3. Word-formation

Papuan Malay has very little productive morphology. Word formation is limited to the two derivational processes of reduplication and affixation. Inflectional morphology is lacking, as nouns and verbs are not marked for any grammatical category such as gender, number, or case. There is also no voice system on verbs. Compounding is a third word-formation process; it remains uncertain, however, to what degree it is a productive process.

In discussing word-formation in Papuan Malay, a major issue is to what degree these processes are productive. Following Booij (2007: 320), a morphological process "is productive if it can be applied to form new (forms of) words". Given this definition, the data in the present corpus indicates that reduplication in Papuan Malay is a very productive process, whereas affixation has only very limited productivity. The productivity of compounding as a word-formation process remains debatable.

This chapter discusses two word-formation processes in detail: affixation in §3.1 and compounding in §3.2. Reduplication is mentioned briefly in §3.3 and more fully discussed in Chapter 4. The main points of this chapter are summarized in §3.4.

3.1. Affixation

3.1.1. Introduction

Affixation is a morphological process whereby an affix is attached to a lexical root to derive new lexemes. As word-forming elements, affixes are bound morphemes which provide additional lexical or syntactic information for their respective base words, which in Papuan Malay are lexical roots. In Papuan Malay, as is typical for derivational devices in general, affixation applies to the open word classes only, namely to nouns and verbs.

The present corpus contains a considerable number of morphologically complex lexical items with the 2,458-item word list mentioned in §1.8.6 including 523 affixed lexemes (21%). The most commonly employed (historical) affixes are the prefixes ***TER-*** 'ACL', ***PE(N)-*** 'AG', and ***BER-*** 'INTR', the suffixes ***-ang*** 'PAT' and ***-nya*** '3POSSR', and the circumfix ***ke-/-ang*** 'NMLZ'.[87]

Before examining these affixes in detail, the remainder of this introduction discusses methodological issues related to examining the productivity of affixation in Papuan Malay.

Morphological patterns are considered to be productive if language users apply them "to create new well-formed complex words" by systematically extending the pattern "to new cases" (Booij 2007: 67, 68). By contrast, a morphological pattern is said to be unproductive when the morphological rule involved "is not used for coining new words" but "has become obsolete" (2007: 68). The productivity of a given pattern is a matter of degree, however, as pointed out by scholars such as Pike

[87] The small caps designate the abstract representation of affixes that have more than one form of realization; prefixes ***TER-***, ***PE(N)-***, and ***BER-***, have two allomorphs each, namely ***ter-*** and ***ta-*** (§3.1.2.1), ***pe(N)-*** and ***pa(N)-*** (small-caps *N* represents the different realizations of the nasal) (§3.1.4.1), and ***ber-*** and ***ba-*** (§3.1.5.1), respectively.

(1967: 169–172) , Bauer (1983: 62–100) , Aikhenvald (2007: 49–58) , or Booij (2007: 67–71). This degree depends on the amount "to which the structural possibilities of a word-formation pattern are actually used" (Booij 2007: 68). That is, depending on their functional load, some patterns are "fully active" or productive, while others are "inactive" or unproductive, with "semi-active" or semi-productive patterns found in-between (Pike 1967: 169–171).[88] Therefore, productivity is best viewed as a "cline" (Bauer 1983: 97) or a "scalar phenomenon" (Bauer 2001: 126).[89] On such a cline of productivity, fully productive patterns are viewed as one end-point, and completely unproductive patterns as the other end-point of the continuum, with semi-productive patterns found in-between.

To investigate whether and to what degree Papuan Malay speakers employ a given affix to create new words, one technique would be to devise a test along the lines of Aronoff and Schvaneveldt's (1978) "Productivity Experiment". This psycholinguistic experiment involved a lexical-decision task which required testees to make judgments about possible but non-occurring affixed words. That is, the testees had to judge whether or not these words were instances of English.

For the present study no productivity tests were conducted to determine whether and to what extent a given affix can be attached to Papuan Malay roots to derive new lexical items. Tests such as the mentioned lexical-decision tasks were considered unworkable due to the sociolinguistic profile of the Papuan Malay speech variety and speech communities, discussed in §1.5:

- Strong and increasing language vitality;
- Functional distribution of Papuan Malay as the LOW variety, and Indonesian as the HIGH variety, in terms of Ferguson's (1972) notion of diglossia;
- Ambivalent language attitudes towards Papuan Malay; and
- Lack of language awareness of Papuan Malay speakers about the status of Papuan Malay as a language distinct from Indonesian.

Given this sociolinguistic profile and the formal setting of a test situation as well as the fairly high degree of linguistic relatedness between Papuan Malay and Indonesian, an undesirable amount of interference from Indonesian was expected. This assumption is based on Weinreich's (1953: 1) definition of "interference" as "instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language, i.e. as a result of language contact". Even in a monolingual test situation, such interference would most likely have had a skewing impact on testees' naïve judgments, given that, when in the "monolingual speech mode […] bilinguals rarely deactivate the other language totally", as Grosjean (1992: 59) points out.

Given these problems, the attested affixes and derived words were instead examined in terms of seven language internal and two language external factors. These factors were deemed relevant in examining the productivity of these affixes.

---

88 Pike (1967: 169–171) talks about the "Activeness of Morphemes" rather than of "morphological patterns".

89 As Bauer (2001: 125) elaborates, however, there is an ongoing discussion among scholars "whether productivity is a gradable/scalar phenomenon or not".

1. Language internal factors

The affixes are examined with respect to the following seven language internal factors: (a) syntactic properties, (b) type frequencies, token frequencies, and hapaxes, (c) relative token frequencies of affixed lexemes and base words, (d) form-function relationship between the derivation and its base word, (e) alternative strategies, (f) affixed lexemes lacking a base, and (g) status of the affixed lexemes as part of the Papuan Malay lexicon or as code-switches with Indonesian.

a) Syntactic properties

Before examining the productivity of the affixes in terms of the above-mentioned language internal factors, the syntactic properties of the base words are described as to whether they can be attached to verbal, nominal, adverbial, and/or other bases. Likewise, the syntactic properties of the affixed lexemes are described, as to which word class they belong to.

b) Type frequencies, token frequencies and hapaxes[90]

If an affix is represented by a large number of words (high type frequency) which, in turn, have low token frequencies, this is taken as evidence that the affixation process under investigation is more productive. (For the purposes of this study, type frequencies of ten or more are considered as "(relatively) high" while token frequencies of less than 20 are considered as "(relatively) low".

As Plag (2006: 542) points out, derived words with low token frequencies "tend to be words that are unlikely to be familiar to the hearer". This applies especially to hapaxes which occur only once in a corpus. They can, however, be understood if "an available word-formation rule allows the decomposition of the newly encountered word into its constituent morphemes and thus the computation of the meaning on the basis of the meaning of the parts" (2006: 542). Hence, productive morphological patterns tend to be characterized by "large numbers of low frequency words and small numbers of high frequency words, with the former keeping the rule alive. In contrast, unproductive morphological categories will be characterized by a preponderance of words with rather high frequencies and by a small number of words with low frequencies" (2006: 542).

Among the derived words with low token frequency, hapaxes are especially useful in determining the productivity of a morphological pattern, as "the highest proportion of neologisms" is found here (Plag 2006: 542); or in other words, "[the] higher the number of hapaxes, the greater the productivity" (2006: 544). Therefore, as Booij (2007: 69–70) points out, "one might define the degree of productivity **P** of a particular morphological process as the proportion between the number of hapaxes of that type (n_1) to

[90] Type frequency is defined as "the number of types of a class of linguistic units in a corpus", while token frequency refers to "the number of tokens of a linguistic unit or a class of linguistic units in a corpus" (Booij 2007: 323). Hapaxes are "new word types that occur only once in the corpus, and clearly do not belong to the set of established words" (Booij 2007: 69).

the total number of tokens N" for that particular affix; a definition which is based on Baayen's (1992: 115) formula $P = n_1/N$.

For the present study, however, it remains unclear to what extent the attested hapaxes are useful in determining productivity. That is, the limited size of the present corpus makes it difficult to verify which hapaxes are neologisms in Papuan Malay and which ones merely reflect the limited size of the corpus. Moreover, the literature does not mention thresholds which would allow interpreting a calculated P value in terms of the degree of productivity of a given morphological pattern. For the interested reader, however, the number of hapaxes and their respective P values for each affix are given in footnotes throughout this chapter.

c) Relative token frequencies of affixed lexemes and base words

If the token frequencies for the affixed lexemes are lower than those for their bases, this is taken as an indication that the affixation process is more likely to be productive, following Hay and Baayen (2002: 230), who give evidence that "[the] frequency with which an affix is activated during processing directly affects the degree to which it is productive".

Hay and Baayen's (2002) findings build on Hay (2001), who shows "that the frequency of the base form is involved in facilitating decomposability. When the base is more frequent than the whole, the word is easily and readily decomposable. However, when the derived form is more frequent than the base it contains, it is more difficult to decompose and appears to be less complex". In terms of processing, morphologically complex words with a low relative frequency are accessed via their parts, that is, via a "decomposed access" or "parsing route". Morphologically complex words with a high relative frequency, by contrast, are accessed as whole words via a "whole-word access" or "direct route" (2001: 1044-1047, 1055).

In their study, Hay and Baayen (2002: 203–204) argue that "for an affix to remain productive, words containing that affix must be parsed sufficiently often that the resting activation level of that affix remains high". The findings of this study confirm this link between productivity and parsing. Hay and Baayen (2002) show that affixes which derive words with low relative frequencies and high rates of decomposition are more likely to be productive. By contrast, affixes which derive words with high relative frequencies and low rates of decomposition are less likely to be productive.

d) Form-function relationship between the derivation and its base

Typical derivational processes include nominalization, verbalization, or class-preserving valency-changing operations, among others. In each case, the derivational process "results in the creation of a new word with a new meaning", as Aikhenvald (2007: 35) points out.

Following Booij (2007: 240, 323), one "necessary" albeit not "sufficient" condition for the productivity of such derivational processes is their transparency, which is defined as "the presence of a systematic form-meaning correspondence in a morphologically complex form". Therefore, if the form-function relationship between the affixed lexemes and their base is

transparent, this is taken as evidence that a given affixation process is more productive. If, by contrast, this relationship is opaque, this is considered evidence that the process is less productive.

If an affix is "polyfunctional", that is, if it can take bases from more than one lexical category, this is also taken as evidence that the process is more productive (Booij 2002: 89–91 ; see also Zwanenburg 2000).

Pairs of words in which the affixed words and their respective bases have the same semantics, are not taken as parts of a larger derivational paradigm. Instead these sets are taken as pairs of words belonging to different speech varieties, namely Papuan Malay and Indonesian. This conclusion is based on the fact that, in general, non-standard varieties of Malay "have lost most or all of this system of affixation", whereas "Standard Malay exhibits a rich system of affixation" (Paauw 2008: 20). Hence, for pairs of words with the same semantics, the unaffixed base words are taken to be the native Papuan Malay lexemes, whereas the affixed words are taken to be code-switches with the corresponding Indonesian lexemes.

e) Alternative strategies

If speakers employ alternative strategies that do not involve affixation and that express the same meanings as the affixed forms, these alternative strategies are taken as evidence that the affixation process is less productive.

f) Affixed lexemes lacking a base

Affixed lexemes for which there is no corresponding base have lost their status as complex words. They are so-called "formally complex words" (Booij 2007: 17). Such a word "behaves as a complex word although there is no corresponding semantic complexity" (2007: 313). A high number of formally complex words are taken as evidence that the affixation process is less productive. Their non-compositional semantics suggest that these lexemes are either lexicalized forms or code-switches with Indonesian. (For each affix, the number of formally complex words is given with a few examples. Given, however, that they have lost their status as complex words, these items are not further discussed.)

g) Status of the affixed lexemes as part of the Papuan Malay lexicon or as code-switches with Indonesian

If a large number of affixed lexemes are not part of the Papuan Malay lexicon but code-switches with Indonesian, this is taken as evidence that the derivation process for a given affix is less productive.

Sources such as Jones (2007), or Tadmor (2009) allow the identification of foreign, non-Malay loan words in the present corpus. They do not, however, allow identifying code-switches with Indonesian. Hence, an alternative approach was deemed necessary to explore whether the affixed lexemes are part of the Papuan Malay lexicon or constitute code-switches with Indonesian.

All 533 attested affixed lexemes were discussed with a Papuan Malay consultant who has a high level of language awareness, both with respect to

Papuan Malay and to Indonesian. Based on his knowledge of both languages, the consultant classified the affixed lexemes as "Papuan Malay" or "borrowings from Indonesian". While his tentative classification is subjective and not necessarily representative, it provides one more piece of evidence as to the potential productivity of the attested affixes. In Table 1 to Table 24, these alleged borrowings or code-switches with (Standard) Indonesian, are underlined.

2. Language external factors: Variables of the communicative event

The affixes were examined as to whether they are employed without sociolinguistic restrictions or whether their use is conditioned by variables of the speech situation in terms of Fishman's (1965: 86) "domains of language choice". The main factors which influence language choices are (1) the topics discussed, (2) the relationships between the interlocutors, and (3) the locations where the communication takes place (1965: 67, 75). Speaker education levels are a fourth pertinent factor.

If the use of the affixes seems to be conditioned by language external factors, this is taken as evidence that the affixation process is less productive. For the present study, the pertinent "domains of language choice" are (a) the topics, (b) speaker education levels, and (c) the relationships between the interlocutors, all of which are discussed in the following. The locations of communication were not considered pertinent domains since all recorded conversations took place in the same informal setting of the home. (For details on the sociolinguistic profile of Papuan Malay, see §1.5.)

a) Speaker education levels

In West Papua, as is typical of diglossic situations, the HIGH variety Indonesian is acquired in school. Given their amount of access to the HIGH variety, better-educated speakers are more likely to display language behaviors influenced by the HIGH variety Indonesian than less-educated speakers. Therefore, if better-educated speakers employ a particular affix considerably more often than less-educated ones, this is taken as evidence that the affixed lexemes are not the result of a productive process but that they constitute code-switches with Indonesian.

b) Topics

Following Fishman (1965: 71), the topics under discussion may also bring "another language to the fore" as "certain topics are somehow handled better in one language than in another". This notion of topical regulation suggests that Papuan Malay speakers consider Indonesian, and not Papuan Malay, the appropriate language to use when discussing HIGH topics associated with formal domains such politics, education, or religion. Therefore, if Papuan Malay speakers use a particular affix much more often when discussing HIGH topics than when discussing casual daily-life issues (LOW topics), this is taken as evidence that the affixed lexemes are code-switches with Indonesian. This applies especially to less-educated Papuans, as better-educated Papuans already display a general tendency to include Indonesian features when speaking Papuan Malay, although this tendency is more pronounced when

the latter discuss high topics. (See also Factor 2 'Topical regulation' in §1.5.1).

c) Relationships between interlocutors

Given the diglossic distribution of Papuan Malay and Indonesian, it is expected that the language behavior of Papuans shows influences from the HIGH variety Indonesian when they interact with fellow-Papuans of higher status or with group outsiders. As discussed under Factor 3 'Relationships between interlocutors' in §1.5.1, the use of features from the HIGH variety serves to signal social inequality, distance, and formality. Therefore, if speakers use a given affix much more often when conversing with interlocutors of higher status or with group outsiders than when interacting with peers, this is taken as evidence that the affixed lexemes are code-switches with Indonesian. Again, this applies especially to less-educated Papuans, given that better-educated Papuans already show a general tendency to "dress-up" their Papuan Malay with Indonesian features, although this tendency is more marked when the latter interact with group outsiders, such as the author. (See also Factor 3 'Relationships between interlocutors' in §1.5.1.)

In examining the attested affixes and affixed lexemes as outlined above, none of the factors was taken in isolation. Instead, the findings pertaining to all nine factors were taken together as an indication of the degree of productivity for the affix in question. The results of this multifaceted investigation indicate that:

- Prefix ***TER-*** 'ACL' and suffix ***-ang*** 'PAT' are somewhat productive;
- Prefix ***PE(N)-*** 'AG' is, at best, marginally productive; and
- Prefix ***BER-*** 'INTR', suffix ***-nya*** '3POSSR', and circumfix ***ke-/-ang*** 'NMLZ' are unproductive.

The unproductive derivations are considered to be lexicalized forms borrowed into the language or code-switches with Indonesian; in the examples, however, no attempt is made to distinguish the two.

In the following, six affixes are discussed in detail in terms of the factors outlined above: ***TER-*** in §3.1.2, ***-ang*** in §3.1.3, ***PE(N)-*** in §3.1.4, ***BER-*** in §3.1.5, ***-nya*** in §3.1.6, and ***ke-/-ang*** in §3.1.7. For the three somewhat productive affixes (***TER-***, ***-ang***, and ***PE(N)-***) the mentioned variables of the communicative event are investigated in detail within the respective sections. For the remaining three affixes (***BER-***, ***-nya***, and ***ke-/-ang***) the variables of the communicative event are summarily discussed in §3.1.8. The main points on affixation are summarized in §3.4.

3.1.2. Prefix *TER-* 'ACL'

Affixation with ***TER-*** 'ACL' derives monovalent verbs from verbal bases.[91] The derived verbs denote accidental or unintentional actions or events, as shown in (1). This derivation process appears to be somewhat productive, as discussed below.

(1) bos pagi su br–angkat ke Sarmi begini adu
boss morning already INTR–leave to Sarmi like.this oh.no!
sial–ang **ter–paksa** tong dua jalang kaki
be.unfortunate–PAT ACL–force 1PL two walk foot
'as the boss had already left for Sarmi in the morning, oh no, damn it!, the two of us were **forced** to walk on foot' [080921-002-Cv.0001]

Prefix ***TER-*** is a reflex of Proto Malayic ****tʌr-***, which, following Adelaar (1992: 155), "contributed the notion of unintentionality or feasibility to the VTR or VDI to which it was affixed".[92] In Standard Malay, "***tər-*** denotes an 'accidental' state, process or action" when affixed to bivalent bases and "a superlative degree" when affixed to monovalent bases (1992: 150–151). In eastern Malay varieties, the prefix also denotes accidental or unintentional actions, or events that happened unexpectedly or unintentionally. These productive uses of the prefix are attested for Ambon Malay (van Minde 1997: 98), Banda Malay (Paauw 2008: 250), Kupang Malay (Steinhauer 1983: 46), Larantuka Malay (Paauw 2008: 256), Manado Malay (Stoel 2005: 22), North Moluccan / Ternate Malay (Taylor 1983: 18,[93] and Litamahuputty 2012: 133).

The present corpus includes 43 monovalent verbs (167 tokens) prefixed with ***TER-***:[94]

1. Verbs with bivalent bases (38 items with 153 tokens)
2. Verbs with monovalent bases (five items with 14 tokens)

The corpus also contains ten formally complex words with non-compositional semantics, such as ***tertawa*** 'laugh', ***tergrak*** 'be moved', or ***trapung*** 'be drifting'.[95]

Before discussing ***TER-***affixation of bivalent bases in §3.1.2.2 and of monovalent bases in §3.1.2.3, the allomorphy of ***TER-*** is examined in §3.1.2.1. Variables of the communicative event that may impact the use of ***TER-*** are explored in §3.1.2.4. The main points on prefix ***TER-*** are summarized and evaluated in §3.1.2.5.

---

[91] The small caps designate an abstract representation of the prefix as it has more than one form of realization, namely the two allomorphs ***ter-*** and ***ta-*** (see §3.1.2.1).

[92] VTR = transitive verb, VDI = ditransitive verb (Adelaar 1992: 8).

[93] While Taylor (1983: 18) considers the prefix to be productive, Voorhoeve (1983: 4) believes that it is unproductive.

[94] The 43 verbs include 21 hapaxes (P=0.1257); the 38 bivalent verbs include 17 hapaxes (P=0.1111); the four monovalent verbs include four hapaxes (P=0.2857).

[95] The historical roots ***tawa***, ***grak***, or ***apung*** do not exist in Papuan Malay.

3.1.2.1. Allomorphy of ***TER-***

Prefix ***TER-*** has two allomorphs, ***ter-*** and ***ta-***. The allomorphs are not governed by phonological processes.

