PRODUCTIVITY OF MORPHOLOGICAL PATTERNS AND SOCIAL DOMAIN ANALYSIS IN PAPUAN MALAY

Angela Kluge

SIL International

<angela_kluge@sil.org>

Abstract

This paper describes a multifaceted approach, including several distinct factors, to examining the productivity of morphological patterns in Papuan Malay, a non-standard variety of Malay, spoken in coastal West Papua. More specifically, the paper discusses the question of how the degree of productivity of affixation can be established in diglossic situations where the target language is the LOW variety which experiences interference from a closely related HIGH variety. This applies to Papuan Malay, which has a great deal of language contact with Indonesian, the official language in the area, and can be exemplified with prefix BER- ‘VBLZ’. This approach is offered as a case study for taking into account both language internal and external factors when examining the productivity of morphological patterns in contexts where more traditional approaches are less likely to provide reliable data. Following an introduction to the topic, nine language internal and external factors are described. These factors were employed to examine to what degree Papuan Malay speakers use prefix BER- ‘VBLZ’ to create new words. These more general remarks are followed by a discussion of prefix BER- ‘VBLZ’ as it is used in Papuan Malay. The paper closes with a summary and conclusions.

Keywords: Papuan Malay, morphology, affixation, social domain analysis

ISO 639-3 codes: pmy

1. Introduction

Papuan Malay is spoken in coastal West Papua, where it is the language of wider communication and the first or second language for an ever-increasing number of people of the area (1,100,000 to 1,200,000 speakers).1 Papuan Malay is in diglossic distribution with another, closely related Malay language, namely Indonesian. Papuan Malay functions as the LOW variety while Indonesian is the HIGH variety. Papuan Malay is not officially recognized, and therefore not used in formal government or educational settings or for religious preaching. The language is used, however, in all other domains, including unofficial use in formal settings, and, to some extent, in the public media.

The discussion here is based on a 16-hour corpus of narratives and spontaneous conversations between Papuan Malay speakers recorded in the Sarmi area, along West Papua’s north coast. The corpus contains a considerable number of morphologically complex lexical items. The most commonly employed (historical) affixes include the prefix BER- ‘VBLZ’, which is the focus of this paper. Other rather frequently occurring affixes are the prefixes PE(N)- ‘AG’ and TER- ‘ACL’, the suffixes -ang ‘PAT’ and -nya ‘3POSSR’, and circumfix ke-/-ang ‘NMLZ’. For each of the six affixes, the base words are lexical roots; typically, the affixed roots are verbs or nouns.

1 This conservative population estimate is based on Kluge’s (2014: 5) assessment.
Generally speaking, morphological patterns are said to be productive if language users employ them ‘to create new well-formed complex words’ by systematically extending the pattern ‘to new cases’ (Booij 2007: 67). Morphological patterns are considered to be unproductive, by contrast, when the respective morphological rule ‘is not used for coining new words’ but ‘has become obsolete’ (2007: 68). The productivity of a given pattern is a matter of degree, though, as scholars such as Aikhenvald (2007: 49–58), Bauer (1983: 62–100), Booij (2007: 67–71), or Pike (1967: 169–172) point out. 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). Therefore, Bauer (1983) concludes that productivity is best viewed as a ‘cline’ or a ‘scalar phenomenon’ (2001: 97).

One technique to explore the productivity of prefix **BER-** ‘VBLZ’ 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, however, no productivity tests were conducted to determine whether and to what extent prefix **BER-** ‘VBLZ’ can be attached to Papuan Malay roots to derive new lexical items. Tests such as the mentioned lexical-decision tasks were considered unfeasible for three reasons, namely the sociolinguistic profile of Papuan Malay, the high degree of linguistic relatedness between both varieties, and the formal setting of a test situation.

The sociolinguistic profile of the Papuan Malay speech variety and its speech communities is characterized by:

- Functional distribution of Papuan Malay as the LOW variety, and Indonesian as the HIGH variety, in terms of Ferguson’s (1972) notion of diglossia;
- Positive to somewhat ambivalent language attitudes toward Papuan Malay; and
- Lack of language awareness of many Papuan Malay speakers about the status of Papuan Malay as a language distinct from Indonesian.

Given this sociolinguistic profile together with the fairly high degree of linguistic relatedness between Papuan Malay and Indonesian, as well as the formal setting of a test situation, an undesirable amount of interference from Indonesian was anticipated.

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

---


3 As Bauer (2001: 125) points out, however, there is an ongoing discussion among scholars ‘whether productivity is a gradable/scalar phenomenon or not’.

4 The sociolinguistic profile of Papuan Malay is discussed in more detail in Kluge (2014: 37–42).
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.

Therefore, an alternative, multifaceted approach was chosen to explore the degree of productivity of prefix BER- ‘VBLZ’. This approach is described in §2.

2. Language factors for exploring the productivity of prefix BER- ‘VBLZ’

To examine whether and to what degree Papuan Malay speakers use BER-affixation to create new words, a multifaceted approach was employed. That is, the prefix and the derived words were examined in terms of six pertinent language internal and three language external factors. None of the factors was taken in isolation, though. Instead, the findings pertaining to all nine factors were taken together as an indication of the degree of productivity for BER-affixation. The language internal factors are discussed in §2.1, and the external ones in §2.2.5

2.1 Language internal factors

Prefix BER- ‘VBLZ’ was examined with respect to the following six language internal factors: (1) syntactic properties, (2) type frequencies, token frequencies, and hapaxes, (3) form-function relationship between the derivations and their base words, (4) alternative strategies, (5) formally complex words with non-compositional semantics, and (6) status of the prefixed lexemes as part of the Papuan Malay lexicon or as code-switches with Indonesian.

1. Syntactic properties

If BER- ‘VBLZ’ is ‘polyfunctional’, that is, if it can take bases from more than one lexical category, this is taken as evidence that the process is more productive (Booij 2002: 90–91; see also Zwanenburg 2000).

2. Type frequencies, token frequencies and hapaxes

If BER- ‘VBLZ’ is represented by a large number of words (high type frequency) which, in turn, have low token frequencies, this is taken as an indication that the affixation process is more likely to be productive. (For the purposes of this analysis, type frequencies of ten or more are considered as ‘(relatively) high’ while token frequencies of less than 20 are considered as ‘(relatively) low’.)

This conclusion is based on Plag’s (2006) study which shows that 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). (See also Hay 2001; Hay and Baayen 2002.)

