-LOC-rule-LF) ‘place that is ruled over; kingdom’ (p.156; cf. Mstem *ma-i-sasu* ‘AF-LOC rule’) may denote a locative nominal.¹³⁶

Suppose the original function of *a*-grade NAF-words in PAn was optative/hortative in nature, these forms would have been excluded from argument nominalizations in PAn. Modern languages that inherit this function (e.g. Mayrinax Atayal, Truku Seediq, and Central Paiwan) also demonstrate the same restriction. The exception in Siraya is most likely a later development that resulted from having extended the semantics of *a*-grade NAF-words from speaker-oriented volition to objective future, a diachronic change that is well documented in grammaticalization studies (e.g. Heine

¹³⁶ All the page numbers are from Adelaar (2011). Both Kstems *pä-i-hahab’* ‘K-LOC-rest’ and *pa-i-sasu* ‘K-LOC-rule’ belong to the same verb class, so the expected Mstem of the former would be *mä-i-hahab’* ‘AF-LOC-rest’, much as the Mstem of the latter is *ma-i-sasu* ‘AF-LOC-rule’. However, the expected Mstem *mä-i-hahab’* is not given here because it is not found in Adelaar (2011).

2003). The indicative-future sense of *a*-grade NAF-words in Siraya, which are dissociated from directive illocutionary forces, might have facilitated their use in argument nominalizations. The Siraya situation is expected along the lines of the Alternative hypothesis, but not of the nominalization-into-verb hypothesis, according to which *a*-grade NAF-words would be strictly verbal.

Finally, *a*-grade NAF-words in Puyuma have not only optative/hortative but also realis-indicative meanings. However, unlike those in Siraya, *a*-grade NAF-words in Puyuma are prohibited from argument nominalizations (see §3.5). This syntactic restriction requires some explanation. Ross (2009: 303) offered two possible accounts. One is that Puyuma innovated by extending the original function of *a*-grade NAF-words in PAN from hortative/optative to realis-indicative. The other one is that *a*-grade NAF-words already expressed both hortative/optative and realis-indicative in PAN, and that this situation is retained by Puyuma. His Nuclear hypothesis supports the retention account because Puyuma is a non-Nuclear language that retains the functions of both Mixed and *a*-grade NAF-words in PAN, where the former were realis nominalizations and the latter both optative/hortative and realis-indicative verbs. Moreover, PNAn innovated by replacing *a*-grade NAF-words (i.e. realis verbs in PAN) with Mixed NAF-words (i.e. realis nominalizations in PAN) via the nominalization-into-verb reanalysis, thus leaving *a*-grade NAF-words with only optative/hortative meanings, a situation inherited by modern Nuclear languages.

However, the putative nominal-verbal contrast between Suffixal and Mixed NAF-words based on the Nuclear hypothesis does not predict their syntactic distributions across Formosan languages, be it Nuclear or not. The prerequisite for the nominalization-

into-verb reanalysis to happen at all is that NAF-words could be reasonably classified into nominalizations (i.e. Mixed ones) on the one hand and verbs (i.e. Suffixal ones) on the other at the PAn period so that PNAn could reanalyze the former (i.e. Mixed NAF-words) into its new-generation verbs. But evidence for such a situation in PAn hinges heavily on the data in Puyuma, the only non-Nuclear language with a robust three-way formal distinction of NAF affixes from both the Suffixal and Mixed set (since Rukai lacks the former and Tsou the latter). Discussions in §3.5.2 have shown that it is problematic to characterize these two types of NAF-words along the nominal-verbal contrast in both Nanwang and non-Nanwang varieties, where both Mixed NAF-words and Suffixal zero-grade ones have both the predicate and argument function and share the same syntagmatic subcategorizations to the exclusion of underived nouns. This current situation in Puyuma then questions a meaningful contrast between nominalizations and verbs among NAF-words in PAn, whose functions are believed to be retained in Puyuma under the Nuclear hypothesis. More importantly, even Suffixal *a*-grade NAF-words may have the nominalization function as Mixed ones do and observe the same Topic-only constraint, provided that the illocutionary force of a NAF-word does not interfere with the presuppositionality required for argument nominalizations, as is found in Siraya. Among languages that retain Suffixal *a*-grade NAF-words, Puyuma and Siraya are both rather exceptional, the former in terms of their semantics (i.e. expressing realis-indicative) and the latter with respect to their syntactic distributions (i.e. having the argument nominalization function). The generalization is that if NAF-words reflecting PAn *STEM-aw/ay/anay can serve as argument nominalizations, it follows that these word forms have non-hortative/optative semantics, as in Siraya, but not vice versa, as in

Puyuma.¹³⁷ The fact that Suffixal *a*-grade NAF-words in Puyuma have realis-indicative semantics but do not have the nominalization function suggests a relatively recent functional extension from hortative/optative to realis-indicative, a point to be further explored below.

Moreover, if the realis function of *a*-grade NAF-words was replaced by Mixed NAF-words at the PNAn period, as the Nuclear hypothesis goes, it is expected that *a*-grade NAF-words have realis meanings only in non-Nuclear languages, but not in Nuclear ones. However, this prediction is not borne out. Kanakanavu, a Nuclear language, uses *Kstem-ai* (< PAN *a*-grade LF **Kstem-ay*) in realis-indicative NAF-constructions with the patientive Topic.¹³⁸ Ross (2009: 313) noticed this issue, but adopted Tsuchida's (1976) analysis that the default realis PF form is *Kstem-ene* (< PAN Mixed PF **Kstem-en*) and that *Kstem-ai* is simply a "special" Focus form used in "narrative" sequences. However, this analysis does not make much a difference on the fact that the Kanakanavu *Kstem-ai*, "special" or not, has realis-indicative meanings. In addition, D. Liu (2014) has recently shown that the two NAF-words are correlated with different tense/aspect markings, with neither of them being more "special" than the other. While the *a*-grade *Kstem-ai* tends to collocate with the perfective marker *ni-*, the Mixed *Kstem-ene* is more likely to co-occur with future or progressive markers. More importantly, evidence from spontaneous texts in Tsuchida (2003) calls into question the idea that only the *a*-grade *Kstem-ai*, but not the Mixed *Kstem-ene*, is used in "narrative"

¹³⁷ According to P. Li (2000), like those in Siraya, Suffixal *a*-grade NAF-words in Pazeh can express indicative-future. However, it is not clear if they have the nominalization function as well. Either way, the generalization here is still valid.

¹³⁸ However, under Zeitoun & S. Teng's (2014) version of the Nuclear hypothesis both Kanakanavu and Saaroa are non-Nuclear languages. This proposal is also based on the nominalization-into-verb reanalysis, as in Ross (2009).

sequences. To illustrate, two snippets from a story about shooting the sun are given in (31) and (32).¹³⁹

(31) Kanakanavu (Tsuchida 2003: 16; #7-8)

- a. **l<um>a~liu'u=kani** sua=nanake
 <AF>IPFV~fish.with.a.scoop.net=REP TOP=woman
 'The woman was fishing with a scoop net.'
- b. **paratea-nene=kani=kiai** sua=sanapisapi
 scoop-PF=EVID=3.ACT TOP=driftwood
 'She scooped a (piece of) driftwood.'
- c. **ala-ene=kani=kiai** t<um>aini valuvalu canumu
 take-PF=EVID=3.ACT <AF>throw.away rapids water
 'She picked it up and threw it into the rapids of water.'

(32) Kanakanavu (Tsuchida 2003: 22-23; #53-54)

- a. **tia=cu** pana-ene=kiai sua=taniare
 FUT=already shoot-PF=3.ACT TOP=sun
 'They were going to shoot the sun.'
- b. **kili-ai=kani='inia** na=taakiringa sua=talisi
 tie-LF=EVID=3.ACT LOC=small.basket TOP=rope
 'He tied the rope to a small basket.'
- c. **tarakari-ai=kani='inia** sua=tamu=ini
 ask-LF=EVID=3.ACT TOP=grandfather=3.GEN
 'He asked his grandfather.'

The sentences where both of the two types of NAF-words occur all describe sequential events that advance the main storyline. It is thus unclear how one is more “narrative” than the other. Therefore, the realis-indicative function of the *a*-grade Kstem-*ai* in Kanakanavu remains a puzzle under the Nuclear hypothesis.

¹³⁹ The transcriptions here follow those in Tsuchida (2003), although the representational format is slightly different and the diacritic for stress is omitted. Where morpheme-by-morpheme glosses are not available in the original source, morphemic boundaries are not indicated here. Moreover, the glosses for the two NAF affixes are not based on the semantic roles of the Topic argument, but on their formal relationship with corresponding PAn etyma. The permutation between PF and LF is common in Formosan languages and independent of the issue discussed here. Finally, person-form Actor indexes for the two types of NAF-words in these two examples come from two different paradigms, as also stated in D. Ho (1997: 254). But this is not always the case based on the data in Tsuchida (2003) as well as D. Liu (2014).

Aside from Kanakanavu, Puyuma is the only other Formosan language where *a*-grade NAF-words have realis-indicative meanings. By contrast, cognate forms in all the others have highly irrealis meanings such as optative/hortative, imperative, or relative/near future. In Kanakanavu, the realis-indicative function is specifically restricted to *Kstem-ai*, but in Puyuma this function is more extensive and available to all the three *a*-grade NAF-words (i.e. *Kstem-aw/ay/anay*). Since these two languages differ from all the others in terms of the semantics of *a*-grade NAF-words, would it be possible that they both innovated (probably independently) by extending the original irrealis function to include the realis? Ross (2009: 304) considered such a semantic change “implausible” whereas both Sagart (2010: 202) and Aldridge (2016: 55) suggested a hypothetical syntactic development that precisely triggered such a semantic change. In what follows, I try to look into this issue from both the semantic and syntactic perspective.

Semantically speaking, it is not entirely unheard of that a linguistic form starts out with only irrealis meanings and is later desemantized to cover realis situations as well. A case in point is the semantic grammaticalization path of the German Infinitive as outlined by Haspelmath (1989). Specifically, *zi*-Infinitive in Old High German was mostly used with irrealis-directive modality (e.g. ‘want to’), and *ze*-Infinitive in Middle High German was expanded to cover irrealis-potential modality (e.g. ‘be able to’), and eventually realis modality much later in the same period, but only limited to collocations with a few verbs. In this period, the Perfect Infinitive, which expresses temporal precedence (e.g. ‘to have done something’), did not exist yet. After the period of Middle High German there was a burgeoning extension of the Infinitive to realis modality, and it was also during this time that the Perfect Infinitive arose. Haspelmath offered an

explanation for this chronological correlation. When the Infinitive expressed only irrealis, its temporal reference was predictable, so there was no need for making finer temporal distinctions, which became important when the Infinitive was widely used with realis modality as well. Suppose Puyuma extended *a*-grade NAF-words from irrealis to realis modality, as was the case for German Infinitive, the new realis function would create a motivation for making finer temporal distinctions. Interestingly, Puyuma does make an aspectual contrast between *Kstem-aw/ay/anay* forms for neutral and punctual event types and α -*Kstem-aw/ay/anay* forms (where α - stands for *Ca*- reduplication or *a*- affixation) for on-going ones (Tsuchida 1980: 217).

As for syntactic developments, both Sagart (2010) and Aldridge (2016) hypothesized there was a period in Puyuma when it had clitic-attracting auxiliaries right before *a*-grade NAF-words. At that time, person-form indexes would be enclitic to auxiliaries (due to their second-position nature), which were lost later on, thus leaving person-form indexes stranded as proclitics to NAF-words. This idea goes back to Starosta et al. (1982), who dubbed the process AUX-axing. To flesh out AUX-axing, affirmative and negative realis-indicative sentences in Rikavung Puyuma are illustrated in (33).

(33) Rikavung Puyuma (H. Jiang & Billings 2015: 97)

- a. **ku=sukun-anay**
 1SG.ACT=push-CF
 ‘I pushed {her/him/them}.’
- b. **'a(zi)=ku sukun-an**
 NEG=1SG.ACT push-CF.DEP
 ‘I didn’t push {her/him/them}.’

If the hypothetical auxiliary ever existed, it would precede the *a*-grade *Kstem-anay* in (33)a and attract the Actor index, much as the negator *'a(zi)* does now in (33)b. Presumably, the hypothetical auxiliary would be semantically hortative/optative such that

it required *a*-grade NAF-words.¹⁴⁰ Accordingly, the *a*-grade *Kstem-anay* would be syntactically dependent on the hypothetical auxiliary, just as the zero-grade *Kstem-an* currently is on the negator. The complex predicate consisting of the hypothetical auxiliary and an *a*-grade NAF-word would then be restricted to express irrealis modality. As the hypothetical auxiliary was lost and person-form indexes switched their direction of attachment (or in the reverse order; see Ross 2015c), *a*-grade NAF-words were generalized to include realis modality and then gradually became the dominant realis-indicative forms in NAF-constructions by replacing the erstwhile indicative Mixed NAF-words.

Although there is no direct evidence for clitic-attracting auxiliaries having ever existed and dictated *a*-grade NAF-words in Puyuma, the possibility of the development outlined above cannot be ruled out, considering that we do find clitic-attracting auxiliaries that dictate *a*-grade NAF-words in Kanakanavu. As has been shown in (31) and (32) above, both the *a*-grade *Kstem-ai* and Mixed *Kstem-ene* in Kanakanavu are used in realis-indicative NAF-constructions with a patientive Topic. According to K. Mei (1992: 227), however, the *a*-grade *Kstem-ai*, but not the Mixed *Kstem-ene*, is required to collocate with the negator *kuu*= ‘never’. This is confirmed by Tsuchida’s (2003) texts, as shown in (34).¹⁴¹

(34) Kanakanavu (Tsuchida 2003: 32; #42)

araanai='inia, **kuu**=kani=cu=kiai **pa-arateken-ai** sua=saruanai
 from=3.OBL NEG=EVID=already=3.ACT CAUS-seperate-LF TOP=man
 ‘Since then she had never let the man leave her.’

¹⁴⁰ To a similar vein, Ross (2013) proposed a hortative/optative preverbal auxiliary that once existed in some pre-Proto Rukai period but is now lost. This is meant to explain why verbs with person-form proclitics have hortative/optative meanings across Rukai dialects whereas those with enclitics have indicative meanings.

¹⁴¹ Everything said in Footnote 139 applies here as well.

The negator in (34) is precisely the kind of hypothetical auxiliaries that is not found in Puyuma. Moreover, two other pieces of circumstantial evidence suggest that compared with the Mixed Kstem-*ene*, the *a*-grade Kstem-*ai* in Kanakanavu is relatively a late comer in expressing realis-indicative meanings. For one thing, the *a*-grade Kstem-*ai* has a lower text frequency than the Mixed Kstem-*ene*, according to D. Liu (2014). For the other, the *a*-grade Kstem-*ai* is more emotionally loaded, conjuring the connotation that the Actor intentionally acts on the Patient-Topic to incur an adversative effect (C. Yang 2015). The intentional reading might be a semantic relic of the original optative/hortative function of *a*-grade NAF-words in PAN.

Despite the lack of evidence for clitic-attracting auxiliaries that were once superordinate to *a*-grade NAF-words in Puyuma, there are still two clues suggesting that their irrealis-optative function might have predated their realis-indicative function. The first one has to do with imperative constructions. According to Tsuchida (1980, 1992a), NAF-words in imperative constructions assume zero-grade NAF-words such as Kstem-*u/i/an* (for PF/LF/CF respectively) in Tamalakaw Puyuma. However, additional data reveal that this is only part of the story. In fact, not only zero-grade but also *a*-grade NAF-words are the required forms in affirmative imperative constructions, with the division of labor determined by the grammatical person of the (non-Actor) Topic. The same pattern has been confirmed by speakers of all non-Nanwang varieties of Puyuma, Tamalakaw included, but to draw relevance to the data presented above Rikavung is illustrated here, as in (35).

(35) Rikavung Puyuma (Fieldnotes)

- a. **{kezeng-u/tapsi'-i/sukun-an}** {intaw/i=misak}
 {pull-PF.IMP/splash.water-LF.IMP/push-CF.IMP} {3.TOP/TOP.SG=M.}
 ‘{Pull/Splash water on/Push} {him/her/them/Misak}!’
- b. **{kezeng-aw/tapsi'-ay/sukun-anay}**=**{ku/mi}**
 {pull-PF.IMP/splash.water-LF.IMP/push-CF.IMP}={1SG.TOP/1EXCL.TOP}
 ‘{Pull/Splash water on/Push} {me/us}!’

Zero-grade and *a*-grade NAF-words are respectively designated for the third-person and first-person Topic. With everything else being equal, interchanging one series of NAF-words with the other would lead to unacceptable sentences, which were immediately rejected by speakers of non-Nanwang Puyuma. In Nanwang Puyuma, by contrast, zero-grade NAF-words are the only acceptable forms in affirmative imperative constructions, irrespective of the grammatical person of the (non-Actor) Topic, as shown in (36) below (see also S. Teng 2008: 217).

(36) Nanwang Puyuma (Fieldnotes)

- a. **{padrek-u/pulrang-i/sulrud-an}**=driya i pilray
 {carry.on.the.back-PF.IMP/help-LF.IMP/push-CF.IMP}=still TOP.SG P.
 ‘{Carry/Help/Push} Pilray!’
- b. **{padrek-u/pulrang-i/sulrud-an}**=ku=driya
 {carry.on.the.back-PF.IMP/help-LF.IMP/push-CF.IMP}=1SG.TOP=still
 ‘{Carry/Help/Push} me!’

Considering that non-Nanwang varieties are morphosyntactically more conservative than Nanwang (S. Teng 2009), it is very likely that Nanwang has innovated by generalizing the zero-grade NAF-words to be the only forms for imperatives and that non-Nanwang varieties preserve the imperative function of the *a*-grade NAF-words (cf. (35)b and (36)b). The structural difference between the imperatives in (35)b and the indicatives in (33)a lies only in the direction of clitic attachment to the *a*-grade NAF-words. Crucially, person-form indexes are enclitic in imperatives, but proclitic in indicatives. Given the overall second-position nature of person-form clitics (see §2.3.3), the imperative function

associated with Puyuma *a*-grade NAF-words hosting enclitics might be original and that the indicative function associated with Puyuma *a*-grade NAF-words hosting proclitics would have been due to the functional expansion from irrealis to realis after AUX-axing, as per the proposal in Sagart (2010) and Aldridge (2016).¹⁴² More importantly, the use of *a*-grade NAF-words in imperative constructions is not specific to Puyuma, but also attested in Siraya (e.g. *uku'-aw* 'Go!'; Adelaar 2013: 221), Kanakanavu (e.g. *kaen-au* 'Eat!'; Tsuchida 1988a: 1208), and Isbukun Bunun (e.g. *astal-av* 'Wait!'; L. Huang 1997: 387). The morphological isomorphism between imperatives and hortatives/optatives, as shown by Puyuma *a*-grade NAF-words, is in line with a widespread typological trend (see van der Auwera et al. 2013) well motivated by their shared directive illocutionary force to expect action as an outcome, as opposed to conveying or soliciting information (Croft 1994).

The second clue for *a*-grade NAF-words in Puyuma having extended their functions from requesting action to passing on information is the structural asymmetry between NAF-words used for main indicative predication and those allowed in argument nominalizations embedded under nominal relation markers. As has been demonstrated in §3.5.2.2, Rikavung Puyuma uses all NAF-words, be it Mixed or Suffixal, for main indicative predication in the non-embedding context, but allows only Mixed and Suffixal zero-grade NAF-words, but not Suffixal *a*-grade ones, in the embedding context (see Table 3.8 above), suggesting that *a*-grade NAF-words entered the territory of main indicative predication relatively late. This is based on the crosslinguistic phenomenon that non-embedded independent structures tend to be more innovative than embedded or

¹⁴² On a related note, Göksel & Kabak (2012) discuss a special construction in Turkish, where imperative verbs are recruited to narrate a past event despite the fact that the language has separate (in fact, many) verb forms for past events.

subordinate ones, and innovation has been shown to take place in such areas as constituent order, grammaticalization, morphological replacement, and morphophonemic changes. The motivation behind this asymmetry, as Bybee (2001) argued, is that non-embedded structures are prone to change because of the much richer pragmatic relations that need to be fulfilled in the main predication while embedded ones hang on to conservatism since they are processed in large chunks. As a result, embedded structures like nominalizations, regardless of how “finite” they may be, are “not just another instance of the main clause ‘S’” (ibid.: 6), a point further expounded by Shibatani (2009). More importantly, the import of this on historical changes is that they would occur earlier in main predications than in embedded ones. If we capitalize on this idea, the functional development of *a*-grade NAF-words in Puyuma could have been as follows. First, they were used for only non-indicatives performing directive speech acts, as they now still are in some constructions of conservative varieties of Puyuma. Then they were recruited into main realis-indicative predication after desemanticization (probably due to loss of optative/hortative auxiliaries), thus in a position to compete with Mixed NAF-words, which are also used for the realis-indicative. Kananavu might have undergone a similar process but to a much lesser degree because only the *a*-grade Kstem-*ai* was affected. Because the new function of *a*-grade NAF-words started out in the innovative main predication, they have not made their way into conservative embedded structures, which explains the restriction against their use in Puyuma argument nominalizations. However, conservative embedded structures might follow suit over time, as is the case in Siraya, where *a*-grade NAF-words did penetrate into argument nominalizations.

Finally, Table 4.2 summarizes the grammatical functions of both *a*-grade and zero-grade NAF-words in Formosan languages where at least one form from either set is attested. Zero-grade NAF-words are either used in affirmative imperatives or immediately after some element that requires them. Accordingly, Ross (2009, 2012) reconstructed the two functions as Imperative and Dependent forms in PAn, which were identical except those for PF. As for *a*-grade NAF-words, they predominantly have such irrealis meanings as optative/hortative, imperative, and future. The realis-indicative function of the Kstem-*ai* in Kanakanavu and all *a*-grade NAF-words in Puyuma are semantic outliers, which might be innovations rather than retentions based on the clues presented in this section.

Table 4.2: Grammatical functions of Suffixal NAF-words in Formosan languages

Languages	Suffixal <i>a</i> -grade	Suffixal zero-grade
Atayal & Seediq	OPT/HORT	AFF-IMP NEG + RLS.IND
Paiwan	OPT/HORT	AFF-IMP NEG + RLS.IND CONJ + RLS.IND
Siraya	OPT/HORT AFF-IMP AFF-IRR.FUT	AFF-IMP
Pazeh	OPT/HORT (-aw < PAn PF *-aw) AFF-IRR.FUT (-ay < PAn LF *-ay)	AFF-IMP (-i < PAn LF *-i)
Kanakanavu	AFF-IMP (-au < PAn PF *-aw) AFF-RLS.IND (-ai < PAn LF *-ay)	AFF-IMP (-i < PAn LF *-i?) ¹⁴³
Nanwang Puyuma	OPT/HORT AFF-RLS.IND	AFF-IMP NEG + RLS.IND
Non-Nanwang Puyuma	OPT/HORT AFF-IMP (TOP=1st person) AFF-RLS.IND	AFF-IMP (TOP=3rd person) NEG + RLS.IND

¹⁴³ K. Mei (1982) only specified Kstem-*au* as the NAF form for (affirmative) imperatives, but both Tsuchida (1988a) and D. Ho (1997) mentioned an additional form Kstem-*i*. However, examples of Kstem-*i* they gave all have a stem-final /a/ vowel, and both of the two *a*-grade forms Kstem-*au* and Kstem-*ai* respectively have Kstem-*u* and Kstem-*i* as their alloforms when the Kstem ends with the /a/ vowel. Thus, it is unclear whether Kstem-*i* actually reflects the *a*-grade *Kstem-*ay* or the zero-grade *Kstem-*i*. If Kanakanavu Kstem-*i* reflects the *a*-grade *Kstem-*ay*, it would resemble non-Nanwang Puyuma by using reflexes of *Kstem-*ay* for both (affirmative) imperatives and realis-indicatives. Finally, recent studies of Kanakanavu (A. Deng 2014; H. Liu 2014; C. Yang 2015) all point out that suffixes -*au* and -*ai* have both undergone sound changes, now rendered as -*oo* (or -*o*) and -*ee* respectively.

4.4.2. Mixed NAF-words

The idea of relative conservatism in embedded structures can also account for syntactic distributions of Mixed NAF-words in Formosan languages regardless of whether or not they belong to the proposed Nuclear subgroup. Along the lines in Sagart (2013), who argued against the Nuclear hypothesis, non-Nuclear languages like Tsou, Rukai, and Puyuma all innovate their main predications in one way or another such that Mixed NAF-words are mostly trapped in embedded structures except in some relic constructions. In extreme scenarios, they either have limited use, such as the reflex of PAn Perfective **<in>* in Rukai dialects (Zeitoun 2007), or are found only in sporadic fossilized forms, such as the reflex of PAn LF *-an in Tsou (Ross 2012).

Moreover, innovations of main predications are not limited to non-Nuclear languages. For instance, Amis, a Nuclear language, has reorganized the distributions of Mixed NAF affixes such that both Conveyance *sa-* (< PAn CF *Sa-) and Locative *-an* (< PAn LF *-an) are now in a syntagmatic rather than paradigmatic relationship with both AF affixes and PF *-en* (< PAn PF *-en). More details will be presented in Chapter 5. Suffice it to say for now that both Patient-Locative Mstem-*an* and Conveyance *sa*-Kstem in Amis are predominantly used as argument nominalizations. Even though these two forms also serve as matrix predicates, they are optionally preceded by the nominal relation marker for underived common nouns functioning as the predicate, and they can only be negated like underived common nouns, suggesting that they are nominal predicates. Also, the perfective-cum-PF *<in>*Kstem in Amis is only found in argument nominalizations but not in main predications (H. Chang 2006: 572). Thus, these three Mixed NAF-words in Amis all predominantly serve as nominalizations by the criteria of

Amis grammar, just as Mixed NAF-words in Puyuma predominantly do by the criteria of Puyuma grammar. However, under the Nuclear hypothesis such a distribution is expected to occur in non-Nuclear languages like Puyuma, but not in Nuclear languages like Amis. In other words, Mixed NAF-words tend to survive in embedded structures such as argument nominalizations, and this trend cuts across the classification of Nuclear vs. non-Nuclear languages. Therefore, the phenomenon can be attributed to the crosslinguistic tendency of conservatism in embedded structures without having to rely on the Nuclear hypothesis.

In the rest of this section, I discuss cognate constructions involving reflexes of PF *-en in particular because the Alternative hypothesis has implications for what PF-words should be reconstructed in PAn in a way that differs from the Nuclear hypothesis.

Ross (2012: 1264) reconstructed PAn Mixed NAF-words as in Table 4.3.

Table 4.3: Reconstructions of Mixed NAF-words in PAn (as per Ross 2012: 1264)

	PF	LF	CF
Neutral	---	*STEM-an	*Sa/Si-STEM
Irrealis	*Ca~STEM-en	*Ca~STEM-an	*Sa/Si-Ca~STEM

The gap in the table, namely, the morphologically unmarked PF *STEM-en, is not reconstructed in PAn, but only in PNAn. This is based on two reasons. For one thing, reflexes of the Neutral *STEM-en are not found in any non-Nuclear language (i.e. Tsou, Rukai, and Puyuma), so it is only reconstructable in PNAn, the ancestor of all Nuclear languages. By contrast, reflexes of the Irrealis *Ca~STEM-en are attested in both Nuclear

and non-Nuclear (specifically Puyuma) languages, so it is reconstructable in PAN. The explanation is that PNAn created the Neutral form *STEM-en out of the Irrealis *Ca~STEM-en based on the analogy of the LF pair *STEM-an and its reduplicated counterpart *Ca~STEM-an, which already existed in PAN. For the other, the Neutral PF *STEM-en is reflected in eleven Formosan languages, including Thao (as STEM-(i)n), Kanakanavu (as STEM-(#)n), Saaroa (as STEM-a), Mayrinax Atayal, Seediq, Ishbukun Bunun (as STEM-un in the last three languages), Saisiyat, Pazeh, Paiwan, Siraya, and Central Amis (as STEM-en [STEM-ən] in the last five languages). Nevertheless, “[o]nly [Pazeh] and Paiwan reflect *STEM-en as a nominalization, suggesting that *STEM-en in PNAn was initially a finite verb form and that the Paiwan and [Pazeh] nominalizations are analogical back-formations” (Ross 2012: 1269). In other words, reflexes of the Neutral PF *STEM-en are highly verbal and do not have the nominalization function, except those in Paiwan and Pazeh, where the new nominalization function has developed through back-formations based on the dual functions (both predicate and argument) of other Focus-words in both languages.

However, the second reason relies heavily on the conclusion that only two out of the eleven Formosan languages where the Neutral PF *STEM-en is retained use the cognate forms as nominalizations. It is unclear how this conclusion was reached, but if we adopt the methodology stated in §4.3 above, a drastically different result will emerge. That is, all the eleven languages except one use reflexes of *STEM-en as argument nominalizations, which is summarized in Table 4.4.

Table 4.4: Nominalization function of reflexes of PF *STEM-en in Formosan languages

Nominalization function	Languages
YES	Thao, Kanakanavu, Saaroa, Mayrinax Atayal, Seediq, Ishbukun Bunun, Saisiyat, Pazeh, Paiwan, Siraya
NO	Central Amis

To support the result, all the languages in the first row of Table 4.4 but Saisiyat are illustrated in (37) through (45), where the nominalization function of a PF-word is compared with an underived noun or pronoun.

(37) Thao¹⁴⁴

- a. tima=[sa **kaytunu-n** lujan]
 who=TOP beat-PF L.ACT
 ‘Who is it that Lujan beat?’ (S. Wang 2004: 169) [cf. Mstem: *k<m>aytunu*]
- b. tima=[sa **izahuy**]
 who=TOP DIST
 ‘Who is that (person)?’ (S. Wang 2004: 296, citing Blust 2003: 990)

(38) Kanakanavu (H. Liu 2014: 39, 80)

- a. iikasu ia, [**vu-un**=maku sinatə]
 2SG.TOP P_{TOP} give-PF=1SG.ACT book.UND
 ‘You are the one I gave the book to.’ [cf. Mstem: *mo-vua*]
- b. iiku ia, **cuma** mu’u
 1SG.TOP P_{TOP} father/uncle M.
 ‘I am Uncle Mu’u.’

(39) Saaroa (C. Pan 2012: 318, 314)

- a. ngahla=isa [**a-rumuk-a**=u um-u]
 what=3.TOP STAT-like-PF=2SG.ACT AF-eat
 ‘What is it that you like to eat?’ [cf. Mstem: *m-a-rumuk*]
- b. pakiaturua=i **ama**=u
 teacher=QP father=2SG.GEN
 ‘Is your father a teacher?’

¹⁴⁴ The PF *-en is reflected as *-in* in Thao, which has *-n* as its allomorph when the stem ends with vowels.

(40) Mayrinax Atayal (Y. Cheng 2013: 105, 7)

- a. **nanuwan**=[ku **vu'-un=nia**]
 what=TOP.DEF shoot-PF=3SG.ACT
 'What is it that {she/he} is shooting?' [cf. Mstem: *s<um>vu*]
- b. **valaiq**=[ku **kisliq=nia**]
 AF.good=TOP.DEF heart=3SG.GEN
 'S/he has a good heart.' (Lit. 'Her/his heart is good.')

(41) Tgdaya Seediq

- a. **maanu ka=[sbet-un na=pawan]**
 what TOP=hit-PF ACT=P.
 'What is it that Pawan will hit?' [cf. Mstem: *s<m>ebuc*] (H. Chang 1997: 43)
- b. **maanu ka=nii**
 what TOP=PROX
 'What is this?' (9-Level Textbooks 1-8)

(42) Isbukun Bunun (C. Shi 2009: 101, 26)

- a. **sa<i>du-an=ku=[a kalat-un=[mas asu]]=a maluspingaz**
 <PST>see-LF=1SG.ACT=TOP bite-PF=ACT dog=LIG woman
 'I saw the woman that the dog bit.' [cf. Mstem: *kalat*]
- b. **ka~kaun-un=ku=[a haising]**
 IPFV~eat-PF=1SG.ACT=TOP cooked.rice
 'I am eating rice.'

(43) Pazeh (P. Li 2000: 92)

- a. [**baked-en** ni sabung] rakihan ka, ma-raxiw lia
 beat-PF ACT S. child PTOF AF-escape ASP
 'The child that Sabung beat has escaped.' [cf. Mstem: *mu-baket*]
- b. **rakihan** rabex ka, iba'-en ni ina
 child little PTOF hold-LF ACT mother
 'The little child {is/was} held by the mother (in her arms).'

(44) Southern (Sinvaudjan) Paiwan (C. Wu 2013: 221, 31)

- a. **anema**=[a **su=umalj-en=a caqis**]
 what=TOP 2SG.ACT=redo-PF=LIG <AF>sew
 'What is it that you sewed again?' [cf. Mstem: *m-umalj*]
- b. **kan-u**=[a **ciqaw**]
 eat-PF.IMP=TOP fish
 'Eat the fish!'

(45) Siraya (Adelaar 2011: 99, 96)

- a. ka-harūm-an ta [ka-muy-ən=au], äsi pa-p'xik-an
 K-compassionate-LF TOP K-wish-PF=1SG.ACT NEG K-sacrify-LF
 'What I wish for is mercy, not sacrifice.' [cf. Mstem: *ma-muy*]
- b. timang ta [Rarenan=au]
 who TOP Mother=1SG.GEN
 'Who is my mother?'

It should be pointed out that Saisiyat presents some challenges to this task. Unlike other Formosan languages, Saisiyat has an extremely flexible word order in NAF-constructions (M. Yeh 2000: 71), and very often dispenses with nominal relations markers. Both properties make it difficult to decide whether a NAF-word is the matrix predicate or embedded in argument nominalizations. Moreover, specialists of Saisiyat (e.g. M. Yeh 2003, 2011) often claim that the PF STEM-*en* does not have a nominalization function, but its irrealis-future counterpart *ka*-STEM-*en* does (where *ka*- is a functional equivalent of *Ca*- reduplication in other languages). To overcome these challenges, I turn to spontaneous data like (46).

(46) Saisiyat¹⁴⁵

- a. ... **mari'-in** noka=ma'iaeh **awpo'-on=ila** 'aehoe'.\
 take-PF ACT=person carry-PF=already dog
 'The person took and carried the dog.'
 (NTU Corpus:Frog 7; F. Hsieh & S. Huang 2006: 94)
- b. ...(0.8) 'akoey 'atomalan=ila [[**kita-en**] ka= .. a takem]
 AF.plentiful very=already see-PF LIG= FIL frog
 '(They) saw many frogs.' (Lit. 'The frogs (they) saw are plentiful.'
 (NTU Corpus:Frog 5; S. Huang 2008: 115)

The two PF-words in (46)a are matrix predicates because they contribute to the main assertion and advance the main storyline. By contrast, the PF-word *kita-en* 'see-PF' in (46)b is part of a complex argument phrase, where it modifies the noun *takem* 'frog'. This

¹⁴⁵ Both *-in* and *-on* are allomorphs of *-en*, a reflex of PF *-en. Besides, the original transcription for 'AF.plentiful' in the cited work is *'akoy*, which is revised to *'akoey* here.

is precisely the same pattern illustrated by the irrealis-future PF-word *ka-hiwa:-en* ‘IRR-cut-PF’ in (47), where it modifies the noun *waliSan* ‘boar’ within a complex argument phrase.

(47) Saisiyat (M. Yeh 2011: 570)

[[**ka-hiwa:-en** rim’an] ka=waliSan] rae’iw=ila
 IRR-cut-PF tomorrow LIG=boar run.away=already
 ‘The boar that is going to be killed tomorrow has run away.’

In both (46)b and (47), the modifier and the modifiee are connected by the attributive ligature *ka*, and the complex argument phrase thus formed is the sole NP of a monadic matrix predicate. Therefore, contrary to the common belief, spontaneous data show that the PF-word *STEM-en* in Saisiyat is in fact used as an argument nominalization just like its irrealis-future counterpart *ka-STEM-en*.

By contrast, among the eleven languages Central Amis is the only one where there is a ban on reflexes of PF **STEM-en* being used as argument nominalizations. The morphologically unmarked PF-word *STEM-en* in Central Amis can be a matrix predicate, but cannot be part of an argument NP, and the latter function has to be fulfilled by its reduplicated counterpart *Ca~STEM-en*, as in (48).

(48) Central Amis (D. Liu 1999: 52)

- a. **cilah-en** aku ku=dateng
 pickle-PF 1SG.ACT TOP.CMN=vegetable
 ‘I will pickle the vegetables.’
- b. u=dateng ku=[{**ca~cilah-en/*cilah-en**} aku]
 CMN=vegetable TOP.CMN=IRR~pickle-PF 1SG.ACT
 ‘What I will pickle is the vegetables.’

It seems to be no coincidence that Amis is also a language where there is a makeover on the way reflexes of PAn Focus affixes are combined with stems (see Chapter 5). Thus, it is reasonable to believe that the exception in Amis is a later development resulting from

innovations in main predications. As a result, only the reduplicated form *Ca~STEM-en* is preserved in argument nominalizations.

Crucially, the new result in Table 4.4 suggests that the Neutral PF-word **STEM-en* in some proto-language had both predicate and argument functions, as its reflexes in modern languages predominantly do. Otherwise analogical back-formations as envisioned by Ross (2012) would have to take place in all the ten languages where reflexes of PF **STEM-en* can be nominalizations, which is very unlikely. Moreover, the present result aligns well with the Alternative hypothesis, and prompts us to rethink in which period the PF **STEM-en* is reconstructable. Reflexes of the Neutral **STEM-en* are not found in any of the three non-Nuclear languages (i.e. Tsou, Rukai, and Puyuma). The complete absence can be explained in terms of loss after the three languages independently innovated their main predications, as per Sagart (2013). On the other hand, within non-Nuclear languages reflexes of the Irrealis PF-word **Ca~STEM-en* are only found in Puyuma, and only as argument nominalizations,¹⁴⁶ which bears some resemblance to the situation in Amis, where only *Ca~STEM-en* but not *STEM-en* can function as a nominalization. This once again attests to the conservatism in embedded structures, in Puyuma and Amis alike. Therefore, the earliest proto-language that had Neutral PF-word **STEM-en* could have been PAn, instead of the proposed PNAn. An advantage of this is that we would not have to reconstruct a morphologically marked form (i.e. **Ca~STEM-en*) without its morphologically unmarked counterpart (i.e. **STEM-en*) in PAn.

¹⁴⁶ Tsuchida (1980: 202) listed the PF form *ka~kerutr-en* ‘IRR~dig-PF’ (or <ka-keRuT-an> in his rendition) in Tamalakaw Puyuma. No sentential examples of this form were provided, but presumably it would mean ‘what is to be dug’.

4.5. Chapter summary

The traditional explanation for the dual predicate-cum-argument function of Mixed NAF-words (i.e. those marked by reflexes of *-en, *-an, *Sa/Si-, and *<in>) involves the nominalization-into-verb reanalysis whereby erstwhile nominal predicates were reanalyzed into indicative verbs either in PAn or in a later period called PNAn according to the Nuclear hypothesis. By contrast, Suffixal NAF-words (i.e. those marked by reflexes of *-a/u/i/ani in the zero-grade series and *-aw/ay/anay in the *a*-grade series) have long been considered highly verbal and do not have much to do with nominalization. This chapter has attempted to revive the Alternative hypothesis in Ross (2002), which dispenses with the nominalization-into-verb reanalysis, and to address some potential questions that emerge from espousing this very hypothesis.

In brief, argument nominalizations in PAn, as envisioned here, would employ the gap strategy and involve not only the AF Mstem and Mixed NAF-words, but also Suffixal zero-grade NAF-words. By contrast, Suffixal *a*-grade NAF-words in PAn would be prohibited from argument nominalizations because their primary function was to call for action (i.e. hortative/optative/imperative) rather than to convey information (i.e. indicative), a property inherited by many modern languages. However, some languages innovated by losing Suffixal *a*-grade NAF-words completely (as in Tsou) or by recruiting them for non-directive functions, starting from main predications (as in Puyuma, and to a much lesser degree, Kanakanavu) and eventually to embedded structures like argument nominalizations (as in Siraya).

|Chapter 5

Verbal-based Nominalization III: A Case Study of Central Amis

This chapter looks into verbal-based nominalizations in Central Amis, spoken mostly on the eastern coast of Taiwan. I supplement, revise, and rectify some previous studies on this topic by systemically examining the verbal and/or nominal properties of various Focus-words, all of which can be morphologically defined.

The Amis language is a dialect cluster, with interrelated varieties that are mostly mutually intelligible. The variations across dialects are to a large extent lexical and phonological in nature. A commonly cited work addressing the internal classification of the Amis language is Tsuchida (1988b), where five dialects were identified, though not much evidence was presented to support the claim. His five dialects from the north to the south are Sakizaya, Northern (or Nanshi) Amis, Tavalong-Vataan Amis, Central Amis (including Coastal and Xiuguluan Amis but excluding Tavalong-Vataan Amis), and

finally Southern Amis (including Beinan and Hengchun Amis).¹⁴⁷ Among them, Sakizaya is considered to be the most “divergent” one (see Tsukida 1993). In this chapter I focus mostly on Central (Coastal) Amis, which is probably the most researched variety.¹⁴⁸

The literature on Amis is copious. Works addressing the grammatical system in general or the verbal system in particular include R. He et al. (1986), T. Chen (1987), Y. Huang (1988), Tsuchida (1988b), Z. Yan (Yan 1992), M. Yang (2005), Tsukida (2008), J. Wu (1996a, 2000, 2006), S. Zeng (1991, 2002), to name just a few. Those that devote much attention to verbal-based nominalization in Amis are J. Wu (1996b, 2003, 2007), M. Lin (1995), D. Liu (1999), E. Liu (2003), M. Chang (2007), and L. Jiang (2009, 2012). Given the abundance of the Amis literature, I shall cite examples from the available treasure trove wherever possible. Only in cases where there are data gaps do I supplement additional examples that I elicited from my Amis language consultants, who are specifically from Donghsin (or *Malalo'on* in the vernacular), Hualian County and Ningpu (or *Kinanoka* in the vernacular), Taitung County.

The organization of this chapter is as follows. First, §5.1 presents an overview of Amis verb forms, and summarizes some syntactic asymmetries among them as observed in the literature. Then §5.2 and §5.3 respectively deal with argument and event nominalization, after which §5.4 draws generalizations over results from the previous two sections and then reevaluates some claims made in previous studies. Finally, §5.5 briefly recapitulates the major points made in this chapter.

¹⁴⁷ Tavalong-Vataan is named after two villages located in the geographical area of Xiuguluan Mountain and River.

¹⁴⁸ I simply use Central Amis to mean Central Coastal Amis unless there is a need to distinguish Coastal from Xiuguluan Amis.

5.1. Overview of Amis verb forms

This section presents an overview of Amis verb forms by not only summarizing observations made in previous studies (mostly D. Liu 1999, J. Wu 2006/2007, and Tsukida 2008), but also presenting some additional data to support the way the language is described here, which differs from the cited works in some respects.¹⁴⁹

5.1.1. Classes of AF verbs

Table 5.1 outlines three classes of AF verbs in Central Amis, with each definable by the morphological alternations between the Mstem and Kstem.

Table 5.1: AF verb classes in Central Amis

	I	II		III	
Mstem	<i>mi-palu</i> 'AF.EXT-beat'	<i>ma-ulah</i> 'AF.INT-like'	<i>k<um>aen</i> 'AF.<UM>eat'	<i>ci-wawa</i> 'AF.have-child'	<i>fangcal</i> 'AF.good'
Kstem	<i>pi-palu</i> 'K.EXT-beat'	<i>ka-ulah</i> 'K.INT-like'	<i>ka-k<um>aen</i> 'K-<UM>eat'	<i>ka-ci-wawa</i> 'K-have-child'	<i>ka-fangcal</i> 'K-good'

The morphological alternations are between *mi*-ROOT and *pi*-ROOT in Class I, and between *ma*-ROOT and *ka*-ROOT in Class II. By contrast, Class III is negatively defined, encompassing any verb forms whose Mstem is neither *mi*-ROOT nor *ma*-ROOT, and its Kstem is formed by prefixing *ka*- to whatever form its Mstem may take, be it unaffixed

¹⁴⁹ Unlike D. Liu (1999) and J. Wu (2006, 2007), Tsukida (2008) is a description of Fata'an (or Vataan) Amis rather than Central Amis. That being the case, most of the observations in Tsukida (2008) still hold true in Central Amis.

roots (e.g. *fangcal* ‘AF.good’) or denominal stems (e.g. *ci-wawa* ‘AF.have-child’).¹⁵⁰ Notice that the Mstem <um>ROOT in Class III behaves quite differently from *mi*-ROOT or *ma*-ROOT in the other two classes. The fact that the Kstem for <um>ROOT is created by adding *ka-* on top of it just like that for unaffixed roots, rather than alternating the AF infix <um> with something, suggests that the <um> morpheme might have become a frozen part of the root. This is also partly supported by the fact that verbs in the form of <um>ROOT are rather limited in number (Tsukida 2008). Due to this, the <um> infix is not given any functional label and is simply glossed based on its exponent form.

The Mstem in Amis is the predicative form in affirmative realis-indicative sentences whereas the Kstem is used in various morphosyntactic contexts, including imperative and negative constructions as well as event nominalizations (see §5.3).¹⁵¹ The two stems are illustrated by the minimal pair in (1), which also shows that Amis is a language with asymmetric standard negation as defined by Milestamo (2013).

¹⁵⁰ Some unaffixed roots, mostly those expressing states, can either be marked by the Introvert *ma-* or dispense with it, thus making them potentially fall under Class II or III. Where such an option applies, according to Tsukida (2008: 287), the *ma*-marked version expresses “temporary states implying that the state was caused through the effect of some other entity” whereas the unaffixed one expresses “inherent or permanent states.” The prefix *ci-* combines quite productively with various nouns to produce verbs whose semantics can be schematically represented as “to have N as a prominent feature of something or somebody”, where N stands for the referent of the incorporated noun. See S. Zeng (1991: 27) for more examples.

¹⁵¹ The Kstem in Amis is referred to as the Connegative form by Tsukida (2008), which is presumably due to its function in negative constructions. But since this form is also used in non-negative contexts, the functionally neutral term Kstem from Ross (2015c) is adopted here instead.

(1) Central Amis¹⁵²

- a. **mi-palu** ci=kilang ci=canglah-an inacila
 AF.EXT-beat TOP.PSN=K. PSN=C.-UND yesterday
 ‘Kilang beat Canglah yesterday.’ (J. Wu 2006: 119)
- b. **caay pi-palu** ci=kilang ci=canglah-an inacila
 NEG K.EXT-beat TOP.PSN=K. PSN=C.-UND yesterday
 ‘Kilang didn’t beat Canglah yesterday.’ (Fieldnotes)

While verbs in Class I are only found in AF-constructions, those in Class II are often said to split into AF and PF subtypes so that there are AF *ma-* and PF *ma-* (D. Liu 1999, 2011; J. Wu 2006, 2007).¹⁵³ The motivation is based on argument marking patterns, whereby AF *ma-* selects its agent-like argument as the Topic NP and encodes its patient-like argument as a non-Topic NP, just as typical AF verbs do in (1), whereas PF *ma-* selects its patient-like argument as the Topic NP and marks its agent-like argument as a non-Topic NP, much as typical PF verbs do (see (13) below for a minimal pair). The contrast is shown in (2), where the Topic NP is underlined.

(2) Central Amis (J. Wu 2007: 131, 134)¹⁵⁴

- a. **ma-ulah** ci=sawmah ci=panay-an
 AF.INT-like TOP.PSN=S. PSN=P.-UND
 ‘Sawmah likes Panay.’
- b. **ma-palu** ni=sawmah ci=mayaw
 AF.INT-beat ACT.PSN=S. TOP.PSN=M.
 ‘Mayaw got beaten by Sawmah.’

¹⁵² Based on the research results in Edmondson et al. (2005), this study does not transcribe the glottal stop in Amis, which predictably occurs in the prevocalic, intervocalic, and postvocalic position. Thus, the negator is transcribed as <caay> [tsa.ʔaj], instead of <ca’ay>, as is commonly found in the literature. The apostrophe <’>, however, is only used in this study to transcribe the epiglottal plosive /ʔ/, “which is realized as [ʔ] initially and medially, and as [ʔʰ] finally” (ibid.: 384). One such example is <faʔinayan> [faʔinajan] ‘man’. Finally, although both [o] and [u] are written as <o> in the Amis conventional orthography, they are transcribed here as <o> and <u> respectively, following J. Wu (2006, 2007).

¹⁵³ The cited works used terms like Actor Voice (AV) and Patient/Undergoer Voice (PV/UV), which I take to be terminological variants of Actor Focus (AF) and Patient Focus (PF) respectively.

¹⁵⁴ Throughout this chapter, Amis sentences without explicit tense/aspect markers or temporal adverbials are translated based on a scenario that the consultants found most appropriate to a particular state of affairs. In many cases, there are more than one possible temporal interpretations. See Zeitoun et al. (1996) for details on TAM marking in Amis and other Formosan languages.

However, such a contrast can be given a semantic account, as has been done by Tsuchida (2008), without resorting to two /ma/ morphemes. Since the way *ma*-verbs are described here deviates from the convention in the literature, some justifications are in order below.¹⁵⁵

Despite the impression of Table 5.1, which may suggest otherwise, roots do not fall neatly into those affixable by *mi*- and others by *ma*- (i.e. either Class I or II). In fact, a great number of roots, regardless of whether they express actions or states, are compatible with both prefixes. Consequently, two roots in combination with two Mstem prefixes potentially give rise to four word forms, and the number is doubled if their corresponding Kstems are also considered, as illustrated by the action roots *palu* ‘beat’ and the state root *ulah* ‘like’ in Table 5.2.¹⁵⁶

Table 5.2: Class I and II verbs in Central Amis

	Action root (e.g. <i>palu</i> ‘beat’)	State root (e.g. <i>ulah</i> ‘like’)
Class I Extrovert	Mstem: <i>mi-palu</i> Kstem: <i>pi-palu</i>	Mstem: <i>mi-ulah</i> Kstem: <i>pi-ulah</i>
Class II Introvert	Mstem: <i>ma-palu</i> Kstem: <i>ka-palu</i>	Mstem: <i>ma-ulah</i> Kstem: <i>ka-ulah</i>

More importantly, the different prefixes in Class I and II are highly correlated with a robust semantic contrast, termed *Extrovert* vs. *Introvert* here (see also M. Yang 2005:

¹⁵⁵ The two verbs *ma-ulah* and *ma-palu* in (2) will be taken as the representative examples of so-called AF and PF *ma*-verbs respectively throughout the discussion.

¹⁵⁶ Not all roots are equally compatible with the prefix *mi*- and *ma*-, and the compatibility is, not surprisingly, subject to the semantics of the root in question.

8)¹⁵⁷. The contrast is by and large in line with of Kemmer’s (1993: 135) characterization of active vs. middle voice across languages: “active intransitive forms imply more volitional and purposeful activity on the part of the Initiator, while middle-marked verbs in this class emphasize the affectedness of the Experiencer.” For instance, the Introvert state predicate *ma-ulah* in (2)a simply asserts the Actor’s (i.e. Sawmah) affection towards someone (i.e. Panay), and the “liker” is not doing anything in the real world to show that. By contrast, its Extrovert counterpart *mi-ulah* asserts that the Actor (i.e. Aki) does something in the real world to gain the affection of someone (i.e. Dongi, a female name), as illustrated in (3).

(3) Central Amis (J. Wu 2006: 293)¹⁵⁸

mi-ulah ci=aki ci=dongi-an
 AF.EXT-like TOP.PSN=A. PSN=D.-UND
 ‘Aki is going to court Dongi.’

Therefore, *mi*-verbs across root types are Extrovert, selecting as their Topic NP participants that exert certain force-dynamics beyond their personal sphere of influence, or the Initiator in Kemmer’s term. That is why *mi*-verbs only participate in AF-constructions, as in (1)a and (3). On the other hand, *ma*-verbs across root types are Introvert, meaning their Topic NP denotes participants that either do not exert force-dynamics at all or only keep it within their personal sphere of influence, or the Experiencer in Kemmer’s term. In other words, while the Topic NP of Extrovert *mi*-verbs

¹⁵⁷ The two semantic terms are borrowed from Schuessler (2007: 38), where he wrote “[i]n *introvert* words, the action is directed toward the subject, or happens to or within the subject...; in *extrovert* words the action originates in or with the subject and is directed out and away to a necessarily external object.” It was also pointed out that due to this semantic contrast Introvert words tend to be monadic whereas Extrovert ones are typically dyadic and often involve causative meanings, both of which are true in Amis. For monadic Introvert verbs, see (5). And the contrast between Introvert non-causative and Extrovert causative action is illustrated by *ma-ruhum* ‘(of plants) become ripe’ and *mi-ruhum* ‘(of people) ripen (plants)’ (J. Wu 2006: 211).

¹⁵⁸ The Extrovert verb *mi-ulah* expresses the action of expressing affection for someone or simply courting someone (hence the free translation) while its Introvert counterpart *ma-ulah* indicates the mental state of having affection towards someone.

is the locus of control, that of Introvert *ma*-verbs is not, a contrast termed agentive semantics versus undergoing semantics by Klaiman (1991: 111). Thus, when state-denoting roots like *ulah* ‘like’ are prefixed by the Introvert *ma*-, the result is a predicate that selects as the Topic NP a participant undergoing the state expressed by the root, as in (2)a. However, when action roots like *palu* ‘beat’ are involved instead, the result is a predicate that selects as the Topic NP a participant undergoing the action expressed by the root, hence a patient-like argument, as in (2)b. In fact, the prefix *ma*- by itself does not really tell whether we are dealing with AF *ma*- or PF *ma*-. Rather, it is the Introvert nature of *ma*-, in conjunction with the semantics of the roots, that renders the split between so-called AF-like and PF-like constructions in (2).

Although the contrast between Extrovert *mi*-verbs and Introvert *ma*-verbs as characterized above deviates from J. Wu’s (2006) treatment of them, it is quite compatible with her semantic generalization of the two groups. While *mi*-verbs portray agentive activities with motional or purposive reading, *ma*-verbs typically express abilitative, emotional, spontaneous, involuntary, or adversative events. For instance, while the Extrovert form *mi-palu* ‘AF.EXT-beat’ in (1)a implies the existence of an intentional/volitional agent, its Introvert counterpart *ma-palu* ‘AF.INT-beat’ in (2)b does not. A comparable semantic contrast is also documented in Cebuano, where AF-constructions with intentionality is marked by *ni*- and those without by *naka*-, as in (4) (respectively called active and inactive affixes in the original source).

(4) Cebuano (Shibatani 1988: 104-105)

- a. **ni-hiwa** si maria ug mangga sa kutsilyo
 AF-cut TOP M. UND mango OBL knife
 ‘Maria cut a mango with a knife.’
- b. **naka-hiwa** si maria sa kutsilyo
 AF-cut TOP M. OBL knife
 ‘Maria got cut on the knife.’

Moreover, a unified semantic account of the Introvert *ma-* also helps to do away with positing unnecessary labels as well as switches among them. In addition to AF *ma-* and PF *ma-*, J. Wu (2006, 2007) proposed Neutral *ma-*, which is neither AF nor PF, for monadic predicates such as those in (5), where she glossed *ma-* as NEUT for Neutral.

(5) Central Amis (J. Wu 2007: 136, 131)

- a. **ma-patay** ku=ra fafuy
 AF.INT-die TOP.CMN=DIST pig
 ‘That pig is dead.’
- b. **ma-ruhem=tu** ku=pawli
 AF.INT-ripen=already TOP.CMN=banana
 ‘The bananas have ripend.’
- c. **ma-laluk** ci=sawmah
 AF.INT-diligent TOP.PSN=S.
 ‘Sawmah is diligent.’

Having designated three labels for *ma-*, J. Wu (2007: 136) did point out that “the same root may appear with more than one *ma-*. Hence, there may be ambiguity for a *ma-* verb in terms of the verb type if no contextual information is provided.” One example given is (6), where *ma-* was considered to be PF, unlike Neutral *ma-* in (5).

(6) Central Amis (J. Wu 2007: 136)

- ma-patay** ni=aki ku=ra fafuy
 AF.INT-die ACT.PSN=A. TOP.CMN=DIST pig
 ‘That pig was killed by Aki.’

In both (5)a and (6), however, the same root *patay* ‘die’ is involved. Admittedly her motivation for assigning PF *ma-* instead of Neutral *ma-* to the Focus-word in (6) was

based on the presence of the additional agentive argument. But even when argument marking patterns are taken into account, the category assignment is not very consistent. Compare, for instance, the examples in (7), where the prefix *ma-* was analyzed by J. Wu (2006: 191) to be Neutral and PF respectively, despite the fact that both involves monadic predicates.

(7) Central Amis (J. Wu 2006: 191)

- a. *ma-kuhting* *ku=pising* *nira*
 AF.INT-black TOP.CMN=face 3SG.GEN
 ‘{His/Her} face got black.’
- b. *ma-faedet* *ku=ni* *a dateng*
 AF.INT-hot TOP.CMN=PROX LIG vegetable
 ‘This dish got heated up.’

Alternatively, one could adopt a semantically unified analysis of Introvert *ma-*, which then dispenses with the switch between the so-called Neutral *ma-* and PF *ma-*, both of which are in sharp contrast to Extrovert *mi-*, as shown in (8), where the Actor (underlined) executes some action to bring about the death of something.

(8) Central Amis (J. Wu 2007: 125)

- mi-patay* *ku=matuasay* *tu=oner*
 AF.EXT-die TOP.CMN=elder UND.CMN=snake
 ‘The elder is {killing/going to kill} a snake.’

In addition to the switch between Neutral *ma-* and PF *ma-*, J. Wu (2006, 2007) identified the switch between AF *ma-* and PF *ma-*, which is only limited to psych-predicates, as contrasted in (9), where the Topic NP is underlined.

(9) Central Amis (J. Wu 2006: 220)

a. **ma-ulah** kaku ci=panay-an
 AF.INT-like 1SG.TOP PSN=P.-UND
 ‘I like Panay.’

b. **ma-ka-ulah** aku ci=panay
 AF.INT-K.INT-like 1SG.ACT TOP.PSN=P.
 ‘I like Panay.’ OR ‘Panay is liked by me.’

While *ma-* in the first example was said to be AF, that in the second was considered PF, which is again based on argument marking patterns. However, such a contrast can be explained by semantics. Basic Introvert psych-predicates like *ma-ulah* select as their Topic NP participants who are the source of the psych-state expressed by the root. In the current example, corresponding to the Mstem *ma-ulah* is its Kstem *ka-ulah*, which is the base from which the more complex Mstem *ma-ka-ulah* is derived. Secondarily derived Introvert psych-predicates like *ma-ka-ulah* are doubly Introvert and thus select as their Topic NP participants who are the target of the psych-state expressed by the root. By contrast, Introvert action predicates like *ma-palu* ‘AF.INT-beat’ already select patient-like arguments as their Topic NP (see (2)b), so it is not possible to make them even more Introvert, thus creating lexical gaps like **ma-ka-palu*, the infelicity of which is expected under the semantic account.

Finally, aside from semantics, there are also grammatical motivations to treat the prefix *ma-* as one coherent grammatical entity. There are various morphological operations targeting at *ma-*verbs in general regardless of whether they are labeled as AF, PF, or Neutral on the ground of argument marking patterns (see also Tsukida 2008: 289). In the first one, verbs of the form *ma-ROOT* for the Mstem always alternate with those of the form *ka-ROOT* for the Kstem. For instance, the negative counterparts of (2), where the Mstem is used, would be (10), where the Kstem is required.

(10) Central Amis (Fieldnotes)

- a. **caay** **ka-ulah** ci=sawmah ci=panay-an
 NEG K.INT-like TOP.PSN=S. PSN=P.-UND
 ‘Sawmah doesn’t like Panay.’
- b. **caay** **ka-palu** ni=sawmah ci=mayaw
 NEG K.INT-beat ACT.PSN=S. TOP.PSN=M.
 ‘Sawmah didn’t beat Mayaw.’

Thus, there is just one single Introvert *ma-* alternating with *ka-*.

Another morphological operation targets at not only Introvert *ma-* verbs (i.e. the Mstem in Class II) but also the Mstem in all the three classes in Table 5.1. The result of this morphological process is referred to as the Factual form by J. Wu (2006, 2007) or the Relative form by Tsukida (2008). More importantly, this morphological operation has syntactic consequences, and will be further investigated in §5.2.1 under Actor nominalizations.

Other morphological operations that target at the Mstem/Kstem in general involve Conveyance *sa-* and Locative *-an*, detailed discussions of which are deferred to §5.2.2.

5.1.2. Verb forms other than AF

In addition to AF verb forms, Amis has others marked by Patient *-en*, Conveyance *sa-*, and Locative *-an*, all of which are typical reflexes of PAn Mixed NAF affixes. The last two are traditionally analyzed as Conveyance Focus (CF) and Locative Focus (LF) affixes respectively, on a par with AF and PF affixes (Z. Yan 1992; J. Wu 1995, 2000; D. Liu 1999, 2011; E. Liu 2003). However, J. Wu (2006, 2007) has argued quite convincingly that unlike their cognates in other Formosan languages, Conveyance *sa-* and Locative *-an* in Amis are better treated as applicative markers that are in syntagmatic, rather than paradigmatic, relationship with either AF or PF affixes. For instance,

Conveyance *sa-* co-occurs with the PF *-en* in (11), where the same word form would be marked by two Focus affixes if Conveyance *sa-* were a CF marker.¹⁵⁹

(11) Central Amis (J. Wu 2007: 110)

aka **sa-pi-litek-en** ku=ra caklis tu=ra kilang
 NEG.IMP CA-K-chop-PF TOP.CMN=MED ax UND.CMN=MED tree
 ‘Don’t chop that tree with that ax!’

Thus, following J. Wu’s (2006, 2007) analysis, I gloss Conveyance *sa-* and Locative *-an* as CA (for Conveyance Applicative) and LA (for Locative Applicative) respectively.¹⁶⁰

However, although these two markers may be applicative at the morphemic level, they accomplish more than what applicative markers typically do at the syntactic level (Shibatani, p.c.). Specifically, Amis word forms containing Conveyance *sa-* still select the instrument or beneficiary as its Topic NP and those containing Locative *-an* still select the location or affected theme as its Topic NP, much as do CF/LF-words in other Formosan languages, where CF/LF affixes do not co-occur with AF/PF affixes. Compare (11), for instance, with (12) from Mayrinax Atayal.

(12) Mayrinax Atayal (L. Huang 2000a: 127)

kaa **ras-ani**=[cu qusia]=[’i watan]
 NEG.IMP bring-CF.DEP=UND water=TOP W.
 ‘Don’t bring Watan water!’