The form ***ter-***, in turn, has three allomorphs that are the effect of, what Booij (2007: 75) calls "morphologically conditioned phonological rules". More specifically, the three allomorphs are conditioned by the word-initial segment of the base word, as shown in Table 1: /tɛr-/, /tɛ-/, and /tr-/. Most commonly, ***ter-*** is realized as /tɛr-/. With onset rhotic /r/, however, it is realized as /tɛ-/. With onset vowels, the prefix is usually realized as /tr-/.

Table 1: Realizations of allomorph ***ter-***

ter-base	Orthogr.	Gloss
/tɛr–pukul/	***terpukul***	'be beaten'
/tɛ–rɛndam/	***terendam***	'be soaked'
/tr–aŋkat/	***trangkat***	'be lifted'

Allomorph ***ta-*** is used in about one third of the affixed items; that is, 17 items of a total of 41 ***ta-*** tokens, listed in Table 2. Some of the derived items are alternatively realized with allomorph ***ter-***. Hence, for each item the frequencies for ***ta-*** and for ***ter-*** are given.[96] If in a greater number of tokens the prefix is realized with /ta-/ than with /tɛr-/, then its orthographic representation is ***ta-*** as in ***tagoyang*** 'be shaken'. If both realizations occur with the same frequency, then the orthographic representation follows its realization in the recorded texts, as in ***terlepas*** 'be loose'.

Table 2: Realizations of allomorph ***ta-***

ta-base	Orthogr.	Gloss	***ta-*** #	***ter-*** #
/ta–gojaŋ/	***tagoyang***	'be shaken'	9	0
/ta–putar/	***taputar***	'be turned around'	7	2
/ta–lipat/	***talipat***	'be folded'	6	1
/ta–lempar/	***talempar***	'be thrown'	4	1
/ta–guliŋ/	***taguling***	'be rolled over'	3	0
/ta–gantoŋ/	***tergantong***	'be dependent'	1	6
/ta–lepas/	***terlepas***	'be loose'	1	1
/ta–balik/	***tabalik***	'be turned upside down	1	0
/ta–bantiŋ/	***tabanting***	'be tossed around'	1	0
/ta–tʃukur/	***tacukur***	'be scalped'	1	0

[96] In addition, the 2,459-item word list (see Chapter 2) contains five items realized with /ta-/ rather than with /tɛr-/: /tabla/ 'be cracked open', /takumpul/ 'be gathered', /takupas/ 'be peeled', /tamasuk/ 'be included', and /tatutup/ 'be closed'. In the present corpus these items are realized with /tɛr-/. Further, the word list also includes three items realized with /tɛr-/ whereas in the corpus these items are most commonly realized with /ta-/: ***talempar*** 'be thrown', ***terlipat*** 'be folded', and ***tarangkat*** 'be lifted up'.

ta-base	Orthogr.	Gloss	***ta***- #	***ter***- #
/ta–gait/	***tagait***	'be hooked	1	0
/ta–hambur/	***tahambur***	'be scattered about'	1	0
/ta–kantʃiŋ/	***takancing***	'be locked'	1	0
/ta–lem/	***talem***	'be glued'	1	0
/ta–sala/	***tasala***	'be mistaken'	1	0
/ta–tikam/	***tatikam***	'be stabbed'	1	0
/ta–toŋkat/	***tatongkat***	'be beaten'	1	0

In realizing the prefix most commonly as ***ter***- rather than as ***ta***-, Papuan Malay differs from other eastern Malay varieties such as Ambon Malay (van Minde 1997: 98), Banda Malay (Paauw 2008: 250), Kupang Malay (Steinhauer 1983: 46), Manado Malay (Stoel 2005: 22), North Moluccan / Ternate Malay (Taylor 1983: 18, Voorhoeve 1983: 4, and Litamahuputty 2012: 133). In these varieties the prefix is always realized as ***ta***-. Instead, the ***TER-***prefixed items have more resemblance with the corresponding items in Indonesian, where the prefix is realized as ***ter***-. In addition, in Larantuka Malay the prefix is also realized as ***tə(r)***- (Paauw 2008: 253). The different behavior of Papuan Malay ***TER-*** supports the conclusion put forward in §1.2 that the history of Papuan Malay is different from that of the other eastern Malay varieties.

3.1.2.2. Prefixed items derived from bivalent verbal bases

The corpus contains 38 ***TER-***prefixed lexemes (with 153 tokens) with bivalent verbal base words (BW), as listed in Table 3. The affixation derives monovalent verbs with non-agent arguments through a valency-changing operation, in which ***TER-*** removes agent arguments. All but one of the derived lexemes are low frequency words (37 lexemes, attested with less than 20 tokens). Besides, the token frequencies for the respective bases are (much) higher for most of the derived words (29 lexemes).

Table 3: Affixation with ***TER-*** of bivalent verbal bases[97]

BW	Gloss	Item	Gloss	***TER-*** #	BW #
jadi	'become'	***terjadi***	'happen'	39	120
paksa	'force'	***terpaksa***	'be forced'	10	10
masuk	'enter'	***termasuk***	'be included'	9	261
putar	'turn around'	***taputar***	'be turned around'	9	33
goyang	'shake'	***tagoyang***	'be shaken'	9	10
gantong	'suspend'	***tergantong***	'be dependent'	7	14
lipat	'fold'	***talipat***	'be folded'	7	1
buka	'open'	***terbuka***	'be opened'	6	1
angkat	'lift'	***trangkat***	'be lifted'	5	81

[97] As discussed in language internal factor (g) on p. 115 in §3.1.1, alleged borrowings or code-switches with (Standard) Indonesian are underlined.

BW	Gloss	Item	Gloss	***TER-*** #	BW #
lempar	'throw'	***talempar***	'be thrown'	5	12
rendam	'soak'	***terendam***	'be soaked'	5	1
pukul	'beat'	***terpukul***	'be beaten'	4	59
bakar	'burn'	***terbakar***	'be burnt'	3	55
guling	'roll over'	***taguling***	'be rolled over'	3	2
tutup	'close'	***tertutup***	'be closed'	3	53
bagi	'divide'	***terbagi***	'be split up'	2	66
tarik	'pull'	***tertarik***	'be pulled'	2	32
lepas	'free'	***talepas***	'be loose'	2	23
kumpul	'gather'	***terkumpul***	'be collected'	2	16
tolak	'reject'	***tertolak***	'be rejected'	2	11
kupas	'peel'	***terkupas***	'be peeled'	2	1
buat	'make'	***terbuat***	'be made'	1	135
kenal	'know'	***terkenal***	'be well-known'	1	57
balik	'turn over'	***tabalik***	'be turned over'	1	37
ganggu	'disturb'	***terganggu***	'be disturbed'	1	18
bla	'split'	***terbla***	'be split'	1	13
pengaru	'influence'	***terpengaru***	'be affected'	1	7
banting	'throw'	***tabanting***	'be tossed around'	1	6
tukar	'exchange'	***tertukar***	'get changed'	1	6
tongkat	'cane'	***tatongkat***	'be beaten up'	1	5
singgung	'offend'	***tersinggung***	'be offended'	1	3
cinta	'love'	***tercinta***	'be beloved'	1	3
cukur	'flatten'	***tacukur***	'be scalped'	1	2
hambur	'scatter'	***tahambur***	'be scattered about'	1	1
wesel	'transfer'	***terwesel***	'be transferred'	1	2
tikam	'stab'	***tatikam***	'be stabbed'	1	2
kancing	'lock'	***takancing***	'be locked'	1	0
lem	'glue'	***talem***	'glued'	1	0

The derived verbs denote accidental or unintentional states, processes, or actions with the term "accidental" covering "such concepts as involuntary, unmotivated, agentless, sudden, and unexpected action (or state resulting therefrom)" (Adelaar 1992: 150). Hence, ***TER-*** is glossed as 'ACL', ('accidental'). Two ***TER-***prefixed items are given in context: ***tagoyang*** 'be shaken' in (2) and ***tertutup*** 'be closed' in (4). Both examples, together with the one in (3), illustrate how ***TER-*** decreases valency by "removing agent-like participants".

(2) de bilang mama sa liat pohong ini de **ta–goyang**
3SG say mother 1SG see tree D.PROX 3SG ACL–shake
'she said, 'mama, I saw this tree, it was **shaking**'' [080917-008-NP.0031]

(3) … bapa Andi K. dosen satu de **goyang** kepala
father Andi K. lecturer one 3SG shake head
'… Mr. Andi K., a certain lecturer, he **shook** (his) head' [080917-010-CvEx.0194]

(4) kalo ko **tutup** pintu berkat juga **ter–tutup** …
if 2SG close door blessing also ACL–close
'if you **close** the door (of your house), the blessing is also **closed off** [(because) guests cannot come into (your) house]' [081110-008-CvNP.0096]

Of the 38 ***TER-***prefixed bivalent verbs, one Papuan Malay consultant classified four as borrowings from Standard Indonesian (SI-borrowings) (see language internal factor (g) on p. 115 in §3.1.1), namely ***terbuat*** 'be made', ***terpengaru*** 'be influenced', ***tercinta*** 'be beloved', and ***terwesel*** 'be transferred' (in Table 3 these items are underlined). The same consultant also stated that Papuan Malay speakers usually employ the respective bases rather than the prefixed forms. One such contrastive set of examples is given in (5) and (6). Instead of using the prefixed form ***terpengaru*** 'be influenced', as in (5), speakers more often employ the base ***pengaru*** 'influence' in the sense of 'be influenced', as in (6).

(5) … tapi de ana juga cepat ikut **ter–pengaru**
but 3SG child also be.fast follow ACL–influence
'… but he/she, a kid, also quickly follows (others) to **be influenced**' [080917-010-CvEx.0001]

(6) de su **pengaru** dengang orang~orang yang minum
3SG already influence with RDP~person REL drink
'he has already **been influenced** by people who drink' [080919-007-CvNP.0018]

3.1.2.3. Prefixed items derived from monovalent verbal bases

The corpus contains five ***TER-***prefixed lexemes (with 14 tokens) with monovalent verbal bases, as listed in Table 4. Contrasting with the affixation of bivalent bases, ***TER-***affixation of monovalent bases is not a valency-changing operation, nor does it derive verbs with non-agent arguments. Instead, ***TER-*** downplays the agentivity of its arguments by deriving monovalent verbs which denote accidental or unintentional states or actions, such as ***terlambat*** 'be late' or ***tersendiri*** 'be separate'. Thereby, the prefix reduces rather than decreases valency. All five lexemes are low frequency words, attested with less than 20 tokens. Moreover, the token frequencies for the respective bases are (much) higher for four of the five derived words.

Table 4: Affixation with ***TER-*** of monovalent verbal bases

BW	Gloss	Item	Gloss	***TER-*** #	BW #
lambat	'be slow'	***terlambat***	'be late'	10	3
sendiri	'be alone'	***tersendiri***	'be separate'	1	232
biasa	'be used to'	***terbiasa***	'be accustomed'	1	186

BW	Gloss	Item	Gloss	***TER-*** #	BW #
jatu	'fall'	***terjatu***	'be dropped, fall'	1	64
sala	'be wrong'	***tasala***	'be mistaken'	1	42

Two items indicating uncontrolled and/or unexpected actions are given in context: ***terjatu*** 'be dropped, fall' in (7) and ***terlambat*** 'be late' in (8). Both examples, along with the example in (9), show that the verbal valency is not further decreased and that the derivation does not result in a loss of agentivity. That is, the referents of the derived verbs ***terjatu*** 'be dropped, fall' and ***terlambat*** 'be late' and the referents of the bases ***jatu*** 'fall' and ***lambat*** 'be slow', respectively, have the same semantic functions which is typically that of an agent. With ***TER-***prefixed verbs, however, the agentivity of the referents is downplayed.

(7) dia **ter–jatu** de **jatu** baru motor tindis dia
3SG ACL–fall 3SG fall and.then motorbike overlap 3SG
'he **fell (off unexpectedly)**, he **fell** (off), and then the motorbike crushed him' [080923-010-CvNP.0012]

(8) kaka tadi **ter–lambat** karna lagi ada duka
oSb earlier ACL–be.slow because again exist grief
'a short while ago I ('older brother') was **(unintentionally) late** because there was (still) mourning (going on)' [080918-001-CvNP.0003]

(9) kalo Niwerawar Aruswar nanti dia agak **lambat** sedikit
if Niwerawar Aruswar very.soon 3SG rather be.slow few
[About a street construction project:] 'as for (the area of) Niwerawar (and) Aruswar, (there) it ('the bulldozer') will be somewhat **slow**' [081006-033-Cv.0051]

3.1.2.4. Variables of the communicative event

To explore the issue of productivity ***TER-*** further, a domain analysis was conducted which focused on the variables of speaker education levels, topics, and role-relations (for details see 'Language external factors' on p. 116 in §3.1.1). In all, 43 ***TER-***prefixed items, totaling 167 tokens, were examined:

- 38 prefixed items derived from bivalent verbal bases (153 tokens)
- Five prefixed items derived from monovalent verbal bases (14 tokens)

For the 43 prefixed lexemes, most tokens (143/167 – 86%) can be accounted for in terms of speaker education levels, topics, and/or role-relations. The remaining 24/167 tokens (14%), however, cannot be explained in terms of these variables of the communicative event. These tokens occurred when less-educated speakers (-EDC-SPK) conversed with fellow-Papuans of equally low social standing (-STAT) about LOW topics, that is, casual daily-life issues.[98] (See Table 5 and Chart 1.)

[98] As mentioned under Factor 3 'Relationships between interlocutors' in §1.5.1, all of the recorded less-educated speakers belonged to the group of Papuans with lower social status

If the prefixed items were the result of a productive affixation process, one would expect the percentage of tokens that cannot be explained in terms of speaker education levels, topics, and/or role-relations to be much higher than 14%. Instead, most tokens (86%) seem to be conditioned by these variables of the communicative event. These findings do not support the conclusion that the respective lexemes result from a productive affixation process. Instead, they appear to be code-switches with Indonesian.

Table 5 and Chart 1 (p. 126) present the token frequencies for ***TER-***prefixed lexemes by speakers and topics/interlocutors. Before discussing the data in more detail, the layouts of Table 5 and Chart 1 are explained.

Table 5: Token frequencies for ***TER-***prefixed lexemes with bi- and monovalent verbal bases by speakers, topics, and interlocutors (43 items)[99]

	Topics (TOP)				Interlocutors (ILCT)			Tokens
Prefixed lexemes with bivalent bases (38 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	6	10	10	15	---	---	9	50
-EDC-SPK	2	1	26	---	45	**23**	6	103
Subtotal	8	11	36	15	45	**23**	15	153
Prefixed lexemes with monovalent bases (5 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	0	0	1	4	---	---	0	5
-EDC-SPK	0	1	5	---	2	**1**	0	9
Subtotal	0	1	6	4	2	**1**	0	14
TOTAL (42 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	6	10	11	19	---	---	9	55
-EDC-SPK	2	2	31	---	47	**24**	6	112
Total	8	12	42	19	47	**24**	15	**167**

Table 5 is divided into three major parts. The top part lists the token frequencies for prefixed lexemes with bivalent bases, while the middle part gives the frequencies for prefixed lexemes with monovalent bases. The bottom part gives the frequencies for all verbal bases. The layout of each of these parts represents the three variables of speaker education levels, topics, and role relations (this layout also applies to the tables and charts presented in §3.1.3.3, §3.1.4.4, and §3.1.8). The token frequencies according to the variable 'Speaker education levels' are given in the rows labeled

(-STAT), while the recorded Papuans with higher social status (-STAT), such as teachers, government officials, or pastors, were all better educated.

[99] Abbreviations: SPK = speaker, +EDC = better educated, –EDC = less educated, POL = politics, EDC = education, REL = religion, LOW = LOW topics, +STAT = higher social status, -STAT = lower social status, OUTSD = outsider.

"+EDC-SPK" and "-EDC-SPK", while the token frequencies according to the variables 'Topic' and 'Role-relations' are presented in the columns labeled "Topics (TOP)" and "Interlocutors (ILCT)", respectively. The token frequencies by speaker education levels are presented in two rows: the first row labeled "+EDC-SPK" gives the token frequencies for better-educated speakers while the second row labeled "-EDC-SPK" lists the token frequencies for less-educated speakers. The token frequencies by topics are presented in the first four columns. The three columns headed POL, EDC, and REL list the frequencies for tokens when speakers conversed about the HIGH topics of politics, education, and religion, respectively. The column headed LOW lists the number of tokens produced during conversations about LOW topics, that is, casual daily-life issues. The token frequencies by role-relations are presented in the next three columns. The columns headed with +STAT, -STAT, and OUTSD give the number of tokens produced during conversations with fellow-Papuans of higher social standing (+STAT), fellow-Papuans of lower social standing (-STAT), and group outsiders (OUTSD), respectively.

The layout of Table 5 is based on four assumptions. First, when discussing HIGH topics, the language behavior of Papuans is likely to show influences from Indonesian, regardless of their own education levels and also regardless of the education levels of their fellow-Papuan interlocutors. Therefore, these token frequencies are totaled in the respective 'Topics' cells and not broken down according to the education levels of their interlocutors. For ***TER-***prefixed lexemes with bivalent bases, the respective token frequencies for better-educated speakers (+EDC-SPK) are as follows: 6 tokens for discussions about politics (POL), 10 about education (EDC), and 10 tokens about religion (REL). For less-educated speakers (-EDC-SPK) the respective frequencies are 2,1, and 26 tokens. (See the left top part of Table 5).

Second, when discussing LOW topics, the language behavior of better-educated speakers (+EDC-SPK) is presumably not affected by the social standing of their fellow-Papuan interlocutors, given that they already have the general tendency to "dress-up" their Papuan Malay with Indonesian features. Therefore, these token frequencies are totaled in the LOW-topic cell of the +EDC-SPK row. That is, in this total are included the token frequencies for interactions with interlocutors of equally high social standing (+STAT) and with those of lower status (-STAT). The columns to the right of the LOW-topic column give the token frequencies according to the social status of the speakers' interlocutors. However, given that for the better-educated speakers (+EDC-SPK), the total in LOW-topic cell includes both +STAT and -STAT interlocutors, the respective cells for +STAT and -STAT interlocutors are left empty. For ***TER-***prefixed lexemes with bivalent bases, the respective token frequency is 15 (see the LOW-topic column in the top part of Table 5), while the +STAT and -STAT cells to the right are left empty.

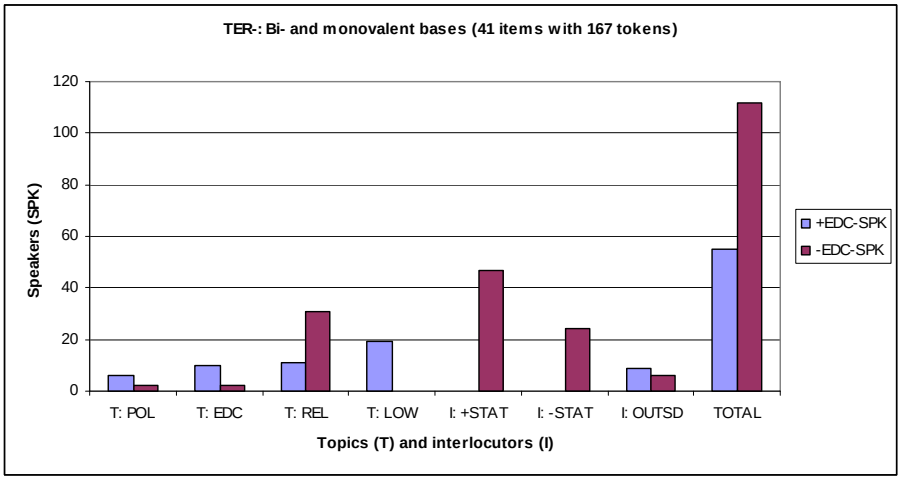
Third, when discussing LOW topics, the language behavior of less-educated speakers (-EDC-SPK) is likely to be affected by the status of their fellow-Papuan interlocutors. Therefore, these total token frequencies are not totaled in the LOW-topic cell of the -EDC-SPK row. Instead the LOW-topic token frequencies are broken down according to the status of their fellow-Papuan interlocutors; hence, the respective LOW-topic cell is left empty. For ***TER-***prefixed lexemes with bivalent bases, the token respective frequencies are 45 for +STAT Papuan interlocutors and 23

for -STAT Papuan interlocutors (see the +STAT- and -STAT -interlocutor columns in the top part of Table 5), while the LOW-topic cell to the left is left empty.