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). Baayen (1992: 115) proposes the following formula to calculate the degree of productivity P of a given morphological pattern: 

\[ P = \frac{n_i}{N} \]

Given the limited size of the corpus, though, it is difficult to determine which of the attested hapaxes are Papuan Malay neologisms and which ones merely reflect the size of the corpus. Furthermore, 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 are given in footnotes throughout this paper.

3. Form-function relationship between the derivations and their base words

Following Booij (2007: 240), one ‘necessary’ albeit not ‘sufficient’ condition for the productivity of 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 prefixed lexemes and their bases is transparent, this is taken as evidence that BER-affixation is more

---

5 The same analysis was applied to the prefixes PE(N)- ‘AG’ and TER- ‘ACL’, the suffixes -ang ‘PAT’ and -nya ‘3POSSR’, and the circumfix ke-/-ang ‘NMLZ’ (for details see Kluge 2014: 111–166.)
productive. If, by contrast, this relationship is opaque, this is taken as evidence that the process is less productive.

Pairs of words in which the prefixed lexemes and their respective bases share the same semantics, are not taken as parts of a larger derivational paradigm. These sets are instead rather 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 2009: 20). Therefore, for pairs of words with the same semantics, the unaffixed bases are taken to be the native Papuan Malay lexemes, while the prefixed words are taken to be code-switches with the corresponding Indonesian lexemes.

4. Alternative strategies
If speakers prefer to use alternative strategies that do not involve BER-affixation and that express the same meanings as the prefixed forms, these alternative strategies are taken as evidence that the affixation process is less productive.

5. Formally complex words with non-compositional semantics
If the corpus contains a large number of formally complex words with non-compositional semantics, this is taken as evidence that BER-affixation is less productive. Their non-compositional semantics suggest that these lexemes are either lexicalized forms or code-switches with Indonesian.

6. Status of BER-prefixes as part of the Papuan Malay lexicon or as borrowings from Indonesian
If a large number of the prefixed lexemes are not part of the Papuan Malay lexicon but constitute code-switches with Indonesian, this is taken as evidence that the derivation process is less productive.

While sources such as Jones (2007), or Tadmor (2009) allow the identification of foreign, non-Malay loan words in the corpus, they do not allow identifying borrowings from Indonesian. Therefore, an alternative approach was chosen to explore whether the prefixed lexemes are part of the Papuan Malay lexicon or constitute code-switches with Indonesian.

All attested BER-prefixes 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 familiarity with both languages, the consultant classified the prefixed lexemes as ‘Papuan Malay’ or ‘borrowings from Indonesian’. The statement that a lexeme is considered to be part of the Papuan Malay lexicon does not imply, however, that the respective lexeme does not exist in other Malay varieties as well. Across Southeast Asia, all Malay varieties have large sets of shared lexical items. This also applies to Papuan Malay, as well as to other eastern Malay varieties, or Indonesian.

The consultant’s tentative classification is of course subjective and not necessarily representative. It provides, however, one more piece of evidence as to the potential productivity of prefix BER- ‘VBLZ’.

2.2 Language external factors
Prefix BER- ‘VBLZ’ was also examined in terms of three language external factors, that is, three variables of the communicative event. The underlying question for this analysis was whether the prefix is employed without sociolinguistic restrictions or whether its use is conditioned by variables of the communicative event, pertaining to Fishman’s (1965: 86) ‘domains of language choice’. The pertinent factors influencing language choices are the topics discussed, the relationships between the interlocutors, and the locations where the communication takes place (1965: 67). Speaker education levels are a fourth pertinent factor.

For the Papuan Malay study, the pertinent ‘domains of language choice’ are (1) the speaker education levels, (2) the topics, and (3) the relationships between the interlocutors, all of which are discussed in the following. The locations of communication were not considered pertinent domains as all recorded conversations took place in the same informal setting of the home. (The sociolinguistic profile of Papuan Malay is discussed in detail in Kluge 2014: 37–42.)

If the use of prefix BER- ‘VBLZ’ seems to be conditioned by these variables of the communicative event, this is taken as evidence that the affixation process is less productive.
1. Speaker education levels
In diglossic situations, the LOW variety is known by everyone while the HIGH variety is acquired through formal education (Ferguson 1972). This also applies to the diglossic distribution of Papuan Malay and Indonesian. While Papuan Malay is known by almost everyone in West Papua’s coastal areas, knowledge of Indonesian depends on the speakers’ education levels. That is, 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. Hence, if better-educated speakers employ prefix \textit{BER-`VBLZ} considerably more often than less-educated ones, this is taken as evidence that the prefixed lexemes are not the result of a productive process but that they constitute code-switches with Indonesian.

2. Topics
The topics under discussion may also bring ‘another language to the fore’ given that ‘certain topics are somehow handled better in one language than in another’ (Fishman 1965: 71). This notion of topical regulation suggests that Papuan Malay speakers consider Indonesian, and not Papuan Malay, the appropriate language to use when they discuss HIGH topics associated with formal domains such as politics, education, or religion. Therefore, if Papuan Malay speakers use prefix \textit{BER-`VBLZ} much more often when discussing HIGH topics than when discussing casual daily-life issues (LOW topics), this is taken as evidence that the prefixed lexemes are code-switches with Indonesian. This applies especially to less-educated Papuans. Better-educated Papuans already display a general tendency to include Indonesian features when speaking Papuan Malay, although this tendency is more pronounced when discussing HIGH topics.

3. Relationships between interlocutors
Language behavior is not only influenced by the topics of communication and speaker education levels, but also by role relations. That is, individual speakers display certain language behaviors depending on the role relations between them (Fishman 1965: 76).

Due to 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. This has to do with the fact that the use of features from the HIGH variety serves to signal social inequality, distance, and formality. The use of the LOW variety, that is Papuan Malay, by contrast, indicates intimacy, informality, and equality. (See also Fishman 1965: 70.)

Therefore, if speakers use prefix \textit{BER-`VBLZ} 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 prefixed lexemes are code-switches with Indonesian. Again, this applies especially to less-educated Papuans. As mentioned, better-educated Papuans already show a general tendency to ‘dress-up’ their Papuan Malay with Indonesian features, although this tendency is more marked when interacting with group outsiders, such as the author.

3. Analysis of prefix \textit{BER-`VBLZ}’
The Papuan Malay corpus contains 78 lexemes, prefixed with \textit{BER-`VBLZ} with a total of 695 tokens. Typically, the prefix derives monovalent verbs from verbal bases, such as \textit{buat ‘make}’ in (1), or from nominal bases such as \textit{malam ‘night}’ in (2). Hence, \textit{BER-} is glossed as ‘VBLZ’ (‘verbalizer’).