Irrespective of their morphological makeup, Focus-words in (11) and (12) both syntactically opt for the Conveyance Topic (instrument in the former and beneficiary in the latter). In other words, the Conveyance form *sa-pi-litek-en* in Amis, despite its morphological complexity, accomplishes the same syntactic effect as the CF-word *ras-ani* in Mayrinax Atayal. Thus, even in Amis we can still speak of CF- and LF-words,

¹⁵⁹ There is another morpheme *-en* in Amis, which apparently has semantics and syntax different from the PF *-en* (J. Wu 2006: 181). In this study, only the PF *-en* is discussed.

¹⁶⁰ Conveyance Applicative here corresponds to J. Wu’s (2006, 2007) Instrument Applicative.

referring to those that respectively single out Conveyance and Locative as the macro-role for the Topic NP.

Taking the action root *adup* ‘hunt’ for instance, Example (13) illustrates six word forms based on it: the Extrovert Mstem (Class I AF-word), the Introvert Mstem (Class II AF-word), the root suffixed by Patient *-en* (PF-word), the Kstem prefixed by Conveyance *sa-* (CF-word), the Mstem suffixed by Locative *-an* (Mstem-based LF-word), and finally the Kstem suffixed by Locative-*an* (Kstem-based LF-word).

(13) Central Amis (J. Wu 2007: 118, 98)

- a. **mi-adup** ci=aki tu=fafuy nu=lutuk
 AF.EXT-hunt TOP.PSN=A. UND.CMN=pig GEN.CMN=mountain
 ‘Aki is {hunting/going to hunt} wild boars.’
- b. **ma-adup** ni=aki ku=ra fafuy nu=lutuk
 AF.INT-hunt ACT.PSN=A. TOP.CMN=DIST pig GEN.CMN=mountain
 ‘Aki hunted that wild boar.’
- c. **adup-en** ni=aki ku=ra fafuy nu=lutuk
 hunt-PF ACT.PSN=A. TOP.CMN=DIST pig GEN.CMN=mountain
 ‘Aki will hunt that wild boar (for sure).’
- d. **sa-pi-adup** ni=mama tu=fafuy nu=lutuk
 CA-K.EXT-hunt ACT.PSN=father UND.CMN=pig GEN.CMN=mountain
 ku=iduc
 TOP.CMN=spear
 ‘Father {hunts/hunted} wild boars with the spear.’
- e. **mi-adup-an** ni=mama ku=fafuy nu=lutuk
 AF.EXT-hunt-LA ACT.PSN=father TOP.CMN=pig GEN.CMN=mountain
 ‘Father {hunts/hunted} the wild boar.’
- f. **pi-adup-an** ni=mama tu=fafuy ku=ni a lutuk
 K.EXT-hunt-LA ACT.PSN=father UND.CMN=pig TOP.CMN=PROX LIG mountain
 ‘Father {hunts/hunted} wild boars on this mountain.’

In all these examples, the Focus-words serve as the matrix predicates of the Topic NPs (underlined).

Given the fact that all the Focus-words containing the root *adup* ‘hunt’ in (13) are predicated of nominal phrases with certain thematic roles, one crucial question to be

asked is whether the very same forms can also serve as denoting expressions and take up typical argument positions on the syntactic level as underived nouns do. In other words, can these word forms serve as argument nominalizations? And if not, what morphosyntactic adjustments would have to be made in order to make that happen. These questions will be answered in §5.2 on argument nominalizations.

Interestingly, the Introvert AF *ma-adup* and PF *adup-en* both select the hunted animal as the Topic NP, but they differ in terms of “agency and unmarked temporal reading” (J. Wu 2006: 110). While *ma-adup* depicts a past event by default and does not necessarily imply the existence of an intentional agent, *adup-en* describes a future event by default and always implies the existence of an intentional agent. The apparently identical pattern of argument marking for verbs like *ma-adup* in (13)b and *adup-en* in (13)c, whereby the agentive argument is marked by *ni* (as well as its paradigmatic counterparts), is the rationale for previous studies to distinguish PF *ma*-verbs from AF ones (see also the contrast in (2) above). However, as already suggested by the low agency of *ni*-phrases in construction with so-called PF *ma*-verbs, *ma*-verbs in general, whether called AF or PF, share more grammatical properties with *mi*-verbs, which are uncontroversial AF-words across the board, than with genuine PF-words in the form of *ROOT-en*. This will become clearer as we investigate how both *ma*-verbs and *mi*-verbs are nominalized in §5.2.1 on Actor nominalizations.

In addition, while Conveyance *sa-* is only prefixed to the Kstem, Locative *-an* is suffixed to either the Mstem or Kstem. In (13), *mi-adup-an* is built from the Mstem *mi-adup* and selects the hunted animal as the Topic NP. By contrast, *pi-adup-an* is constructed from the Kstem *pi-adup* and selects the hunting location as the Topic NP

instead. Additional treatment of CF- and LF-words will be presented in §5.2.2 on Circumstantial nominalizations.

Finally, the syntagmatic requirements on affirmative and negative predication are good indicators for distinguishing a verbal predicate from a nominal one. In contrast to the verbal pattern that has been demonstrated so far, common nouns for affirmative predication are preceded by *u* whereas those for negative predication by *ku* in addition to the negator, as in (14).

(14) Central Amis (M. Chang 2007: 79, 80)

- a. **u=amis** ci=mita
 CMN=Amis TOP.PSN=M.
 ‘Mita is Amis.’
- b. **caay** **ku=amis** ci=mita
 NEG TOP.CMN=Amis TOP.PSN=M.
 ‘Mita is not Amis.’
- c. **u=wawa** ni=putal ci=mita
 CMN=child GEN.PSN=P. TOP.PSN=M.
 ‘Mita is Putal’s child.’
- d. **caay** **ku=wawa** ni=putal ci=mita
 NEG TOP.CMN=child GEN.PSN=P. TOP.PSN=M.
 ‘Mita is not Putal’s child.’

Patterns like those in (14) will be referred to as nominal predication. It is on the basis of these two predication patterns that the nominal or/and verbal nature of various Focus-words in Amis have been determined in previous studies (D. Liu 1999: 43; J. Wu 2007: 108; M. Chang 2007: 111), some observations of which are summarized in the next section.

5.1.3. Asymmetries between AF/PF and CF/LF

Morphosyntactic asymmetries between AF/PF-words on the one hand and CF/LF-words on the other have long been observed. Some particularly relevant studies are reviewed here before I explore this topic further in subsequent sections.

First, D. Liu (1999: 47-57) concluded that AF/PF-words are always verbal because they have the predicate function, but no argument function whereas CF/LF-words are both verbal and nominal because they have both functions. Specifically, the AF Mstem cannot serve as an argument unless it is suffixed by *-ay* (glossed as NMLZ, which will be further justified in §5.2.1.1). This is so regardless of which class the Mstem belongs to. Class I Extrovert Mstem, Class II Introvert Mstem (including either the so-called AF or PF *ma-*), and Class III Mstem are demonstrated in (15) through (17) in that order.¹⁶¹

(15) Central Amis (D. Liu 1999: 48)

- a. **mi-kalat** ku=ra oner tu=wacu
 AF.EXT-bite TOP.CMN=DIST snake UND.CMN=dog
 ‘That snake is biting a dog.’
- b. u=ra oner ku={**mi-kalat-ay**/***mi-kalat**} tu=wacu
 CMN=DIST snake TOP.CMN=AF.EXT-bite-NMLZ UND.CMN=dog
 ‘What is biting a dog is that snake.’

¹⁶¹ Transcriptions of <l> (for [l]) and <r> (for [r]) in D. Liu (1999) are sporadically messed up, as found in words like *oner* [ʔonər] ‘snake’ in (15) (originally transcribed as <’onel>), *tangal* [taŋal] ‘head’ in (17) (originally transcribed as <taNar>), and *cilah* [tɛilah] ‘pickle, salt’ in (18) below (originally transcribed as <cirah>). These typos have been corrected in the data presented here.

(16) Central Amis (D. Liu 1999: 48, 52)

- a. **ma-futi'** ku=ra kulung i=facal
 AF.INT-sleep TOP.CMN=DIST water.buffalo LOC=paddy.field
 'That water buffalo is sleeping in the paddy field.'
- b. u=ra kulung ku={**ma-futi'-ay/*ma-futi'**} i=facal
 CMN=DIST water.buffalo TOP.CMN=AF.INT-sleep-NMLZ LOC=paddy.field
 'What is sleeping in the paddy field is that water buffalo.'
- c. **ma-kalat** nu=wacu kaku
 AF.INT-bite ACT.CMN=dog 1SG.TOP
 'The dog bit me.'
- d. kaku ku={**ma-kalat-ay/*ma-kalat**} nu=wacu
 1SG.TOP TOP.CMN=AF.INT-bite-NMLZ ACT.CMN=dog
 'The one bitten by the dog is me.'

(17) Central Amis (D. Liu 1999: 48-49)

- a. **ci-tangkal** ku=ra salu
 AF.have-head TOP.CMN=DIST monkey
 'That monkey is clever (lit. have brains).'
- b. u=ra salu ku={**ci-tangkal-ay/*ci-tangkal**}
 CMN=DIST monkey TOP.CMN=AF.have-head-NMLZ
 'The clever one (lit. the one having brains) is that monkey.'

Similarly, PF-words in the shape of *ROOT-en* do not have the argument function unless they first undergo *Ca-* reduplication, a morphological process that creates irrealis forms, as shown in (18).

(18) Central Amis (D. Liu 1999: 52)

- a. **cilah-en** aku ku=dateng
 pickle-PF 1SG.ACT TOP.CMN=vegetable
 'I will pickle the vegetables.'
- b. u=dateng ku={**ca~cilah-en/*cilah-en**} aku
 CMN=vegetable TOP.CMN=IRR~pickle-PF 1SG.ACT
 'What I will pickle is the vegetables.'

The irrealis PF-word *Ca~ROOT-en* additionally has the predicate function, like the morphologically simpler form *ROOT-en*. Shown in (19) are a minimal pair of the PF-word *ROOT-en* and its irrealis counterpart, both serving as the matrix predicate.

(19) Central Amis (J. Wu 2006: 134, 126)

- a. **palu-en** ni=mayaw ci=dongi
 beat-PF ACT.PSN=M. TOP.PSN=D.
 ‘Mayaw will beat Dongi.’
- b. **pa~palu-en** ni=sera ci=kuyu
 IRR~beat-PF ACT.PSN=S. TOP.PSN=K.
 ‘Sera will beat Kuyu.’

Based on (19), the two forms may not seem to differ much in functions, but in fact they demonstrate different syntactic distributions. This is equally true between the AF Mstem and their irrealis counterpart *Ca*~Mstem. Syntactic differences between basic AF/PF-words and their irrealis counterparts will be explored in §5.2.1.

By contrast, unlike basic AF/PF-words, CF/LF-words have both the predicate and argument function without any concomitant morphological changes. CF- and LF-words are illustrated in (20) and (21) respectively.

(20) Central Amis (D. Liu 1999: 55)

- a. **sa-pi-cicih** nira ku=funus tu=ra riko’
 CA-K.EXT-tear 3SG.ACT TOP.CMN=machete UND.CMN=DIST clothes
 ‘{He/She} tore those clothes with the machete.’
- b. u=funus ku=**sa-pi-cicih** nira tu=ra riko’
 CMN=machete TOP.CMN=CA-K.EXT-tear 3SG.ACT UND.CMN=DIST clothes
 ‘What {he/she} tore those clothes with is the machete.’

(21) Central Amis (D. Liu 1999: 55)

- a. **ka-futi’-an** isu ku=anengan
 K.INT-sleep-LA 2SG.ACT TOP.CMN=chair
 ‘You sleep in the chair.’
- b. u=anengan ku=**ka-futi’-an** aku
 CMN=chair TOP.CMN=K.INT-sleep-LA 1SG.ACT
 ‘The place where I sleep is the chair.’

Additionally, D. Liu (1999: 43-45) observed another morphosyntactic asymmetry between AF/PF and CF/LF forms, which has to do with how the predicate is negated. It was found that AF/PF-words are negated by the verbal pattern but CF/LF-words by the

nominal pattern, as contrasted by the AF Mstem *mi-futing* in (22) and the Kstem-based LF-word *pi-licay-an* in (23).

(22) Central Amis (D. Liu 1999: 43)

- a. **mi-futing** ci=panay anudafak
 AF.EXT-fish TOP.PSN=P. tomorrow
 ‘Panay is going fishing tomorrow.’
- b. **caay pi-futing** ci=panay anudafak
 NEG K.EXT-fish TOP.PSN=P. tomorrow
 ‘Panay is not going fishing tomorrow.’

(23) Central Amis (D. Liu 1999: 45)

- a. **pi-licay-an** nu=ra imeng ku=kakitaan
 K.EXT-ask-LA ACT.CMN=DIST guard TOP.CMN=chief
 ‘That guard consulted the chief.’
- b. **caay ku=pi-licay-an** nu=ra imeng ku=kakitaan
 NEG TOP.CMN=K.EXT-ask-LA ACT.CMN=DIST guard TOP.CMN=chief
 ‘The chief was not the one who that guard consulted.’

The distribution facts in (15) through (23), when taken together, seem to suggest that CF/LF-words are more nominal than AF/PF-words because the former, but not the latter, have both the predicate and argument function, and also because the former, but not the latter, are negated like underived nouns when acting as predicates. All these distribution facts have been corroborated by J. Wu (2006, 2007), who made one additional observation. Specifically, CF-words, but not LF-words, have the option to be negated by the verbal pattern. This seems to suggest LF-words are the most “nominal” of all on the nominal-verbal spectrum because even as predicates they behave like underived nouns. This topic will be further explored in §5.2.2.2, where the syntactic functions of CF- and LF-words are compared.

Finally, in a similar vein Tsukida (2008: 278) analyzed AF/PF-words as part of the verbal Focus system on the one hand and CF/LF-words as “deverbal nouns” on the other. However, her claim of CF/LF forms being deverbal nouns is based on Fata’an

Amis whereas the observations of both D. Liu (1999) and J. Wu (2006, 2007) come from Central Amis. It remains to be investigated whether CF/LF-words in Fata'an Amis are indeed deverbal nouns as such, but those in Central Amis are clearly not lexical nouns in the same sense as uncontroversial nouns due to the verbal properties illustrated by CF/LF-words but not by typical nouns, as we shall see in §5.2.2.

In summary, according to previous studies, the morphological requirements on Focus-words that serve as arguments create an asymmetry between AF/PF and CF/LF forms. Moreover, the predicate negation pattern generates yet another asymmetry, but this time between AF/PF forms, which can only be negated by the verbal pattern, and LF ones, which can only be negated by the nominal pattern. Of particular interest to the present study is the reported distribution that CF, but not LF, forms are susceptible to verbal negation in addition to nominal negation. As it shall become clear later, the two negation patterns are in fact due to the different scopes of negation they each involve. And after I present an update on the data, the aforementioned studies will be further evaluated where appropriate.

5.2. Argument nominalizations

In this section on argument nominalizations, I address issues relating to the previously observed asymmetries between AF/PF and CF/LF forms on the one hand as well as those between CF and LF forms on the other. For expository purposes, AF-words are taken as representatives of the AF/PF group and investigated under Actor nominalizations in §5.2.1. On the other hand, CF- and LF-words are compared side by side under Circumstantial nominalizations in §5.2.2.

Combinations of various affixes with the Mstem or Kstem in Amis can be highly stratified and thus rather complicated, and it is not always clear how a given multi-affixed Focus-word is to be used in a complete sentence (e.g. S. Zeng 1991, 2002). Thus, the present section on argument nominalizations limits the shape of the Mstem to those consisting of the root and no more than one affix. On such a basis, word forms also investigated include the corresponding Kstem, the root suffixed by PF *-en*, the Kstem prefixed by Conveyance *sa-*, and finally the Mstem or Kstem suffixed by Locative *-an*. These are also the word forms frequently discussed in the literature and would suffice to illustrate the points to be made here.

5.2.1. Actor nominalizations

Since the Topic NP of an Mstem verb is collectively referred to as the Actor, any expression denoting the Actor role of that Mstem is then an Actor nominalization, where the verb assumes the form of Mstem-*ay* for realis events (§5.2.1.1) or *Ca~*Mstem for irrealis ones (§5.2.1.2).

5.2.1.1. The realis Mstem-*ay*

As mentioned above, the Mstem serves as the matrix predicate in affirmative indicative-realis AF-constructions, and is required to be suffixed by *-ay* when occurring within an Actor nominalization, which denotes the Actor-Topic regardless of its specific thematic roles. Actor nominalizations are used as either referential or restricting expressions. For instance, the Actor nominalization in (24) denotes the agent of a biting

event whereas that in (25) the patient of an eating event, but in both cases what is denoted by the Mstem-*ay* corresponds to the Actor-Topic of the Mstem.

(24) Central Amis

- a. **mi-kalat** ku=ra oner tu=wacu
 AF.EXT-bite TOP.CMN=MED snake UND.CMN=dog
 ‘That snake is biting a dog.’ (D. Liu 1999: 48)
- b. u=ra oner ku=[{mi-kalat-ay/*mi-kalat} tu=wacu]
 CMN=MED snake TOP.CMN=AF.EXT-bite-NMLZ UND.CMN=dog
 ‘What is biting a dog is that snake.’ (D. Liu 1999: 48)
- c. ma-patay=tu ku=[ya {mi-kalat-ay/*mi-kalat} ci=aki-an]
 AF.INT-die=already TOP.CMN=DIST AF.EXT-bite-NMLZ PSN=A.-UND
a wacu
 LIG dog
 ‘That dog which bit Aki is already dead’.’ (J. Wu 2007: 105)

(25) Central Amis (J. Wu 2006: 362, 366, 96)

- a. **ma-kaen** ni=aki ku=ya tali
 AF.INT-eat ACT.PSN=A. TOP.CMN=DIST taro
 ‘Aki ate that taro.’
- b. u=maan ku=[ma-kaen-ay ni=aki]
 CMN=what TOP.CMN=AF.INT-eat-NMLZ ACT.PSN=A.
 ‘What is it that Aki ate?’
- c. tatiuh ku=[ya ma-kaen-ay ni=aki] a tali
 AF.bad TOP.CMN=DIST AF.INT-eat-NMLZ ACT.PSN=A. LIG taro
 ‘That taro which Aki ate was bad.’

As denoting expressions, the external syntax of Actor nominalizations follows that of underived nouns in their NP-use and modification-use. Like underived nouns, Actor nominalizations serving as the Topic NP are preceded by *ku*, as compared by the underlined Topic NPs in both (24) and (25). In addition, Actor nominalizations that modify another nominal, as in the third example of both (24) and (25), follow one of the three coding strategies for one noun to modify another in Amis. Namely, the modifier noun precedes the modified one with the attributive ligature *a* occurring in between, as in

takaw a tamdaw (Kaohsiung *LIG* person) ‘person born and raised in Kaohsiung’ and *kilang a kayakay* (tree *LIG* bridge) ‘wooden bridge’ (J. Wu 2012).

Importantly, the *-ay* suffixation process is operated on all classes of the *Mstem* in Table 5.1, which is nicely illustrated by (26), where a series of Actor nominalizations in disjunction are constructed from different classes of the *Mstem*.

(26) Central Amis (Supplementary Materials, Daily Conv. L27)

ma-sa-maan-ay a fa'inayan ku=ka-ulah-an
 AF.INT-CA-do.what-NMLZ *LIG* man TOP.CMN=K.INT-like-LA

isu? **u=takaraw-ay** **u=ma-susu-ay**
 2SG.ACT CMN=AF.tall-NMLZ CMN=AF.INT-fat-NMLZ

u=ci-pida-ay ri **u=ma-laluk-ay?**
 CMN=AF.have-money-NMLZ or CMN=AF.INT-diligent-NMLZ

‘What kind of men do you like? Tall ones, fat ones, rich ones, or diligent ones?’

While the distribution of the marker *-ay* is rather clear, how to analyze it has been the focus of contention among specialists of Amis. Both M. Lin (1995) and D. Liu (1999, 2011) analyzed it as a nominalizer whereas J. Wu (2003, 2006) specifically argued against the nominalizer analysis and proposed to identify *-ay* as a factual mood marker.¹⁶² On this issue, I align myself with M. Lin (1995) and D. Liu (1999, 2011) in adopting the nominalizer analysis, which will be justified with additional evidence in this section.

Tsukida (2008: 278) referred to forms in the shape of *Mstem-ay* as “Neutral relative,” which “is used as the head of a referential phrase, or as a modifier in a referential phrase.” The two functions she described correspond to the NP-use and modification-use of grammatical nominalizations in this study. Thus, aside from the different labels adopted, Tsukida’s (2008) analysis essentially matches M. Lin’s (1995) and D. Liu’s (1999) in spirits.

¹⁶² On a related note, based on data from Southern Amis (specifically Beinan), S. Zeng (1991, 2002), M. Yang (2005), and L. Jiang (2009, 2012) all treat *-ay* as a perfective aspect maker.

The factual mood analysis for the formative *-ay* is for the most part motivated by the fact that the Mstem-*ay* form can also serve as the matrix predicate and express a realized event. The minimal pair in (27) shows that adding the formative *-ay* changes the default ongoing or imminent reading of the Extrovert Mstem into one that emphasizes the happening of a realized event.

(27) Central Amis (J. Wu 2006: 125)

- a. **mi-kilim** kaku ci=panay-an
 AF.EXT-search 1SG.TOP PSN=A.-UND
 ‘I am {looking for/going to look for} Panay.’
- b. **mi-kilim-ay** kaku ci=panay-an
 AF.EXT-search-NMLZ 1SG.TOP PSN=A.-UND
 ‘I did look for Panay.’

Granted that the formative *-ay* imports a factual/perfective interpretation, it does not rule out the possibility that the Mstem-*ay* form is a nominalized structure functioning as a nominal predicate and that the factual/perfective overtone is derived from the presupposition associated with nominalization, as was first suggested by M. Liu (1995: 167). This line of thought is supported by the syntactic consequences brought about by the *-ay* affixation, to be demonstrated below. Some observations from previous studies are also repeated here so as to present a complete picture.

First, (28) is another minimal pair of the Mstem and the Mstem-*ay*, but unlike that in (27), this pair contains a past-time adverbial.

(28) Central Amis (Fieldnotes)

- a. **mi-palu** tu=wawa aku inacila ci=mayaw
 AF.EXT-beat UND.CMN=child 1SG.GEN yesterday TOP.PSN=M.
 ‘Mayaw beat my child yesterday.’
- b. **mi-palu-ay** tu=wawa aku inacila ci=mayaw
 AF.EXT-beat-NMLZ UND.CMN=child 1SG.GEN yesterday TOP.PSN=M.
 ‘Mayaw did beat my child yesterday (for sure).’

Since the Mstem with the past-time adverbial already asserts a realized event, the Mstem-*ay*, with everything else being held equal, is now associated with emphatic overtones. But even so, both (27) and (28) may give the impression that the marker *-ay* only contributes to modify the semantics of the Mstem it is attached to, be it factual or emphatic. However, the Mstem-*ay* in fact demonstrates syntactic potentials different from the Mstem. As first pointed out by M. Lin (1995), the Mstem-*ay* form, but not the Mstem, can be preceded by the prenominal marker *u*, thus following the nominal predication pattern, as shown in (29).

(29) Central Amis

- a. **u={mi-palu-ay/*mi-palu}** tu=wawa aku inacila] ci=mayaw
 CMN=AF.EXT-beat-NMLZ UND.CMN=child 1SG.GEN yesterday TOP.PSN=M.
 ‘Mayaw is the one who beat my child yesterday.’ (Fieldnotes; cf. D. Liu 1999: 112)
- b. **u=ngayngay** ci=aki
 CMN=Hakka TOP.PSN=A.
 ‘Aki is a Hakka person.’ (D. Liu 1999: 46)

The Mstem-*ay*, together with its patientive argument and temporal adverbial, is then a denoting expression syntactically treated on an equal footing with an underived noun. If so, the Mstem-*ay* form in both (27) and (28) from above might as well be treated in the same manner; only that the prenominal marker *u* is absent. However, the absence of the prenominal marker *u* from the Mstem-*ay* should not be taken as evidence against the nominal property of the Mstem-*ay* because the same option also applies to nominal predicates consisting of underived nouns (see J. Wu 2006: 76).

Second, the nominal nature of the Mstem-*ay* form is further demonstrated by how it is negated. As D. Liu (1999) has shown extensively, the Mstem-*ay* form, but not the Mstem, is negated by the nominal pattern, as shown in (30).

(30) Central Amis

- a. **caay ku={mi-palu-ay/*mi-palu}** tu=wawa aku
 NEG TOP.CMN=AF.EXT-beat-NMLZ UND.CMN=child 1SG.GEN
 inacila] ci=mayaw
 yesterday TOP.PSN=M.
 ‘Mayaw is not the one who beat my child yesterday.’
 (Fieldnotes; cf. D. Liu 1999: 112)
- b. **caay ku=[ngayngay]** ci=aki
 NEG TOP.CMN=Hakka TOP.PSN=A.
 ‘Aki is not a Hakka person.’ (D. Liu 1999: 46)

Therefore, the Mstem *-ay* form, but not the Mstem, follows the nominal pattern in both affirmative and negative predication. Crucially, the negative predication pattern, as in (30), is a stronger indicator for a nominal construction than the affirmative one, where the prenominal marker *u* is sometimes left out, thus leaving little evidence for something being nominal (cf. (28) and (29) above).

In addition, the Mstem-*ay* is required in the argument position, which has been demonstrated above. For the sake of fair comparisons, (31) shows that the Mstem *-ay*, but not the Mstem, is permitted in the argument position, as is a typical noun.

(31) Central Amis (Fieldnotes)

- a. **caay ci=mayaw ku=[ya {mi-palu-ay/*mi-palu}** tu=wawa
 NEG PSN=M. TOP.CMN=DIST AF.EXT-beat-NMLZ UND.CMN=child
 aku inacila]
 1SG.GEN yesterday
 ‘That one who beat my child yesterday is not Mayaw.’
- b. **caay ci=mayaw ku=[ya fafahiyan]**
 NEG PSN=M. TOP.CMN=DIST man
 ‘That man is not Mayaw.’

Comparisons in (29) through (31) make it clear that predicate negation is a good syntactic operation through which to unveil the morphosyntactic differences between the the Mstem and the Mstem-*ay*, which can be accounted for if the formative *-ay* is analyzed as a nominalizer that creates Actor nominalizations out of the Mstem.

Third, I would like to show that negation can take place within a nominalization just as it can outside one. In (30), the Mstem-*ay* is negated by the nominal pattern. This suggests that what is negated is the entity denoted by a nominalized structure consisting of a presupposed predication headed by the Mstem-*ay*, so the negation happens outside the nominalization structure. For a descriptive purpose, this can be referred to as *external negation*. As opposed to that, there is *internal negation*, which occurs within a nominalization that as a whole denotes an entity characterized by a negative predication. To show this, it is necessary to first review the verbal pattern of predicate negation. For instance, to negate the Mstem *mi-palu* ‘AF.EXT-beat’ as in (28), its corresponding Kstem is required after the negator *caay*, as in (32).

(32) Central Amis (Fieldnotes)

caay **pi-palu** tu=wawa aku inacila ci=mayaw
 NEG K.EXT-beat UND.CMN=child 1SG.GEN yesterday TOP.PSN=M.
 ‘Mayaw didn’t beat my child yesterday.’

Moreover, an often neglected fact is that a negated predicate phrase like that in (32) can also function as a denoting expressing, thus giving rise to negation internal to a nominalization. Nominalizing a negative predicate phrase is similar to nominalizing an affirmative one, except that the nominalizer *-ay* is attached to the Mstem in the affirmative context, as in (29) and (30), but to the negator instead in the negative context, as in (33).

(33) Central Amis (Fieldnotes)

- a.* **u=caay** pi-palu tu=wawa aku inacila
 CMN=NEG K.EXT-beat UND.CMN=child 1SG.GEN yesterday
 ci=mayaw
 TOP.PSN=M.
- b. **(u)=[caay-ay** pi-palu tu=wawa aku inacila]
 CMN=NEG-NMLZ K.EXT-beat UND.CMN=child 1SG.GEN yesterday
 ci=mayaw
 TOP.PSN=M.
 ‘Mayaw is the one who didn’t beat my child yesterday.’

Much like regular Mstems shown above, the negator *caay* cannot be preceded by the prenominal marker *u* unless it is marked by *-ay*.¹⁶³ Moreover, aside from being a nominal predicate on the phrasal level, the internally negated predicate phrase as in (33) can also serve as an argument just like regular Mstems nominalized by *-ay*, as illustrated in (34).

(34) Central Amis (Fieldnotes)

- udengan ci=mayaw ku=[{**caay-ay**/***caay**} pi-palu tu=wawa
 only PSN=M. TOP.CMN=NEG-NMLZ K.EXT-beat UND.CMN=child
 aku inacila]
 1SG.GEN yesterday
 ‘The only one who didn’t beat my child yesterday is Mayaw.’

Both (33) and (34) accentuate the nominal pattern of a negated Focus-word, which is not permitted at all if the negator is not marked by *-ay* in the first place.

In addition, the nominalizer analysis of *-ay* gives a straightforward account of why there are two options to negate a given Mstem. Take the Introvert Mstem *ma-tuka* ‘AF.INT-lazy’ for instance. The verbal negation pattern demands its corresponding Kstem

¹⁶³ M. Chang (2007) mentioned three forms of the negator, *ca*, *caay*, and *caay-ay*, all of which immediately precede the Kstem. It seems that the form *caay* historically came from the negator *ca* suffixed by *-ay*, just as a typical Mstem is suffixed by *-ay*. This is supported by the fact that the negator *ca* also has its Kstem form *ka-ca* (thus making the negator morphologically a Type III verb), which is used when a negated predicate is further negated (i.e. double negation; see her work for examples), just as a typical Kstem is required under verbal negation. However, since *caay* is further suffixed by *-ay*, giving rise to *caay-ay* and that *caay* seems to be much more common than simply *ca*, I treat the *-ay* as in *caay* a historical relic, and thus do not put a morpheme boundary in *caay*. By contrast, *caay-ay* is segmented into two morphemes since it shows syntactic differences from the basic negator *caay*.

ka-tuka whereas the nominal negation pattern requires the Mstem to be marked by *-ay*, as contrasted in (35).

(35) Central Amis (M. Chang 2007: 93)

- a. **caay ka-tuka kaku**
 NEG K.INT-lazy 1SG.TOP
 ‘I am not lazy.’
- b. **caay ku=[ma-tuka-ay] kaku**
 NEG TOP.CMN=AF.INT-lazy-NMLZ 1SG.TOP
 ‘I am not a lazy person.’

While the verbal pattern in the first example negates the state of affairs expressed by the Kstem, the nominal pattern in the second negates the equational/identificational relationship between two denoting expressions, one of which is an Actor nominalization characterized by the state of affairs expressed by the Mstem. The same generalization holds true of the Extrovert Mstem as well (in fact to all Mstems), as illustrated in (36), where the Mstem/Kstem subcategorizes for a non-Topic Undergoer NP.

(36) Central Amis (Fieldnotes)

- a. **caay [pi-tilid tu=ngangan aku] ci=panay**
 NEG K.EXT-write UND.CMN=name 1SG.GEN TOP.PSN=P.
 ‘Panay didn’t write my name.’
- b. **caay ku=[mi-tilid-ay tu=ngangan aku] ci=panay**
 NEG TOP.CMN=AF.EXT-write-NMLZ UND.CMN=name 1SG.GEN TOP.PSN=P.
 ‘Panay is not the one who wrote my name.’

It should be clear by now that the formative *-ay* does not only contribute perfective, factual, or/and emphatic overtones to the Mstem it is attached to, but also enables the Mstem to be treated like a denoting nominal on the phrasal level, which is then used as nominal predicates, arguments, or modifiers of arguments (i.e. so-called relative clauses). AF-constructions nominalizable by *-ay* can be as simple as just the lone Mstem, as in (35), or as complex as an internally negated predication (consisting of the negator and the Kstem) with its own patientive arguments and temporal adverbials, as in

(33) and (34). The morphosyntactic differences between negation external and internal to AF nominalizations are summarized in Table 5.3.

Table 5.3: External vs. internal negation of Actor nominalizations in Central Amis

External	AFF	NEG
Internal		
AFF	<i>u=mi-palu-ay</i> (CMN=AF.EXT-beat-NMLZ) ‘the one who beat X’ [(29) and (31)]	<i>caay ku=mi-palu-ay</i> (NEG TOP.CMN=AF.EXT-beat-NMLZ) ‘not the one who beat X’ [(30)]
NEG	<i>u=caay-ay pi-palu</i> (CMN=NEG-NMLZ K.EXT-beat) ‘the one who didn’t beat X’ [(33) and (34)]	NO DATA

Note: AFF and NEG stand for affirmative and negative contexts respectively, and X for the patientive NP encoded as the non-Topic Undergoer. Examples with both internal and external nominalizations are logically possible, but no such data have been successfully elicited (probably due to some pragmatic conflicts). When a nominal is externally negated (the upper-right cell), the prenominal marker *ku* is required. In all the other contexts, the prenominal marker can be *u*, *ku*, or *tu*, depending on the grammatical relations of the NPs it marks. See the referenced examples for complete sentences.

The constructions in brackets are all denoting expressions headed by the nominalized Mstem or the negator *caay*, both marked as such by the formative *-ay*. Their external distributions follow those of underived common nouns, with nominal relation markers occurring at the left periphery, which is not possible if the formative *-ay* is absent. On the other hand, their internal morphosyntax resembles in all aspects the verbal pattern of the Mstem, taking argument NPs, collocating with temporal adverbials, and switching to its corresponding Kstem so as to be negated by *caay*. That is to say, functional differences between the Mstem and its nominalized form Mstem-*ay* are only demonstrable through

their external interaction with larger constituents. When nominal predicates consisting of the Mstem-*ay* form start to dispense with the prenominal marker *u*, just as common-noun predicates sometimes do, the Mstem and Mstem-*ay* appear to be identical save for the formative -*ay*. Without syntactic clues from larger constituents, the two forms are left with only semantic differences, as in (27) and (28). From there, it is just a step away for a marker originally meant to create denoting expressions to end up marking non-denoting notions such as TAM or speaker's stance, which is a common theme of numerous contributions in Yap et al. (2011).

One last illustration of the nominal property of the Mstem-*ay* form is its compatibility with GEN-phrases, a distribution that draws little attention in the Amis literature. For instance, the Extrovert Mstem *mi-palu* 'AF.EXT-beat' selects the beater as the Topic NP and the beaten one as the non-Topic Undergoer, as in (37).

(37) Central Amis (J. Wu 2006: 119)

mi-palu ci=kilang ci=canglah-an inacila
 AF.EXT-beat TOP.PSN=K. PSN=C.-UND yesterday
 'Kilang beat Canglah yesterday.'

When the beater role is nominalized and denoted by *mi-palu-ay*, the beaten one maintains the same marking, as in (38), where *mi-palu-ay* demonstrates the nominal pattern for both affirmative and negative predication.

(38) Central Amis (Fieldnotes)

- a. **u=[mi-palu-ay** ci=canglah-an] ci=kilang
 CMN=AF.EXT-beat-NMLZ PSN=C.-UND TOP.PSN=K.
 'Kilang is the one who beat Canglah.'
- b. **caay ku=[mi-palu-ay** ci=canglah-an] ci=kilang
 NEG TOP.CMN=AF.EXT-beat-NMLZ PSN=C.-UND TOP.PSN=K.
 'Kilang is not the one who beat Canglah.'

Given its pattern in (37), the marking for the non-Topic Undergoer in (38) is expected. However, what is less anticipated is the fact that the Mstem-*ay* also collocates with a GEN-phrase, as in (39).

(39) Central Amis (Fieldnotes)

- a. **u=[mi-palu-ay** **ni=canglah]** **ci=kilang**
 CMN=AF.EXT-beat-NMLZ GEN.PSN=C. TOP.PSN=K.
 ‘Kilang is the one who beat (someone) on Canglah’s behalf.’
- b. **caay** **ku=[mi-palu-ay** **ni=canglah]** **ci=kilang**
 NEG TOP.CMN=AF.EXT-beat-NMLZ GEN.PSN=C. TOP.PSN=K.
 ‘Kilang is not the one who beat (someone) on Canglah’s behalf.’

Here *mi-palu-ay* still denotes the beater role (i.e. Kilang), and the GEN-phrase is interpreted as someone (i.e. Canglah) for or on behalf of whom the beating action is executed. One possible scenario, according to my consultants, is that Canglah ordered or hired Kilang to beat someone else. Unlike the non-Topic Undergoer, the GEN-marked phrase is not even permitted in the argument structure of the Mstem *mi-palu*. Thus, one reasonable explanation for its presence in (39) is that the word form *mi-palu-ay*, unlike the Mstem *mi-palu*, is a nominalized denoting expression and as such allows the modification of a reference-point entity marked in GEN (cf. *wacu ni=canglah* ‘Canglah’s dog’; more on this in Chapter 8).¹⁶⁴ In other words, the denoting expression found in (39) is a nominal predicate, and the same nominal functions as the Topic NP just as readily, as in (40).¹⁶⁵

¹⁶⁴ However, Focus-words collocating with *ni*-phrases are not necessarily nominal, such as the Introvert AF-word *ma*-verb (see (2)b) and the PF-word ROOT-*en* (see (13)c).

¹⁶⁵ Given the appropriate contextual factors, AF-words in Tagalog may also combine with a phrase that elsewhere marks the possessor. See Kaufman (2009b: 23) for an example.

(40) Central Amis (Fieldnotes)

- a. u=ya tamdaw ku=[**mi-palu-ay** ni=canglah]
 CMN=DIST person TOP.CMN=AF.EXT-beat-NMLZ GEN.PSN=C.
 ‘The one who beat (someone) on Canglah’s behalf is that person.’
- b. u=ya tamdaw ku=[**mi-palu-ay** ni=canglah ci=panay-an]
 CMN=DIST person TOP.CMN=AF.EXT-beat-NMLZ GEN.PSN=C. PSN=P.-UND
 ‘The one who beat Panay on Canglah’s behalf is that person.’

Finally, the nominalizer analysis for the marker *-ay* is compatible with its collocations with numerals, a distribution fact that is hard to reconcile with the perfective aspect or factual mood analysis. In (41), for instance, the marker *-ay* is attached to not only the Mstem *mi-palu* ‘AF.EXT-beat’ but also to the numeral *ta~tusa* ‘HUM-two’.

(41) Central Amis (J. Wu 2006: 98)

- paka-araw ci=sawmah-an ku=[ya **ta~tusa-ay** a
 ABLT-see PSN=S.-UND TOP.CMN=DIST HUM-two-NMLZ LIG
 fa’inayan] a sinsi, u=[ya **mi-palu-ay** ci=mayaw-an]
 man LIG teacher CMN=DIST AF.EXT-beat-NMLZ PSN=M.-UND
 ‘Those two male teachers saw Sawmah, i.e. those who beat Mayaw.’

Under the nominalizer analysis, *ta~tusa-ay* is a denoting expression for two human beings characterized by the number of two, just as is *mi-palu-ay* a denoting expression characterized by the Actor role of a beating event. In both cases, the marker *-ay* helps to create a nominalized structure out of a base, be it a numeral or a verbal stem. The two denoting expressions, both modified by the distal demonstrative, are even coreferential in the present example. However, the perfective/factual analysis fails to account for expressions like *ya ta~tusa-ay* ‘those two (human beings)’, where the marker *-ay* is used regardless of whether the number of two is interpreted as perfective/imperfective or factual/non-factual.

5.2.1.2. The irrealis *Ca~Mstem*

Other than the fact that the *Mstem-ay* form can serve as the matrix predicate (without being preceded by the prenominal marker *u*), another reason why J. Wu (2003, 2006) found the nominalizer analysis of *-ay* unsatisfactory is that the formative in question does not co-occur with the irrealis *Ca~Mstem* form. The irrealis form describes “a not-yet-happening event/state in the future... or non-happening event/state in the past” (J. Wu 2006: 127), and thus it is rather unexpected, as the logic goes, that a putative nominalizer should not nominalize a verb simply because it has a different modal interpretation.

Take the Class III *Mstem tayra* ‘AF.go’ for instance, whose *Kstem* is *ka-tayra*, as respectively illustrated by the affirmative indicative and affirmative imperative sentences in (42).

(42) Central Amis (J. Wu 2006: 119, 139)

- a. **tayra** ci=aki
 AF.go TOP.PSN=A.
 ‘Aki is {on his way/going to leave}.’
- b. **ka-tayra** i=taypak
 K-go LOC=Taipei
 ‘Go to Taipei!’

Like all the other *Mstems*, *tayra* ‘AF.go’ is nominalized by *-ay* and the result form can function, among others, as the modifier of an argument. By contrast, its irrealis form *ta~tayra* ‘IRR~AF.go’, when serving the same modifier function, is prohibited from the same morphological operation. The discrepancy between the two forms is contrasted in (43).

have been shown in (43) to some extent, where they both serve as the modifier of an argument.

To further illustrate the point, here is another example of the Class III Mstem. Both the Mstem *ci-wawa* ‘AF.have-child’ and its irrealis counterpart *ca~ci-wawa* ‘IRR~AF.have-child’ show the verbal predication pattern, as in (44) (see also (19) above).

(44) Central Amis (Fieldnotes)

- a. **ci-wawa** ci=panay
 AF.have-child TOP.PSN=P.
 ‘Panay has {a child/children}.’
- b. **ca~ci-wawa** ci=panay
 IRR~AF.have-child TOP.PSN=P.
 ‘Panay will have {a child/children}.’

However, when it comes to serving as arguments, the Mstem has to be marked by *-ay* while the *Ca~Mstem* undergoes no concomitant changes other than being placed in the typical argument slot, as in (45), where the forms in question head a complete NP.

(45) Central Amis (Fieldnotes)

- a. cima ku={**ci-wawa-ay**/***ci-wawa**}
 who TOP.CMN=AF.have-child-NMLZ
 ‘Who is the one that has {a child/children}?’
- b. cima ku=**ca~ci-wawa**
 who TOP.CMN=IRR~AF.have-child
 ‘Who is the one that will have {a child/children}?’

Like (43) above, (45) shows that the *Ca~Mstem* has the same nominal distribution as the *Mstem-ay*. Nevertheless, given the discrepancy between (44) and (45), one would have to come to the conclusion that the Mstem has only the verbal use whereas the *Ca~Mstem* has both the verbal and nominal use. This conclusion is additionally supported by negation patterns. The Mstem *ci-wawa* can be negated by the verbal pattern, with its Kstem *ka-ci-wawa* following the negator, but not by the nominal one, with the Mstem preceded by the prenominal marker *ku*, as contrasted in (46).

(46) Central Amis (Fieldnotes)

- a. **caay** ka-ci-wawa ci=panay
 NEG K-have-child TOP.PSN=P.
 ‘Panay does not have any child.’
- b.* **caay** ku=ci-wawa ci=panay
 NEG TOP.CMN=AF.have-child TOP.PSN=P.

The second example is not acceptable because the nominal negation pattern requires common nouns as well as the nominalized Mstem-*ay* in the slot after *ku* (see also (35) above). By contrast, the *Ca~Mstem ca~ci-wawa* is amenable to both negation patterns, with its Kstem *ka-ca~ci-wawa* immediately following the negator in the verbal pattern and the same *Ca~Mstem* coming after the prenominal marker *ku* in the nominal pattern, as in (47).

(47) Central Amis (Fieldnotes)

- a. **caay ka-ca~ci-wawa** ci=panay
 NEG K-IRR~have-child TOP.PSN=P.
 ‘Panay will not have any child.’
- b. **caay ku=ca~ci-wawa** ci=panay
 NEG TOP.CMN=IRR~have-child TOP.PSN=P.
 ‘Panay is not the one who will have {a child/children}.’

Finally, just as in Class III, verb forms in Class I and II all have to be either Mstem-*ay* or *Ca~Mstem* in order to semantically denote the Actor-Topic and syntactically assume the function of arguments, modifiers of arguments, or nominal predicates, which are negated by the nominal pattern like underived nouns. To continue the examples from Table 5.2 above, two roots (*palu* ‘beat’ and *ulah* ‘like’) and two verb class prefixes (Extrovert *mi-* and Introvert *ma-*) would potentially generate four Mstem forms, which when operated upon by two morphological processes (*-ay* suffixation and *Ca-* reduplication) would then give rise to eight word forms in total. All of them can take up typical argument positions in Amis, as summarized in Table 5.4 (cf. J. Wu 2007: 104).

Table 5.4: Actor nominalizations of Class I and II verbs in Central Amis¹⁶⁷

Word forms	Root types	Extrovert (Class I <i>mi</i> -verbs)	Introvert (Class II <i>ma</i> -verbs)
Mstem-ay	Action root (e.g. <i>palu</i> ‘beat’)	<i>mi-palu-ay</i> ‘the who beats B’	<i>ma-palu-ay</i> ‘the one who gets beat up (by A)’
	State root (e.g. <i>ulah</i> ‘like’)	<i>mi-ulah-ay</i> ‘the one who courts B’	<i>ma-ulah-ay</i> ‘the one who likes B’
Ca~Mstem	Action root (e.g. <i>palu</i> ‘beat’)	<i>ma~mi-palu</i> ‘the one who’ll beat B’	<i>ma~ma-palu</i> ‘the one who’ll get beat up (by A)’
	State root (e.g. <i>ulah</i> ‘like’)	<i>ma~mi-ulah</i> ‘the one who’ll court B’	<i>ma~ma-ulah</i> ‘the one who’ll like B’

Note: A and B stand for event participants, which are encoded as the non-Topic Actor and non-Topic Undergoer respectively.

Examples of the *Ca~Mstem* functioning as Actor nominalizations are illustrated in (48).

(48) Central Amis (Fieldnotes)

- a. cima ku=[**ma~mi-palu** ci=panay-an]
who TOP.CMN=IRR~AF.EXT-beat PSN=P.-UND
‘Who is the one that will beat Panay?’
- b. cima ku=[**ma~mi-ulah** ci=panay-an]
who TOP.CMN=IRR~AF.EXT-like PSN=P.-UND
‘Who is the one that will court Panay?’
- c. cima ku=[**ma~ma-palu** ni=panay]
who TOP.CMN=IRR~AF.INT-beat ACT.PSN=P.
‘Who is the one that will get beat up by Panay?’
- d. cima ku=[**ma~ma-ulah** ci=panay-an]
who TOP.CMN=IRR~AF.INT-like PSN=P.-UND
‘Who is the one that will like Panay?’

¹⁶⁷ Word forms in this table are translated with the meanings they would have when acting as argument NPs in contexts like (48).

Moreover, like the *Ca*~Mstem based on Class III Mstem (e.g. *ca~ci-wawa* in (47) above), the *Ca*~Mstem based on Class I and II Mstem permits both the verbal and nominal type of negation, as illustrated by *ma~mi-palu* in (49).

(49) Central Amis (Fieldnotes)

- a. **caay ka-ma~mi-palu** ci=panay-an ci=mayaw
 NEG K-IRR~AF.EXT-beat PSN=P.-UND TOP.PSN=M.
 ‘Mayaw will not beat Panay.’
- b. **caay ku=ma~mi-palu** ci=panay-an ci=mayaw
 NEG TOP.CMN=IRR~AF.EXT-beat PSN=P.-UND TOP.PSN=M.
 ‘Mayaw is not the one who will beat Panay.’

This once again shows the *Ca*~Mstem has the dual function of being equally verbal and nominal, unlike the basic Mstem.

Another aspect where the *Ca*~Mstem differs from the basic Mstem concerns how their corresponding Kstems are formed. Given a basic Mstem, its Kstem is formed in three different ways, depending on which class the Mstem belongs to (see Table 5.1 above). By contrast, the irrealis *Ca*~Mstem is consistently prefixed by *ka-* to produce its Kstem, which is the required form for verbal negation. This is so regardless of the classes of the basic Mstem from which the *Ca*~Mstem is constructed, as summarized in Table 5.5 below. This is as if all forms of the *Ca*~Mstem belonged to Class III Mstem, which should come as no surprise since Class III is a negatively defined miscellaneous category in the first place.

Table 5.5: Basic and reduplicated Mstem/Kstem pairs in Central Amis

	I	II	III
Basic Mstem and its Kstem	<i>mi-palu</i> <i>pi-palu</i>	<i>ma-palu</i> <i>ka-palu</i>	<i>ci-wawa</i> <i>ka-ci-wawa</i>
Reduplicated Mstem and its Kstem	<i>ma~mi-palu</i> <i>ka-ma~mi-palu</i>	<i>ma~ma-palu</i> <i>ka-ma~ma-palu</i>	<i>ca~ci-wawa</i> <i>ka-ca~ci-wawa</i>

Note: *mi-palu* ‘AF.EXT-beat’; *ma-palu* ‘AF.INT-beat’; *ci-wawa* ‘AF.have-child’.

In short, the Mstem has to be marked by *-ay* to become denoting expressions and subsequently take up typical argument slots, which justifies the nominalizer analysis of *-ay*. The irrealis *Ca~Mstem*, on the other hand, is amphibious over verbal and nominal construals (which also holds true to other forms in Amis; see §5.2.2). And since the irrealis *Ca~Mstem* can be a nominalization by itself, it is not compatible with the nominalizer *-ay*. Moreover, the Mstem-*ay* and *Ca~Mstem* both share the dual properties of grammatical nominalizations, heading a phrase that is externally nominal based on the properties it demonstrates when interacting with larger constituents and at the same time internally verbal because of the properties it shares with verb forms outside the nominalization context. Due to the dual properties of nominalizations, negation can easily take place within them, thus having scope over their main predication, or outside of them, targeting at the entities denoted by nominalizations as a whole. Last, both the Mstem-*ay* and *Ca~Mstem* deserve to be called nominalizations on functional grounds, despite the fact that there is a dedicated nominalizer in the case of the Mstem-*ay* but there is none in the case of the *Ca~Mstem*.

More importantly, the difference between the Mstem and its corresponding *Ca~Mstem* is not only morphological (with or without *Ca-* reduplication) and semantic

(with or without the irrealis meaning), but also syntactic. While the Mstem has only the predicate function and can only be negated by the verbal pattern, the *Ca*~Mstem shows both the predicate and argument function and can be negated by both the verbal and nominal pattern.

Finally, syntactic differences are also found between the PF-word *ROOT-en* and its reduplicated counterpart *Ca*~*ROOT-en*, but along somewhat different lines. Specifically, both forms can serve as the matrix predicate, as in (50).

(50) Central Amis (J. Wu 2006: 134, 126) [= (19)]

- a. **palu-en** ni=mayaw ci=dongi
 beat-PF ACT.PSN=M. TOP.PSN=D.
 ‘Mayaw will beat Dongi.’
- b. **pa~palu-en** ni=sera ci=kuyu
 IRR~beat-PF ACT.PSN=S. TOP.PSN=K.
 ‘Sera will beat Kuyu.’

This minimal pair seems to vary only in terms of whether or not the root is partially reduplicated (as well as the different personal names that happen to be used, which is irrelevant to the issue in question), and the reduplication does not seem to contribute much to semantics because the non-reduplicated PF-word *ROOT-en* already describes a future event by default (J. Wu 2006: 110). However, once the two forms are negated, their syntactic differences start to emerge. The basic PF-word *ROOT-en* allows only the verbal negation pattern whereas its reduplicated counterpart *Ca*~*ROOT-en* only the nominal one, as contrasted in (51).

(51) Central Amis (Fieldnotes)

- a. caay **ka-palu-en** ni=sera ci=kuyu
 NEG K-beat-PF ACT.PSN=S. TOP.PSN=K.
 ‘Sera won’t beat Kuyu.’
- b.* caay **ku=palu-en** ni=sera ci=kuyu
 NEG TOP.CMN=beat-PF ACT.PSN=S. TOP.PSN=K.
- c.* caay **ka-pa~palu-en** ni=sera ci=kuyu
 NEG K-IRR~beat-PF ACT.PSN=S. TOP.PSN=K.
- d. caay **ku=pa~palu-en** ni=sera ci=kuyu
 NEG TOP.CMN=IRR~beat-PF ACT.PSN=S. TOP.PSN=K.
 ‘Kuyu is not the one Sera will beat.’

The result from negation is consistent with the fact that only the irrealis *Ca~ROOT-en*, but not the basic *ROOT-en*, has the argument function (see (18) above). Thus, syntactic distributions of these two forms point to the following generalization: unlike the basic *ROOT-en*, the irrealis *Ca~ROOT-en* is always a denoting expression that can serve as an argument as well as a nominal predicate, which is negated like underived nouns.¹⁶⁸ In this regard, the irrealis PF-word *Ca~ROOT-en* is no less a nominalized structure than the irrealis AF-word *Ca~Mstem*, which additionally allows the verbal negation pattern.

5.2.2. Circumstantial nominalizations

Conveyance and Locative nominalizations refer to denoting expressions involving CF- and LF-words respectively, which may subcategorize for other constituents (e.g. event participants and spatio-temporal adverbials). CF- and LF-words are compared in this section due to their similarities and differences observed in previous studies. Since Conveyance and Locative nominalizations denote the instrument, reason, location, or

¹⁶⁸ Hence, a better translation of (50)b would be ‘Kuyu is someone Sera will beat.’

time of a state of affairs, the two types are collectively referred to as Circumstantial nominalizations following Drude (2011).

As has been reviewed in §5.1.3, D. Liu (1999) showed that CF-words take up argument positions without concomitant morphosyntactic changes and that they are negated by the nominal pattern. Both properties hold true for LF-words as well. The examples in (52) show that like the AF-word *Mstem-ay*, which heads Actor nominalizations, CF- and LF-words head Conveyance and Locative nominalizations respectively, which syntactically serve as NP arguments or modifiers of arguments.

(52) Central Amis (D. Liu 1999: 71)

- a. *fangcal* *ku=ni* **sa-ka-r<um>adiw** *ni=usay*
 AF.good TOP.CMN=PROX CA-K-<UM>sing ACT.PSN=U.
 ‘What Usay sang with is good.’
- b. *fangcal* *ku=ni* **sa-ka-r<um>adiw** *ni=usay* *a* *papah*
 AF.good TOP.CMN=PROX CA-K-<UM>sing ACT.PSN=U. LIG leaf
 ‘The leaf that Usay sang with is good.’
- c. *kuhting* *ku=ni* **ka-nawnaw-an** *nu=kakunah*
 AF.black TOP.CMN=PROX K.INT-float-LA ACT.CMN=ant
 ‘(The place) where the ants are floating is black.’
- d. *kuhting* *ku=ni* **ka-nawnaw-an** *nu=kakunah* *a* *nanum*
 AF.black TOP.CMN=PROX K.INT-float-LA ACT.CMN=ant LIG water
 ‘The water on which the ants are floating is black.’

However, J. Wu (2006, 2007) pointed out that CF-words, but not LF-words, can be additionally negated by the verbal pattern. Before dealing with such a difference in negation between CF- and LF-words, I first expound how the two forms are derived and at the same time demonstrate that a semantically unified account of the Introvert AF *ma-*, as argued for in §5.1.1, better anticipates its corresponding CF and LF forms, some of which have not hitherto been documented. Once the derivation of CF and LF forms (§5.2.2.1) is made clear, we will be in a better position to investigate the reported syntactic differences between them (§5.2.2.2).

5.2.2.1. Derivations of forms

CF and LF forms are generally predictable from the Mstem/Kstem pair for AF-words. J. Wu (2006: 113, 2007: 116) summarized their derivational correspondences as in Table 5.6, where the three AF markers belong to the three AF classes shown in Table 5.1 above.¹⁶⁹

Table 5.6: Morphological correspondence among AF, CF, and LF forms in Central Amis (after J. Wu 2006, 2007)

Actor Focus markers		<i>mi-</i>	<i>ma-</i>	< <i>um</i> >
Conveyance Focus forms		<i>sa-pi-</i>	<i>sa-ka-</i>	<i>sa-ka-...<um></i>
Locative Focus forms	Goal-Locative	<i>mi-...-an</i>	---	---
	Patient-Locative	<i>mi-...-an</i>	<i>ka-...-an</i>	< <i>um</i> >...- <i>an</i>
	Location-Locative	<i>pi-...-an</i>	<i>ka-...-an</i>	<i>ka-...<um>...-an</i>

Table 5.6 shows that CF forms are built from the Conveyance *sa-* prefixed to the Kstem whereas LF forms are constructed from the Locative *-an* suffixed to either the Mstem (shaded cells) or the Kstem. Semantically, CF forms may denote “an instrument, a reason, a motivation, or an indirect cause” (J. Wu 2006: 413), depending on the semantics of the root as well as what affixes are involved. On the other hand, LF forms may denote the goal of a motion event (i.e. “the thing that one goes somewhere to get”), the patient of an action, the stimulus of emotion, the transported theme or recipient of object transferral, or the location/time of an activity (among others), all of which J. Wu (2007: 116)

¹⁶⁹ Table 5.6 contains information drawn from both J. Wu (2006: 113) and J. Wu (2007: 116), and it differs from the original tables only in terms of the terminology and layout of the data.

generalized into Goal-Locative, Patient-Locative, and Location-Locative.¹⁷⁰ Her three types of LF forms will be illustrated and then slightly revised where appropriate.

However, there are some problems with Table 5.6. Since J. Wu (2006, 2007) distinguished AF *ma-*, PF *ma-*, and Neutral *ma-* based on argument marking patterns (see §5.1.1), Table 5.6 would imply that only her AF *ma-* verbs, but not the other two types, have corresponding CF forms in the shape of *sa-ka-ROOT* and LF forms in the shape of *ka-ROOT-an*. If this were true, there would be good motivation to single out AF *ma-* from either PF or Neutral *ma-*. But a closer inspection shows this is not the case. In addition, it is explicitly indicated in Table 5.6 that while LF forms corresponding to *mi-* verbs are based on the Mstem or Kstem, those corresponding to *ma-* verbs are only based on the Kstem so that LF forms in the shape of *ka-ROOT-an* are used for both Patient-Locative and Location-Locative (to be illustrated below). It turns out that this is so only because of some data gaps in the literature. Finally, Table 5.6 is not inclusive enough in that only verbs with explicit AF markers (such as *mi-*, *ma-*, and *<um>*) are shown to have corresponding CF and LF forms. In fact, verbs infixed with *<um>* are just a subtype of Class III verbs (as in Table 5.1), many of which do not have explicit AF markers but nonetheless derive their CF and LF forms in the same manner as verbs infixed with *<um>*. All these points, some of which are inspired by Tsukida's (2008: 289) work on Fata'an Amis, are addressed one by one below.

To start with, the four Mstem forms in Table 5.2 are illustrated in (53), with the Topic NP underlined.¹⁷¹

¹⁷⁰ These are all called Locative because of the shared suffix *-an*, the most common function of which across Austronesian languages is to indicate the location of an event (Blust 2013: 394-395). In addition, shared marking of locations and affected patients is a widespread phenomenon in Formosan languages, which has been explained in terms of the conceptual contiguity between location and object (S. Huang 2005: 789).

(53) Central Amis (Fieldnotes)

- a. **mi-palu** ci=kacaw ci=panay-an i=lutuk
 AF.EXT-beat TOP.PSN=K. PSN=P.-UND LOC=mountain
 ‘Kacaw is beating Panay in the mountains.’ (based on J. Wu 2006: 73)
- b. **mi-ulah** ci=kacaw ci=panay-an i=taypak
 AF.EXT-like TOP.PSN=K. PSN=P.-UND LOC=Taipei
 ‘Kacaw is going to court Panay in Taipei.’ (based on J. Wu 2006: 293)
- c. **ma-palu** ni=kacaw ci=panay i=lutuk
 AF.INT-beat ACT.PSN=K. TOP.PSN=P. LOC=mountain
 ‘Panay got beat up in the mountains by Kacaw.’ (based on J. Wu 2006: 73)
- d. **ma-ulah** ci=kacaw ci=panay-an
 AF.INT-like TOP.PSN=K. PSN=P.-UND
 ‘Kacaw likes Panay.’ (based on J. Wu 2006: 83)

Among them, *ma-ulah* ‘AF.INT-like’ and *ma-palu* ‘AF.INT-beat’ are the so-called AF and PF *ma*-verbs respectively due to their different argument marking patterns. However, both forms derive their corresponding CA-words in the same manner as their Extrovert counterparts *mi-ulah* ‘AF.EXT-like’ and *mi-palu* ‘AF.EXT-beat’, that is, by prefixing Conveyance *sa-* to the Kstem, as illustrated in Table 5.7 below. In other words, Conveyance *sa-* prefixation is operated on whatever Kstems available, irrespective of the argument marking patterns of their corresponding Mstems.

¹⁷¹ These examples are only slightly different from J. Wu’s (2006), and all of them have been confirmed by my consultants.

Table 5.7: CF-words deriving from Class I & II AF-words in Central Amis¹⁷²

Verb classes	Action root (e.g. <i>palu</i> ‘beat’)	State root (e.g. <i>ulah</i> ‘like’)
	AF Mstem: <i>mi-palu</i> Kstem: <i>pi-palu</i>	Mstem: <i>mi-ulah</i> Kstem: <i>pi-ulah</i>
Class I Extrovert	CF <i>sa-pi-palu</i> (CA-K.EXT-beat) ‘{tool/reason} for A’s beating B’	<i>sa-pi-ulah</i> (CA-K.EXT-like) ‘{tool/reason} for A’s courting B’
	AF Mstem: <i>ma-palu</i> Kstem: <i>ka-palu</i>	Mstem: <i>ma-ulah</i> Kstem: <i>ka-ulah</i>
Class II Introvert	CF <i>sa-ka-palu</i> (CA-K.INT-beat) ‘reason for B’s getting beat up’	<i>sa-ka-ulah</i> (CA-K.INT-like) ‘reason for A’s liking B’

Note: Both A and B stand for event participants, which are encoded as the non-Topic Actor and Undergoer respectively.

Moreover, semantic interpretations of CF-words also support a semantically coherent category of Introvert *ma*-verbs as opposed to splitting them into AF, PF, and Neutral. Specifically, while CF-words based on Extrovert verbs are interpreted as either instrument/means or reason/cause, those built on Introvert ones only denote reason/cause, as contrasted in Table 5.7 and further illustrated in (54).

¹⁷² CF forms in this table are translated with the meanings they would have when acting as arguments in contexts like (54). In addition to forms listed here, M. Chang (2007: 44) also mentioned those in the shape of *sa*-ROOT and *sa-ki*-ROOT, which do not seem to be quite as productive and thus fall outside the current discussion. Additionally, there are also forms in the shape of *sa*-CVCV~ROOT, but not much is known about them at this stage. One example I came across is *sa-lica~licay* ‘CA-IPFV~ask’, which means “any questions” (cf. *mi-licay* ‘AF.EXT-ask’; J. Wu 2006: 115).