Fourth, the language behavior of better and less-educated speakers is likely to be affected when they converse with a non-Papuan outsider, regardless of the topic under discussion. Therefore, all tokens produced during conversations with an outsider, namely the author, are totaled in the OUTSD column of the +EDC-SPK and -EDC-SPK rows. For ***TER-***prefixed lexemes with bivalent bases, this token frequency is 9 for better-educated speakers and 6 for less-educated speakers (see the OUTSD-interlocutor column in the top part of Table 5)

Chart 1 gives a graphic representation of the data listed in Table 5. The horizontal category (X) axis presents the different categories according to which the token frequencies are listed, that is, the four topic categories and the three interlocutor categories. The vertical value (Y) axis gives the token totals for each of these categories, according to speaker education levels. The light (or blue) shaded columns denote the token frequencies for the better-educated speakers, while the dark (or red) shaded columns indicate the frequencies for the less-educated speakers.

Chart 1: Token frequencies for ***TER-***prefixed lexemes with bi- and monovalent verbal bases by speakers, topics, and interlocutors



The data given in Table 8 and Chart 2 show that for the 43 ***TER-***prefixed lexemes, most tokens (143/167 – 86%) can be explained in terms of speaker education levels, topics, and/or role-relations between the speakers and their interlocutors; this total includes 130/153 tokens (85%) with bivalent bases, and 13/14 tokens (93%) with monovalent bases.

Only 55/167 tokens (33%) were produced by better-educated speakers (+EDC-SPK) while most tokens (112/167 – 67%) were produced by less-educated speakers (-EDC-SPK). The +EDC-SPK produced half of their tokens (27/55 – 49%) during discussions about HIGH topics, that is, political, educational or religious affairs (POL, EDC and REL, respectively). Another 19 tokens (35%) occurred during conversations with fellow-Papuans (both +STAT and -STAT speakers) about LOW topics. The

remaining nine tokens (16%) occurred while conversing with an outsider, namely the author (OUTSD).

The -EDC-SPK produced most of their tokens (53/112 – 47%) while discussing LOW topics with +STAT speakers (47 tokens) or the author (6 tokens). Another 35/112 tokens (31%) where produced during discussions about HIGH topics. The remaining 24/112 tokens (21%) occurred when -EDC-SPK discussed LOW topics with -STAT Papuans, and therefore cannot be explained in terms of speaker education levels, topics, and/or role-relations. This total of 24 tokens refers to 14% of all 167 ***TER-***tokens, including 23/153 tokens (15%) with bivalent bases and 1/14 tokens (7%) with monovalent bases.[100]

3.1.2.5. Summary and conclusions

For most of the derived verbs with bivalent bases, the data suggest a productive form-function relationship between the derived lexemes and their bases. This conclusion is based on four observations: (1) the valency-decreasing or -reducing function of ***TER-*** of removing or downplaying agent-like participants, (2) the transparent form-function relationships between derived lexemes and bases, (3) the large number of low frequency words and small number of high frequency words, and (4) the relative token frequencies with most bases having higher frequencies than the affixed lexemes.

For the prefixed verbs with monovalent bases, the derivation process also seems to be productive, given (1) the transparent form-function relationships between derived lexemes and bases, (2) the comparatively large number of low frequency words and small number of high frequency words, and (3) the relative token frequencies with most bases having higher frequencies than the affixed lexemes. However, the low type frequency, with only five derived verbs, suggests that ***TER-***prefixation of monovalent bases plays a minor role.

As for the speech situations during which the derived lexemes occurred, a sizable number of verbs with bivalent bases cannot be explained in terms of pertinent variables of the communicative event. Most tokens, however, including those with bivalent bases, seem to be conditioned by the variables of speaker education levels, topics, and/or role-relations and therefore are best explained as code-switches with Indonesian.

These findings suggest that ***TER-***affixation is a productive process to derive monovalent verbs that denote accidental or unintentional actions. The degree of productivity appears to be limited, however, given that most of the attested tokens are best explained as code-switches with Indonesian.

---

[100] As for the 21 hapaxes (17 with bivalent and four with monovalent bases), 18 appear to be conditioned by the variables of speaker education levels, topics, and/or role-relations, and therefore seems to be code-switches with Indonesian. This leaves only three hapaxes (with bivalent bases) that are unaccounted for in terms of language external factors and that might result from a productive derivation process. For three hapaxes P=0.0180 as opposed to P=0. 1257 for 21 hapaxes (N=167).

3.1.3. Suffix *-ang* 'PAT'

Affixation with ***-ang*** 'PAT' typically derives nominals from verbal bases. The derived nouns denote the patient or result of the action specified by the verbal base, as illustrated in (10). Some lexical items are also derived from nominal and quantifier bases. The derivation process seems to be productive to some degree, as discussed below.

(10) **pake–ang** itu basa smua
use–PAT D.DIST be.wet all
'all those **clothes** were wet' [080917-008-NP.0139]

Suffix ***-ang*** is a reflex of Proto Malayic ***-ʌn**, which "was a noun-forming suffix occurring on the basis of VTRs and denoting the goal or result of an act" (Adelaar 1992: 174). In Standard Malay, when affixed to monovalent bases, the suffix designates "something that has the quality of" the monovalent base, while with transitive bases it denotes the "goal or result of an action, or place where the action takes place" or "the instrument" (1992: 172–173). As for the eastern Malay varieties, the suffix is only mentioned for Ambon Malay. Also realized as ***-ang***, it "refers to the object of the transitive verb or an instrument used in an act of V" (van Minde 1997: 106). It is left unclear, however, whether and to what degree the Ambon Malay suffix is productive. These observations are again an indication of the distinct history of Papuan Malay vis-à-vis the other Malay varieties, discussed in §1.2. Moreover, the similarities between Papuan Malay and Ambon Malay reflect the link between both speech communities, also discussed in §1.2.

The present corpus contains 84 nouns (441 tokens) suffixed with ***-ang***:[101]

1. Nouns with verbal bases (69 items with 403 tokens)
2. Nouns with nominal and quantifier bases (15 items with 38 tokens)

The corpus also includes 28 formally complex words that have non-compositional semantics, such as ***kasiang*** 'pity', ***lapangang*** 'field', or ***grakang*** 'movement'.

Suffixed items with verbal bases are examined in §3.1.3.1, and those with nominal bases in §3.1.3.2. Variables of the communicative event that may impact the use of ***-ang*** are explored in §3.1.3.3. The main findings on suffix ***-ang*** are summarized and evaluated in §3.1.3.4.

3.1.3.1. Suffixed items derived from verbal bases

The present corpus contains 69 ***-ang***-suffixed items (with 403 tokens) with verbal bases, including bases such as bivalent ***pake*** 'use', monovalent dynamic ***jalang*** 'walk', or monovalent stative ***dulu*** 'be prior'. Affixation with ***-ang*** typically derives nouns that denote the object of the action indicated by the verbal base.

Derived words with token frequencies of five or more are listed in Table 6. Most of the affixed lexemes are low frequency words (63 lexemes, attested with less than

[101] The 84 nouns include 28 hapaxes (P=0.0635); the 69 nouns with verbal bases include 23 hapaxes (P=0.0571); the 15 nouns with nominal or quantifier bases include five hapaxes (P=0.1316).

20 tokens). Moreover, the token frequencies for the respective bases are (much) higher for most of the derived words (64 lexemes). While all 69 derived lexemes are structurally nouns, three of them have other than nominal functions in their actual uses as illustrated in (13) to (14): ***jualang*** 'merchandise', ***latiang*** 'practice', and ***duluang*** 'prior to others'.

Seven of the 69 lexemes were tentatively classified as borrowings from Standard Indonesian (SI-borrowings) (for more details see language internal factor (g) on p. 115 in §3.1.1). As their token frequencies are four or less, they are not included in Table 6.

Table 6: Affixation with ***-ang*** of verbal bases

BW	Gloss	Item	Gloss	***-ang*** #	BW #
makang	'eat'	***makangang***	'food'	57	414
pake	'use'	***pakeang***	'clothes'	38	218
dulu	'be prior'	***duluang***	'prior to others'	29	351
bagi	'divide'	***bagiang***	'part'	28	63
pikir	'think'	***pikirang***	'thought'	23	102
uji	'test'	***ujiang***	'examination'	21	1
lati	'practice'	***latiang***	'practice'	17	3
kubur	'burry'	***kuburang***	'grave'	14	8
atur	'arrange'	***aturang***	'regulation'	8	24
ikat	'tie up'	***ikatang***	'tie'	8	14
jual	'sell'	***jualang***	'merchandise'	8	14
turung	'descend'	***turungang***	'descendant'	8	192
ulang	'repeat'	***ulangang***	'repetition'	8	16
bantu	'help'	***bantuang***	'help'	7	34
alas	'put down as base'	***alasang***	'reason'	6	7
bangung	'build'	***bangungang***	'building'	6	25
libur	'have vacation'	***liburang***	'vacation'	6	10
campur	'mix'	***campurang***	'mixture'	5	5
jalang	'walk'	***jalangang***	'route'	5	485
lapor	'report'	***laporang***	'report'	5	14
tulis	'write'	***tulisang***	'writing'	5	12

Affixing verbal bases with ***-ang*** typically derives nouns that denote the object of the action specified by the verbal base. The suffixed nouns include patients such as ***makangang*** 'that which is eaten' or 'food', or results such as ***bagiang*** 'that which is divided' or 'part'. "Objective nominalization" that derives "nouns designating the result, or the typical or 'cognate' object of an action" has also been observed for other languages (Comrie and Thompson 2007: 340). This polysemy can be explained in terms of a "domain shift" in that "one may go from one semantic domain to another, related one, and thus derive new interpretations" (Booij 2007: 221). Hence, suffix ***-ang*** is glossed as 'PAT' ('patient') in the sense of 'patients or results which are BASE-ed'.

Two derived nouns together with their bases are given in context: ***makangang*** ‘food’ with its bivalent base ***makang*** ‘eat’ in (11), and ***jalangang*** ‘route’ with its monovalent base ***jalang*** ‘walk’ in (12).

(11) maytua bilang, **makang** karna **makang–ang** suda masak
wife say eat because eat–PAT already cook
‘(my) wife said, ‘**eat**, because the **food** had already been cooked’ [080919-004-NP.0039]

(12) trus kitong dua pulang, sampe di **jalang–ang** sa istirahat,
next 1PL two go.home reach at walk–PAT 1SG rest
de bilang, kitong dua **jalang** suda!
3SG say 1PL two walk already
‘and then we two went home, on the **way** I rested, he said, ‘let the two of us **walk** (on)!’’ [081015-005-NP.0036]

Some of the suffixed items, listed in Table 6, differ from the other suffixed items, as for example ***jual-ang*** ‘sell-PAT’, or ***dulu-ang*** ‘be.prior-PAT’. Suffixed with ***-ang***, these items are structurally nouns. In a sentence, however, ***jualang*** also functions as the verb ‘sell’ in the same way as its base ***jual*** ‘sell’, as shown in (13). Likewise (14) ***duluang*** ‘be.prior-PAT’ is used as the verb ‘be prior to others’.

(13) mama saya pergi **jual** pinang, sa pu mama **jual–ang**
mother 1SG go sell betel.nut 1SG POSS mother sell–PAT
pinang
betel.nut
‘my mother went to **sell** betel nuts, my mother **sells** betel nuts’ [081014-014-NP.0002]

(14) kaka sa **dulu–ang** dalam dulu e?
oSb 1SG be.prior–PAT inside be.prior eh
‘older brother, I (go) ahead, (I go) inside for now, eh?’ [080918-001-CvNP.0030]

3.1.3.2. Suffixed items derived from nominal and quantifier bases

The corpus contains 13 ***-ang***-suffixed lexemes with nominal bases (36 tokens) and two derived lexemes with quantifier bases (2 tokens), as listed in Table 7. In most cases, the bases and the derived nouns differ in their semantics. In some cases, the affixed nouns designate a magnification of the base, such as ***laut*** ‘sea’ and ***lautang*** ‘ocean’, or ***ruang*** ‘room’ and ***ruangang*** ‘large room’. In some cases, the meanings of the derived nouns are an extension of the meanings of their bases with suffix ***-ang*** having generalizing function, as for instance ***ana*** ‘child and ***anaang*** ‘offspring’, or ***musim*** ‘season’ and ***musimang*** ‘each season’. In yet other cases, the affixed nouns have unpredictable meanings compared to the semantics of their bases, such as ***rambut*** ‘hair’ and ***rambutang*** ‘rambutan’, and ***obat*** ‘medicine’ and ***obatang*** ‘magic

spell'. And in a few cases, the base and the derived noun have the same semantics, as in ***pasang*** 'pair' and ***pasangang*** 'pair' or ***pangkal*** 'base' and ***pangkalang*** 'base'.

All 13 derived lexemes are low frequency words, attested with less than 20 tokens. Moreover, the token frequencies for the respective bases are (much) higher for most of the derived words (10 lexemes); for one lexeme, the base is unattested in the present corpus, although it does exist. Four of the 15 derived nouns were tentatively classified as SI-borrowings; in Table 7 these items are underlined (for more details see language internal factor (g) on p. 115 in §3.1.1).

Table 7: Affixation with ***-ang*** of nominal and quantifier bases

BW	Gloss	Item	Gloss	***-ang*** #	BW #
bayang	'image'	***bayangang***	'shadow'	6	2
ana	'child'	***anaang***	'offspring'	4	741
tingkat	'floor'	***tingkatang***	'level'	4	5
hukum	'law'	***hukumang***	'punishment'	4	3
rambut	'hair'	***rambutang***	'rambutan'	3	23
obat	'medicine'	***obatang***	'magic spell'	3	9
pasang	'pair'	***pasangang***	'pair'	3	2
laut	'sea'	***lautang***	'ocean'	2	68
pinggir	'border'	***pinggirang***	'edges'	2	23
ruang	'room'	***ruangang***	'large room'	2	3
kandung	'womb'	***kandungang***	'womb'	1	8
musim	'season'	***musimang***	'each season'	1	5
pangkal	'base'	***pangkalang***	'base'	1	0
pulu	'tens'	***puluang***	'tens'	1	78
ratus	'hundreds'	***ratusang***	'hundreds'	1	34

The data listed in Table 7 show that most of the nominal bases and affixed nouns differ in their semantics. The magnifying function of suffix ***-ang*** is illustrated in (15) and (16), the generalizing function in (17), and its unpredictable semantics in (18) and (19).

The magnifying function of ***-ang*** is demonstrated with ***laut*** 'sea' in (15) and ***lautang*** 'ocean' in (16). While ***laut*** refers to the 'sea' close to the coast, ***lautang*** denotes the open and deep 'ocean' off the coast.

Suffix ***-ang***: Magnifying function

(15) dong dua pergi mancing di **laut**
3PL two go fish.with.rod at sea
'the two of them went fishing on the **sea**' [081109-005-JR.0005]

(16) banyak mati … di pulow~pulow banyak mati di **laut–ang**
sea–PAT
many die at RDP~island many die at ocean
'many died … on the islands, many died on the (open) **ocean**' [081029-002-Cv.0025]

The generalizing function of ***-ang*** is illustrated with ***ana*** 'child' and ***anaang*** 'offspring' in (17).

Suffix ***-ang***: Generalizing function

(17) kalo mo antar **ana** **prempuang** ke **ana** **laki~laki** … kitorang
if want bring child woman to child RDP~husband 1PL
itu harus … bawa **ana–ang** pinang **ana–ang** sagu
child–PAT child–PAT
D.DIST have.to bring offspring betel.nut offspring sago
[About wedding preparations:] 'if we want to bring our **daughter** to (their) **son** … we have to … bring betel nut **seedlings** (and) sago **seedlings**' (Lit. 'female/male **child**; betel nut/sago **offspring**') [081110-005-CvPr.0055-0057]

In some cases, the semantics of the affixed nouns are unpredictable, although a connection between the base word and the derived word can still be seen. This is demonstrated with ***rambut*** 'hair' in (18) and ***rambutang*** 'rambutan' in (19), which refers to the fruit of the rambutan tree (*Nephelium lappaceum*). The leathery reddish skin of the fruit is covered with numerous hairy protuberances, which is depicted by the label ***rambut-ang*** 'hair-PAT'.

Suffix ***-ang***: Unpredictable semantics

(18) sa mo cuci de pu **rambut**
1SG want wash 3SG POSS hair
'I want to wash her **hair**' [081025-001-CvHt.0006]

(19) di sini ada jambu di sini ada ada **rambut–ang**
hair–PAT
at L.PROX exist rose.apple at L.PROX exist exist rambutan
'here are rose apples, here are are **rambutan**' [081029-001-Cv.0006]

3.1.3.3. Variables of the communicative event

To further investigate the issue of productivity of ***-ang***, a domain analysis was conducted which focused on the variables of speaker education levels, topics, and/or role-relations (for details see 'Language external factors' on p. 116 in §3.1.1). In all, 84 items suffixed with ***-ang***, totaling 441 tokens, were investigated:

- 69 suffixed items derived from verbal bases (403 tokens)
- 15 suffixed items derived from nominal and quantifier bases (38 tokens)

For the 84 suffixed lexemes, 352/441 all tokens (80%) can be accounted for in terms of speaker education levels, topics, and/or role-relations. The remaining 89/441 tokens (20%) occurred when less-educated speakers (-EDC-SPK) talked with fellow-Papuans of equally low social standing (-STAT) about LOW topics.[102] (See Table 8 and Chart 2.)

[102] As mentioned under Factor 3 'Relationships between interlocutors' in §1.5.1, all of the recorded less-educated speakers belonged to the group of Papuans with lower social status

That is, a considerable number of tokens (20%) cannot be explained in terms of these variables of the communicative event. Therefore, it cannot be concluded that the respective lexemes are code-switches with Indonesian. This total of 89/441 tokens (20%) includes 80/403 tokens (20%) with verbal bases and 9/38 tokens (24%) with nominal or quantifier bases. The vast majority of ***-ang***-suffixed tokens (352/441 – 80%), however, seems to be conditioned by variables of the communicative event.

As for the rather high number of unaccounted tokens with nominal or quantifier bases (9/38 – 24%), one observation is made. Four of the nine tokens refer to the same lexeme produced by the same speaker during three conversations about the same topic, namely the death of a young mother. This speaker has a reputation of speaking incoherently due to his unsuccessful attempts to approximate Standard Indonesian. Excluding these four tokens brings down the number of unaccounted lexemes to 13% (5/38). If affixation of nominal bases was a productive process, however, one would expect this percentage to be much higher. In turn, this finding does not support the conclusion that the suffixed lexemes with nominal or quantifier bases result from a productive derivation process. Instead, they seem to be code-switches with Indonesian.

The data presented in Table 8 and Chart 2 are discussed in more detail below.

Table 8: Token frequencies for ***-ang***-suffixed lexemes with verbal and nominal bases by speakers, topics, and interlocutors (84 items)[103]

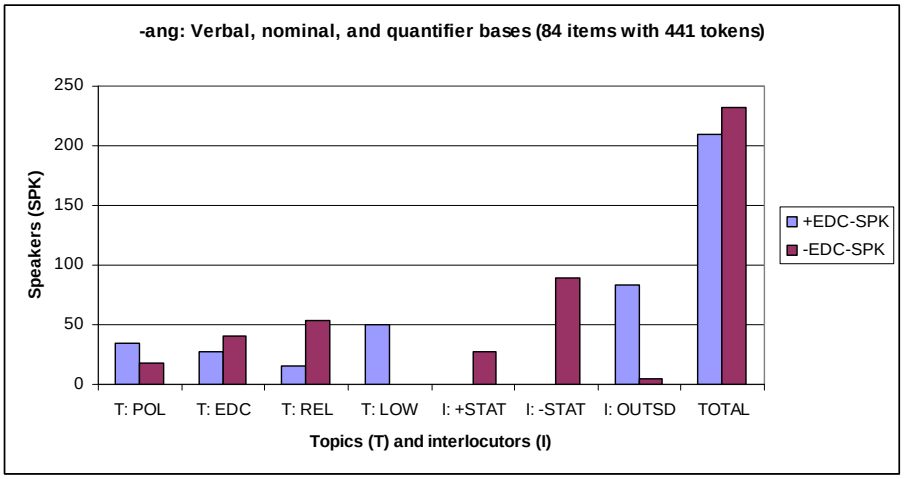
	Topics (TOP)				Interlocutors (ILCT)			Tokens
Suffixed lexemes with verbal bases (69 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	30	26	15	46	---	---	75	192
-EDC-SPK	15	40	47	---	26	**80**	3	211
Subtotal	45	66	62	46	26	**80**	78	403
Suffixed lexemes with nominal or quantifier bases (15 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	4	1	0	4	---	---	8	17
-EDC-SPK	3	0	6	---	1	**9**	2	20
Subtotal	7	1	6	4	1	**9**	10	38

(-STAT), while the recorded Papuans with higher social status (-STAT), such as teachers, government officials, or pastors, were all better educated.