(1) \textit{kalo ko ber–buat baik dalam kluarga itu …}
   if 2SG VBLZ–make be.good in family D.DIST
   ‘if you behaved well in that family (you won’t have difficulties)’ [081115-001a-Cv.0303]

(2) \textit{sampe di Webro tong TRU-nai naik ber–malam satu malam situ}
   reach at Webro 1PL ascend ascend VBLZ–night one night L.MED
   ‘having arrived in Webro, we went up[TRU], went up (the beach) to \textit{overnight} there (for) one
   night’ [080917-008-NP.0118]

More specifically, the corpus includes 29 prefixed lexemes with verbal bases (227 tokens) and 30 lexemes with nominal bases (362 tokens). In addition, two lexemes are derived from numeral bases (7
Angela KLUGE | Productivity of Morphological Patterns & Social Domain Analysis in Papuan Malay | JSEALS 9 (2016)

tokens) and one from a quantifier base (6 tokens).\textsuperscript{6} In addition, the corpus also contains 16 formally complex words with non-compositional semantics (93 tokens). Examples are \textit{bertriak} ‘scream’, \textit{berjuang} ‘struggle’, or \textit{berlabu} ‘anchor’.

Before discussing BER-affixation in terms of the six internal and three external language factors in §3.2 and §3.3, respectively, the allomorphy of BER- is investigated in §3.1. The main findings on prefix BER- are summarized and evaluated in §4.

3.1 Allomorphy of BER- ‘VBLZ’
Prefix BER- ‘VBLZ’ has two allomorphs, \textit{ber}- and \textit{ba}-.

Allomorph \textit{ber}- has four realizations which are effected by morphologically conditioned phonological rules, namely /\textit{ber}/, /\textit{br}/, /\textit{bl}/, and /\textit{bɛr}/. In particular, these realizations are conditioned by the word-initial segment of the base word, as shown in \textit{Table 1}. Typically, the prefix is realized as /\textit{ber}/. With an onset vowel, however, the prefix is very commonly realized as /\textit{br}/. When prefixed to \textit{ajar} ‘teach’ the prefix is realized as /\textit{bl}/. Finally, when affixed to \textit{kerja} ‘work’ or \textit{brapa} ‘several’, it is realized as /\textit{bɛ}/.

\textbf{Table 1: Realizations of allomorph \textit{ber}-}

<table>
<thead>
<tr>
<th>\textit{ber}-Base</th>
<th>Orthogr.</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/\textit{ber}–\textit{dʒuaŋ}/</td>
<td>\textit{berjuang}</td>
<td>‘struggle (for)’</td>
</tr>
<tr>
<td>/\textit{br}–\textit{aŋkat}/</td>
<td>\textit{brangkat}</td>
<td>‘leave’</td>
</tr>
<tr>
<td>/\textit{bl}–\textit{adʒar}/</td>
<td>\textit{blajar}</td>
<td>‘study’</td>
</tr>
<tr>
<td>/\textit{bɛ}–\textit{kerdʒa}/</td>
<td>\textit{bekerja}</td>
<td>‘work’</td>
</tr>
<tr>
<td>/\textit{bɛ}–\textit{brapa}/</td>
<td>\textit{bebrapa}</td>
<td>‘be several’</td>
</tr>
</tbody>
</table>

Allomorph \textit{ba}- occurs much less frequently. The corpus includes only ten items with a total of 32 tokens, as listed in \textit{Table 2}. In the corpus, some of these items are also realized with allomorph \textit{ber}-.

In realizing prefix BER- ‘VBLZ’ typically as allomorph \textit{ber}- rather than as \textit{ba}-, Papuan Malay is distinct from other eastern Malay varieties such as Ambon Malay (van Minde 1997: 95), Banda Malay (Paauw 2009: 249), Kupang Malay (Steinhauer 1983: 46), Manado Malay (Stoel 2005: 18), and North Moluccan / Ternate Malay (Taylor 1983: 18; Voorhoeve 1983: 4; Litamahuputty 2012: 125). In these varieties, the prefix is always realized as \textit{ba}-.

\textit{In the context of the Papuan Malay study, a 2,459-item word list was elicited (for details see Kluge 2014: 57–58).}
### 3.2 Language internal factors

Prefix **BER-** ‘VBLZ’ was examined in terms of six language internal factors, namely its syntactic properties (§3.2.1), its type frequencies and token frequencies (§3.2.2), the form-function relationship between the derivations and their base words (§3.2.3), alternative strategies (§3.2.4), formally complex words with non-compositional semantics (§3.2.5), and the status of the affixed lexemes as part of the Papuan Malay lexicon or as code-switches with Indonesian (§3.2.6).

#### 3.2.1 Syntactic properties

Prefix **BER-** ‘VBLZ’ is polyfunctional in that it can be attached to different types of bases, namely verbal, nominal, numeral, and quantifier ones. Typically, the derivations are monovalent verbs. The present corpus includes 29 **BER-** prefixed lexemes (227 tokens) with verbal bases and 30 prefixed lexemes (362 tokens) with nominal bases. Another three prefixed lexemes have numeral or quantifier bases (13 tokens).

Of the 29 lexemes with verbal bases, 11 bases are monovalent such as stative **diam** ‘be quiet’ or dynamic **jalang** ‘walk’. The derived lexemes are monovalent verbs. The remaining 18 items have bivalent bases, such as **buru** ‘hunt’ or **kerja** ‘work’. Five of the derived lexemes with bivalent bases can be used mono- or intransitively, such as **berpikir** ‘think’ or **berbuat** ‘make’. The same applies to their bases. This has to do with the fact that in Papuan Malay bivalent verbs allow but do not require two syntactic arguments (Kluge 2014: 221–223). Another 11 derivations have intransitive uses only, such as **berbicara** ‘speak’ or **brikut** ‘follow’. These 16 items have the same semantics as their bivalent bases. For the remaining two prefixed lexemes, **blajar** ‘study’ and **brangkat** ‘leave’, the semantics are distinct from those of their respective bases **ajar** ‘teach’ and **angkat** ‘lift’. The former derivation has monotransitive as well as intransitive uses, while the latter has intransitive uses only. These and other examples are presented in Table 3.