(54) Central Amis (Fieldnotes)

- a. u=maan ku=[**sa-pi-palu** ni=kacaw ci=panay-an]
 CMN=what TOP.CMN=CA-K.EXT-beat ACT.PSN=K. PSN=P.-UND
 ‘What is the tool that Kacaw used to beat Panay?’
 ‘What is the reason why Kacaw beat Panay?’
- b. u=maan ku=[**sa-pi-ulah** ni=kacaw ci=panay-an]
 CMN=what TOP.CMN=CA-K.EXT-like ACT.PSN=K. PSN=P.-UND
 ‘What is Kacaw’s means of courting Panay?’
 ‘What is the reason why Kacaw is courting Panay?’
- c. u=maan ku=[**sa-ka-palu** ni=panay]
 CMN=what TOP.CMN=CA-K.INT-beat UND.PSN=P.
 ‘What is the reason why Panay got beat up?’
- d. u=maan ku=[**sa-ka-ulah** ni=kacaw ci=panay-an]
 CMN=what TOP.CMN=CA-K.INT-like ACT.PSN=K. PSN=P.-UND
 ‘What is the reason why Kacaw likes Panay?’

These preferred interpretations are in line with J. Wu’s (2006: 417) observation, and they are well-motivated if we recognize an overall semantic contrast between Extrovert *mi*-verbs and Introvert *ma*-verbs. Since Extrovert verbs highlight the application of force dynamics beyond the personal sphere of the Initiator, their CF forms can denote the external source of force, that is, the instrument/means for carrying out certain action, just as easily as the internal one, namely, the reason/cause for carrying out certain action. By contrast, since Introvert verbs contain the force dynamics (Talmy 2000: 409) within the personal sphere of the Experiencer, their CF forms are then limited to denote only the internal motivation of force dynamics.

It should also be pointed out that while the *ni*-phrase of the Introvert verb *ma-palu* ‘AF.INT-beat’, as in (53)c, can only be the agent, but not the patient, of a beating event, that of its CF form *sa-ka-palu*, as in (54)c, can only be the patient, but not the agent. Moreover, it seems impossible to integrate both the agent and patient phrase with the Introvert CF form, as shown in (55).

(55) Central Amis (Fieldnotes)

- a. u=maan ku=**sa-ka-palu** ni=kacaw
 CMN=what TOP.CMN=CA-K.INT-beat UND.PSN=K.
 ‘What is the reason why Kacaw got beat up?’ (cf. (54)c)
- b.* u=maan ku=**sa-ka-palu** ni=kacaw ni=panay
 CMN=what TOP.CMN=CA-K.INT-beat UND.PSN=K. ACT.PSN=P.

Without considering (55), one would be tempted to conclude that the *ni*-phrase of CA forms always encodes the agentive argument. But it turns out that the *ni*-phrase of CA forms consistently matches the Topic NP of their corresponding Mstem, be it agentive or patientive (cf. the underlined Topic NPs in (53) and the NPs marked by *ni* in (54) above). Thus, it is better to think of Introvert *ma*-verbs, with action and state roots alike, as one coherent grammatical class, the Topic NP of which is mapped to the *ni*-phrase in their corresponding CF-words.

Compared with CF-words, LF-words are more complicated because the latter can be derived from two types of stems. However, like CF-words, LF-words are derivable from various types of Introvert *ma*-verbs (be it labeled as AF, PF, or Neutral), just as they are from Extrovert *mi*-verbs. This is to be illustrated by the four pairs of AF Mstem/Kstem in Table 5.7, starting with those involving the Extrovert *mi*-, whose corresponding LF forms are clear from previous studies, then moving on to those involving the Introvert *ma*-, whose LF forms call for some adjustments to Table 5.6.

First, the Extrovert Mstem *mi-palu* ‘AF.EXT-beat’ potentially takes a patientive argument and a locative adverbial, and when the verb is nominalized the two roles are respectively denoted by the Mstem-based LF form *mi-palu-an* and the Kstem-based LF form *pi-palu-an*, as illustrated in (56).¹⁷³

¹⁷³ To facilitate comparisons between the Mstem and its LF forms, examples in (53) are repeated over the (a) examples in (56) through (59).

(56) Central Amis (Fieldnotes)

- a. **mi-palu** ci=kacaw ci=panay-an i=lutuk
 AF.EXT-beat TOP.PSN=K. PSN=P.-UND LOC=mountain
 ‘Kacaw is beating Panay in the mountains.’ (Based on J. Wu 2006: 73)
- b. ci=panay ku=**mi-palu-an** ni=kacaw i=lutuk
 PSN=P. TOP.CMN=AF.EXT-beat-LA ACT.PSN=K. LOC=mountain
 ‘The one who Kacaw beat in the mountains is Panay.’
- c. u=ya lutuk ku=**pi-palu-an** ni=kacaw ci=panay-an
 CMN=DIST mountain TOP.CMN=K.EXT-beat-LA ACT.PSN=K. PSN=P.-UND
 ‘The place where Kacaw beat Panay is on that mountain.’

Next, the Extrovert Mstem *mi-ulah* ‘AF.EXT-like’ demonstrates a similar pattern, as in (57).

(57) Central Amis (Fieldnotes)

- a. **mi-ulah** ci=kacaw ci=panay-an i=taypak
 AF.EXT-like TOP.PSN=K. PSN=P.-UND LOC=Taipei
 ‘Kacaw is going to court Panay in Taipei.’ (Based on J. Wu 2006: 293)
- b. ci=panay ku=**mi-ulah-an** ni=kacaw i=taypak
 PSN=P. TOP.CMN=AF.EXT-like-LA ACT.PSN=K. LOC=Taipei
 ‘The one who Kacaw courted in Taipei is Panay.’
- c. u=taypak ku=**pi-ulah-an** ni=kacaw ci=panay-an
 CMN=Taipei TOP.CMN=K.EXT-like-LA ACT.PSN=K. PSN=P.-UND
 ‘The place where Kacaw courted Panay is Taipei.’

Thus, as far as Extrovert *mi*-verbs are concerned, the Mstem- and Kstem-based LF forms respectively denote a patient-like argument and a location, hence J. Wu’s (2006, 2007) terms Patient-Locative and Location-Locative in Table 5.6.

Similarly, one Introvert *ma*-verb may potentially produce two LF forms, one of which is not covered in Table 5.6. Whether a given *ma*-verb permits an LF form built from the Mstem is subject to variations across roots, which, however, do not align well with the putative differences among AF, PF, or Neutral *ma*-verbs.

The Introvert Mstem *ma-palu* ‘AF.INT-beat’ potentially takes an agent-like argument and a locative adverbial, and the locative argument is denoted by the Kstem-based LF form when the verb is nominalized, as in (58).

(58) Central Amis (Fieldnotes)

- a. **ma-palu** ni=kacaw ci=panay i=lutuk
 AF.INT-beat ACT.PSN=K. TOP.PSN=P. LOC=mountain
 ‘Panay got beat up in the mountains by Kacaw.’ (based on J. Wu 2006: 73)
- b. u=ya lutuk ku=**ka-palu-an** ni=panay
 CMN=DIST mountain TOP.CMN=K.INT-beat-LA UND.PSN=P.
 ‘The place where Panay got beat up is on that mountain.’

However, the Mstem-based LA form **ma-palu-an* is not acceptable, either by itself or in a context, a point to be taken up later.

By contrast, the Introvert Mstem *ma-ulah* ‘AF.INT-like’ potentially takes a patient-like argument (i.e. the stimulus of emotion), and when the verb is nominalized that same role is denoted by either the Mstem-based LF *ma-ulah-an* or the Kstem-based LF *ka-ulah-an*, as in (59).

(59) Central Amis (Fieldnotes)

- a. **ma-ulah** ci=kacaw ci=panay-an
 AF.INT-like TOP.PSN=K. PSN=P.-UND
 ‘Kacaw likes Panay (a woman).’ (based on J. Wu 2006: 83)
- b. ci=panay ku=**ma-ulah-an** ni=kacaw
 PSN=P. TOP.CMN=AF.INT-like-LA ACT.PSN=K.
 ‘The one who Kacaw once loved is Panay.’
- c. ci=panay ku=**ka-ulah-an** ni=kacaw
 PSN=P. TOP.CMN=K.INT-like-LA ACT.PSN=K.
 ‘The one who Kacaw likes most (i.e. his favorite one) is Panay.’

Recall that in Table 5.6 J. Wu (2006: 113, 2007: 116) distinguished Patient-Locative in the form of *mi-ROOT-an* from Location-Locative in the form of *pi-ROOT-an* among Class I *mi*-verbs (as in (57) above), but only identified the LF form *ka-ROOT-an* for both functions among Class II *ma*-verbs. J. Wu (2006: 421) pointed out two implicational

indicators by which one could tell apart the Patient-Locative reading from the Location-Locative reading of *ka-ROOT-an*. First, if a *ma*-verb is dyadic (as evidenced by its taking a non-Topic Undergoer NP), then its LA form *ka-ROOT-an* illustrates Patient-Locative, rather than Location-Locative. The example she gave was precisely *ka-ulah-an*, which denotes someone's favorite person as in (59)c. However, this observation does not apply to *ma-radiw* 'AF.INT-sing', whose Kstem-based LA form *ka-radiw-an* exemplifies Location-Locative (i.e. the location for singing) rather than Patient-Locative (i.e. what is sung) (examples shown in (66) below). The other indicator she pointed out is that if a verb has an LF form in the shape of *mi-ROOT-an*, then its LF form in the shape of *ka-ROOT-an* is likely to be Location-Locative. An immediate counterexample to this rule of thumb is illustrated by the Extrovert AF *mi-ulah* 'AF.EXT-like'. It has the LA form *mi-ulah-an* as in (57)b, but its LA form *ka-ulah-an* was judged by her to be Patient-Locative instead of Location-Locative. Thus, neither of her indicators works very well.

J. Wu's (2006, 2007) decision to conflate Patient-Locative and Location-Locative, which are consistently expressed by *mi-ROOT-an* and *pi-ROOT-an* respectively among Class I *mi*-verbs, onto the single LF form *ka-ROOT-an* among Class II *ma*-verbs is presumably due to the lack of LF forms in the shape of *ma-ROOT-an* in her data, which would correspond well with its Class I counterpart *mi-ROOT-an*. With Mstem-based LF forms like *ma-ulah-an* 'AF.INT-like-LA' brought into the picture, it is suggested here that whatever semantic contrast holds between *mi-ROOT-an* and *pi-ROOT-an* among Class I *mi*-verbs is generally reflected between *ma-ROOT-an* and *ka-ROOT-an* among Class II *ma*-verbs. For one thing, like Class I *mi-ROOT-an*, Class II *ma-ROOT-an* consistently denotes the entity acted upon, hence Patient-Locative. For another, like Class I *pi-ROOT-an*, Class

II *ka*-ROOT-*an* denotes a location when combined with an action root (e.g. *ka-palu-an* ‘where one got beat up’ in (58)b above) or an activity root (e.g. *ka-radiw-an* ‘where one sings well’ in (66)d below). It is only when *ka*-ROOT-*an* contains a state root like *ulah* ‘like’ that the result form *ka-ulah-an* ‘someone’s favorite’ seem to fail to denote a location, as in (59)c. Nevertheless, aside from locations, Kstem-based LF forms (both Class I *pi*-ROOT-*an* and Class II *ka*-ROOT-*an*) denote temporal durations, which J. Wu (2006: 423) already treated as a subtype of Location-Locative. Given the spatio-temporal link found in Kstem-based LF forms, we might as well consider the Kstem-based LF form *ka-ulah-an* as the metaphorical extension of Location-Locative, whereby an essentially locative form is used to express the testimony or manifestation of an abstract state. This is somewhat motivated by the fact that abstract states, unlike actions and activities, generally do not take place within a spatio-temporal domain. Thus, the “location” of states would be their testimony or manifestation in the real world. Another example in the shape of *ka*-ROOT-*an* with a state root is *ka-laluk-an* ‘K.INT-diligent-LA’ (cf. the Mstem *ma-laluk* ‘AF.INT-diligent’), which can potentially denote the testimony of someone’s diligence (e.g. working hard in the field). Meanwhile, unlike *ma-ulah-an* ‘the one once loved’, the Mstem-based LF **ma-laluk-an* is not a legitimate form, which is presumably because the Mstem *ma-laluk* does not take any patient-like arguments, unlike the Mstem *ma-ulah* ‘AF.INT-like’ (see (59)a). Therefore, in the current analysis, we arrive at the generalization that Mstem-based LF forms, whether Class I *mi*-ROOT-*an* or Class II *ma*-ROOT-*an*, are used for Patient-Locative (denoting various patient-like arguments) whereas Kstem-based LF forms, whether Class I *pi*-ROOT-*an* or Class II *ka*-ROOT-*an*, are

reserved for Location-Locative (denoting the location or time of an action or activity as well as manifestations of a state).

Interestingly, the Mstem-based LF form *ma-ulah-an* is associated with a past-state episodic reading while the Kstem-based LF form *ka-ulah-an* with a current-state dispositional interpretation, as has been shown in (59) above.¹⁷⁴ The same semantic contrast obtains when the stimulus of emotion is a city instead of a person, as in (60).

(60) Central Amis (Fieldnotes)

- a. **ma-ulah** ci=kacaw tu=pusong
 AF.INT-like TOP.PSN=K. UND.CMN=Taitung
 ‘Kacaw likes Taitung (a city).’ (based on J. Wu 2006: 83)
- b. u=pusong ku=**ma-ulah-an** ni=kacaw
 CMN=Taitung TOP.CMN=AF.INT-like-LA ACT.PSN=K.
 ‘What Kacaw once loved is Taitung.’
- c. u=pusong ku=**ka-ulah-an** ni=kacaw
 CMN=Taitung TOP.CMN=K.INT-like-LA ACT.PSN=K.
 ‘What Kacaw likes most (i.e. his favorite place) is Taitung.’

However, the temporal contrast observed here is not generalizable to other Introvert LF forms.¹⁷⁵ More importantly, the fact that an LF form can be built from the Mstem *ma-ulah* (with a state root) but not from the Mstem *ma-palu* (with an action root) is not correlated with the putative difference between AF and PF *ma*-verbs. Both points are to be illustrated with the activity root *radiw* ‘sing’ below. Meanwhile, the various LF forms with the action root *palu* ‘beat’ and the state root *ulah* ‘like’ are summarized in Table 5.8.

¹⁷⁴ In addition, my consultants consistently gave a superlative reading for *ka-ulah-an* but not for *ma-ulah-an*. There is also one instance of *ka-ulah-an* in J. Wu (2006: 347) where she translated it as “the one who somebody likes most.”

¹⁷⁵ It is not quite clear what may cause such a temporal contrast.

Table 5.8: LF-words deriving from Class I & II AF-words in Central Amis (I)¹⁷⁶

Verb classes	Stem types	Action root (e.g. <i>palu</i> ‘beat’)	State root (e.g. <i>ulah</i> ‘like’)
Extrovert (Class I)	Mstem-based	<i>mi-palu-an</i> (AF.EXT-beat-LA) ‘the one that A beats’	<i>mi-ulah-an</i> (AF.EXT-like-LA) ‘the one that A courts’
	Kstem-based	<i>pi-palu-an</i> (K.EXT-beat-LA) ‘where A beats B’	<i>pi-ulah-an</i> (K.EXT-like-LA) ‘where A courts B’
Introvert (Class II)	Mstem-based	* <i>ma-palu-an</i> (AF.INT-beat-LA)	<i>ma-ulah-an</i> (AF.INT-like-LA) ‘what A once liked’
	Kstem-based	<i>ka-palu-an</i> (K.INT-beat-LA) ‘where B gets beat up’	<i>ka-ulah-an</i> (K.INT-like-LA) ‘what A likes best’

Note: Both A and B stand for event participants, which are encoded as the non-Topic Actor and Undergoer respectively.

Based on its argument marking patterns, J. Wu (2006: 189, 2007: 136) analyzed the Mstem *ma-radiw* ‘AF.INT-sing’ to be either a Neutral *ma*-verb or a PF one, as respectively shown in the (a) and (b) example of (61).

¹⁷⁶ LF forms in this table are translated with the meanings they would have when acting as arguments. Tsukida (2008: 289) reported LF forms only in the shape of *pi*-ROOT-*an* and *ka*-ROOT-*an* (i.e. Kstem-based), and J. Wu (2006: 113, 2007: 116) added those in the shape of *mi*-ROOT-*an*, but not of *ma*-ROOT-*an*. This current table shows that the Mstem in general, whether Extrovert *mi*-verbs or Introvert *ma*-verbs, can undergo *-an* suffixation, although whether a given Mstem-based LA form is acceptable varies from one root to another.

(61) Central Amis (J. Wu 2006: 189, 2007: 136)

- a. **ma-radiw** ku=ra kaying
 AF.INT-sing TOP.CMN=MED young.lady
 ‘That young lady sings (well) (i.e. she is good at singing).’
- b. **ma-radiw** ni=aki ku=radiw aku
 AF.INT-sing ACT.PSN=A. TOP.CMN=song 1SG.GEN
 ‘My song was sung by Aki.’

However, the Mstem *ma-radiw* has both Mstem- and Kstem-based LA forms just like its Extrovert counterpart *mi-radiw*. In other words, like the state root *ulah* ‘like’ presented above, the activity root *radiw* ‘sing’ in combination with two AF prefixes produces four LA forms in total, each of which is illustrated below.

The Extrovert *mi-radiw* potentially takes musical compositions (e.g. songs, tunes, hymns) as the patient-like argument, and that same role is denoted by the Mstem-based LF form, but not the Kstem-based one, which is only good for the location of a singing event, as in (62).

(62) Central Amis (Fieldnotes)

- a. **mi-radiw** ci=panay tu=ya ’ulic
 AF.EXT-sing TOP.PSN=P. UND.CMN=DIST hymn
 ‘Panay is singing that hymn.’
- b. u=ya ’ulic ku=**mi-radiw-an** ni=panay
 CMN=DIST hymn TOP.CMN=AF.EXT-sing-LA ACT.PSN=P.
 ‘What Panay is singing is that hymn.’
- c.* u=ya ’ulic ku=**pi-radiw-an** ni=panay
 CMN=DIST hymn TOP.CMN=K.EXT-sing-LA ACT.PSN=P.
- d. u=ya kyokay ku=**pi-radiw-an** ni=panay
 CMN=DIST church TOP.CMN=K.EXT-sing-LA ACT.PSN=P.
 ‘The place where Panay sings is that church.’

Moreover, the patient-like argument of the Extrovert *mi-radiw* can alternatively be something that one praises or proclaims through singing, such as the church. In this case then, the Mstem-based LF form denotes the praised entity, as in (63).

(63) Central Amis (Fieldnotes)

- a. **mi-radiw** ci=panay tu=ya kyokay
 AF.EXT-sing TOP.PSN=P. UND.CMN=DIST church
 ‘Panay is singing of that church.’
- b. u=ya kyokay ku=**mi-radiw-an** ni=panay
 CMN=DIST church TOP.CMN=AF.EXT-sing-LA ACT.PSN=P.
 ‘What Panay sang of is that church.’ (cf.(62)d)

Thus, from (62) and (63) we see that the Kstem-based LF form consistently denotes the location of a singing event whereas its Mstem-based counterpart may refer to what is sung or what is sung of, both roles are marked as the non-Topic Undergoer argument of their corresponding Mstem.

Moreover, there is yet another possible interpretation for the Mstem-based LF form *mi-radiw-an*, which J. Wu (2007) dubbed as Goal-Locative, as opposed to Patient-Locative. The two interpretations are illustrated in (64).

(64) Central Amis (J. Wu 2006: 424, 2007: 120)

- mi-radiw-an** aku ku=ni
 AF.EXT-sing-LA 1SG.ACT TOP.CMN=PROX
 ‘I sang this.’ OR ‘This is what I sang.’ (Patient-Locative) (cf. (62)b)
 ‘I sang for (getting) this.’ OR ‘This is what I got by singing.’ (Goal-Locative)

Given these two labels, what should we make of the same LF form *mi-radiw-an* in (63)b, where it denotes what is sung of? Should it be an instance of Patient-Locative or Goal-Locative? Neither seems to be appropriate. In fact, this particular LF form is at least polysemous over three readings: what is sung (i.e. musical compositions), what is sung of (i.e. things or people that one praises by singing), or things that one obtains by singing. More readings might come up if we keep searching. Thus, there seems to be no strong motivation to single out Goal-Locative from Patient-Locative, which is already a cover term for various patient-like arguments anyway (e.g. the transported theme or recipient of object transferral) under J. Wu’s (2006, 2007) analysis. Besides, not many examples of

Goal-Locative were found in her works. The only one other than (64) is (65), where the Mstem-based LF form *mi-cikay-an* ‘AF.EXT-run-LA’ expresses the goal of someone’s motion.

(65) Central Amis (J. Wu 2006: 424)

mi-cikay-an aku i=pitilidan ku=ni a ciudad
 AF.EXT-run-LA 1SG.ACT LOC=school TOP.CMN=PROX LIG book
 ‘This book is what I ran to school to get.’

However, in this example again, there is an alternative interpretation (offered by my Amis consultants), which reads “This book is what I got (as a prize) by running (a race) at school.” Therefore, considering the Mstem-based LF form is typically polysemous over various patient-like arguments, we might as well see the Goal-Locative reading as an instance of the Patient-Locative reading (or the other way around). And it seems that if a given Mstem allows a non-Topic Undergoer in its argument structure, then its Mstem-based LF form is permitted to denote that particular argument. If so, we would have a minimal pair of nominalized forms, with the Mstem-*ay* denoting the Actor-Topic of an Mstem (§5.2.1) and the Mstem-*an* denoting the non-Topic Undergoer of the same Mstem.

Like its Extrovert counterpart, the Introvert Mstem *ma-radiw* ‘AF.INT-sing’ potentially takes musical compositions (e.g. songs, tunes, hymns) as its non-Topic Undergoer NP. As a result, that very same Undergoer argument is denoted by the Mstem-based LF form *ma-radiw-an* while the Kstem-based LF form denotes the location of a singing event, as in (66) below.

(66) Central Amis (Fieldnotes)

- a. **ma-radiw** ci=panay tu=ya 'ulic
 AF.INT-sing TOP.PSN=P. UND.CMN=DIST hymn
 'Panay is good at singing that hymn.'
- b. u=ya 'ulic ku=**ma-radiw-an** ni=panay
 CMN=DIST hymn TOP.CMN=AF.INT-sing-LA ACT.PSN=P.
 'What was sung by Panay is that hymn.'
- c.* u=ya 'ulic ku=**ka-radiw-an** ni=panay
 CMN=DIST hymn TOP.CMN=K.INT-sing-LA ACT.PSN=P.
- d. u=ya kyokay ku=**ka-radiw-an** ni=panay
 CMN=DIST church TOP.CMN=K.INT-sing-LA ACT.PSN=P.
 'The place where Panay sings well is that church.'

Although both the Extrovert *mi-radiw* and the Introvert *ma-radiw* allow musical compositions to be their patient-like argument, the former emphasizes that someone is engaged in a singing activity while the latter highlights someone's ability to sing. This once again illustrates the semantic contrast between Class I Extrovert and Class II Introvert verbs (cf. the (a) example in (62) above and (66) here).

Table 5.9 below summarizes the four LF forms based on the activity root *radiw* 'sing', all of which have been illustrated above.

Now we turn to the CF and LF forms of Class III AF verbs. Like Class I and II, Class III verbs derive their corresponding CF forms by prefixing *sa-* to the Kstem and LF forms by suffixing *-an* to the Mstem or Kstem. This is so regardless of whether the Mstem in Class III has the (historically AF) infix <um>. An often cited example of verbs with <um> is *k<um>aen* 'AF.<UM>eat', the Kstem of which is *ka-k<um>aen*, as in (67).

Table 5.9: LF-words deriving from Class I & II AF-words in Central Amis (II)¹⁷⁷

Prefix types	Stem types	Activity root (e.g. <i>radiw</i> ‘sing’)
Extrovert (Class I)	Mstem-based	<i>mi-radiw-an</i> (AF.EXT-sing-LA) ‘what A sings (of), or what A gets by singing’
	Kstem-based	<i>pi-radiw-an</i> (K.EXT-sing-LA) ‘where A sings B’
Introvert (Class II)	Mstem-based	<i>ma-radiw-an</i> (AF.INT-sing-LA) ‘what is sung by A’
	Kstem-based	<i>ka-radiw-an</i> (K.INT-sing-LA) ‘where A sings well’

Note: Both A and B stand for event participants, which are encoded as the non-Topic Actor and Undergoer respectively.

(67) Central Amis (Fieldnotes)

- a. **k<um>aen** kaku tu=pawli
 AF.<UM>eat 1SG.TOP UND.CMN=banana
 ‘I am eating a banana.’ (J. Wu 2006: 118)
- b. **caay ka-k<um>aen** kaku tu=pawli
 NEG K-<UM>eat 1SG.TOP UND.CMN=banana
 ‘I am not eating any banana.’

Thus, its CF form is *sa-ka-k<um>aen*, which denotes the instrument for eating, and its Mstem- and Kstem-based LA forms are *k<um>aen-an* and *ka-k<um>aen-an* respectively, which denote the food eaten and the location for eating in that order, as shown in (68) below.

¹⁷⁷ LF forms in this table are translated with the meanings they would have when acting as arguments.

(68) Central Amis

- a. u=ni alapit ku=**sa-ka-k<um>aen** ni=aki tu=futing
 CMN=PROX chopstick TOP.CMN=CA-K-<UM>eat ACT.PSN=A. UND.CMN=fish
 ‘What Aki used to eat fish is this (pair of) chopsticks.’
 (Fieldnotes; based on J. Wu 2006: 75)
- b. taiih ku=ya **k<um>aen-an** ni=aki a tali
 AF.bad TOP.CMN=DIST <UM>eat-LA ACT.PSN=A. LIG taro
 ‘That taro which Aki ate tasted bad.’ (J. Wu 2006: 116)
- c. u=ra lutuk ku=**ka-k<um>aen-an** ni=aki tu=tali
 CMN=MED mountain TOP.CMN=K-<UM>eat-LA ACT.PSN=A. UND.CMN=taro
 ‘(The place) where Aki ate taros is on that mountain.’ (Fieldnotes)

Likewise, the denominal Mstem *ci-wawa* ‘AF.have-child’, despite not having any explicit AF markers, derives its Kstem in the same manner as *k<um>aen* ‘<UM>eat’, that is, by prefixing *ka-* to the Mstem (a pattern that defines Class III verbs to begin with; see Table 5.1). The Mstem and Kstem of this denominal verb are illustrated in (69).

(69) Central Amis (Fieldnotes)

- a. **ci-wawa** ci=panay
 AF.have-child TOP.PSN=P.
 ‘Panay has {a child/children}.’
- b. caay **ka-ci-wawa** ci=panay
 NEG K-have-child TOP.PSN=P.
 ‘Panay does not have any child.’

Following the patterns shown above, its CF and Kstem-based LF form are respectively *sa-ka-ci-wawa* and *ka-ci-wawa-an*, as in (70).

(70) Central Amis (Fieldnotes)

- a. u=maan ku=**sa-ka-ci-wawa** ni=panay
 CMN=what TOP.CMN=CA-K-have-child ACT.PSN=P.
 ‘What is the reason that Panay give birth to a baby?’
- b. i=cuwa ku=**ka-ci-wawa-an** ni=panay
 AF.LOC=where TOP.CMN=K-have-child-LA ACT.PSN=P.
 ‘Where is the place Panay gave birth to a baby?’

The Mstem-based LF form **ci-wawa-an* is not acceptable, either by itself or in a context.

This is rather expected since *ci-wawa* ‘AF.have-child’ is a monadic predicate that does not

allow non-Topic Undergoer NPs in its argument structure, unlike *k<um>aen* ‘AF.<UM>eat’, which does so and thus permits the Mstem-based LF form (cf. (67) and (68)b).

Table 5.10 summarizes the CF and LF forms of the two Class III AF verbs illustrated above.

Table 5.10: CF- and LF-words deriving from Class III AF-words in Central Amis¹⁷⁸

	Stem types	CF	LF
Subtype I	Mstem-based	---	<i>k<um>aen-an</i> (<UM>eat-LA) ‘A’s eaten thing’
	Kstem-based	<i>sa-ka-k<um>aen</i> (CA-K-<UM>eat) ‘instrument for A’s eating B’	<i>ka-k<um>aen-an</i> (K-<UM>eat-LA) ‘location for A’s eating B’
Subtype II	Mstem-based	---	* <i>ci-wawa-an</i> (HAVE-child-LA)
	Kstem-based	<i>sa-ka-ci-wawa</i> (CA-K-HAVE-child) ‘reason for A’s giving birth’	<i>ka-ci-wawa-an</i> (K-HAVE-child-LA) ‘location for A’s giving birth’

Note: Both A and B stand for event participants, which are encoded as the non-Topic Actor and Undergoer respectively.

Finally, generalized over Table 5.7 through Table 5.10, Table 5.11 below summarizes the CF and LF forms across Class I through III AF verbs in Central Amis.

¹⁷⁸ CA and LA forms in this table are translated with the meanings they would have when acting as denoting expressions.

Table 5.11: CF and LF forms in Central Amis¹⁷⁹

AF verb classes	Stem types	CF	LF
Class I Extrovert	Mstem-based	---	{ <i>mi</i> -ROOT}- <i>an</i>
	Kstem-based	<i>sa</i> -{ <i>pi</i> -ROOT}	{ <i>pi</i> -ROOT}- <i>an</i>
Class II Introvert	Mstem-based	---	{ <i>ma</i> -ROOT}- <i>an</i> [†]
	Kstem-based	<i>sa</i> -{ <i>ka</i> -ROOT}	{ <i>ka</i> -ROOT}- <i>an</i>
Class III Others	Mstem-based	---	{ <i>X</i> }- <i>an</i> [†]
	Kstem-based	<i>sa</i> -{ <i>ka</i> - <i>X</i> }	{ <i>ka</i> - <i>X</i> }- <i>an</i>

Note: Curly brackets indicate the Mstem or Kstem. A superscript dagger means the relevant form might not always be acceptable and is subject to cross-root variations. X in Class III is a schematic symbol for whatever form the Mstem in that class may assume, be it unaffixed roots or multi-affixed stems.

Some recapitulations in this section are as follows. First, all forms in Table 5.11 have the potential to take up typical argument positions, where CF forms denote either the internal source of force dynamics (i.e. reason/cause) or/and the external one (i.e. instrument/means), and Mstem- and Kstem-based LF forms denote Patient-Locative and Location-Locative respectively. Second, Patient-Locative targets at patient-like arguments permitted in the argument structure of the Mstem, from which Patient-Locative is derived, whereas Location-Locative encompasses the spatio-temporal aspect of actions or activities as well as external manifestations of abstract states. Third, in terms of participant encoding, two generalizations are found. For one thing, the Topic NP of an Mstem, be it agentive or patientive, is always mapped to the *ni*-phrase (or its

¹⁷⁹ ROOT in this table can be generalized to include morphologically more complex stems. See Footnote 184 for examples.

paradigmatic counterparts) of its corresponding CF/LF form. For the other, non-Topic Undergoer arguments are compatible with the Kstem-based Location-Locative, but not with the Mstem-based Patient-Locative. This is somewhat expected considering the latter already denotes patient-like affected participants.

5.2.2.2. Syntactic functions

It has been shown in the previous section that Circumstantial nominalizations consist of minimally a CF/LF form and additionally its optional event participants as well as spatio-temporal adverbials. As far as the argument function is concerned, the two forms show no syntactic differences in their interactions with higher-order syntax.

However, asymmetries between CF and LF forms do exist when it comes to the predicate function, especially in the negative context. Recall that nominal predicates made up of common nouns are preceded by *u* in the affirmative context and by *ku* in the negative one (see (14) above), which applies to CF/LF (in fact all) nominalizations as well. Nevertheless, the marker *u* in the affirmative context is sometimes left out, from underived nouns and Focus nominalizations alike. Such examples of Actor nominalizations have been given in §5.2.1.1, and additional examples of CF/LF forms are demonstrated in (71).

(71) Central Amis (Fieldnotes)

- a. **(u)=sa-pi-palu** ni=mayaw ci=dongi-an ku=ni a sasti'
 CMN=CA-K.EXT-beat ACT.PSN=M. PSN=D.-UND TOP.CMN=PROX LIG stick
 'This stick is what Mayaw beat Dongi with.'
 'Mayaw beat Dongi with this stick.' (See also J. Wu 2006: 136)
- b. **(u)=pi-palu-an** aku tu=wawa ku=luma'
 CMN=K.EXT-beat-LA 1SG.ACT UND.CMN=child TOP.CMN=house
 ni=mayaw
 GEN.PSN=M.
 'Mayaw's house is where I beat (my) child.'
 'I beat (my) child in Mayaw's house.'

When the prenominal marker *u* is present, there is good evidence that CF/LF forms are just as nominal as underived nouns. When it is absent, however, there is little decisive evidence for whether CF/LF forms are showing the verbal or nominal use. And translations (with two versions given for each example) are not quite helpful either because what is at issue is the syntactic status of CF/LF forms, not their semantics.¹⁸⁰ Consequently, CF/LF forms have generated disagreeing conclusions in the literature. D. Liu (1999: 55), for instance, assumed that their predicate function without the *u* marker, as in (71), is verbal. J. Wu (2006, 2007), however, judged the same distributional fact to be nominal instead. Under the nominal analysis, CF/LF forms in (71) are essentially nominal predicates, and this is motivated by the fact that they are denoting expressions elsewhere that can easily take up argument positions, a property not shared by the AF-word Mstem or PF-word ROOT-*en*, both of which are highly verbal. Thus, considering highly nominal structures (e.g. underived nouns) can dispense with the prenominal marker *u* and that highly verbal ones (e.g. AF Mstem) cannot be preceded by it, it seems much simpler to uphold, as J. Wu (2006, 2007) did, the nominal analysis for CF/LF forms, even when they function as predicates.

¹⁸⁰ Similar analytical problems with predicates and predicate phrases in Tagalog are discussed by Sabbagh (2009).

In spite of embracing the nominal analysis for CF/LF forms, J. Wu (2006: 136-137) did identify a context where CF but not LF forms demonstrate a verbal pattern, which has to do with negative predication. Specifically, although both CF and LF forms can be negated by the nominal pattern, only the former, but not the latter, can be alternatively negated by the verbal pattern, as contrasted in (72) and (73).

(72) Central Amis (J. Wu 2006: 136, 137)¹⁸¹

a. **caay ku=sa-pi-palu** ni=mayaw ci=dongi-an
 NEG TOP.CMN=CA-K.EXT-beat ACT.PSN=M. PSN=D.-UND

ku=ni a sasti'
 TOP.CMN=PROX LIG stick

‘This stick is not what Mayaw beat Dongi with.’

b. **caay ka-sa-pi-palu** ni=mayaw ci=dongi-an
 NEG K.INT-CA-K.EXT-beat ACT.PSN=M. PSN=D.-UND

ku=ni a sasti'
 TOP.CMN=PROX LIG stick

‘Mayaw didn’t use this stick to beat Dongi.’

(73) Central Amis (J. Wu 2006: 136, 137)

a. **caay ku=pi-palu-an** ni=mayaw
 NEG TOP.CMN=K.EXT-beat-LA ACT.PSN=M.

ci=dongi-an ku=ni anudafak
 PSN=D.-UND TOP.CMN=PROX tomorrow

‘This (place) is not where Mayaw is going to beat Dongi tomorrow.’

b.* **caay ka-pi-palu-an** ni=mayaw
 NEG K.INT-K.EXT-beat-LA ACT.PSN=M.

ci=dongi-an ku=ni anudafak
 PSN=D.-UND TOP.CMN=PROX tomorrow

As has been shown earlier, the nominal negative pattern involves negation external to nominalizations, whereby what is negated is the identificational/equational relationship

¹⁸¹ J. Wu (2006: 137) presented the verbal pattern in (72)b with confidence, but she (J. Wu 2007: 107) later did so with a caveat that the verbal type with CA forms “is rendered ungrammatical by some speakers.” However, my consultants have consistently confirmed the verbal type (with various roots in the form of *ka-sa-pi-ROOT* after the negator). Considering the fact the J. Wu’s consultants and mine are from two different villages (Ciwkangan [Amis]/Changuang [Chinese] and Kinaluka [Amis]/Ningpu [Chinese] respectively, both in Taitung County), there is good reason to believe that the verbal type is a widespread and stable pattern.

between two denoting expressions. Thus, given the nominal nature of CF/LF forms, the nominal negative pattern is quite expected. The verbal negative pattern, however, is not, and thus requires further inquiry, to be presented below.

First, the fact that the verbal negative pattern applies to CF but not LF forms is correlated with the fact that CF but not LF forms can be prefixed by the Class II Introvert *ma-*. To put it schematically, *ma-sa-Kstem* is a legitimate word form whereas *ma-Kstem-an* is not. Examples of *ma-sa-Kstem* are shown in (74).¹⁸²

(74) Central Amis (Fieldnotes)

- a. **ma-sa-pi-palu** ni=mayaw ci=dongi-an
 AF.INT-CA-K.EXT-beat ACT.PSN=M. PSN=D.-UND
 ku=ni a sasti'
 TOP.CMN=PROX LIG stick
 'Mayaw will beat Dongi with this stick.' (based on J. Wu 2006: 137)
- b. **ma-sa-pi-tukad** ni=ina tu=futing
 AF.INT-CA-K.EXT-slice ACT.PSN=mother UND.CMN=fish
 ku=ni a pu'ut
 TOP.CMN=PROX LIG knife
 'Mother will slice fish with this knife.' (based on J. Wu 2006: 134)

By contrast, neither of the *ma-Kstem-an* forms **ma-pi-palu-an* and **ma-pi-tukad-an* is acceptable. Recall that the verb form immediately following the negator has to be the Kstem, and that the Kstem of Class II *ma*-verbs is formed by replacing the prefix *ma-* with *ka-* (see Table 5.1 above). Given the Mstem *ma-sa-pi-palu* 'AF.INT-CA-K.EXT-beat' in (74)a, its Kstem is accordingly *ka-sa-pi-palu* 'K.INT-CA-K.EXT-beat', which is then the required form in the verbal negation pattern in (72)b. This is comparable to the fact that the morphologically simpler Mstem *ma-palu* 'AF.INT-beat' has to change to its Kstem *ka-palu* 'K.INT-beat' when negated by the verbal pattern (see (10) above). Thus, the verbal

¹⁸² The preferred interpretation for the two examples prompts a future-event reading, hence the free translations. The full range of temporal interpretations of *ma-CA* verbs awaits further studies since not many examples of them are found in the current literature. The original transcription for "slice (animals)" in J. Wu (2006: 134) is *tukas*, which is corrected here to *tukad* [tukaʔ].

pattern found in (72)b is in fact a property of the Mstem/Kstem pair *ma/ka-sa*-Kstem, rather than that of the CF form *sa*-Kstem.

Second, the process whereby a CF form, which has denoting functions, is prefixed by Class II Introvert *ma-* to derive an Mstem verb can be likened to the one whereby an entity-denoting root undergoes the same prefixation to derive an Mstem verb. Both processes are denominal in nature, turning a nominal into a verbal, and the only difference is whether the source nominal is a root or a derived stem like the CF form. Some examples of Mstems deriving from nominal roots are illustrated in (75).

(75) Central Amis (J. Wu 2006: 231-232)

- a. *kimad* ‘story’ > *ma-kimad* ‘tell stories’
- b. *duka* ‘wound’ > *ma-duka* ‘get wounded’
- c. *fali* ‘wind’ > *ma-fali* ‘be windy’

Once the Mstem/Kstem pair *ma/ka-sa*-Kstem is considered a denominal derivation from the CF form *sa*-Kstem, it becomes clear that the reason CF forms can be alternatively negated by the verbal pattern is that they have been verbalized after being prefixed by *ma/ka-*, much as do nominal roots in (75).

Third, given the unacceptability of (73)b, does that mean the verbal negation pattern is prohibited from LF forms altogether? It is found the answer is negative and that what makes (73)b problematic is the wrong Kstem form tested (i.e. **ka-pi-palu-an*, which is not even a legitimate form by itself). Recall that the Kstem starts with either *pi-* or *ka-* (see Table 5.1 above), that the LF form in (73) is essentially the Kstem suffixed by Locative *-an*, and that CF forms are the Kstem prefixed by Conveyance *sa-* (see Table 5.7 above). As a result, Kstem-based LF forms still start with either *pi-* or *ka-*, which may qualify them as the Kstem, whereas Kstem-based CF forms do not start with *pi-* or *ka-* any more, which makes CF forms unsuitable for the Kstem. The idea that the LF form by

itself is a legitimate Kstem that can immediately follow the negator is supported by an example with internal negation in (76), to be contrasted with the one illustrating external negation in (77).

(76) Central Amis (Fieldnotes)

udengan u=luma' ni=mayaw ku=[caay-ay pi-palu-an aku
 only CMN=house GEN.PSN=M. TOP.CMN=NEG-NMLZ K.EXT-beat-LA 1SG.ACT
tu=wawa]
 UND.CMN=child

'The place where I didn't beat (my) child is only in Mayaw's house.'

(77) Central Amis (Fieldnotes)

caay ku=[pi-palu-an aku tu=wawa]
 NEG TOP.CMN=K.EXT-beat-LA 1SG.ACT UND.CMN=child
ku=luma' ni=mayaw
 TOP.CMN=house GEN.PSN=M.

'Mayaw's house is not where I beat (my) child.'

The same LF form *pi-palu-an* 'K.EXT-beat-LA', together with its event participants, is negated by the verbal pattern in the first example, where negation occurs within a nominalization, but by the nominal pattern in the second, where negation falls outside a nominalization. Thus, once the wrong Kstem is fixed, LF forms do permit the verbal (as well as nominal) type of negation.

Therefore, syntactic asymmetries between CF and LF forms do exist, but not along the lines observed in J. Wu (2006, 2007), where CF forms were concluded to be both verbal and nominal whereas LF forms to be only nominal in nature. Instead, it has been shown in this section that (i) CF forms are nominal because they are negated by the nominal pattern; (ii) the verbal negation pattern CF forms seem to illustrate is a property of the Mstem/Kstem deriving from nominal CF stems and Class II AF verbal prefixes (*ma-* for the Mstem and *ka-* for the Kstem), a process that resembles deriving AF verbs

out of nominal roots; and (iii) Kstem-based LF forms show both nominal and verbal negation patterns.

Finally, Table 5.12 summarizes the morphosyntactic requirements of negation both external and internal to a Location-Locative nominalization consisting of the action root *palu* ‘beat’.

Table 5.12: External vs. internal negation of Locative nominalizations in Central Amis

	AFF	NEG
External		
Internal		
AFF	<i>u=pi-palu-an</i> (CMN=K.EXT-beat-LA) ‘where A beat B’ [(71)]	<i>caay ku=pi-palu-an</i> (NEG TOP.CMN=K.EXT-beat-LA) ‘not where A beat B’ [(73) and (77)]
NEG	<i>u=caay-ay pi-palu-an</i> (CMN=NEG-NMLZK.EXT-beat-LA) ‘where A didn’t beat B’ [(76)]	NO DATA

Note: AFF and NEG stand for affirmative and negative contexts respectively. A and B refer to event participants, which are encoded as the non-Topic Actor and Undergoer respectively. Examples with both internal and external nominalizations are logically possible, but no such data have been successfully elicited (probably due to some pragmatic conflicts). When a nominal is externally negated (the upper-right cell), the pronominal marker *ku* is required. In all the other contexts, the pronominal marker can be *u*, *ku*, or *tu*, depending on the grammatical relations of the NPs it marks. See the referenced examples for complete sentences.

The results in Table 5.12 are consistent with how Actor nominalizations are negated, both internally and externally (cf. Table 5.3). When nominalizations are internally affirmative, both the AF Mstem-*ay* and Kstem-based LF form are treated like underived nouns on the phrasal level, following the same syntax required of common nouns. Thus, if the Kstem-

based LF form is a nominal, so should the Mstem-*ay* be, which again reinforces the nominalizer analysis for the marker *-ay*. On the other hand, when nominalizations are internally negative but externally affirmative, the Kstem-based LF form shows a verbal property shared by the AF Kstem because both are the required forms immediately following the negator *caay*, which has to be marked by *-ay* in order to create nominalizations with internal negative predication.

5.3. Event nominalizations

Compared with the last section on argument nominalizations, this section on event nominalizations is relatively brief because the forms used to express the latter are much simpler than those for the former.

At least two forms are involved in event nominalizations. The simpler one is the root alone, which J. Wu (2006: 142) has concluded to be “syntactically nominal” in Amis. This is illustrated by *keru* ‘dance’ in the answer part of (78), where it serves as the sole argument of a Class III Mstem *fangcal* ‘AF.good’, to be contrasted with its Introvert Mstem *ma-keru*, which is the second predicate of a serial verb construction (J. Wu 1996a) in the question part.

(78) Central Amis (9 Level Textbooks: 3-6)

Q *ma-fana’ kisu a ma-keru=haw*
 AF.INT-know 2SG.TOP LIG AF.INT-dance=QP
 ‘Do you know how to dance?’

A *hay, fangcal ku=keru aku*
 yes AF.good TOP.CMN=dance 1SG.ACT
 ‘Yes. I dance well.’ (Lit. ‘My dancing is good.’)

The other form involved in event nominalizations is the Kstem, which, as has been demonstrated excessively, is the required form for the verbal type of negation.

Another function of the Kstem is to serve as the matrix predicate in imperative sentences, as shown in (79) through (81) for Class I through III AF Mstem/Kstem pairs respectively.

(79) Central Amis(J. Wu 2006: 138)

- a. **mi-canuy** kaku tu=safa
 AF.EXT-swing 1SG.TOP UND.CMN=younger.sibling
 ‘I’m going to swing (my) younger sibling.’
- b. **pi-canuy** tu=safa
 K.EXT-swing UND.CMN=younger.sibling
 ‘(Go to) swing (your) younger sibling!’

(80) Central Amis (J. Wu 2006: 138-139)

- a. **ma-tayal** kaku i=taypak
 AF.INT-work 1SG.TOP LOC=Taipei
 ‘I work in Taipei.’
- b. **ka-tayal** i=taypak
 K.INT-work LOC=Taipei
 ‘Work in Taipei!’

(81) Central Amis(J. Wu 2006: 139)

- a. **tayra** kaku i=taypak
 AF.go 1SG.TOP LOC=Taipei
 ‘I am going to Taipei.’
- b. **ka-tayra** i=taypak
 K-go LOC=Taipei
 ‘Go to Taipei!’

According to M. Chang (2007), if the addressee of AF imperatives, as in the (b) examples of (79) through (81), is to be expressed at all, it takes on the same marking as the Actor of their indicative counterparts. This is true of both affirmative and negative imperatives, as in (82).

(82) Central Amis (M. Chang 2007: 38, 52)

- a. ka-tayra **(kisu)** a dafak
 K-go 2SG.TOP LIG tomorrow
 ‘Go (there) tomorrow!’
- b. aka ka-tayra **(kisu)** i=taypak
 NEG.IMP K-go 2SG.TOP LOC=Taipei
 ‘Don’t go to Taipei!’

Meanwhile, the Kstem demonstrates nominal syntax when used to denote a state of affairs. This can be illustrated by comparative constructions, which have been extensively investigated by J. Kuo (2008). One of the comparative constructions in Amis makes use of the prefix *ki-*, which is attached to a predicate root that denotes scalar properties, and the *ki*-marked stem is additionally prefixed by Class I *mi-* or Class II *ma-*. In the case of Class I *mi-ki-ROOT*, the comparee and standard of comparison are encoded as the Topic and non-Topic respectively, as shown in (83), where the Kstem occurs in typical argument positions as underived nouns do.

(83) Central Amis (J. Kuo 2008: 57, 77)

- a. mi-ki-kereteng ku=widang aku cingranan
 AF.EXT-CMPR-heavy TOP.CMN=friend 1SG.GEN 3SG.OBL
 ‘My friend is heavier than him/her.’
- b. mi-ki-lihanaw ku=**pi-adup** tu=**ka-lingad**
 AF.EXT-CMPR-dangerous TOP.CMN=K.EXT-hunt OBL.CMN=K.INT-farm
 ‘Hunting is more dangerous than farming.’

In addition to arguments, event nominalizations expressed by the Kstem also function as matrix predicates, which are negated by the nominal pattern, as in (84).

(84) Central Amis (Fieldnotes)

- caay ku=**pi-salama** ku=**pi-adup**
 NEG TOP.CMN=K.EXT-play TOP.CMN=K.EXT-hunt
 ‘Hunting is not (children’s) game.’

As has been demonstrated earlier, the nominal type of negation denies the equational/identification relationship holding between two entities. This is in sharp contrast to the verbal type of negation, in which the Kstem also participates, as in (85).

(85) Central Amis (Fieldnotes)

caay **pi-salama** ku=ya wawa
 NEG K.EXT-play TOP.CMN=DIST child
 ‘That child is not playing.’

The Kstem now immediately follows the negator, and what is negated by the verbal type is the state of affairs expressed by the Kstem. Therefore, the Kstem allows both nominal and verbal uses, just as does its morphological derivative Kstem-*an* (i.e. the Kstem-based LF form) discussed in §5.2.2.2.

When expressing event nominalizations, the Kstem can be further expanded to incorporate event participants. For instance, the agent and patient of the Extrovert Kstem *pi-palu* ‘K.EXT-beat’ are respectively encoded as the non-Topic Actor and Undergoer, as in (86), to be contrasted with the non-nominalized context in (87), where the agent role is encoded as the Actor-Topic instead.

(86) Central Amis (J. Kuo 2008: 77)

mi-ki-palifet ku=[**pi-palu** aku ci=umus-an]
 AF.EXT-CMPR-serious TOP.CMN=K.EXT-beat 1SG.ACT PSN=U.-UND

tu=[**pi-palu** isu ci=mayaw-an]
 OBL.CMN=K.EXT-beat 2SG.ACT PSN=M.-UND

‘I beat Umus harder than you beat Mayaw.’ (Lit. ‘My beating Umus is more serious than your beating Mayaw.’)

(87) Central Amis (Fieldnotes)

- a. **mi-palu** kaku ci=umus-an inacila
 AF.EXT-beat 1SG.TOP PSN=U.-UND yesterday
 ‘I beat Umus yesterday.’
- b. **caay pi-palu** kaku ci=umus-an inacila
 NEG K.EXT-beat 1SG.TOP PSN=U.-UND yesterday
 ‘I didn’t beat Umus yesterday.’

In terms of semantics, the Kstem may profile a state of affairs as an event, a fact, or some sort of manner or result, depending on the matrix predicate it collocates with, as is typical of event nominalizations across languages. For example, the event nominalization in (88), marked in brackets, is ambiguous between a factive and a resultative reading.

(88) Central Amis (Fieldnotes)

- ma-raraw** ku=[**pi-palu** ni=mayaw ci=panay-an]
 AF.INT-erraneous TOP.CMN=K.EXT-beat ACT.PSN=M. PSN=P.-UND
 ‘It’s ungraceful (i.e. morally wrong) that Mayaw beat Panay.’
 ‘Mayaw’s beating Panay is a mistake.’ (i.e. Mayaw beat the wrong person.)

In the factive reading, the Mstem *ma-raraw* ‘AF.INT-erraneous’ is a moral judgment on the fact that Mayaw beat Panay. In the resultative reading, however, the same Mstem asserts that the beating result (i.e. Panay got beaten up) is wrong, but not necessarily the beating action per se. In both readings, the presupposition is that Mayaw already beat Panay. In addition, the same Kstem *pi-palu* ‘K.EXT-beat’ denotes the beating manner when collocating with the predicate *tataak* ‘AF.big’, which elsewhere describes the large size of physical objects, as in (89).

(89) Central Amis (J. Wu 2006: 69, 68)

- a. **tataak** **ku=pi-palu** **aku**
 AF.big TOP.CMN=K.EXT-beat 1SG.ACT
 ‘My way of beating (people) is severe.’
- b. **tataak** **ku=’ayam** **aku**
 AF.big TOP.CMN=chicken 1SG.GEN
 ‘My chicken is big.’

A more illustrative example of the Kstem denoting the result associated with an event comes from a snippet of the Frog story narratives as shown in (90), which also nicely demonstrates the functional difference between the Mstem and Kstem.

(90) Central Amis (NTU Corpus|Frog: IU.44-46)

ma-tefad **na=itira** **tu=ra** **kilang** **a** **ma-kunkun** **cingra**
 AF.INT-fall from=there OBL.CMN=DIST tree LIG AF.INT-stumble 3SG.TOP
u=ya **wacu** **sa=tu** **i,** **paka-tengil=tu** **cingra**
 CMN=DIST dog say.so=already PTOP ABLT-listen=already 3SG.TOP
tu=[ka-tefad **nu=ya...(1.8)** **nu=ya** **tamdaw]**
 UND.CMN=K.INT-fall ACT.CMN=DIST ACT.CMN=DIST person

‘He (i.e. the boy) stumbled and fell down from the tree. As for that dog, it could have heard (the sound of) that guy’s falling down.’

The Introvert Mstem *ma-tefad* ‘AF.INT-fall’ is syntactically one of the two matrix predicates in the first sentence, and pragmatically asserts the downfall of a boy looking for his missing frog on a tree. By contrast, its Kstem *ka-tefad* is syntactically an argument of the abilitative predicate *paka-tengil* ‘ABLT-listen’ in the second sentence, and pragmatically assumed the boy’s downfall. More importantly, the Kstem does not refer to the falling as an event or fact, but to the accompanying outcome of the falling event, which is specifically the sound made during the boy’s downfall as coerced by the semantics of the abilitative predicate.¹⁸³

¹⁸³ J. Wu (2006: 243) distinguishes abilitative predicates prefixed by *paka-* from causative ones prefixed by *pa-ka-*.

5.4. Generalizations and reevaluations

This section recapitulates some syntactic properties of various verb forms in Central Amis and reevaluates some claims made in previous studies, some of which are conflicting due to the confusion between the internal syntax of a nominalized structure and its external syntax. I start with the external syntax in §5.4.1, which illustrates the nominal or/and verbal properties of various verb forms. All the verb forms discussed in this section have been amply demonstrated with examples in previous sections, so the discussion here will be elaborated more on the schematic level than specific word forms. Then I move on to the internal syntax in §5.4.2, which argues that verb forms that have the argument function are grammatical nominalizations (Shibatani 2009) rather than lexical ones or “deverbal nouns” (Tsukida 2008: 278) unless there is otherwise compelling evidence to show that these nominalizations have been lexicalized.

5.4.1. External syntax

Previous sections have shown that many verb forms do not behave homogeneously when they interact with higher-order constituents on the phrasal level. While some of them demonstrate syntactic distributions comparable to those of underived nouns, others do not. Still others display both nominal and verbal properties, depending on the context. One of the most telling syntactic operations for teasing apart the nominal and verbal use of a given verb form is negation. While all verb forms can serve as the matrix predicate, some of them are negated only by the verbal pattern, others only by the nominal pattern, and still others by both, as summarized in Table 5.13 below.

Table 5.13: Negation patterns of some verb forms in Central Amis¹⁸⁴

Word forms	Roles in Focus	Verbal NEG	Nominal NEG
I. Mstem	Actor	yes	no
II. ROOT- <i>en</i>	Patient	yes	no
III. Mstem- <i>ay</i>	Actor	no	yes
IV. <i>Ca</i> ~ROOT- <i>en</i>	Patient	no	yes
V. Mstem- <i>an</i>	Patient-Locative	no	yes
VI. <i>sa</i> -Kstem	Conveyance	no	yes
VII. <i>Ca</i> ~Mstem	Actor	yes	yes
VIII. Kstem	Event	yes	yes
IX. Kstem- <i>an</i>	Location-Locative	yes	yes

The negation pattern that a given verb form permits is a good indicator of how it would behave in the larger syntactic environment. Verb forms III~VI must be and VII~IX can be negated by the nominal pattern, whereby the prenominal marker *ku* obligatorily intervenes between the negator *caay* and a verb form. They all share with underived common nouns various properties in terms of serving as arguments, modifiers of arguments, and nominal predicates. These nominal properties are summarized in (91).

¹⁸⁴ Although I have mostly discussed Mstem/Kstem forms in the shape of *mi/pi*-ROOT (for Extrovert AF) and *ma/ka*-ROOT (for Introvert AF), the results in Table 5.13 are generalizable to morphologically more complex Mstem/Kstem forms, such as the causative Mstem *mi-pa-nanum* 'AF.EXT-CAUS-drink', whose Patient-Locative form is *mi-pa-nanum-an* (i.e. Mstem-*an*). Similarly, PF forms consisting of complex stems and the PF -*en* (e.g. *pa-nanum-en* 'CAUS-drink-PF') illustrate the same negation patterns as those made up of just the root and the PF -*en* (e.g. *nanum-en* 'drink-PF').

(91) Nominal properties of V-based nominalizations in Central Amis

- a. They are preceded by nominal relation markers for common nouns, including *ku*, *nu*, *tu*, and *u*, with the last one used before nominal predicates or dislocated NPs.
- b. They can be modified by nominal demonstratives, including PROX *ni*, MED *ra*, and DIST *ya*.
- c. They can modify another nominal within the same NP, with the modifier and modifiee often linked by the attributive ligature *a*.
- d. When serving as nominal predicates, they are optionally preceded by *u*.
- e. When negated by *caay*, they are obligatorily marked by *ku* for external negation.

The properties shared between verb forms III~IV and underived nouns are presumably motivated by the fact that they all denote entity concepts, which qualifies these verb forms as nominalizations irrespective of their morphological composition.

By contrast, the AF-word Mstem and the PF-word ROOT-*en* are the only two forms in Table 5.13 that do not show externally nominal syntax. While the ultimate explanation for why this should be the case awaits further studies, some immediate causes for such a situation can be identified. As has been shown in §4.3, the Mstem across Formosan languages predominantly has the argument function, and exceptions to this generalization are only found in a few languages, with Amis being one of them. Thus, it is very likely that these exceptional languages are innovators. If so, as Central Amis innovated the Mstem-*ay* form for Actor nominalizations, the unmarked Mstem was deprived of the argument function and thus became a highly verbal form. Consequently, we end up having the complementary distribution between the Mstem, which is only negatable by the verbal pattern, and the Mstem-*ay*, which is only negatable by the nominal pattern. By contrast, the irrealis *Ca*~Mstem is somehow immune to the innovation process and has preserved both the predicate and argument function, and is thus negatable by both the nominal and verbal pattern. Similarly, the fact that the PF-word ROOT-*en* in Central Amis does not have the argument function is most likely due to

functional loss, considering that Formosan languages where the PF-word *STEM-en is retained all have the argument function except for Central Amis (see §4.4.2). Finally, there is also dialectal evidence suggesting that the AF-word Mstem and PF-word ROOT-en in Central Amis might have lost their argument function, which remains intact in other varieties of Amis. For instance, Southern (Beinan) Amis is such a variety, as shown in (92).

(92) Southern (Beinan) Amis (L. Jiang 2012: 65)¹⁸⁵

- a. u=tumay ku=[**ma-'adup** ni=tumay]
 CMN=bear TOP.CMN=AF.INT-hunt ACT.PSN=T.
 'What got hunted by Tumay is a bear.'
- b. u=tumay ku=[**'adup-en** ni=tumay]
 CMN=bear TOP.CMN=hunt-PF ACT.PSN=T.
 'What Tumay (intentionally) hunted is a bear.'

The equal footing between a nominalization and an underived noun on the phrasal level is also observed in the flexible order between the two, as illustrated by the two possible word orders in (93), where four nominals are conjoined together by three instances of the ligature *a*.

(93) Central Amis (J. Wu 2006: 97)

- a. ma-araw aku ku=ya [**mi-repel-an ni=mayaw**]
 AF.INT-see 1SG.ACT TOP.CMN=DIST AF.EXT-catch-LA ACT.PSN=M.
 a ta~tulu a tawinaan a kulong
 LIG PL~three LIG female.livestock LIG water.buffalo
 'I saw those three female water buffalos caught by Mayaw.'
- b. ma-araw aku ku=ya ta~tulu a tawinaan
 AF.INT-see 1SG.ACT TOP.CMN=DIST PL~three LIG female.livestock
 a [**mi-repel-an ni=mayaw**] a kulong
 LIG AF.EXT-catch-LA ACT.PSN=M. LIG water.buffalo
 'I saw those three female water buffalos caught by Mayaw.'

¹⁸⁵ Her data are mostly drawn from R. He et al. (1986), which is based on the Amis spoken in Tuli (or *Turik* in the vernacular) and Hsingchang (or *Kanifangar* in the vernacular), Taitung County.

In both examples, the nominalization in bold precedes the head noun *kulong* ‘water buffalo’ and restricts its reference, and the order among modifiers is relatively flexible so long as the head noun appears as the last constituent of the whole NP. On the other hand, nominalizations can be non-restrictive as well, in which case they follow the head noun that they are appositive with. The pair of examples in (94) shows the contrast between a restrictive and non-restrictive use of nominalizations (a topic explored in J. Wu 1996b).

(94) Central Amis (J. Wu 2003: 74)

- a. **ma-araw** **aku** **ku=ya** **[mi-palu-an** **ni=aki]**
 AF.INT-see 1SG.ACT TOP.CMN=DIST AF.EXT-beat-LA ACT.PSN=A.
 a **wawa**
 LIG child
 ‘I saw that child Aki beat.’
- b. **ma-araw** **aku** **ku=ya** **wawa,** **ya** **[mi-palu-an**
 AF.INT-see 1SG.ACT TOP.CMN=DIST child DIST AF.EXT-beat-LA
 ni=aki]
 ACT.PSN=A.
 ‘I saw that child, that (one) Aki beat.’

Crucially, in the second example with an appositive nominalization, the nominalization is modified by the distal demonstrative, just as the head noun it is coreferential with, thus illustrating again the syntactic parallel between underived nouns and nominalizations.¹⁸⁶

On the other hand, verb forms I~II must be and VII~IX can be negated by the verbal pattern, whereby the negator *caay* immediately precedes a verb form, which is required to change into the Kstem if the affirmative form is not already one. The affirmative and negative forms are summarized in Table 5.14 below.¹⁸⁷

¹⁸⁶ Croft (2007: 28-29) distinguished true appositions from alleged appositive modifiers. Their major differences include independent grammatical specifications and separate intonation units, both of which are predominantly phenomena for true appositions but not for alleged appositive modifiers. By these criteria, (94)b seems to demonstrate a true apposition because both the head noun and the appositive phrase made up by a nominalization are each marked by a demonstrative and there is a pause between them (indicated by the comma).