[103] Abbreviations: SPK = speaker, +EDC = better educated, –EDC = less educated, POL = politics, EDC = education, REL = religion, LOW = LOW topics, +STAT = higher social status, -STAT = lower social status, OUTSD = outsider.

	Topics (TOP)				Interlocutors (ILCT)			Tokens
TOTAL (84 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	34	27	15	50	---	---	83	209
-EDC-SPK	18	40	53	---	27	**89**	5	232
Total	52	67	68	50	27	**89**	88	**441**

Chart 2: Token frequencies for ***-ang***-suffixed lexemes with verbal and nominal bases by speakers, topics, and interlocutors



The data given in Table 8 and Chart 2 show that for the 84 ***-ang***-suffixed lexemes, 352/441 all tokens (80%) can be explained in terms of speaker education levels, topics, and/or role-relations between the speakers and their interlocutors; this includes 323/403 tokens (80%) with verbal bases, and 29/38 tokens (76%) with nominal bases.

The better-educated speakers (+EDC-SPK) produced 209/441 tokens (47%), while the less-educated speakers (-EDC-SPK) produced 232/441 (53%) tokens.

In terms of topics (TOP), less than half of the tokens (187/441 – 42%) occurred during conversations about HIGH topics, that is, political, educational or religious affairs (POL, EDC and REL, respectively). This includes 76/209 +EDC-SPK tokens (36%) and 111/232 -EDC-SPK tokens (48%). This leaves 254/441 tokens (58%) that were produced when the interlocutors discussed LOW topics: 133/254 tokens were produced by +EDC-SPK (53%) and 121/254 tokens by -EDC-SPK tokens (48%).

The 133 tokens produced by +EDC-SPK when discussing LOW topics are divided as follows: 50 tokens occurred during conversations with fellow-Papuans (both +STAT and -STAT speakers) (that is, 50/209 +EDC-SPK tokens – 24%) and 83 tokens were produced during conversations with an outsider, namely the author (OUTSD) (that is, 83/209 +EDC-SPK tokens – 40%). The 121 tokens produced by -EDC-SPK

when discussing LOW topics are distributed as follows. When conversing with +STAT Papuans, 27 tokens were produced (that is, 27/232 -EDC-SPK tokens – 12%), while five tokens occurred during conversations with the author (that is, 5/232 -EDC-SPK tokens – 2%). The remaining 89/232 tokens (38%) occurred when -EDC-SPK discussed LOW topics with -STAT Papuans, and therefore cannot be explained in terms of speaker education levels, topics, and/or role-relations. This total of 89 tokens refers to 20% of all 441 ***-ang*** tokens. It includes 80/403 tokens (20%) with verbal bases and 9/38 tokens (24%) with nominal or quantifier bases.[104]

3.1.3.4. Summary and conclusions

Suffix ***-ang*** is polyfunctional in that it derives nouns from verbal, nominal, and quantifier bases. This polyfunctionality suggests that affixation with ***-ang*** is a somewhat productive process (see language internal factor (d) on p. 114 in §3.1.1).

Concerning ***-ang***-affixation of verbal bases, four other observations support this conclusion: (1) the transparent form-function relationship between the derived nouns and their respective bases, (2) the large number of low frequency words and small number of high frequency words, and (3) the relative token frequencies with most bases having higher frequencies than the affixed lexemes, and (4) the low number of derived lexemes tentatively classified as SI-borrowings.

To a lesser extent, the same observations apply to ***-ang***-affixation of nominal bases: (1) the form-function relationships between derived lexemes and bases is more or less transparent, (2) all derived lexemes are low frequency words, (3) most bases have higher token frequencies than the affixed lexemes, and (3) the number of derived lexemes tentatively classified as SI-borrowings is rather low. These findings suggest that ***-ang***-affixation of nominal bases is a somewhat productive process.

With respect to the speech situations during which the derived nouns occurred, the following patterns emerge. For affixed nouns with verbal bases, one fifth of the attested tokens cannot be accounted for in terms of pertinent variables of the communicative event; that is, for these items there are no indications that they are code-switches with Indonesian. However, the vast majority of tokens with verbal bases (80%) seem to be conditioned by the variables of speaker education levels, topics, and/or role-relations and are best explained as code-switches with Indonesian. The same applies to nouns with nominal bases for which most tokens also appear to be conditioned by the three mentioned variables of the communicative event. Hence, these items are also best explained as code-switches with Indonesian. These findings suggest that ***-ang***-affixation is a productive process to derive nouns from verbal and nominal bases. The degree of productivity appears to be limited, however, as most tokens seem to be code-switches with Indonesian.

---

[104] As for the 28 hapaxes (23 with verbal bases, and three with nominal bases), 17 appear to be conditioned by the variables of speaker education levels, topics, and/or role-relations, and therefore are best explained as code-switches with Indonesian. This leaves 11 hapaxes that are unaccounted for in terms of language external factors and that might be the result of a productive word-formation process. For 11 hapaxes, P=0.0249 as opposed to P=0.0635 for 28 hapaxes (N=441). The total of 11 hapaxes includes nine with verbal bases (N=403, P=0.0223) and two with nominal or quantifier bases (N=38, P=0.0526).

3.1.4. Prefix *PE(N)-* 'AG'

Affixation with ***PE(N)-*** 'AG' typically derives nominals from verbal bases.[105] The derived nouns denote the agent or instrument of the action, event, or state specified by the verbal base, as in (20). Some lexemes are also derived from nominal bases. The affixation process appears to be marginally productive, at best, as discussed below.

(20) pokoknya orang **pen–datang** pulang
the.main.thing.is person AG–come go.home
'the main thing is (that) the **strangers** return home' (Lit. '**the one who comes**') [081029-005-Cv.0048]

Suffix ***PE(N)-*** is a reflex of Proto Malayic ****pAN-***, which "formed deverbal nouns that were used attributively, predicatively, and in prepositional phrases, and that had a nominal as head or subject. They denoted a purpose or instrument when prefixed to VDIs and VTRs. Moreover, ****pAN-*** denoted an inclination or characteristic when prefixed to VSIs"[106] (Adelaar 1992: 193). In Standard Malay, derived lexemes with a monovalent base "denote a characteristic" while forms with a bivalent base "usually denote an actor or instrument" or "a goal or result, or they form an abstract noun. Furthermore ***pa*N-** forms are used attributively, and, on the basis of VSIs, they can function as VSIs" (1992: 183).

In some of the eastern Malay varieties, the prefix is also found. In Ambon Malay, the prefix occurs but it is unproductive (van Minde 1997: 109). In Manado Malay ***paŋ-*** also occurs and is productive (in addition, a non-unproductive form ***pa-*** exists) (Stoel 2005: 18, 24). Likewise, in North Moluccan / Ternate Malay ***pa(*N*)-*** occurs, but its status is uncertain. While Voorhoeve (1983: 4) maintains that it "is no longer morphologically distinct", Litamahuputty (2012: 30) states that prefixation "with pang-" is a "productive morphological" process. In these varieties, the prefix usually denotes the actor or instrument of the event expressed by the base. In addition, however, some of prefixed forms can receive a verbal reading, as discussed in more detail in §3.1.4.2.

The present corpus contains 34 nouns (186 tokens) prefixed with ***PE(N)-***:[107]

1. Nouns with verbal bases (29 items with 153 tokens)
2. Nouns with nominal bases (five items with 33 tokens)

The corpus also contains nine formally complex words with non-compositional semantics, such as ***peserta*** 'participant' or ***panggayu*** '(a/to) paddle'.

Before discussing ***PE(N)-***affixation of verbal bases in §3.1.4.2 and of nominal bases in §3.1.4.3, the allomorphy of ***PE(N)-*** is investigated in §3.1.4.1. Two variables of the communicative event that may impact the use of ***PE(N)-*** are explored in

---

[105] The small caps designate an abstract representation of the prefix as it has more than one form of realization, namely the two allomorphs ***pe(N)-*** and ***pa(N)-*** (small-caps *N* represents the different realizations of the nasal) (see §3.1.4.1).

[106] VSI = intransitive stative verb (Adelaar 1992: 8).

[107] The 34 nouns include 11 hapaxes (P=0.0591); the 29 nouns with verbal bases include nine hapaxes (P=0.0588); the five nouns with nominal bases include two hapaxes (P=0.0606).

§3.1.4.4. The main points on prefix ***PE(N)-*** are summarized and evaluated in §3.1.4.5.

3.1.4.1. Allomorphy of ***PE(N)-***

Prefix ***PE(N)-*** has two allomorphs, ***pe(N)-*** and ***pa(N)-*** (small-caps N represents the different realizations of the nasal). The allomorphs are not governed by phonological processes.

The form ***pe(N)-***, in turn, has seven allomorphs that result from morphologically conditioned phonological rules. More specifically, they are conditioned by the word-initial segment of the base word, as shown in Table 9: /pɛm-/, /pɛn-/, /pɛɲ-/, pɛŋ-/, /pɛ-/, /p-/, and /pl-/. The prefix is realized as /pɛm-/ when the initial segment of the base is a bilabial stop. Onset voiced stops are retained, while voiceless stops are deleted. With onset bilabial /m/, the prefix is realized as /pɛ-/. With alveolar stops, the prefix is very commonly realized as /pɛn-/. Again, the onset voiced stop is retained, while the onset voiceless stop is deleted. Alternatively, however, the onset voiceless stop can also be retained, in which case the prefix is realized as /pɛ-/. With onset fricative /s/, the prefix is realized as /pɛɲ-/, with /s/ being deleted. With onset palato-alveolar affricates, ***pe(N)-*** is realized as /pɛn-/. With onset rhotic /r/, the affix is realized as /pɛ-/. With onset velar stops and onset vowels, the prefix is realized as /pɛŋ-/. Finally, when prefixed to ***ajar*** 'teach', ***pe(N)-*** is realized as /pl-/.

Table 9: Realizations of allomorph ***pe(N)-***

pe(N)-base	Orthogr.	Gloss
/pɛm–bantu/	***pembantu***	'house-helper'
/pɛm–pili/	***pemili***	'voter'
/pɛ–muda/	***pemuda***	'youth'
/pɛn–dataŋ/	***pendatang***	'new-comer'
/pɛn–tumpaŋ/	***penumpang***	'passenger'
/pɛ–tugas/	***petugas***	'official'
/pɛɲ–sakit/	***penyakit***	'disease'
/pɛn–tʃuri/	***pencuri***	'thief, to steal (EMPH)'
/pɛn–dʒaga/	***penjaga***	'guard'
/pɛ–rentʃana/	***perencana***	'planner'
/pɛŋ–acara/	***pengacara***	'master of ceremony'
/pɛŋ–ganti/	***pengganti***	'replacement'
/pl–adʒar/	***plajar***	'teacher'

The allomorph ***pa(N)-*** occurs considerably less frequently. Attested are only the four items listed in Table 10 with a total of 18 ***pa(N)-*** tokens. Form ***pa(N)-*** has two attested allomorphs: /pan-/ and /pa-/. The phonological processes involved in the allomorphy are the same as those for ***pe(N)-***, discussed above. For two of the items,

the prefix is alternatively realized as allomorph ***pe(N)-***. Therefore, for each item the token frequencies for ***pa(N)-*** and for ***pe(N)-*** are given. If the prefix is realized with /pe(N)-/ in a greater number of tokens than with /pa(N)-/, then its orthographic representation is ***PE(N)-*** as in ***pencuri*** 'thief, steal (EMPH)'.

Table 10: Realizations of allomorph ***pa(N)-***

pa(N)-base	Orthogr.	Gloss	***pa(N)-*** #	***pe(N)-*** #
/pa–malas/	***pamalas***	'listless person, be very listless'	12	2
/pan–diam/	***pandiam***	'taciturn person, be very quiet'	2	0
/pan–takut/	***panakut***	'coward, feel afraid (of) (EMPH)'	3	0
/pan–tʃuri/	***pencuri***	'thief, steal (EMPH)'	1	11

In realizing the prefix typically as ***pe(N)-*** rather than as ***pa(N)-***, Papuan Malay differs from other eastern Malay varieties such as Ambon Malay (van Minde 1997: 109), Manado Malay (Stoel 2005: 23), North Moluccan / Ternate Malay (Voorhoeve 1983: 4, and Litamahuputty 2012: 30). In these varieties the prefix is always realized as ***pa(N)-***. Instead, the ***PE(N)-***prefixed items have more resemblance with the corresponding items in Standard Indonesian where the prefix is realized as ***pe(N)-***. This is again an indication of the distinct history of Papuan Malay vis-à-vis the other eastern Malay varieties, discussed in §1.2.

3.1.4.2. Prefixed items derived from verbal bases

The present corpus includes 29 ***PE(N)-***prefixed nouns (with 153 tokens) with verbal bases, listed in Table 11. Included are items with verbal bases such as bivalent ***curi*** 'steal', monovalent dynamic ***duduk*** 'sit', or monovalent stative ***muda*** 'be young'. The affixation process derives nouns that designate the subject of the action, event, or state specified by the verbal base.

All but one of the derived words are low frequency words (28 lexemes, attested with less than 20 tokens). In addition, the token frequencies for the respective bases are (much) higher for most of the derived words (24 lexemes). While the 29 prefixed items are structurally nouns, four of them also have verbal functions in their actual uses: ***pemalas*** 'listless person, be very listless', ***pencuri*** 'thief, steal (EMPH)', ***penakut*** 'coward, be very afraid', and ***pendiam*** 'taciturn person, be very quiet'. These items are investigated in more detail in (25) to (28).

Of the 29 derived lexemes, more than half (17 items) were tentatively classified as borrowings from Standard Indonesian (SI-borrowings) (for details see language internal factor (g) on p. 115 in §3.1.1); in Table 11 these items are underlined.

Table 11: Affixation with ***PE(N)-*** of verbal bases

BW	Gloss	Item	Gloss	***PE(N)-*** #	BW #
muda	'be young'	***pemuda***	'youth'	46	24
malas	'be list-less'	***pamalas***	'listless person, be very listless'	14	19
curi	'steal'	***pencuri***	'thief, steal	12	4

BW	Gloss	Item	Gloss	***PE(N)-*** #	BW #
			(EMPH)'		
pimping	'lead'	***pemimping***	'leader'	11	8
datang	'come'	***pendatang***	'new-comer'	10	447
sakit	'be sick'	***penyakit***	'disease'	7	155
duduk	'sit'	***penduduk***	'inhabitant'	5	167
tunggu	'wait'	***penunggu***	'tutelary spirit'	5	92
pili	'choose'	***pemili***	'voter'	5	25
tanggung-jawap	'be responsible'	***penanggung-jawap***	'responsible person'	5	6
tumpang	'join in'	***penumpang***	'passenger'	5	1
takut	'feel afraid (of)'	***panakut***	'coward, feel afraid (of) (EMPH)'	3	154
tokok	'pound'	***penokok***	'pounder'	3	44
antar	'bring'	***pengantar***	'escort'	2	130
diam	'be quiet'	***pandiam***	'taciturn person, be very quiet'	2	58
jaga	'guard'	***penjaga***	'guard'	2	41
ajar	'teach'	***plajar***	'teacher'	2	41
bantu	'help'	***pembantu***	'house helper'	2	34
urus	'arrange'	***pengurus***	'manager'	2	28
bicara	'speak'	***pembicara***	'speaker'	1	332
ikut	'follow'	***pengikut***	'follower'	1	253
dengar	'hear'	***pendengar***	'listener'	1	130
pikir	'think'	***pemikir***	'thinker'	1	102
ganti	'replace'	***pengganti***	'replacement'	1	40
tolong	'help'	***penolong***	'helper'	1	39
tunjuk	'show'	***petunjuk***	'guide'	1	32
tendang	'kick'	***penendang***	'kicker'	1	4
iris	'slice'	***pengiris***	'slicer'	1	3
tinju	'box'	***petinju***	'boxer'	1	1

Affixing verbal bases with ***PE(N)-*** derives nouns that denote the subject of the action, event, or state specified by the verbal base. The prefixed nouns include personal agents such as ***pendatang*** 'new comer', impersonal agents such as ***penyakit*** 'disease', or instruments such as ***penokok*** 'pounder'. This polysemy can be explained in terms of Booij's (1986: 509) "extension scheme" which shows that "the conceptual category Agent […] derived from verbs with an Agent subject can be extended" to instruments" such that "Personal Agent > Impersonal Agent > Instrument". In Papuan Malay, this extension schema also includes less typical agents derived from stative verbs, that is, "attributants" (van Valin 2005: 55) such as ***pemuda*** 'youth', derived from ***muda*** 'be young'. Hence, prefix ***PE(N)-*** is glossed as 'AG' ('agent') in the sense of 'agents or instruments who/which habitually do BASE or have the characteristics of BASE'.

Two of the derived nouns together with their verbal bases are given in context: ***pemimping*** 'leader' and its bivalent base ***pimping*** 'lead' in (21) and (22), and

pemuda ‘youth’ and its monovalent base ***muda*** ‘be young’ in (23) and (24), respectively.

(21) **pemimping** mati, yo smua mati
pem–pimping
AG–lead die yes all die
‘(when) the **leader** dies, yes, all die’ [081010-001-Cv.0026]

(22) o kenal karna bapa kang biasa **pimping** kor
oh! know because father you.know usual lead choir
‘oh, (I) know (him), because, you know, the gentleman usually **leads** the choir’ [081011-022-Cv.0243]

(23) sa liat **pe–muda** di Takar banyak skali
1SG see AG–be.young at Takar many very
‘I see (there are) very many **young people** in Takar’ [080925-003-Cv.0176]

(24) kasi–ang masi **muda** baru janda
love–PAT still be.young and.then widow
‘poor thing, (she’s) still **young** but now (she’s) a widow’ [081006-015-Cv.0032]

Four of the prefixed lexemes listed in Table 11 are nouns that can also receive an intensified verbal reading: ***pamalas*** ‘be very listless’ (25), ***pencuri*** ‘steal (EMPH)’ (26), ***panakut*** ‘fear (EMPH)’ (27), and ***pandiam*** ‘be very quiet’ (28). In (25) ***pamalas*** ‘be very listless’ receives a verbal reading given that a nominal reading of ***pamalas kerja*** ‘the lazy males work’ is inappropriate. In (26), ***pencuri*** ‘steal (EMPH)’ has verbal function as only verbs are negated with ***tra*** ‘NEG’ (see §5.3.6 and §13.1.1). In (27) ***panakut*** ‘fear (EMPH)’ functions as a verb, which is intensified with ***sampe*** ‘reach’. The utterance in (28) is ambiguous, as ***pandiam*** can receive the nominal reading ‘taciturn person’ or the verbal reading ‘be very quiet’.

(25) jadi sampe skarang laki~laki **pa–malas** kerja
so until now RDP~husband AG–be.listless work
‘so until now the men are **too listless** / **don’t like it at all** to work’ [081014-007-CvEx.0087]

(26) dong tra **pen–curi**
3PL NEG AG–steal
‘(nowadays), they don’t **steal (EMPH)**!’ [081011-022-Cv.0298]

(27) … i biasa–nya **panakut** sampe bagemana
pan–takut
ugh! be.usual–3POSSR AG–feel.afraid(.of) reach how
[About a frightening event at night:] ‘[she started (running) past (us),] ugh, usually (she’s) **very fearful** beyond words’ [081025-006-Cv.0330]

(28) Sofia de bilang begini, sa ini **pan–diam**
Sofia 3SG say like.this 1SG D.PROX AG–be.quiet
'Sofia said like this, 'I'm a **taciturn person** / I'm **very quiet**'' [081115-001a-Cv.0190]

As discussed in the introductory remarks in §3.1.4, the corresponding prefix in Proto Malayic and Standard Malay also has verbal function. That is, with monovalent stative bases, the derived lexemes "can function as VSIs" (Adelaar 1992: 183). This prefix does not, however, have the intensifying verbal function that Papuan Malay ***PE(N)-*** has. This intensified verbal reading of mono- and bivalent verbal bases prefixed with ***PE(N)-*** could be an extension of the original functions of ***pəN-*** found in Standard Malay or of ****pAN-*** found in Proto Malayic.