#### Table 2: Realizations of allomorph ba-

<table>
<thead>
<tr>
<th>ba-Base</th>
<th>Orthogr.</th>
<th>Gloss</th>
<th>ba- #</th>
<th>ber- #</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ba–kala/</td>
<td>bakalay</td>
<td>‘fight’</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>/ba–taria/</td>
<td>bertriak</td>
<td>‘scream’</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>/ba–bini/</td>
<td>babingung</td>
<td>‘be confused’</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>/ba–diam/</td>
<td>badiam</td>
<td>‘be quiet’</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>/ba–dir/</td>
<td>berdiri</td>
<td>‘stand’</td>
<td>1</td>
<td>54</td>
</tr>
<tr>
<td>/ba–gaja/</td>
<td>bagaya</td>
<td>‘put on airs’</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/ba–git/</td>
<td>bagigit</td>
<td>‘bite’</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>/ba–kumis/</td>
<td>bakumis</td>
<td>‘have a beard’</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>/ba–isi/</td>
<td>baisi</td>
<td>‘be muscular’</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>/ba–mekap/</td>
<td>bamekap</td>
<td>‘wear make-up’</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

9 The root is realized as /triak/ when speakers employ allomorph ber-, whereas it is realized as /taria/ when speakers use allomorph ba-.
Table 3: BER-affixation of verbal bases

<table>
<thead>
<tr>
<th>BER-Base</th>
<th>Item</th>
<th>Gloss</th>
<th>BER-#</th>
<th>Base #</th>
</tr>
</thead>
<tbody>
<tr>
<td>tobat</td>
<td>bertobat</td>
<td>‘repent’</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>beda</td>
<td>berbeda</td>
<td>‘be different’</td>
<td>7</td>
<td>34</td>
</tr>
<tr>
<td>bahaya</td>
<td>berbahaya</td>
<td>‘be dangerous’</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>diam</td>
<td>badiam</td>
<td>‘be quiet’</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>bingung</td>
<td>berbingung</td>
<td>‘be confused’</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>jalang</td>
<td>berjalang</td>
<td>‘walk’</td>
<td>1</td>
<td>480</td>
</tr>
<tr>
<td>sandar</td>
<td>bersandar</td>
<td>‘lean’</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

Bivalent bases: Bases and prefixed lexemes have same semantics

| Prefixed lexemes: Derivations with monotransitive and intransitive uses |
|---------------------------|-----|-----|-----|
| beruru                  | berburu | ‘hunt’ | 10 | 5 |
| beru                    | berbuat | ‘make’ | 7 | 100 |
| berpikir                | berpikir | ‘think’ | 8 | 102 |
| berharap                | berharap | ‘hope’ | 1 | 8 |
| bribut                  | bribut   | ‘trouble’ | 1 | 5 |

Prefixes: Derivations with intransitive uses

| bicara               | berbicara | ‘speak’ | 7 | 333 |
| kerja                | bekerja   | ‘work’ | 5 | 191 |
| uba                   | bruba     | ‘change’ | 5 | 9 |
| gabung               | bergabung | ‘join’ | 4 | 3 |
| maing                | bermaing  | ‘play’ | 3 | 113 |
| tindak               | bertindak | ‘act’ | 2 | 1 |
| ikut                 | brikut    | ‘follow’ | 1 | 259 |

Bivalent bases: Bases and prefixed lexemes have distinct semantics

| ajar (‘teach’)    | blajar | ‘study’ | 51 | 41 |
| angkat (‘lif’)   | brangkat | ‘leave’ | 82 | 81 |

In addition, the corpus contains 30 BER-prefixed lexemes with nominal bases (362 tokens), such as doa ‘prayer’ or arti ‘meaning’. Also attested are two lexemes with numeral bases (7 tokens), namely satu ‘one’ and empat ‘four’, as well as one lexeme with a quantifier base (6 tokens), namely brapa ‘several’. The derived lexemes are monovalent verbs. Examples are presented in Table 4.

Table 4: BER-affixation of nominal, numeral, or quantifier bases

<table>
<thead>
<tr>
<th>BER-Base</th>
<th>Gloss</th>
<th>Item</th>
<th>Gloss</th>
<th>BER-#</th>
<th>Base #</th>
</tr>
</thead>
<tbody>
<tr>
<td>doa</td>
<td>‘prayer’</td>
<td>berdoa</td>
<td>‘pray’</td>
<td>136</td>
<td>20</td>
</tr>
<tr>
<td>arti</td>
<td>‘meaning’</td>
<td>brarti</td>
<td>‘mean’</td>
<td>89</td>
<td>7</td>
</tr>
<tr>
<td>dosa</td>
<td>‘sin’</td>
<td>berdosa</td>
<td>‘sin’</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>kebung</td>
<td>‘garden’</td>
<td>berkebung</td>
<td>‘do farming’</td>
<td>3</td>
<td>61</td>
</tr>
<tr>
<td>malam</td>
<td>‘night’</td>
<td>bermalam</td>
<td>‘overnight’</td>
<td>2</td>
<td>191</td>
</tr>
<tr>
<td>ana</td>
<td>‘child’</td>
<td>brana</td>
<td>‘give birth’</td>
<td>1</td>
<td>739</td>
</tr>
<tr>
<td>dara</td>
<td>‘blood’</td>
<td>berdara</td>
<td>‘bleed’</td>
<td>1</td>
<td>27</td>
</tr>
</tbody>
</table>

| satu     | ‘one’ | bersatu | ‘be one’ | 6 | 516 |
| empat    | ‘four’ | berempat | ‘be four’ | 1 | 66 |
| brapa    | ‘several’ | bebrapa | ‘be several’ | 6 | 109 |
3.2.2 Type and token frequencies

The vast majority of the derivations has rather low token frequencies of 10 or less (56/62 – 90%). This applies to 27/29 lexemes with verbal bases (93%), as shown in (3) to (5). (For more examples, see Table 3.) The exceptions are blajar ‘study’ (51 tokens) and brangkat ‘leave’ (82 tokens). The high frequencies for both derivations are due to the fact that they have different semantics than their bases, as discussed in §3.2.3.1.

BER-derivations with verbal bases: Low token frequencies

(3) berburu ‘hunt’ (10 tokens)
(4) bertobat ‘repent’ (8 tokens)
(5) berbicara ‘speak’ (7 tokens)

Likewise, most of the prefixed lexemes with nominal, numeral, or quantifier bases (29/33 – 88%) have rather low token frequencies, as illustrated in (6) to (10). (For more examples, see Table 4.) Only four lexemes with nominal bases have higher token frequencies, namely berdoa ‘pray’ (136 tokens), brarti ‘mean’ (89 tokens), berdiri ‘stand’ (55 tokens), and berusaha ‘attempt’ (25 tokens). The monovalent verb berdiri ‘stand’ is an exception. Historically related to the noun diri ‘self’, it does not have a transparent form-function relationship to its base.