¹⁸⁷ The negative forms of *Ca~Mstem* and *Kstem-an* are not reported in J. Wu (2006: 133), according to

Table 5.14: Negative verb forms for the verbal negation in Central Amis

Affirmative forms	Negative forms
I. Mstem	NEG + Kstem
II. ROOT- <i>en</i>	NEG + (<i>ka-</i>)ROOT- <i>en</i>
VII. <i>Ca</i> ~Mstem	NEG + <i>ka-Ca</i> ~Mstem
VIII. Kstem	NEG + Kstem
IX. Kstem- <i>an</i>	NEG + Kstem- <i>an</i>

The two negation patterns help to distinguish strictly verbal forms from strictly nominal ones in terms of their syntactic interactions with the larger syntax. Verb forms that allow verbal negation but not nominal negation at the same time (i.e. I~II) do not have the functions shared by nominalizations and nouns, namely, serving as arguments, modifiers of arguments, and nominal predicates. They are thus considered strictly verbal. By the same token, verb forms that allow nominal negation but not verbal negation at the same time (i.e. III~VI) are strictly nominal, at least with respect to their external syntax. Only those allowing both the verbal and nominal negation (i.e. VII~XI) are taken to be amphibious over verbal and nominal construals. For instance, when the *Ca*~Mstem *ma~mi-palu* is negated, the verbal construal negates the beating process whereas the nominal construal negates the beater, as in (95).

which the Kstem marker *ka-* as in *ka*-ROOT-*en* for the negative form is optional (hence the parenthesis). M. Chang (2007: 78) also summarized various affirmative and negative verb forms in a table, but her table conflates the verbal and nominal negation, which obscures the syntactic differences among the forms investigated.

(95) Central Amis (Fieldnotes) [= (49)]

- a. **caay ka-ma~mi-palu** ci=panay-an ci=mayaw
 NEG K-IRR~AF.EXT-beat PSN=P.-UND TOP.PSN=M.
 ‘Mayaw will not beat Panay.’
- b. **caay ku=ma~mi-palu** ci=panay-an ci=mayaw
 NEG TOP.CMN=IRR~AF.EXT-beat PSN=P.-UND TOP.PSN=M.
 ‘Mayaw is not the one who will beat Panay.’

The analysis above helps to elucidate some disagreeing opinions on Amis nominalization, in terms of both the syntactic nature of some verb forms and the functions of certain morphemes. First, D. Liu (1999: 114) concluded that the AF-word Mstem and PF-word *ROOT-en* are only verbal, as is also verified in this study, whereas the CF-word *sa-Kstem* and LF-word Mstem-*an* are both verbal and nominal, which differs from the results observed in Table 5.13. The differences between her study and the present one stem from what criteria are taken to be verbal. For her, as long as a verb form functions as the matrix predicate (without being preceded by the marker *u*, as in (65) above), it is then sufficient to grant it the verbal status. In other words, she equated the predicative use of a word form with its syntactic status as a verbal. However, this approach is not fine-grained enough due to two reasons. For one thing, all nominals, including underived nouns, can be the matrix predicate, but not all matrix predicates show the same syntactic distributions. For the other, treating forms like the CF *sa-Kstem* and LF Mstem-*an* as verbals on a par with those like VII~XI in Table 5.13 fails to capture the generalization that only the latter group, but not the former group, can be negated by the verbal pattern just like strictly verbal forms such as the AF Mstem and PF *ROOT-en*. In the present analysis, however, the CF *sa-Kstem* and LF Mstem-*an*, together with the AF Mstem-*ay* and PF *Ca~ROOT-en*, are strictly nominal because their external

syntax follows that of uncontroversial nouns by illustrating the properties in (91) and by prohibiting the verbal negation.

Second, recognizing the syntactic implications of the two negation patterns, J. Wu (2006, 2007) came to a conclusion different from D. Liu's (1999) regarding the nominal/verbal nature of the CF *sa*-Kstem and LF Kstem-*an*. Specifically, she concluded that both forms are nominal because they are negated by the nominal pattern, which is confirmed in this study and many others, but that only the CF *sa*-Kstem, but not the LF Kstem-*an*, is also verbal because the former can be alternatively negated by the verbal pattern. However, the results in Table 5.13 indicate that it is the LF Kstem-*an* rather than the CF *sa*-Kstem that is alternatively verbal. As has been reasoned in §5.2.2.2, the negative form *ka-sa*-Kstem is the corresponding Kstem of the morphologically complex Mstem *ma-sa*-Kstem, so the observed verbal negation pattern should be considered a property of the complex Mstem/Kstem pair (i.e. *ma-sa*-Kstem/*ka-sa*-Kstem) rather than of the CF *sa*-Kstem. By contrast, *ka-Ca*~Mstem is the negative/Kstem form of the AF irrealis *Ca*~Mstem, not of *ma-Ca*~Kstem, which is not even a legitimate form. The alternation between affirmative *Ca*~Mstem and negative/Kstem *ka-Ca*~Mstem is comparable to the non-reduplicated affirmative Mstem *ci-wawa* 'AF.have-child' and its negative/Kstem *ka-ci-wawa* 'K-HAVE-child' (see Table 5.5). On the other hand, contrary to J. Wu's (2006, 2007) claim, the LF Kstem-*an* does participate in the verbal negation in addition to the nominal one, provided that the correct Kstem form is used after the negator (as in (76) above). It is also found that in addition to the LF Kstem-*an*, the Kstem is likewise well-suited for both the nominal and verbal negation. Thus, both the Kstem and its morphological derivative Kstem-*an* share the same syntactic properties.

Third, the minimal contrast between the Kstem and LF Kstem-*an*, together with that between the AF Mstem and LF Mstem-*an*, leads to the question of what role the suffix -*an* should play. According to M. Lin (1995: 170), the answer is a nominalizer for non-Topic arguments, “including patient, goal, locative, and temporal.” For instance, given the Mstem/Kstem pair *mi-palu* ‘AF.EXT-beat’ and *pi-palu* ‘K.EXT-beat’, the morphological derivatives *mi-palu-an* (i.e. Patient-Locative Mstem-*an*) and *pi-palu-an* (i.e. Location-Locative Kstem-*an*) denote the beaten individual and the beating place respectively (see Table 5.8 and Table 5.9 for more examples).¹⁸⁸ The nominalizer analysis seems well-supported, but D. Liu (1999: 56) specifically argued against it on the grounds that the suffix -*an* “plays no role in shifting a verbal element into a nominal one in Amis.” Instead, she analyzed the suffix as part of circumfixes like *mi-...-an* for Patient Focus. However, the Focus circumfix analysis is flawed in many ways. First, both the Mstem and Kstem are legitimate word forms that function independently of the suffix -*an*, so there is no need to posit a circumfix such as *mi-...-an*, which can be systematically derived from the Mstem (consisting of *mi-* and the verb root) suffixed by -*an*. In addition, as shown in Table 5.13, the Mstem is strictly verbal whereas the derived LF Mstem-*an* is strictly nominal with regards to its external syntax, so the suffix -*an* does shift a verbal into a nominal, contrary to what is stated in D. Liu’s (1999) quote above. The circumfix analysis is even more problematic when the morpheme -*an* is suffixed to unaffixed Mstems expressing states (a subtype of Class III in Table 5.1), where no circumfixes can be identified. For instance, the Mstem *tataang* ‘AF.big’ is strictly verbal whereas its

¹⁸⁸ F. Yap & J. Wang (2011) outlined the development of *suo* in Chinese, which started out as a generic noun for places and then grammaticalized into a nominalizer which derives nominals denoting places, goals of motion, and non-locative patients. A similar development might have happened to -*an* in Amis, which indicates patient-like roles when combined with the Mstem and spatio-temporal roles when attached to the Kstem.

morphological derivative *tataang-an* is a nickname for someone who is big in size (J. Wu 2006: 161). In this sense, M. Lin's analysis of the suffix *-an* as a nominalizer should be defended. Even between the Kstem and Kstem-*an*, both of which can be entity-denoting nominals and treated as such in syntax, the suffix *-an* still functions as a nominalizer, in the sense of creating a new nominal (Kstem-*an*) out of an existing one (Kstem). While the Kstem denotes events (see §5.3), Kstem-*an* denotes contingent spatio-temporal roles of events (such as place and time) as well as their metaphorical extensions. Additional examples of the suffix *-an* creating new nominals out of existing ones also come from underived nouns, which denote generic terms after the suffixation, such as *futing* 'fish' vs. *futing-an* 'fish kind' and *fafahi* 'wife' vs. *fafahi-an* 'woman' (J. Wu 2006: 162). Therefore, the suffix *-an* qualifies as a nominalizer that derives a nominal out of either a strictly verbal form (Mstem) or an existing nominal (Kstem or underived nouns).¹⁸⁹

Fourth, another morpheme whose function is also the subject of contentious debate is the suffix *-ay*, which can only be attached to the Mstem (but not the irrealis *Ca~Mstem*). Proposed functions of *-ay* include a nominalizer (M. Lin 1995; D. Liu 1999, 2011) and a factual mood marker (J. Wu 2003, 2006). It has been shown in §5.2.1.1 that although the Mstem-*ay* does introduce import on TAM modifications when compared

¹⁸⁹ However, when *-an* is suffixed to the CF *sa*-Kstem, the result form expresses the optative mood, a function that cannot be easily attributed to nominalization or the component morphemes. Whence the optative meaning has arisen remains a puzzle. For instance, the Mstem *mi-nanum* 'AF.EXT-drink' and its optative counterpart *sa-pi-nanum-an* are contrasted in (i) and (ii) respectively.

- (i) *mi-nanum=ho* *ci=panay* *tu=sayta*
 AF.EXT-drink=still TOP.PSN=P. UND.CMN=soda
 'Panay is still drinking soda.' (J. Wu 2006: 176)
- (ii) *sa-pi-nanum-an* *ku=wawa* *tu=ni* *sayta*
 CA-K.EXT-drink-OPT TOP.CMN=child UND.CMN=PROX soda
 'The child wants to drink this soda.' (J. Wu 2006: 441)

with the Mstem, the two forms demonstrate different syntactic distributions: while the Mstem-*ay* can be arguments, modifiers of arguments, and matrix predicates just like uncontroversial nouns, the Mstem can only be matrix predicates. And when they both serve as the matrix predicate, only the Mstem-*ay*, but not the Mstem, can be preceded by *u* in affirmatives and by *ku* in negatives (see (29) and (30) above), both of which are nominal relation markers for common nouns. Thus, the most elegant explanation for these syntactic differences is then to argue that the Mstem-*ay* is a nominal expression denoting the Actor argument of the Mstem, just as the LF Mstem-*an* is a nominal expression denoting the non-Actor argument of the same Mstem. This justifies the nominalizer analysis of -*ay*, which offers a better explanation for the syntax of the Mstem-*ay* than alternative proposals. More importantly, even a negated predicate can have the same syntactic distributions as the Mstem-*ay* or underived nouns if the negator is marked by -*ay*, thus creating a nominalized construction with internal negation (see Table 5.3 and Table 5.12).

Finally, the marker *u* preceding uncontroversial nominal predicates also helps to elucidate the functional differences between the PF-word ROOT-*en* and its reduplicated counterpart *Ca~ROOT-en*, which are seemingly synonymous in their predicate function when the marker *u* is not used (see (19) above). However, it has been shown in §5.2.1.2 that the reduplicated *Ca~ROOT-en* has the full range of nominal properties listed in (91) whereas the basic PF-word ROOT-*en* is strictly verbal. This parallels the syntactic differences between the strictly verbal AF Mstem and its nominalized counterpart Mstem-*ay*. In both cases, only the strictly nominal member of either pair, namely, AF Mstem-*ay* and PF *Ca~ROOT-en*, can be optionally marked by *u* in the predicate function,

suggesting the two forms are nominal predicates, as opposed to AF Mstem and PF ROOT-*en*, both of which are verbal predicates. Other word forms that behave like the AF Mstem-*ay* and PF *Ca*~ROOT-*en* include the CF *sa*-Kstem and LF Mstem-*an* and, both of which can be the matrix predicate with or without the marker *u*, but only allows the nominal negation pattern.

5.4.2. Internal syntax

The previous section has shown that word forms III~IX in Table 5.13 all demonstrate syntactic functions comparable to NPs consisting of uncontroversial nouns, including serving as arguments, modifiers of arguments, and nominal predicates. But does that mean these forms should belong to the same lexical categories as nouns? For instance, Tsukida (2008: 278) claimed that Conveyance *sa*- and Locative *-an*, which derive the CF-word *sa*-Kstem and LF-word Mstem/Kstem-*an* respectively, “function rather to derive deverbal nouns than as part of the Focus system.” While there is good evidence that these two affixes are not the same animal as Focus affixes in other languages, as J. Wu (2006, 2007) also argues, it is doubtful that derived forms with them are nouns as such. In this section, I argue that word forms III~IX in their argument function, including the CF *sa*-Kstem and LF Mstem/Kstem-*an*, are better analyzed as grammatical nominalizations, or “noun phrases without nouns” in Dryer’s (2007: 193) term, instead of deverbal nouns.

The major evidence for grammatical nominalizations comes from their internal syntax, which they share with strictly verbal forms but not with underived nouns. First, word forms in Table 5.13, all morphologically defined, have the potential to take

argument NPs that are neither Topic nor Actor as long as there is no semantic conflict. Second, they all allow temporal adverbials. Both points are illustrated in (96) by the Mstem-*ay* for Actor nominalizations and in (97) by the Kstem-*an* for Location-Locative nominalizations. Both forms are also compared with the strictly verbal Mstem.

(96) Central Amis (Fieldnotes)

- a. **mi-palu** tu=wawa aku inacila ci=mayaw
 AF.EXT-beat UND.CMN=child 1SG.GEN yesterday TOP.PSN=M.
 ‘Mayaw beat my child yesterday.’ [= (28)a]
- b. caay ku=**mi-palu-ay** tu=wawa aku
 NEG TOP.CMN=AF.EXT-beat-NMLZ OBL.CMN=child 1SG.GEN
 inacila ci=mayaw
 yesterday TOP.PSN=M.
 ‘Mayaw is not the one who beat my child yesterday.’ [= (30)a]

(97) Central Amis

- a. **mi-palu** ci=kilang ci=canglah-an anudafak
 AF.EXT-beat TOP.PSN=K. PSN=C.-UND tomorrow
 ‘Kilang is going to beat Canglah tomorrow.’ (J. Wu 2006: 119)
- b. caay ku=**pi-palu-an** ni=mayaw
 NEG TOP.CMN=K.EXT-beat-LA ACT.PSN=M.
 ci=dongi-an ku=ni anudafak
 PSN=D.-UND TOP.CMN=PROX tomorrow
 ‘This (place) is not where Mayaw is going to beat Dongi tomorrow.’ [= (73)a]

While Tsukida (2008) considered the LF Kstem-*an* a deverbal noun, she called the Mstem-*ay* a Relative form, which seems to assume that the former is lexically a noun whereas the latter is lexically a verb but a relativized one. However, such an assumption is not well supported by syntactic evidence. First, both the AF Mstem-*ay* and LF Kstem-*an* constitute the syntactic head of an NP which denotes a specific participant role in the event expressed by the Mstem/Kstem, and such an NP shares the same nominal properties as one made up of underived nouns, as has been shown in the previous section. Second, the two forms demonstrate the same verbal properties within the denoting phrase

they are part of, including the two points shown in (96) and (97). Third, there is, however, one property that seems to distinguish the LF Kstem-*an* from the AF Mstem-*ay*, and it has to do with collocational compatibility with the GEN-marked phrase, which is often cited as a nominal feature since all uncontroversial nouns collocate with GEN-marked phrases expressing the possessor. But even this factor fails to distinguish one form from the other. The GEN-marked phrase collocating with the LF Kstem-*an* can easily encode the Actor of its corresponding Mstem, as in (97)b, but given the right situational context even the Mstem-*ay*, which itself denotes the Actor, is compatible with a GEN-marked phrase, which expresses the instigator of an action, as in (98).

(98) Central Amis (Fieldnotes) [= (40)]

- a. u=ya tamdaw ku=[**mi-palu-ay** ni=canglah]
 CMN=DIST person TOP.CMN=AF.EXT-beat-NMLZ GEN.PSN=C.
 ‘The one who beat (someone) on Canglah’s behalf is that person.’
- b. u=ya tamdaw ku=[**mi-palu-ay** ni=canglah ci=panay-an]
 CMN=DIST person TOP.CMN=AF.EXT-beat-NMLZ GEN.PSN=C. PSN=P.-UND
 ‘The one who beat Panay on Canglah’s behalf is that person.’

Finally, even stronger evidence against the LF Kstem-*an* being a deverbal noun comes from nominalizations with internal negation. As has been summarized in Table 5.3 and Table 5.12 above, Actor nominalizations with internal negation take the form of NEG-*ay* Kstem, and Location-Locative ones with internal negation almost bear the same form, except that the Kstem is additionally suffixed by *-an*, giving rise to NEG-*ay* Kstem-*an*. Since both the Kstem and Kstem-*an* are immediately preceded by the negator, which is a verbal negation pattern not possible for underived nouns, both forms have to be equally verbal. Therefore, the LF Kstem-*an* is no more a deverbal noun than the AF Mstem-*ay*. They both constitute grammatical nominalizations with internal verbal properties shared

with strictly verbal forms and at the same time external nominal properties shared with underived nouns.

Perhaps more telling of the internal syntax of nominalizations is the way complex predicates are nominalized. In Formosan serial verb constructions, the overall pattern is that the initial verb is the primary locus for Focus and TAM marking as well as argument-indexing whereas non-initial ones are rather constrained with respect to these operations (H. Chang 2010: 192). According to J. Wu (2006a), when the initial verb in Amis is phasal (e.g. BEGIN, FINISH), manner-expressing (e.g. FAST, HARD), or psychological (e.g. LIKE, AFRAID), the second verb can only be the AF Mstem. This AF-only restriction on the second verb is observed irrespective of the form of the initial one. For instance, the initial verb in imperative sentences can be either AF (taking the required Kstem) or PF, but the second verb can only be the AF Mstem, as shown in (99).

(99) Central Amis (J. Wu 2006: 140)

- a. **pi-lingatu=tu** {k<um>aen/*kaen-en} tu=futing
 K.EXT-begin=already AF.<UM>eat/eat-PF UND.CMN=fish
 ‘Start to eat fish!’ [AF imperative]
- b. **lingatu-en=tu** {k<um>aen/*kaen-en} ku=futing
 begin-PF=already AF.<UM>eat/eat-PF TOP.CMN=fish
 ‘Start to eat the fish!’ [PF imperative]

Importantly, the same restriction applies when complex predicates are nominalized. In (100), for instance, the Mstem *ma-ulah* ‘AF.INT-like’ and a following Mstem form a complex predicate in the first example, and when such a complex predicate is nominalized only the initial Mstem is changed into its Kstem-based LF form *ka-ulah-an* ‘K.INT-like-LA’ while the second verb remains intact, as in the second example.

(100) Central Amis

- a. **ma-ulah** ku=oner a **ma-sadak** tu=dadaya
 AF.INT-like TOP.CMN=snake LIG AF.INT-appear OBL.CMN=evening
 ‘Snakes like to show up in the evening. (9-Level Textbooks 7-8)
- b. u=maan ku=[**ka-ulah-an** isu a **mi-nanam**]
 CMN=what TOP.CMN=K.INT-like-LA 2SG.ACT LIG AF.EXT-learn
 ‘What is it that you like to learn?’ (Supplementary Materials, Daily Conv. L23)

If the LF Kstem-*an* were a deverbal noun in contrast to the Mstem, to which no one seems to have attributed any nominal properties, then we would have a complex predicate consisting of two verbs in (100)a but a similar construction made up of one deverbal noun and a verb in (100)b, which is bizarre. But if the LF Kstem-*an* is not a deverbal noun, as is argued here, then all that needs to be stated is that when a serial verb construction is nominalized only the initial verb has to be one of those forms that have the nominal use (i.e. III~IX in Table 5.13). This is comparable to the fact that in imperative sentences only the initial verb of a serial verb construction needs to assume the required form (e.g. the Kstem for AF verbs).

In spite of the supporting evidence for grammatical nominalizations and against deverbal nouns, some word forms do seem to have lexicalized into something that is syntactically indistinguishable from underived nouns and often semantically much more specialized than what a grammatical nominalization would denote.

Take Actor nominalizations for instance. As has been shown in §5.2.1.1, Actor nominalizations maximally consist of a negated predicate with patient-like arguments and temporal adverbials and minimally of just the Mstem-*ay*, as repeated in (101).

(101) Central Amis

- a. udengan ci=mayaw ku=[**caay-ay** pi-palu tu=wawa
 only PSN=M. TOP.CMN=NEG-NMLZ K.EXT-beat UND.CMN=child
 aku inacila]
 1SG.GEN yesterday
 ‘The only one who didn’t beat my child yesterday is Mayaw.’ (Fieldnotes) [= (34)]
- b. caay ku=[**ma-tuka-ay** kaku
 NEG TOP.CMN=AF.INT-lazy-NMLZ 1SG.TOP
 ‘I am not a lazy person.’ (M. Chang 2007: 93) [= (35)b]

Given the wide range of syntactic complexities within Actor NPs, it is conceivable that those consisting of just the Mstem-*ay* would be more susceptible to lexicalization once their meanings are conventionalized. A good example is *matuasay*, which is a term for senior members of a specific “age grade” (Chen 1989), and has lexicalized from *ma-tuas-ay* ‘AF.INT-grow-NMLZ’, which could potentially denote anyone who grows. But the semantics of *matuasay* has become so specialized that it is now inappropriate to use it to refer to someone other than people of certain age. A comparable example in English would be *grown-up*, which denotes adults and can be pluralized just like lexical nouns. Other than social ranks, the Mstem-*ay* is also a common source for names of tribal villages, often dubbed after the most prominent local biological attribute. For instance, in her Amis dictionary Fey (1986: 63) lists four village names in the form of *ci-X-ay*, where *ci-X* is a denominal verb (AF Mstem Class III) meaning “having or filled with X” (see (17) and (26) above). They include *ci-lamit-ay* (cf. *lamit* ‘root’) and *ci-lengac-ay* (cf. *lengac* ‘shell ginger’) in Hualien County and *ci-kuwa’-ay* (cf. *kuwa* ‘papaya’) and *ci-likes-ay* (cf. *likes* ‘mosquito’) in Taitung County.

Since lexicalization is often a solution to answer the need to create new labels for novel concepts, it is not surprising that terms for modern facilities tend to result from

lexicalization of grammatically and semantically less restricted structures. For instance, the Mstem/Kstem pair *mi-tilid/pi-tilid* means to write (or paint), as in (102).

(102) Central Amis (Fieldnotes) [= (36)]

- a. caay pi-tilid tu=ngangan aku ci=panay
 NEG K.EXT-write UND.CMN=name 1SG.GEN TOP.PSN=P.
 ‘Panay didn’t write my name.’
- b. caay ku=[mi-tilid-ay tu=ngangan aku] ci=panay
 NEG TOP.CMN=AF.EXT-write-NMLZ UND.CMN=name 1SG.GEN TOP.PSN=P.
 ‘Panay is not the one who wrote my name.’

And the pair also means to study or acquire knowledge in the modern sense, and it is based on this sense that its Kstem-based LF form *pi-tilid-an* is recruited for a term meaning an educational establishment, not just any location where one happens to be studying or writing. This term is a relatively recent lexicalization not only because the notion it expresses is modern, but also because the same notion is expressed by LF forms based on different roots across Amis varieties, with each choosing a root (native or borrowed) that can be associated with studying, as contrasted in (103) through (120).

(103) Central Coastal Amis (Fieldnotes)

mi-tilid kaku i=**pi-tilid-an**
 AF.EXT-study 1SG.TOP LOC=K.EXT-study-LA
 ‘I am studying at school.’

(104) Central Xiuguluan Amis (9-Level Textbooks 2-1)

mi-cudad kaku i=**pi-cudad-an**
 AF.EXT-study 1SG.TOP LOC=K.EXT-study-LA
 ‘I am studying at school.’

(120) Southern Hengchun Amis (9-Level Textbooks 2-1)

mi-takci kaku i=**pi-takci-an**
 AF.EXT-study 1SG.TOP LOC=K.EXT-study-LA
 ‘I am studying at school.’

In short, verb forms in their argument function are better analyzed as grammatical nominalizations rather than deverbal nouns. They are syntactically headed by verbs (or

the negator, which is a highly grammaticalized verb; see Footnote 163) but semantically denote event participants or the event itself expressed by verbs. Since grammatical nominalizations can be as simple as one single word form, they are effective ways for creating new lexical items with specialized meanings (see also Shibatani & Bin Makhshen 2009).

5.5. Chapter summary

In this chapter, I first define Amis verbs in terms of their morphological potentials. AF verbs fall into at least three classes, depending on the morphological alternations between the Mstem and Kstem (see Ross 2015a), two abstract forms that make the description of Amis verbs much more economical and elegant than those in previous studies. In addition, while previous studies split *ma*-verbs into AF and PF (or even Neutral) types based on argument encoding patterns, I characterize all *ma*-verbs as AF Introvert, in contrast to *mi*-verbs, which are AF Extrovert. A unified semantic account of Introvert *ma*-verbs has the advantages of not only avoiding unnecessary labels like AF and PF *ma*- as well as unmotivated switches between them, but also accounting for some systematic morphological operations across all *ma*-verbs. Importantly, the Extrovert/Introvert semantic contrast is also compatible with both J. Wu's (2006) and Tsukida's (2008) generalizations.

Verb forms other than AF are identified according to the affixes or/and the Mstem/Kstem they take. In total, nine verb forms are investigated and it is found that some are strictly verbal, others are strictly nominal, and that still others are both verbal and nominal in terms of their external syntax, especially with regards to negation

patterns. The AF Mstem and PF ROOT-*en* are strictly verbal and prohibited from serving as nominalizations. By contrast, the AF Mstem-*ay*, irrealis PF *Ca~ROOT-en*, CF *sa-Kstem*, Patient-Locative Mstem-*an* and are strictly nominal and share various nominal properties with NPs consisting of underived nouns. Finally, the AF irrealis *Ca~Mstem*, eventive Kstem, and Location-Locative Kstem-*an* are equally verbal and nominal, depending on the context. Against these results, disagreeing opinions on the syntactic functions of verb forms as well as the functions of certain morphemes thereof are reexamined.

Finally, verb forms that has the argument function are argued to be grammatical nominalizations rather than deverbal nouns because they internally demonstrate properties shared with strictly verbal forms but not with underived nouns and externally behave like typical NPs, which serve as arguments, modifiers of arguments, or nominal predicates. Nevertheless, these grammatical structures may be lexicalized, thus creating word forms with specialized meanings and syntactic behaviors indistinguishable from those of underived nouns.

|Chapter 6

Nominal-based Nominalization I: General Issues

The present study limits the investigation of grammatical nominal-based nominalization, more commonly known as genitive/possessive constructions, to fifteen Formosan languages/dialects, including Budai and Taromake Rukai, Isbukun and Takibakha Bunun, Pngawan and Squliq Atayal, Tgdaya Seediq, Rikavung Puyuma, Northern Paiwan, Central Amis, Kavalan, Saisiyat, Tsou, Thao, and Saaroa.¹⁹⁰ The locations where linguistic data on these languages were collected are indicated in Figure 6.1. This is more or less a convenient sample because I happened to have direct access to native speakers of these languages at the time of research. Nevertheless, the sample is still believed to be diverse and representative enough because it covers all the first-order

¹⁹⁰ To avoid the cumbersome use of “languages/dialects”, I will simply refer to them as languages in the sense of linguistic systems, and not be concerned with where to draw the line between languages and dialects.

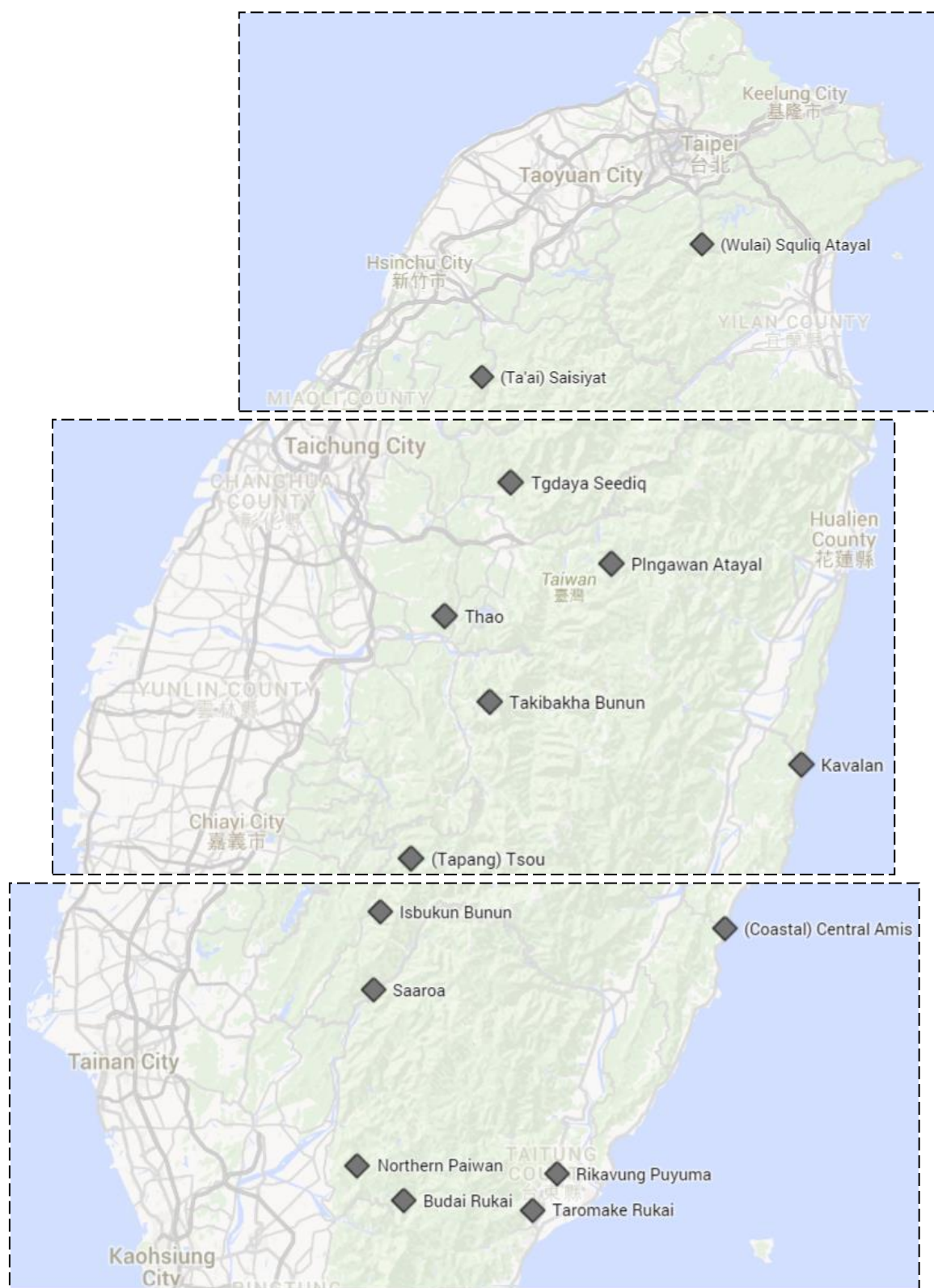


Figure 6.1: Locations of fifteen Formosan languages

subgroups of Austronesian languages that are Formosan (i.e. non-Malayo-Polynesian) in Blust's (1999) subgrouping account.

Specific examples from each language will be presented later in Chapter 7. This chapter deals with general issues regarding nominals that contain a reference entity but denote entities associated with it.

6.1. POSS and GEN in the nominalization account

In previous chapters, prenominal and person-form markers that indicate the possessor were glossed as GEN, following the tradition in Austronesian linguistics (Reid & H. Liao 2004; Ross & S. Teng 2005a). In this and subsequent chapters, we will look into genitive/possessive constructions in more detail, and thus start to adopt glosses such as NMLZ (for nominalizer) and NMRK (for nominalization marker) from the nominalization-based perspective (Shibatani & Shigeno 2013).

In general typology, there has been some terminological confusion with regard to terms like “possessive” and “genitive” (respectively abbreviated as POSS and GEN). As Partee & Borschev (2001: 91) put it, “[t]he terminology surrounding “possessives” and “genitives” is confusing, since the correspondences among morphological forms, syntactic positions, grammatical relations, and semantic interpretations are complex and debated, and vary considerably across languages.” Moreover, the marker that specifies the possessor in a language often has non-possessive functions. When this happens, “[t]here is...no terminological consensus on the label used ... and different local traditions solve this problem in different ways” (Koptjevskaja-Tamm 2001: 961). The labeling challenge is in fact expected if we accept the idea of categorical particularism, according

to which linguistic categories are language-specific rather than universal (Haspelmath 2010).

A common practice among typologists is to distinguish GEN from POSS following Dryer's (2007: 178) suggestion: "It is important not to confuse the two sorts of affixes...The genitive affix...is a case affix and signals that the possessor noun it occurs with is functioning as a possessor. The possessive affix... in contrast, is a pronominal morpheme, varying for pronominal features of the possessor." He illustrates such a distinction with Maybrat in (1), where *ro-* indicates the nominal it attaches to is the possessor and hence a GEN marker whereas *m-* indexes the number and gender of the possessor and is thus a POSS.

(1) Maybrat (Dryer 2007a: 185)

- a. amah ro-Petrus
house GEN-P.
'Petrus's house'
- b. Sely m-me
S. 3SG.NM.POSS-mother
'Sely's mother'

Moreover, a language may have both types of markers (i.e. so-called double-marking languages), such as Turkish in (2).

(2) Turkish (Dryer 2007a: 181)

- Ahmed-in ođl-u
A.-GEN son-3SG.POSS
'Ahmed's son'

However, there is a skewed distribution of functional intricacies between the type of information encoded by markers in construction with the possessor (POR) and those in construction with the possessum (PUM) across languages. PUM-collocating markers either vary with respect to the gender, number, or person of the POR, as in Maybrat and

Turkish, or remain invariant, as in Western Tarahumara (e.g. *kanteláριο upí-la* [Candelario wife-POSS] ‘Candelario’s wife’; Burgess 1984: 61). Invariant markers of the latter type are often called “construct” possessive affix, a term commonly used in Afro-Asiatic languages but also adopted in Austronesian (e.g. Lichtenberk 2009: 257). By contrast, POR-collocating markers provide much diverse and richer information on the PUM. They may be invariant, as in Maybrat, Turkish (excluding changes triggered by vowel harmony), and English. They may also “agree” with the PUM in terms of its gender, number, or other grammatical features, as in Hindi and Albanian (Spencer 2007). In the present nominalization-based perspective, these PUM-agreeing markers are nominalizers that create a schematic nominal (i.e. the PUM) metonymically associated with a plain nominal (i.e. the POR). For instance, nominals marked by *kaa* in Hindi can be put into the NP-use or modification-use, as in (3).

(3) Hindi (Spencer 2007; citing McGregor 1995: 9)

- a. yah makaan [us strii=**kaa**] hai
 PROX house.M.SG that woman=NMLZ.M.SG COP
 ‘This house is that woman’s.’ [NP-use]
- b. [us strii=**kaa**] beTaa
 that woman=NMLZ.M.SG son.M.SG
 ‘that woman’s son’ [Modification-use]

Hindi *kaa* differs from English *'s* only in terms of the denotation potentials of the nominalized nominal, which have to fall into one of the few grammatical noun classes in Hindi but are unconstrained in English. Markers like Hindi *kaa* are sometimes called “genitive markers that agree” (Haspelmath 2014), but the phrase *us strii=kaa* in (3)a does not even form a syntactic constituent with a noun such that the marker *kaa* could agree with it.

Moreover, there are many other markers that substantialize the PUM like Hindi *kaa*, but never go by GEN. For instance, POR-collocating markers can be a sizable number of noun-like free forms that indicate the physical or functional properties of the PUM (called “possessive nouns”; e.g. *u-yekîn kaware* [1SG-pet horse] ‘my horse (as a pet)’ in Macushi; Abbott 1991: 85) or a few grammaticalized forms specifying that the PUM is meant for alimentary or other general purposes (called “possessive classifiers”; e.g. *ga-ku moli* [POSS.CLF-1SG.POSS citrus] ‘my orange/pomelo (as fruits to eat)’ in Lolovoli; Hyslop 2001: 185). Aikhenvald (2000: 125) refers to POR-collocating markers of the Macushi and Lolovoli type as “possessed classifiers” and “relational classifiers” respectively, which are two of the three ways nouns in possessive constructions are categorized according to her typology on classifiers. However, Shibatani (2015, 2016) argues that classifiers in diverse language families are better understood as nominalizers because they derive new nominals with denotations evoked by the classificatory features of classifiers, often in terms of the physical properties of worldly objects. Importantly, expressions nominalized by classifiers can be syntactically independent of a noun, which makes dubious the general belief that classifiers are meant to classify nouns. For instance, *kuv* ‘I/me’ in White Hmong refers to the speaker, and *kuv lub* ‘mine’, where *lub* is a classifier, denotes any body parts or table-like objects associated with the speaker. The nominalized nominal *kuv lub* can in turn modify a noun like *rooj* ‘table’, as in *kuv lub rooj* ‘my table’ (ibid.; citing Nerida Jarkey, p.c.). Like Hindi *kaa*, classifiers narrow down the denotation range of a nominal associated with the POR. Therefore, POR-collocating markers reify the PUM with varying degrees of specificity, ranging from something as concrete as a pet in Macushi to something as vague as a generic entity in English. It is in

this sense that they all have to do with nominal-based nominalization, where a reference nominal (i.e. the POR) is nominalized one way or another to denote a metonymically associated nominal (i.e. the PUM), which is either referential by itself or restrictive when modifying another nominal. Although few studies other than Shibatani & Shigeno (2013) apply the term nominalizer to what is commonly called a genitive case marker, Noonan (2008b: 139) suggested that the marker *-ye* in Chantyal, a Bodic language, “functions... as a sort of *nominalizer*, permitting the adnominal NP to fill a noun slot requiring another case,” as in (4).

(4) Chantyal (Noonan 2008b: 130)¹⁹¹

nə-ye-sə **kfi-ye-ra** jfi-i
 1SG-NMLZ-ERG 2SG-NMLZ-DAT bite-PFV
 ‘Mine bit yours.’ (e.g. ‘My dog bit your dog.’)

He also commented that “simple headless adnominals” as in (4) are attested in all the Bodic languages he had investigated but were rarely discussed in the literature. A similar situation can be said about the Formosan literature, which is one of the reasons why we are looking into this topic in the present study.

Therefore, POR-collocating markers in the nominalization account are nominal-based nominalizers that derive new nominals via metonymic processes, be it invariant markers like English *'s*, paradigmatic ones like Hindi *kaa*, or a sizable number of classifiers. On the other hand, PUM-collocating markers like bound person forms simply index the gender/number/person of a reference nominal that is being nominalized, and thus can be called nominalized person indexes (i.e. POR indexes). When nominalized indexes are clitics, they may attach to hosts other than a PUM noun (see §2.3.3). Finally,

¹⁹¹ The original gloss for *jfi* is “hit”, which seems to be a typo for “bite” judging from other examples in Noonan (2008b).

“construct state” forms of nouns simply indicate the existence of such a reference nominal without specifying its grammatical features.

6.2. Variations of nominalized person indexes in Formosan

This section demonstrates the variations of nominalized person indexes in Formosan languages, namely, bound person forms that index the number/person of the reference nominal that is being nominalized. I focus on paradigmatic distributions between nominalized person indexes and nominalized full nominals on the one hand and those between plain person indexes for the Topic and nominalized ones on the other.

Nominalized indexes and nominalized full nominals are typically in complementary distributions, often with the former being clitics directly attached onto the PUM (or other potential hosts; see §2.3.3) and the latter marked by something or simply juxtaposed with the PUM. Some languages, however, allow nominalized non-SAP indexes to optionally co-occur with nominalized full nominals, as illustrated in (5) through (7).

- (5) Kavalan (Fieldnotes)
- a. lepaw={ku/na}
house={1SG.NMLZ/3.NMLZ}
'{my/his or her or their} house'
 - b. lepaw=(na) ni=abas
house=3.NMLZ NMLZ=A.
'Abas's house'

(6) Tsou (Fieldnotes)

- a. ceopngu={u/si}
 hat={1SG.NMLZ/3SG.NMLZ}
 ‘{my/his or her} hat’
- b. ceopngu=(si)=[to paicɯ]
 hat=3SG.NMLZ=NMLZ P.
 ‘Paicɯ’s hat (said when Paicɯ is not visible to the speaker)’

(7) Saaroa (Fieldnotes)¹⁹²

- a. {suhlati=ku/suhlate=isa}
 {book=1SG.NMLZ/book=3.NMLZ}
 ‘{my/his or her or their} book’
- b. suhlate=(isa) amahle
 book=3.NMLZ A.NMLZ
 ‘Amahle’s book’

Much less common is for SAP indexes to co-occur with their coreferential free nominals, which is found in Seediq (see §2.3.3). Thus, the generalization is that if a language allows person indexes to co-occur with their coreferential free nominals (i.e. cross-indexes in Haspelmath’s 2013 terminology), it always starts with non-SAP indexes.

As far as SAP indexes are concerned, Topic person indexes and nominalized ones can be completely identical in form, as in Tsou.¹⁹³ The complete paradigms in Tsou are given in Table 6.1 (Zeitoun 2005: 277; G. Lin 2010: 95), with 1SG forms illustrated in (8).

¹⁹² The Saaroa word for “book” is *suhlate* [sulati] when used in isolation, but the word-final central vowel [i] is fronted to [i] when immediately followed by a syllable with the high back rounded vowel [u] in it (see Pan 2012: 48 for details).

¹⁹³ Non-SAP indexes are excluded because most Formosan languages do not have Topic indexes for non-SAPs.

Table 6.1: Topic vs. nominalized person forms in Tsou¹⁹⁴

	1SG	1EXCL	1INCL	2SG	2PL
Topic indexes	= <i>'o</i>	= <i>mia</i>	= <i>to</i>	= <i>su</i>	= <i>mu</i>
Nominalized indexes	= <i>'u</i>			= <i>ko</i>	

(8) Tsou (G. Lin 2010: 284, 86)

a. **la='u** **smoyo=[to av'u]**
 HAB=1SG.TOP AF.afraid=OBL dog
 'I am afraid of dogs.'

b. **mi='o=cu** **suyusɬ=[ne hopo='u]**
 RLS.AF=1SG.TOP=already AF.dress=LOC room=1SG.NMLZ
 'I already got dressed in my room.'

The other extreme is that the two types of indexes are utterly distinct in form, as in Kavalan. Table 6.2 (P. Li 1978: 352; Tsuchida 1993: 92; D. Yen 2012: 30) shows the complete paradigms in Kavalan, and Example (9) illustrates 1SG forms.

Table 6.2: Topic vs. nominalized person forms in Kavalan

	1SG	1EXCL	1INCL	2SG	2PL
Topic indexes	= <i>iku</i>	= <i>imi</i>	= <i>ita</i>	= <i>isu</i>	= <i>imu</i>
Nominalized indexes	= <i>ku</i>	= <i>niq</i>	= <i>ta</i>	= <i>su</i>	= <i>numi</i>

¹⁹⁴ The table shows forms in the Tapang variety of Tsou. Equivalent forms in the Tfuya variety are almost identical except for =*mza* '1EXPL.PL'. Also, the two variant forms for 1SG and 2SG are allomorphs conditioned by phonology, not by the roles they index.

(9) Kavalan

- a. aiku nani kbalan=**iku**
 1SG.TOP P_{TOP} Kavalan=1SG.TOP
 ‘As for me, I am a Kavalan.’ (P. Li 2006: 31)
- b. yau ta=bibuR ya=lepaw=**ku**
 EX LOC=south TOP=house=1SG.NMLZ
 ‘My house is in the south.’ (H. Jiang 2006: 112)

Finally, between the two extremes, there are various intermediate cases with partial overlaps in form. Regardless of the overlap patterns, distinct forms are much more often found in 1SG, as in Tgdaya Seediq. Table 6.3 (Holmer 1996: 32; Holmer & Billings 2014: 114) lists the complete paradigms in Tgdaya Seediq, the 1SG forms of which are exemplified in (10).

Table 6.3: Topic vs. nominalized person forms in Tgdaya Seediq

	1SG	1EXCL	1INCL	2SG	2PL
Topic indexes	= <i>ku</i>				
Nominalized indexes	= <i>mu</i>	= <i>nami</i>	= <i>ta</i>	= <i>su</i>	= <i>namu</i>

(10) Tgdaya Seediq

- a. iwan=**ku** tado
 I.=1SG.TOP T.
 ‘I am Iwan Tado.’ (Ochiai 2009: 31)¹⁹⁵
- b. laqi=**mu** mqedin
 child=1SG.NMLZ woman
 ‘my daughter’ (Holmer 1996: 73)

In sum, bound person forms that index someone (or something) and those that index entities associated with someone (or something) range from being identical to completely distinct and anywhere in between. Interestingly, the variations found in the 15

¹⁹⁵ Tado is a patronymic surname.

Formosan languages are diverse enough to compete with Siewierska's (2009) results drawn from a sample of 157 languages.

6.3. Existential vs. equational type of possession

There are two major types of constructions that assert a possessive relationship in the western Austronesian area (Himmelmann 2005). They can be conveniently referred to as the *existential type* and *equational type*. The two types are highly correlated with the modification-use and NP-use of nominalized nominals respectively. The existential type is illustrated by Muna in (11) and the equational type by Tetun Dili in (12).

(11) Muna (Himmelmann 2005a: 139; citing van den Berg 1989: 161)

miina bhe doi-ku
 NEG EX money-1SG.NMLZ
 'I do not have any money.' [The existential type with the modification-use]

(12) Tetun Dili (Himmelmann 2005a: 139; citing Hull & Eccles 2001: 34)¹⁹⁶

ne'e sira-nia-n la'ós ami-nia-n
 PROX 3PL-NMLZ-NMRK NEG 1INCL-NMLZ-NMRK
 'This is theirs, not ours.' [The equational type with the NP-use]

¹⁹⁶ In the cited reference, the original gloss for *-nian* is POSS, which is not in the same sense as the gloss is suggested by Dryer (2007: 178), meaning person forms on the PUM that index the person/number of the POR. The present glossing is based on the following distribution patterns, which are taken from Williams-van Klinken (2015: 71):

- (i) *nee hau-nia kareta*
 PROX 1SG-NMLZ car
 'This is my car.'
- (ii) *kareta nee hau-nia-n*
 car PROX 1SG-NMLZ-NMRK
 'This car is mine.'
- (iii) *nee José-nia uma*
 PROX J.-NMLZ house
 'This is José's house.'
- (iv) *uma nee José-nia-n*
 house PROX J.-NMLZ-NMRK
 'This house is José's.'

Heine (1997) identifies eight event schemas that form the basis of predicative possessive constructions in languages around the world, and the Austronesian existential and equational type respectively fall into her Genitive and Equation schema, which respectively give rise to two macro-constructions dubbed as HAVE- and BELONG-Constructions, as schematically represented in (13).

(13) Event schemas of predicative possession (Heine 1997)

- a. Genitive Schema: X's Y exists > X has Y
- b. Equation Schema: Y is X's (property) > Y belongs to X

Himmelmann (2005: 139) mentioned that the existential type is “by far the most common”, and that the equational type is the “major alternative.” However, it is not clear whether the equational type is less common because it is less researched and thus less found in the literature or it is so because it is less attested in languages as compared with the existential type.

Among Formosan languages, the existential type is frequently researched but the equational type is generally ignored, as revealed by typological studies such as Zeitoun et al. (1999) and Zeitoun (2000a). The lack of attention to the equational type leads to a lack of understanding of the NP-use of nominalized nominals and its paradigmatic relationship with the modification-use. To fill the gap, this and the subsequent chapter explore both the modification-use and NP-use in Formosan languages. Paradigmatic comparisons across uses will help to elucidate their possible developments over time and how some languages acquire multiple modification patterns.

The existential and equational type in two Formosan languages are illustrated in (14) and (15).

(14) Central Amis (Fieldnotes)

- a. ira ku=**cu**dad ni=**ka**caw
 EX TOP.CMN=book NMLZ=K.
 ‘Kacaw has a book.’ [Existential type]
- b. ni=**ka**caw ku=ni a cudad
 NMLZ=K. TOP.CMN=PROX LIG book
 ‘This book is Kacaw’s.’ [Equational type]

(15) Kavalan (Fieldnotes)

- a. yau ya=**wa**su=**ku**
 EX TOP=dog=1SG.NMLZ
 ‘I have a dog.’ [Existential type]
- b. za=**ku** ya=wasu zau
 NMRK=1SG.NMLZ TOP=dog PROX
 ‘This dog is mine.’ [Equational type]

The examples show that the equational type is a great construction to identify the NP-use of nominalized nominals because they function as the matrix predicate by themselves without presenting potential parsing problems caused by interacting with other constituents. It is found that the equational type is attested in all the fifteen Formosan languages investigated. Supporting examples from each of them are deferred until Chapter 7. In the rest of this chapter, I discuss some general properties of the possessive predicate found in the equational type.

6.4. Syntactic functions of possessive substantives

Despite the convenient name “equational type”, it is not always the case that the sole argument and predicate in this type are truly equational in the sense that the two constituents are equally nominal, especially when the POR is marked by some sort of locative or oblique markers. For instance, there are languages like Japanese, where the possessive predicate is a true entity-denoting nominal that can function as both arguments

and predicates (e.g. *kono hoN=wa watashi=no desu* ‘This book is **mine**.’). There are also languages like French, where the possessive predicate does not fulfill argument functions (e.g. *Le livre est à moi* ‘The book is **mine**.’). In addition, one single language may have a grammatical way to distinguish one scenario from the other. Spanish is such a language, where *mia* ‘mine’, without the definite article, is predicative but not nominal, as in *Esa pluma es mia* ‘The pen is **mine**.’ whereas *la mia* ‘mine’, with the definite article, is both predicative and nominal, as in *Esa pluma es la mia* ‘The pen is **mine**.’ (see Lyons 1999: 25).¹⁹⁷

It is found that in the fifteen Formosan languages investigated, possessive predicates in equational-type constructions like those in (14) and (15) are all true nominals that have both predicate and argument functions, which will be demonstrated in §6.4.1 and §6.4.1 respectively. They come close to what Ultan (1978: 27) calls possessive substantives, meaning “nonattributive, independent possessive pronouns or nouns” (e.g. English *mine*). However, in spite of the traditional term “possessive pronoun”, expressions like *mine* are strictly speaking no more pronominal than *what I bought* because both are grammatical nominalizations denoting a heterogeneous set of entities and that the identity of their referents is unspecified and not entirely dependent on an antecedent (see §1.1.2). While person forms referring to speech-act participants (SAPs) may be pronominal, such as English *I*, nominalized person forms referring to entities metonymically associated with SAPs are not, such as English *mine*. Thus, so-called possessive pronouns are the NP-use of nominalized person forms (Shibatani 2014).

¹⁹⁷ Zribi-Hertz (1997) characterizes the former as a construction with a “relation” meaning and the latter as the one with “possession” meaning.

Nevertheless, I shall use possessive substantives as a shorthand for forms that can be put into the NP-use of nominalized nominals.

6.4.1. Predicate functions

Possessive substantives function as affirmative predicates simply by occurring in the predicate position, as has been demonstrated by Central Amis in (14). This very example presents a challenge for Kaufman's (2009b: 29) claim that a ban on "genitive predicates" in many Philippine languages provides a basis for Austronesian extraction restrictions whereby only the Topic argument can be questioned, clefted, or nominalized. For instance, in a Tagalog NAF-construction only the non-Actor Topic can be questioned but not the non-Topic Actor, as shown in (16).

(16) Tagalog (Kaufman 2009b: 30-31)¹⁹⁸

- | | | | |
|-----|------------------------------|---------------------------|------------------|
| a. | b<in>ili | ni=Boboy | ang=libro |
| | <RLS.PF>buy | GEN=B. | TOP=book |
| | 'Boboy bought the book.' | | |
| b.* | nino | ang=b<in>ili | |
| | GEN.who | TOP=<RLS.PF>buy | |
| | Intended: 'Who bought (it)?' | | |
| c. | sino | ang=b<in>ili | |
| | TOP.who | TOP=<RLS.PF>buy | |
| | 'Who was bought?' | | |

According to him, the restriction in (16)b is due to the fact that genitive predicates are prohibited in the first place, as in (17)b.

¹⁹⁸ To highlight the reasoning in Kaufman (2009b), I keep the original gloss for *ni* in both (16) and (17).

(17) Tagalog (Kaufman 2009b: 28)

- a. ang=koponan **ni=Juan**
 TOP=team GEN=J.
 ‘Juan’s team’
- b.* **ni=Juan** ang=koponan
 GEN=J. TOP=team
 Intended: ‘The team is Juan’s.’

The rationale behind this is that both the non-Topic Actor and the POR are marked by Genitive *ni* (see §4.1 for the overall Actor/POR syncretism). Thus, if (17)b is not even acceptable, it follows that (16)b is ungrammatical as well. However, the same extraction restrictions apply to Formosan languages, but Amis does allow so-called genitive predicates, and so do Paiwan and Seediq, as shown in (18) and (19) respectively.

(18) Northern Paiwan (Fieldnotes)

- a. sunatj **ni=camak**
 book NMLZ=C.
 ‘Camak’s book’
- b. aza=(a) sunatj, **ni=camak**
 TOP.PROX=LIG book NMLZ=C.
 ‘This book is Camak’s.’

(19) Tgdaya Seediq (Fieldnotes)

- a. sapah **na=mona**
 house NMLZ=M.
 ‘Mona’s house’
- b. **na=mona** ka=[sapah nii] han
 NMLZ=M. TOP=house PROX PRT
 ‘This house is Mona’s.’

Importantly, markers like *ni* or *na* in these languages introduce both non-Topic Actor and the POR, and the *ni* marker in Tagalog, Amis, and Paiwan are even cognate forms (< PAn *ni).¹⁹⁹ Therefore, it is problematic to base Austronesian extraction restrictions on so-called genitive predicates, which are independent nominals in light of Formosan data.

¹⁹⁹ Like that in Amis and Paiwan, *ni*-phrase in Pazeh can also be the predicate, as in the following two

Moreover, possessive substantives are negated like underived nouns. In indicative sentences, Formosan languages either use the same negator regardless of predicate types or require different negators depending on whether the predicate is verbal or nominal. For ease of reference, the former is termed *share-NEG languages* and the latter *split-NEG languages* (modified from Stassen 1997). These two types are respectively shown in the upper and lower sections of Table 6.4 below (separated by double lines). In all the split-NEG languages, possessive substantives are negated like a nominal, not a verbal, thus illustrating their nominal nature in all these languages.

Split-NEG languages come in two subtypes. In Budai Rukai, Central Amis, and Saisiyat, the negative morpheme is identical for both verbal and nominal predicates, but different syntagmatic requirements have to be met, depending on whether the negated predicate is verbal or nominal.²⁰⁰ For instance, in Budai Rukai the morpheme *ka* is prohibited from verbal negation but is required for nominal negation, and importantly possessive substantives marked by *-ane*, for person forms and full nominals alike, are negated like nominals, as in (20) below.

examples:

Pazeh (P. Li 2000: 96-97)

- (i) *ni taruat ki babizu*
 NMLZ T. TOP book
 'The book Taruat's.'
- (ii) *imini a syatu ka ni rakihan*
 PROX LIG clothes PTOPTOP NMLZ child
 'These clothes are the child's.'

Pazeh is not included in the fifteen languages covered in this and next chapter simply because it has become extinct and that direct access to language consultants is impossible.

²⁰⁰ This may not seem to be clear in Saisiyat, but according to M. Yeh (2000b) the verbal negator *'okay* and the nominal negator *'okik* might have resulted from the coalescence between a shared negator *'oka* and two predicate-introducing morphemes *'i* and *'ik*, respectively. See Zeitoun (2001) for additional analysis of the last morpheme.

Table 6.4: Verbal vs. nominal negators in some Formosan languages²⁰¹

Languages	Verbal NEG	Nominal NEG	NEG for possessive substantives
Thao	<i>'antu</i>	<i>'antu</i>	<i>'antu</i>
Tsou	<i>o'a</i>	<i>o'a</i>	<i>o'a</i>
Saaroa	<i>ku=</i>	<i>ku=</i>	<i>ku=</i>
Northern Paiwan	<i>ini=ka</i>	<i>ini=ka</i>	<i>ini=ka</i>
Takibakha Bunun	<i>ni</i>	<i>ni</i>	<i>ni</i>
Isbukun Bunun	<i>ni(=tu)</i>	<i>ni(=tu)</i>	<i>ni(=tu)</i>
Budai Rukai	<i>kai=</i>	<i>kai=ka</i>	<i>kai=ka</i>
Central Amis	<i>caay + Kstem</i>	<i>caay + ku</i>	<i>caay + ku</i>
Saisiyat	<i>'okay + Dep.</i>	<i>'okik</i>	<i>'okik</i>
Kavalan	<i>mai</i>	<i>usa</i>	<i>usa</i>

²⁰¹ Nominal negation in this table is defined as one that negates the equational/identificational relationship between two underived nominals. These languages typically have only one coding strategy for nominal negation. By contrast, several languages allow different negators to negate the same (or morphologically related) verbal forms in indicative main clauses (called standard negation by Miestamo 2005). For comparative purposes, verbal negation in this table is restricted to one that negates realis events with a pragmatically non-contrastive or non-emphatic overtone. The abbreviation Dep. stands for the Dependent verb form that is required after the negator in some languages. While the data for the negation of possessive predicates mostly come from my fieldnotes, the information on verbal and nominal negation is extracted from the following sources: Thao and Tsou (Y. Chen 2000), Saaroa (C. Pan 2012), Northern Paiwan (A. Chang 2006), Takibakha Bunun (Y. Jiang 2012), Isbukun Bunun (L. Huang 1997), Budai Rukai (Y. Tang 2008), Central Amis (M. Chang 2007), Saisiyat (M. Yeh 2000b), Kavalan (Y. Yeh 2005), Rikavung Puyuma (Fieldnotes; cf. Tsuchida 1980: 292 for Tamalakaw Puyuma, which is closely related to Rikavung), Taromake Rukai (P. Li 1973), Tgdaya Seediq (Holmer 1996), Pngawan Atayal (S. Shih 2007), Sqliq Atayal (L. Huang 1993). Finally, crosslinguistic discussions and historical reconstructions of Formosan negators can be found in M. Yeh et al. (1998) and S. Lin (2011), respectively.

Languages	Verbal NEG	Nominal NEG	NEG for possessive substantives
Rikavung Puyuma	'azi + Dep.	<i>melri</i>	<i>melri</i>
Taromake Rukai ²⁰²	<i>kadro(a)</i> + Kstem	<i>maela</i>	<i>maela</i>
Tgdaya Seediq	<i>ini</i> + Dep.	<i>uxe</i>	<i>uxe</i>
Pngawan Atayal	<i>ini</i> + Dep.	<i>arat</i>	<i>arat</i>
Squliq Atayal	<i>ini</i> + Dep.	<i>iyat</i>	<i>iyat</i>

(20) Budai Rukai

- a. **kai=naku wa-malra** ku paisu=su
 NEG=1SG.NOM RLS-take OBL money=2SG.NMLZ
 'I didn't take your money.' (Y. Tang 2008: 55)
- b. **kai=[*(ka) hungu=li]** kuini
 NEG=DET book=1SG.NMLZ PROX
 'This is not my book.' (Y. Tang 2008: 84)
- c. kai silu, **kai=[ka nakuane]** **kai=[ka musuane]** la
 PROX glass.bead NEG=DET 1SG.NMLZ NEG=DET 2SG.NMLZ then
kai=[ka takanav-ane]
 NEG=DET T.-NMLZ
 'This glass bead, (it's) not mine, not yours, and not Takanau's.' (C. Chen 2008: 50)

It is worth pointing out that in Budai Rukai possessive substantives that serve as the affirmative predicate have two structural patterns, depending on whether the plain nominal is a single word form or a complex nominal. If the former is involved, it is directly nominalized by *-ane* and no copula is required. However, if the plain nominal is a complex one, such as one modified by demonstratives or a verbal-based grammatical

²⁰² In teaching materials of Taromake Rukai and Puyuma, <l> stands for a retroflex lateral and <lr> an alveolar lateral. However, in all the other dialects of Rukai, it is the opposite that is the convention, namely, <lr> for a retroflex lateral and <l> for an alveolar lateral, which is the way I transcribe Taromake Rukai in this study. This tweak will make the sound value of <l> and <lr> consistent across all Rukai dialects.

argument nominalization, the copula *amani* is required, to which *-ane* affixed. The two patterns are illustrated in (21), where mixing one marking strategy with the other would lead to unacceptable results.

(21) Budai Rukai

- a. *kikai dane ka takanav-ane*
 PROX house NOM T.-NMLZ
 ‘This house belongs to Takanau.’ (C. Chen 2008: 74)
- b. *kikai senate amani-ane [kikai vavalake]*
 PROX book COP-NMLZ PROX child
 ‘This book belongs to this child.’ (Fieldnotes)
- c. *kikai senate amani-ane [kudra wa-sulrau ki camak]*
 PROX book COP-NMLZ DIST RLS-cure OBL C.
 ‘This book belongs to that one who cured Camak.’ (Fieldnotes)

Like those in Budai Rukai, possessive substantives in Saisiyat are negated like nominals by the negator *'okik*, in contrast to verbals, which are negated by *'okay*, as in (22).

(22) Saisiyat²⁰³

- a. *'oya' 'okay Sebet ka=korkoring*
 mother NEG beat UND=child
 ‘Mother didn’t beat the child.’ (M. Yeh 2000: 259)
- b. *yako 'okik SaySiyat*
 1SG.TOP NEG Saisiyat
 ‘I am not Saisiyat.’ (M. Yeh 2000: 260)
- c. *hini' tawmo' 'okik 'in=siya=a*
 PROX banana NEG NMLZ=3SG=NMLZ
 ‘This banana is not {his/hers}.’ (Kaybaybaw 2009: 13)

In all the other split-NEG languages, the verbal and nominal negators are not morphologically related, but they also negate possessive substantives the way they do nominals rather than verbals, as illustrated in Pngawan Atayal in (23).

²⁰³ See Footnote 217 below for the rationale behind the decision to treat both elements of the NMLZ *'in....a*, as well as its variant form *'an....a*, as (phonological) clitics.

(23) Plngawan Atayal

- a. **ini=cu** **aras=[ci tarasi]**
 NEG=1SG.TOP carry=UND umbrella
 ‘I didn’t bring an umbrella (with me).’ (Supplementary Materials, S.P. 5-15)
- b. **arat kur** kaca
 NEG owl that
 ‘That is not an owl.’ (Supplementary Materials, S.P. 5-1)
- c. **arat kinang/iwal-an=[ka lukus hani]**
 NEG 1SG.NMLZ/I.-NMLZ=TOP clothes PROX
 ‘This (piece of) clothing is not mine/Iwal’s.’ (Fieldnotes)

Like Plngawan Atayal, Squliq Atayal is a split-NEG language (examples not shown here, which would be similar to those in (23)a-b), and uses the nominal negator to negate possessive substantives, as in (24), where the NMRK *rwa* is in a paradigmatic relationship with a typical PUM noun in both affirmative and negative contexts.