In other eastern Malay varieties, lexical items prefixed with ***pa-*** can also receive a verbal reading. For Ambon Malay, van Minde (1997: 109) presents a number of examples, noting that "the word class of the ***pa(N)-*** formation varies between transitive verb, intransitive verb and noun". For North Moluccan / Ternate Malay, Voorhoeve (1983: 4) presents two prefixed items with a basic verbal reading: ***pamalas*** 'lazy' and ***panggayung*** 'row'. Likewise, Litamahuputty (1994: 40) presents two such items: ***pamalas*** 'lazy' and ***panako*** 'afraid'; both "are considered to be monomorphemic", however. For Manado Malay, Stoel (2005: 24) also presents two such items: ***pancuri*** 'steal' and ***pandusta*** 'lie'. As mentioned, though, prefix ***pa-*** is unproductive in Manado and North Moluccan / Ternate Malay.

3.1.4.3. Prefixed items derived from nominal bases

The corpus contains five ***PE(N)-***prefixed nouns (with 33 tokens), listed in Table 12, which are derived from nominal bases and denote abstract concepts. In general, the derived lexemes denote an 'agent who executes what BASE indicates'. Four of the five lexemes are low frequency words, attested with less than 20 tokens. Moreover, the token frequencies for the respective bases are (much) higher for three of the five derived words. In addition, four items were tentatively classified as SI-borrowings (for details see language internal factor (g) on p. 115 in §3.1.1); in Table 12 these items are underlined.

Table 12: Affixation with ***PE(N)-*** of nominal bases

BW	Gloss	Item	Gloss	***PE(N)-*** #	BW #
printa	'command'	***pemrinta***	'government'	23	5
tugas	'duty'	***petugas***	'official'	5	19
usaha	'effort'	***pengusaha***	'entrepreneur'	3	2
acara	'ceremony'	***pengacara***	'master of ceremony'	1	40
rencana	'plan'	***perencana***	'planner'	1	17

In (29) and (30) one of the prefixed nouns and its nominal base are given in context, ***pemrinta*** 'government', and ***printa*** 'command', respectively.

(29) kalo de bilang spulu milyar **pemrinta** sanggup bayar
pem–printa
if 3SG say ten billion AG–command be.capable pay
'if he demands ten billion (then) the **government** is capable of paying' [081029-004-Cv.0073]

(30) masi banyak yang melangar **printa~printa** Tuhan
still many REL collide.with RDP~command God
'(there are) still many who violate God's **commands**' [081014-014-NP.0050]

3.1.4.4. Variables of the communicative event

To examine the issue of productivity of ***PE(N)-*** from a different perspective, a domain analysis was conducted which focused on the variables of speaker education levels, topics, and/or role-relations (for details see 'Language external factors' on p. 116 in §3.1.1). In all, 34 items prefixed with ***PE(N)-***, totaling 186 tokens, were investigated:

- 29 prefixed items with verbal bases (153 tokens)
- Five prefixed items with nominal bases (33 tokens)

For the 34 prefixed lexemes, most tokens (167/186 – 90%) can be accounted for in terms of speaker education levels, topics, and/or role-relations. The remaining 19/186 tokens (10%) cannot be explained in terms of these variables of the communicative event. These tokens occurred when less-educated speakers (-EDC-SPK) conversed with fellow-Papuans of equally low social standing (-STAT) about LOW topics, that is, casual daily-life issues.[108] (See Table 13 and Chart 3.)

If the prefixed lexemes were the result of a productive affixation process, one would expect the percentage of tokens that cannot be explained in terms of speaker education levels, topics, and/or role-relations to be much higher than 10%. Instead, most tokens (90%) seem to be conditioned by these variables of the communicative event. These findings do not support the conclusion that the respective lexemes are the result of a productive derivation process. Instead, they seem to be code-switches with Indonesian.

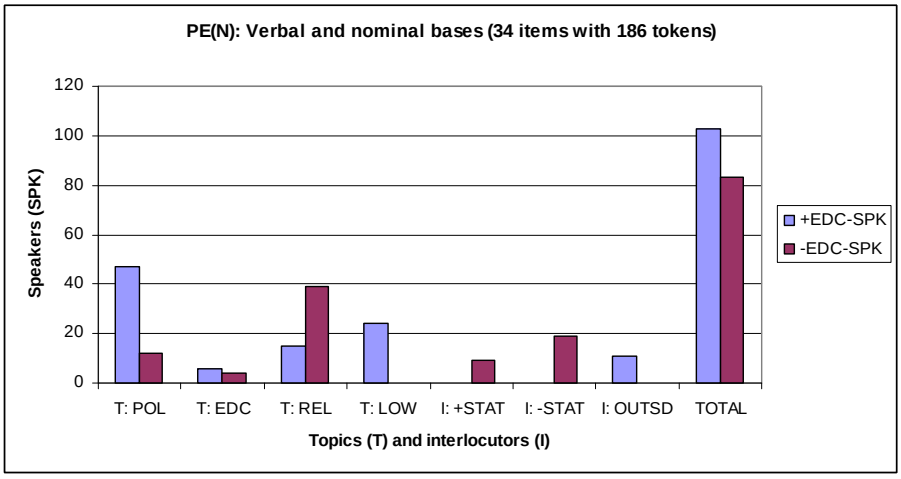
The data presented in Table 13 and Chart 3 is discussed in more detail below.

[108] As mentioned under Factor 3 'Relationships between interlocutors' in §1.5.1, all of the recorded less-educated speakers belonged to the group of Papuans with lower social status (-STAT), while the recorded Papuans with higher social status (-STAT), such as teachers, government officials, or pastors, were all better educated.

Table 13: Token frequencies for ***PE(N)***-prefixed lexemes with verbal and nominal bases by speakers, topics, and interlocutors (34 items)[109]

	Topics (TOP)				Interlocutors (ILCT)			Tokens
Prefixed lexemes with verbal bases (29 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	37	6	3	19	---	---	11	76
-EDC-SPK	11	2	37	---	9	18	0	77
Subtotal	48	8	40	19	9	**18**	11	153
Prefixed lexemes with nominal bases (5 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	10	0	12	5	---	---	0	27
-EDC-SPK	1	2	2	---	0	**1**	0	6
Subtotal	11	2	14	5	0	**1**	0	33
TOTAL (34 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	47	6	15	24	---	---	11	103
-EDC-SPK	12	4	39	---	9	**19**	0	83
Total	59	10	54	25	9	**19**	12	**186**

Chart 3: Token frequencies for ***PE(N)***-prefixed lexemes with verbal bases by speakers, topics, and interlocutors



[109] Abbreviations: SPK = speaker, +EDC = better educated, –EDC = less educated, POL = politics, EDC = education, REL = religion, LOW = LOW topics, +STAT = higher social status, -STAT = lower social status, OUTSD = outsider.

The data given in Table 13 and Chart 3 shows that for the 34 ***PE(N)-***prefixed lexemes, most tokens (167/186 – 90%) can be explained in terms of speaker education levels, topics, and/or role-relations between the speakers and their interlocutors; this total includes 135/153 (88%) tokens with verbal and 32/33 tokens (97%) with nominal bases.

More than half of the tokens were produced by better-educated speakers (+EDC-SPK) (103/186 – 55%), while less-educated speakers (-EDC-SPK) produced 83/186 tokens (45%).

Two thirds of the 186 tokens (123/186 – 66%) occurred during conversations about HIGH topics, that is, political, educational or religious affairs (POL, EDC and REL, respectively). This includes 68/103 tokens (66%) produced by +EDC-SPK and 55/83 tokens (66%) produced by -EDC-SPK.

Of the remaining 63/186 tokens, 37 were produced by +EDC-SPK when discussing LOW topics. This includes 24/103 tokens (23%) produced by +EDC-SPK during conversations with fellow-Papuans (both +STAT and -STAT speakers) and 11/103 tokens (11%) produced when conversing with an outsider, namely the author (OUTSD). This leaves 28 tokens that -EDC-SPK produced when discussing LOW topics with fellow-Papuans. More specifically, -EDC-SPK produced 9/83 tokens (11%) when conversing with +STAT Papuans. The remaining 19/83 tokens (23%) occurred when --EDC-SPK discussed LOW topics with -STAT Papuans, and therefore cannot be explained in terms of topics and/or role-relations. This total of 19 tokens refers to 10% of all 186 ***PE(N)-***tokens, including 18/153 tokens (12%) with verbal bases and 1/33 tokens (3%) with nominal bases.[110]

3.1.4.5. Summary and conclusions

Prefix ***PE(N)-*** is polyfunctional, in that it derives nouns from verbal, nominal, and quantifier bases. This polyfunctionality suggests that affixation with ***PE(N)-*** is a somewhat productive process (see language internal factor (d) on p. 114 in §3.1.1).

Concerning ***PE(N)-***affixation of verbal bases, three other observations support this conclusion: (1) the transparent form-function relationship between the derived nouns and their respective bases, (2) the large number of low frequency words and small number of high frequency words, and (3) the relative token frequencies with most bases having higher frequencies than the affixed lexemes. On the other hand, more than half of the derived lexemes were tentatively classified as SI-borrowings. These observations suggest that productivity of the affixation process is rather limited.

As for ***PE(N)-***affixation of nominal bases, two observations suggest that this is a productive process: (1) most of the derived lexemes are low frequency words, and

[110] Concerning the 11 hapaxes (nine with verbal and two with nominal bases), the data suggests that seven are conditioned by the variables of speaker education levels, topics, and/or role-relations, and therefore are best explained as code-switches with Standard Indonesian. This leaves only four hapaxes (with verbal bases) that cannot be accounted for in terms of language external factors and that are likely to result from a productive word-formation process. For four hapaxes P=0.0215 as opposed to P=0.0591 for 11 hapaxes (N=186).

(2) most bases have higher token frequencies than the affixed lexemes. On the other hand, almost all derived lexemes were tentatively classified as SI-borrowings. These findings suggest that ***-an***-affixation of nominal bases has limited productivity

As for the speech situations during which the derived nouns occurred, the vast majority of the attested tokens are conditioned by the variables of speaker education levels, topics, and/or role-relations. Hence, these items are best explained as code-switches with Indonesian.

These findings suggest that ***PE(N)***-affixation has, at best, marginal productivity.

3.1.5. Prefix *BER-* 'INTR'

Affixation with ***BER-*** 'INTR' typically derives monovalent verbs from verbal bases, as illustrated in (31).[111] Most commonly, the derived verbs have the same semantics as their respective bases. Some lexical items are also derived from nominal bases. The derivation process does not seem to be productive, as discussed below.

(31) … waktu saya **ber–buru** saya perlu makang pinang
time 1SG INTR–hunt 1SG need eat betel.nut
'… when I **hunt** I need to chew betel nuts' [080919-004-NP.0011]

The present corpus contains 62 derived nouns (602 tokens) prefixed with ***BER-***:[112]

1. Verbs with verbal bases (29 items with 227 tokens)
2. Verbs with nominal, numeral, and interrogative bases (33 items with 375 tokens)

The corpus also includes 16 formally complex words with non-compositional semantics, such as ***bertriak*** 'scream', ***berjuang*** 'struggle', or ***berlabu*** 'anchor'.

Before discussing ***BER-***affixation of verbal bases in §3.1.5.2 and of nominal bases in §3.1.5.3, the allomorphy of ***BER-*** is investigated in §3.1.5.1. Pertinent variables of the communicative event that may impact the use of ***BER-*** are explored in §3.1.8. The main findings on prefix ***BER-*** are summarized and evaluated in §3.1.5.4.

3.1.5.1. Allomorphy of ***BER-***

Prefix ***BER-*** has two allomorphs, ***ber***- and ***ba***-. The allomorphs are not governed by phonological processes.

The form ***ber***-, in turn, has four realizations that are effected by morphologically conditioned phonological rules. More specifically, the four allomorphs are conditioned by the word-initial segment of the base word, as illustrated in Table 14: /bɛr-/, /br-/, /bl-/, and /bɛ-/. The prefix is typically realized as /bɛr-/. With an onset vowel, however, ***ber***- is very commonly realized as /br-/. When prefixed to ***ajar***

[111] The small caps designate an abstract representation of the prefix as it has more than one form of realization, namely the two allomorphs ***ber***- and ***ba***- (see §3.1.5.1).

[112] The 62 verbs include 25 hapaxes (P=0.0415); the 29 verbs with verbal bases include 11 hapaxes (P=0.0484); the 33 verbs with nominal, numeral, or interrogative bases include 14 hapaxes (P=0.0373).

'teach' the prefix is realized as /bl-/, while it is realized as /bɛ-/ when affixed to ***kerja*** 'work' or ***brapa*** 'how many, several'.

Table 14: Realizations of allomorph ***ber-***

ber-base	Orthogr.	Gloss
/bɛr–dʒuaŋ/	***berjuang***	'struggle (for)'
/br–aŋkat/	***brangkat***	'leave'
/bl–adʒar/	***blajar***	'study'
/bɛ–kɛrdʒa/	***bekerja***	'work'
/bɛ–brapa/	***bebrapa***	'be several'

Allomorph ***ba-*** occurs much less frequently. Attested are only the ten items listed in Table 15 with a total of 32 tokens. Some of these items are alternatively realized with allomorph ***ber-***. Therefore, for each item the token frequencies for ***ba-*** and for ***ber-*** are given.[113] If in a greater number of tokens the prefix is realized with /ba-/ rather than with /bɛr-/, then its orthographic representation is ***ba-*** as in ***bakalay*** 'fight'. If both realizations have the same token frequencies, then the orthographic representation follows its realization in the recorded word list, as in ***bagaya*** 'put on airs'.

Table 15: Realizations of allomorph ***ba-***

ba-base	Orthogr.	Gloss	***ba-*** #	***ber-*** #
/ba–kalaj/	***bakalay***	'fight'	19	0
/ba–taria/	***bertriak***[114]	'scream'	3	18
/ba–biŋuŋ/	***babingung***	'be confused'	2	0
/ba–diam/	***badiam***	'be quiet'	2	0
/ba–diri/	***berdiri***	'stand'	1	54
/ba–gaja/	***bagaya***	'put on airs'	1	1
/ba–gigit/	***bagigit***	'bite'	1	0
/ba–kumis/	***bakumis***	'have a beard'	1	0
/ba–isi/	***baisi***	'be muscular'	1	0
/ba–mɛkap/	***bamekap***	'wear make-up'	1	0

In realizing prefix ***BER-*** most commonly as allomorph ***ber-*** rather than as ***ba-***, Papuan Malay contrasts with other eastern Malay varieties such as Ambon Malay (van Minde 1997: 95), Banda Malay (Paauw 2008: 249), Kupang Malay (Steinhauer

[113] In addition, the 2,459-item word list (see Chapter 2) contains five lexemes which are realized with /ba-/ rather than with /bɛr-/: /badara/ 'be bloody', /baduri/ 'be thorny', /bagisi/ 'be nutritious', /bajalan/ 'walk', and /baribut/ 'be noisy'. In the corpus these five items are realized with /bɛr-/ (each item occurs only once in the entire corpus). By contrast, the word-list informant realized ***badiri*** 'stand' with /bɛr-/ as /bɛrdiri/ 'stand'.

[114] The root is realized as /triak/ when speakers employ allomorph ***ber-***, whereas it is realized as /taria/ when speakers use allomorph ***ba-***.

1983: 46), Manado Malay (Stoel 2005: 18), North Moluccan / Ternate Malay (Taylor 1983: 18, Voorhoeve 1983: 4, and Litamahuputty 2012: 125). In these varieties the prefix is always realized as ***ba-***. Instead, the items prefixed with ***BER-*** have more resemblance with the corresponding Indonesian items where the prefix is realized as ***ber-***. In addition, in Larantuka Malay the prefix is also realized as ***bə(r)-*** (Paauw 2008: 253). Again this difference between Papuan Malay and the other eastern Malay varieties points to the distinct histories of both, discussed in §1.2.

3.1.5.2. Prefixed items derived from verbal bases

The present corpus includes 29 ***BER-***prefixed lexemes (with 227 tokens) with verbal bases, as listed in Table 16. Of the 29 lexemes, 11 have monovalent bases such as stative ***diam*** 'be quiet' or dynamic ***jalang*** 'walk'. Another 18 items have bivalent bases. Four of the 18 prefixed lexemes have monotransitive as well as intransitive uses, while 12 lexemes have intransitive uses only. These 16 lexemes have the same semantics as their bivalent bases, as shown in (34) to (42). For the remaining two prefixed lexemes the semantics are distinct from those of their bases, as shown in (43) to (48). One of them has monotransitive as well as intransitive uses, while the other one has intransitive uses only.

Almost all of the derived lexemes are low frequency words (27 lexemes, attested with less than 20 tokens). Moreover, the token frequencies for the respective bases are (much) higher for most of the derived words (22 lexemes). This is due to the fact that the affixed lexemes and the bases have the same semantics and that, overall, speakers tend to use the bases rather than the prefixed forms, as shown in (34) to (33). Also, most of the 29 prefixed lexemes (24 items) were tentatively classified as borrowings from Standard Indonesian (SI-borrowings) (for details see language internal factor (g) on p. 115 in §3.1.1); in Table 16 these items are underlined.

Table 16: Affixation with ***BER-*** of verbal bases

BW	Item	Gloss	***BER-*** #	BW #
Monovalent bases: Bases and prefixed lexemes have same semantics				
tobat	***bertobat***	'repent'	8	1
beda	***berbeda***	'be different'	7	34
tanggung-jawap	***bertanggung-jawap***	'be responsible'	5	6
bahaya	***berbahaya***	'be dangerous'	3	3
diam	***badiam***	'be quiet'	2	60
bingung	***berbingung***	'be confused'	2	30
jalang	***berjalang***	'walk'	1	480
ibada	***beribada***	'worship'	1	11
sandar	***bersandar***	'lean'	1	6
hati~hati	***berhati~hati***	'be careful'	1	5
pisa	***berpisa***	'be separate'	1	4

BW	Item	Gloss	***BER-*** #	BW #
Bivalent bases: Bases and prefixed lexemes have same semantics				
Prefixed lexemes: Monotransitive and intransitive uses				
buru	***berburu***	'hunt'	10	5
buat	***berbuat***	'make'	7	100
harap	***berharap***	'hope'	1	8
ribut	***bribut***	'trouble'	1	5
Prefixed lexemes: Monotransitive uses				
pikir	***berpikir***	'think'	8	102
bicara	***berbicara***	'speak'	7	333
kerja	***bekerja***	'work'	5	191
tahang	***bertahang***	'hold (out/back)'	5	48
uba	***bruba***	'change'	5	9
gabung	***bergabung***	'join'	4	3
maing	***bermaing***	'play'	3	113
tindak	***bertindak***	'act'	2	1
ikut	***brikut***	'be following'	1	259
kumpul	***berkumpul***	'gather'	1	16
bentuk	***berbentuk***	'form'	1	12
gigit	***bergigit***	'bite'	1	10
Bivalent bases: Bases and prefixed lexemes have distinct semantics				
Prefixed lexeme: Monotransitive and intransitive uses				
ajar ('teach')	***blajar***	'study'	51	41
Prefixed lexeme: Monotransitive uses				
angkat ('lift')	***brangkat***	'leave'	82	81

Affixation with ***BER-*** of verbal bases derives lexemes that typically have the same semantics as their respective bases, with ***BER-*** being glossed as 'VBLZ' ('verbalizer'). This applies to ***BER-***prefixed lexemes with monovalent and with bivalent bases.

The fact that monovalent bases derive ***BER-***prefixed lexemes with the same semantics is illustrated with stative ***bingung*** 'be confused' and prefixed ***berbingung*** 'be confused' in (32) and (33), and with dynamic ***ibada*** 'worship' and prefixed ***beribada*** 'worship' in (34) and (35), respectively.