BER-derivations with nominal, numeral, or quantifier bases: Low token frequencies

(6) berkebung ‘do farming’ (3 tokens)
(7) bermalam ‘overnight’ (2 tokens)
(8) brana ‘give birth’ (1 token)
(9) bersatu ‘be one’ (6 tokens)
(10) bebrapa ‘be several’ (6 tokens)

Moreover, for most of the derivations, the respective bases have higher token frequencies than the prefixed lexemes (43/62 items – 70%). This applies to 22/29 lexemes with verbal bases (76%), as shown in (11) to (13). Of the seven exceptions, six derivations have higher tokens frequencies than their bases, including items such as berburu ‘hunt’, bertobat ‘repent’, blajar ‘study’, or brangkat ‘leave’. The remaining exception is berbahaya ‘be dangerous’; both the derivation and its base are attested with three tokens. (See also Table 3.)

BER-derivations with verbal bases: Token frequencies of derivations versus bases

(11) berjalan ‘walk’ (1 derived token vs. 480 base tokens)
(12) berbicara ‘speak’ (71 derived tokens vs. 333 base tokens)
(13) bekerja ‘work’ (5 derived tokens vs. 191 base tokens)

A similar pattern applies to derivations with nominal, numeral, or quantifier bases: for 21/33 derivations (64%), the respective bases have higher token frequencies than the prefixed lexemes, as illustrated in (14) to (17).

BER-derivations with nominal, quantifier, or numeral bases: Token frequencies of derivations versus bases

(14) bermalam ‘overnight’ (2 derived tokens vs. 191 base tokens)
(15) berbahasa ‘speak’ (2 derived tokens vs. 136 base tokens)
(16) berempat ‘be four’ (1 derived token vs. 66 base tokens)
(17) bebrapa ‘be several’ (6 derived tokens vs. 109 base tokens)

Among the 12 exceptions, six derivations have higher token frequencies than their bases. The six exceptions include items such as berdoa ‘pray’, brarti ‘mean’, or berdosa ‘sin’. For the remaining eight exceptions, the token frequencies for the derivations and their bases are equally low, namely one. (See also Table 4.)
3.2.3 Form-function relationship between derivations and bases

Affixation with BER- of verbal bases derives lexemes that typically have the same semantics as their respective bases, as described in §3.2.3.1. Derivations with nominal, numeral, or quantifier bases, by contrast, have different semantics than their bases, as discussed in §3.2.3.2.

3.2.3.1 Semantics of derivations with verbal bases

Almost all attested derivations with verbal bases (27/29 – 93%) have the same semantics as their bases. This applies to the 11 BER-prefixed lexemes with monovalent bases and to 16 of the 18 prefixed lexemes with bivalent bases, as shown in (18) to (28). The two exceptions are presented in (29) to (34).

The shared semantics between BER-prefixed lexemes and their monovalent bases are illustrated with stative bingung ‘be confused’ and its derivation berbingung ‘be confused’ in (18) and (19), and with dynamic ibada ‘worship’ and its derivation beribada ‘worship’ in (20) and (21), respectively.

**BER-affixation of monovalent verbal bases**

(18) 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]

(19) nanti di skola baru kamu ba-bingung dengang bahasa Inggris
very.soon at school and.then 2PL VBLZ–be.confused with language English
[Addressing lazy students:] ‘later in school, then you’ll be **confused** about English’ [081115-001a-Cv.0151]

(20) 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]

(21) nanti kita ber-ibada selesay malam ka baru sa pergi
very.soon 1PL VBLZ–worship finish night maybe and.then 1SG go
‘later, after we have **worshipped**, maybe in the evening, and then I’ll go (there)’ [080918-001-CvNP.0016]10

The shared semantics between BER-prefixed lexemes and their bivalent bases are demonstrated in (22) to (26). As mentioned in §3.2.1, Papuan Malay bivalent verbs can be used mono- or intransitively. The same applies to some of their BER-derivations. This is shown with pikir ‘think’ and berpikir ‘think’ in (22) to (25). The monotransitive uses of the base and its derivation are shown in (22) and (23), and their intransitive uses in (24) and (25), respectively.

**BER-affixation of bivalent verbal bases: Mono- and intransitive uses**

(22) 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]

(23) … 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]

---

10 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’.
(24) *skarang orang su tra pikir tentang hal ke–benar–ang*  
now person already NEG think about thing NMLZ–be.true–NMLZ  
‘nowadays, the people already don’t *think* about things (related to) truth’ [081006-032-Cv.0016]

(25) … *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]

Most *BER*-prefixed lexemes with bivalent bases, however, have only intransitive uses, whereas their respective bases can be used mono- or intransitively. This is illustrated with *bicara* ‘speak’ and *berbicara* ‘speak’ in (26) to (28).

**BER-affixation of bivalent verbal bases: Intransitive uses only**

(26) *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]

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

(28) *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]

Only two of the attested *BER*-prefixed lexemes with bivalent bases have semantics that are distinct from those of their bases: bivalent *ajar* ‘teach’ and prefixed *blajar* ‘study’, and bivalent *angkat* ‘lift’ and prefixed *brangkat* ‘leave’, as shown in (29) to (34). Both *ajar* ‘teach’ and *blajar* ‘study’ are used intransitively as in (29) and (30), or monointransitively as in (31) and (32), respectively; in each case the two lexemes maintain their distinct semantics.

**BER-affixation of bivalent verbal bases: Mono- and intransitive uses and distinct semantics vis-à-vis the base**

(29) *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]

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

(31) *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]

(32) *Ise de … ikut bahasa Inggris, bl–ajar kursus bahasa Inggris dulu*  
Ise 3SG follow language English VBLZ–teach course language English be.prior  
‘Ise will participate in an English course, *(she’ll)* *study* an English language course first’ [081025-003-Cv.0223]
Bivalent *angkat* ‘lift’ and prefixed *brangkat* ‘leave’ also have distinct semantics. In addition, they also have distinct distributions. The base *angkat* ‘lift’ is always used montransitively, as in (33), whereas *brangkat* ‘leave’ only has intransitive uses, as in (34).