(24) (Wulai) Squliq Atayal (Fieldnotes)

- a. **qani ga,** **biru={ku/mu/maku}**
 PROX PTOP book=1SG.NMLZ
 ‘This is my book.’
- b. **qani ga,** **iyat={ku/mu/maku}** **biru**
 PROX PTOP NEG=1SG.NMLZ book
 ‘This is not my book.’
- c. **qani ga,** **rwa={ku/mu/maku}**
 PROX PTOP NMRK=1SG.NMLZ
 ‘This is mine.’
- d. **qani ga,** **iyat={ku/mu/maku}** **rwa**
 PROX PTOP NEG=1SG.NMLZ NMRK
 ‘This is not mine.’

In almost all the languages investigated, predicative possessive substantives bear distinct forms from predicative plain nominals. In Thao, however, two exceptional situations are found. One is when a personal noun (including personal names and some kinship terms) or non-SAP person form is involved. Given the phrase *ti=X*, where X is a

personal noun, it may denote X or entities associated with X, which is largely determined by pragmatic inferences, as shown in (25).²⁰⁴

(25) Thao

- a. **haya=wa** binanaw'az, **ti=shawi**
 PROX=LIG woman PSN=S.
 'This woman is Shawi.' (S. Wang 2004: 314)
- b. **haya=wa** tamuhun, **ti=shawi**
 PROX=LIG hat PSN.NMLZ=S.
 'This hat is Shawi's.' (Fieldnotes)

The same principle applies to non-SAP person forms as well, including *thithu* for SG and *thaythuy* for PL, both of which may serve as a plain or nominalized nominal, depending on the context, as in (26).

(26) Thao

- a. **thithu** muapaw=iza
 3SG.TOP AF.appear=already
 '{She/he/it} has come out already.' (Blust 2003: 299)
- b. Q: **huya=wa** patashan manu
 DIST=LIG book whose
 'Whose book is that?' (Lit. 'That book is whose?')
- A: **thithu**=[s izui]
 3SG.NMLZ=TOP DIST
 'That is {his/hers}.' (Fieldnotes)

The other situation concerns the person form *manium*, which denotes either 2PL or entities associated with 2PL. It is the only syncretic form among SAPs (see Table 7.1 in §7.2.1). Compare, for instance, the syncretic 2PL with the non-syncretic 1EXCL in (27).

²⁰⁴ With a design like this, there might be ambiguous examples, but I have no confirmed data at hand.

(27) Thao

- a. *izai=ya patashan antu manion, nam*
 PROX=LIG book NEG 2PL.NMLZ 1EXCL.NMLZ
 ‘This book is not yours, (but) ours.’ (Fieldnotes)
- b. *alhakaiza manion latusha amusha=iza*
 when.IRR 2PL.TOP two AF.go.IRR=already
 ‘When will you two go?’ (S. Wang 2004: 298; citing Blust 2003: 433)
- c. *kahiwan yamin azazak=uan masa rima ya k<m>an*
 old.times 1EXCL.TOP child=still AF.use hand when <AF>eat
 ‘When we were still children, we would use (our hands) when eating.’
 (S. Wang 2004: 370)

6.4.2. Argument functions

For argument functions, possessive substantives can at least make up complete NPs (§6.4.2.1) and refer to something associated with the POR. In some languages, they can additionally modify another nominal (§6.4.2.2), which lead to a periphrastic modification pattern that is pragmatically more contrastive or emphatic than a more compact modification pattern, if the latter strategy is available in the first place.

In general, markers that create possessive substantives out of plain nominals are not used for verbal-based nominalizations. Budai Rukai and Plngawan Atayal are two exceptions, where reflexes of PAN *-an are used for both nominal-based and verbal-based nominalizations, as respectively illustrated in (28) and (29).

(28) Budai Rukai

- a. kabang ku ki-acebe=li, la=[ka manemane]
 bag NOM PASS-give.presents=1SG.NMLZ and.then=DET what
 ku **cegav-ane**
 NOM C.-NMLZ
 ‘My gift (i.e. what I was given) is a bag, and what is Cegau’s?’ (Fieldnotes)
- b. ka lasu, **ta-ka-dalam-ane** ki kui
 NOM guy RLS-K-love-NMLZ NMLZ K.
 ‘This guy was the one Kui loved.’ (C. Chen 2008: 105)
- c. ma-ulai=nga ku **ta-ka-tuas-ane** ki takanau
 RLS-a.while.ago=already NOM RLS-K-leave-NMLZ NMLZ T.
 ‘It’s been a while since Takanau left.’ (Lit. ‘Takanau’s leaving was already a while ago.’) (C. Chen 2008: 15)

(29) Pngawan Atayal

- a. ini pintana b<in>iniy=min=[ka tamuku]
 NEG different <PFV>buy=1EXCL.ACT=TOP hat
 ma-tanah=[ka **iwal-an**], ma-kalux=[ka **kinang**]
 AF-red=TOP I.-NMLZ AF-black=TOP 1SG.NMLZ
 ‘The hats we bought are different. Iwal’s is red (while) mine is black.’ (Fieldnotes)
- b. amol=[ka **wah-an=su** hani]
 what=TOP come-LF=2SG.ACT here
 ‘What is (the reason) you come here?’ (C. Shih 2007: 72)
- c. inu=[ka **t<in>ahk-an=su**]
 where=TOP <PFV>cook-LF=2SG.ACT
 ‘Where is (the place) you cooked?’ (C. Chen 2008: 73)

However, a verbal-based argument nominalization can be the plain nominal that becomes nominalized in the same way as basic nouns. For instance, compare a plain nominal that is a basic noun with one consisting of Focus nominalizations in (30) from Northern Paiwan and in (31) from Tsou.

(30) Northern Paiwan (Fieldnotes)

- a. vutulj n(u)a=kuka
meat NMLZ=chicken
'meat of chickens'
- b. na-pacun=a'en t(u)a=zua kaka n(u)a=[**na-ma-parang**
PFV-AF.see=1SG.TOP UND=DIST brother NMLZ=PFV-AF-bully
tjay=camak tatiav]
UND=C. yesterday
'I saw that brother of the one who bullied Camak yesterday.'
- c. na-pacun=a'en t(u)a=zua kaka n(u)a=['**<in>a-parang-an**
PFV-AF.see=1SG.TOP UND=DIST brother NMLZ=<PFV>K-bully-LF
ni=camak tatiav]
ACT=C. yesterday
'I saw that brother of the one who Camak bullied yesterday.'

(31) Tsou (Fieldnotes)²⁰⁵

- Q: zou ceopngu=[no siya] eni
EMPH hat=NMLZ who TOP.PROX
'Whose hat is this?'
- A: 'a nu=[to pasuya]
AFF NMRK=NMLZ P.
'(It's) Pasuya's.'
- A: 'a nu=[to **moso=la** **eaobako=[to mo'o]**
AFF NMRK=NMLZ AF.RLS=DSTT AF.beat=UND M.
'(It belongs to) the one who beat Mo'o.'
- A: 'a nu=[to **oh=ta=la** **eaobak-a=[to paicu]**
AFF NMRK=NMLZ NAF.RLS=3SG.ACT=DSTT beat-PF=ACT P.
'(It belongs to) the one who Paicu beat.'

Focus nominalizations (marked in bold) in both languages serve as the POR nominal just as readily as underived nouns. This once again shows that it is highly problematic to posit a head noun for Focus argument nominalizations, which are denoting expressions on their own right.

²⁰⁵ There seems to be no agreement on how the NMRK /nu/ should be transcribed. Zeitoun (2000a: 241) had *nu*, but Zeitoun (2000c: 89, 2005: 278) used *nuu* instead. And the two forms were even found in the same study (cf. S. Huang et al. 2001: 92, 198). I choose *nu* on the grounds that *nu=to* (NMRK=1INCL.NMLZ) 'ours' has the same pitch pattern as *putu* 'hammer', as opposed to *puutu* 'Han people'. More importantly, this /nu/ is believed to be related to the /nu/ as in *nenu* 'where', which is discussed in §6.6 below.

6.4.2.1. NP-use

NPs consisting only of possessive substantives are referential expressions like those consisting of underived nouns, and accordingly they fill various argument positions in the same manner as underived nouns do. As Topic NPs, possessive substantives often occur in a contrastive context, as illustrated in (32) through (34) from three languages.

(32) Amis (T. Wei 2008)

ka-ulah-an haw, u=pusi ku=**maku**
 K-like-LA PTP CMN=cat TOP.CMN=1SG.NMLZ

u=wacu ku=**misu**
 CMN=dog TOP.CMN=2SG.NMLZ

‘As for preferences, mine are cats (and) yours are dogs.’

(33) Kavalan (A. Lee 1997: 50)

misi ya=[taquq=su], qiwat a=[**za=ku**]
 AF.fat TOP=chicken=2SG.NMLZ AF.thin TOP=NMRK=2SG.NMLZ

‘Your chickens are fat (and) mine are thin.’

(34) Pngawan Atayal (Fieldnotes)

ini pintana b<in>iniy=min=[ka tamuku]
 NEG different <PFV.PF>buy=1EXCL.ACT=LIG hat

ma-tanah=[ka **iwal-an**], ma-kalux=[ka **kinang**]
 AF-red=TOP I.-NMLZ AF-black=TOP 1SG.NMLZ

‘The hats we bought are different. Iwal’s is red (and) mine is black.’

As non-Topic NPs, possessive substantives are often found in the following two contexts. One is when they serve as the argument of a denominal or a noun-incorporated verb, as shown by Paiwan in (35).

(35) Northern Paiwan (A. Chang 2006: 444; Text *tjuvak* ‘sea shells’ #63)²⁰⁶

manasika nu=sane-kava tua=[nua=va<vaya~>vayan]
therefore if=<AF>make-clothes UND=NMLZ=<PL~>woman

ini=ka pu-tjuva~tjuvak angataa aya
NEG1=NEG2 AF.have-PL~sea.shell absolutely AF.say

‘Therefore, (it was) said that if we made women’s clothes, we couldn’t possibly pick sea shells at all.’ (Lit. ‘Therefore, (it was) said that if we clothes-made women’s, we couldn’t possibly pick sea shells at all.’)

The possessive substantive *nua=va<vaya~>vayan* in (35) heads a complete NP marked by the prenominal marker *tua*, right after which is the position where the putative head noun *kava* ‘clothes’ would occur (i.e. between *tua* and *nua*). It is particularly problematic to posit such a head in this example because the noun *kava* is incorporated into the verbal predicate, which does not even permit any argument headed by the same noun. Moreover, the marking pattern of possessive substantives (i.e. nominalized nominals) follows that of plain nominals. Compare, for instance, (35) with (36), where the argument is a plain nominal.

(36) Northern Paiwan (Fieldnotes)

na-pu-aljak ti=muni tua=vavayan
PFV-AF.have-child TOP=M. UND=woman
‘Muni gave birth to a girl.’

The other context is comparative constructions, where the Topic NP serves as the comparee and a possessive substantive as the standard of comparison. This is illustrated by Isbukun Bunun in (37) and Budai Rukai in (38), where possessive substantives in the (b) examples receive the same prenominal marking as underived nouns in the (a) examples.

²⁰⁶ See W. Huang (2012: 207) for more examples of the Paiwan prefix *pu-* ‘have/contain/bear/own’.

(37) Isbukun Bunun (C. Hsieh 2011: 1, 3)

- a. ma-stán ma-taiklas=[a abus=a]=[mas aping=tia]
 AF-exceed AF-smart=TOP A.=DIST.TOP=OBL A.=DIST.NTOP
 ‘Abus is smarter than Aping.’
- b. ma-stán ma-sling=[a is-tahai=a sin-kahuzas]=[mas
 AF-exceed AF-melodious=TOP NMLZ-T.=DIST.TOP CF.PFV-sing=OBL
is-anu=tia]
 NMLZ-A.=DIST.NTOP
 ‘Tahai’s singing is more melodious than Anu’s.’

(38) Budai Rukai (Fieldnotes)

- a. ma-ku<cia~>cingalre kai senate=li ki cegau
 RLS-<CMPR~>precious PROX book=1SG.NMLZ OBL C.
 ‘My book is more precious than Cegau (i.e. the person).’
- b. ma-ku<cia~>cingalre kai senate=li ki cegav-ane
 RLS-<CMPR~>precious PROX book=1SG.NMLZ OBL C.-NMLZ
 ‘My book is more precious than Cegau’s (i.e. his book).’

6.4.2.2. Modification-use

This section demonstrates that possessive substantives can modify yet another nominal (i.e. the PUM) and help to restrict the denotation of a complex NP consisting of two nominals, thus giving rise to structurally more periphrastic possessive NPs. Importantly, whenever such a periphrastic structure is available, possessive substantives are simply juxtaposed with the PUM phrase or modify it following the morphosyntax for noun-noun modification. Moreover, when multiple modification patterns are available in a language and when there is a consistent semantic contrast between the two choices, it is always the one involving possessive substantives that is loaded with strengthened, emphatic, or contrastive meanings.²⁰⁷

²⁰⁷ On a related note, Allen (2002) calls English possessive substantives like *mine* “strengthened possessives.”

Table 6.5 demonstrates three languages where possessive substantives can modify another nominal.

Table 6.5: Possessive NPs in Kaohsiung Isbukun Bunun, Taromake Rukai, and Saisiyat

	Kaohsiung Isbukun Bunun	Taromake Rukai	Saisiyat
NP-use	<i>is-biung</i> (NMLZ-B.) 'Biung's	<i>ya=li</i> (NMRK=1SG.NMLZ) 'mine'	<i>'an='iban=a</i> (NMLZ=I.=NMLZ) 'Iban's'
Mod.-use I	<i>is-biung=[tu ahil]</i> (NMLZ-B.=LIG book) 'Biung's book'	<i>taw'ong=li</i> (dog=1SG.NMLZ) 'my dog'	<i>tatpo' ni='iban</i> (hat NMLZ=I.) 'Iban's hat'
Mod.-use II		<i>ya=li ka taw'ong</i> (NMRK=1SG.NMLZ LIG dog) 'my dog'	<i>'an='iban=a tatpo'</i> (NMLZ=I.=NMLZ hat) 'Iban's hat'

In Kaohsiung Isbukun Bunun, there is just one modification pattern whereby a possessive substantive modifies a head noun with the ligature *tu* in between. The order between a possessive substantive and the noun it modifies can be flexible, as shown in (39).²⁰⁸

(39) Kaohsiung Isbukun Bunun (C. Shi 2009: 128)

- a. saikin hai, mazima=[mas is-puan=tu uváz]
1SG.TOP P TOP AF.like=UND NMLZ-P.=LIG child
'I like Puan's child.'
- b. saikin hai, mazima=[mas uváz=tu is-puan]
1SG.TOP P TOP AF.like=UND child=LIG NMLZ-P.
'I like Puan's child.'

By contrast, there are two modification patterns in Saisiyat and Taromake Rukai, where the structurally more complex one (i.e. Mod.-use II) involves possessive substantives. In

²⁰⁸ However, my Isbukun consultants consistently prefer the order in (39)a.

Taromake Rukai, possessive substantives are connected with the noun they modify by the marker *ka*, which is also required between two underived nouns, as shown in (40).

(40) Taromake Rukai²⁰⁹

- a. *ya=li* *ka* *taw'ong*
 NMRK=1SG.NMLZ LIG dog
 ‘my dog (or the dog that is mine)’ (P. Li 1973: 79)
- b. *lrolay ka ababay*
 kid LIG *female*
 ‘girl’ (Fieldnotes)

Finally, possessive substantives in Saisiyat are marked by *'an....a*, and the derived nominal is simply juxtaposed to the noun they modify, as shown in Table 6.5.²¹⁰

A more complicated situation is found in Paiwan, where possessive substantives comprised of free nominals are allowed to modify a noun either prenominally or postnominally, but with different syntagmatic requirements, as shown in (41) (see also §7.2.3 for a similar situation in Central Amis).

(41) Northern Paiwan (C. Tang 2006: 951-952; citing C. Tang et al. 1998)

- a. *kun* *ni=kai*
 skirt NMLZ=K.
 ‘Kai’s skirt’
- b.* *kun=a* *ni=kai*
 skirt=LIG NMLZ=K.
- c.* *ni=kai* *kun*
 NMLZ=K. *skirt*
- d. *ni=kai=a* *kun*
 NMLZ=K.=LIG *skirt*
 ‘Kai’s skirt’

²⁰⁹ See Footnote 202 for the tweak on the sound value of <l> and <lr> in Taromake Rukai.

²¹⁰ The NMLZ *'an...a* is in free variations with *'in...a* (see (22)c on p.366). In this study, except for data taken from the literature, only the former form is illustrated because it is what my consultant uses spontaneously. See Footnote 217 below for the rationale behind the decision to treat both elements of the NMLZ *'an...a* as (phonological) clitics.

The attributive ligature *a* is required in prenominal modification but prohibited in postnominal modification. Elsewhere, the attributive ligature is used in various modification constructions, including noun-noun modification in (42).

(42) Northern Paiwan (Fieldnotes)²¹¹

- a. *kasiv=a tjekeza*
 wood=LIG bridge
 ‘wood bridge’
- b. *tjuvu=a siav*
 bamboo.shoot=LIG soup
 ‘bamboo shoot soup’

Question-answer pairs in (43) illustrate how NP structures in (41) fit into larger syntax.

(43) Northern Paiwan (Fieldnotes)

- Q: *anema[=a su=ki~kim-en]?*
 what=TOP 2SG.ACT=IPFV~search-PF
 ‘What are you looking for?’
- A: (*ki~kim=a’en tua*)=[**sunatj ni=camak**]
 <AF>IPFV~search=1SG.TOP UND=book NMLZ=C.
 ‘(I’m looking for) Camak’s book.’
- A: (*ki~kim=a’en tua*)=[**ni=camak=a sunatj**]
 <AF>IPFV~search=1SG.TOP UND=NMLZ=C.=LIG book
 ‘(I’m looking for) Camak’s book (not somebody else’s).’

Although (41)b is not an acceptable NP, the syntagm of a noun followed by /a/ and a *ni*-phrase can be felicitous in a context like (44), where the *ni*-phrase serves as the sole argument of a nominal predicate.

²¹¹ An alternative expression for “bamboo shoot soup” is *p<in>u-tjuvu-an=a siav* (<PFV>have-bamboo.shoot-LF=LIG soup), which is again made up of two nominals connected by the ligature *a*.

(44) Northern Paiwan (Fieldnotes)

Q: azua=(a) 'u=umiyaki, kabang. anema=[(a) **ni=camak**]
 DIST=LIG 1SG.NMLZ=gift bag what=TOP NMLZ=Camak
 'As for my gift, (it's) a bag. What about Camake's?'

A: sunatj=[a **ni=camak**]
 book=TOP NMLZ=C.
 'Camak's is a book.'

A:* **ni=camak=a** sunatj
 NMLZ=C.=LIG book
 [Not accepted in this context, but accepted as an answer in the context of (43)]

The key factor here is that the attributive ligature and the Topic marker for common nouns bear the same form in Paiwan, and the /a/ in the second example of (44) illustrates the Topic-marking function.

On the other hand, Paiwan possessive substantives composed of bound person forms share the same structural patterns as those consisting of free nominals. Besides, there is an additional synthetic modification pattern available for SAP person forms (but not for non-SAP ones), where nominalized person clitics are directly attached to the PUM. Example (45) demonstrates three types of modification patterns with 1SG as the POR.

(45) Northern Paiwan (Fieldnotes)

Q: anema=[(a) su=ki~kim-en]?
 what=TOP 2SG.ACT=IPFV~search-PF
 'What are you looking for?'

A: (ki~kim=a'en tua)=[**'u=sunatj**]
 <AF>IPFV~search=1SG.TOP OBL=1SG.NMLZ=book
 '(I'm looking for) my book.'

A: (ki~kim=a'en tua)=[sunatj **ni=a'en**]
 <AF>IPFV~search=1SG.TOP OBL=book NMLZ=1SG
 '(I'm looking for) my book.'

A: (ki~kim=a'en tua)=[**ni=a'en=a** sunatj]
 <AF>IPFV~search=1SG.TOP OBL=NMLZ=1SG=LIG book
 '(I'm looking for) my book (not somebody else's).'

Unlike that in Taromake Rukai, the synthetic indexing strategy in Paiwan does not have the NP-use, which can only be expressed by *ni*-phrases, as in (46).

(46) Northern Paiwan (Fieldnotes)

Q: *azua=(a) 'u=umiyaki, kabang. anema=[(a) ni=sun]*
 TOP.DIST=LIG 1SG.NMLZ=gift bag what=TOP NMLZ=2SG
 'As for my gift, (it's) a bag. What about yours?'

A: *sunatj=[a ni=a'en]*
 book=TOP NMLZ=1SG
 'Mine is a book.'

A: *azua=(a) ni=a'en, sunatj*
 TOP.DIST=LIG NMLZ=1SG book
 'As for that one of mine, (it's) a book.'

The three structural types of possessive NPs in Paiwan are summarized in Table 6.6.

Table 6.6: Three types of possessive NPs in Northern Paiwan

POR types	Direct indexing (I)	Postnominal modification (II)	Prenominal modification (III)
SAP person forms	<i>'u=sunatj</i> (1SG.NMLZ=book) 'my book'	<i>sunatj ni=a'en</i> (book NMLZ=1SG) 'my book'	<i>ni=a'en=a sunatj</i> (NMLZ=1SG=LIG book) 'my book'
Non-SAP person forms	---	<i>sunatj ni=madju</i> (book NMLZ=3SG) 'his/her book'	<i>ni=madju=a sunatj</i> (NMLZ=3SG=LIG book) 'his/her book'
Full nominals	---	<i>sunatj ni=camak</i> (book NMLZ=C.) 'Camak's book'	<i>ni=camak=a sunatj</i> (NMLZ=C.=LIG book) 'Camak's book'

The Paiwan data in Table 6.6 show that SAP person forms have the most coding possibilities available. Direct indexing is presumably the oldest pattern since the structure is widely found in Austronesian languages and that the bound person forms involved can be traced back to PAn (Dyen 1974). By contrast, bound person forms collocating with *ni*

have been argued to be innovations from a set of free forms in a proto-language (see Ross 2006).

Moreover, comparisons across dialects also suggest how person forms might have acquired new modification patterns. Table 6.7 compares possessive NPs in Kaohsiung Isbukun Bunun and Takibakha Bunun.

Table 6.7: Possessive NPs in two Bunun dialects

Languages	POR types	Mod.-use I	NP-use	Mod.-use II
Kaohsiung Isbukun Bunun	SAP person forms	---	<i>i-nák</i> (NMLZ-1SG) 'mine'	<i>i-nák=tu ahil</i> (NMLZ-1SG=LIG book) 'my book'
	Non-SAP person forms	---	<i>i-sai=cia</i> (NMLZ-3SG=DIST.NTOP) 'his/hers'	<i>i-sai=cia=tu ahil</i> (NMLZ-3SG=DIST.NTOP=LIG book) 'his/her book'
	Full nominals	---	<i>is-biung</i> (NMLZ-B.) 'Biung's'	<i>is-biung=tu ahil</i> (NMLZ-B.=LIG book) 'Biung's book'
Takibakha Bunun	SAP person forms	<i>tamuhung=nak</i> (hat=1SG.NMLZ) 'my hat'	<i>i=nak</i> (NMRK=1SG.NMLZ) 'mine'	<i>i=nak tamuhung</i> (NMRK=1SG.NMLZ hat) 'my hat'
	Non-SAP person forms	<i>tamuhung=cia</i> (hat=3SG.NMLZ) 'his/her hat'	<i>i=cia=ta</i> (NMRK=3SG.NMLZ=DIST) 'his/hers'	<i>i=cia=ta tamuhung</i> (NMRK=3SG.NMLZ=DIST hat) 'his/her hat'
	Full nominals	---	<i>i=cia savi</i> (NMRK=3SG.NMLZ S.) 'Savi's'	<i>i=cia savi tamuhung</i> (NMRK=3SG.NMLZ S. hat) 'Savi's hat'

In Takibakha Bunun, person forms (both SAPs and non-SAPs) have two modification patterns, with one involving direct indexing (i.e. Mod.-use I) and the other based on

possessive substantives (i.e. Mod.-use II). In Kaohsiung Isbukun Bunun, however, direct indexing is not attested (R. He et al. 1986; S. Huang 1997: 364; S. Huang et al. 1999: 169), though it has been reported to exist in Nantou Isbukun Bunun (P. Li 1997a: 317, 1997b: 364). L. Huang (1997: 370) thus raised the question of whether the Nantou variety has innovated a set of nominalized person indexes that attach directly to nouns or the Kaohsiung variety has lost such a set. Considering that direct indexing with similar person forms is found not only in Northern dialects (including Takibakha and Takituduh),²¹² which are the most conservative among Bunun (P. Li 1988), but also in Nantou Isbukun Bunun, it seems much more likely that Kaohsiung Isbukun Bunun has lost the direct indexing strategy, which predated the more recent development of a modification pattern based on possessive substantives.

6.5. A typology of possessive substantives

This section reinterprets and expands Ultan's (1978) typology of possessive substantives in light of the nominalization-based account and Formosan data. The general remark in this section is primarily based on full-nominal possession, although similar principles also hold true for person-form possession.

Drawing evidence from a sample of 75 languages, Ultan (1978) concluded that possessive substantives fall into two types. In Type 1 the possessive substantive results from eliding the PUM in the POR-PUM syntagm, such as English *Tom's* as in *John's hat is cheaper than Tom's*. In Type 2 the possessive substantive derives from replacing the PUM in the POR-PUM syntagm with what Ultan called "determination markers",

²¹² POR-indexes on the PUM in Takituduh Bunun are documented by Tsuchida (1992b: 732).

including determiner-like markers, demonstratives, definiteness markers, or markers that agree with the PUM. The assumption in his typology is that possessive substantives are morphosyntactic derivatives of the POR-PUM syntagm. In the present nominalization-based account, his Type 1 languages have a nominal-based NMLZ, like English 's, which creates a new nominal that can have either the NP-use or modification-use. Determination markers in his Type 2 languages are rather heterogeneous, with some of them corresponding to NMLZ and others to NMRK. For instance, one example he gave is Italian agreement markers as in *la tu-a* (DEF.F.SG 2SG-F.SG) 'yours'. However, that particular form also has the modification-use, as in *la tu-a amic-a* (DEF.F.SG 2SG-F.SG friend-F.SG) 'your (female) friend'. Thus, these Italian agreement markers are NMLZ in the present terminology, much like those in Hindi (see (3) above). On the other hand, there are other types of agreement markers that only occur in the NP-use, such as those in German, as shown in (47).

(47) German (Lehmann et al. 2004: 56)

- a. Der Bleistift ist mein-er.
 DEF.M.SG.NOM pencil COP.3SG.PRS 1SG-M.SG.NOM
 'The pencil is mine.' (cf. *mein Bleistift* 'my pencil' in NOM)
- b. Das Buch ist mein-(e)s.
 DEF.N.SG.NOM book COP.3SG.PRS 1SG-N.SG.NOM
 'The book is mine.' (cf. *mein Buch* 'my book' in NOM)

These agreement markers are NMRK in the present terminology since they only occur in the NP-use.²¹³

In addition to Ultan's two types, Formosan data show that a third type needs to be recognized. This is so even if we only consider languages that retain PAN *ni, reflexes of which mark personal nouns as the POR (see §7.1 for all the three types irrespective of

²¹³ Lehmann et al. (2004: 57) is one of the few studies that conceptualize possession as nominalization because they refer to the German possessive substantives in (47) as "nominalized possessive attribute."

cognacy). Six out of the fifteen Formosan languages investigated here retain PAN *ni. They fall into three types as schematically represented in Table 6.8, and one language from each type is illustrated in Table 6.9 below.²¹⁴

Table 6.8: Three types of Formosan languages that retain PAN *ni

Language types	Modification-use	NP-use	Languages
Type A	PUM + [<i>ni</i> +POR]	[<i>ni</i> +POR]	Northern Paiwan Central Amis
Type B	PUM + [<i>ni</i> +POR]	NMRK + [<i>ni</i> +POR]	Kavalan Rikavung Puyuma
Type C	PUM + [<i>ni</i> +POR]	[X+POR]	Plngawan Atayal Saisiyat

Note: X is any marker that is not a reflex of PAN *ni. The brackets indicate morphosyntactic affiliations, not necessarily phonological ones, and [X+POR] represents a schematic constituent unit without being committed to a fixed linear morpheme order.

Type A and B correspond to Ultan's Type 1 and 2. Type C is different from either of them because its possessive substantive is neither an elliptical nor substitutional version of the POR-PUM syntagm. Notice that the suffix *-an* for the NP-use in Plngawan Atayal could not have replaced the PUM in the modification-use because they occur on opposite sides of the POR (see Table 6.9 below).

More importantly, the nominalization-based perspective also reveals how the three types in Table 6.8 might have arisen. If we only look at the modification-use, the POR-PUM syntagm in the six Formosan languages looks so similar that the formula

²¹⁴ Moreover, Pazeh is also a Type A language with the reflex of PAN *ni according to P. Li's (2000) data (see Footnote 199). It is just that Pazeh falls outside the fifteen languages investigated here. In fact, Blust & Trussel's Austronesian Comparative Dictionary (http://www.trussel2.com/acd/acd-s_n1.htm#8781) lists only six Formosan languages that have reflexes of PAN *ni, including Pazeh and all the languages in Table 6.8 minus Plngawan Atayal.

Table 6.9: Possessive NPs in Type A through C languages

Languages	Modification-use	NP-use
Type A Northern Paiwan	<i>sunatj ni=camak</i> book NMLZ=C. 'Camak's book'	<i>ni=camak</i> NMLZ=C. 'Camak's'
Type B Kavalan	<i>lepaw ni=abas</i> house NMLZ=A. 'Abas's house'	<i>*ni=abas</i> NMLZ=A. <i>za=ni=abas</i> NMRK=NMLZ=A. 'Abas's'
Type C Pngawan Atayal	<i>lukus=[ni iwal]</i> clothes=NMLZ I. 'Iwal's clothes'	<i>*ni=iwal</i> NMLZ=I. <i>iwal-an</i> I.-NMLZ 'Iwal's'

“*PUM ni POR” can be reconstructed in PAN with much confidence, as has been done in previous studies (Reid 1981, 2007; Ross 2002; Blust 2005, 2015). However, a question rarely asked is what PAN was like when the POR was not modifying any lexical noun. When the NP-use is taken into consideration, distributions of the so-called genitive marker start to diverge. It is by comparing both the NP- and modification-use pattern that the three types are definable. Table 6.8 shows that presence of the /ni/ marker in the NP-use implies its presence in the modification-use, but not vice versa. In other words, the /ni/ marker is found everywhere except for the NP-use in Type C languages, suggesting it is where innovations might have taken place. If so, chances are that both Pngawan Atayal and Saisiyat have innovated a new mechanism for marking the NP-use, thus leaving the /ni/ marker trapped in the modification context (Shibatani, p.c.).

Pngawan Atayal uses the suffix *-an* (< PAn **-an*) to create possessive substantives, but this morphological process is not applicable to all semantic classes of nominals. Considering cognate forms with reflexes of PAn **-an* in other languages predominantly express location, Pngawan Atayal might have undergone the common grammaticalization change from location to possession (see Heine 1997). The fact that only a subset of nominals can be nominalized by *-an* suggests that the grammaticalization process is not fully complete yet. Specifically, *-an* suffixation in Pngawan is only applicable to personal nouns, but not to common nouns like *kanel* ‘woman’. Personal nouns in this language include at least person-form roots, personal names, and personal interrogative words (i.e. *ima* ‘who’). According to L. Huang (2006: 217), personal nouns also cover kinship terms. However, it is found that kinship terms in Pngawan do not have consistent grammatical properties (see §6.8 and Dahl & Koptjevskaja-Tamm 2001). While some kinship terms are treated like personal nouns, others behave like common nouns. For instance, while *yaba* ‘father’ can be suffixed by *-an* to become a nominalized nominal in the NP-use, *suse* ‘younger sibling’ cannot, as contrasted in (48).

(48) Pngawan Atayal (Fieldnotes)

Q: *ima-n patas hani*
 who-NMLZ book PROX
 ‘Whose book is this?’

A: ***yaba-n=mu***
 father-NMLZ=1SG.NMLZ
 ‘my father’s’ [NP-use]

A: *patas=[ni yaba=mu]*
 book=NMLZ father=1SG.NMLZ
 ‘my father’s book’ [Modification-use]

- A:* **suse-n=mu**
 younger.sibling-NMLZ=1SG.NMLZ
- A: **patas=[na suse=mu]**
 book=NMLZ younger.sibling=1SG.NMLZ
 ‘my younger sibling’s book’ [Modification-use]

While nouns eligible for *-an* suffixation in the NP-use select the NMLZ *ni* in the modification-use, those ineligible ones opt for the NMLZ *na* in the modification-use. Aside from *yaba* ‘father’, kinship terms eligible for *-an* suffixation include *yaya* ‘mother’ (as in *yaya-n=mu* ‘my mother’s’), *yutas* ‘grandfather’ (as in *yatas-an=mu* ‘my grandfather’s’), *yaki* ‘grandmother’ (as in *yake-n=mu* ‘my grandmother’s’), *mama* ‘uncle’ (as in *mama-n=mu* ‘my uncle’s’), *yata* ‘aunt’ (as in *yata-n=mu* ‘my aunt’s’). On the other hand, like *suse* ‘younger sibling’, *asuran* ‘older sibling’ cannot undergo *-an* suffixation (hence **asuran-an=mu*).²¹⁵ Given the current data, it seems that only ascending kin terms, but not descending or horizontal ones, are treated like personal nouns, though additional research is required to confirm this. At any rate, there is a lexical restriction on what nouns can be suffixed by *-an* to be used as possessive substantives in the NP-use.²¹⁶ This situation suggests a recent development of this

²¹⁵ Aside from the rationale enumerated in Footnote 25, another reason for not transcribing the word-final glottal stop in Pngawan Atayal (e.g. <yaba> rather than <yabaʔ> for [jaβaʔ] ‘father’) is that the glottal stop disappears after suffixation, namely, when it is not word-final. If the preglottal stop vowel is /i/, the sequence /ia/ is then monophthongized to /e/ (see C. Shih 2007: 16). Hence, after undergoing *an*-suffixation, *yaki* [jakiʔ] ‘grandmother’ becomes *yaken* [jaken] due to monophthongization.

²¹⁶ By contrast, Budai Rukai also uses reflexes of PAN **-an* to create possessive substantives, but it does not demonstrate the kind of lexical restriction found in Pngawan Atayal. All kinds of nouns, including inanimate ones, can be nominalized by *-ane*, as shown below.

Budai Rukai (H. Jiang 2013)

- (i) *cegav-ane kikai talrupunu*
 C.-NMLZ PROX hat
 ‘This hat is Cegau’s.’
- (ii) *taupong-ane kikai valisi*
 dog-NMLZ PROX tooth
 ‘This tooth belongs to a dog.’
- (iii) *tatukul-ane kikai vasaw*
 Taiwan.acacia-NMLZ PROX leaf

encoding strategy in Pngawan Atayal, which seems to be the only Atayalic language that uses reflexes of PAN *-an for such a purpose, given the data available to date.

In Saisiyat, on the other hand, a personal noun is nominalized by *ni* in the modification-use, but by *'an....a* in the NP-use, a unique marking not found in any other Formosan languages, as shown in (49).

(49) Saisiyat (Fieldnotes)

- a. **tatpo' ni='iban**
 hat NMLZ=I.
 'Iban's hat' [Modification-use]
- b. **hini' tatpo' (mina), 'an='iban=a**
 PROX hat EMPH NMLZ=I.=NMLZ
 'This hat is (indeed) Iban's.' [NP-use]

By contrast, a common noun is nominalized by *noka* in the modification-use, and unlike its counterpart *ni*, *noka* may optionally collocate with *'an....a* in the NP-use, as in (50).

(50) Saisiyat (Fieldnotes)²¹⁷

- a. **tatpo' noka=korkoring**
 hat NMLZ=child
 '{a/the} child's hat' [Modification-use]
- b. **hini' tatpo' (mina), 'an=(noka)=korkoring=a**
 PROX hat EMPH NMLZ1=NMLZ2=child=NMLZ1
 'This hat is (indeed) the child's.' [NP-use]
- c. **hini' tatpo' (mina), 'an=(noka)=hini' korkoring=a**
 PROX hat EMPH NMLZ1=NMLZ2=PROX child=NMLZ1
 'This hat is (indeed) this child's.' [NP-use]

The distributions above suggest that markers in the modification-use (i.e. *ni* and *noka*) are giving way to those newly arising ones in the NP-use (i.e. *'an....a*) such that *noka* is now optional and *ni* prohibited in the NP-use. Moreover, judging from the paradigmatic

'This leaf belongs to the Taiwan acacia.'

²¹⁷ Co-occurrence of *'an....a* and *noka* is also found in M. Yeh (2003: 16). Example (50)c also shows that the beginning element of *'an....a* phonologically attaches to the left periphery of a complex phrase (i.e. *hini' korkoring* 'this child'), and its ending element, to the right edge of that phrase. This property, together with the fact that both of the elements are unaccented on their own, supports the current analysis of both formatives as phonological clitics, although they are commonly transcribed as free forms in the Saisiyat literature (e.g. M. Yeh 1991, 2003).

alternation between a PUM noun in (50)a and the *'an* marker in (50)b, it can be inferred that the *'an* marker may have started out as a NMRK indicating the NP-use, much like *za* in Kavalan (see Table 6.9 above), at a time when *noka* was consistently used in both the modification- and NP-use. It is only when *noka* later became optional in the NP-use, as in (50)b, that the *'an* marker replaced it as the new-generation NMLZ. Last, the /a/ part of the NMLZ *'an....a* has a rather curious history. A reviewer of L. Huang et al. (1998: 46) pointed out that the /a/ marker might be an attributive ligature, which they considered “rather plausible” but did not accept. I agree with L. Huang et al.’s (1998) view, but on different grounds. While their reason was that person forms marked by *'an* all end with /a/ (see Table 7.8 for all the SAP forms), I take the obligatory presence of /a/ in the NP-use to be the main argument against the ligature analysis. In the NP-use like (50)b~c, there is no modifiee following the /a/ part, so it does not qualify as an attributive ligature by definition. In languages where /a/ does function as a ligature, such as Thao, Amis, Kavalan, and Paiwan (see (42) above), it always links two constituents. Nevertheless, it is still possible that the /a/ part was historically a ligature in the modification context like *'an='iban=a tatpo* ‘Iban’s hat’, but has been reanalyzed as an inseparable part of the first constituent due to phonological affiliation. Consequently, the erstwhile ligature /a/ sticks around with or without a following constituent. If that is the case, whether possessive person forms all end with /a/ is irrelevant since *'an....a* is a fixed frame for a nominal phrase, be it a full nominal or person-form root. All the points above suggest the *'an....a* marking for the NP-use is a relatively later development.

If Type C is an innovation as reasoned out above, PAN would most likely belong to either Type A or Type B, the decision of which hinges upon whether or not PAN used a

NMRK in the NP-use on top of the schema “*ni POR”. To address this question, we need to look into functions of another PAn etymon, namely *nu, which is also found in possessive NPs. This is to be discussed in the next section.

6.6. PAn *nu reflexes and their implication on the development across types

While there is a consensus among historical linguists that the form *nu is reconstructable in PAn, opinions are divided with regards to its function in PAn. Ross (2002) and Blust (2005, 2015) both reconstructed the function of PAn *nu as “genitive of common nouns” (as opposed to genitive of personal nouns, which was PAn *ni), but they differ in what kinds of common nouns PAn *nu marked. By contrast, Reid (1981, 2007) proposed that PAn *nu was not a genitive marker at all, but a “non-referential noun” meaning “thing” (Reid 2007: 250). He cited two major pieces of evidence to support his proposal. One is that reflexes of *nu form the base of interrogative words in many Austronesian languages, both outside and within Formosan, with the latter illustrated in Table 6.10 below.²¹⁸

²¹⁸ Reid (2007: 250) in turn cited Ferrell (1969).

Table 6.10: Formosan interrogative words with the PAn *nu (after Reid 2007: 250)

Squliq Atayal	<i>na-nu</i> <i>i-nu</i>	‘what’ ‘where’
Seediq	<i>ma-nu</i> <i>i-nu</i> <i>ka-nu-wan</i>	‘what’ ‘where’ ‘when’
Favorlang	<i>nu-mma</i>	‘what’
Thao	<i>nu-ma</i>	‘what’
Saisiyat	<i>ka-no</i> <i>hay-no</i> <i>i-no-wan</i>	‘what’ ‘where’ ‘when (future)’

The other piece of evidence is that *nu “could be followed by a genitive NP (‘thing of the man’, etc.) and could have genitive clitics attached to it to signify nouns of absolute possession” (Reid 2007: 250). While he did not illustrate the first function in the quote or specify which language has it, he exemplified the second function with Amis forms such as *nu=maku* ‘mine’, *nu=misu* ‘yours (SG)’, and *nu=mita* ‘ours (INCL)’, although these person forms without *nu* are in fact independent nominals rather than clitics (see Table 7.12).²¹⁹ In this section, I present additional data from Formosan languages, which lend more support to Reid’s (1981, 2007) functional reconstruction of PAn *nu than Ross’s (2002) or Blust’s (2005, 2015).

Blust (2015) cited three Formosan languages in support of his functional reconstruction of PAn *nu as a genitive marker for common nouns, including Saisiyat (where the formula is “PUM noka=POR”), Amis (where the formula is “PUM

²¹⁹ Reid (2007: 250) mistakenly specified *nu=mita* as the exclusive form, which has been corrected here.

nu=POR”), and Paiwan (where the formula is “PUM *nu*a=POR”). On the other hand, according to his Austronesian Comparative Dictionary, the *nu found in interrogative words (see Table 6.10) was reconstructed as a “marker of uncertainty” in PAn.²²⁰ In other words, under Blust’s reconstructions, the *nu in possessive NPs and that in interrogative words just happen to be homophonous. However, if we look beyond the POR-PUM syntagm and include possessive substantives (i.e. the NP-use of nominalized nominals), even Tsou and Seediq can be potentially shown to retain PAn *nu, in both possessive NPs and interrogative words, just like Saisiyat and Paiwan, as summarized in Table 6.11 below. As more languages are brought into the picture, the homophonous account becomes more tenuous than the single-morpheme account.

First, the Tsou marker *nu* is only found in the NP-use and thus in a paradigmatic relationship with PUM nouns, as illustrated in (51), so it falls outside the usual radar of so-called genitive markers.

(51) Tsou (Fieldnotes)²²¹

- a. (zou) **nu='u**=[’e ceopngu]
 EMPH NMRK=1SG.NMLZ=TOP hat
 ‘The hat (here) is mine.’ [cf. ’oko=’u ‘my child’]
- b. (zou) **nu=(si)=[to paicu]**=[’e ceopngu]
 EMPH NMRK=3SG.NMLZ=NMLZ P.=TOP hat
 ‘The hat (here) is Paicu’s (with her not being around).’
 [cf. ’oko=(si)=to paicu ‘Paicu’s child’]

²²⁰ http://www.trussel2.com/acd/acd-s_n1.htm#3687

²²¹ Markers that alternate with *to* include *ta* and *no*. They only differ in terms of whether the NP referent they mark is (in)visible or/and (un)witnessed (H. Chang 2011). Only *to* is illustrated in Table 6.4 because it has the highest text frequency among the three markers based on H. Huang’s (2010: 95) natural discourse data. See Footnote 205 regarding the transcription of the NMRK *nu*.

Table 6.11: Possessive NPs and interrogative words in some Formosan languages

Languages	Possessive NPs		Interrogative words with the PAn *nu
	NP-use	Modification-use	
Tsou	<i>nu</i> =[<i>to</i> + POR]	PUM=[<i>to</i> + POR]	<i>ne-nu</i> ‘where; which one’ <i>mai-ne-nu</i> ‘how, what kind’
Seediq	<i>n</i> =POR	PUM + (<i>n</i>)=POR	<i>i-nu</i> ‘where’ <i>ma-nu</i> ‘what’ <i>ka-nu-wan</i> ‘when’
Saisiyat	<i>'an</i> =(<i>noka</i> =)POR= <i>a</i>	PUM + <i>noka</i> =POR	<i>ka-no</i> ‘what’ <i>kay-no</i> ‘where’ <i>i-no-wan</i> ‘when’
Paiwan	<i>nua</i> =POR	PUM + <i>nua</i> =POR	<i>i-zai-nu</i> ‘where’

Note: POR and PUM in this table stand for a full-nominals possessor and possessum NP respectively.

However, it does not seem to be accidental that Tsou interrogative words like *nenu* ‘where, which one’ and *mainenu* ‘how, what kind’ both contain /nu/, just as in other Formosan languages where reflexes of *nu do function as a so-called genitive marker (i.e. by having the modification-use) and as a component morpheme of interrogative words. Although the /nu/ part in *nenu* ‘where, which one’ is rarely connected to the NMRK *nu* in possessive substantives, their syntactic environments are far too strikingly similar to be a mere coincidence. Compare, for instance, (51)b with (52).

(52) Tsou (S. Huang et al. 2001: 200)

nenu=(si)=[*na mo ea-chumu*]
 where=3SG.TOP=TOP AF.RLS have-water
 ‘Where is (the place that) has water?’

Like the NMRK *nu* in (51)b, *nenu* ‘where, which one’ can be optionally followed by /si/, which is relatable to the same /si/ that indexes the third-person singular Topic, non-Topic Actor, and possessor. This would then explain why there are both *nenu* and *nenusi*, although the latter was treated as an unanalyzable alternative form of the former by S. Huang et al. (2001: 200).²²² As for the /ne/ part in *nenu* ‘where, which one’, it can be related to the locative marker *ne*, as in (53).

(53) Tsou (H. Huang 2010: 158)

- a. **ta=ta=la** **aoko** **uh=[ne** **taihoku]**
 AF.IRR=3SG.TOP=DSTT AF.always AF.go=LOC Taipei
 ‘{She/He} will often go to Taipei.’

Hence, *nenu* in Tsou would be literally “at what place,” thus showing a morphological makeup comparable to *inu* ‘where’ in Seediq and *izainu* ‘where’ in Paiwan, all composed of a locative marker (/ne/ in Tsou and /i/ in Seediq and Paiwan) and a reflex of PAn *nu denoting unspecified entities, which is also used to create possessive substantives in all the three languages. More importantly, the connection between reflexes of PAn *nu in possessive NPs and those in interrogative words goes well with the widespread typological trend that interrogative and indefinite words are “either identical in form, or derivationally related, in the majority of the world’s languages” (Bhat 2004: 226).

Second, Table 6.4 also shows that if reflexes of PAn *nu are applicable to the modification-use, it follows that they are also applicable to the NP-use, but not vice versa. This suggests that the original habitat of PAn *nu was in the NP-use, as is still the case now for the Tsou marker *nu*, and that it is only when reflexes of PAn *nu were later expanded to the modification-use that they started to function like a genitive marker. This

²²² Moreover, *nenusi no* is used at the sentence-initial position as a frozen expression to “convey epistemic modality” (see S. Huang & H. Huang 2003: 24).

development is still somewhat observable in Tgdaya Seediq, where *n* (realized as [nu] or [ne] depending on phonological environments), when combining with a full nominal, is obligatory in the NP-use but optional in the modification-use, as illustrated in (54).

(54) Tgdaya Seediq (Fieldnotes)

- a. **n=mona** ka=sapah nii han
 NMRK=M.NMLZ TOP=house PROX PRT
 ‘This house is Mona’s.’
- b.* **mona** ka=sapah nii han
 M. TOP=house PROX PRT
- c. sapah **mona**
 house M.NMLZ
 ‘Mona’s house’
- d. sapah **n=mona**
 house NMRK=M.NMLZ
 ‘house of Mona’s’

The *n* in (54)a for the NP-use is in a paradigmatic relationship with the PUM noun *sapah* ‘house’ in (54)c for the modification-use, suggesting that *n* was originally meant to denote unspecified entities in the NP-use. As *n*-marked nominals are allowed to modify a PUM noun, the more marked modification pattern in (54)d arises, a step that the *nu*-phrase in Tsou has not taken.

Interestingly, despite its functions, *n* in Tgdaya Seediq is typically not considered a genitive marker in the literature (e.g. H. Chang 1997, 2000; but see Ochiai 2009: 23), thus rendering obscure its possible connection with the PAn *nu. The one referred to as the genitive marker in the Seediq literature is *na*, which also derives possessive substantives (see §7.1.1). The differential treatment of the two markers is mostly due to the fact that *n*-marked phrases do not also encode the (non-Topic) Actor of a NAF-word whereas *na*-marked phrases do. In other words, *n*-marked phrases collocate with nouns but not with verbs, and this aligns well with the idea that the marker *n* has historically

evolved from a generic thing-denoting formative such that even synchronically speaking expressions like *sapah n=mona* ‘car of Mona’s’ are appositive in nature (lit. ‘car, Mona’s thing’). In addition, the idea that Seediq *n* is a reflex of PAN **nu* is indirectly supported by the fact that the marker attaches to a complete paradigm of free person-based forms that already denote entities associated with persons (e.g. *naku* ‘mine’ vs. *n=naku* ‘mine’; see Table 7.11 for more), just as does Amis *nu*, which is no doubt a reflex of PAN **nu* (e.g. *maku* ‘mine’ vs. *nu=maku* ‘mine’; see Table 7.12 for more).

Third, the strongest evidence against PAN **nu* being a genitive markers (for common nouns) that alternate with PAN **ni* (for personal nouns) probably comes from Amis, one of the three Formosan languages where reflexes of PAN **ni* and **nu* are both found and demonstrate paradigmatic alternations (the other two being Saisiyat and Paiwan). As a Type A language, Amis selects *ni* and *nu* to create nominalized nominals based on personal and common nouns respectively, and these nominalized nominals have both the modification-use, as in (55), and the NP-use, as in (56).

(55) Central Amis (J. Wu 2006: 80, 306)

- a. *mi-la’up* *ku=wacu* *tu=[wawa* ***ni=panay]***
 AF.EXT-chase TOP.CMN=dog UND.CMN=child NMLZ.PSN=P.
 ‘The dog is chasing Panay’s child.’
- b. *aka* *caliw-en* *ku=[paysu* ***nu=wawa]***
 NEG.IMP borrow-PF TOP.CMN=money NMLZ.CMN=child
 ‘Don’t borrow the child’s money!’

(56) Central Amis (Fieldnotes)

- a. ira ku='udax aku k<um>aen=ho kaku
 EX TOP.CMN=candy 1SG.NMLZ AF.<UM>eat=still 1SG.TOP

tu=[ni=kacaw]

UND.CMN=NMLZ.PSN=K.

'I have candies, (but) I'm eating Kacaw's (regardless).' (Based on Huang 1995: 231)

- b. ira ku='udax aku, k<um>aen=ho kaku
 EX TOP.CMN=candy 1SG.NMLZ AF.<UM>eat=still 1SG.TOP

tu=[nu=wawa]

UND.CMN=NMLZ.CMN=child

'I have candies, (but) I'm eating the child's (regardless).'

(Based on L. Huang 1995: 231)

The paradigmatic alternation between *ni* and *nu* is the rationale for treating both as genitive markers (or nominalizers in the present terminology) in the previous literature. However, what has not been hitherto reported is that the two markers, in Central Amis at least, can also be in a syntagmatic relationship and mark the same noun, as in (57).

(57) Central Amis (Fieldnotes)²²³

- ira ku='udax aku, k<um>aen=ho kaku
 EX TOP.CMN=candy 1SG.NMLZ AF.<UM>eat=still 1SG.TOP

tu=[nu=ni=kacaw]

UND.CMN=NMRK=NMLZ.PSN=K.

'I have candies, (but) I'm eating those of Kacaw's (regardless).'

(Based on L. Huang 1995: 231)

Compared with *ni=Kacaw* in (56)a, the phrase *nu=ni=Kacaw* in (57) conveys an emphatic or contrastive overtone (hence the slightly different free translations). Moreover, choosing one form over the other sometimes also brings about the difference between literal and metaphorical meanings, as contrasted by *u=mata ni=kacaw* (CMN=eye NMLZ.PSN=K.) 'Kacaw's eye' vs. *u=mata nu=ni=kacaw* (CMN=eye NMRK=NMLZ.PSN=K.) 'Kacaw's insight'. Although more detailed studies on the semantics of these two

²²³ This example presents a particular challenge to the "headless genitive" analysis, according to which two PUM nouns would have to be deleted or elided here, one between *tu* and *nu* and the other between *nu* and *ni*. More arguments against the "headless" analysis will be presented in §6.7.

structures have yet to be done, what is important for our present purpose is that both are legitimate structures. Crucially, the syntagmatic combination of *ni* and *nu* would illustrate what Reid (2007: 250) meant by **nu* being “followed by a genitive NP”, and adds support for his functional reconstruction of PAN **nu* as a thing-denoting marker.

In addition, recall that *n*-marked phrases in Tgdaya Seediq do not encode the Actor in NAF-constructions. By contrast, this restriction does not apply to *ni/nu*-marked phrases in Amis, as shown in (58).

(58) Central Amis (J. Wu 2006: 134, 74)

- a. palu-en **ni=mayaw** ci=dongi
 beat-PF ACT.PSN=M. TOP.PSN=D.
 ‘Mayaw will beat Dongi.’
- b. la’op-en **nu=kuyu** ku=takulil
 chase-PF ACT.CMN=leopard.cat TOP.CMN=rabbit
 ‘A leopard cat will chase the rabbit.’

However, the restriction does apply when the same noun is marked by both *ni* and *nu*, as in (59), where the addition of *nu* makes an otherwise perfect example ungrammatical.

(59) Central Amis (Fieldnotes)

fafa-en **(*nu)=ni=kacaw** ku=ra wawa
 carry.on.back-PF ??=ACT.PSN=K. TOP.CMN=MED child
 ‘Kacaw will carry that child on (his) back.’ (Based on Huang 1995: 231)

Accordingly, it can be inferred that the syntagmatic combination of *ni* and *nu*, as in (57), preserves the earlier function of *nu* as a thing-denoting marker, a stage that Tgdaya Seediq *n* is still at right now. As Amis *nu* further grammaticalized into a marker in a paradigmatic alternation with *ni*, the grammatical contrast between personal/common nouns started to emerge and the same contrast was carried over when the two markers were recruited to mark the Actor in NAF-constructions as well, a route Seediq *n* has not undertaken. One additional clue for Amis *nu* being more grammaticalized than Seediq *n*

is that *n*-marked phrases in Seediq are restricted to denote thing-like objects whereas *nu*-marked phrases in Amis do not have this semantic restriction, as contrasted in (60) and (61).

(60) Tgdaya Seediq (Fieldnotes)

- a. **n=naku** ka=sapah nii
 NMRK=1SG.NMLZ TOP=house PROX
 ‘This house is mine.’
- b.* **n=naku** ka=laqi nii
 NMRK=1SG.NMLZ TOP=child PROX

(61) Central Amis (J. Wu 2006: 87)

- Q: nima wawa ku=ni
 who.NMLZ child TOP.CMN=PROX
 ‘Whose child is this?’
- A: **nu=maku**
 NMRK=1SG.NMLZ
 ‘Mine.’

Table 6.12 summarizes the functions of PAN **nu* reflexes in Tsou, Tgdaya Seediq, and Central Amis.

Table 6.12: Functions of PAN **nu* reflexes in Tsou, Seediq, and Amis

Languages	Possessive NPs		Actor of NAF-words
	NP-use	Modification-use	
Tsou	<i>nu</i> =[<i>to</i> + POR]	---	---
Tgdaya Seediq	<i>n</i> =POR	PUM + (<i>n</i>)=POR	---
Central Amis	<i>nu</i> =POR <i>nu=ni</i> =POR	PUM + <i>nu</i> =POR PUM + <i>nu=ni</i> =POR	NAF-words + <i>nu</i> =ACT *NAF-words + <i>nu=ni</i> =ACT

Note: POR and PUM in this table stand for a full-nominal possessor and possessum NP respectively. ACT refers to a full-nominal Actor in NAF-constructions. The construction marked with an asterisk is ungrammatical.

As is clear from the organization of the table and the progression of our current discussion, the three languages seem to demonstrate three different stages of PAN *nu reflexes where they progressively grammaticalize from a thing-denoting marker, as in Tsou, to a fully grammaticalized marker indicating both the possessor and non-Topic Actor, as in Amis.

Therefore, Reid's (1981, 2007) functional reconstruction of PAN *nu as a generic thing-denoting marker not only establishes a morphological connection between possessive substantives and interrogative words in many modern languages, but also accounts for the functional developments of its possible reflexes in more Formosan languages. Crucially, if all the reasoning is on the right track, this functional reconstruction of PAN *nu would allow us to reconstruct the formula “*nu ni POR” for the NP-use in PAN, which then would make PAN a Type B language since the formula for the modification-use was incontrovertibly “*PUM ni POR”. Moreover, assuming “*nu ni POR” was the formula in PAN also makes it relatively easy to account for synchronic variations. First, Tsou would innovate new markers like *to* to replace the functional load of *ni, which has now been lost. Second, Tgdaya Seediq would simply lose *ni, thus giving rise to the current contrast between *sapah mona* ‘Mona’s house’ for the modification-use and *n=mona* ‘Mona’s (thing)’ for the NP-use. Since the two languages both lost *ni, they did not have a chance to development the grammatical contrast between personal and common nouns, unlike all the six languages that retain *ni in Table 6.8. Third, Amis would then be the only Formosan language that retains PAN “*nu ni POR”, albeit in a vestigial manner. Fourth, speaking in terms of language types, Type B languages that have lost PAN *nu would innovate new markers to compensate for its

functional load. This applies to possessive substantives based on full nominals and person forms alike. Compare, for instance, *nu='u* ‘mine’ in Tsou, where both morphemes are reconstructable in PAn (i.e. the **nu* under discussion and *=*ku* for 1SG), with *za=ku* ‘mine’ in Kavalan, where only the 1SG bound person form /*ku*/ is conservative and the NMRK /*za*/ is an innovation in form but not in function.²²⁴ Another example of functional replacement is observable by comparing *nu=maku* ‘mine’ in Amis, where both morphemes are reconstructable in PAn (i.e. the **nu* under discussion and **m-aku* for 1SG; see Ross 2015b: 119), with *rwa=maku* ‘mine’ in (Wulai) Squliq Atayal, where only the 1SG bound person form /*maku*/ is conservative and the NMRK *rwa* is an innovated form in place of **nu*. Finally, as for Type A languages, they would innovate by not requiring the NMRK for the NP-use. It seems no coincidence that Type A languages in Table 6.8 (i.e. Amis and Paiwan) are also two of the three Formosan languages that demonstrate paradigmatic alternations between reflexes of PAn **ni* and **nu* (the other one being Saisiyat, which has innovated in a different way; see §6.5).

6.7. More arguments against the “headless” analysis

The thrust of arguments against the “headless” analysis for possessive substantives, even in Type A languages, where they do seemingly result from omitting the PUM noun, is that possessive substantives are neither syntactically nor semantically dependent on a putative PUM head noun, a claim expounded by Shibatani (2009) and will be further illustrated with Formosan data in this section.

²²⁴ On a related note, the functional equivalent in Proto-Malayo-Polynesian was reconstructed as **a-nu-ku* ‘my unnamed thing, mine’ in the Comparative Dictionary of Austroneisan. See http://www.trussel2.com/acd/acd-s_n1.htm#3694.

First, it has been shown in §6.4.2.2 that possessive substantives in some languages are allowed to further modify another nominal, thus resulting in periphrastic constructions where the POR and PUM are juxtaposed or follow the morphosyntax required for modification-by-noun in those languages (see Nikolaeva and Spencer 2013). In Paiwan, for instance, claiming that the possessive substantive *ni*-phrase in (62)a is derived from eliding or deleting the PUM in (62)b is tantamount to claiming that the underived noun in (62)c is derived from eliding or deleting the second noun in the complex nominal expression in (62)d, which is clearly untenable.

(62) Northern Paiwan (Fieldnotes)

- a. *ni=camak*
NMLZ=C.
'what pertains to Camak'
- b. *ni=camak=a sunatj*
NMLZ=C.=LIG book
'Camak's book'
- c. *tjuvu*
bamboo
'bamboo'
- d. *tjuvu=a siav*
bamboo=LIG soup
'bamboo soup'

Thus, the *ni*-phrase is better analyzed as an independent nominal just like *tjuvu* 'bamboo', which can head a complete NP or modify another nominal. What makes the *ni*-phrase special is that it does not denote the nominal marked by the NMLZ *ni*, but what pertains to that nominal. Moreover, the plain nominal undergoing nominal-based nominalization can even be a complex nominal like *vavayan=a kakedrian* (female=LIG kid) 'young girl', which selects the NMLZ *nua=*, as in *nua=[vavayan=a kakedrian]=a kava* (NMLZ=female=LIG kid=LIG clothes) 'clothes of young girls'.

The same kind of argument also applies to Budai Rukai, where the possessive substantive and another nominal form a complex NP in the same way two underived nouns would, as respectively shown in (63)a and (63)b.

(63) Budai Rukai (Fieldnotes)²²⁵

- a. wa-lrangadh=aku ku **ababadh-ane** ku laimai
 RLS-buy=1SG.NOM OBL female-NMLZ OBL clothes
 ‘I bought women’s clothes.’
- b. wa-kane=aku ku **lrubu** ku lapanai
 RLS-eat=1SG.NOM OBL porridge OBL corn
 ‘I ate corn porridge.’

Again, it is not desirable to analyze the *ane*-marked nominal as deriving from eliding or deleting the PUM in (63)a any more than it is to analyze the underived noun *lrubu* ‘porridge’ as deriving from eliding or deleting the second noun in the modification-by-noun construction in (63)b.

Second, even if in many cases the “missing” PUM can reasonably be “restored” from the linguistic context, there are situations where a possessive substantive simply denotes what is contextually relevant to the base nominal (which is metonymy at work) and the putative PUM is never mentioned explicitly, let alone being elided or deleted. For instance, the two Kavalan examples in (64) are both the opening sentences of two spontaneous narratives (told by the same speaker on different days).

²²⁵ In Budai Rukai, there is a morphophonemic alternation between the postvocalic glide in {ai} (realized as [aj]; cf. *lrangai* ‘buy’ and *ababai* ‘woman’) and {dh} (realized as [ð]), which also happens in Taromake Rukai (see P. Li 1977: 381).

presumably refer to the addressee's physical condition. But no explicit nouns can serve as the putative PUM head of either *misu* '2SG.NMLZ' or *nu=misu* 'NMRK=2SG.NMLZ'.

(65) Central Amis (Fieldnotes)

- a. *ngaay=ho kisu*
good=still 2SG.TOP
- b. *ngaay=ho ku=misu*
good=still TOP.CMN=2SG.NMLZ
- c. *ngaay=ho ku=nu=misu*
good=still TOP.CMN=NMRK=2SG.NMLZ
'How are you?'