(32) memang sa punya ade sa juga **bingung** dengang dia
indeed 1SG POSS ySb 1SG also be.confused with 3SG
'indeed (he was) my younger cousin, I'm also **confused** about him'
[080918-001-CvNP.0014]

(33) nanti di skola baru kamu **ba–bingung** dengang
very.soon at school and.then 2PL VBLZ–be.confused with

bahasa Inggris
language English
[Addressing lazy students:] 'later in school, then you'll be **confused** about English' [081115-001a-Cv.0151]

(34) orang jalang itu mo pergi **ibada**
person walk D.DIST want go worship
[About a youth retreat:] 'people doing that traveling want to go (and) **worship**' [081006-016-Cv.0017]

(35) nanti kita **ber–ibada** selesay malam ka
very.soon 1PL VBLZ–worship finish night maybe
baru sa pergi
and.then 1SG go
'later, after we have **worshipped**, maybe in the evening, and then I'll go (there)' [080918-001-CvNP.0016][115]

Bivalent bases also derive ***BER-***prefixed lexemes with the same semantics, as shown in (36) to (40). As discussed in §11.1.2, bivalent verbs also have intransitive uses, as well as monotransitive uses. The same applies to some of the ***BER-***prefixed lexemes, as illustrated in (36) to (39).

(36) jadi kitorang bingung **pikir** itu pen–jaga kubur–ang
so 1PL be.confused think D.DIST AG–guard bury–PAT
'so we're confused to **think (about)**, what's-its-name, a guard (for) the grave' [080923-007-Cv.0024]

(37) … tapi ana~ana ni dong tida taw **ber–pikir** itu
but RDP~child D.PROX 3PL NEG know VBLZ–think D.DIST
[About impolite teenagers:] '… but these kids they don't know (how) to **think (about)** those (feelings of mine)' [081115-001b-Cv.0037]

(38) skarang orang su tra **pikir** tentang
now person already NEG think about
hal ke–benar–ang
thing NMLZ–be.true–NMLZ
'nowadays, the people already don't **think** about things (related to) truth' [081006-032-Cv.0016]

(39) … karna dia **ber–pikir** tentang dia punya badang
because 3SG VBLZ–think about 3SG POSS body
'[she doesn't think about serving my or her guests] because she **thinks** about her body' [081006-032-Cv.0062]

Other ***BER-***prefixed lexemes, such as ***berbicara*** 'speak', only have intransitive uses, as shown in (42) to (40).

[115] The original recording says ***kita i beribada selesay***. Most likely the speaker wanted to say ***kita ibada selesay*** 'after we have worshipped' but cut himself off to replace ***ibada*** 'worship' with ***beribada*** 'worship'.

(40) baru nanti **ber–bicara** untuk nika
and.then very.soon VBLZ–speak for marry.officially
[About wedding customs:] 'and then very soon (they'll) **talk** about marrying' [081110-006-CvEx.0050]

(41) de **bicara** trus
3SG speak be.continuous
'he kept **talking**' [080922-010a-CvNF.0145]

(42) baru de **bicara** sa deng bahasa Inggris
and.then 3SG speak 1SG with language English
'and then she **talked** to me in English' [081115-001a-Cv.0229]

The corpus includes only two ***BER-***prefixed lexemes that have distinct semantics vis-à-vis their bivalent bases, namely ***ajar*** 'teach' and prefixed ***blajar*** 'study', and ***angkat*** 'lift' and prefixed ***brangkat*** 'leave' as shown in (43) to (48). Both ***ajar*** 'teach' and ***blajar*** 'study' are used monotransitively as in (43) and (44), as well as intransitively as in (45) and (46), respectively; in each case both lexemes maintain their distinct semantics.

(43) de **ajar** dorang tu untuk baik
3SG teach 3PL D.DIST for be.good
'she **teaches** them there for (their own) good' [081115-001a-Cv.0216]

(44) Ise de … ikut bahasa Inggris **bl–ajar** kursus
Ise 3SG follow language English VBLZ–teach course
bahasa Inggris dulu
language English be.prior
'Ise will participate in an English course, (she'll) **study** an English language course first' [081025-003-Cv.0223]

(45) de suda **ajar** bagus tiap sore itu
3SG already teach be.good every afternoon D.DIST
'she's already been **teaching** well, each and every afternoon' [081115-001a-Cv.0126]

(46) dong tida **bl–ajar** baik
3PL NEG VBLZ–teach be.good
'they don't **study** well' [081115-001b-Cv.0067]

Bivalent ***angkat*** 'lift' and prefixed ***brangkat*** 'leave' have distinct semantics and a distinct distribution. While ***angkat*** 'lift' is always used monotransitively as in (47), ***brangkat*** 'leave' is always used intransitively as in (48).

(47) bapa de **angkat** rotang besar
father 3SG lift rattan be.big
'he father **picked-up** a big rattan (stick)' [080921-004a-CvNP.0084]

(48) skarang de mo **br–angkat**
now 3SG want VBLZ–lift
'then he wanted to **leave**' [080919-007-CvNP.0023]

In sum, with the exception of the last two lexemes, ***BER-***prefixed verbs have the same semantics as their respective bases. This suggests that affixation of verbal bases with prefix ***BER-*** is not a productive process. Instead, the attested prefixed lexemes and their bases are taken as pairs of words from two different speech varieties: the unaffixed items are native Papuan Malay lexemes whereas the corresponding affixed items are Indonesian lexemes.

Given these properties, Papuan Malay ***BER-*** contrasts with the corresponding prefix in other Malay varieties. In most eastern Malay varieties, corresponding prefix ***ba-*** forms verbs with a variety of meanings. The most common ones are durative and reflexive meanings, which are reported for Ambon Malay (van Minde 1997: 96–98) , Banda Malay (Paauw 2008: 249–250) ,[116] Manado Malay (Stoel 2005: 18–22) , North Moluccan / Ternate Malay (Taylor 1983: 18, and Litamahuputty 2012: 125–127). In Kupang Malay (Steinhauer 1983: 46–49) and Larantuka Malay (Paauw 2008: 249–254-255) , the prefix typically signals durative and reciprocal meanings. In Standard Indonesian, the main function of corresponding prefix ***ber-*** is to create monovalent verbs with reflexive meaning (Englebretson 2003: 131, 2007: 96, and Ewing 2005: 251).

3.1.5.3. Prefixed items derived from nominal, numeral or interrogative bases

The present corpus contains 33 ***BER-***prefixed lexemes (with 375 tokens), as listed in Table 17: 30 lexemes with nominal bases, two lexemes with numeral bases, and one lexeme with an interrogative base. Generally speaking, the affixation process derives monovalent verbs with the meaning of 'be/have/do what BASE indicates'.

Most of the derived lexemes are low frequency words (29 lexemes, attested with less than 20 tokens). Besides, the token frequencies for the respective bases are (much) higher for most of the derived words (21 lexemes). Again, this is due to the fact that Papuan Malay speakers typically use alternative analytical constructions to convey the meanings of the prefixed lexemes. Further, most of the 33 items (25 items) were tentatively classified as SI-borrowings (for details see language internal factor (g) on p. 115 in §3.1.1); in Table 17 these items are underlined.

Table 17: Affixation with ***BER-*** of nominal, numeral, and interrogative bases

BW	Gloss	Item	Gloss	***BER-*** #	BW #
Nominal bases					
doa	'prayer'	***berdoa***	'pray'	136	20
arti	'meaning'	***brarti***	'mean'	89	7
diri	'self'	***berdiri***	'stand'	55	14

[116] For Banda Malay, Paauw (2008: 249) reports that ***ba-*** does not form verbs with reflexive meaning.

BW	Gloss	Item	Gloss	***BER-*** #	BW #
usaha	'effort'	***berusaha***	'attempt'	25	2
dosa	'sin'	***berdosa***	'sin'	6	4
saksi	'witness'	***bersaksi***	'testify'	6	2
hasil	'result'	***berhasil***	'succeed'	6	13
kwasa	'power'	***berkwasa***	'be powerful'	4	25
hak	'right'	***berhak***	'have right'	4	15
sodara	'sibling'	***bersodara***	'be siblings'	3	127
kebung	'garden'	***berkebung***	'do farming'	3	61
ade-kaka	'siblings'	***brade-kaka***	'be siblings'	3	25
malam	'night'	***bermalam***	'overnight'	2	191
bahasa	'language'	***berbahasa***	'speak'	2	136
temang	'friend'	***bertemang***	'be friends'	2	85
kluarga	'family'	***berkluarga***	'have family'	2	49
gaya	'manner'	***bagaya***	'put on airs'	2	7
ana	'child'	***brana***	'give birth'	1	739
bua	'fruit'	***berbua***	'have fruit'	1	38
dara	'blood'	***berdara***	'bleed'	1	27
sifat	'character-istic'	***bersifat***	'have character-istics of'	1	18
duri	'thorn'	***berduri***	'have thorns'	1	8
harga	'value'	***berharga***	'be valuable'	1	4
syukur	'thanks'	***bersyukur***	'give thanks'	1	2
fungsi	'function'	***berfungsi***	'function'	1	1
gisi	'nutrient'	***bergisi***	'be nutritious'	1	1
isi	'content'	***baisi***	'be muscular'	1	1
komunikasi	'communica-tion'	***berkomuni-kasi***	'communicate'	1	1
kumis	'beard'	***bakumis***	'have a beard'	1	1
mekap	'make-up'	***bamekap***	'wear make-up'	1	1
Numeral bases					
satu	'one'	***bersatu***	'be one'	6	516
empat	'four'	***berempat***	'be four'	1	66
Interrogative base					
brapa	'how many, several'	***bebrapa***	'be several'	6	109

Affixation with ***BER-*** derives monovalent verbs from nominal, numeral, and interrogative bases. The derived verbs have the general meaning of 'be/have/do BASE'. Examples are ***brarti*** 'have the meaning of' or 'mean', ***berdoa*** 'do prayer' or 'pray', ***bersatu*** 'be one', or ***bebrapa*** 'be several'.[117] The monovalent verb ***berdiri*** 'stand' is an exception. Historically related to the noun ***diri*** 'self', it does not have a

[117] See §5.8.4 for a discussion of interrogative ***brapa*** 'how many' and §5.10 regarding its uses as the quantifier 'several'.

transparent form-function relationship to its base. The transparent form-function relationship between the remaining 32 items and their bases suggests that these lexemes are the result of a productive affixation process. Two observations are made, however.

First, the data indicates that Papuan Malay speakers prefer to employ analytical constructions to express the meanings conveyed by the prefixed items, as illustrated in (49) to (54). To communicate 'have BASE', speakers typically use the existential verb ***ada*** 'exist' rather than the prefixed form as shown in (49) with ***ada duri*** vs. ***berduri*** 'have thorns'.

(49) ada … dua macang jenis ada yang **ber–duri** ada yang
exist two variety kind exist REL VBLZ–thorn exist REL
tida … kang ada sagu yang tida **ada** **duri**
NEG you.know exist sago REL NEG exist thorn
'there are … two kinds (of sago palms), ones that **have thorns** and ones that don't (have thorns) … you know (there are) sago palms that don't **have thorns**' [081014-006-CvPr.0007/0009]

To express 'be BASE', speakers rather use a nominal predicate such as ***ade-kaka*** 'siblings' in (50), than the prefixed form such as ***brade-kaka*** 'be siblings' in (51).

(50) jadi saya dengang dia **ade-kaka** sunggu
ySb-oSb
so 1SG with 3SG siblings be.true
'so I and she **are** full **siblings**' [080927-009-CvNP.0044]

(51) jadi saya dengang kaka Nofita masi **br–ade-kaka**
so 1SG with oSb Nofita still VBLZ–siblings
'so I and older sister Nofita **are** still **siblings**' [080927-007-CvNP.0022]

To communicate 'do BASE', speakers typically employ alternative verbs. They tend to say, for example, ***biking kebung*** 'make/work a garden' as in (52), rather than use prefixed ***berkebung*** 'do farming' as in (53). Likewise, it is more common to say ***taw bahasa X*** 'speak language X' than to use prefixed ***berbahasa X*** 'speak language X' as in (54).

(52) kalo di Arbais prempuang bisa **biking** **kebung**
if at Arbais woman be.able make garden
'as for Arbais, (there) the women can **work a garden**' [081014-007-CvEx.0035]

(53) bapa pergi **ber–kebung** saya ikut
father go VBLZ–garden 1SG follow
'(whenever my) father went to **do farming** I went with (him)' [081110-008-CvNP.0002]

(54) jadi tong cuma **taw** **bahasa** **Yali** … tapi sa bilang
so 1PL just know language Yali but 1SG say

kamu	ber–syukur	karna	bisa	…	**ber–bahasa**	**Yali**
2PL	VBLZ–thank.God	because	be.able		VBLZ–language	Yali

'so we only **spoke Yali** … but I said, 'you (should) be grateful because (you) can **speak Yali**'' [081011-022-Cv.0101/0184]

Second, the exchange in (55) suggests that the high frequency items listed in Table 17 may well have non-compositional semantics for Papuan Malay speakers. In a conversation about religious affairs, the speaker produced ***diberdoa*** 'be prayed for'. This item is ungrammatical in both Papuan Malay and Standard Indonesian. Papuan Malay does not have a morphologically marked undergoer voice. The Standard Indonesian undergoer voice marker ***di-*** cannot co-occur with prefix ***ber-***, but always replaces it. This example show that the speaker perceives ***berdoa*** 'pray' as a monomorphemic word to which she affixed the Indonesian undergoer voice marker ***di-*** in an attempt to approximate Indonesian.

(55)

bebang	masala	de	punya	dia	perlu	…	harus	**di–ber–doa**
burden	problem	3SG	POSS	3SG	need		have.to	UV–VBLZ–prayer

[Conversation about problems of a church congregation:] '(all) burdens (and) problems (that) it has, (the congregation) needs … has to **be prayed for**' [080917-008-NP.0089/0091]

3.1.5.4. Summary and conclusions

Prefix ***BER-*** is a polyfunctional affix that derives lexemes from verbal and nominal bases. This polyfunctionality suggests that affixation with ***BER-*** is a somewhat productive process (see language internal factor (d) on p. 114 in §3.1.1). Two other observations support this conclusion: (1) the large number of low frequency words and small number of high frequency words, and (2) the relative token frequencies with most bases having higher frequencies than the affixed lexemes.

Three other observations, however, do not support the conclusion that affixation with ***BER-*** is a productive process: (1) for the prefixed lexemes with verbal bases, the derived lexemes have the same semantics as their bases, and (2) for lexemes with nominal bases, speakers prefer to use alternative analytical constructions rather than the affixed lexemes, and (3) most of the lexemes with verbal or nominal bases were tentatively classified as SI-borrowings.

Taken together, these findings indicate that Papuan Malay speakers do not employ prefix ***BER-*** as a productive device to derive new words. This conclusion is also supported by the findings of a domain analysis which indicate that most of the attested tokens can be accounted for in terms of the variables of speaker education levels, topics, and/or role-relations (details are discussed in §3.1.8, together with the findings for suffix ***-nya*** '3POSSR' and circumfix ***ke-/-ang*** 'NMLZ'). Therefore, these lexemes are best explained as code-switches with Indonesian. (See Adelaar 1992, Mintz 1994, and Sneddon 2010 for a detailed discussion of prefix ***ber-*** in Standard Indonesian and Standard Malay.)

The conclusion that prefix ***BER-*** is unproductive again sets Papuan Malay apart from other eastern Malay varieties. In regional varieties such as Ambon Malay (van Minde 1997: 96–98) , Banda Malay (Paauw 2008: 249–250) , Larantuka Malay

(Paauw 2008: 253–255), Manado Malay (Stoel 2005: 18–22) , North Moluccan / Ternate Malay (Taylor 1983: 18), and (Litamahuputty 2012: 125–127) the prefix is a productive derivational device.[118] This distinction between Papuan Malay and the other eastern Malay varieties once again hints at the separate histories of both, discussed in §1.2.

3.1.6. Suffix *-nya* '3POSSR'

Suffix ***-nya*** '3POSSR' is typically attached to nominal bases to indicate possessive relations, as illustrated in (56). In addition, a considerable number of lexemes are derived from verbal bases, while a small number of lexemes have prepositional, adverbial, locative, or demonstrative bases. The affixation process does not seem to be productive, as discussed below.

(56)

jadi	**ana–nya**	hidup	ana	itu	masi	ada
so	child–3POSSR	live	child	D.DIST	still	exist

'so **her child** lives, that child still exists' [080921-005-CvNP.0007]

The present corpus contains 123 lexical items (387 tokens) suffixed with ***-nya***:[119]

1. Suffixed items with nominal bases (82 items with 285 tokens)
2. Suffixed items with verbal bases (36 items with 82 tokens)
3. Suffixed items with other bases (five items with 20 tokens)

The corpus also contains seven formally complex words with non-compositional semantics. All seven items have adverbial function, such as ***misalnya*** 'for example' or ***akirnya*** 'finally'.

Suffixed lexemes with nominal bases are discussed in §3.1.6.1, those with verbal bases in §3.1.6.2, and those with other bases in §3.1.6.3. Pertinent variables of the communicative event that may impact the use of ***-nya*** are explored in §3.1.8. The main findings on suffix ***-nya*** are summarized and evaluated in §3.1.6.4.

3.1.6.1. Suffixed items derived from nominal bases

The present corpus contains 81 ***-nya***-suffixed lexemes (with 215 tokens) with nominal bases, where ***-nya*** typically signals possession. As an extension of the possessive-marking function, some of the derived items listed in Table 18 function as sentence adverbs, namely ***maksutnya*** '(someone) meant' (literally 'the purpose of'), ***katanya*** '(someone) said' (literally 'the word of'), and ***artinya*** 'that means' (literally 'the meaning of').

Derived words with token frequencies of five or more are listed in Table 18. All but one of the derived lexemes are low frequency words (80 items, attested with less than 20 tokens). Besides, the token frequencies for the respective bases are (much) higher for most of the derived words (65 lexemes). Of the 81 suffixed lexemes, 76

[118] Voorhoeve (1983: 4) considers prefix ***ba-*** to be unproductive.

[119] The 123 suffixed lexemes include 68 hapaxes (P=0.1757); the 82 lexemes with nominal bases include 44 hapaxes (P=0.1543); the 36 lexemes with verbal bases include 21 hapaxes (P=0.2561); the five lexemes with other bases include three hapaxes (P=0.1500).

were tentatively classified as borrowings from Standard Indonesian (SI-borrowings) (for details see language internal factor (g) on p. 115 in §3.1.1); in Table 18 these items are underlined. The exceptions are the three derived lexemes that function as sentence adverbs, two of which are presented in context in (59) and (60).

The low token frequencies for the derived lexemes result from the fact that Papuan Malay speakers usually use an alternative strategy to express possessive relations. Instead of suffixing ***-nya*** to a nominal base, Papuan Malay encodes adnominal possession by an analytical construction with ***punya***, or reduced ***pu***, 'POSS' (see Chapter 9). The '***punya*** #' column in Table 18 lists the token frequencies for adnominal possessive constructions with ***punya***/***pu*** 'POSS'. Examples are given in (57) and (58).

Table 18: Affixation with ***-nya*** of nominal bases

BW	Gloss	Item	Gloss	***-nya*** #	***punya*** #
nama	'name'	***namanya***	'the name of'	23	38
istri	'wife'	***istrinya***	'the wife of'	11	22
ana	'child'	***ananya***	'the child of'	7	119
orang	'person'	***orangnya***	'the person of'	6	8
ruma	'house'	***rumanya***	'the house of'	5	43
hasil	'product'	***hasilnya***	'the product of'	5	2
istila	'term'	***istilanya***	'the term of/for'	5	1
dalam	'inside'	***dalamnya***	'the inside of'	5	---
maksut	'purpose'	***maksutnya***	'(someone) meant'	70	3
kata	'word'	***katanya***	'(someone) said'	19	---
arti	'meaning'	***artinya***	'that means'	17	---

In (57), ***-nya*** is suffixed to the nominal base ***nama*** 'name', giving the possessive reading ***namanya*** 'her name'. By contrast, (58) shows the inherited analytical strategy of expressing the same meaning with possessive marker ***pu*** 'POSS'.

(57) **nama–nya** Madga
name–3POSSR Madga
'**her name** is Madga' [081011-005-Cv.0027]

(58) **de** **pu** **nama** Martin
3SG POSS name Martin
'**his name** is Martin' [081011-022-Cv.0241]

The examples in (59) and (60) illustrate the uses of ***maksutnya*** '(someone) meant' and ***katanya*** '(someone) said', respectively, as sentence adverbs.