**BER-affixation of bivalent verbal bases: Intransitive uses only and distinct semantics vis-à-vis the base**

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

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

In summary, with the exception of *blajar* ‘study’ and *brangkat* ‘leave’, *BER*-prefixed verbs and their respective bases have the same semantics. This indicates that *BER*-affixation of verbal bases 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 code-switches with Indonesian.

Given these properties, Papuan Malay *BER*- contrasts with the corresponding prefix in other Malay varieties. In most eastern Malay varieties, the 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 2009: 249–250),11 Manado Malay (Stoel 2005: 18–22), and North Moluccan / Ternate Malay (Taylor 1983: 18; Litamahuputty 2012: 125–127). In Kupang Malay (Steinhauer 1983: 46–49) and Larantuka Malay (Paauw 2009: 249–254–255), the prefix typically signals durative and reciprocal meanings. In Standard Indonesian, the main function of the corresponding prefix *ber-* is to create monovalent verbs with reflexive meanings (Englebretson 2003: 131; 2007: 96; Ewing 2005: 251).

### 3.2.3.2 Semantics of derivations with nominal, numeral, or quantifier bases

Affixation with *BER-* of nominal, numeral, or quantifier bases derives lexemes that have distinct semantics vis-à-vis their bases. The derived monovalent verbs have the general meaning of ‘be/have/do BASE’. This is shown with *bersodara* ‘be siblings’ in (35), *berduri* ‘have thorns’ in (36), and *berkebung* ‘do farming’ in (37).12

**BER-affixation of nominal bases**

(35) *… de punya bapa deng bapa ber–sodara*  
3SG POSS father with father VBLZ–sibling  
‘his father and I (‘father’) are siblings’ [081110-008-CvNP.0004]

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

(37) *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]

---

11 For Banda Malay, Paauw (2009: 249) reports that *ba-* does not form verbs with reflexive meanings.

12 The corpus includes one exception, namely the monovalent verb *berdiri* ‘stand’. Historically related to the noun *diri* ‘self’, it does not have a transparent form-function relationship to its base.
The examples in (35) to (37) illustrate the transparent form-function relationship between the derivations and their respective bases. Two observations are made, however, both of which suggest that BER-affixation of nominal bases is not a productive process. The first observation relates to the language internal factor 4 (p. 223) and has to with ‘Alternative strategies’ (see §3.2.4). The second observation relates to the language internal factor 5 (p. 223) and has to do with ‘Formally complex words with non-compositional semantics’ (see §3.2.5).

3.2.4 Alternative strategies

The first observation regarding the form-function relationship between the attested BER-prefixed lexemes and their nominal bases is that Papuan Malay speakers rather use alternative strategies than BER-derivations. That is, the attested data indicate that Papuan Malay speakers prefer to employ analytical constructions to express the meanings conveyed by the prefixed items, as illustrated in (38) to (43). To express ‘be BASE’, speakers typically employ a nominal predicate such as ade-kaka ‘siblings’ in (38), rather than the prefixed form such as brade-kaka ‘be siblings’ in (39).

BER-affixation of nominal bases: Alternative strategy for ‘be BASE’

(38) jadi saya dengang dia ade-kaka sunggu
so 1SG with 3SG siblings be.true
‘so I and she are full siblings’ [080927-009-CvNP.0044]

(39) 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 ‘have BASE’, speakers typically use the existential verb ada ‘exist’ rather than the prefixed form as shown in (40) with ada duri versus berduri ‘have thorns’.

BER-affixation of nominal bases: Alternative strategy for ‘have BASE’

(40) 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 exist
‘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 communicate ‘do BASE’, speakers prefer to employ alternative verbs. They tend to say, for example, biking kebung ‘make/work a garden’ as in (41), rather than use prefixed berkebung ‘do farming’ as in (42). Likewise, it is more common to say taw bahasa X ‘speak language X’ than to use prefixed berbahasa X ‘speak language X’ as in (43).

BER-affixation of nominal bases: Alternative strategy for ‘do BASE’

(41) 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]

(42) 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]
(43)  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]

3.2.5 Formally complex words with non-compositional semantics
The second observation concerning the form-function relationship between the attested BER-derivations and their nominal, numeral, or quantifier bases is that the high frequency derivations listed in Table 4 may well have non-compositional semantics.

In a conversation about religious affairs, the speaker produced *diberdoa* ‘be prayed for’. This item is ungrammatical in both Papuan Malay and Indonesian. Papuan Malay does not have a morphologically marked undergoer voice. The Indonesian undergoer voice marker *di-* cannot co-occur with prefix *ber-*, but always replaces it. This example shows that the speaker perceives *berdoa* ‘pray’ as a monomorphemic word to which she affixed the Indonesian undergoer voice marker *di-* in an attempt to dress-up her Papuan Malay.

BER-affixation: Non-compositional semantics

(44)  bebang masala de punya dia perlu ... harus *di–ber–doa*  
burden problem 3SG POSS 3SG need have.to UVVBLZ–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]

As for ‘Formally complex words with non-compositional semantics’ in more general terms, another observation is made. The present corpus includes a substantial number of formally complex words that are lacking a base, namely 16 lexemes with 93 tokens, presented in Table 5. These 16 items make up 20% of the attested total of 78 BER-prefixed lexemes.