Finally, there are even more extreme cases where possessive substantives are used as taboo-avoiding expressions for sex-related referents, which makes doubtful the existence of a putative PUM once there in the syntactic representation but then elided or deleted for some reason. In Paiwan, possessive substantives marked by the NMLZ *nua* may refer to human genitals, which is also true for possessive substantives marked by the NMLZ *-ane* in Budai Rukai. The following examples contrast typical possessive NPs with possessive substantives used as substitutes for taboo expressions.

(66) Northern Paiwan (Fieldnotes)

- a. *tja-ma-san-pazangal=[a 'u=zitusia] tua=['u=cekelj]*
CMPR-AF-become-important=TOP 1SG.NMLZ=car OBL=1SG.NMLZ=spouse
'My car has become more important than my spouse.'
- b. *tja-ma-san-pazangal=[a 'u=zitusia] tua=[nua='u=cekelj]*
CMPR-AF-become-important=TOP 1SG.NMLZ=car OBL=NMLZ=1SG.NMLZ=spouse
'My car has become more important than my spouse's (private parts).'

(67) Budai Rukai (Fieldnotes)

- a. *kai senate=li, ma-ku<cia~>cingalre ki senate=su*
PROX book=1SG.NMLZ RLS-<CMPR~>precious OBL book=2SG.NMLZ
'My book is more precious than your book.'
- b. *kai senate=li, ma-ku<cia~>cingalre ki musuane*
PROX book=1SG.NMLZ RLS-<CMPR~>precious OBL 2SG.NMLZ
'My book is more precious than yours (i.e. your private parts).'

I stumbled upon these examples when trying to elicit possessive substantives, and was first suspicious of the interpretations that my Paiwan and Rukai consultants gave me, thinking the sexual connotation might be due to personal interpretations rather than part of the conventional meaning. It was only later when I found more similar examples in the literature that I became convinced of the taboo-avoiding function of possessive substantives. For instance, Tsukida (2009: 181) pointed out that the possessive substantive in Truku Seediq is formed by adding *ne-* to a base nominal (e.g. *ne-rubiq* (NMLZ-R.) ‘what pertains to Rubiq’ and *ne-laqi* (NMLZ-child) ‘what pertains to a/the child’), and expressions like *ne-senaw* (NMLZ-man) and *ne-kuyuh* (NMLZ-woman) are euphemisms for male and female genitals respectively (or secretions of human genitals). Another similar example comes from Nanwang Puyuma as illustrated in (68), which is an excerpt about vagina dentata.

(68) Nanwang Puyuma (S. Teng 2008: 273-274)

ulriya kadri i ami a salaw bulay, amuna
EX here LOC north TOP.INDF very beautiful but

mi-a-walri **na=n=taw** ka
AF.have-IPFV-teeth NMRK=LNK=3.NMLZ <AF>say

‘It was said that here in the north there was a very beautiful (woman) but hers (i.e. her vagina) was toothed.’

The possessive substantive *nantaw* ‘his/hers/theirs’, which syntactically serves as the sole argument NP of the predicate *mi-a-wali* ‘have teeth’ in the second clause, refers to the private part of a woman introduced in the previous clause.

Therefore, based on morphological, syntactic, and semantic considerations, possessive substantives are better viewed as independent nominals on their own, which are nominalized nominals created out of plain nominals. Possessive substantives have the same syntactic distributions as underived nouns, but they lack the denotational constancy

of underived nouns since they denote whatever is contextually relevant to or conventionally associated with the referent of a base nominal.

6.8. Semantic constraints on possessive substantives

The phenomenon where the possessive relationship is coded by different constructions in a language depending on various factors is referred to as split possession (see Stolz et al. 2008). The contrast between alienable and inalienable possession is a common type of semantically motivated splits. Blust (2013: 482) commented that “[i]n the AN [*Austronesian*] languages of Taiwan, the Philippines and western Indonesia possessive relationships generally are *simple and uninteresting*, but in Oceanic languages the situation is quite different, since a fundamental distinction between obligatorily or inalienably possessed nouns, and alienably possessed nouns is commonplace” (emphasis mine). While it is true that the alienability split prevalent in Oceanic languages is complex and fascinating, it remains to be seen whether possessive relationships in Formosan languages are really “simple and uninteresting.” I will show in this section that this impression is in part due to the general neglect of possessive substantives, which, when taken into consideration, do reveal some degrees of semantically motivated possession splits.

Except for Tsuchida (1995) on Tamalakaw Puyuma, S. Teng (2008: 96) on Nanwang Puyuma, and Saillard (1995: 67) on Maga Rukai, the Formosan literature rarely discusses semantically motivated splits. Among Formosan languages, however, such splits turn out to be more pervasive than has been previously identified in the literature.

Before I present the Formosan data, a brief excursion onto Guaraní would set the scene. Velázquez-Castillo (1996: 83) argues that the “equational possessive construction” in Guaraní is “primarily an expression of alienable possession” because it applies to non-relational nouns like “table”, but not to relational ones like “child”, as contrasted in (69).

(69) Guaraní (Velázquez-Castillo 1996)²²⁶

- a. ko-mesa che-mba'e
 this-table 1SG.NMLZ-NMRK
 ‘This table is mine.’
- b.* ko-memby che-mba'e
 this-child 1SG.NMLZ-NMRK

Similar splits like those in Guaraní are also found in Squliq Atayal and Tsou. Example (70) illustrates Squliq Atayal, where the NMRK *rwa* can refer to non-relational nouns like “dog” and “book,” but not relational ones like “child” and “father.”

(70) (Wulai) Squliq Atayal (Fieldnotes)

- a. {hozín/bíru} qani ga, **rwa=maku**
 {dog/book} PROX PTOP NMRK=1SG.NMLZ
 ‘This dog/book is mine.’
- b.* {laqi/yaba} qani ga, **rwa=maku**
 {child/father} PROX PTOP NMRK=1SG.NMLZ

To rectify (70)b, relational nouns would have to be expressed twice, as in (71).

(71) (Wulai) Squliq Atayal (Fieldnotes)

- {laqi/yaba} qani ga, {laqi/yaba}=maku
 {child/father} PROX PTOP {child/father}=1SG.NMLZ
 ‘This {child/father} is my {child/father}.’

Importantly, we know the contrast is between relational and non-relational nouns rather than between human and non-human ones because non-human relational body-part terms behave like human relational nouns, as shown in (72).

²²⁶ The original gloss for *che-* is 1IN, where IN stands for the “inactive” set of person forms indexing the POR, P of transitive constructions, and S of intransitive constructions with predicates of low-dynamicity (Velázquez-Castillo 1996: 13). Moreover, the original gloss for *mba'e* is “thing”, reglossed as NMRK here.

(72) Squliq Atayal (Fieldnotes)

- a. kyahin qani ga, kyahin=[na yapit]
 animal.skin PROX PTOP animal.skin=NMLZ flying.squirrel
 ‘This skin is the skin of flying squirrels.’
- b.* kyahin qani ga, **rwa=[na yapit]**
 animal.skin PROX PTOP NMRK=NMLZ flying.squirrel

Similarly, the NMRK *nu* in Tsou individuates non-relational nouns like “hat” but not relational ones like “sibling,” as contrasted in (73).²²⁷

(73) Tsou

- a. mo-yonghu=he=[ta ceopngu=su]=[’e ceopngu=’u]
 AF-pretty=CMPR=OBL hat=2SG.NMLZ=TOP hat=1SG.NMLZ
 ‘My hat is prettier than your hat.’
- b. mo-yonghu=he=[ta ceopngu=su]=[’e **nu=’u]**
 AF-pretty=CMPR=OBL hat=2SG.NMLZ=TOP NMRK=1SG.NMLZ
 ‘My hat is prettier than yours.’
- c. mo-yonghu=he=[ta ohaeva=su]=[’e ohaeva=’u]
 AF-pretty=CMPR=OBL older.sibling=2SG.NMLZ=TOP older.sibling=1SG.NMLZ
 ‘My older sibling is prettier than your older sibling.’
- d.* mo-yonghu=he=[ta ohaeva=su]=[’e **nu=’u]**
 AF-pretty=CMPR=OBL older.sibling=2SG.NMLZ=TOP NMRK=1SG.NMLZ

The restrictions against Squliq and Tsou possessive substantives being used to denote relational entities may have to do with the etymology of the NMRK in these languages, which was presumably a generic noun denoting non-human objects, much like *mba’e* in Guaraní as well as *mun* in Amami Ryukyuan (see (5) on p.14). Synchronically, however, the NMRK *rwa* in Squliq and *nu* in Tsou are highly grammaticalized morphemes and do not have the same degree of syntactic freedom as regular nouns. For instance, regular nouns like *ceopngu* ‘hat’ in Tsou can serve as the Topic NP without collocating with nominalized person indexes, but the NMRK *nu* cannot.

²²⁷ The morpheme /he/ is glossed as CMPR, following G. Lin (2010). It might be historically related to the bound person form for 3PL.INVIS, which functions as an “inclusory pronominal” indexing a third-person *singular* Topic NP in construction with an included NP (see Lichtenberk 2000). Similar use of inclusory pronominals in comparative constructions is also found in Plngawan Atayal. See (77) below for relevant examples.

On the other hand, not all the languages investigated here impose on possessive substantives such semantic restrictions as those found in Squliq and Tsou. A case in point is Kavalan. For instance, (74) is an excerpt from a narrative of the Frog story, where the speaker used *za=na* ‘NMRK=3.NMLZ’ to refer anaphorically to *lazat=na* ‘person=3.NMLZ’, meaning “the dog’s master.”

(74) Kavalan (NTU Corpus|Frog.Buya: IU.27-29)

a. ... (2.0) *supaR qaya wasu 'nay*
 AF.know also dog MED

b. *p<m>upuk tu..=lazat=na, /*
 <AF>comfort UND.CMN=person=3.NMLZ

c. *tu= tu=za=na nani. *
 UND.CMN UND.CMN=NMRK=3.NMLZ DM

‘The dog also knew (that its master was mad), (so it) comforted its master.’

Tsuchida (1993a) translated Kavalan NMRK /za/ with Japanese *mono* ‘thing’, but (74) clearly shows Kavalan possessive substantives can denote human entities, unlike their counterparts in Squliq and Tsou.²²⁸

In addition to Maga Rukai, which has been shown by Saillard (1995: 67) to demonstrate alienability splits, both of the two Rukai varieties included in this study behave likewise. In Taromake Rukai, certain kinship terms can only be modified by direct indexing whereas other nouns can be additionally modified by possessive substantives, as contrasted in (75).

²²⁸ F. Hsieh (2011: 507) showed another similar example, where the possessive substantive *za=ku* ‘NMRK=1SG.NMLZ’ modifies the relational noun *sunis* ‘child’, illustrating again that Kavalan possessive substantives do not have the semantic constraints found in Squliq and Tsou.

- (75) Taromake Rukai (P. Li 1973: 79)
- a. {taw'ong/dan/maca/talragi}=li
 {dog/house/eye/friend}=1SG.NMLZ
 'my {dog/house/eyes/friend}'
- b. **ya=li** ka {taw'ong/dan/maca/talragi}
 NMRK=1SG.NMLZ LIG {dog/house/eye/friend}
 'my {dog/house/eyes/friend}'
- c. {tama/tina}=li
 {father/mother}=1SG.NMLZ
 'my {father/mother}'
- d.* **ya=li** ka {tama/tina}
 NMRK=1SG.NMLZ LIG {father/mother}

By the same token, in Budai Rukai possessive substantives marked by *-ane* denote only alienably possessed entities, but never inalienably possessed ones such as “sibling” or “tooth”, which would have to be modified by the *ki*-phrase, as illustrated in (76).

- (76) Budai Rukai (H. Jiang 2013)²²⁹
- a. yakai ku paisu ki camake
 EX.RLS NOM money NMLZ C.
 'Does Camak have money?'
- b. yakai ku **camak-ane** ku paisu
 EX.RLS NOM C.-NMLZ NOM money
 'Is there Camake's share of money?'
- c. yakai ku {taka/valrisi} ki camake
 EX.RLS NOM {older.sibling/tooth} NMLZ C.
 'Does Camak have (any) {older siblings/teeth}?'
- d.* yakai ku **camak-ane** ku {taka/valrisi}
 EX.RLS NOM C.-NMLZ NOM {older.sibling/tooth}

Compared with the POR introduced by the *ki*-phrase, the same POR marked by *-ane* is semantically more emphatic or contrastive. In addition, unlike the former, the latter syntactically stands on an equal footing with the PUM that it modifies, as indicated by the fact that both are marked by the NOM *ku* in (76)b.

²²⁹These examples can be statements or polar questions, which are distinguished by prosody alone.

Like Budai Rukai, Pngawan Atayal uses reflexes of PAn *-an to derive possessive substantives from plain nominals. Example (77) shows that the nominalized person forms ending with *-an* denote non-relational nouns like “house” (see Table 7.6 for all the SAP forms), but not relational ones like “father” or “friend”.

(77) Pngawan Atayal²³⁰

- a. umabas babawi morong=mu laha=[cu morong=su]
 excede high house=1SG.NMLZ 3PL=OBL house=2SG.NMLZ
 ‘My house is higher than your house.’
- b. umabas babawi morong=mu laha=[cu **sinang**]
 excede high house=1SG.NMLZ 3PL=OBL 2SG.NMLZ
 ‘My house is higher than yours.’
- c. umabas takokuw {yaba/rangi}=mu laha=[cu
 excede tall {father/friend}=1SG.NMLZ 3PL=OBL
 {yaba/rangi}=su]
 {father/friend}=2SG.NMLZ
 ‘My {father/friend} is taller than your {father/friend}.’
- d.* umabas takokuw {yaba/rangi}=mu laha=[cu **sinang**]
 excede tall {father/friend}=1SG.NMLZ 3PL=OBL 2SG.NMLZ

Moreover, possessive substantives in Pngawan are generally not used to denote a body part, as expected for forms incompatible with relational nouns, unless the body part is external to the POR. The contrast is illustrated in (78).

(78) Pngawan Atayal

- a. tunux=[ni watan]=[ka hani]
 head=NMLZ W.=TOP PROX
 ‘This is Watan’s head.’ [The head is Watan’s inalienable body part.]
- b. **watan-an**=[ka tunux=[na barok hani]]
 W.-NMLZ=TOP head=NMLZ pig PROX
 ‘This pig’s head is Watan’s (share).’ [The head is Watan’s share of a hunted animal.]

²³⁰ The two Pngawan consultants I worked with showed different preferences over what markers they used to introduce the standard of comparison in their spontaneous examples (prompted by Mandarin translations). For the speaker of (77) (Yumin Nawi, born in 1948), it is *laha=cu* ‘3PL=OBL’. However, for the other speaker (Kumuy Nawi, born in 1939), it is either *laha=ci* or *laha=cika* instead. It seems that *cu* is an alternative form of the UND/OBL marker *ci*, though distributions of *cu* are much less clear. In addition to the non-SAP form *laha*, SAP person forms also function as inclusory pronominals (see Lichtenberk 2000), as in *cami=[ci watan]* (1EXCL=OBL W.) ‘I and Watan.’

Finally, Puyuma is perhaps the language where the split phenomenon is the most complicated and intriguing of all Formosan languages. Tsuchida (1995) first observed the phenomenon in Tamalakaw Puyuma, to which Rikavung Puyuma is most closely related, both genetically and geographically. Then S. Teng (2011, 2015) did a comprehensive study of person forms, including those for the POR, in three Puyuma dialects (i.e. Nanwang, Katripul, and Tamalakaw). Dialectal differences are rather conspicuous in this regard, so I will only focus on Rikavung Puyuma in the following discussion.

Nominalized person indexes in Rikavung Puyuma attach either right before the PUM (e.g. *ku=valray* [1SG.NMLZ=book] ‘my book’) or combine with the nominal relation marker that precedes the PUM (e.g. *na=n=ku valray* [DEF.TOP=LNK=1SG.NMLZ book] ‘my book’). In either case, nominalized person indexes precede the PUM noun in the linear order, and thus can be called prenominal. Possessive substantives like *nanku* constitute one of the many paradigms of free person forms in Rikavung, all consisting of nominal relation markers (see Table 2.2 for all the forms) and person-form indexes. Depending on the nominal relation markers involved, the result forms are either plain nominals (denoting persons) or nominalized nominals (denoting entities associated with persons), as respectively illustrated by Set I-II and Set III-V in Table 6.13 below.

The functional difference between Set I-II for plain person forms and among Set III-V for nominalized ones lies in the grammatical relations they assume. Set I and III both serve as the Topic, Set II and IV both as the definite non-Topic Undergoer/Oblique, and finally Set V as the indefinite counterpart of Set IV. Nominalized nominals in Set III-V all have the potential to fulfill a complete NP or modify another nominal, unlike plain

Table 6.13: Plain vs. nominalized person forms in Rikavung Puyuma²³¹

	1SG	1EXCL	1INCL	2SG	2PL
I	<i>i=n=ku</i>	<i>i=n=niam</i>	<i>i=n=ta</i>	<i>i=n=nu</i>	<i>i=n=mu</i>
II	<i>ka=ni=n=ku</i>	<i>ka=n=niam</i>	<i>ka=n=ta</i>	<i>ka=n=nu</i>	<i>ka=n(e)=mu</i>
III	<i>na=n=ku</i>	<i>na=n=niam</i>	<i>na=n=ta</i>	<i>na=n=nu</i>	<i>na=n(e)=mu</i>
	<i>ni=n=ku</i>	<i>ni=n=niam</i>	<i>ni=n=ta</i>	<i>ni=n=nu</i>	<i>ni=n=mu</i>
IV	<i>ka=na=n=ku</i>	<i>ka=na=n=niam</i>	<i>ka=na=n=ta</i>	<i>ka=na=n=nu</i>	<i>ka=na=n(e)=mu</i>
V	<i>za=n=ku</i>	<i>za=n=niam</i>	<i>za=n=ta</i>	<i>za=n=nu</i>	<i>za=n=mu</i>

nominals in Set I-II, which do not modify another nominal, as contrasted by the three forms for 1SG in (79) below.

²³¹ In fast speech, two consecutive alveolar nasals are reduced to one. For instance, *nannu* [nan.nu] for 2SG becomes *nanu* [na.nu]. Second, the 1SG form in Set II (i.e. *kaninku*) is an outlier since all the other forms in the same paradigm take the shape of *ka=n=PRO*, where PRO stands for bound person forms. The expected form *kanku* is found in other Puyuma varieties, but not in Rikavung. Finally, S. Teng (2015: 415) reports forms in the shape of both *za=PRO* and *za=na=n=PRO* for Set V in Katripul Puyuma. My Rikavung consultants recognized the *za=PRO* series, but not the *za=na=n=PRO* series. However, as far as Rikavung data are concerned, the *za=PRO* series is better analyzed as bound forms that require a host, unlike the *za=n=PRO* series in the table, which is syntactically independent. Thus, the *za=PRO* series is not listed in Table 6.13. The syntactic contrast between the *za=PRO* and *za=n=PRO* series is illustrated by relevant forms for 1SG in the following examples:

- (i) *'azi=ku* *m-na'u* **zanku**
 NEG=1SG.TOP AF-see 1SG.NMLZ.INDF.UND
 'I didn't see any one of mine.'
- (ii) *'azi=ku* *m-na'u* **zanku** *suan*
 NEG=1SG.TOP AF-see 1SG.NMLZ.INDF.UND dog
 'I didn't see any dog of mine.'
- (iii)* *'azi=ku* *m-na'u* **za=ku**
 NEG=1SG.TOP AF-see UNDF.INDF=1SG.NMLZ
- (iv) *'azi=ku* *m-na'u* **za=ku=suan**
 NEG=1SG.TOP AF-see UNDF.INDF=1SG.NMLZ=dog
 'I didn't see any dog of mine.'

(79) Rikavung Puyuma (Fieldnotes)

- a. mara-vulay nannu alrak **kaninku** (*alrak)
 AF.CMPR-beautiful 2SG.NMLZ.TOP child 1SG.OBL child
 ‘Your child is more beautiful than me.’
- b. mara-vulay nannu alrak **kananku** (alrak)
 AF.CMPR-beautiful 2SG.NMLZ.TOP child 1SG.NMLZ.DEF.OBL child
 ‘Your child is more beautiful than {my child/mine}.’
- c. mara-vulay nannu alrak **zanku** (alrak)
 AF.CMPR-beautiful 2SG.NMLZ.TOP child 1SG.NMLZ.INDF.OBL child
 ‘Your child is more beautiful than {a child/one} of mine.’

In addition to prenominal nominalized indexes, which attach directly to the PUM or to nominal relation markers, there are postnominal ones. Crucially, prenominal and postnominal nominalized indexes are not entirely identical in form and they select different semantic types of the PUM. The two sets of nominalized indexes, together with Topic indexes, are listed in Table 6.14.²³²

Table 6.14: Topic vs. nominalized person forms in Rikavung Puyuma

	1SG	1EXCL	1INCL	2SG	2PL
Topic indexes	= <i>ku</i>	= <i>mi</i>	= <i>ta</i>	= <i>u</i>	= <i>mu</i>
Prenominal nominalized indexes	<i>ku</i> = = <i>ku</i>	<i>niam</i> = = <i>niam</i>	<i>ta</i> = = <i>ta</i>	<i>nu</i> = = <i>nu</i>	<i>mu</i> = = <i>mu</i>
Postnominal nominalized indexes	= <i>lri</i>	---	= <i>ta</i>	= <i>u</i>	---

²³² The data are essentially identical to those in Tamalakaw Puyuma as reported by Tsuchida (1980, 1992a), except that he also listed the postnominal nominalized indexes for 1EXCL and 2PL (= *mi* and = *mu* respectively), neither of which is recognizable to my Rikavung consultants. Despite the fact that postnominal nominalized indexes are much less mobile than Topic indexes, which are no doubt Wackernagel enclitics, I assume that the former are also enclitics like those for the latter due to the overall formal identity between the two sets (except for 1SG) and the similar prosodic pattern they illustrate when attaching to a host. The alternations between /*mi*/ and /*niam*/ for 1EXCL and that between /*u*/ and /*nu*/ for 2SG are not due to syntactic functions or thematic roles, but to positional allomorphy instead, as has been argued by H. Jiang & Billings (2014, 2015).

Depending on their compatibility with prenominal or/and postnominal nominalized indexes, at least three classes of nouns can be distinguished. The first class constitutes the largest of the three, including most non-relational nouns. Its members favor prenominal nominalized indexes and generally reject postnominal ones, as illustrated in (80).

(80) Rikavung Puyuma (Fieldnotes)

- a. 'uze'uzem na=n=ku {talupung/suan}
 AF.black TOP.DEF=LNK=1SG.NMLZ {hat/dog}
 'My {hat/dog} is black.'
- b.* 'uze'uzem na={talupung/suan}=lri
 AF.black TOP.DEF={hat/dog}=1SG.NMLZ

By contrast, the second class is compatible with postnominal nominalized indexes but not with prenominal ones, and it consists of only a handful of consanguineal and senior kin terms (i.e. those for relatives by blood who are older than the anchor person), as shown in (81).

(81) Rikavung Puyuma (Fieldnotes)²³³

- a.* tihasar na=n=ku {mu/ama/ina/va}
 AF.tall TOP.DEF=LNK=1SG.NMLZ {grandparent/father/mother/older.sibling}
- b. ti<a>hasar-an na={mu/ama/ina/va}={lri/ni}
 <A>AF.tall-COL TOP.PL={grandparent/father/mother/older.sibling}=1SG.NMLZ
 'My {grandparents/uncles/aunts/older siblings} are tall.'
- c. tihasar i={mu/ama/ina/va}={lri/ni}
 AF.tall TOP.SG={grandparent/father/mother/older.sibling}=1SG.NMLZ
 'My {grandparent/father/mother/older sibling} is tall.'

²³³ When the PUM is *ina* 'mother', the 1SG nominalized index changes from =*lri* to =*ni* due to assimilation (Tsuchida 1995: 797). Moreover, *ama* and *ina* usually refer to one's father and mother respectively when the denotation is singular and to one's uncles and aunts respectively when plural (since people generally don't have more than one father or mother). Incidentally, the counterparts of these four kinship terms in Mantauran Rukai are precisely those that each have five sets of morphologically related forms. Among them is a set that is bound and has to be followed by nominalized person indexes (Zeitoun 2008: 4). All members except for one in this set share formal resemblance to the Puyuma equivalents (cf. *omo*- 'grandparent'; *ama*- 'father'; *ina*- 'mother'; *taka*- 'older sibling').

The last class covers relational nouns other than those in the second class and allows both prenominal and postnominal nominalized indexes, as in (82).

(82) Rikavung Puyuma (Fieldnotes)²³⁴

- a. *tihasar na=n=ku* {kuravak/wadiyan/'ali'an}
 AF.tall TOP.DEF=LNK=1SG.NMLZ {sister's.husband/younger.sibling/male.peer}
 'My {sister's husband/younger sibling/male friend} is tall.'
- b. *ti<a>hasar-an na={kuravak/wadi/'ali'}=lri*
 <A>AF.tall-COL TOP.PL={sister's.husband/younger.sibling/male.peer}=1SG.NMLZ
 'My {sisters' husbands/younger siblings/male friends} are tall.'
- c. *tihasar i={kuravak/wadi/'ali'}=lri*
 AF.tall TOP.SG={sister's.husband/younger.sibling/male.peer}=1SG.NMLZ
 'My {sister's husband/younger sibling/male friend} is tall.'

The structural contrast in (80) and (81) resembles the typical alienability split between a synthetic construction for inalienable possession, where nominalized indexes are directly marked on the PUM, and a periphrastic one for alienable possession, where nominalized indexes are marked on something other than the PUM. As noted by Koptjevskaja-Tamm (2001: 966), "there is sufficient evidence that alienability splits often involve an opposition between the archaic, inalienable construction and the innovative, alienable construction." However, the dual nature of those nouns in (82) complicates the clear semantic contrast that could have emerged. Nevertheless, considering grammatical differences among kin terms are pervasive across languages and that they often "form a residual domain, which is the last one to be conquered by a new, expanding construction" (ibid.: 213), it can still be said with fair confidence that the synthetic construction with postnominal nominalized indexes may have predated its

²³⁴ The choice between prenominal and postnominal nominalized indexes seems to have effect on the form of some PUM nouns. The consultants consistently used *wadi* 'younger sibling' and *'ali* 'male peer' with postnominal indexes but *wadiyan* 'younger sibling' and *'ali'an* 'male peer' instead with prenominal ones. The latter two terms are both suffixed by *-an*, which elsewhere imports a collective/plural meaning to a base noun (see S. Teng 2008: 284), but both of them can denote singular individuals, as in (82)a.

periphrastic counterpart involving prenominal ones.²³⁵ If so, nominalized person indexes in Puyuma would have developed from being adjacent to the PUM to phrasal clitics second to the nominal relation markers (NRM) of NPs, or schematically from NRM=PUM=PRO to NRM=PRO PUM, where the combination of NRM and person-form indexes (i.e. PRO) end up forming possessive substantives like Set III-V in Table 6.13. Also, these possessive substantives might have expanded their territory starting from denoting low-animacy entities and then gradually encroached upon the domain of the older synthetic construction, so much so that only archaic expressions and certain recalcitrant kin terms are immune from such a takeover in present-day Puyuma.²³⁶

The general pattern seen in this section is that in languages where semantically motivated splits are observable possessive substantives are always reserved for non-human, non-relational, or alienably possessed entities. This is somewhat expected because possessive substantives are exocentric expressions containing reference entities (i.e. the POR) as well as some highly grammaticalized markers that provide schematic nominal information (e.g. animacy) about the target entities that the whole expressions denote (i.e. those associated with the POR). These markers make target entities more individuated and thus prone to alienable possession, where the PUM is not inherently relational to the POR. Moreover, possessive substantives that additionally modify the PUM could potentially become the precursor of a grammaticalized construction for

²³⁵ This conclusion is also congruent with speakers' intuition. For instance, my Rikavung consultants (both almost 80 years old) commented that sentences like (80)b are rarely used and sound like "old people's speech." This is consistent with the judgement of Tsuchida's (1980: 797) Tamalakaw consultant, who reported the noun *'azin* 'spouse' with postnominal nominalized indexes sounded more archaic than the same noun with prenominal ones.

²³⁶ Among Formosan languages, nominalized person indexes attaching to nominal relation markers are unique to Puyuma. Interestingly, a strikingly similar construction is reported in Talubin Bontok (spoken in the Northern Philippines), where nominalized person indexes additionally attach right before or after the PUM (see Kikusawa & Reid 2003).

alienable possession. This development is also in line with the historical source of the alienable-inalienable distinction in Oceanic languages proposed by Lichtenberk (2005: 357), where the possessive substantive (with such meanings as “my food”, “my drink”, or more generally “my thing”) in apposition with the PUM was reanalyzed into a periphrastic possessive NP, paradigmatically contrasting with direct indexing on the PUM, as illustrated in (83) by Manam (spoken on Manam Island, Papua New Guinea).

(83) Manam (Lichtenberk 2005: 341)

- a. **ʔusi-gu**
skin-1SG.NMLZ
'my skin (the skin of my body)'
- b. **ʔusi ʔana-gu**
skin ALIM.CLF-1SG.NMLZ
'my skin (for me to eat, e.g., chicken skin)'
- c. **ʔusi ne-gu**
loincloth NALIM.CLF-1SG.NMLZ
'my loincloth' (worn wrapped tightly around one's body)

If direct indexing in Taromake Rukai were only limited to relational nouns (see (75) above), the language would end up having a grammaticalized contrast between direct indexing for inalienable possession and indirect indexing for alienable possession, structurally similar to the pattern in many Oceanic languages.

6.9. Chapter summary

This chapter has started out by contrasting the existential type with the equational type of possessive constructions in Austronesian, and then focusing on the possessive predicates of the equational type in fifteen Formosan languages because the equational type is much less researched compared with the existential type. These possessive predicates have been demonstrated to be true nominals, which minimally constitute

complete NPs and function as nominal predicates or arguments, or may additionally modify another nominal within NPs. Functionally speaking, they are nominalized nominals based on a full nominal or person form, previously called possessive substantives in Ultan's (1978: 27) terminology.

While Ultan proposed a two-way typology of possessive substantives based on a sample of 75 languages, the fifteen Formosan languages investigated here show that a third type needs to be recognized, even if we only look at cognate constructions with reflexes of PAN *ni. Specifically, Type A employs /ni/ as a NMLZ in both the NP- and modification-use, Type B is similar to Type A except that a NMRK is required in the NP-use, and finally Type C makes use of /ni/ only in the modification-use and adopts a different coding strategy for the NP-use, which is in all likelihood a later innovation. Moreover, I have argued for Reid's (1981, 2007) functional reconstruction of PAN *nu as a generic thing-denoting marker, or a NMRK in the present terminology, based on its reflexes in Tsou, Seediq, and Amis. Interestingly, the three languages seem to capture three different progressive stages of a grammaticalization process whereby a NMRK is only employed in the NP-use of possessive substantives (as in Tsou), then expands to the modification-use to introduce the POR after a PUM noun (as in Seediq), and finally morphs into a full-fledged grammatical marker that indicates both the POR after a PUM noun and the non-Topic Actor after a NAF-word (as in Amis). One implication of adopting Reid's functional reconstruction is that we could hypothesize PAN to be a Type B language. Following this hypothesis, I have also outlined the transitions from Type B in PAN with both *nu and *ni to any of the three types of its descendant languages, with their reflexes often lost and replaced by innovated functional equivalents.

Finally, I have presented several arguments against the “headless” analysis for possessive substantives with so-called genitive markers but without lexical PUM nouns that denote the type of entities the whole nominal phrases designate. In short, they are not syntactic derivatives of possessive NPs with both POR and PUM nouns, but rather independent nominals on their own right, with schematic denotations specified in relation to a base nominal (i.e. the POR). Although their actual reference is often determined anaphorically, situational cues or even conventional usage may also help to establish the reference, which is essentially metonymy at work. Moreover, possessive substantives in many Formosan languages can only refer to non-relational or alienably possessed nouns, which is presumably because the NMLZ/NMRK that creates possessive substantives was historically a generic thing-denoting noun, although these markers are synchronically so grammaticalized that they barely enjoy the syntactic freedom of regular nouns. Once possessive substantives with schematic nominal information (and hence semantic constraints) are allowed to modify a PUM noun, the new modification pattern may turn out to be the forerunner of a grammaticalized construction for alienable possession, in contrast to an older modification pattern that is often more compact.

|Chapter 7

Nominal-based Nominalization II: Structural Types

This chapter presents data from fifteen Formosan languages in terms of their POR-PUM syntagms and possessive substantives (see §6.3 for their uses in two types of possessive constructions), with the latter containing the POR but denoting entities associated with it.

For POR-PUM syntagms, only an NP is presented, which can practically assume any grammatical relation. The NP status has been verified by the fact that it can be a short answer to content questions like “What are you looking for?”, as illustrated by Northern Paiwan in (1), where the expression in the answer part within the parenthesis is contextually understood and thus optional.

(1) Northern Paiwan (Fieldnotes)

Q: anema=[(a) su=ki~kim-en]
 what=TOP 2SG.ACT=IRR~search-PF
 ‘What is it that you’re looking for?’

A: (ki~kim=a’en tua)=sunatj ni=camak
 <AF>IPFV~search=1SG.TOP UND=book NMLZ=C.
 ‘(I’m looking for) Camak’s book.’

As for possessive substantives, a complete sentence is shown instead so as to demonstrate the argument-predicate organization of the equational-type construction (see §6.3). Although instances of possessive substantives presented in this chapter all function as the predicate, the same forms elsewhere can also serve various argument functions (see §6.4).

Crosslinguistically speaking, person forms tend to come in idiosyncratic paradigms that often defy a systematic morphological analysis (at least synchronically; Siewierska 2004). Moreover, a person-form POR typically demonstrates morphosyntactic behaviors quite different from a full-nominal POR. For these reasons, the two types of POR deserve separate treatment. I start with full-nominal possession (§7.1) because it is less complicated, and then move on to person-form possession (§7.2).

7.1. Full-nominal possession

Many Formosan languages have more than one marker that introduces a full-nominal POR, depending on the POR’s semantic features, such as personal vs. common, definite vs. indefinite, visible vs. invisible, or/and singular vs. plural (Reid & H. Liao 2004: 468). To simplify data presentation, singular personal names are chosen as a representative type of full-nominal POR in this section.

Three types of languages are identified based on the way a full nominal is coded as a modifier in a syntagmatic relationship with the PUM as opposed to as a possessive

substantive in the NP-use. In other words, the three types are defined constructionally and paradigmatically in each language (see also Table 6.8). Two parameters are crucial in determining the types. The first one is whether the POR-PUM syntagm and the possessive substantive have some marking in common. Languages where they do are termed *share-NMLZ*, and those where they do not are referred to as *split-NMLZ* languages (modified from Stassen 1997). The second parameter concerns share-NMLZ languages, which are further distinguished by whether there exists a special morpheme marking the NP-use, which is called the NMRK. Languages requiring the NMRK are said to be marked, as opposed to unmarked ones, which dispense with the NMRK.

7.1.1. Type A: Unmarked share-NMLZ languages

Unmarked share-NMLZ languages are called Type A. Five languages in the sample belong to this type.²³⁷ Northern Paiwan, Central Amis, and Tgdaya Seediq are respectively illustrated in (2) through (4).

(2) Northern Paiwan (Fieldnotes)

a. sunatj **ni**=camak
 book NMLZ=C.
 ‘Camak’s book’

b. aza=(a) sunatj, **ni**=camak
 this=LIG book NMLZ=C.
 ‘This book is Camak’s.’ (See also W. Huang 2012: 43 for another Paiwan variety)

²³⁷ Judging from the data in P. Li (2000: 234), Pazeh also belongs to this type. Because it already became extinct in 2010, the language is not included in this study.

(3) Central Amis (Fieldnotes)

- a. **cudad ni=kacaw**
 book NMLZ=K.
 ‘Kacaw’s book’
- b. **ni=kacaw ku=ni a cudad**
 NMLZ=K. TOP.CMN=PROX LIG book
 ‘This book is Kacaw’s.’

(4) Tgdaya Seediq (Fieldnotes)

- a. **sapah na=mona**
 house NMLZ=M.
 ‘Mona’s house’
- b. **na=mona ka=sapah nii han**
 NMLZ=M. TOP=house PROX PRT
 ‘This house is Mona’s.’

The other two languages are slightly different from the previous three in that they typically require an additional ligature in the POR-PUM syntagm, which is never used in the NP-use. Isbukun Bunun and Thao exemplify these twists, as shown in (5) and (6) respectively.²³⁸

(5) Isbukun Bunun (Fieldnotes)

- a. **is-biung=tu ahil**
 NMLZ-B.=LIG book
 ‘Biung’s book’
- b. **sain=tu ahil hai, is-biung**
 PROX=LIG book TOP NMLZ-B.
 ‘This book is Biung’s.’ (See also C. Shi 2009: 43)

²³⁸ In addition to *tu*, the ligature between the POR and PUM can also be *a* (see C. Shi 2009 for more details). Moreover, the ligature can be also be optional in some cases, as in the following example:

inak busul hai, ciahun-un=[mas tama]
 1SG.NMLZ gun PTOPTOP rectify-PF=ACT father
 ‘As for my gun, Father fixed it.’ (ODIL; from the entry *ciahun* ‘rectify’)

(6) Thao (Fieldnotes)²³⁹

- a. **ti=kilash=a** tamuhun
 PSN.NMLZ=K.=LIG hat
 ‘Kilash’s hat’
- b. **haya=wa** tamuhun, **ti=kilash**
 PROX=LIG hat PSN.NMLZ=K.
 ‘This hat is Kilash’s.’

In both languages, the preferred word order is for the POR to precede the PUM, and the attributive ligature, which does not bear word stress, phonologically attaches to its preceding host. In Thao, phonological liaison happens between the ligature and its host, thus giving rise to such allomorphs as *a*, *wa*, and *ya*.

7.1.2. Type B: Marked share-NMLZ languages

Marked share-NMLZ languages are called Type B, to which six languages belong. While the NMLZ found in both the modification- and NP-use often involves cognate forms, the NMRK in the NP-use varies considerably from one language to another.

In Kavalan and Saaroa, the NMRK seems to be recruited from nominal demonstratives, as respectively shown in (7) and (8).

²³⁹ Recall that the phrase *ti=X* in Thao may refer to X or entities associated with X, depending on the context (see (25) on p.368). In addition, S. Wang (2004: 300) pointed out that the demonstrative *haya* refers to entities that are “semi-distal” from the speaker, and thus can be translated as “this” or “that”, depending on the context. However, since *haya* is deictically contrastive to only the distal term *huya*, which refers to entities that are far from the speaker, the two-way contrastive glosses PROX and DIST would be enough for *haya* and *huya* respectively.

(7) Kavalan (Fieldnotes)²⁴⁰

- a. **lepaw=(na)** **ni=abas**
 house=3.NMLZ NMLZ=A.
 ‘Abas’s house’
- b. **za=(na)** **ni=abas** ya=lepaw ’nay
 NMRK=3.NMLZ NMLZ=A. TOP=house MED
 ‘That house is Abas’s.’ (See also A. Lee 1997: 52)

(8) Saaroa (Fieldnotes)

- a. **suhlate=(isa)** amahle
 book=3.NMLZ A.NMLZ
 ‘Amahle’s book’
- b. suhlate kani’i=na ia, **isikana=(isa)** amahle
 book PROX=DEF PTOP NMRK=3.NMLZ A.NMLZ
 ‘This book is Amahle’s.’ (See also C. Pan 2012: 262)

Kavalan has three-way deictically contrastive nominal demonstratives: *zau* for PROX, *yau* for MED, and *wiu* for DIST (H. Jiang 2009), which can be analyzed as consisting of monosyllabic demonstrative bases and the vowel /u/, the function of which is unclear. The analysis is line with Reid’s (2007: 243) reconstructions of demonstrative bases in pre-Amis (i.e. *ni, *za, and *ya), an ancestral language of Amis, which belongs to East Formosan as Kavalan does. Thus, the Kavalan NMRK /za/ would then have grammaticalized from one of the three demonstrative bases. Similarly, /kana/ as in the Saaroa NMRK *isikana* is most likely relatable to the distal demonstrative *kana’a*, in contrast to its proximal counterpart *kani’i* (see Radetzky 2004).

In both Squiliq Atayal and Tsou, the NMLZ phonologically attaches to its preceding element rather than the POR, namely, either the PUM noun or the NMRK, which is *rwa* and *nu* in the respective languages, as in (9) and (10).²⁴¹

²⁴⁰ Parentheses around person indexes indicate that indexing is optional, which only happens to non-SAP PORs. In addition, when the person index /na/ in the second example does not occur, the NMRK /za/ forms a prosodic word with the following phrase *ni=abas*.

²⁴¹ See Footnote 205 for the rationale behind transcribing the Tsou NMRK as *nu* rather than *nuu*.

(9) (Wulai) Squliq Atayal (Fieldnotes)²⁴²

- a. *biru*=[(na) *yukan*]
 book=NMLZ Y.
 ‘Yukan’s book’
- b. *biru qani ga, rwa*=[(na) *yukan*]
 book PROX TOP NMRK=NMLZ Y.
 ‘This book is Yukan’s.’

(10) Tsou (Fieldnotes)²⁴³

- a. *ceopngu*=(si)=[to *paicɤ*]
 hat=3SG.NMLZ=NMLZ P.
 ‘Paicɤ’s hat (with her not being around)’
- b. (zou) *nu*=(si)=[to *paicɤ*]=[’e *ceopngu*]
 EMPH NMRK=3SG.NMLZ=NMLZ P.=TOP hat
 ‘The hat (here) is Paicɤ’s (with her not being around).’
- c. (zou) *nu*=(hin’i)=[ta *paicɤ*]=[’e *ceopngu*]
 EMPH NMRK=3PL.NMLZ=NMLZ P.=TOP hat
 ‘The hats (here) belong to Paicɤ and her associates (with them being around).’

Possessive substantives with Squliq *rwa* and Tsou *nu* share the same semantic restrictions against denoting human entities (as well as relational non-human ones; see §6.8), so it seems reasonable to assume that the two markers historically came from a generic noun denoting non-human objects. Synchronically, however, the two markers are highly grammaticalized and thus do not have the same degree of syntactic freedom as regular nouns.

In Rikavung Puyuma, the NMRK capitalizes on its rich system of nominal relation markers, which host nominalized person indexes and function as possessive substantives (see Table 6.13 for three such paradigms), as in (11).

²⁴² See Footnote 25 for the rationale behind not transcribing word-final glottal stops in Atayal.

²⁴³ The markers *ta* and *to* introduce the POR nominal that is visible and invisible respectively (see Zeitoun 2005; H. Chang 2011).

(11) Rikavung Puyuma (Fieldnotes)

- a. **taw=**valray ni=misak
3.NMLZ=book NMLZ=M.
'Misak's book'
- b. izu na=valray mu, **na=n=taw** ni=misak
MED TOP.DEF=book PTOP NMRK=LNK=3.NMLZ NMLZ=M.
'That book is Misak's.'

The NMRK in (11) is the same nominal relation marker for definite NPs, be it the patientive Topic or the agentive non-Topic Actor in NAF-constructions, as in (12).

(12) Rikavung Puyuma (Fieldnotes)

- a. ku=trima'-ay=lra **na=**valray
1SG.ACT=buy-LF=already TOP.DEF=book
'I already bought the book.'
- b. **taw=**ku=sukun-anay **na=**alrak
INV=1SG.TOP=push-CF ACT.DEF=child
'The child pushed me.'

The last Type B language is Taromake Rukai, as shown in (13).

(13) Taromake Rukai

- a. bo'a'e **ki** tanebake
sheep NMLZ T.
'Tanebake's sheep' (9-Level Textbooks 7-7)
- b. **ya=[ki** tama=so] kaivai
NMRK=NMLZ father=2SG.NMLZ PROX
'This is your father's.' (FLDA: Legends 06-026)

The NMRK *ya*, to which the NMLZ *ki* attaches phonologically, seems to have a locative origin. In Taromake (as well as Budai) Rukai, a locative NP is introduced by the existential/locative verb *yakai* or its Kstem *ikai*, consisting of the locative *i-*, the realis marker *a-*, and the proximal demonstrative *kai* (see also Zeitoun et al. 1999: 19), as illustrated in (14).²⁴⁴

²⁴⁴ Much as the Mstem *wa-tobi* has *tobi* as its Kstem, the Mstem *yakai* has *ikai* as its Kstem, consisting of the locative *i-* and the demonstrative *kai*. See also Table 3.3 for verb classes in Budai Rukai, which is closely related to Taromake Rukai.

(14) Taromake Rukai

- a. **yakai** lrgelrege kwadra dane=li
 EX.RLS moutain there house=1SG.NMLZ
 ‘My house is on the mountains over there.’ (9-Level Textbooks 2-4)
- b. wa-tobi **ikai** dane kai lrolay
 RLS-cry EX.K house PROX child
 ‘This child cried in a house.’ (P. Li 1973: 120)

The locative *i-* is reconstructed as *i in PAN, the function of which was to mark both location and possessor (Starosta et al. 1982: 155).²⁴⁵ The location-possession isomorphism expressed by reflexes of PAN *i is also found in Takibakha Bunun, to be illustrated in §7.1.3 below.

7.1.3. Type C: Split-NMLZ languages

Split-NMLZ languages, called Type C, differ from the previous two types in that they make use of different NMLZ forms for the POR-PUM syntagm and possessive substantives. Four languages are identified as Type C. First, in Saisiyat the relator in the POR-PUM syntagm is *ni* whereas the markers for possessive substantives are *'an...a*, as illustrated in (15).

(15) Saisiyat (Fieldnotes)

- a. tatpo' **ni='iban**
 hat NMLZ=I.
 ‘Iban’s hat’
- b. hini' tatpo' (mina), **'an='iban=a**
 PROX hat EMPH NMLZ=I.=NMLZ
 ‘This hat is (indeed) Iban’s.’

²⁴⁵ Their original terms for location and possessor were Locus and Correspondent respectively, and both case relations are defined semantically and morphosyntactically in Starosta’s (1988b) Lexicase Grammar.

While possessive substantives marked by *'an...a* either make up an independent NP or additionally modify a noun, *ni*-marked phrases have to modify a noun, as contrasted in (16).

(16) Saisiyat

- a. *hiza' tatpo' ni='iban*
 DIST hat NMLZ=I.
 'That is Iban's hat.' (M. Yeh 1991: 45)
- b.* *hiza' ni='iban*
 DIST NMLZ=I.
 Intended 'That is Iban's.' (Fieldnotes)
- c. *hiza' 'an='iban=a tatpo'*
 DIST NMLZ=I.=NMLZ hat
 'That is Iban's hat.' (M. Yeh 1991: 44)
- d. *hiza' 'an='iban=a*
 DIST NMLZ=I.=NMLZ
 'That is Iban's.' (Fieldnotes)

In Takibakha Bunun, the relator *it*, which is a ditropic clitic, occurs in the POR-PUM syntagm whereas possessive substantives are formed by *i*, which hosts person-form clitics that index the POR, as in (17).

(17) Takibakha Bunun (Fieldnotes)

- a. *tamuhung=[it savi]*
 hat=NMLZ S.
 'Savi's hat'
- b. *i=cia savi=[at tamuhung di]*
 NMRK=3SG.NMLZ S.=TOP hat PROX
 'This hat is Savi's.' (See also Starosta 1988: 569)

In (17)b, *=cia* indexes the personal name *savi*, without which *=cia* could alternate with other person-form clitics in the same paradigm (see §7.2.1 for details). In other words, whether there is a juxtaposed full nominal right after the possessive substantive is the only structural difference between full-nominal possession and non-SAP person-form possession (cf. *i=cia savi* 'Savi's' vs. *i=cia* 'hers'). Like those in Saisiyat, possessive

substantives in Takibakha can additionally modify a noun, as in *i=cia savi tamuhung* ‘Savi’s hat’.²⁴⁶ Finally, like the NMRK *ya* in Taromake Rukai (see (13) and (14) above), the NMRK *i* in Takibakha Bunun seems to have a locative origin. In addition to associated entities, *i* also marks location, as in (18).

(18) Takibakha Bunun (9-Level Textbooks: 2-6)

i-lumaq=cak **cin** **masituqas** **papaqainan**
 LOC-house=1SG.TOP CONJ older.sibling have.fun
 ‘(My) older {sister/brother} and I had fun at home.’

The last two Type C languages are Pngawan Atayal and Budai Rukai, both of which make use of reflexes of PAN *-an for possessive substantives. In Pngawan Atayal, the relator *ni* is found in the POR-PUM syntagm whereas the suffix *-an* in possessive substantives, as shown in (19).

(19) Pngawan Atayal (Fieldnotes)

- a. **lukus=[ni iwal]**
 clothes=NMLZ I.
 ‘Iwal’s clothes’
- b. **iwal-an=[ka lukus hani]**
 I.-NMLZ=TOP clothes PROX
 ‘This (piece of) clothing is Iwal’s.’

Aikhenvald (2013) distinguishes possession-related morphemes marked on the POR from those marked on the PUM, and calls the former *genitive* and the latter *pertensive*. She illustrates the pertensive marker with the following example from the Lolovoli dialect of North-East Ambae spoken on Vanuatu.

²⁴⁶ In Takibakha Bunun, the attributive ligature *tu* may optionally occur between the possessive substantive and a noun, as in *i=cia savi=[tu tamuhung]* ‘Savi’s hat’. It is unclear what conditions the presence of the ligature. Y. Jiang (2012: 72) mentioned three contexts where the ligature is used, but none of them covers such a situation.

(20) North-East Ambae (Hyslop 2011: 167; cited in Aikhenvald 2013: 7)

gamali-**ni** Robert
 club.house-PRTN R.
 ‘Robert’s club house’

Despite the structural similarity between (19)a from Plngawan Atayal and (20) from North-East Ambae, *ni* in Plngawan is better not analyzed as a pertensive marker because its phonological affiliation with the PUM is only a contingent result of the more general distribution whereby *ni*, like other nominal relation markers in the language, attaches to whatever happens to come before it. In other words, the marker *ni* is a ditropic clitic. This is made more obvious when the same morpheme *ni* marks the non-Topic Actor, as shown in (21), where the *ni*-phrase and its preceding noun *tokey* ‘watch’ hold a relationship of adjacency but not of constituency.

(21) Plngawan Atayal (9-Level Textbooks 7-5)

si-biniy=**cu** tutux ga’arus tokey=**[ni** yaba=**mu]**
 CF-give=1SG.TOP one new watch=ACT father=1SG.NMLZ
 ‘My father gave me a new watch.’

Similarly, Budai Rukai uses the prenominal marker *ki* in the POR-PUM syntagm but the suffix *-ane* for possessive substantives, as in (22).

(22) Budai Rukai (Fieldnotes)²⁴⁷

- a. senate=**(ini)** **ki** cegau
 book=3SG.NMLZ NMLZ C.
 ‘Cegau’s book’
- b. cegav-**ane** kikai senate
 Cegau-NMLZ PROX book
 ‘This book is Cegau’s.’ (See also C. Chen 2008: 50)

Finally, both *-ane* in Budai Rukai and *-an* in Plngawan Atayal are multifunctional morphemes that do more than just to mark a full nominal as the possessive substantive.

²⁴⁷ In Budai Rukai, there is a morphophonemic alternation between the postvocalic glide in /au/ and /v/, which also happens in Taromake Rukai (see Li 1977: 381).

Suffice it to say that they are component morphemes of a paradigm of free person forms in both languages (see Table 7.6 and Table 7.7), and that phrases marked by them, be it person forms or full nominals, assume several locative roles (e.g. goal of motion) in addition to denoting associated entities, thus demonstrating the location-possession isomorphism like reflexes of PAn *i in Taromake Rukai and Takibakha Bunun.

7.2. Person-form possession

Compared with full nominals, person forms typically have more coding potentials as the POR (Handschuh 2014: 170), which is also the case in Formosan languages. For an expository purpose, 1SG is taken as the representative of all SAP person forms in the same paradigm because 1SG is where exceptional forms most frequently occur when Topic and nominalized person indexes are nearly identical (see §6.2), which is probably due to the high frequency of 1SG. Moreover, non-SAP person forms are not taken into consideration for the present purpose because they are treated as full nominals in some languages but as SAP person forms in others (see also Bhat 2004).

In terms of person-form possession, languages are classified into three types, called Type I, II, and III, to be distinguished from Type A, B, and C in full-nominal possession. If a language uses the same paradigm of person forms for the POR-PUM syntagm and for possessive substantives, it is referred to as a *share-person language*, and as a *split-person language* if otherwise (modified from Stassen 1997). Like share-NMLZ languages, share-person languages may be marked or unmarked, depending on whether the NMRK is required to form possessive substantives. Despite the structural correspondences between the three types in full-nominal possession (Type A through C)

and those in person-form possession (Type I through III), languages do not always show matching coding patterns for both types of possession. For instance, a language may belong to Type A in terms of full-nominal possession but to Type III with regards to person-form possession (e.g. Paiwan). This is precisely why full-nominal and person-form possession are treated separately.

7.2.1. Type I: Unmarked share-person languages

Unmarked share-person languages are called Type I, to which two languages belong. The first one is Thao, as in (23).

(23) Thao (Fieldnotes)

- a. **nak=a** tamuhun
 1SG.NMLZ=LIG hat
 ‘my hat’
- b. **nak** izay=a tamuhun
 1SG.NMLZ PROX=LIG hat
 ‘This hat is mine.’ (See also Blust 2003: 445)

Unlike those in most other Formosan languages, Topic and nominalized person forms in Thao are phonologically and morphosyntactically free. Among SAPs, the two functions are expressed by two sets of distinct forms except for the syncretic 2PL, as shown in Table 7.1 (cf. Blust 2003: 207; S. Wang 2004: 188; P. Li 2011: 7).²⁴⁸

²⁴⁸ Blust (2003: 207) listed *yamin* as both the Topic and nominalized person form for 1EXCL, but S. Wang (2004: 188) explicitly pointed out that the nominalized person form for 1EXCL should be *nam*, rather than *yamin*. Both later studies (e.g. P. Li 2011: 7) and my two Thao consultants have confirmed that S. Wang’s description was correct. Thus, the 2PL form *manium* is the only syncretic one among SAPs.

Table 7.1: Topic vs. nominalized person forms in Thao

	1SG	1EXCL	1INCL	2SG	2PL
Topic	<i>yaku</i>	<i>yamin</i>	<i>ita</i>	<i>ihu</i>	
Nominalized (for NP- and modification-use)	<i>nak</i>	<i>nam</i>	<i>mita</i>	<i>mihu</i>	<i>maniun</i>

The other unmarked share-person language is the Kaohsiung variety of Isbukun Bunun, as in (24).

(24) Kaohsiung Isbukun Bunun (Fieldnotes)²⁴⁹

- a. **i-nák=**tu ahil
 NMLZ-1SG=LIG book
 ‘my book’
- b. **sain=**tu ahil hai, **i-nák**
 this=LIG book TOP NMLZ-1SG
 ‘This book is mine.’ (See also L. Huang 1997: 375)

All person-form possessive substantives in this variety are composed of the morpheme *i* and person-form roots, which are formally distinct from Topic indexes among SAPs, as shown in Table 7.2 (cf. R. He et al. 1986: 52; L. Huang 1997: 364; L. Huang et al. 1999: 169).

²⁴⁹ The acute accent indicates the stress, which is only specified when it does not fall on the penultimate syllable.

Table 7.2: Topic vs. nominalized person forms in Kaohsiung Isbukun Bunun

	1SG	1EXCL	1INCL	2SG	2PL
Topic indexes	= <i>ik</i>	= <i>im</i>	= <i>ta</i>	= <i>as</i>	= <i>am</i>
Nominalized nominals (for NP- and modification-use)	<i>i-nák</i>	<i>i-nám</i>	<i>i-mita</i>	<i>i-sú</i>	<i>i-mú</i>

P. Li (1997a: 317, 1997b: 364) reported that in the Nantou variety of Isbukun Bunun there is a bound set of nominalized person indexes (e.g. *lumah=su* [house=2SG.NMLZ] ‘your house’), in addition to free forms like *isu* ‘2SG.NMLZ’. However, the bound set was not mentioned in Zeitoun’s (2000b: 72) reference grammar of the same variety. In addition, neither previous studies (R. He et al. 1986: 52; L. Huang 1997: 364; L. Huang et al. 1999: 169) nor my consultants recognize such a bound set in the Kaohsiung variety. Finally, the bound set is not found in Ismahasan et al.’s (1998) texts (digitalized in FLDA) of Kaohsiung Isbukun Bunun. Due to lack of such a bound set in the Kaohsiung variety, free forms such as *inák* are analyzed as person-form roots marked by the NMLZ *i-*, which elsewhere marks locations, as in (25).²⁵⁰

(25) Isbukun Bunun (ODIL; under the entry *i-*)

adu=aiza bunun **i-lumah=tan**
 QP=EX person LOC-house=MED.NTOP
 ‘Is there anybody in that house?’

The possession-location syncretism is attested in Isbukun and Takibakha Bunun, and marked by the morpheme /i/ (< PAn *i) in both dialects. However, relevant forms in Takibakha Bunun call for an analysis different from their counterparts in Kaohsiung

²⁵⁰ See L. Li (2010: 9) and H. Jiang (2012) for the clitic analysis on bound demonstratives in Isbukun Bunun.

Isbukun, to be introduced in the next section (see also Table 6.7 for a comparison of the two systems).

As has been reasoned out in §6.4.2.2, it seems that Kaohsiung Isbukun Bunun has lost the direct indexing strategy whereby person-form indexes attach to the PUM noun, a strategy that is still vigorous in conservative Bunun dialects like Takibakha as well as in the majority of Formosan languages. Such a loss was perhaps facilitated by the possessive substantives taking over the modification-use. A similar development can also be outlined for Thao. Interestingly, both languages still have bound person forms that index the non-Topic Actor in NAF-constructions, as in (26) and (27).

(26) Isbukun Bunun (ODIL; under the entry *pit'unu*)²⁵¹

pit'unu-un=**ku**=[a lulubunun]
 steam-PF=1SG.ACT=TOP egg
 'I steamed the eggs.'

(27) Thao (S. Wang 2004: 189)

azazak in-apa=**ku**
 child NAF.PFV-carry.on.the.back=1SG.ACT
 'I carried the child.'

Since the same set of bound person forms indexes both the POR and non-Topic Actor in almost all Formosan languages (see §4.1), these non-Topic Actor indexes can be considered historical relics for the direct indexing strategy.

7.2.2. Type II: Marked share-person languages

Marked share-person languages are dubbed Type II, which covers all the six Type B marked share-NMLZ languages (see §7.1.2) and Takibakha Bunun. Person-form

²⁵¹ In the writing system of Isbukun, the hyphen is used as a symbol of syllable boundaries. Since I am using the hyphen as a symbol of morphemic boundaries throughout this work, syllable boundaries in Isbukun are indicated by the apostrophe instead.

possession in the six Type II & B languages is illustrated in (28) through (33), where possessive substantives are built from nominalized person indexes attached to the NMRK.

(28) Taromake Rukai

- a. **taw'ong=li**
 dog=1SG.NMLZ
 'my dog' (P. Li 1973: 79)
- b. **ya=li** kaivai taw'ong
 NMRK=1SG.NMLZ PROX dog
 'This dog is mine.' (Fieldnotes)

(29) Kavalan (Fieldnotes)

- a. **wasu=ku**
 dog=1SG.NMLZ
 'my child'
- b. **za=ku** ya=[wasu zau]
 NMRK=1SG.NMLZ TOP=dog PROX
 'This dog is mine.' (See also P. Li 1978: 355)

(30) Saaroa (Fieldnotes)²⁵²

- a. **suhlati=ku**
 book=1SG.NMLZ
 'my book'
- b. **suhlate** kani'i=na ia, **isikana=ku**
 book PROX=DEF PTOPTOP NMRK=1SG.NMLZ
 'This book is mine.' (See also C. Pan 2012: 11)

(31) Tsou (Fieldnotes)

- a. **ceopngu='u**
 hat=1SG.NMLZ
 'my hat'
- b. (zou) **nu='u='e** ceopngu]
 EMPH NMRK=1SG.NMLZ=TOP hat
 'The hat (here) is mine.' (See also Zeitoun 2000a: 241)

²⁵² See Footnote 192 for the allomorphic alternation between *suhlate* and *suhlati*.

(32) Rikavung Puyuma (Fieldnotes)

- a. **ku=valray**
 1SG.NMLZ=book
 ‘my book’
- b. izu na=valray mu, **na=n=ku**
 DIST TOP.DEF=book PTOP NMRK=LNK=1SG.NMLZ
 ‘That book is mine.’

(33) (Wulai) Squliq Atayal (Fieldnotes)

- a. biru={**ku/mu/maku**}
 book=1SG.NMLZ
 ‘my book’
- b. biru qani ga, **rwa={ku/mu/maku}**
 book PROX TOP NMRK=1SG.NMLZ
 ‘This book is mine.’

Table 7.3 below lists all the SAP person forms that combine with the NMRK in the six languages demonstrated above. Every bound person form in the table can attach to the NMRK to produce a free possessive substantive for the NP-use, except that there should be a homorganic nasal between the NMRK and a person form in Puyuma (see Footnote 93).

Among the six languages demonstrated above, Rikavung Puyuma and Squliq Atayal attest what Nichols & Bickel (2013) call floating marking of a possessive NP since nominalized person indexes in both languages may “float,” as it were, to potential hosts other than the PUM noun. In Rikavung, nominalized person indexes either attach right before the PUM, as in (34)a, or after a nominal phrase marker like *na*, which combines with the indexes to form possessive substantives (see Table 6.13 above), as in (34)b.

Table 7.3: Nominalized person forms in six Type II languages

	NMRK	1SG	1EXCL	1INCL	2SG	2PL
Taromake Rukai	<i>ya</i>	<i>=li</i>	<i>=nai</i>	<i>=ta</i>	<i>=so</i>	<i>=nomi</i>
Kavalan	<i>za</i>	<i>=ku</i>	<i>=niq</i>	<i>=ta</i>	<i>=su</i>	<i>=numi</i>
Saaroa	<i>isikana</i>	<i>=ku</i>	<i>=hlamu</i>	<i>=ta</i>	<i>=u</i>	<i>=mu</i>
Tsou	<i>nu</i>	<i>=’u</i>	<i>=mia</i>	<i>=to</i>	<i>=su</i>	<i>=mu</i>
Rikavung Puyuma	<i>na</i>	<i>=ku</i>	<i>=niam</i>	<i>=ta</i>	<i>=nu</i>	<i>=(e)mu</i>
Squliq Atayal	<i>rwa</i>	<i>=ku/mu/maku</i>	<i>=miyan</i>	<i>=ta</i>	<i>=su</i>	<i>=mamu</i>

(34) Rikavung Puyuma (Fieldnotes)

- a. *sazu* **ku=anger** *kani=misak*
 plentiful 1SG.NMLZ=heart OBL=M.
 ‘I feel sad for Misak.’ (Lit. ‘My heart is plentiful towards Misak.’)
- b. *sazu* **na=n=ku** *anger kani=misak*
 plentiful TOP.DEF=LNK=1SG.NMLZ heart OBL=M.
 ‘I feel sad for Misak.’ (Lit. ‘My heart is plentiful towards Misak.’)