(59) … **maksut–nya** saya harus dayung dulu dengang prahu
purpose–3POSSR 1SG have.to row be.prior with boat
[I'm getting ready, I take my bow and arrows and an oar,] **that means**, I have to row first with a boat' [080919-004-NP.0008]

(60) **kata–nya** orang Sulawesi smua
word–3POSSR person Sulawesi all
'**it's being said** (that) they are all Sulawesi people' [081006-032-Cv.0102]

3.1.6.2. Suffixed items derived from verbal bases

The present corpus contains 36 ***-nya***-suffixed lexemes (with 82 tokens) with verbal bases. Affixation with ***-nya*** derives nominals from verbal bases. The derived nominals have a possessive reading, such as ***ceritranya*** 'the telling of' or 'his/her telling'. As an extension of the nominalizing and possessive-marking function of ***-nya***, eight of the derived lexemes function as adverbs, such as ***biasanya*** 'usually' (literally 'its being usual') or ***kususnya*** 'especially' (literally 'its being special'). Derived words with token frequencies of three or more are listed in Table 19.

All 36 affixed lexemes are low frequency words, attested with less than 20 tokens. Moreover, the token frequencies for the respective bases are (much) higher for all but one of the derived words (35 lexemes). This is due to the fact that Papuan Malay speakers tend to use the respective bases, as in (61) to (64), rather than the suffixed forms. Of the 36 derived lexemes, nine were tentatively classified as SI-borrowings (for details see language internal factor (g) on p. 115 in §3.1.1); in Table 19 these items are underlined.

Table 19: Affixation with ***-nya*** of verbal bases

BW	Gloss	Item	Gloss	***-nya*** #	BW #
mo	'want'	***mawnya***	'the wanting of'	6	972
ceritra	'tell'	***ceritranya***	'the telling of'	6	162
pegang	'hold'	***pegangnya***	'the holding of'	3	114
hidup	'live'	***hidupnya***	'the living of'	3	74
biasa	'be usual'	***biasanya***	'usually'	18	181
harus	'have to'	***harusnya***[120]	'appropriately'	7	379
kusus	'be special'	***kususnya***	'especially'	3	30

In (61), ***-nya*** is suffixed to the verbal base ***mo*** 'want' giving the nominalized form ***mawnya*** 'the wanting of'. The example in (62) illustrates the preferred strategy of expressing the same meaning in a verbal clause with the base ***mo*** 'want'.

(61) **maw–nya** ke kampung maw biking apa di sana
want–3POSSR to village want make what at L.DIST
[Addresses a teenager playing hooky:] '**your wish** (is to go) to the village, what do (you) want to do there?' (Lit. '**his wanting** (is) to the village') [081115-001a-Cv.0046]

[120] Included in the six ***harusnya*** 'appropriately' tokens is one ***seharusnya*** 'properly' token. According to one consultant, both forms have the same semantics with ***harusnya*** 'appropriately' being the more common form.

(62) **ko** **mo** ke kampung tapi ko skola
2SG want to village but 2SG go.to.school
'**you want** (to go) to the village but you're going to school' [080922-001a-CvPh.0734]

In (63), ***-nya*** is suffixed to the verbal base ***biasa*** 'be usual' with adverbially used ***biasanya*** 'usually' modifying the verb ***dansa*** 'dance'. More commonly, however, speakers employ the base ***biasa*** 'be usual' as in (64) with adverbially used ***biasa*** 'be usual' modifying the verb ***maing*** 'play'.

(63) … dansa lemon-nipis itu **biasa–nya** dansa lemon-nipis
dance citron D.DIST be.usual–3POSSR dance citron
'[they make a ceremony, they sing on and on,] (they) dance that citron (group dance), **usually** (they) dance the citron (group dance)' [081110-005-CvPr.0098]

(64) Herman dorang **biasa** maing di sini tu
Herman 3PL be.usual play at L.PROX D.DIST
'Herman and the others **usually** play right here' [080923-009-Cv.0017]

3.1.6.3. Suffixed items derived from other bases

The present corpus contains five lexemes (with 20 tokens) which are derived from a number of different bases. Two lexemes have prepositional bases and one has an adverbial base, listed in Table 20, with ***-nya*** having adverb-marking function. In addition, one lexeme has a demonstrative base and one a locative base, listed in Table 21, with ***-nya*** having emphasizing function.

The two lexemes with prepositional bases and the one with an adverbial base have distinct meanings vis-à-vis their bases. These items usually function as sentence adverbs as shown in (65) and (66). Again, the adverbial-marking function of ***-nya*** seems to be an extension of its nominalizing and possessive-marking function. For instance, ***spertinya*** 'it seems' can be literally translated as 'its being similar to'. All five affixed lexemes are low frequency words, attested with less than 20 tokens. In addition, the token frequencies for the respective bases are (much) higher for all of the derived words. All five suffixed lexemes were tentatively classified as SI-borrowings; in Table 20 these items are underlined.

Table 20: Affixation with ***-nya*** of prepositional and adverbial bases

BW	Gloss	Item	Gloss	***-nya*** #	BW #
Prepositional base					
sperti	'similar to'	***spertinya***	'it seems'	12	217
kaya	'like'	***kayanya***	'it looks like'	5	61
Adverbial bases					
memang	'indeed'	***memangnya***	'actually'	1	143

The examples in (65) and (66) illustrate the respective uses of ***spertinya*** 'it seems' and ***kayanya*** 'it looks like' as sentence adverbs.

(65) **sperti–nya** de suda tinggalkang de punya orang-tua
similar.to–3POSSR 3SG already leave 3SG POSS parent
'**it seems** she already left her parents behind' [081110-005-CvPr.0086]

(66) **kaya–nya** munta~munta
like–3POSSR RDP-vomit
'**it looked like** (he was going to) vomit' [081025-008-Cv.0051]

When suffixed to demonstrative or locative bases, ***-nya*** functions as an emphasizer. This usage of ***-nya*** is very rare, however; attested are only the two lexemes listed in Table 21. Instead, to signal emphasis, Papuan Malay speakers typically employ a modifying demonstrative (see §7.1.2.3); this is shown with the token frequencies given in the 'DEM #' column, which refer to modification with a demonstrative. Examples are presented in (67) and (68).

Table 21: Affixation with ***-nya*** of demonstrative and locative bases

BW	Gloss	Item	Gloss	***-nya*** #	DEM #
itu	'D.DIST'	***itunya***	'it!'	1	19
sini	'L.PROX'	***sininya***	'right here'	1	18

In (67), ***-nya*** is suffixed to locative ***sini*** 'L.PROX', giving the emphatic reading ***sininya*** 'right here'. In (68) the same meaning is expressed with an analytical construction in which a demonstrative modifies the locative.

(67) jatu di sana di sini di **sini–nya** ter–kupas
fall at L.DIST at L.PROX at L.PROX–3POSSR ACL–peel
[About a motorbike accident:] 'he fell (with his bike) over there, here, **right here** (his skin) was peeled off' [081014-013-NP.0001]

(68) a di **sini** **tu** bahaya
ah! at L.PROX D.DIST be.dangerous
'ah, **right here** it is dangerous' [081011-001-Cv.0138]

3.1.6.4. Summary and conclusions

Suffix ***-nya*** is a polyfunctional affix that derives lexemes from nominal, verbal and a number of other bases. Three observations indicate that affixation with ***-nya*** is a productive process: (1) the polyfunctionality of the suffix and the transparent form-function relationship between the derived lexemes and their respective bases, (2) the large number of low frequency words and small number of high frequency words, and (3) the relative token frequencies with most bases having higher frequencies than the affixed lexemes.

Two other observations, however, do not support this conclusion: (1) speakers usually employ alternative strategies that express the same meanings as the suffixed items, and (2) most of the suffixed items were tentatively classified as SI-

borrowings. Also, the findings of a domain analysis suggest that most of the attested tokens can be accounted for in terms of the variables of speaker education levels, topics, and/or role-relations. (Details are discussed in §3.1.8, together with the findings for prefix ***BER-*** 'INTR' and circumfix ***ke-/-ang*** 'NMLZ'.)

In considering these conflicting observations, two findings are given special weight, namely the fact that speakers prefer alternative strategies without affixation, and the findings of the domain analysis. Therefore, it is concluded that in Papuan Malay affixation with ***-nya*** is not used as a productive derivation device. Instead, the suffixed lexemes are best explained as code-switches with Indonesian. (See Mintz 1994 and Sneddon 2010 for a detailed discussion of suffix ***-nya*** in Standard Indonesian and Standard Malay.)

3.1.7. Circumfix *ke-/-ang* 'NMLZ'

Circumfixation with ***ke-/-ang*** 'NMLZ' in Papuan Malay typically derives nouns from nouns. The circumfixed nouns typically denote stable conditions or attributes, as in (69). Some lexical items are also derived from nominal, numeral, or quantifier bases. The affixation process does not appear to be productive, as discussed below.

(69)	jadi	itu	suda	**ke–biasa–ang**	dari	dulu
	so	D.DIST	already	NMLZ–be.usual–NMLZ	from	be.prior

'so already that (has become) a **habit** from the past' [081014-007-CvEx.0063]

The present corpus includes 65 lexical items (258 tokens) circumfixed with ***ke-/-ang***:[121]

1. Circumfixed items with verbal bases (57 items with 239 tokens)
2. Circumfixed items with nominal or numeral/quantifier bases (eight items with 19 tokens)

The corpus also contains three formally complex words with non-compositional semantics, ***kebaktiang*** 'religious service', ***kecelakaang*** 'accident', and ***kegiatang*** 'activity'.

Circumfixed items with verbal bases are discussed in §3.1.7.1, and those with nominal, numeral, or quantifier bases in §3.1.7.2. Pertinent variables of the communicative event that may impact the use of ***ke-/-ang*** are examined in §3.1.8. The main findings on circumfix ***ke-/-ang*** are summarized and evaluated in §3.1.7.3.

3.1.7.1. Circumfixed items derived from verbal bases

The corpus includes 57 ***ke-/-ang*** -circumfixed lexemes (with 238 tokens) with verbal bases, such as bivalent ***turung*** 'descend' or monovalent ***biasa*** 'be usual'. Of the 57 lexemes 52 are nouns and five are accidental verbs.

The 52 circumfixed nouns typically denote stable conditions or attributes in the sense of 'state/quality of being BASE'. Derived words with token frequencies of four

[121] The 65 circumfixed lexemes include 22 hapaxes (P=0.0853); the 57 lexemes with verbal bases include 17 hapaxes (P=0.0711); the ten lexemes verbs with nominal, numeral, or quantifier bases include five hapaxes (P=0.2631).

or more are listed in Table 22. Examples are presented in (70) and (71). All but one of the affixed lexemes are low frequency words (51 lexemes, attested with less than 20 tokens). Moreover, the token frequencies for the respective bases are (much) higher for most of the derived words (41 lexemes). Of the 52 circumfixed nouns, more than half (27 items) were tentatively classified as borrowings from Standard Indonesian (SI-borrowings) (for details see language internal factor (g) on p. 115 in §3.1.1); in Table 22 these items are underlined.

Table 22: Affixation with ***ke-/-ang*** of verbal bases

BW	Gloss	Item	Gloss	***ke-/-ang*** #	BW #
biasa	'be usual'	***kebiasaang***	'habit'	21	185
merdeka	'be independent'	***kemerdekaang***	'freedom'	14	42
baik	'be good'	***kebaikang***	'goodness'	13	182
trang	'be clear'	***ketrangang***	'explanation'	11	4
tindis	'overlap'	***ketindisang***	'k.o.trap'	10	13
turung	'descend'	***keturungang***	'descent'	9	192
sempat	'have enough time'	***kesempatang***	'opportunity'	9	2
benar	'be true'	***kebenarang***	'truth'	9	16
hidup	'live'	***kehidupang***	'life'	8	74
nyata	'be obvious'	***kenyataang***	'reality'	8	1
takut	'feel afraid (of)'	***ketakutang***	'fear'		
sehat	'be healthy'	***kesehatang***	'health'	7	11
jahat	'be bad'	***kejahatang***	'evilness'	7	10
inging	'wish'	***keingingang***	'wish'	6	6
laku	'do'	***kelakuang***	'behavior'	6	5
mo	'want'	***kemawang***	'will'	5	972
lebi	'be more'	***kelebiang***	'surplus'	5	467
saksi	'testify'	***kesaksiang***	'testimony'	5	2
ada	'exist'	***keadaang***	'condition'	4	1,742
betul	'be true'	***kebetulang***	'chance'	4	123
kurang	'lack'	***kekurangang***	'shortage'	4	40

One ***ke-/-ang***-lexeme and its base are given in context: ***kebaikang*** 'goodness' in (70) and its base ***baik*** 'be good' in (71).

(70) dong masi ingat de pu **ke–baik–ang**
2PL pray 1PL 3SG POSS NMLZ–be.good–NMLZ
'they still remember his/her **goodness**' [081110-008-CvNP.0261]

(71) knapa orang bilang, adu ko pu sifat **baik**
why person say oh.no! 2SG POSS characteristic be.good
'why do people say, 'oh no, your character is **good**'' [081110-008-CvNP.0134]

As an extension of its function to derive nouns that denote stable states or attributes, five ***ke-/-ang***-circumfixed lexemes with verbal bases receive an accidental verbal reading, as listed in Table 23.[122] That is, these items indicate that the referent has undergone an accidental or unintentional action or event, such as ***keliatang*** 'be visible' or ***ketinggalang*** 'be left behind'. An example is presented in (72). All five affixed lexemes are low frequency words, attested with less than 20 tokens. Besides, the token frequencies for the respective bases are (much) higher for all of the derived words. Two of the five accidental verbs were tentatively classified as SI-borrowings; in Table 23 these items are underlined.

Table 23: Verbs with circumfix ***ke-/-ang*** with verbal bases

BW	Gloss	Item	Gloss	***ke-/-ang*** #	BW #
liat	'see'	***keliatang***	'be visible'	6	467
tinggal	'stay'	***ketinggalang***	'be left behind'	5	515
taw	'know'	***ketawang***	'be found out'	1	603
lewat	'pass by'	***kelewatang***	'be overly abundant'	1	140
masuk	'enter'	***kemasukang***	'be possessed'	1	261

One ***ke-/-ang***-lexeme and its base are given in context: ***keliatang*** 'be visible' in (72) and ***liat*** 'see' in (73). The verbal status of ***keliatang*** 'be visible' is evidenced by the fact that it is negated with ***tida*** 'NEG' (nominals cannot be negated with ***tida*** 'NEG'; see §5.2 and §13.1.1).

(72) taw~taw orang itu tida **ke–liat–ang**
suddenly person that NEG NMLZ–see–NMLZ
'suddenly, that person wasn't **visible** (any longer)' [080922-002-Cv.0123]

(73) tukang ojek ini dia tida **liat** kolam ini
craftsman motorbike.taxi D.PROX 3SG NEG see big.hole D.PROX
'this motorbike taxi driver, he didn't **see** this big hole' [081015-005-NP.0009]

3.1.7.2. Circumfixed items derived from nominal, numeral or quantifier bases

The corpus includes six ***ke-/-ang***-circumfixed nouns with nominal bases (with eight tokens), and two nouns with numeral or quantifier bases (with 11 tokens), listed in Table 24. The lexemes with nominal bases express 'abstract concepts associated with BASE', while those with numeral or quantifier bases denote stable conditions in the sense of 'state of being BASE'.

All eight affixed lexemes are low frequency words, attested with less than 20 tokens. Moreover, the token frequencies for the respective bases are (much) higher

[122] In discussions about ***ke-/-ang***-circumfixed lexemes with a verbal reading in standard Indonesian or Malay, the circumfix is typically glossed as 'ADVRS' ('adversative') (Englebretson 2003, and Kroeger 2005) or 'NONVOL' ('nonvolitional') (Englebretson (2007).

for all of the derived words. Seven of the eight derived lexemes were tentatively classified as SI-borrowings (for details see language internal factor (g) on p. 115 in §3.1.1); in Table 24 these items are underlined.

Table 24: Affixation with ***ke-/-ang*** of nominal, numeral, and quantifier bases

BW	Gloss	Item	Gloss	***ke-/-ang*** #	BW #
Nominal bases					
budaya	'culture'	***kebudayaang***	'civilization'	2	18
untung	'fortune'	***keuntungang***	'advantage'	2	26
camat	'subdistrict head'	***kecamatang***	'subdistrict'	1	22
hutang	'forest'	***kehutangang***	'forestry'	1	42
pegaway	'civil servant'	***kepegawayang***	'civil service'	1	16
uang	'money'	***keuangang***	'finances'	1	139
Numeral and quantifier bases					
banyak	'many'	***kebanyakang***	'majority'	10	184
satu	'one'	***kesatuang***	'unity'	1	514

One ***ke-/-ang-***item and its base are given in context: ***kebanyakang*** 'majority' in (74) and ***banyak*** 'many' in (75).

(74) smua orang **ke–banyak–ang** mempunyai masala tapi …
all person NMLZ–many–NMLZ have problem but
'all people, the **majority** have problems but …' [080917-010-CvEx.0162]

(75) bua apel di sini **banyak**
fruit apple at L.PROX many
'here are **many** apples' (Lit. 'the apples are **many**') [080922-001a-CvPh.0408]

3.1.7.3. Summary and conclusions

Circumfix ***ke-/-ang*** is a polyfunctional affix that derives lexemes from verbal, nominal, numeral, and quantifier bases. Three observations suggest that affixation with ***ke-/-ang*** is a productive process: (1) the polyfunctionality of the circumfix and the transparent form-function relationship between the derived lexemes and their respective bases, (2) the large number of low frequency words and small number of high frequency words, and (3) the relative token frequencies with most bases having higher frequencies than the affixed lexemes. On the other hand, however, more than half of the circumfixed lexemes were tentatively classified as SI-borrowings.

These findings are further qualified by the results of a domain analysis. These results suggest that most of the attested tokens (with verbal, nominal and numeral/quantifier bases and including the hapaxes) can be accounted for in terms of the variables of speaker education levels, topics, and/or role-relations. Hence, it

cannot be concluded that these items are the result of a productive derivation process. (Details are given in §3.1.8, together with the findings for prefix ***BER-*** 'INTR' and suffix ***-nya*** '3POSSR'.)

Considering the conflicting observations, and taking the findings of the domain analysis as the main decisive factor, it is concluded that in Papuan Malay affixation with ***ke-/-ang*** is not used as a productive derivation device. Instead, the circumfixed lexemes are best explained as code-switches with Indonesian. (See Adelaar 1992, Mintz 1994, and Sneddon 2010 for a detailed discussion of circumfix ***ke-/-ang*** in Standard Indonesian and Standard Malay.)

3.1.8. Variables of the communicative event: Affixes *BER-* 'INTR', *-nya* '3POSSR', and *ke-/-ang* 'NMLZ'

To further investigate the degrees of productivity for prefix ***BER-*** 'INTR', suffix ***-nya*** '3POSSR', and circumfix ***ke-/-ang*** 'NMLZ', a domain analysis was conducted. This analysis focused on the variables of topics and/or role-relations. In all, 244 items[123] with a total of 764 tokens were examined in terms of the variables of topics and/or role-relations:

- 27 items prefixed with ***BER-*** with verbal bases (94 tokens)
- 30 items prefixed with ***BER-*** with nominal, numeral, and interrogative bases (95 tokens)
- 81 items suffixed with ***-nya*** with nominal bases (215 tokens)
- 36 items suffixed with ***-nya*** with verbal bases (82 tokens)
- 5 items suffixed with ***-nya*** with other bases (20 tokens)
- 57 items circumfixed with ***ke-/-ang*** with verbal bases (239 tokens)
- 8 items circumfixed with ***ke-/-ang*** with nominal and numeral/quantifier bases (19 tokens)

For the 244 affixed lexemes, most tokens (698/764 – 91%) can be explained in terms of topics and/or role-relations. The remaining 66/764 tokens (9%) cannot be accounted for in terms of these variables of the communicative event. These tokens occurred when less-educated speakers (-EDC-SPK) conversed with fellow-Papuans of equally low social standing (-STAT) about LOW topics, that is, casual daily-life issues.[124] (See Table 25 and Chart 4; see also Appendix F for detailed tables and charts for each of the three affixes.)