Table 5: Formally complex words lacking a base

<table>
<thead>
<tr>
<th>Base</th>
<th>BER-Base</th>
<th>Gloss</th>
<th>BER-#</th>
</tr>
</thead>
<tbody>
<tr>
<td>*triak</td>
<td>bertriak</td>
<td>‘scream’</td>
<td>21</td>
</tr>
<tr>
<td>*kalai</td>
<td>bakalai</td>
<td>‘fight’</td>
<td>19</td>
</tr>
<tr>
<td>*juang</td>
<td>berjuang</td>
<td>‘struggle (for)’</td>
<td>17</td>
</tr>
<tr>
<td>*henti</td>
<td>berhenti</td>
<td>‘stop’</td>
<td>8</td>
</tr>
<tr>
<td>*gaul</td>
<td>bergaul</td>
<td>‘associate’</td>
<td>6</td>
</tr>
<tr>
<td>*grak</td>
<td>bergrak</td>
<td>‘move’</td>
<td>5</td>
</tr>
<tr>
<td>*labu</td>
<td>berlabu</td>
<td>‘anchor’</td>
<td>4</td>
</tr>
<tr>
<td>*saling</td>
<td>bersaling</td>
<td>‘give birth’</td>
<td>3</td>
</tr>
<tr>
<td>*tele-tele</td>
<td>bertele-tele</td>
<td>‘talk excessively’</td>
<td>2</td>
</tr>
<tr>
<td>*gumul</td>
<td>bergumul</td>
<td>‘struggle’</td>
<td>2</td>
</tr>
<tr>
<td>*sedia</td>
<td>bersedia</td>
<td>‘be prepared’</td>
<td>1</td>
</tr>
<tr>
<td>*sina</td>
<td>bersina</td>
<td>‘commit adultery’</td>
<td>1</td>
</tr>
<tr>
<td>*tengkar</td>
<td>bertengkar</td>
<td>‘quarrel’</td>
<td>1</td>
</tr>
<tr>
<td>*debar</td>
<td>berdebar</td>
<td>‘pulsate’</td>
<td>1</td>
</tr>
<tr>
<td>*lindung</td>
<td>berlindung</td>
<td>‘shelter’</td>
<td>1</td>
</tr>
<tr>
<td>*temu</td>
<td>bertemu</td>
<td>‘meet’</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2.6 Status of the affixed lexemes as part of the Papuan Malay lexicon or as code-switches with Indonesian
Most of the BER-prefixed lexemes (49/62 – 79%) were tentatively classified as borrowings from Standard Indonesian (SI-borrowings). This includes 24/29 lexemes (83%) with verbal bases and 22/30 lexemes (73%) with nominal bases and the three lexemes with numeral/quantifier bases. Examples are presented in Table 6. Also included are BER-prefixed lexemes, identified as native Papuan Malay lexemes.
Table 6: Affixation with BER- of nominal, numeral, or quantifier bases

<table>
<thead>
<tr>
<th>BER- Base</th>
<th>Gloss</th>
<th>BER- Base</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER- prefixed lexemes with verbal bases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesian code-switches</td>
<td>Papuan Malay lexemes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bekerja</td>
<td>‘work’</td>
<td>berjalan</td>
<td>‘walk’</td>
</tr>
<tr>
<td>berpikir</td>
<td>‘think’</td>
<td>berbicara</td>
<td>‘speak’</td>
</tr>
<tr>
<td>berbuat</td>
<td>‘make’</td>
<td>berbagai</td>
<td>‘mean’</td>
</tr>
<tr>
<td>berkebang</td>
<td>‘do farming’</td>
<td>bersaksi</td>
<td>‘testify’</td>
</tr>
<tr>
<td>bersyukur</td>
<td>‘give thanks’</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>berbua</td>
<td>‘have fruit’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>berduri</td>
<td>‘have thorns’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>berkebang</td>
<td>‘do farming’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bersaksi</td>
<td>‘testify’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bersyukur</td>
<td>‘give thanks’</td>
</tr>
</tbody>
</table>

3.3 Language external factors
To further investigate the degree of productivity for prefix BER- ‘VBLZ’, a domain analysis was conducted. This analysis focused on the variables of speaker education levels, topics, and relationships between interlocutors. In all, 56 items with a total of 164 tokens were examined:

- 27 ber- prefixed lexemes verbal bases (94 tokens)
- 26 ber- prefixed lexemes with nominal bases (57 tokens)
- ber- prefixed lexemes with numeral or quantifier bases (13 tokens)

Five BER- prefixed lexemes with token frequencies of more than 50 were excluded from the analysis: two derivations with verbal bases, namely blajar ‘study’ and brangkat ‘leave’, and three derivations with nominal bases, namely berdoa ‘pray’, brarti ‘mean’, and berdiri ‘stand’. 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.

In addition, the derivation berusaha ‘attempt’ was excluded due to questions concerning the reliability of the recorded tokens. Of its 25 occurrences, 11 were produced by the same speaker during a phone conversation which was characterized by many repetitions due to a bad connection.

For the 56 BER- prefixed lexemes, most tokens (148/164 – 90%) can be accounted for in terms of speaker education levels, topics, and/or role-relations. The remaining 16/164 tokens (10%), however, cannot be explained in terms of these variables of the communicative event. These tokens occurred when less-educated speakers conversed with fellow-Papuans of equally low social standing about LOW topics, that is, casual daily-life issues, as illustrated in Table 7. More specifically, almost two thirds of the tokens were produced by better-educated speakers (+EDC-SPK) (103/164 – 63%), while the less-educated speakers (-EDC-SPK) produced 61/164 tokens (37%).

---

13 All of the recorded less-educated speakers belonged to the group of Papuans with lower social status, while the recorded Papuans with higher social status, such as teachers, government officials, or pastors, were all better educated.
Table 7: Tokens for BER-prefixed words with verbal bases (27 items)

<table>
<thead>
<tr>
<th>Interlocutors/Topics</th>
<th>Speakers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+EDC-SPK</td>
<td>-EDC-SPK</td>
</tr>
<tr>
<td>Interlocutors: OUTSIDER</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Topics: HIGH</td>
<td>71</td>
<td>24</td>
</tr>
<tr>
<td>Topics: LOW</td>
<td>18</td>
<td>45</td>
</tr>
<tr>
<td>Interlocutors: +STAT</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Interlocutors: -STAT</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>103</td>
<td>61</td>
</tr>
</tbody>
</table>

More than half of the tokens (95/164 – 58%) occurred during conversations about HIGH topics, that is, political, educational or religious affairs. This includes 71/103 tokens (69%) produced by +EDC-SPK and 24/61 tokens (40%) produced by -EDC-SPK. In addition, 24/164 tokens (15%) occurred during conversations with an outsider, namely the author, including 14/103 +EDC-SPK tokens (14%) and 10/61 -EDC-SPK tokens (16%).

This leaves 45/164 tokens (27%) that were produced when the interlocutors discussed LOW topics. This includes 18/103 +EDC-SPK tokens (17%) and 27/61 -EDC-SPK tokens (44%). The 27 LOW topic tokens produced by -EDC-SPK are distributed as follows. Eleven tokens occurred during conversations with fellow-Papuans of higher social status (+STAT) (that is, 11/61 -EDC-SPK tokens – 18%). The remaining 16 tokens (that is, 16/61 tokens -EDC-SPK tokens – 26%) occurred when -EDC-SPK discussed LOW topics with fellow-Papuans of equally low social standing (-STAT), and therefore cannot be explained in terms of speaker education levels, topics, and/or role-relations.