In a similar fashion, nominalized person indexes in Squliq attach to the PUM in affirmative sentences, but to the negator instead in negative ones, as contrasted in (35).

(35) (Wulai) Squliq Atayal (Fieldnotes)

- a. *qani ga, biru={ku/mu/maku}*
 PROX TOP book=1SG.NMLZ
 ‘This is my book.’
- b. *qani ga, iyat={ku/mu/maku} biru*
 PROX TOP NEG=1SG.NMLZ book
 ‘This is not my book.’

The last Type II language is Takibakha Bunun, which is a Type C split-NMLZ language in terms of full-nominal possession. Unlike Kaohsiung Isbukun Bunun,

Takibakha Bunun has a complete paradigm of nominalized person indexes in construction with the PUM, and these very same indexes attach to the NMRK *i* to serve as possessive substantives, as in (36).

(36) Takibakha Bunun (Fieldnotes)

- a. **tamuhung=nak**
 hat=1SG.NMLZ
 ‘my hat’
- b. **i=nak=[at tamuhung di]**
 NMRK=1SG.NMLZ=TOP hat PROX
 ‘This hat is mine.’ (cf. Y. Jiang 2012: 63)

Table 7.4 (cf. I. Chen 2009: 16; Takivatan 2011: 26; Y. Jiang 2012: 59) compares Topic person forms with nominalized ones in Takibakha, where Topic indexes are not regularly predictable from their free counterparts whereas nominalized indexes are consistently possessive substantives minus the NMRK *i*.²⁵³

²⁵³ The three studies mentioned here are rather consistent with respect to free person forms for the Topic, but differ a great deal in terms of bound ones. While I. Chen (2009: 16) listed only one Topic index (for 1SG), Y. Jiang (2012: 59) identified two (for 1SG and 1EXCL) and Takivatan (2011: 26) presented three (for 1SG, 1EXCL, and 1INCL). In this table, I add the Topic index for 2SG, as illustrated in the following question-answer pair:

Takibakha Bunun (9-Level Textbook: 8-1)

- (i). *qanciap=as* *malas-bunun*
 can=2SG.TOP AF.speak-Bunun
 ‘Can you speak Bunun?’
- (ii). *qanciap=cak* *tukuc*
 can=1SG.TOP a.little
 ‘I can (speak) a little.’

Moreover, all the three studies gave *imita* (with a bilabial nasal) for 1INCL, but one of my consultants (Tiang Maitangan, born in 1949) consistently produced *inita* (with an alveolar nasal) instead. Hence, the latter alternative form is also included in the table. On a related note, its counterpart in Takituduh, which is a Northern Bunun dialect like Takibakha, is *inita* (with an alveolar nasal) after consonants, but *kinita* after vowels (see Tsuchida 1992: 732; Y. Su 2008: 5).

Table 7.4: Topic vs. nominalized person forms in Takibakha Bunun

Roles	Bound/Free	1SG	1EXCL	1INCL	2SG	2PL
Topic	Bound	= <i>cak</i>	= <i>cam</i>	= <i>cata</i>	= <i>as</i>	---
	Free	<i>azak</i>	<i>azam</i>	<i>ata</i>	<i>asu</i>	<i>amu</i>
Nominalized	Bound	= <i>nak</i>	= <i>nam</i>	= <i>mita</i> = <i>nita</i>	= <i>su</i>	= <i>mu</i>
	Free	<i>i=nak</i>	<i>i=nam</i>	<i>i=mita</i> <i>i=nita</i>	<i>i=su</i>	<i>i=mu</i>

Like Topic indexes, nominalized indexes in Takibakha are Wackernagel morphosyntactic phrasal clitics because their positions vary depending on the availability of potential hosts. In the case of Topic indexes, they attach to either a lexical verb or to the negator that precedes the verb, as in (37).

(37) Takibakha Bunun (Y. Jiang 2012: 78)

- a. **qu=cak** danum
 AF.drink=1SG.TOP water
 ‘I drink water.’
- b. **ni=cak** qu danum
 NEG=1SG.TOP AF.drink water
 ‘I don’t drink water.’

Similarly, in addition to attaching to the PUM noun, as in (38)a, nominalized indexes may attach to a sentence-initial predicate, with which they do not even form a syntactic constituent, as in (38)b-c.

(38) Takibakha Bunun

- a. **sedu asu hung=nak**
 AF.PFV.see 2SG.TOP book=1SG.NMLZ
 ‘{Did you see my book?/You already saw my book.}’ (Y. Jiang 2012: 119)
- b. **macakbit=nak=[a bungu]**
 AF.ache=1SG.NMLZ=TOP head
 ‘I have a headache.’ (Lit. ‘My head aches.’) (Y. Jiang 2012: 54)
- c. **mastan=nak=[at tama] ma-daqvas malka su’u**
 AF.surpass=1SG.NMLZ=TOP father AF-tall AF.act.upon 2SG.UND
 ‘My father is taller than you.’ (Fieldnotes)

This is so despite the fact that expressions like *bungu=nak* ‘my head’ and *tama=nak* ‘my father’ are legitimate phrases elsewhere. Although it is not clear at this point what else can be acceptable hosts for nominalized indexes, Takibakha Bunun clearly has floating marking for possessive NPs, just as do Rikavung Puyuma and Squaliq Atayal.

7.2.3. Type III: Split-person languages

Split-person languages, called Type III, feature two sets of nominalized person forms, one for the modification-use only and the other for the NP-use in all languages and additionally for the modification-use in some. Nominalized person forms for the NP-use are always phonologically no shorter than those exclusively for the modification-use. Type III includes six languages and comes in two varieties, depending on whether person-form possessive substantives contain the same NMLZ as full-nominal ones do (specifically personal names illustrated in §7.1).

In the first variety, which includes four languages, person-form possessive substantives are made up of person-form roots plus the same NMLZ as found in full-nominal ones. Northern Paiwan (in fact all Paiwan varieties) is such a language, as shown in (39).

(39) Northern Paiwan (Fieldnotes)

- a. **'u=sunatj**
 1SG.NMLZ=book
 'my book'
- b. **aza=(a) sunatj, ni=a'en**
 PROX=LIG book NMLZ=1SG
 'This book is mine.'

While nominalized indexes are proclitics in Paiwan, which is rare among Formosan, Topic indexes are enclitics, and the two types of indexes are distinct in form. Possessive substantives are then built from Topic indexes and the NMLZ *ni*, which also nominalizes personal nouns. SAP person forms for the three functional paradigms are listed in Table 7.5 (cf. A. Chang 2000: 87, 2006: 68).²⁵⁴

Table 7.5: Topic vs. nominalized person forms in Northern Paiwan²⁵⁵

	1SG	1EXCL	1INCL	2SG	2PL
Topic indexes	=a'en	=amen	=itjen	=sun	=mun
Nominalized indexes (for modification-use only)	'u=	nia=	tja=	su=	nu=
Nominalized nominals (for NP- and modification-use)	ni=a'en	ni=amen	ni=itjen	ni=sun	ni=mun

Second, Pngawan Atayal also falls into the first variety, as illustrated in (40).

²⁵⁴ The glottal stop (represented as <'> here) in person forms corresponds to /k/ in other varieties of Paiwan (hence 'u vs. ku '1SG.NMLZ' and ni=a'en vs. ni=aken 'NMLZ=1SG'). While Northern varieties like Sandi/Stimur and Saijia/Tjayljaking are reported to use /k/ (A. Chang 2006), Southern varieties like Gaoshi/Kuskus, Shimen/Tjuaqaciljai (S. Wu 2010), and Mudan/Sinvaudjan (C. Wu 2013) opt for the glottal stop. Although Majia/Makazayazaya Paiwan is often considered a Northern variety, my consultant consistently uses the glottal stop for person forms, though she does have /k/ in some other lexical items.

²⁵⁵ See Table 6.6 for three structural types of possessive NPs in Paiwan.

(40) Pngawan Atayal (Fieldnotes)

- a. **lukus=mu**
clothes=1SG.NMLZ
'my clothes'
- b. **kinang=[ka lukus hani]**
1SG.NMLZ=TOP clothes PROX
'This (piece of) clothing is mine.' (See also C. Shih 2007: 25)

In Pngawan, nominalized indexes bear the same form as Topic indexes among SAPs except for 1SG. Possessive substantives, on the other hand, are constructed from a different set of person-form roots marked by the NMLZ *-an*, which also nominalizes personal nouns, as shown in Table 7.6 (cf. L. Huang 2006: 236; C. Shih 2007: 24).²⁵⁶

Table 7.6: Topic vs. nominalized person forms in Pngawan Atayal

	1SG	1EXCL	1INCL	2SG	2PL
Topic indexes	= <i>cu</i>				
Nominalized indexes (for modification-use only)	= <i>mu</i>	= <i>min</i>	= <i>ta</i>	= <i>su</i>	= <i>mamu</i>
Nominalized nominals (for NP-use only)	<i>kinan(g)</i>	<i>caminan</i>	<i>itan</i>	<i>sinan(g)</i>	<i>cimunan</i>

Third, a situation strikingly similar to that in Pngawan Atayal is also found in Budai Rukai, as shown in (41).

²⁵⁶ L. Huang (2006: 236) listed both *kinan* and *kinang* for 1SG as well as *sinan* and *sinang* for 2SG, but C. Shih (2007: 24) gave only the form ending with the velar nasal for both 1SG and 2SG, that is, *kinang* and *sinang* respectively, which are also the forms I elicited from my two Pngawan consultants. The alternative forms ending with the alveolar nasal (i.e. *kinan* and *sinan*) would have made the whole paradigm consistently suffixed by *-an*, and are listed here for the record (hence the parentheses in the table). It seems safe to treat the alveolar-ending variant as the person-form equivalent of full nominals marked by *-an*, though it is not clear why only 1SG and 2SG allow the alternation between the alveolar and velar nasal. Like many Atayalic languages, Pngawan has portmanteaux person-form clitics, for which interested readers are referred to the two aforementioned references.

(41) Budai Rukai (Fieldnotes)

- a. **senate=li**
 book=1SG.NMLZ
 ‘my book’
- b. **nakuane** *kikai* **senate**
 1SG.NMLZ PROX book
 ‘This book is mine.’ (See also C. Chen 2008: 74)

In Budai, nominalized indexes and subject indexes are identical in form except for 1SG, and possessive substantives are made up of a different set of person-form roots marked by the NMLZ *-ane*, which is a cognate with Plngawan *-an* and nominalizes all sorts of nouns. The three functional paradigms are listed in Table 7.7 (cf. P. Li 1996: 210-211; Zeitoun 1997: 316; Y. Tang 2008: 16).²⁵⁷

²⁵⁷ The person forms in Table 7.7 are generally in agreement with those listed in the three references mentioned above, except that subject and nominalized indexes are treated here as enclitics rather than suffixes. Zeitoun (1997: 337) explicitly states that subject indexes in Budai are clitics, but is non-committal to the cliticness of nominalized indexes. I base the cliticness of nominalized indexes on two criteria. One is the promiscuity criterion, whereby nominalized indexes attach to not only nouns but also verbs (not restricted to those marked by the NMLZ *-ane*), as in (i) below.

- (i) *kabange* *ku* ***ki-acebe=li***
 bag NOM PASS-give.presents=1SG.NMLZ
 ‘What I was given is a bag.’ (Fieldnotes)

The second criterion is relative order, whereby nominalized indexes are added to a sequence already containing a clitic. In (ii), the temporal clitic *=nga* occurs closer to the host than nominalized indexes, much as it does when subject indexes are involved, as in (iii).

- (ii) *ta-kabarenger-ane=nga=li*
 RLS-miss-NMLZ=already=1SG.NMLZ
 ‘someone who I miss already’ (C. Chen 2008: 96)
- (iii) *wa-kane=nga=naku* *ku* *urasi*
 RLS-eat=already=1SG.NOM OBL yam
 ‘I already ate the yam.’ (C. Chen 2008: 180)

Table 7.7: Subject vs. nominalized person forms in Budai Rukai²⁵⁸

	1SG	1EXCL	1INCL	2SG	2PL
Subject indexes	= <i>aku</i> / <i>naku</i>				
Nominalized indexes (for modification-use only)	= <i>li</i>	= <i>nai</i>	= <i>ta</i>	= <i>su</i>	= <i>numi</i>
Nominalized nominals (for NP- and modification-use)	<i>nakuane</i>	<i>naiane</i>	<i>mitane</i>	<i>musuane</i>	<i>numiane</i>

Possessive substantives are sometimes regularly formed by subject/nominalized indexes combined with the NMLZ *-ane*, as in 1EXCL and 2PL, but show morphological idiosyncrasies at other times, as in 1INCL and 2SG.

The last language in the first variety of Type III is Saisiyat, as illustrated in (42).

(42) Saisiyat (Fieldnotes)

- a. *tatpo'* ***ma'an***
 hat 1SG.NMLZ
 'my hat' (See also P. Li 1978: 382)
- b. *hini'* *tatpo'* (*mina*), ***'anmana'a***
 PROX hat EMPH 1SG.NMLZ
 'This hat is (indeed) mine.'

While person forms in the same paradigm as *ma'an* have only the modification-use, those in the same paradigm as *'anmana'a* have the NP- and modification-use, all consisting of person-form roots sandwiched by the NMLZ *'an...a*, which also nominalizes various full nominals. SAP person forms for these two paradigms, together with those for the Topic, are listed in Table 7.8 (cf. Tsuchida 1989: 6; M. Yeh 2003: 17; Kaybaybaw 2009: 11).²⁵⁹

²⁵⁸ See §3.4.1 for how the grammatical category subject is defined in Budai Rukai. For reasons stated in Footnote 84, /*nakuane*/ and /*naiane*/ are phonetically realized as [na.kwa.nə] and [naj.ja.nə] respectively while the form /*mitane*/ results from the coalescence of /*mita*/ and /*ane*/.

²⁵⁹ Due to the free variations between the NMLZ *'an...a* and *'in...a* (see Footnote 210), nominalized

As can be seen from the table, person-form roots that are nominalized by *'an...a* are sometimes identical to nominalized nominals exclusively for the modification-use (*niya'om* for 1EXCL, *mita'* for 1INCL, and *nimon* for 2PL), but othertimes idiosyncratic (as in 1SG and 2SG). Moreover, the Topic pronominal for 2PL is the only SAP base that can be nominalized by *'an...a*.

Table 7.8: Topic vs. nominalized person forms in Saisiyat

	1SG	1EXCL	1INCL	2SG	2PL
Topic	<i>yako</i> <i>ya:o</i>	<i>yami</i>	<i>'ita'</i>	<i>So'o</i>	<i>moyo</i>
Nominalized nominals (for modification-use only)	<i>ma'an</i>	<i>niya'om</i>	<i>mita'</i>	<i>niSo'</i>	<i>nimon</i>
Nominalized nominals (for NP- and modification-use)	<i>'anmana'a</i>	<i>'anniya'oma</i>	<i>'anmita'a</i>	<i>'anSo'a</i>	<i>'anmoyo'a</i> <i>'annimona</i>

In the second variety of Type III languages, person-form possessive substantives are so specialized in form that no consistent NMLZ can be synchronically isolated from person-form roots. Two languages belong to this variety. One is Tgdaya Seediq, as illustrated in (43).

pronominals like *'anmana'a* '1SG.NMLZ' all have variant forms like *'inmana'a*. These variant forms are not listed here, but can be found in Tsuchida (1989: 6). Other than this, the major difference between Table 7.8 and similar tables in the three previous studies lies in the forms of nominalized pronominals for the NP-use. For instance, both Tsuchida (1989: 6) and M. Yeh (2003: 17) transcribed the 1EXCL form as *'anya'oma*, but I choose to transcribe it as *'anniya'oma* [ʔan.nja.ʔo.ma] instead (cf. its variant form *'inniya'oma* in Kaybaybaw 2009: 11) because its corresponding form for the modification-use only is conventionally represented as *niya'om* [nja.ʔom] (as in M. Yeh 2003). Second, both Tsuchida (1989: 6) and M. Yeh (2003: 17) gave the 2SG form *'anSo'a'a* [ʔan.fo.ʔa.ʔa], but the form I got from my consultant (born in 1945) is simply *'anSo'a* [ʔan.fo.ʔa]. Thus, the latter simpler form is listed in the table. Finally, in addition to the 2PL form *'anmoyo'a*, my consultant also provided *'annimona* [ʔan.ni.mo.na], which is not documented in previous studies.

(43) Tgdaya Seediq (Fieldnotes)

a. **sapah=mu**

house=1SG.NMLZ

‘my house’

b. **naku** ka=[sapah nii] han

1SG.NMLZ TOP=house PROX PRT

‘This house is mine.’ (See also Ochiai 2009: 24)

Table 7.9 (cf. Holmer 1996: 32; H. Chang 1997: 13; Ochiai 2009: 13; Holmer & Billings 2014: 114) shows that independent nominalized nominals like *naku* in (43)b are not regularly derivable from nominalized indexes like =*mu* in (43)a, which are identical in form to Topic indexes except for 1SG.

Table 7.9: Topic vs. nominalized person forms in Tgdaya Seediq

	1SG	1EXCL	1INCL	2SG	2PL
Topic indexes	= <i>ku</i>				
Nominalized indexes (for modification-use only)	= <i>mu</i>	= <i>nami</i>	= <i>ta</i>	= <i>su</i>	= <i>namu</i>
Nominalized nominals (for NP- and modification-use)	<i>naku</i>	<i>nami</i>	<i>nita</i>	<i>nisu</i>	<i>namu</i>

Independent nominalized nominals (i.e. those in the third row of Table 7.9) may additionally modify a PUM noun, as in *sapah naku* (house 1SG.NMLZ) ‘my house’ (H. Chang 1997: 15), which has a meaning similar to that of (43)a, but more emphatic or contrastive. Moreover, in all the three sets of forms those for 1EXCL and 2PL (respectively /*nami*/ and /*namu*/) are identical. However, distributionally speaking, they are on one hand free forms that are in a paradigmatic relationship with independent nominalized nominals like *naku* for 1SG, as in (43)a. On the other hand, the two syncretic forms are

Wackernagel clitics indexing the Topic, just like the bound person form =*ku* for 1SG, as in (44).

(44) Tgdaya Seediq

- a. **b**<n>**be-an=ku=daha**
 <PFV>hit-LF=1SG.TOP=3PL.ACT
 ‘They hit me.’ (Holmer & Billings 2014: 118)
- b. **q**<n>**ta-an=namu=daha**
 <PFV>see-LF=2PL.TOP=3PL.ACT
 ‘They saw you.’ (Holmer 2002: 344)

The other language in the second variety is Central Amis, as shown in (45).

(45) Central Amis (Fieldnotes)

- a. **cudad aku**
 book 1SG.NMLZ
 ‘my book’
- b. **maku ku=ni a cudad**
 1SG.NMLZ TOP.CMN=PROX LIG book
 ‘This book is mine.’ (See also J. Wu 2006: 87)

Similarly, nominalized nominals for the NP-use, such as *maku* in (45)b, are not always predictable from those exclusively for the modification-use, such as *aku* in (45)a. SAP person forms for these two paradigms, together with those for the Topic, are listed in Table 7.10 (cf. L. Huang 1995b: 236; J. Wu 2006: 85).

Table 7.10: Topic vs. nominalized person forms in Central Amis

	1SG	1EXCL	1INCL	2SG	2PL
Topic	<i>kaku</i>	<i>kami</i>	<i>kita</i>	<i>kisu</i>	<i>kamu</i>
Nominalized nominals (for modification-use only)	<i>aku</i>	<i>niyam</i>	<i>ita</i>	<i>isu</i>	<i>namu</i>
Nominalized nominals (for NP- and modification-use)	<i>maku</i>	<i>niyam</i>	<i>mita</i>	<i>misu</i>	<i>namu</i>

The *maku*-series also has the modification-use, but its modification pattern differs from that of the *aku*-series. The latter allows only postnominal modification, as in (45)a, whereas the former only prenominal modification, as in *maku a wawa* (1SG.NMLZ LIG child) ‘my child’ (J. Wu 2006: 87).

Finally, the last two languages share two intriguing properties. One is that among SAPs nominalized person forms for the NP-use are identical to those exclusively for the modification-use only in the case of 1EXCL and 2PL. This is analogous to English *his*, which is used for both the NP- and modification-use (as in *This book is his.* and *This is his book.* respectively), unlike other person forms in the language. The other property Tgdaya Seediq and Central Amis share is that free nominalized person forms can be additionally marked by something that is most likely a reflex of PAN *nu (see §6.6), thus giving rise to their long-form counterparts.

Table 7.11 below summarizes the short and long nominalized person forms in Tgdaya Seediq, where long forms are consistently marked by *n* (realized as [nu] or [ne]) on top of short forms.²⁶⁰ Documentation of these forms is not complete in previous

²⁶⁰ There is also a phonologically similar marker /n/ (realized as [nu] or [nə] depending on speakers), which

Table 7.11: Free nominalized person forms in Tgdaya Seediq

	1SG	1EXCL	1INCL	2SG	2PL
Short forms	<i>naku</i>	<i>nami</i>	<i>nita</i>	<i>nisu</i>	<i>namu</i>
Long forms	<i>n=naku</i>	<i>n=nami</i>	<i>n=nita</i>	<i>n=nisu</i>	<i>n=namu</i>

studies. H. Chang (1997: 13) mentioned only the short forms. Holmer (1996: 32, 2002: 343) listed only short forms for 1SG/1INCL/2SG, but only long forms for 1EXCL/2PL. Ochiai (2009: 13) additionally gave long forms for 1SG/1INCL/2SG. It turns out that both short and long forms are attested for all SAPs, as has been confirmed by my consultants. One possible reason for the scattered documentation of these forms is that it is unclear how the short and long forms differ in terms of functions. For instance, like their short counterparts, the long forms have both the NP- and modification-use, as shown in (46).

adds past-time reading to certain stative predicates, as in the following example, where both instances of /n/ are realized as [nu].

Tgdaya Seediq (H. Chang 2000: 97)

- (i). *n=biciq laqi nii, para saya=di*
 PST=small child PROX big now=already
 ‘This child was small, (but) is big now.’
- (ii). *n=naku sapah nii, uxe saya=di*
 PST=1SG.NMLZ house PROX NEG now=already
 ‘This house used to be mine, (but) not anymore now.’

Probably due to the phonological similarity between /n/ as a NMRK and as a past tense marker, my three Tgdaya consultants did not have unanimous judgements for a given instance of /n/. Take the short possessive substantive *naku* for instance. Two of them (Dakis Pawan, born in 1954, and Lubi Neyung, born in 1934) realized the past tense *n* as [nə] in front of *naku* (i.e. *n=naku* [nənaku] ‘PST=1SG.NMLZ’ meaning ‘what used to be mine’) while the other consultant (Ape Neyung, born in 1946) expressed the same meaning as [nunaku], which was however interpreted by one of the first two speakers as ‘NMRK=1SG.NMLZ’ meaning ‘mine’ (without the past-time reading), interchangeable with simply *naku*. More research is needed in order to find out the conditioning factors of the phonetic realizations of /n/ as a NMRK and as a past tense marker. Despite the potential ambiguities between the two functions, it is believed that we are dealing with two separate morphemes, rather than one, because the two can occur consecutively, as in *n=n=naku* [nununaku] ‘PST=NMRK=1SG.NMLZ’ meaning ‘what used to be mine’ and *n=n=heya* [nuneheya] ‘PST=NMRK=3SG’ meaning ‘what used to be his/hers.’ In Table 7.11, I specifically discuss the /n/ as a NMRK. Finally, short and long possessive substantives like the forms listed here are also found in Truku Seediq. See L. Hsu (2008) and Tsukida (2009) for details.

(46) Tgdaya Seediq (Ochiai 2009: 24)

- a. rulu **n=naku**
 car NMRK=1SG.NMLZ
 ‘my car (or car of mine)’
- b. **n=naku** ka=[rulu ga]
 NMRK=1SG.NMLZ TOP=car DIST
 ‘That car is mine.’

This synchronic situation presents a challenge for how to gloss *n* properly. Here NMRK is chosen simply because of its possible historical connection with PAn **nu*, as has been explicated in §6.6.

The same principle applies to *nu* in Central Amis, which derives long forms out of short nominalized person forms, as summarized in Table 7.12 below. Like those in Tgdaya Seediq, the short and long forms in Central Amis do not seem to vary much in functions, in terms of either their NP-use (serving as arguments or nominal predicates), as in (47), or prenominal modification-use, as in (48).

Table 7.12: Free nominalized person forms in Central Amis

	1SG	1EXCL	1INCL	2SG	2PL
Short forms	<i>maku</i>	<i>niyam</i>	<i>mita</i>	<i>misu</i>	<i>namu</i>
Long forms	<i>nu=maku</i>	<i>nu=niyam</i>	<i>nu=mita</i>	<i>nu=misu</i>	<i>nu=namu</i>

(47) Central Amis

- a. mi-ala ci=aki **tu=maku** atu misu
 AF.EXT-take TOP.PSN=A. UND.CMN=1SG.NMLZ and 2SG.NMLZ
 ‘Aki is going to take mine and yours.’ (J. Wu 2006: 86)
- b. ira ku=nira a ’udax, k<um>aen=ho
 EX TOP.CMN=3SG.NMLZ LIG candy AF.<UM>eat=still
tu=nu=maku
 UND.CMN=NMRK=1SG.NMLZ
 ‘{She/He} has candies, (but) is still eating mine (regardless).’ (L. Huang 1995: 231)
- c. **(nu)=maku** ku=ni a cudad
 NMRK=1SG.NMLZ TOP.CMN=PROX LIG book
 ‘This book is mine.’ (Fieldnotes; see also J. Wu 2006: 87)

(48) Central Amis

- a. **(nu)=maku** a ’udax
 NMRK=1SG.NMLZ LIG candy
 ‘my candy’ (L. Huang 1995: 230)
- b. **(nu)=maku** a wawa
 NMRK=1SG.NMLZ LIG child
 ‘my child’ (J. Wu 2006: 87)

However, the long-form *nu=maku*-series does share some functions with the *aku*-series, which only has the modification-use, to the exclusion of the short-form *maku*-series. Specifically, the *nu=maku*- and *aku*-series, but not the *maku*-series, can indicate the possessor after a noun, as in (49), or the non-Topic Actor after a NAF-verb, as in (50). Removing *nu* from the (b) examples in both (49) and (50) would produce unacceptable sentences that are otherwise just as felicitous as the (a) examples.

(49) Central Amis (J. Wu 2006: 86)

- a. fangcal ku=wawa **aku**
 AF.good TOP.CMN=child 1SG.NMLZ
 ‘My child is good.’
- b. fangcal ku=wawa ***(nu)=maku**
 AF.good TOP.CMN=child NMRK=1SG.NMLZ
 ‘My child is good.’

(50) Central Amis (L. Huang 1995: 231)

- a. fafa-en **aku** kisu
 carry.on.back-PF 1SG.ACT 2SG.TOP
 ‘I will carry you on (my) back.’
- b. fafa-en ***(nu)=maku** kisu
 carry.on.back-PF ACT=1SG 2SG.TOP
 ‘I will carry you on (my) back.’

Table 7.13 below summarizes the functional distributions of the three series of Amis person forms that have been demonstrated above, with each series illustrated by 1SG. Some observations can be made from the table. First, the *aku*-series is syntactically dependent, either on a noun or verb, functionally comparable to bound person forms in other Formosan languages. Second, the short-form *maku*-series semantically denotes only entities associated with person forms, but never refers to the Actor, much like the independent nominalized person forms in Seediq (both short and long; see Table 7.11). Finally, the long-form *nu=maku*-series has the maximal number of functions, some of which are shared with the *aku*-series only and others with the short-form *maku*-series only. This convergence of functions might have been because the long-form *nu=maku*-series, which contains the PAn *nu etymon, was later recruited to mark the Actor after Amis *nu* became further grammaticalized (see §6.6), considering this function is not available to the long-form *n=naku*-series in Tgdaya Seediq, which contains a promising reflex of the same PAn etymon.

Table 7.13: Functional distributions of three sets of person forms in Central Amis

	<i>aku</i> -series	<i>maku</i> -series	<i>nu=maku</i> -series
NP-use	* <i>aku</i>	<i>maku</i>	<i>nu=maku</i>
Prenominal modification-use	* <i>aku</i> + <i>a</i> + N	<i>maku</i> + <i>a</i> + N	<i>nu=maku</i> + <i>a</i> + N
Postnominal modification-use	N + <i>aku</i>	*N + <i>maku</i>	N + <i>nu=maku</i>
Postverbal Actor	V + <i>aku</i>	*V + <i>maku</i>	V + <i>nu=maku</i>

Note: N stands for underived PUM nouns and V is short for NAF-verbs. All person forms in this table concern 1SG, which is meant to represent the three series of person forms. Constructions marked with an asterisk are ungrammatical.

7.3. Chapter summary

Table 7.14 below summarizes the three types (A through C) of full-nominal possession discussed in §7.1 and another three (I through III) of person-form possession presented in §7.2 across the fifteen languages investigated. The four languages in italics (Northern Paiwan, Tgdaya Seediq, Central Amis, and Takibakha Bunun) are those where full-nominal possession and person-form possession do not make use of the same repertoire of coding possibilities for the POR-PUM syntagm as opposed to possessive substantives.

Table 7.14: Full-nominal and person-form possession in Formosan languages

Full-nominal possession	Languages	Person-form possession	Languages
A: Unmarked share-NMLZ	Isbukun Bunun Thao <i>Northern Paiwan</i> <i>Tgdaya Seediq</i> <i>Central Amis</i>	I: Unmarked share-person	Isbukun Bunun Thao
B: Marked share-NMLZ	Rikavung Puyuma Taromake Rukai Squliq Atayal Kavalan Saaroa Tsou	II: Marked share-person	Rikavung Puyuma Taromake Rukai Squliq Atayal Kavalan Saaroa Tsou <i>Takibakha Bunun</i>
C: Split-NMLZ	Pngawan Atayal Budai Rukai Saisiyat <i>Takibakha Bunun</i>	III: Split-person	Pngawan Atayal Budai Rukai Saisiyat <i>Northern Paiwan</i> <i>Tgdaya Seediq</i> <i>Central Amis</i>

In the nominalization-based account, the so-called POR is the reference entity with which the denotation of a nominalized nominal is metonymically related. As a full nominal in its modification-use modifying a PUM noun, the POR is marked by a relator called NMLZ in most languages or simply juxtaposed to the PUM in only a few (e.g. Saaroa). In both types of languages, nominalized person forms indexing a full nominal may be optional (cf. *senate=(ini) ki cegau* [book=3.NMLZ NMLZ C.] ‘Cegau’s book’ in Budai Rukai and *suhlata=(isa) amahle* [book=3.NMLZ A.NMLZ] ‘Amahle’s book’ in Saaroa). Moreover, when it is not modifying any lexical noun, the full-nominal POR is almost always marked by relators called NMLZ/NMRK such that a plain nominal and its

nominalized counterpart bear distinct forms. One exception is found in Thao, where the phrase *ti=X* may refer to X (i.e. a plain nominal) or what is metonymically associated with X (i.e. a nominalized nominal), depending on the context.

On the other hand, the languages investigated predominantly have nominalized SAP person clitics that attach directly to a PUM noun while only two of them (Saisiyat and Amis) have nominalized SAP person forms which are phonologically free but morphosyntactically bound and that are exclusively reserved for the modification-use. Only two languages (Kaohsiung Isbukun Bunun and Thao) are exceptional in this regard because they have neither nominalized SAP person clitics nor modification-only nominalized SAP person forms. This is presumably due to their loss of nominalized SAP clitics, which are now only observable as SAP clitics indexing the (non-Topic) Actor, though only vestigially so in Thao. Finally, SAP person-form roots are nominalized for the NP-use either by combining with NMLZ/NMRK or taking on specialized forms that are synchronically indecomposable. In either case, a plain person form and its nominalized counterpart almost always bear distinct forms. Exceptions are found in only two languages and are only specific to certain person forms rather than generalizable to all SAP forms in the same paradigm. Specifically, the 2PL *manium* in Thao as well as the 1EXCL *nami* and 2PL *namu* in Tgdaya Seediq may refer to persons (i.e. a plain person form) or entities associated with persons (i.e. a nominalized person form), depending on contextual cues.

|Chapter 8

Conclusion

This dissertation has investigated two broad types of grammatical nominalizations in Formosan languages under Shibatani's (2009, 2010, 2011) functional framework where nominalization is essentially a metonymic operation yielding denoting expressions. The first type is verbal-based, where denoting expressions are created in relation to a state of affairs. Within verbal-based nominalizations, I have specifically concentrated on argument nominalizations, those that denote event participants, which are expressed by word forms containing Focus affixes (or their cognates in the case of Rukai) together with their potential argument-structure materials. The Mstem (containing PAn AF **<um>* or **ma-*) is attested in all Formosan languages and predominantly used as argument nominalizations except for a handful of languages that are likely to be innovators. By contrast, distributions of Suffixal (PAn **-aw/ay/anay* for the *a*-grade series and **-u/a/i/ani* for the zero-grade series) and Mixed (PAn **-en*, **-an*, **Sa/Si-*, and **<in>*) NAF affixes vary greatly among Formosan languages, with both sets retained in only a few

conservative languages like Atayal, Paiwan, and Puyuma. Both the SPR hypothesis (Starosta et al. 1982) and Nuclear hypothesis (Ross 2009, 2012) propose that Mixed NAF-words were originally nominalizations and later reanalyzed as verbs in either PAN or some later stage after that (i.e. the nominalization-into-verb reanalysis) whereas Mixed NAF-words have been verbal throughout the Austronesian history. However, this study has shown that Suffixal NAF-words are just as capable of creating argument nominalizations as Mixed ones and that the putative verbal-nominal contrast between these two types of NAF-words does not predict their syntactic distributions across Formosan languages, in either the proposed Nuclear or non-Nuclear group.

Specifically, Suffixal zero-grade NAF-words in at least Tsou, Puyuma, and Atayalic (i.e. Atayal proper and Seediq) are component structures of argument nominalizations just as Mixed NAF-words are if the latter are also retained in those languages. Similarly, Suffixal *a*-grade NAF-words in Siraya are just as amenable to argument nominalizations as its Mixed NAF-words, although all the other Formosan languages strictly prohibit their Suffixal *a*-grade NAF-words from participating in argument nominalizations irrespective of the semantics of those NAF-words. More importantly, in languages where Suffixal NAF-words are allowed to create argument nominalizations, they are subject to the same Topic-only constraint as those made up of Mixed NAF-words. That is, these argument nominalizations only denote the Topic NP of the Focus-word involved. All these distributions and constraints suggest that both Suffixal and Mixed NAF-words in PAN were a single word class, which could function equally as predicates or arguments. If so, PAN would create argument nominalizations by making use the gap strategy whereby the Topic NP of a Focus-word is gapped. In fact,

the same situation is continued in the majority of modern Formosan languages with only some exceptions. In this regard, PAn would be typologically similar to many Native American languages (see Comrie & Estrada-Fernández 2012).

However, there is a caveat regarding the generalization above. The illocutionary forces of a Focus-word should not interfere with the presuppositionality required for argument nominalizations, which are meant to denote an event participant in terms of a presupposed state of affairs. Otherwise, the pragmatic incongruence would prevent that Focus-word from serving as argument nominalizations. Specifically, Suffixal *a*-grade NAF-words in Formosan languages predominantly have hortative/optative/imperative meanings, all of which share the property of expressing the speaker's volition to have a prescribed action executed (Xrakovskij 2001). These directive meanings are pragmatically at odds with argument nominalizations, which may account for the nearly exceptionless ban on Suffixal *a*-grade NAF-words from occurring in argument nominalizations. The same situation can be inferred to obtain in PAn as well. To the extent of data availability, the only language where there is no such ban is Siraya, where Suffixal *a*-grade NAF-words can be as entity-denoting as Mixed NAF-words. The exception in Siraya is most likely due to a later grammaticalization process that extended the semantics of *a*-grade NAF-words from speaker-oriented volition to objective future such that these forms can be dissociated from directive illocutionary forces. Thus, unlike previous studies, this present one seeks the explanation for the overall ban on Suffixal *a*-grade NAF-words from argument nominalizations in the constraints on their illocutionary forces rather than their presumably verbal nature. This is additionally supported by the crosslinguistic trend that "constraints on/loss of illocutionary elements" happen prior to

anything else in a nominalization process (Lehmann 1988). Crucially, if an *a*-grade NAF-word can be used for argument nominalizations, it implies that the same word form also has non-directive meanings, as in Siraya, but not vice versa, as in Puyuma. Several pieces of evidence suggest that *a*-grade NAF-word in Puyuma might have undergone a desemanticization process where the directive meanings of *a*-grade NAF-words were expanded to include indicative-realis meanings, triggered by the loss of a directive auxiliary that used to dictate and precede *a*-grade NAF-word (per Sagart 2010 and Aldridge 2016). Perhaps because this functional extension was relatively recent, *a*-grade NAF-words in Puyuma still maintain the earlier restriction against them for argument nominalizations even though they have acquired indicative meanings. Finally, two *a*-grade NAF-words are attested in Kananavu, but one of them (i.e. STEM-*ai* < PAn LF *STEM-*ay*) has indicative-realis meanings whereas the other (i.e. STEM-*aw* < PAn PF *STEM-*aw*) has imperative meanings, showing that the acquisition of indicative-realis meanings among *a*-grade NAF-words can be sporadic. Since there is no established evidence for Puyuma and Kananavu being in the same subgroup, the semantic change from directive to indicative-realis in both languages is most likely an independent innovation.

Just as arguing for the Nuclear hypothesis has implications for genetic classifications, so too does arguing against it. Based on phonological evidence of exclusively shared innovations, Blust (1999, 2009) concluded that Formosan languages constitute no fewer than nine first-order subgroups within Austronesian, and that non-Formosan Austronesian languages altogether form another subgroup called Malayo-Polynesian, which has well-defined shared innovations. Subsequently, Ross's (2009,

2012) Nuclear hypothesis proposed a modification on Blust's subgrouping by putting forth the subgroup called Nuclear Austronesian (i.e. all Austronesian but Puyuma, Rukai, and Tsou), which is solely based on the morphosyntactic change of reanalyzing Mixed Focus-words from nominalizations into verbs. While there are ten first-order subgroups of the Austronesian family in Blust's subgrouping, there are only four in Ross's. However, this study has shown that even by morphosyntactic criteria Nuclear languages (specifically Formosan Nuclear) do not share properties to the exclusion of non-Nuclear ones, thus questioning the validity of the Nuclear subgroup. For instance, Chapter 5 investigates many Mixed Focus-words in Central Amis, a Nuclear language, and some of them exclusively have external nominal syntax just as Mixed Focus-words in Puyuma, a non-Nuclear language, were claimed to do in the Nuclear hypothesis. In other words, synchronic variations on the pattern of argument nominalizations cut across the divide between the proposed Nuclear and non-Nuclear languages. More importantly, an interdisciplinary study (Gray et al. 2009) based on lexical cognacy, database technologies, and computational phylogenetic methods has arrived at a Maximum Clade Credibility tree of 400 Austronesian languages, of which 18 are Formosan.²⁶¹ The tree shows strong support for a divide between Formosan and Proto-Malayo-Polynesian, which corresponded with a long settlement pause in Taiwan before the early Austronesian people migrated to the Philippines, as argued by Blust (1999). More pertinent to the point is that the tree confirms Formosan languages form several first-order subgroups as per Blust, although one of his first-order subgroup is only weakly supported and another one not supported at all. Crucially, the tree does not support a

²⁶¹ They are Paiwan, Squaliq Atayal, Ci'uli Atayal, Seediq, Saisiyat, Favorlang, Thao, Pazeh, Kananabu, Saaroa, Tsou, Rukai, Kavalan, Basai, Bunun, Central Amis, Puyuma, and Siraya.

subgroup consisting of Malayo-Polynesian plus any Formosan language, let alone the proposed Nuclear subgroup consisting of Malayo-Polynesian plus all Formosan languages but Puyuma, Rukai, and Tsou. Therefore, various types of converged evidence, including lexical, phonological, and morphosyntactic, all corroborate Blust's long-rake structure for first-order subgroups of the Austronesian family.

The second type of nominalization investigated in this study is nominal-based, where denoting expressions are created with respect to a reference entity. In this type, I have focused on variations of possessive substantives in a sample of 15 Formosan languages. The variations are generalized into three structural types, one of which was not identified in Ultan's (1978) two-way typology of possessive substantives based on a sample of 75 languages. The three types are attested regardless of whether the reference entity is expressed by full nominals or person forms. The majority of languages demonstrate consistent types between full-nominal and person-form possession while a few others may belong to one type in terms of full-nominal possession and to another with respect to person-form possession.

The significance of the identified three types are at least two fold. On the one hand, they unveil different syntactic functions of so-called genitive case markers across Formosan languages even when cognate forms are involved. In the majority of Formosan languages where PAN *ni is retained, its reflexes can mark a personal name as the possessor without having to modify any lexical possessum noun. Specifically, the schema "*ni X*", where X stands for a personal name, is all it takes to form independent NPs that are as readily to take on additional marking required by syntax as NPs consisting of regular nouns in some languages (Type A). In some others (Type B), the schema is

obligatorily marked by a highly grammaticalized morpheme (often recruited from demonstratives or definite markers) in order to create independent NPs. Thus, *ni* deserves to be called a nominal-based nominalizer (NMLZ) in that it creates new nominals with denotations different from those of the base nominal X, and the grammaticalized morpheme marking the NP status in Type B is called a nominalization marker (NMRK) following Shibatani & Shigeno (2013). In still other languages (Type C), the same schema has to modify a possessum noun, and *ni* is not involved in the creation of possessive substantives for the NP-use, thus causing a complementary distribution of two nominalized forms, one for the NP-use and the other for the modification-use.

On the other hand, the three types identified in modern languages present some clues as to which of them PAN might belong to as well as the historical development across types. Some evidence suggests that Type C may be a later development resulting from innovating a new coding strategy for the NP-use. Moreover, functional distributions of the reflexes of PAN *nu in Tsou, Tgdaya Seediq, and Central Amis seem to indicate that PAN would use *nu as a NMRK on top of the schema “*ni X”, which would make it a Type B language. This conclusion is in line with Reid’s (1981, 2007) functional reconstruction of PAN *nu as an indefinite entity-denoting marker, which not only establishes a morphological connection between possessive substantives and interrogative words in many modern languages but also accounts for the functional developments of its reflexes in more Formosan languages. By contrast, these advantages would be not available according to the alternative view (Ross 2002; Blust 2005, 2015) that PAN *nu was as a genitive marker (for common nouns) that paradigmatically alternated with PAN *ni (for personal nouns).

In terms of general typology, situating possessive/genitive constructions in the broader definition of nominalization is additionally supported by several implicational universals in the literature on Suffixaufnahme, which is characterized by Moravcsik (1995: 452) as the phenomenon where “an attributive nominal carries two distinct case markers: one appropriate to its own function as an attributive, and the other appropriate to the function of the NP that includes both the attributive and the head.” First, the head nominal phrase is optional in the definitional schema he gave, as in (1).

(1) Definitional schema of Suffixaufnahme (Moravcsik 1995: 452)²⁶²

[[Head Nominal, (External Case)], [Attributive Nominal, Internal case, External case]

This is based on the implicational universal that “[i]n all languages, if Suffixaufnahme occurs with the head present, it also occurs with the head missing” (ibid.: 469). Although Formosan languages have not hitherto been reported to demonstrate Suffixaufnahme, the Amis NP *tu=ni=Kacaw* in (2)a precisely satisfies Moravcsik’s definitional schema, parallel to the simpler NP *tu=futing* in (2)b.

(2) Central Amis

a. *ira ku='udax aku k<um>aen=ho kaku*
 EX TOP.CMN=candy 1SG.NMLZ AF.<UM>eat=still 1SG.TOP

tu=[ni=kacaw]

UND.CMN=NMLZ.PSN=K.

‘I have candies, (but) I’m eating Kacaw’s (regardless).’

(Fieldnotes) [= (56)a in Chapter 6]

b. *ma-ulah ci=kulas a k<um>aen tu=futing*
 AF.INT-like TOP.PSN=K. LIG AF.<UM>eat UND.CMN=fish
 ‘Kulas likes eating fish.’ (ODIL; under the entry *kumaen*)

Since the “attributive nominal” is not syntactically dependent on the head nominal, we might as well treat it and its “internal case” as an independent nominal on its own right,

²⁶² The schema here is a simplified version of the original one, where “[c]ommas indicate co-occurrence without commitment to linear order; square brackets delimit constituents; parentheses indicate optional presence” (Moravcsik 1995: 452).

which is ready to take any “external case” just like an underived nominal, as suggested by Noonan (2008b) and argued for by Shibatani (2013). Second, “[i]n almost all languages, if the internal case involved in Suffixaufnahme is a case other than that of the possessor, the case of the possessor may also be involved in Suffixaufnahme” (Moravcsik 1995: 471). In other words, if a language allows Suffixaufnahme at all, the “internal case” almost always involves a marker for the possessor. The ubiquitousness of possessor marking in Suffixaufnahme can be accounted for if the so-called “internal genitive case” is not an adnominal case as such but a nominal-based nominalizer that makes further case marking possible. Finally, Suffixaufnahme often historically derives from constructions involving pronominal elements. Specifically, “[i]f Suffixaufnahme is historically preceded by a different construction, one possibility for that construction is to involve a *pronoun-like constituent*. This constituent may be an *indefinite pronoun* ‘one’ which is coreferential with the head and which serves as the immediate head for the attributive nominal; or it may be *the definite article* of the head” (ibid.: 477; emphasis mine). These pronominal elements are precisely the precursor of NMLZ or NMRK, which turns attributive expressions into syntactically independent nominals that we call nominalizations, denoting entities characterized by those attributes (e.g. *the poor*, *the poor one*, etc. in English). More importantly, regardless of whether pronominal elements are involved, the same coding strategy may be used to derive a nominalization out of both a nominal and verbal base, as in many Asian languages (see Yap et al. 2011). Within Formosan, nominal- and verbal-based nominalization are also found to share the same marking in Budai Rukai and Plngawan Atayal (both using reflexes of PAn *-an),

although the dominant pattern in most other Formosan languages is to maintain separate coding mechanisms for these two types of nominalization.

Last but not least, the functional-typological approach adopted in this study has made it possible to look into nominalization in constructional terms. That means making reference to something in terms of something else that is mentally more accessible, be it a state of affairs or a reference entity, via metonymic connections between the two. This agrees with the idea of part-whole integration in Cognitive Linguistics, in the terminology of which products of a nominalization process are composite structures integrated by component structures that facilitate the identification of their composite structures (Langacker 2009a). Seeing nominalizations as gestalt composite structures helps to ensure that when two constructions across languages are compared the comparability is rooted in what propositional acts (Croft 2001: 88) those constructions carry out rather than what word forms are involved in those constructions. For instance, while the Mstem in Budai Rukai has been claimed to form a “relative clause” (L. Sung 2011), the *ta*-Kstem in Mantauran Rukai is considered to be a “derived nominal” in a paradigmatic relationship with underived nouns (Zeitoun 2002). However, both word forms are component structures of composite denoting expressions that target at the grammatical subject as defined in both linguistic systems. In terms of component structures, both forms take patientive arguments in the same way as verbal forms in matrix predication would (i.e. the Mstem in both Budai and Mantauran), and, unlike underived nouns, both do not host possessor-indexing clitics. Meanwhile, composite structures consisting of either form are used as referring expressions in a paradigmatic relationship with underived nouns (the NP-use) or as modifying expressions in a

syntagmatic relationship with a nominal (the modification-use). These propositional acts are the basis for comparing the composite structures in Budai and Mantaoran and consequently for treating them equally as nominalizations. However, whether component structures of nominalizations are also used elsewhere as the verbal form in main predication (yes in Budai, but no in Mantaoran) is subject to language variations and historical changes, and is thus largely irrelevant to nominalization. On the other hand, the gestalt view of nominalization also avoids some pitfalls in previous studies. A common description of argument nominalizations in many Formosan languages is that “[e]xcept for [F]ocus and tense/aspect/mood markers...there are no productive morphological devices to produce lexical nominals; nominalized elements and verb forms are...identical” (L. Huang 2002: 197). However, a less misleading way to characterize the same observation is that nominalized elements are composite structures consisting of component structures like verb forms marked for tense/aspect/mood (i.e. Focus-words). This is supported by the fact that Focus-words show the same subcategorization properties to the exclusion of underived nouns whether they function as matrix predicates or arguments. This solution also avoids having to draw a line between lexical and clausal nominalizations involving the same Focus-word when there is no rigorous basis for such a distinction in numerous Formosan languages other than the contingent realizations of argument-structure materials. That is to say, nominalized composite structures have nominal distributions while their component structures maintain verbal properties of Focus-words to a large extent. Nevertheless, when nominalized composite structures consist of nothing but Focus-words as their component structures, they become the immediate constituent of a nominal relation marker on a par with underived nouns, hence

opening up the chance of Focus-words being lexicalized and associated with specialized semantics. One such example is the PF-word *c<in>avu* ‘<PF.PFV>wrap’ in Paiwan, which refers to rice dumplings wrapped in (banana or shell ginger) leaves, but not just any wrapped thing. The same analysis appears even more evident in Tsou, which, unlike all the other Formosan languages, requires Focus-words to be almost always preceded by auxiliaries. This grammatical constraint eliminates the possibility of Focus-words being the immediate constituent of a nominal relation marker on a par with underived nouns, thus shielding Focus-words from undergoing lexicalization. However, even in such a unique language nominalized composite structures consisting of Focus-words may still lexicalize into nouns as defined in the grammar Tsou, such as dispositional agentive nominals in the form of *le*-Mstem. Therefore, within a nominalized composite structure, its component structures like Focus-words may inherit the same clause-like properties as found in non-nominalized contexts. This result resonates well with the conclusion in Cognitive Linguistics that “nominalizations cannot be reduced to clauses, but consist of relationships that also occur at clause level” (Heyvaert 2010: 73).

References

- Abbott, Miriam. 1991. Macushi. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages*, vol. 1, 23–160. Berlin: Mouton de Gruyter.
- Adelaar, Alexander. 2013. Reviving Siraya: A case for language engineering. *Language Documentation & Conservation* 7. 212–234.
- Adelaar, K. Alexander. 2011. *Siraya: Retrieving the phonology, grammar and lexicon of a dormant Formosan language*. Berlin: Walter de Gruyter.
- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2011. Word-class changing derivations in typological perspective. In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Languages at large: Essays on syntax and semantics*, 221–289. Leiden: Brill.
- Aikhenvald, Alexandra Y. 2013. Possession and ownership: A cross-linguistic perspective. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Possession and ownership: A cross-linguistic typology*, 1–64. Oxford: Oxford University Press.
- Aldridge, Edith. 2004a. Ergativity and word order in Austronesian languages. Ithaca: Cornell University (Ph.D. dissertation).
- Aldridge, Edith. 2004b. Internally headed relative clauses in Austronesian languages. *Language and Linguistics* 5(1). 99–129.
- Aldridge, Edith. 2012. Antipassive and ergativity in Tagalog. *Lingua* 122. 192–203.
- Aldridge, Edith. 2013. Wh-clefts and verb-initial word order in Austronesian languages. In Katharina Hartmann & Tonjes Veenstra (eds.), *Cleft structures*, 71–96. Amsterdam: John Benjamins.
- Aldridge, Edith. 2014. Subjunctive and the emergence of ergativity in Austronesian languages. Presented at the Institute of Linguistics, Academia Sinica, Taipei, March 17.
- Aldridge, Edith. 2016. Ergativity from subjunctive in Austronesian languages. *Language and Linguistics* 17(1). 27–62.
- Alexiadou, Artemis & Florian Schäfer. 2010. On the syntax of episodic vs . dispositional -er nominals. In Artemis Alexiadou & Monika Rathert (eds.), *The syntax of nominalizations across languages and frameworks*, 9–38. Berlin: Mouton de Gruyter.
- Allen, Cynthia L. 2002. The development of “strengthened” possessive pronouns in English. *Language Sciences* 24(3–4). 189–211.
- Anderson, Stephen. 2005. *Aspects of the theory of clitics*. (Oxford Studies in Theoretical Linguistics 11). Oxford: Oxford University Press.
- Andrews, Avery D. 2007. The major functions of the noun phrase. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. 1: Clause structure*, 132–223. Cambridge: Cambridge University Press.
- Aristar, Anthony Rodrigues. 1991. On diachronic sources and synchronic pattern: An investigation into the origin of linguistic universals. *Language* 67(1). 1–33.
- Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT Press.
- Baron, Irène, Michael Herslund & Finn Sørensen (eds.). 2001. *Dimensions of possession*. (Typological Studies in Language 47). Amsterdam: John Benjamins.
- Bhat, Darbhe Narayana Shankara. 2004. *Pronouns*. Oxford: Oxford University Press.
- Bickel, Balthasar. 1999. Nominalization and focus constructions in some Kiranti languages. In Yogendra P. Yadava & Warren G. Glover (eds.), *Topics in Nepalese linguistics*, 271–296. Kathmandu: Royal Nepal Academy.

- Blake, Frank R. 1936. Expression of case by the verb in Tagalog. *Journal of the American Oriental Society* 27. 183–189.
- Bloomfield, Leonard. 1917. *Tagalog texts with grammatical analysis*. Urbana: University of Illinois.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt.
- Blust, Robert. 1998a. A note on the Thao Patient Focus perfective. *Oceanic Linguistics* 37(2). 346–353.
- Blust, Robert. 1998b. Ca- reduplication and Proto-Austronesian grammar. *Oceanic Linguistics* 37(1). 29–64.
- Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected papers from the 8th International Conference on Austronesian Linguistics*, 31–94. Taipei: Academia Sinica.
- Blust, Robert. 2002. Notes on the history of “focus” in Austronesian languages. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 63–78. Canberra: Australian National University.
- Blust, Robert. 2003. *Thao Dictionary*. (Language and Linguistics Monograph Series A-5). Taipei: Academia Sinica.
- Blust, Robert. 2005. A note on the history of genitive marking in Austronesian languages. *Oceanic Linguistics* 44(1). 215–222.
- Blust, Robert. 2009. *The Austronesian Languages*. (Pacific Linguistics 602). Canberra: Australian National University.
- Blust, Robert. 2013. *The Austronesian languages*. (Asia-Pacific Linguistics 8). Canberra: Australian National University.
- Blust, Robert. 2015. The case-markers of Proto-Austronesian. *Oceanic Linguistics* 54(2). 436–491.
- Bois, John W. Du, Stephan Schuetze-Coburn, Susanna Cumming & Danae Paolino. 1993. Outline of discourse transcription. In Jane A. Edwards & Martin D. Lampert (eds.), *Talking data: Transcription and coding in discourse research*, 45–89. Hillsdale, NJ: Lawrence Erlbaum.
- Burgess, Don. 1984. Western Tarahumara. In Ronald Langacker (ed.), *Studies in Uto-Aztecan grammar: Southern Uto-Aztecan grammatical sketches*, 1–149. Dallas: Summer Institute of Linguistics.
- Bussmann, Hadumod. 1996. *Routledge dictionary of language and linguistics*. London: Routledge.
- Bybee, Joan. 2001. Main clauses are innovative, subordinate clauses are conservative: Consequences for the nature of constructions. In Joan Bybee & Michael Noonan (eds.), *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson*, 1–17. Amsterdam: John Benjamins.
- Bybee, Joan & Suzanne Fleischman. 1995. Modality in grammar and discourse: An introductory essay. In Joan L. Bybee & Suzanne Fleischman (eds.), *Modality in grammar and discourse*, 1–14. Amsterdam: John Benjamins.
- Capell, Arthur. 1964. Verbal systems in Philippines languages. *Philippine Journal of Science* 93. 231–249.
- Chang, Anna Hsiou-chuan. 2000. *A reference grammar of Paiwan*. (Series on Formosan Languages 9). Taipei: Yuanliu.
- Chang, Anna Hsiou-chuan. 2006. *A reference grammar of Paiwan*. Canberra: Australian National University (Ph.D. dissertation).
- Chang, Henry Yung-li. 1997. Voice, case, and agreement in Seediq and Kavalan. Hsinchu: National Tsing Hua University (Ph.D. dissertation).
- Chang, Henry Yung-li. 1999. Seediq bound pronouns: Pronominal clitics or agreement affixes. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected papers from the 8th International Conference on Austronesian Linguistics*, 355–370. Taipei: Academia

- Sinica.
- Chang, Henry Yung-li. 2000. *A reference grammar of Seediq*. (Series on Formosan Languages 6). Taipei: Yuanliu.
- Chang, Henry Yung-li. 2004. AF verbs: Transitive, intransitive, or both? In Ying-jin Lin, Fang-min Hsu, Chun-chih Lee, Jackson T.-S. Sun, Hsiu-fang Yang & Dah-an Ho (eds.), *Studies on Sino-Tibetan languages: Papers in honor of professor Hwang-cherng Gong on his seventieth birthday*, 95–120. Taipei: Academia Sinica.
- Chang, Henry Yung-li. 2006. Rethinking the Tsouic subgroup hypothesis: A morphosyntactic perspective. In Henry Y. Chang, Lillian M. Huang & Dah-an Ho (eds.), *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, 565–583. Taipei: Institute of Linguistics, Academia Sinica.
- Chang, Henry Yung-li. 2009. Adverbial verbs and adverbial compounds in Tsou: A syntactic analysis. *Oceanic Linguistics* 48(2). 439–476.
- Chang, Henry Yung-li. 2010. On the syntax of Formosan adverbial verb constructions. In Raphael Mercado, Eric Potsdam & Lisa Travis (eds.), *Austronesian and theoretical linguistics*, 183–121. Amsterdam: John Benjamins.
- Chang, Henry Yung-li. 2011. Tsou case markers revisited: Visibility, proximity, and beyond. *Diversity of languages: Papers in honor of Professor Feng-fu Tsao on the occasion of his retirement*, 91–123. Taipei: Crane.
- Chang, Henry Yung-li & Amy Pei-jung Lee. 2002. Nominalization in Kavalan. *Language and Linguistics* 3(2). 349–368.
- Chang, Henry Yung-li, Chih-chen Jane Tang & Dah-an Ho. 1998. A study of noun-class markers in Kavalan. *The Tsing Hua Journal of Chinese Studies, New Series* 28(3). 275–298.
- Chang, Melody Ya-yin. 1998. Wh-constructions and the problem of Wh-movement in Tsou. Hsinchu: National Tsing Hua University (M.A. thesis).
- Chang, Melody Ya-yin. 2002. Nominalization in Tsou. *Language and Linguistics* 3(2). 335–348.
- Chang, Melody Ya-yin. 2004. Subjecthood in Tsou. Hsinchu: National Tsing Hua University (Ph.D. dissertation).
- Chang, Mita Yu-Lung. 2007. Negation in Central Amis. Hsinchu: National Hsin-Chu University of Education.
- Chang, Yiyang Ann. 2012. Pronominal ordering in Pngawan Atayal. In Enrico Boone, Kathrin Linke & Maartje Schulp (eds.), *Proceedings of ConSOLE XIX*, 177–190. Leiden: Student Organization of Linguistics in Europe.
- Chen, Cheng-fu. 2008. Aspect and tense in Rukai: Interpretation and interaction. Austin: University of Texas at Austin (Ph.D. dissertation).
- Chen, Irene Shen-Tzu. 2009. Word formation in Takibakha Bunun. Hsinchu: National Hsin-Chu University of Education.
- Chen, Kang & Rongsheng Ma. 1986. *高山族语言简志 (排湾语) [A grammatical sketch of languages of the mountain tribes (Paiwan)]*. Beijing: Minzu Publishing House.
- Chen, Sih-Wei. 2007. Applicative constructions in Atayal. Hsinchu: National Tsing Hua University (M.A. thesis).
- Chen, Teresa M. 1987. *Verbal constructions and verbal classification in Nataoran-Amis*. (Pacific Linguistics C-85). Canberra: Australian National University.
- Chen, Tingchun. 2010. Restructuring in Mayrinax Atayal. Montréal: McGill University B.A. honours thesis.
- Chen, Wen-te. 1989. The study of age-set systems and its theoretical implications: A Taiwanese case. *Bulletin of the Institute of Ethnology, Academia Sinica* 68. 105–144.
- Chen, Youmehim. 2000. Negation in Thao and Tsou. Chiayi: National Chung Cheng

- University.
- Cheng, Yi-Yang. 2013. Modality in Mayrinax Atayal. Taipei: National Taiwan University (M.A. thesis).
- Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs & Peter Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Waltham, MA: Ginn and Company.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*. 2nd ed. Oxford: Basil Blackwell.
- Comrie, Bernard. 1998. Rethinking the typology of relative clauses. *Language Design* 1. 59–85.
- Comrie, Bernard. 2013. Ergativity: Some recurrent themes. In Edith L. Bavin & Sabine Stoll (eds.), *The acquisition of ergativity*, 15–34. Amsterdam: John Benjamins.
- Comrie, Bernard & Zarina Estrada-Fernández (eds.). 2012. Relative clauses in languages of the Americas: A typological overview. (Typological Studies in Language 102). Amsterdam: John Benjamins.
- Comrie, Bernard & Sandra Thompson. 2007. Lexical nominalization. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. 3: Grammatical categories and the lexicon*, 334–381. Cambridge: Cambridge University Press.
- Cooreman, Ann, Barbara A. Fox & Talmy Givón. 1984. The discourse definition of ergativity. *Studies in Language* 8(1). 1–34.
- Croft, William. 1994. Speech act classification, language typology, and cognition. In Savas L. Tsohatzidis (ed.), *Foundations of speech act theory: Philosophical and linguistic perspectives*, 460–490. London: Routledge.
- Croft, William. 1999. Modern syntactic typology. In Masayoshi Shibatani & Theodora Bynon (eds.), *Approaches to language typology*, 85–144. Oxford: Oxford University Press.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William. 2003. *Typology and universals*. 2nd ed. Cambridge: Cambridge University Press.
- Croft, William. 2007. Typology and linguistic theory in the past decade: A personal view. *Linguistic Typology* 11(1). 79–91.
- Dahl, Östen & Maria Koptjevskaja-Tamm. 2001. Kinship in grammar. In Irène Baron, Michael Herslund & Finn Sørensen (eds.), *Dimensions of possession*, 201–225. Amsterdam: John Benjamins.
- Dahl, Otto Christian. 1973. *Proto-Austronesian*. (Scandinavian Institute of Asian Studies Monograph Series 15). Lund: Studentlitteratur.
- Daniel, Michael & Edith Moravcsik. 2013. The associative plural. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/36>.
- Deng, An-Ting. 2014. Interrogative constructions in Kanakanavu. Taipei: National Taiwan University (M.A. thesis).
- Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Drude, Sebastian. 2011. Nominalization as a possible source for subordination in Awetí. *Amérindia* 35. 189–218.
- Dryer, Matthew S. 2004. Noun phrases without nouns. *Functions of Language* 1(1). 43–76.
- Dryer, Matthew S. 2007. Noun phrase structure. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. 2: Complex constructions*, 151–205. Cambridge: Cambridge University Press.
- Dyen, Isidore. 1962a. The position of the Malayopolynesian languages of Formosa. *Asian Perspectives* 7(1–2). 261–271.
- Dyen, Isidore. 1962b. The lexicostatistical classification of the Malayopolynesian

- languages. *Language* 38(1). 38–46.
- Dyen, Isidore. 1965. *A lexicostatistical classification of the Austronesian languages*. Baltimore: Wavely Press.
- Dyen, Isidore. 1974. The Proto-Austronesian enclitic genitive pronouns. *Oceanic linguistics. Papers of the first international conference on comparative Austronesian linguistics, 1974: Proto-Austronesian and Western Austronesian* 13(1/2). 17–31.
- Edmondson, Jerold A., John H. Esling, Jimmy G. Harris & Tung-chiou Huang. 2005. A laryngoscopic study of glottal and epiglottal/pharyngeal stop and continuant articulations in Amis: An Austronesian language of Taiwan. *Language and Linguistics* 6(3). 381–396.
- Elizabeth, Zeitoun & Stacy Fang-ching Teng. 2009. From ki-N “get N” in Formosan languages to ki-V “get V-ed” (passive) in Rukai, Paiwan and Puyuma. In Bethwyn Evans (ed.), *Discovering history through language: Papers in honour of Malcolm Ross*, 479–500. Canberra: Australian National University.
- Embick, David & Rolf Noyer. 1999. Locality in post-syntactic operations. *MIT Working Papers in Linguistics* 34. 265–317.
- Farrell, Patrick. 2005. *Grammatical relations*. Oxford: Oxford University Press.
- Ferrell, Raleigh. 1969. *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. (Institute of Ethnology Monograph 17). Taipei: Academia Sinica.
- Fey, Virginia A. 1986. *Amis dictionary*. Taipei: The Bible Society.
- Foley, William A. 2014. A comparative look at nominalizations in Austronesian. In I Wayan Arka & N. L. K. Mas Indrawati (eds.), *Argument realisations and related constructions in Austronesian languages: Papers from 12-ICAL, vol.2*, 1–51. Canberra: Australian National University.
- Foley, William Auguste. 1976. Comparative syntax in Austronesian. Berkeley: University of California Berkeley (Ph.D. dissertation).
- Frajzyngier, Zygmunt & Erin Shay. 2002. *A grammar of Hdi*. (Mouton Grammar Library 21). Berlin: Mouton de Gruyter.
- Gildea, Spike. 1998. *One reconstructing grammar: Comparative Cariban morphosyntax*. Oxford: Oxford University Press.
- Givón, Talmy. 2001. *Syntax: An introduction, vol.2*. Amsterdam: John Benjamins.
- Givón, Talmy. 2012. Toward a diachronic typology of relative clause. In Bernard Comrie & Zarina Estrada-Fernández (eds.), *Relative clauses in languages of the Americas: A typological overview*, 3–25. Amsterdam: John Benjamins.
- Göksel, Asli & Barış Kabak. 2012. Covert modality and context shift: The “perplexive” construction. In Werner Abraham & Elisabeth Leiss (eds.), *Covert patterns of modality*, 106–132. Newcastle upon Tyne: Cambridge Scholars Publishing.
- González, Albert Álvarez. 2012. Relative clauses and nominalizations in Yaqui. In Bernard Comrie & Zarina Estrada-Fernández (eds.), *Relative clauses in languages of the Americas: A typological overview*, 67–95. Amsterdam: John Benjamins.
- Gray, R. D., A. J. Drummond & S. T. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323(5913). 479–483.
- Halliday, M. A. K. & Christian M. I. M. Matthiessen. 2004. *An introduction to functional grammar*. 3rd ed. London: Hodder Arnold.
- Halpern, Aaron L. 2001. Clitics. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*. Oxford: Wiley.
- Hamawand, Zeki. 2011. *Morphology in English: Word formation in Cognitive Grammar*. London: Continuum International Publishing Group.
- Handschuh, Corinna. 2014. *A typology of marked-S languages*. (Studies in Diversity Linguistics 1). Berlin: Language Science Press.
- Haspelmath, Martin. 1989. From purposive to infinitive: A universal path of grammaticization. *Folia Linguistica Historica* 10(1–2). 287–310.
- Haspelmath, Martin. 2002. *Understanding morphology*. London: Arnold.

- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.
- Haspelmath, Martin. 2012. How to compare major word-classes across the world's languages. *UCLA Working Papers in Linguistics, Theories of Everything* 17. 109–130.
- Haspelmath, Martin. 2013. Argument indexing: A conceptual framework for the syntactic status of bound person forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 197–226. Berlin: Mouton de Gruyter.
- Haspelmath, Martin. 2014. The three adnominal possessive constructions in Egyptian-Coptic: Three degrees of grammaticalization. In Martin Haspelmath, Eitan Grossman & Tonio Sebastian Richter (eds.), *Egyptian-Coptic linguistics in typological perspective*, 103–144. Berlin: Mouton de Gruyter.
- Haudricourt, André G. 1965. Problems of Austronesian comparative philology. *Lingua* 14. 315–329.
- Hayung, Tali'. 2008. 尖石泰雅語詞綴研究[Affixes in Jianshi Squiliq Atayal]. Hsinchu: National Hsinchu University of Education (M.A. thesis).
- He, Rufen, Siqi Zeng, Citamih & Dengxian Lin. 1986. 高山族語言簡志(阿眉斯語) [A grammatical sketch of languages of the mountain tribes (Amis)]. Beijing: Minzu Publishing House.
- He, Rufen, Siqi Zeng, Wensu Li & Qingchun Lin. 1986. 高山族語言簡志(布嫩語)[A grammatical sketch of languages of the mountain tribes (Bunun)]. Beijing: Minzu Publishing House.
- Heath, Jeffre. 1999. *A grammar of Koyra Chiini: The Songhay of Timbuktu*. (Mouton Grammar Library 19). Berlin: Mouton de Gruyter.
- Heine, Bernd. 1997. *Possession: Cognitive sources, forces, and grammaticalization*. (Cambridge Studies in Linguistics 83). Cambridge: Cambridge University Press.
- Heine, Bernd. 2003. Grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 575–601. Malden: Blackwell Publishing.
- Heyvaert, Liesbet. 2003. *A cognitive-functional approach to nominalization in English*. (Cognitive Linguistic Research 26). Berlin: Mouton de Gruyter.
- Heyvaert, Liesbet. 2010. A cognitive-functional perspective on deverbal nominalization in English: Descriptive findings and theoretical ramifications. In Monika Rathert & Artemis Alexiadou (eds.), *The semantics of nominalizations across languages and frameworks*, 51–81. Berlin: Mouton de Gruyter.
- Himmelman, Nikolaus P. 2002. Voice in western Austronesian: An update. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 7–16. Canberra: Australian National University.
- Himmelman, Nikolaus P. 2004. Tagalog (Austronesian). In Geert E. Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphology: An international handbook on inflection and word-formation*, vol. 2, 1473–1490. Berlin: Walter de Gruyter.
- Himmelman, Nikolaus P. 2005a. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Alexander Adelaar & Nikolaus P. Himmelman (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. London: Routledge.
- Himmelman, Nikolaus P. 2005b. Tagalog. In Alexander Adelaar & Nikolaus P. Himmelman (eds.), *The Austronesian languages of Asia and Madagascar*, 350–376. London: Routledge.
- Himmelman, Nikolaus P. 2008. Lexical categories and voice in Tagalog. In Peter K. Austin & Simon Musgrave (eds.), *Voice and grammatical functions in Austronesian languages*, 247–293. Stanford: CSLI.
- Ho, Dah-an. 1997. The Kakanavu language. In Paul Jen-kuei Li (ed.), *The*

- Austronesian languages of Kaohsiung County*, 228–271. Kaohsiung: Kaohsiung County Government.
- Holmer, Arthur. 1996. *A parametric grammar of Seediq*. (Travaux de l'Institut de Linguistique de Lund 30). Lund: Lund University Press.
- Holmer, Arthur. 2002. The morphology and syntax of Seediq focus. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 333–354. Canberra: Australian National University.
- Holmer, Arthur. 2006. Seediq: Adverbial heads in a Formosan language. In Hans-Martin Gärtner, Paul S. Law & Joachim Sabel (eds.), *Clause structure and adjuncts in Austronesian languages*, 83–124. Berlin: Mouton de Gruyter.
- Holmer, Arthur & Loren Billings. 2014. Clitic pronouns in Seediq. In I Wayan Arka & N. L. K. Mas Indrawati (eds.), *Argument realisations and related constructions in Austronesian languages: Papers from 12-ICAL*, vol. 2, 111–139. Canberra: Australian National University.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
- Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in Universal Grammar. *Language* 60(4). 703–752.
- Hsieh, Chin-Fu. 2011. Comparative constructions in Isbukun Bunun. Kaohsiung: National Kaohsiung Normal University (M.A. thesis).
- Hsieh, Fuhui. 2011. The functions of -an and =ay in Kavalan. In Foong Ha Yap, Karen Grunow-Härsta & Janick Wrona (eds.), *Nominalization in Asian languages: Diachronic and typological perspectives*, 499–522. Amsterdam: John Benjamins.
- Hsieh, Fuhui & Shuanfan Huang. 2006. The pragmatics of case marking in Saisiyat. *Oceanic Linguistics* 45(1). 91–109.
- Hsieh, Fuhui & Shuanfan Huang. 2007. Documenting and revitalizing Kavalan. In Der-Hwa V. Rau & Margaret Florey (eds.), *Documenting and revitalizing Austronesian languages*, 93–110. Honolulu: University of Hawai'i Press.
- Hsu, Lowking Wei-cheng. 2008. Word Formation in Truku. Hsinchu: National Hsinchu University of Education (M.A. thesis).
- Huang, Huei-ju. 2003. The tense, aspect, and reality in Tsou and Saisiyat. Taipei: National Taiwan University (M.A. thesis).
- Huang, Huei-ju. 2010. The syntax and pragmatics of clausal constituents in Tsou discourse. Taipei: National Taiwan University (Ph.D. dissertation).
- Huang, Hui-Chuan J. 2006. Resolving vowel clusters: A comparison of Isbukun Bunun and Sqliq Atayal. *Language and Linguistics* 7(1). 1–26.
- Huang, Lillian M. 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3(2). 197–225.
- Huang, Lillian M. 1993. *A study of Atayal syntax*. Taipei: The Crane Publishing.
- Huang, Lillian M. 1994. Ergativity in Atayal. *Oceanic Linguistics* 33(1). University of Hawai'i Press. 129–143.
- Huang, Lillian M. 1995a. *A study of Mayrinax syntax*. Taipei: The Crane Publishing.
- Huang, Lillian M. 1995b. The case markers and pronominal system in Amis. *The Journal of National Chengchi University* 70. 217–258.
- Huang, Lillian M. 1997. The Bunun language in Kaohsiung County. In Paul Jen-kuei Li (ed.), *The Austronesian languages of Kaohsiung County*, 351–409. Kaohsiung: Kaohsiung County Government.
- Huang, Lillian M. 2000a. *A reference grammar of Atayal*. (Series on Formosan Languages 1). Taipei: Yuanliu.
- Huang, Lillian M. 2000b. *A reference grammar of Puyuma*. (Series on Formosan Languages 10). Taipei: Yuanliu.
- Huang, Lillian M. 2001. Focus system of Mayrinax Atayal: A syntactic, semantic and pragmatic perspective. *Journal of Taiwan Normal University: Humanities & Social*

- Science* 46. 51–69.
- Huang, Lillian M. 2006. Case marking system in Pngawan Atayal. In Henry Y. Chang, Lillian M. Huang & Dah-an Ho (eds.), *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, 205–238. Taipei: Academia Sinica.
- Huang, Lillian M. & Tali' Hayung. 2011. Coordination and comitativity in Squliq Atayal. *Language and Linguistics* 12(1). 1–47.
- Huang, Lillian M., Marie M. Yeh, Elizabeth Zeitoun, Anna Hsiou-chuan Chang & Jing-lan Joy Wu. 1998. A typological overview of nominal case marking systems of some Formosan languages. In Shuanfan Huang (ed.), *Selected papers from the second International Symposium on Languages in Taiwan*, 21–48. Taipei: Crane.
- Huang, Lillian M., Elizabeth Zeitoun, Marie Mei-li Yeh, Anna Hsiou-chuan Chang & Jing-lan Joy Wu. 1999. A typological overview of pronominal systems of some Formosan languages. In Hsu Samuel Wang, Feng-fu Tsao & Chin-fa Lien (eds.), *Selected papers from the fifth International Conference on Chinese Linguistics*, 165–198. Taipei: Crane.
- Huang, Shuanfan. 2005. Split O in F Formosan languages: A localist interpretation. *Language and Linguistics* 6(4). 783–806.
- Huang, Shuanfan & Huei-ju Huang. 2003. On the status of reality marking in Tsou. *Taiwan Journal of Linguistics* 1(2). 1–34.
- Huang, Shuanfan, Lily I-wen Su & Li-may Sung. 2001. *A functional grammar of Tsou*. National Science Council Annual Report (NSC 89-2411-H-002-092-M7).
- Huang, Shuping. 2008. Analysis of polysemy and its applications to corpus documentation: A study ased on Saisiyat. Taipei: National Taiwan University (Ph.D. dissertation).
- Huang, Wei-chen. 2012. A study of verbal morphology in Puljetji Paiwan. Hsinchu: National Tsing Hua University.
- Huang, Ya-jun. 1988. Amis verb cassification. Taipei: Fu Jen Catholic University (M.A. thesis).
- Hull, Geoffrey & Lance Eccles. 2001. *Tetum reference grammar*. Winston Hills: Sebastião Aparício da Silva Project.
- Hyslop, Catriona. 2001. *The Lolovoli dialect of the North-East Ambae language, Vanuatu*. (Pacific Linguistics 515). Canberra: Australian National University.
- Ismahasan, Dahu, Wen-shu Li & Bukun Ismahasan. 1998. *Dainasbae tus'a tu buan [The moon that walks through space and time]*. (Taiwan Aborigines Series 27). Taichung: Morning Star.
- Jiang, Haowen. 2006. Spatial conceptualizations in Kavalan. Taipei: National Taiwan University (M.A. thesis).
- Jiang, Haowen. 2009. Spatial deixis as motion predicates and aspect markers: The case in Kavalan. Presented at the 11th International Conference on Austronesian Linguistics (ICAL-11), Aussois, France, Jun 22-26.
- Jiang, Haowen. 2011. On the grammaticalization of nominalization marker =ay in Kavalan and Amis: A contrastive study. *LSA Annual Meeting Extended Abstracts (ExtAbs)* 2. 19:1-5.
- Jiang, Haowen. 2012. Demonstratives and nominalization: On the differential scope of the two sets of demonstratives in Isbukun Bunun. Presented at the 19th Annual Meeting of the Austronesian Formal Linguistics Association (AFLA-19), Academia Sinica, Taipei, Taiwan, June 26-30.
- Jiang, Haowen. 2013. The possession-modification scale and a reappraisal of “genitive” in Paiwan, Rukai, and Puyuma. Presented at the 2nd NTU Workshop on Discourse and Grammar in Formosan Languages, Graduate Institute of Linguistics, National Taiwan University, January 1.
- Jiang, Haowen & Loren Billings. 2014. Case closed: Positionally conditioned pronoun

- allomorphy (+ an inverse clitic) in Rikavung Puyuma. In Juhee Lee (ed.), *Proceedings of the 5th international conference on phonology and morphology (ICPM-5)*, 132–135. Gwangju: The Phonology-Morphology Circle of Korea.
- Jiang, Haowen & Loren Billings. 2015. Person-based ordering of pronominal clitics in Rikavung Puyuma: An inverse analysis. In Amber Camp, Yuko Otsuka, Claire Stabile & Nozomi Tanaka (eds.), *AFLA 21: The proceedings of the 21st meeting of the Austronesian Formal Linguistics Association*, 87–106. Canberra: Australian National University.
- Jiang, Lifang. 2009. 阿美语偏正短语研究 [Modifier-head phrases in Amis]. Beijing: Central University for Nationalities (Ph.D. dissertation).
- Jiang, Lifang. 2012. 阿美语名物化的途径 [Nominalization devices in the Amis language]. *Minzuyuwen* 2. 64–67.
- Jiang, Yijuan. 2012. A sketch grammar of Takibakha Bunun. Hsinchu: National Tsing Hua University (M.A. thesis).
- Johns, Alana. 1992. Deriving ergativity. *Linguistic Inquiry* 23(1). 57–87.
- Kalmár, Ivan. 1979. The antipassive and grammatical relations in Eskimo. In Frans Plank (ed.), *Ergativity: Towards a theory of grammatical relations*, 117–144. London: Academic Press.
- Kaufman, Daniel. 2009a. Austronesian typology and the nominalist hypothesis. In Alexander Adelaar & Andrew Pawley (eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, 197–226. Canberra: Australian National University.
- Kaufman, Daniel. 2009b. Austronesian nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35(1). 1–49.
- Kaufman, Daniel. 2011. Exclamatives and temporal nominalizations in Austronesian. In Foong Ha Yap, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in Asian languages: Diachronic and typological perspectives*, 721–753. Amsterdam: John Benjamins.
- Kaybaybaw, Ya'aw Kalahae'. 2009. A morphological and semantic study on word formation in SaySiyat. Hsinchu: National Hsinchu University of Education (M.A. thesis).
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam: John Benjamins.
- Kikusawa, Ritsuko & Lawrence A. Reid. 2003. A Talubin text with a wordlist and grammatical notes. *Journal of Asian and African Studies* 65. 89–148.
- Klaiman, M. H. 1991. *Grammatical voice*. (Cambridge Studies in Linguistics 59). Cambridge: Cambridge University Press.
- Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 61(1). 95–120.
- Koptjevskaja-Tamm, Maria. 1995. Possessive and relational forms in Chukchi. In Frans Plank (ed.), *Double case: Agreement by Suffixaufnahme*, 301–321. Oxford: Oxford University Press.
- Koptjevskaja-Tamm, Maria. 2001. Adnominal possession. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, vol. 2, 960–970. Berlin: Walter de Gruyter.
- Koptjevskaja-Tamm, Maria. 2006. Nominalization. In Keith Brown (ed.), *Encyclopedia of languages and linguistics*, vol. 8, 652–659. 2nd ed. Oxford: Elsevier.
- Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI.
- Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press.
- Kuo, Ching-hua. 1979. Budai complementation. Taipei: Fu Jen Catholic University (M.A. thesis).
- Lander, Yury A. 2009. Varieties of genitive. In Andrej Malchukov & Andrew Spencer

- (eds.), *The Oxford handbook of case*, 581–592. Oxford: Oxford University Press.
- Langacker, Ronald W. 1993. Reference-point constructions. *Cognitive Linguistics* 4(1). 1–38.
- Langacker, Ronald W. 1999. *Grammar and conceptualization*. (Cognitive Linguistics Research 14). Berlin: Mouton de Gruyter.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.
- Langacker, Ronald W. 2009a. Metonymic grammar. In Klaus-Uwe Panther, Linda L. Thornburg & Antonio Barcelona (eds.), *Metonymy and metaphor in grammar*, 45–71. Amsterdam: John Benjamins.
- Langacker, Ronald W. 2009b. *Investigations in Cognitive Grammar*. (Cognitive Linguistics Research 42). Berlin: Mouton de Gruyter.
- Lee, Amy Pei-jung. 1997. The case-marking and focus systems in Kavalan. Hsinchu: National Tsing Hua University (M.A. thesis).
- Lee, Ting-wu Terry. 2010. On relative clause constructions in Saisiyat. Taipei: National Taiwan Normal University (M.A. thesis).
- Lees, Robert B. 1963. *The grammar of English nominalizations*. Bloomington: Indiana University.
- Lehmann, Christian. 1988. Towards a typology of clause linkage. In John Haiman & Sandra Thompson (eds.), *Clause combining in grammar and discourse*, 181–225. (Typological Studies in Language 18). Amsterdam: John Benjamins.
- Lehmann, Christian. 2002. *Thoughts on grammaticalization*. 2nd ed. (Arbeitspapiere Des Seminars Für Sprachwissenschaft Der Universität Erfurt (ASSidUE) 9). Erfurt: Universität Erfurt.
- Lehmann, Christian, Yong-Min Shin & Elisabeth Verhoeven. 2004. *Person prominence and relation prominence: On the typology of syntactic relations with particular reference to Yucatec Maya*. 2nd ed. (Arbeitspapiere Des Seminars Für Sprachwissenschaft Der Universität Erfurt (ASSidUE) 12). Erfurt: Universität Erfurt.
- Lemaréchal, Alain. 1982. Sémantisme des parties du discours et sémantisme des relations. *Bulletin de la Société de Linguistique de Paris* 77(1). 1–39.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
- Li, Lilian Li-ying. 2010. Clitics in Nantou Isbukun Bunun (Austronesian). Nantou: National Chi Nan University (M.A. thesis).
- Li, Paul Jen-kuei. 1973. Rukai structure. Honolulu: University of Hawai‘i at Mānoa (Ph.D. dissertation).
- Li, Paul Jen-kuei. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History & Philology* 48. 1–92.
- Li, Paul Jen-kuei. 1978. The case-marking systems of the four less-known Formosan languages. *Proceedings of the 2nd International Conference on Austronesian Linguistics, Fascicle 1*, 569–615. Canberra: Australian National University.
- Li, Paul Jen-kuei. 1979. Variations in the Tsou dialects. *Bulletin of the Institute of History & Philology* 50(2). 273–300.
- Li, Paul Jen-kuei. 1988. A comparative study of Bunun dialects. *Bulletin of the Institute of History & Philology* 59(2). 479–508.
- Li, Paul Jen-kuei. 1996. The pronominal systems in Rukai. In Bern Nothofer (ed.), *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*, 210–230. Hamburg: Abera Verlag.
- Li, Paul Jen-kuei. 1997a. A syntactic typology of Formosan languages: Case markers on nouns and pronouns. In Tseng Chiu-yu (ed.), *Chinese languages and linguistics IV: Typological studies of languages in China*, 343–378. Taipei: Academia Sinica.
- Li, Paul Jen-kuei. 1997b. Southern Bunun of Hsinyi Township, Nantou County. In Paul

- Jen-kuei Li (ed.), *The Austronesian languages of Kaohsiung County*, 300–350. Kaohsiung: Kaohsiung County Government.
- Li, Paul Jen-kuei. 2000. Some aspects of Pazeh syntax. In Videia P. de Guzman & Byron Bender (eds.), *Grammatical analysis: Morphology, syntax, and semantics. Studies in honor of Stanley Starosta*, 89–108. Honolulu: University of Hawai'i Press.
- Li, Paul Jen-kuei. 2001. The dispersal of the Formosan aborigines in Taiwan. *Language and Linguistics* 2(1). 271–278.
- Li, Paul Jen-kuei. 2002. Nominalization in Pazih. *Language and Linguistics* 3(2). 227–239.
- Li, Paul Jen-kuei. 2006. Introduction. In Paul Jen-kuei Li & Shigeru Tsuchida (eds.), *Kavalan dictionary*, 1–62. (Language and Linguistics Monograph Series A-19). Taipei: Academia Sinica.
- Li, Paul Jen-kuei. 2008a. The great diversity of Formosan languages. *Language and Linguistics* 9(3). 523–546.
- Li, Paul Jen-kuei. 2008b. Time perspective of Formosan aborigines. *Past human migrations in East Asia: Matching archaeology, linguistics and genetics*, 211–218. London: Routledge.
- Li, Paul Jen-kuei. 2011. *Thao texts and songs*. (Language and Linguistics Monograph Series 44). Taipei: Academia Sinica.
- Liao, Hsiu-chuan. 2002. The interpretation of tu and Kavalan ergativity. *Oceanic Linguistics* 41(1). 140–158.
- Liao, Hsiu-chuan. 2004. Transitivity and ergativity in Formosan and Philippine languages. Honolulu: University of Hawai'i at Mānoa (Ph.D. dissertation).
- Liao, Hsiu-chuan. 2005a. Pronominal forms in central Cagayan Agta: Clitics or agreement features? In Hsiu-chuan Liao & Carl R. Galvez Rubino (eds.), *Current issues in Philippine linguistics and anthropology: Parangal kay Lawrence A. Reid*, 346–363. Manila: Linguistic Society of the Philippines and SIL Philippines.
- Liao, Hsiu-chuan. 2005b. Another look at the order of clitic pronouns in Wulai Atayal. *Concentric: Studies in Linguistics* 31(1). 47–63.
- Lichtenberk, Frantisek. 2000. Inclusive pronominals. *Oceanic Linguistics* 39(1). 1–32.
- Lichtenberk, Frantisek. 2005. Inalienability and possessum individuation. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic diversity and language theories*, 339–362. Amsterdam: John Benjamins.
- Lichtenberk, Frantisek. 2009. Attributive possessive constructions in Oceanic. In William B. McGregor (ed.), *The expression of possession*, 249–292. Berlin: Mouton de Gruyter.
- Lin, Chiao-Chun Beryl. 2005. Interrogatives in Sqliq Atayal. Hsinchu: National Tsing Hua University (M.A. thesis).
- Lin, Gujing. 2010. Argument structure of Tsou: Simplex and complex predicates. Houston: Rice University (Ph.D. dissertation).
- Lin, Michelle Hsueh-o. 1995. Two Amis suffixes -ay and -an. *Studies in English Literature and Linguistics* 21. 159–173.
- Lin, Shu-yi Mars. 2011. Reconstructing negative morphemes in Proto-Austronesian: Evidence from Formosan languages. National Taiwan Normal University M.A. thesis.
- Liu, Adlay Kun-long. 2005. The structure of relative clauses in Jianshi Sqliq Atayal. *Concentric: Studies in Linguistics* 31(2). 89–110.
- Liu, Chong-yu Harry. 2008. Echo vowels in Budai Rukai. Hsinchu: National Tsing Hua University (M.A. thesis).
- Liu, Dorinda Tsai-hsiu. 1999. Cleft constructions in Amis. Taipei: National Taiwan University (M.A. thesis).
- Liu, Dorinda Tsai-hsiu. 2011. Complementation in three Formosan languages: Amis, Mayrinax Atayal, and Tsou. Honolulu: University of Hawai'i at Mānoa (Ph.D. dissertation).

- dissertation).
- Liu, Dorinda Tsai-hsiu. 2014. Neutral and imperfective forms in Kanakanavu. In I Wayan Arka & N. L. K. Mas Indrawati (eds.), *Argument realisations and related constructions in Austronesian languages: Papers from 12-ICAL*, vol. 2, 175–195. Canberra: Australian National University.
- Liu, Emma En-hsing. 2003. Conjunction and modification in Amis. Hsinchu: National Tsing Hua University (M.A. thesis).
- Liu, Hsing-Chen. 2014. A semantic and discourse study on the voice system in Kanakanavu. Taipei: National Taiwan University (M.A. thesis).
- Lopez, Cecilio. 1941. *A manual of the Philippine national language*. Manila: Bureau of Printing.
- Maas, Utz. 2004. Finite and nonfinite from a typological perspective. *Linguistics* 42(2). 359–385.
- Martínez Fabián, Constantino & Terence D. Langendoen. 1996. Sobre las llamadas cláusulas relativas en yaqui. In Z. Estrada Fernández, M. Figueroa Esteva & G. López Cruz (eds.), *Memorias del Tercer Encuentro de Lingüística en el Noroeste, Tomo 1: Lenguas Indígenas*, vol. 2, 443–464. Hermosillo: Editorial Unison.
- Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Press.
- McGregor, Ronald Stuart. 1995. *Outline of Hindi grammar: With exercises*. 3rd ed. Oxford: Oxford University Press.
- McGregor, William B. (ed.). 2009. *The expression of possession*. (The Expression of Cognitive Categories 2). Berlin: Mouton de Gruyter.
- Mei, Kuang. 1982. Pronouns and verb inflection in Kanakanavu. *The Tsing Hua Journal of Chinese Studies, New Series* 14. 207–32.
- Miestamo, Matti. 2005. *Standard negation: The negation of declarative verbal main clauses in a typological perspective*. (Empirical Approaches to Language Typology 31). Berlin: Walter de Gruyter.
- Miestamo, Matti. 2013. Symmetric and asymmetric standard negation. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/113>.
- Milsark, Gary. 2006. Gerundive nominalizations. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. 2, 436–459. Oxford: Wiley-Blackwell.
- Moravcsik, Edith A. 1995. Summing up Suffixaufnahme. In Frans Plank (ed.), *Double case: Agreement by Suffixaufnahme*, 451–484. Oxford: Oxford University Press.
- Moseley, Christopher (ed.). 2010. *Atlas of the world's languages in danger*. Paris: UNESCO. <http://www.unesco.org/culture/en/endangeredlanguages/atlas>.
- Nagaya, Naonori. 2007. Information structure and constituent order in Tagalog. *Language and Linguistics* 8(1). 343–372.
- Nagaya, Naonori. 2011. The Lamaholot language of Eastern Indonesia. Houston: Rice University (Ph.D. dissertation).
- Nichols, Johanna & Balthasar Bickel. 2013. Locus of marking in possessive noun phrase. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/24>.
- Noonan, Michael. 2008a. Nominalization in Bodic languages. In Marí José López-Couso & Elena Seoane (eds.), *Rethinking grammaticalization: New perspectives*, 219–237. Amsterdam: John Benjamins.
- Noonan, Michael. 2008b. Case compounding in the Bodic languages. In Greville G. Corbett & Michael Noonan (eds.), *Case and grammatical relations: Studies in honor of Bernard Comrie*, 127–166. Amsterdam: John Benjamins.
- Noonan, Michael P. 1992. *A grammar of Lango*. (Mouton Grammar Library 7). Berlin:

- Mouton de Gruyter.
- Ochiai, Izumi. 2009. A study of Seediq pronouns. Taipei: National Taiwan University (M.A. thesis).
- Pan, Chia-jung. 2012. A grammar of lha'alua, an Austronesian language of Taiwan. Cairns: James Cook University (Ph.D. dissertation).
- Partee, Barbara H. & Vladimir Borschev. 2001. Some puzzles of predicate possessives. In István Kenesei & Robert M. Harnish (eds.), *Perspectives on semantics, pragmatics, and discourse: A Festschrift for Ferenc Kiefer*, 91–117. Amsterdam: John Benjamins.
- Peterson, David A. 2007. *Applicative constructions*. Oxford: Oxford University Press.
- Quakenbush, Stephen J. 2005. Philippine linguistics from an SIL perspective: Trends and prospects. In Hsiu-chuan Liao & Carl R. Galvez Rubino (eds.), *Current issues in Philippine linguistics and anthropology: Parangal kay Lawrence A. Reid*, 3–27. Manila: Linguistic Society of the Philippines and SIL Philippines.
- Radetzky, Paula Kadose. 2004. Grammaticalisation d'un marqueur de définitude en saaroa. *Fait de langues: Revue de linguistique. N. 23-24: Les langues austronésiennes*. 213–230.
- Rau, D. Victoria & Margaret Florey (eds.). 2007. *Documenting and revitalizing Austronesian languages*. (Language Documentation & Conservation Special Publication 1). Honolulu: University of Hawai'i Press.
- Reid, Lawrence A. 1978. Problems in the reconstruction of Proto-Philippine construction markers. In Stephen A. Wurm & Lois Carrington (eds.), *Second International Conference on Austronesian Linguistics: Proceedings, Fascicle I, Western Austronesian*, 33–66. Canberra: Australian National University.
- Reid, Lawrence A. 1981. Proto-Austronesian genitive determiners. In Andrew Gonzalez & David Thomas (eds.), *Linguistics across continents: Studies in honor of Richard S. Pittman*, 97–105. Manila: Summer Institute of Linguistics and Linguistic Society of the Philippines.
- Reid, Lawrence A. 2002. Determiners, nouns, or what? Problems in the analysis of some commonly occurring forms in Philippine languages. *Oceanic Linguistics* 41(2). 295–309.
- Reid, Lawrence A. 2005. Tagalog and Philippine Languages. In Philipp Strazny (ed.), *Encyclopedia of linguistics*, vol. 2, 1–5. New York: Routledge.
- Reid, Lawrence A. 2007. Another look at the marking of plural personal noun constructions in Austronesian languages. *Oceanic Linguistics* 46(1). 232–252.
- Reid, Lawrence A. & Hsiu-chuan Liao. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5(2). 433–490.
- Richards, Norvin. 2009. Nouns, verbs, and hidden structure in Tagalog. *Theoretical Linguistics* 35(1). 139–152.
- Rissanen, Matti. 1997. The pronominalization of one. In Matti Rissanen, Merja Kytö & Kirsi Heikkonen (eds.), *Grammaticalization at work: Studies of long-term developments in English*, 87–144. Berlin: Walter de Gruyter.
- Ross, John Robert. 1973. Nouniness. In Osamu Fujimura (ed.), *Three dimensions of linguistic theory*, 137–257. Tokyo: TEC.
- Ross, Malcolm. 1995. Reconstructing Proto Austronesian verbal morphology: Evidence from Taiwan. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang & Cheng-hwa Tsang (eds.), *Austronesian studies relating to Taiwan*, 727–791. Taipei: Institute of History and Philology, Academia Sinica.
- Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice-marking. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 17–62. Canberra: Australian National University.
- Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of

- Proto Austronesian. In Henry Y. Chang, Lillian M. Huang & Dah-an Ho (eds.), *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, 521–564. Taipei: Academia Sinica.
- Ross, Malcolm. 2009. Proto Austronesian verbal morphology: A reappraisal. In Alexander Adelaar & Andrew Pawley (eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, 295–326. Canberra: Australian National University.
- Ross, Malcolm. 2012. In defense of nuclear Austronesian (and against Tsouic). *Language and Linguistics* 13(6). 1253–1330.
- Ross, Malcolm. 2013. *How Rukai became accusatively aligned*. Unpublished draft.
- Ross, Malcolm. 2015a. The argument indexing of early Austronesian verbs: A reconstructional myth? In Dag T. T. Haug (ed.), *Historical Linguistics 2013*, 257–280. Amsterdam: John Benjamins.
- Ross, Malcolm. 2015b. Revising the reconstruction of early Austronesian personal pronouns. In Malcolm D. Ross & I Wayan Arka (eds.), *Language change in Austronesian languages: Papers from I2-ICAL, vol. 3*, 109–121. Canberra: Australian National University.
- Ross, Malcolm. 2015c. Reconstructing Proto Austronesian verb classes. *Language and Linguistics* 16(3). 279–345.
- Ross, Malcolm & Stacy Fang-ching Teng. 2005a. Formosan languages and linguistic typology. *Language and Linguistics* 6(4). 739–781.
- Ross, Malcolm & Stacy Fang-ching Teng. 2005b. Clause constructions in Nanwang Puyuma. *Concentric: Studies in Linguistics* 31(1). 119–158.
- Sabbagh, Joseph. 2009. The category of predicates and predicate phrases in Tagalog: Commentary on Kaufman’s article. *Theoretical Linguistics* 35(1). 153–165.
- Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2). 411–444.
- Sagart, Laurent. 2010. Is Puyuma a primary branch of Austronesian? *Oceanic Linguistics* 49(1). 194–204.
- Sagart, Laurent. 2013. Is Puyuma a primary branch of Austronesian? A rejoinder. *Oceanic Linguistics* 52(2). 481–492.
- Saillard, Claire. 1995. Is Maga accusative or ergative? Evidence from case marking. *Papers from the First International Symposium on Languages in Taiwan (ISLT-I)*, 59–72. Taipei: Crane.
- Sasse, Hans Jürgen. 2009. Nominalism in Austronesian: A historical typological perspective. Comments on Daniel Kaufman’s “Austronesian nominalism and its consequences.” *Theoretical Linguistics* 35(1). 167–181.
- Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic or none of the above. In Charles Li (ed.), *Subject and topic*, 491–518. New York: Academic Press.
- Schachter, Paul & Fe T. Otanes. 1972. *Tagalog reference grammar*. (California Library Reprint Series 1983). Berkeley, CA: University of California Press.
- Schachter, Paul & Lawrence A. Reid. 2009. Tagalog. In Bernard Comrie (ed.), *The world’s major languages*, 833–855. 2nd ed. London: Routledge.
- Schachter, Paul & Timothy Shopen. 2007. Parts-of-speech systems. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. 1: Clause structure*, 1–60. Cambridge: Cambridge University Press.
- Schuessler, Axel. 2007. *ABC etymological dictionary of Old Chinese*. Honolulu: University of Hawai‘i Press.
- Shi, Chaokai. 2009. The linker tu in Isbukun Bunun. Kaohsiung: National Kaohsiung Normal University (M.A. thesis).
- Shibatani, Masayoshi. 1988. Voice in Philippine languages. In Masayoshi Shibatani (ed.), *Passive and voice*, 85–142. Amsterdam: John Benjamins.

- Shibatani, Masayoshi. 1991. Grammaticalization of topic into subject. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization, vol.2: Focus on types of grammatical markers*, 93–134. Amsterdam: John Benjamins.
- Shibatani, Masayoshi. 2008. Relativization in Sasak and Sumbawa, Eastern Indonesia. *Language and Linguistics* 9(4). 865–916.
- Shibatani, Masayoshi. 2009. Elements of complex structures, where recursion isn't: the case of relativization. In Talmy Givón & Masayoshi Shibatani (eds.), *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution*, 163–198. Amsterdam: John Benjamins.
- Shibatani, Masayoshi. 2010. On the typology of relative clauses: A nominalization-based counterproposal. Presented at the Graduate Institute of Linguistics, National Taiwan University, June 25.
- Shibatani, Masayoshi. 2011. Toward the functional definitions of clauses and sentences. Presented at Atelier international: Finitude et nominalisation, Paris, October 12-14.
- Shibatani, Masayoshi. 2013. What can Japanese dialects tell us about the function and development of the nominalization particle “no”? *Japanese/Korean Linguistics* 20. 421–444.
- Shibatani, Masayoshi. 2014. What is nominalization? Toward the theoretical foundations of nominalization. Presented at International Workshop on Nominalization in the Languages of the Americas, Pontificia Universidad Católica del Perú, Lima, Peru, Aug 28-30.
- Shibatani, Masayoshi. 2015. (Numeral) classifiers and nominalization. Presented at IX Congresso Internacional da ABRALIN, Belém PA, Brazil, February 27.
- Shibatani, Masayoshi. 2016. (Numeral) classifiers and nominalization in Japanese and the world's languages. Presented at Reischauer Institute of Japanese Studies, Harvard University, April 8.
- Shibatani, Masayoshi & Khaled Awadh bin Makhshen. 2009. Nominalization in Soqotri, a South Arabian language of Yemen. In W. Leo Wetzels (ed.), *The Linguistics of endangered languages: Contributions to morphology and morpho-syntax*, 311–332. Utrecht: Landelijke Onderzoekschool Taalwetenschap.
- Shibatani, Masayoshi & Hiromi Shigeno. 2013. Amami nominalizations. *International Journal of Okinawan Studies: Special issue on Ryukyuan languages* 4(1). 107.
- Shih, Cindy Peiru. 2007. Interrogative constructions in Plngawan Atayal. Taipei: National Taiwan Normal University (M.A. thesis).
- Shih, Winslow Chia-lin. 2012. Are there case markers in Budai Rukai. Taipei: National Taiwan University (M.A. thesis).
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Siewierska, Anna. 2009. Person asymmetries in zero expression and grammatical function. In Franck Floricic (ed.), *Essais de linguistique generale et de typologie linguistique offerts au Professeur Denis Creissels à l'occasion de ses 65 ans*, 425–438. Paris: Presses de L'Ecole Normale Supérieure.
- Simpson, Jane & Margaret Withgott. 1986. Pronominal clitic clusters and templates. In Hagit Borer (ed.), *The Syntax of pronominal clitics*, 149–174. Orlando: Academic Press.
- Sin-huei Wu. 2010. Negation in Southern Paiwan. Kaohsiung: National Kaohsiung Normal University (M.A. thesis).
- Spencer, Andrew. 2005. Towards a typology of “mixed categories.” In Cemil Orhan Orgun & Peter Sells (eds.), *Morphology and the web of grammar: Essays in memory of Steven G. Lapointe*, 95–138. Stanford: Center for the Study of Language and Information.
- Spencer, Andrew. 2007. The possessum-agreement construction or “Does Albanian have a genitive case?” *Essex Research Reports in Linguistics* 53. 219–246.
- Spruiell, William Craig. 1990. Perspectives on nominalization. Houston: Rice University

- (Ph.D. dissertation).
- Starosta, Stanley. 1988a. A grammatical typology of Formosan languages. *Bulletin of the Institute of History and Philology* 59(2). 541–576.
- Starosta, Stanley. 1988b. *The case for Lexicase: An outline of Lexicase Grammatical Theory*. London: Pinter Publishers/Cassell.
- Starosta, Stanley. 1993. The case-marking system of Proto-Formosan. In Luksaneeyanawin Sudaporn (ed.), *PAN-Asiatic linguistics: Proceedings of the 3rd International Symposium on Language and Linguistics, vol. 3*, 1207–1221. Bangkok: Chulalongkorn University.
- Starosta, Stanley. 1995. A grammatical subgrouping of Formosan languages. In Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng (eds.), *Austronesian studies relating to Taiwan*, 683–726. Taipei: Academia Sinica.
- Starosta, Stanley. 1997. Formosan clause structure: Transitivity, ergativity, and case marking. In Chiu-yu Tseng (ed.), *Chinese languages and linguistics, vol. 4: Typological studies of languages in China*, 125–154. Taipei: Academia Sinica.
- Starosta, Stanley. 1999. Transitivity, ergativity, and the best analysis of Atayal case marking. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected papers from the 8th International Conference on Austronesian Linguistics*, 371–392. Taipei: Academia Sinica.
- Starosta, Stanley. 2009. Verbal inflection versus deverbal nominalization in PAN: The evidence from Tsou. In Elizabeth Zeitoun (ed.), *Formosan linguistics: Stanley Starosta's contributions*, 483–521. Taipei: Academia Sinica.
- Starosta, Stanley, Andrew K. Pawley & Lawrence A. Reid. 1982. The evolution of focus in Austronesian. In Amran Halim, Lois Carrington & S. A. Wurm (eds.), *Papers from the 3rd International Conference on Austronesian Linguistics, vol. 2: Tracking the travellers*, 145–170. Canberra: Australian National University.
- Stassen, Leon. 1985. *Comparison and Universal Grammar*. Oxford: Basil Blackwell.
- Stassen, Leon. 1997. *Intransitive predication*. Oxford: Clarendon Press.
- Stolz, Thomas, Sonja Kettler, Cornelia Stroh & Aina Urdze. 2008. *Split possession: An areal-linguistic study of the alienability correlation and related phenomena in the languages of Europe*. (Studies in Language Companion Series 101). Amsterdam: John Benjamins.
- Su, Yi-fan. 2008. Adverbials in Takituduh Bunun. Hsinchu: National Tsing Hua University.
- Sung, Li-may. 2011. Clausal nominalization in Budai Rukai. In Foong Ha Yap, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in Asian languages: Diachronic and typological perspectives*, 523–559. Amsterdam: John Benjamins.
- Sung, Li-may. 2015. Why exclamatives in Budai Rukai. In Elizabeth Zeitoun, Stacy F. Teng & Jing-lan Joy Wu (eds.), *New advances in Formosan linguistics*, 291–311. Canberra: Australian National University.
- Szakos, Józef. 1994. Die Sprache der Cou: Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan. Bonn: Rheinische Friedrich-Wilhelms-Universität Bonn (Ph.D. dissertation).
- Takivatan, Atul Manqoqo. 2011. A preliminary study on noun formation in Takibakha Bunun. Hsinchu: National Hsinchu University of Education (M.A. thesis).
- Talmy, Leonard. 2000. *Toward a cognitive semantics, vol. 1: Concept structuring systems*. Cambridge: MIT Press.
- Tang, Chih-chen Jane, Henry Yung-li Chang & Dah-an Ho. 1998. On noun phrase structures in Paiwan. *Tsing Hua Journal of Chinese Studies* 28(3). 335–384.
- Tang, Yaoming. 2008. Negators in Budai Rukai. Kaohsiung: National Kaohsiung Normal University (M.A. thesis).
- Taylor, John R. 1996. *Possessives in English: An exploration in Cognitive Grammar*. Oxford: Oxford University Press.

- Teng, Stacy Fang-ching. 2007. A grammar of Puyuma, an Austronesian language of Taiwan. Canberra: Australian National University (Ph.D. dissertation).
- Teng, Stacy Fang-Ching. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. (Pacific Linguistics 595). Canberra: Australian National University.
- Teng, Stacy Fang-ching. 2009. Case syncretism in Puyuma. *Language and Linguistics* 10(4). 819–844.
- Teng, Stacy Fang-ching. 2011. A comparative study of the personal pronominal systems in three Puyuma dialects. Presented at NTU Workshop on Discourse and Grammar in Formosan Languages.
- Teng, Stacy Fang-ching. 2012. Formation of instrumental nouns in Katripul Puyuma. Presented at the linguistics colloquium of Academia Sinica, Taiwan.
- Teng, Stacy Fang-ching. 2015. A comparative study of the personal pronominal systems in three Puyuma dialects. In Elizabeth Zeitoun, Stacy F. Teng & Jing-lan Joy Wu (eds.), *New advances in Formosan linguistics*, 407–429. Canberra: Australian National University.
- Teng, Stacy Fang-ching & Malcolm Ross. 2010. Is Puyuma a primary branch of Austronesian? A reply to Sagart. *Oceanic Linguistics* 49(2). 543–558.
- Thim, Stefan. 2012. *Phrasal verbs: The English verb-particle construction and its history*. (Topics in English Linguistics 78). Berlin: Mouton de Gruyter.
- Trask, Robert Lawrence. 1993. *A dictionary of grammatical terms in linguistics*. London: Routledge.
- Tryon, Darrell. 2006. Proto-Austronesian and the major Austronesian subgroups. In Peter Bellwood, James J. Fox & Darrell Tryon (eds.), *The Austronesians: Historical and comparative perspectives*, 19–41. Canberra: Research School of Pacific Studies, the Australian National University.
- Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsuoic phonology*. Tokyo: Institute for the Study of Languages, Cultures of Asia, and Africa.
- Tsuchida, Shigeru. 1980. ブユマ語 (タマラカオ方言) 語彙: 付・語法概説およびテキスト [Puyuma (Tamalakaw dialect) vocabulary: With grammatical notes and texts]. In 黒潮文化の會 [Committee of Cultures along the Kuroshio Current] (ed.), *黒潮の民族・文化・語言 [Peoples, cultures, and languages along the Kuroshio Current]*, 183–307. Tokyo: Kadokawa Shoten.
- Tsuchida, Shigeru. 1988a. カナカナブ語 [The Kananavu language]. In Takashi Kamei, Rokurō Kōno & Eiichi Chino (eds.), *言語学大辞典 [The Sanseido encyclopedia of linguistics]*, vol. 1, 1203–1210. Tokyo: Sanseido.
- Tsuchida, Shigeru. 1988b. アミ語 [The Amis language]. In Takashi Kamei, Rokurō Kōno & Eiichi Chino (eds.), *言語学大辞典 [The Sanseido encyclopedia of linguistics]*, vol. 1, 447–449. Tokyo: Sanseido.
- Tsuchida, Shigeru. 1989. サイシャット語 [The Saisyat language]. In Takashi Kamei, Rokurō Kōno & Eiichi Chino (eds.), *言語学大辞典 [The Sanseido encyclopedia of linguistics]*, vol. 2, 5–7. Tokyo: Sanseido.
- Tsuchida, Shigeru. 1992a. ブユマ語 [The Puyuma language]. In Takashi Kamei, Rokurō Kōno & Eiichi Chino (eds.), *言語学大辞典 [The Sanseido encyclopedia of linguistics]*, vol. 3, 742–753. Tokyo: Sanseido.
- Tsuchida, Shigeru. 1992b. ブヌン語 [The Bunun language]. In Takashi Kamei, Rokurō Kōno & Eiichi Chino (eds.), *言語学大辞典 [The Sanseido encyclopedia of linguistics]*, vol. 3, 731–734. Tokyo: Sanseido.
- Tsuchida, Shigeru. 1993a. カバラン語 [The Kavalan language]. In Takashi Kamei, Rokurō Kōno & Eiichi Chino (eds.), *言語学大辞典 [The Sanseido encyclopedia of linguistics]*, vol. 5, 89–99. Tokyo: Sanseido.
- Tsuchida, Shigeru. 1993b. パゼツヘ語 [The Pazeh language]. In Takashi Kamei, Rokurō Kōno & Eiichi Chino (eds.), *言語学大辞典 [The Sanseido encyclopedia of linguistics]*, vol. 5, 89–99. Tokyo: Sanseido.

- linguistics*], vol. 5, 302–310. Tokyo: Sanseido.
- Tsuchida, Shigeru. 1995. Alienable and inalienable distinction in Puyuma. In Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng (eds.), *Austronesian studies relating to Taiwan*, 793–804. Taipei: Academia Sinica.
- Tsuchida, Shigeru. 2000. Lexical prefixes and prefix harmony in Siraya. In Videa P. de Guzman & Byron Bender (eds.), *Grammatical analysis: Morphology, syntax, and semantics. Studies in honor of Stanley Starosta*, 109–128. University of Hawai'i Press.
- Tsuchida, Shigeru (ed.). 2003. *カナカナブ語 (台湾オーストロネシア語族) テキスト [Kanakanavu texts (Austronesian Formosan)]*. (Endangered Languages of the Pacific Rim Series A3-014). Osaka: Osaka Gakuin University.
- Tsukida, Naomi. 1993. A brief sketch of the Sakizaya dialect of Amis. *Tokyo University Linguistics Papers (TULIP)* 13. 375–413.
- Tsukida, Naomi. 2000. The CX-un and CX-an forms in Seediq. In Ritsuko Kikusawa & Kan Sasaki (eds.), *Modern approaches to transitivity*, 53–78. Tokyo: Kurosio.
- Tsukida, Naomi. 2005. Seediq. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 291–325. London: Routledge.
- Tsukida, Naomi. 2008. Verb classification in Amis. In Mark Donohue & Søren Wichmann (eds.), *The typology of semantic alignment*, 277–293. Oxford: Oxford University Press.
- Tsukida, Naomi. 2009. *セデック語 (台湾) の文法 [Grammar of the Seediq language]*. Tokyo: University of Tokyo (Ph.D. dissertation).
- Tsukida, Naomi & Shigeru Tsuchida. 2007. Indigenous Languages of Formosa. In Osahito Miyaoka, Osamu Sakiyama & Michael E. Krauss (eds.), *The vanishing languages of the Pacific Rim*, 285–300. Oxford: Oxford University Press.
- Tsunoda, Tasaku. 1988. Antipassives in Warrungu and other Australian languages. In Masayoshi Shibatani (ed.), *Passive and voice*, 595–650. Amsterdam: John Benjamins.
- Tung, T'ung-ho. 1964. *A descriptive study of the Tsou language*. (Special Publications 48). Taipei: Academia Sinica.
- Ulan, Russell. 1978. Toward a typology of substantival possession. In Joseph Harold Greenberg, Charles Albert Ferguson & Edith A. Moravcsik (eds.), *Universals of human language, vol. 4: Syntax*, 221–248. Stanford: Stanford University Press.
- van den Berg, René. 1989. *A grammar of the Muna language*. Dordrecht-Holland: Foris.
- van der Auwera, Johan, Nina Dobrushina & Valentin Goussev. 2013. Imperative-hortative systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structure online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/72>.
- van der Voort, Hein. 2009. Possessive expressions in the Southwestern Amazon. In William B. McGregor (ed.), *The expression of possession*, 343–388. Berlin: Mouton de Gruyter.
- Velázquez-Castillo, Maura. 1996. *The grammar of possession: Inalienability, incorporation, and possessor ascension in Guaraní*. (Studies in Language Companion Series 33). Amsterdam: John Benjamins.
- Wang, May Hsiu-mei. 2005. Case markers in Tona Rukai revisited: the distinction between ko and na. *Concentric: Studies in Linguistics* 31(2). 1–29.
- Wang, Shan-Shan. 2004. An ergative view of Thao syntax. Honolulu: University of Hawai'i at Mānoa (Ph.D. dissertation).
- Wang, Shu-Chen. 2005. The grammatical representations of causatives and resultatives in Paiwan. Chiayi: National Chung Cheng University (M.A. thesis).
- Wei, Ting-Chi. 2008. *Typology of ellipsis in Formosan languages*. National Science Council technical report (NSC 96-2411-H-017-012).
- Williams-van Klinken, Catharina. 2015. *Peace Corps East Timor Tetun language course*.

- 3rd ed. Dili: Peace Corps East Timor.
- Wolff, John U. 1973. Verbal inflection in Proto-Austronesian. *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*, 71–91. Quezon: Linguistics Society of the Philippines.
- Wolff, John U. 2002. Final words: The development of the focus system. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 437–449. Canberra: Australian National University.
- Wouk, Fay & Malcolm Ross (eds.). 2002. *The history and typology of western Austronesian voice systems*. (Pacific Linguistics 518). Canberra: Australian National University.
- Wu, Chunming. 2013. The syntax of linking constructions in Mayrinax Atayal and Sinvaudjan Paiwan. Hsinchu: National Tsing Hua University (Ph.D. dissertation).
- Wu, Jing-lan Joy. 1995. Complex sentences in Amis. Taipei: National Taiwan Normal University (M.A. thesis).
- Wu, Jing-lan Joy. 1996a. Serial verb construction in Amis. *Bulletin of National Taiwan Normal University* 41. 317–354.
- Wu, Jing-lan Joy. 1996b. Relative clauses in Amis. *Studies in English Literature and Linguistics* 22. 115–146.
- Wu, Jing-lan Joy. 2000. *A reference grammar of Amis*. (Series on Formosan Languages 11). Taipei: Yuanliu.
- Wu, Jing-lan Joy. 2003. Clausal modifiers in Amis. *Concentric: Studies in English Literature and Linguistics* 29(2). 59–81.
- Wu, Jing-lan Joy. 2006. Verb classification, case marking, and grammatical relations in Amis. Buffalo: State University of New York at Buffalo (Ph.D. dissertation).
- Wu, Jing-lan Joy. 2007. Voice markers in Amis: A Role and Reference Grammar analysis. *Language and Linguistics* 8(1). 95–142.
- Wu, Jing-lan Joy. 2012. On the N-N constructions in Amis. Presented at the 11th Conference on Austronesian Linguistics (ICAL-11), Bali, Indonesian, July 2–6.
- Xrakovskij, Viktor S. 2001. Hortative constructions. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, vol. 2, 1028–1038. Berlin: Walter de Gruyter.
- Yan, Zhi-kuang. 1992. Syntactic structure of Amis: A study of participants and events. Taipei: National Chengchi University (M.A. thesis).
- Yang, Chia-Ying. 2015. Linguistic representations of benefaction and malefaction in Kananavu. Taipei: National Taiwan University (M.A. thesis).
- Yang, Mei. 2005. 阿美语动词的语义特征及结构分析 [Semantic features and structural analyses of Amis verbs]. Beijing: Central University for Nationalities (Ph.D. dissertation).
- Yap, Foong Ha, Karen Grunow-hårsta & Janick Wrona. 2011. Introduction: Nominalization strategies in Asian languages. In Foong Ha Yap, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in Asian languages: Diachronic and typological perspectives*, 1–57. Amsterdam: John Benjamins.
- Yap, Foong Ha, Karen Grunow-Hårsta & Janick Wrona (eds.). 2011. *Nominalization in Asian Languages: Diachronic and typological perspectives*. (Typological Studies in Language 96). Amsterdam: John Benjamins.
- Yap, Foong Ha & Jiao Wang. 2011. From light noun to nominalizer and more: The grammaticalization of zhe and suo in Old and Middle Chinese. In Foong Ha Yap, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in Asian languages: Diachronic and typological perspectives*, 61–107. Amsterdam: John Benjamins.
- Yeh, Marie Mei-li. 1991. Saisiyat structure. Hsinchu: National Tsing Hua University (M.A. thesis).
- Yeh, Marie Mei-li. 2000a. *A reference grammar of Saisiyat*. (Series on Formosan

- Languages 2). Taipei: Yuanliu.
- Yeh, Marie Mei-li. 2000b. The syntax and semantics of Saisiyat negators. In Videa P. de Guzman & Byron Bender (eds.), *Grammatical analysis: Morphology, syntax, and semantics. Studies in honor of Stanley Starosta*, 258–273. Honolulu: University of Hawai'i Press.
- Yeh, Marie Mei-li. 2003. A syntactic and semantic study of Saisiyat verbs. Taipei: National Taiwan Normal University (Ph.D. dissertation).
- Yeh, Marie Mei-li. 2011. Nominalization in Saisiyat. In Foong Ha Yap, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in Asian languages: Diachronic and typological perspectives*, 561–588. Amsterdam: John Benjamins.
- Yeh, Marie Mei-li, Lillian M. Huang, Elizabeth Zeitoun, Anna Hsiou-chuan Chang & Jing-lan Joy Wu. 1998. A Preliminary study on the negative constructions in some Formosan languages. In Shuanfan Huang (ed.), *Selected papers from the second International Symposium on Languages in Taiwan*, 79–110. Taipei: Crane.
- Yeh, Maya Yuting. 2013. Event conceptualization and verb classification in Squiliq Atayal. Taipei: National Taiwan University (Ph.D. dissertation).
- Yeh, Yu-Ting. 2005. Negation in Kavalan: A syntactic study. Taipei: National Taiwan University (M.A. thesis).
- Yen, Doris Ching-jung. 2012. Clitics in Kavalan (East Formosan, Austronesian). Nantou: National Chi Nan University (M.A. thesis).
- Yen, Doris Ching-jung & Loren Billings. 2011. Sequences of pronominal clitics in Mantauran Rukai: V-deletion and suppletion. Presented at the 18th annual meeting of the Austronesian Formal Linguistics Association (AFLA-18), Harvard University, Cambridge, March 4–6.
- Yen, Doris Ching-jung & Loren Billings. 2014. The cluster-internal ordering of clitics in Kavalan (East Formosan, Austronesian). *Proceedings of the 38th annual meeting of the Berkeley Linguistics*, 523–538.
- Zeitoun, Elizabeth. 1995. Problèmes de linguistique dans les langues aborigènes de Taiwan. Paris: Université René Diderot Paris 7 (Ph.D. dissertation).
- Zeitoun, Elizabeth. 1997. The pronominal system of Mantauran (Rukai). *Oceanic Linguistics* 36(2). University of Hawai'i Press. 312–346.
- Zeitoun, Elizabeth. 2000a. Notes on a possessive construction in the Formosan languages. In Videa P. de Guzman & Byron Bender (eds.), *Grammatical analysis: Morphology, syntax, and semantics. Studies in honor of Stanley Starosta*, 241–257. Honolulu: University of Hawai'i Press.
- Zeitoun, Elizabeth. 2000b. *A reference grammar of Bunun*. (Series on Formosan Languages 5). Taipei: Yuanliu.
- Zeitoun, Elizabeth. 2000c. *A reference grammar of Tsou*. (Series on Formosan Languages 7). Taipei: Yuanliu.
- Zeitoun, Elizabeth. 2001. Negation in Saisiyat: Another perspective. *Oceanic Linguistics* 40(1). 125–134.
- Zeitoun, Elizabeth. 2002. Nominalization in Mantauran (Rukai). *Language and Linguistics* 3(2). 241–282.
- Zeitoun, Elizabeth. 2003. Toward a reconstruction of Proto-Rukai morpho-syntax. Keynote address at the 10th annual meeting of the Austronesian Formal Linguistics Association (AFLA-10), University of Hawai'i at Mānoa, March 28–30.
- Zeitoun, Elizabeth. 2004. Typologie des langues austronésiennes de Taïwan. *Fait de langues: Revue de linguistique no 23–24: Les langues austronésiennes*. 41–58.
- Zeitoun, Elizabeth. 2005. Tsou. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 259–290. London: Routledge.
- Zeitoun, Elizabeth. 2007. *A grammar of Mantauran (Rukai)*. Taipei: Academia Sinica.
- Zeitoun, Elizabeth. 2008. Kinship terms in Mantauran Rukai. *Academia Sinica Weekly* 1191. 4–6.

- Zeitoun, Elizabeth, Tai-hwa Chu & Lalo a Tahesh Kaybaybaw. 2011. ki as a marker of coordination and comitativity in Saisiyat. *Language and Linguistics* 12(1). 77–107.
- Zeitoun, Elizabeth & Lillian M. Huang. 1997. Toward a typology of tense, aspect and modality in the Formosan languages: A preliminary study. In Tseng Chiu-yu (ed.), *Chinese languages and linguistics IV: Typological studies of languages in China*, 595–618. Taipei: Academia Sinica.
- Zeitoun, Elizabeth & Lillian M. Huang. 2000. Concerning ka- , an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39(2). 391–414.
- Zeitoun, Elizabeth & Lillian M. Huang. 2006. Preliminary remarks on nominalization in Formosan languages. Presented at the 13th annual meeting of the Austronesian Formal Linguistics Association (AFLA-13), Tsing Hua University, Hsinchu, Taiwan, March 24–26.
- Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh & Anna H. Chang. 1999. Existential, possessive, and locative constructions in Formosan languages. *Oceanic Linguistics* 38(1). 1–42.
- Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh, Anna H. Chang & Jing-lan Joy Wu. 1996. The temporal, aspectual, and modal systems of some Formosan languages: A typological perspective. *Oceanic Linguistics* 35(1). University of Hawai'i Press. 21–56.
- Zeitoun, Elizabeth & Stacy Fang-ching Teng. 2014. The position of Kanakanavu and Saaroa within the Formosan languages revisited. Presented at the 14th International Symposium on Chinese Languages and Linguistics (IsCLL-14) & the 10th Anniversary of The Institute of Linguistics, Academia Sinica, Taipei, June 4–6.
- Zeitoun, Elizabeth & Chen-huei Wu. 2005. Saisiyat reduplication revisited. *Concentric: Studies in Linguistics* 31(2). 31–56.
- Zeitoun, Elizabeth & Chen-huei Wu. 2006. An overview of reduplication in Formosan languages. In Henry Y. Chang, Lillian M. Huang & Dah-an Ho (eds.), *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, 97–142. Taipei: Academia Sinica.
- Zeng, Siqu. 1991. 台湾阿眉斯语语法[*Amis grammar*]. Beijing: Minzu Publishing House.
- Zeng, Siqu. 2002. 阿眉斯语的基本词缀及其复合结构分析[An analysis of basic affixes and their combinations in Amis]. *Minzuyuwen* 1. 34–44.
- Zribi-Hertz, Anne. 1997. On the dual nature of the “possessive” marker in Modern English. *Journal of Linguistics* 33(2). 511–537.