If the affixed items were the result of a productive affixation process, one would expect the percentage of tokens that cannot be explained in terms of speaker education levels, topics, and/or role-relations to be much higher than 9%. Instead,

---

[123] Six items with high token frequencies are excluded from the analysis: five ***BER-***prefixed items with more than 50 tokens and one ***-nya***-suffixed item with 70 tokens. Given their high token frequencies, it was assumed that speakers employ these items regardless of the variables of speaker education levels, topics, and/or role-relations.

[124] As mentioned under Factor 3 'Relationships between interlocutors' in §1.5.1, all of the recorded less-educated speakers belonged to the group of Papuans with lower social status (-STAT), while the recorded Papuans with higher social status (-STAT), such as teachers, government officials, or pastors, were all better educated.

most tokens (91%) seem to be conditioned by these variables of the communicative event. In turn, these findings do not support the conclusion that the affixed lexemes are the result of a productive derivation process. Instead, they seem to be code-switches with Indonesian.

The data presented in Table 25 and Chart 4 are discussed in more detail below.

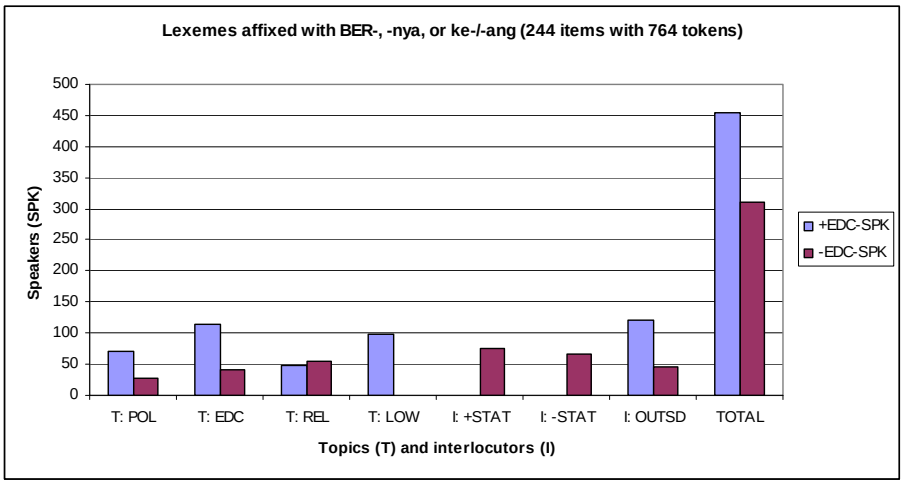
Table 25: Token frequencies for lexemes affixed with ***BER-***, ***-nya***, and ***ke-/-ang*** by speakers, topics, and interlocutors (246 items)[125]

	Topics (TOP)				Interlocutors (ILCT)			Tokens
Affixation with ***BER-*** (57 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	21	34	22	20	---	---	20	117
-EDC-SPK	4	9	11	---	11	**27**[126]	10	72
Subtotal	25	43	33	20	11	**27**	30	189
Affixation with ***-nya*** (122 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	33	35	12	41	---	---	57	178
-EDC-SPK	16	11	28	---	30	**26**	28	139
Subtotal	49	46	40	41	30	**26**	85	317
Affixation with ***ke-/-ang*** (65 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	16	46	14	38	---	---	45	159
-EDC-SPK	7	20	16	---	35	**13**	8	99
Subtotal	23	66	30	38	35	**13**	53	258
TOTAL (246 items)								
	POL	EDC	REL	LOW	+STAT	-STAT	OUTSD	Total
+EDC-SPK	70	115	48	99	---	---	122	454
-EDC-SPK	27	40	55	---	76	**66**	46	310
Total	97	155	103	99	76	**66**	168	**764**

[125] Abbreviations: SPK = speaker, +EDC = better educated, –EDC = less educated, POL = politics, EDC = education, REL = religion, LOW = LOW topics, +STAT = higher social status, -STAT = lower social status, OUTSD = outsider.

[126] This total of 27 tokens includes ten ***berusaha*** 'attempt' tokens produced by the same speaker during the same conversation.

Chart 4: Token frequencies for lexemes affixed with ***BER-***, ***-nya***, or ***ke-/-ang*** by speakers, topics, and interlocutors



The data presented in Table 25 and Chart 4 show that for the 244 affixed lexemes, most tokens (698/764 – 91%) can be explained in terms of speaker education levels, topics, and/or role-relations between the speakers and their interlocutors (ILCT).

Most tokens (454/764 – 59%) were produced by +EDC-SPK, while 310/764 tokens (41%) were produced by -EDC-SPK.

The +EDC-SPK produced half of their tokens (233/454 – 51%) when talking about HIGH topics. The remaining 221 tokens (49%) occurred when +EDC-SPK discussed LOW topics with different interlocutors: 99/221 tokens (45%) with fellow-Papuans (both +STAT and -STAT speakers), and 122 tokens (55%) with an outsider, namely the author (OUTSD).

The -EDC-SPK produced 39% of their tokens (122/310) while discussing HIGH topics. Most of their tokens were produced during conversations about LOW topics (188/310 – 61%). More specifically, 76/188 tokens (40%) occurred when discussing LOW topics with +STAT Papuans, and another 46/188 tokens (24%) were produced during conversations with the author. The remaining 66/188 tokens (35%) occurred during conversations with -STAT Papuans, and therefore cannot be explained in terms of speaker education levels, topics, and/or role-relations. This total of 66 tokens refers to 9% of the 764 tokens attested in the present corpus.[127]

[127] As for the attested hapaxes, the findings of this domain analysis suggest that most are conditioned by the variables of speaker education levels, topics, and/or role-relations: 22/25 hapaxes prefixed with ***BER-***, 59/68 hapaxes suffixed with ***-nya***, and 18/22 hapaxes circumfixed with ***ke-/-ang***. These items are best explained as code-switches with Indonesian. This leaves only three ***BER-***hapaxes, nine ***-nya***-hapaxes, and four ***ke-/-ang***-hapaxes that are unaccounted for in terms of language external factors and that are likely to be the result of a productive word-formation process. This, in turn, decreases the respective P values:

3.2. Compounding

Compounding denotes the "formation of a new lexeme by adjoining two or more lexemes" (Bauer 2003: 40). That is, a compound is "a word that is itself the combination of two or more words" (Booij 2007: 310). In Papuan Malay, however, the demarcation between compounds and phrasal expressions is unclear. That is, neither phonological, morphological, morphosyntactic, nor semantic criteria allow classifying word sequences unambiguously as either compounds or phrasal expressions, as shown in §3.2.1. Therefore, the term "collocation" (Krishnamurthy 2006) is employed as a cover term for such juxtaposed word sequences.

Such collocations always have a binary structure in that they consist of two juxtaposed lexemes. More specifically, three types of collocations can be distinguished, namely endocentric, exocentric, and coordinative collocations, as discussed in §3.2.2. The main points on compounding are summarized in §3.4.

3.2.1. Demarcation of compounds from phrasal expressions

Four different criteria have been suggested to distinguish compounds from phrasal expressions: phonological, morphological, morphosyntactic, and semantic criteria (Aikhenvald 2007: 24). They are discussed in turn in this section.

On phonological grounds, compounds can be distinguished from phrasal expressions in terms of their stress behavior. Compounds typically contain one primary stress, whereas in phrasal expressions each phonological word carries its own stress (Aikhenvald 2007: 25). This criterion also applies to Papuan Malay, as shown in (76) and (77). In the compound ***kacang-hijow*** 'mung bean' in (76), the penultimate syllable carries primary stress, while secondary stress is assigned to the alternating syllable preceding the one carrying the primary stress. By contrast, in the phrasal expression ***kacang hijow*** 'green bean' in (77) each constituent carries its own stress. In fast speech, however, it is difficult to distinguish both constructions on phonological grounds. Instead, the context is the determining factor to establish the intended meaning.

Phonological criteria

(76)	/ˌka.tʃaŋ.ˈhi.dʒow/	**kacang-hijow**	'mung bean'
		bean-be.green	
(77)	/ˈka.tʃaŋ ˈhi.dʒow/	**kacang hijow**	'green bean'
		bean be.green	

In terms of morphological criteria (Aikhenvald 2007: 26), Papuan Malay compounds are not distinct from phrasal expressions. As illustrated in (76) and (77),

---

(1) for three ***BER-***hapaxes P=0.0156 (N=192) as opposed to P=0.0415 for 25 hapaxes (N=602) (N differs for the two P values, as five of the derived lexemes were excluded from the domain analysis),
(2) for nine ***-nya***-hapaxes P=0.0284 (N=317) as opposed to P=0.1757 for 68 hapaxes (N=387) (N differs for the two P values, as one derived lexeme was excluded from the domain analysis), and
(3) for four ***ke-/-ang***-hapaxes P=0.0155 as opposed to P=0.0853 for 22 hapaxes (N=258).

neither construction has an additional morpheme that would mark it as a compound or phrasal expression. Neither are the two constructions distinct in terms of their constituent order as in each case the head precedes the modifier.

On morphosyntactic grounds, compounds are usually distinct from phrasal expressions in that the components of a compound cannot be separated by inserting other morphemes (Aikhenvald 2007: 26). Such an insertion leads to the loss of their compound sense. This criterion also applies to Papuan Malay as shown in (78) and (79). When, for instance, the relativizer ***yang*** 'REL' is inserted in the compound ***lemon-manis*** 'orange' in (78), the compound sense is lost. The result is the phrasal expression ***lemon yang manis*** 'lemon which is sweet' or 'sweet lemon' in (79).

Morphosyntactic criteria

(78) **lemon-manis** 'orange'
lemon-be.sweet

(79) **lemon yang manis** 'citrus which is sweet'
lemon REL be.sweet

In cases such as the compound ***orang-tua*** 'parent' in (80) or the phrasal expression ***orang tua*** 'old person' in (81), however, it is difficult to distinguish both constructions on morphosyntactic grounds. Again, the context is the determining factor to establish the intended meaning.

Ambiguities with respect to morphosyntactic criteria

(80) **orang-tua** 'parent'
person-be.old

(81) **orang tua** 'old person'
person be.old

Semantically, compounds and phrasal expressions can be arranged on a scale from less to more compositional (Aikhenvald 2007: 28); the present corpus does not contain non-compositional compounds with idiosyncratic semantics.[128] This is illustrated in (82) to (86). Less compositional compounds are expressions such as ***kampung-tana*** 'home village' in (82), or ***paduan-swara*** 'choir' in (83). Compounds that are more compositional are those whose meaning is predictable from the meanings of its parts, such as ***air-mata*** 'tears' in (85) or ***tali-prut*** 'intestines' in (85). Very transparent compounds blend into phrasal expressions such as ***uang jajang*** 'pocket money' or 'money for snacks' in (86). On the one hand one could say that ***uang jajang*** 'pocket money' is a compound with an idiosyncratic meaning. On the other hand one could argue that this construction has a phrasal structure that denotes a purpose relation between the nominal head ***uang*** 'money' and its nominal modifier ***jajang*** 'snack'; hence, the construction ***uang jajang*** 'money for snacks' is a phrasal expression and not a compound. Finally, there are phrasal expressions with clear

[128] While Aikhenvald (2007: 28) suggests that compounds can also be compositional, Dryer (2007b: 175) maintains that compounds have "an idiosyncratic meaning not predictable from the meaning of the component parts, as compared with syntactic compounds, in which one noun is modifying a second noun in a productive syntactic construction".

compositional semantics, such as ***air sagu*** 'liquid of the sago palm tree' in (87). (For details on noun phrases with nominal modifiers see §8.2.2.)

Semantic criteria

(82) **kampung-tana** 'home village'
village-ground

(83) **paduang-swara** 'choir'
fusion-voice

(84) **air-mata** 'tears'
water-eye

(85) **tali-prut** 'intestines'
cord-stomach

(86) **uang jajang** 'pocket money' / 'money for snacks'
money snack

(87) **air sagu**
water sago 'liquid of the sago palm tree'

The data presented in this section shows that in Papuan Malay the demarcation between compounds and phrasal expressions is unclear. That is, neither phonological, morphological, morphosyntactic, nor semantic criteria allow the unambiguous classification of word sequences as either compounds or phrasal expressions. What can be said, however, is that some word combinations are "more compoundlike" while other "less compoundlike […] with no clear categorical distinction" along this "cline" (Lieber and Štekauer 2009: 14). Word combinations or collocations range from less compositional two-word expressions such as ***kampung-tana*** 'home village' to those with compositional transparent semantics such as ***air sagu*** 'liquid of the sago palm tree'.

3.2.2. Classification of collocations

Collocations are defined as "word combinations which have developed an idiomatic semantic relation based on their frequent co-occurrence" (Bussmann 1996: 200; see also Krishnamurthy 2006). In Papuan Malay, three types of collocations are found: endocentric, exocentric, and coordinative ones. In the following they are discussed one by one.

In endocentric collocations, one component has head function while the subordinate component has modifying, content-specifying function, denoting "a sub-class of the items denoted by one of their elements" (Bauer 2003: 42). In Papuan Malay endocentric collocations, the head component or "fixed base" always precedes the modifier component or "replaceable 'collocate'" (Kavka 2009: 22), which can be a noun or a stative verb. Semantically, these 'N N' or 'N V' collocations encode different types of relationships between their components such as 'Element/Part-of', 'Subtype-of', or 'Characteristic-of' relations, as illustrated in Table 26. In addition, the corpus contains one collocation in which the modifying component is a numeral: ***segi-empat*** 'quadrangle' (literally 'side-four').

Table 26: Endocentric 'N N/V' collocations

Item	Gloss	Literal translation	Semantic relation
tali-prut cord-stomach	'intestines'	'cord of the stomach'	'Element/Part-of'
lemon-manis lemon-be.sweet	'orange'	'sweet citrus'	'Subtype-of'
kreta-api carriage-fire	'train'	'carriage of fire'	'Characteristic-of'

In exocentric collocations, none of the constituents "functions as its head" (Booij 2007: 313). They "denote something which is not a sub-class" of either of their components; that is, "they are not hyponyms of either of their elements" (Bauer 2003: 42), as shown in Table 27. The collocation ***kepala-batu***, literally 'head-stone', for instance, does not refer to some kind of head. Instead, it denotes a 'pig-headed person'. Likewise, in ***bapa-ade***, literally 'father-younger.sibling' neither of the two components serves as the content-specifying element. These examples also illustrate that exocentric collocations typically consist of two juxtaposed nouns.

Table 27: Exocentric 'N N' collocations

Item	Gloss
kepala-batu head-stone	'pig-headed person'
mata-hari eye-day	'sun'
bapa-ade father-ySb	'father's younger brother' (FyB) / 'mother's younger sister's husband' (MyZH)

The distinction between endocentric and exocentric collocations is not always clear-cut, however, as shown in Table 28. The kinship terms ***bapa-tua*** 'uncle' (literally 'father-be.old') and ***mama-tua*** 'aunt' (literally 'mother-be.old') qualify as exocentric collocations on semantic grounds but as endocentric collocations on syntactic grounds. Both terms are exocentric in that they "denote something which is not a sub-class" of either of their components (Bauer 2003: 42): ***bapa-tua*** does not refer to an 'old father', neither does ***mama-tua*** refer to an 'old mother'. Instead, ***bapa-tua*** denotes a 'parent's older brother' (PoB) or a 'parent's older sister's husband' (PoZH), while ***mama-tua*** designates a 'parent's older sister' (PoZ) or a 'parent's older brother's wife' (PoBW). Syntactically, however, ***tua*** 'be old' is subordinate to the head ***bapa***/***mama*** 'father/mother' and has modifying content-specifying function. Hence, both kinship terms also qualify as endocentric collocations.

Table 28: Endocentric versus exocentric collocations: Ambiguities

Item	Gloss
bapa-tua father-be.old	'parent's older brother' (PeB) / 'parent's older sister's husband' (PeZH)
mama-tua mother-be.old	'parent's older sister' (PeZ) / 'parent's older brother's wife' (PeBW)

Coordinative collocations designate entities made up of two nominal components that "can be interpreted as being joined by 'and'" (Bauer 2009: 351). That is, in such collocations both components "are of semantically equal weight" (Bussmann 1996: 221). The nominal components can be antonyms, synonyms, or different parts or aspects of the designated concept, as shown in Table 29.

Table 29: Coordinative 'N N' collocations

Item	Gloss	Semantic relation
ade-kaka ySb-oSb	'siblings'	Antonyms
kasi-sayang love-love	'deep love'	Synonyms
guntur-kilat thunder-lightening	'thunderstorm'	Different parts/aspects
tete-moyang grandfather-ancestor	'ancestors'	Different parts/aspects

3.3. Reduplication

Reduplication refers to the "repetition of phonological material within a word for semantic or grammatical purposes" (Rubino 2011: 1). In Papuan Malay, as in other Austronesian languages, reduplication is a very productive morphological device to derive new words.

In terms of lexeme formation, three different types of reduplication are attested: full, partial, and imitative reduplication. The most common type is full reduplication which involves the repetition of an entire root, or derived lexeme; bound morphemes cannot be reduplicated. Full reduplication typically applies to content words, although some function words can also be reduplicated. Partial and imitative reduplication occur only very rarely. As for lexeme interpretation, the reduplicated lexemes express a variety of meanings, such as plurality, variety, intensity, continuation, repetitiveness, or vagueness. Reduplication, including lexeme formation and interpretation, is discussed in detail in Chapter 4.

3.4. Summary

This section briefly summarizes the main points on affixation and compounding.

1. Affixation

Affixation in Papuan Malay has very limited productivity. This conclusion is based on an investigation of six affixes: the prefixes ***TER-*** 'ACL', ***PE(N)-*** 'AG', and ***BER-*** 'INTR', the suffixes ***-ang*** 'PAT' and ***-nya*** '3POSSR', and the circumfix ***ke-/-ang*** 'NMLZ'. Given the sociolinguistic profile of Papuan Malay (lack of language awareness, diglossic distribution with Indonesian, negative language attitudes towards Papuan Malay, substantial amount of language contact, and high degree of bilingualism) no productivity testing was conducted, as a substantial amount of interference from Indonesian was expected. This interference would have skewed testees' naïve judgments. Instead, the six affixes were examined in terms of seven language internal and two language external factors considered relevant in establishing the degree of productivity of these affixes.

The results of this investigation are as follows:

1. Papuan Malay ***TER-*** 'ACL' has limited productivity; it indicates accidental or unintentional actions or events. In other eastern Malay varieties and in Standard Malay, the prefix is rather productive; here it also signals accidental or unintentional actions or events.
2. Papuan Malay ***-an*** 'PAT' has limited productivity; it typically designates the patient or result of an action. As for other eastern Malay varieties, the suffix is only mentioned for Ambon Malay; its degree of productivity is unclear. In Standard Malay the suffix is very productive. In both varieties, the suffix also indicates the patient or product of an action.
3. Papuan Malay ***PE(N)-*** 'AG' has marginal productivity, at best. It typically denotes the subject of the action, event, or state specified by the verbal base; some of the affixed lexemes also receive an intensified intransitive or monotransitive reading. As for other eastern Malay varieties, the prefix seems to have retained its productivity only in Ternate Malay. In Standard Malay, the suffix is very productive. In all the other Malay varieties the prefix also denotes the subject of the action or state specified by the verbal base. A verbal interpretation, but not the intensified reading, is also reported for other eastern Malay varieties. In Standard Malay, by contrast, only lexemes with a monovalent stative base can function as monovalent stative verbs.
4. Papuan Malay ***BER-*** 'INTR' is unproductive, whereas in other eastern Malay varieties and Standard Malay the prefix is very productive.
5. Papuan Malay ***-nya*** '3POSSR' and ***ke-/-ang*** 'NMLZ' are unproductive. The same applies to other eastern Malay varieties, while both affixes are very productive in Standard Malay.

2. Compounding

In Papuan Malay the demarcation between compounds and phrasal expressions is unclear. Neither phonological, morphological, morphosyntactic, nor semantic criteria allow the unambiguous classification of two juxtaposed nouns as compounds and phrasal expressions. Therefore, the term "collocation" is employed as a cover term for such word combinations that differ in transparency from non-compositional idiosyncratic semantics to compositional transparent semantics. Three different

types of collocations are attested, endocentric, exocentric, and coordinative ones. Given the lack of a clear demarcation between compounds and phrasal expressions, it remains unclear to what degree compounding is a productive process.