This total of 16 tokens refers to 10% of all 164 BER-tokens, including 8/94 tokens (9%) with verbal bases and 8/70 tokens (11%) with nominal bases. Interestingly, 13 of these 16 derivations were tentatively classified as code-switches with Indonesians. This includes items such as berbeda ‘be different’, bersaudara ‘be siblings’, and behapa ‘be several’. Only three items were tentatively classified as Papuan Malay lexemes, namely berhasil ‘be successful’, bruba ‘change’, and bagaya ‘put on airs’.

4. Summary and conclusions

This paper has described how the degree of productivity of morphological patterns can be established in diglossic situations where the target language is the LOW variety which experiences interference from a closely related HIGH variety. This is the case with Papuan Malay which functions as the LOW variety in the coastal areas of West Papua, where Indonesian is used as the HIGH variety. In light of this diglossic distribution, the high degree of linguistic relatedness between both varieties, and the formal setting of a test situation, more traditional analysis approaches, such as productivity tests, were considered unsuitable.

Instead, prefix BER- ‘VBLZ’ and its derivations were investigated in terms of nine language internal and external factors. That is, given the mentioned linguistic and sociolinguistic situation of Papuan Malay it was deemed necessary to not only investigate different language internal factors. Rather, it seemed indispensable to take into account pertinent variables of the communicative event that are likely to influence the use patterns of the prefix.

Summarizing the findings for the language internal factors, four observations suggest that affixation with BER- ‘VBLZ’ is a rather productive process:

1. Polyfunctionality: BER- ‘VBLZ’ is polyfunctional in that the prefix derives lexemes from verbal, nominal, numeral, and quantifier bases.
2. Form-function relationship: For BER-derivations with nominal, numeral, and quantifier bases the form-function relationship is transparent.
3. Type and token frequencies: BER- ‘VBLZ’ is represented by a large number of derivations which, in turn, have low token frequencies.
4. Relative token frequencies: Most BER-derivations have lower frequencies than their bases.

Four other observations, however, do not support the conclusion that BER-affixation is a productive process:

1. Polyfunctionality: BER- ‘VBLZ’ is polyfunctional in that the prefix derives lexemes from verbal, nominal, numeral, and quantifier bases.
2. Form-function relationship: For BER-derivations with nominal, numeral, and quantifier bases the form-function relationship is transparent.
3. Type and token frequencies: BER- ‘VBLZ’ is represented by a large number of derivations which, in turn, have low token frequencies.
4. Relative token frequencies: Most BER-derivations have lower frequencies than their bases.
1. Form-function relationship: For the prefixed lexemes with verbal bases, the derived lexemes have the same semantics as their bases.
2. Alternative strategies: For lexemes with nominal bases, speakers prefer to use alternative analytical constructions rather than the affixed lexemes.
3. Formally complex words: Derivations may well have non-compositional semantics for Papuan Malay speakers.
4. Indonesian code-switches: Most of the lexemes with verbal or nominal bases were tentatively classified as code-switches with Indonesian.

These findings for the language internal factors suggest that the productivity of BER-affixation is rather questionable.

As for the language external factors, the findings show that 90% of the prefixed lexemes (148/164) can be accounted for in terms of variables of the communicative event, namely speaker education levels, topics, and/or role relations. Only 10% of the tokens (16/164) are unaccounted for. They occurred when less-educated speakers conversed with fellow-Papuans of equally low social standing about LOW topics. These findings for the language external factors indicate that BER-affixation is rather unproductive.

Taken together, the findings for the language internal and external factors suggest that in Papuan Malay BER-affixation is rather unproductive, at least as far as the data in the 16-hour corpus are concerned. Hence, it is concluded that lexemes such as berbuat ‘make’, as in (1), or bermalam ‘overnight’ as in (2) do not result from a productive affixation process. They are either Papuan Malay lexemes with non-compositional semantics or code-switches with Indonesian.

The multifaceted approach presented here may not be as reliable as well-established productivity tests. It does, however, offer an alternative to analyzing morphological patterns in situations where formal productivity tests are unsuitable and less likely to provide reliable data. This scenario applies especially to language continua, as for instance, in the present case, the Malay continuum. Language continua are typically comprised of a number of nonstandard languages and one (or more) standard languages, all of which are more or less closely related. Moreover, the nonstandard and standard languages are likely to be sociolinguistically intertwined with respect to their functional distribution. That is, the nonstandard languages function as the LOW varieties, while the standard language functions as the HIGH variety, with the LOW varieties experiencing interference from the HIGH variety. This is the case with Papuan Malay, which has a great deal of language contact with Indonesian, the official language in the area.

In taking into account both language internal and external factor, this paper has exemplified how the sociolinguistic setting of the target language can be taken into account in examining the productivity of morphological patterns of the target language.

This paper has dealt with but a single language. In conclusion, the approach presented here is offered as a case study for taking into account both language internal and external factors when examining the productivity of morphological patterns in contexts where more traditional approaches are less likely to provide reliable data.

### List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
<td>1st, 2nd, 3rd person</td>
</tr>
<tr>
<td>ACL</td>
<td>accidental</td>
</tr>
<tr>
<td>AG</td>
<td>agent</td>
</tr>
<tr>
<td>D,DIST</td>
<td>demonstrative, distal</td>
</tr>
<tr>
<td>D,PROX</td>
<td>demonstrative, proximal</td>
</tr>
<tr>
<td>EDC</td>
<td>education</td>
</tr>
<tr>
<td>+EDC</td>
<td>better educated</td>
</tr>
<tr>
<td>-EDC</td>
<td>less educated</td>
</tr>
<tr>
<td>L,DIST</td>
<td>locative, distal</td>
</tr>
<tr>
<td>L,MED</td>
<td>locative, medial</td>
</tr>
</tbody>
</table>

---

As for the other five, rather commonly employed affixes, the findings presented in Kluge (2014: 111–166) suggest that (a) prefix TER- ‘ACL’ and suffix -ang ‘PAT’ are somewhat productive; (b) prefix PE(N)- ‘AG’ is, at best, marginally productive; and (c) suffix -nya ‘3POSSR’, and circumfix ke-/-ang ‘NMLZ’ are unproductive.
Abbreviations

L.PROX locative, proximal
NEG negation
NMLZ nominalizer
Orthogr. orthography
oSb older sibling
pat patient
PL plural
POSS possessive
POSSR possessor
RDP reduplicant
REL relativizer
SG singular
STAT social status
+STAT higher social status
-STAT lower social status
TRU truncated
UV undergoer voice
VBLZ verbalizer
ySb younger sibling

References


