



## University of Pennsylvania Working Papers in Linguistics

---

Volume 6

Issue 2 *Selected Papers from NWAV 27*

Article 7

---

1-1-1999

# The emergence of creole subject-verb agreement

Miriam Meyerhoff

---

This paper is posted at ScholarlyCommons. <http://repository.upenn.edu/pwpl/vol6/iss2/7>

For more information, please contact [libraryrepository@pobox.upenn.edu](mailto:libraryrepository@pobox.upenn.edu).

---

# The emergence of creole subject-verb agreement

# The Emergence of Creole Subject-Verb Agreement

Miriam Meyerhoff

This paper undertakes two tasks: first, it resolves an outstanding question as to the proper structural description of Bislama; second, it will unify this analysis of the verb system with the distribution of phonetically null subjects in Bislama. In this way, it will show that resolving patterns of variation in one domain of the grammar contributes to a meaningful account of what underlies variation in another domain.<sup>1</sup>

We first turn to the data. Table 1 shows the paradigm for finite verbs in Bislama (the English-lexified creole spoken by c.180,000 people in Vanuatu, SW Pacific) in all persons and numbers with the verb *karem* 'to carry; to bring'.

|           | SINGULAR           | DUAL                    | TRIAL                    | PLURAL                  |
|-----------|--------------------|-------------------------|--------------------------|-------------------------|
| 1 (incl.) | --                 | <i>yumitu karem</i>     | <i>yumitri karem</i>     | <i>yumi karem</i>       |
| 1 (excl.) | <i>mi karem</i>    | <i>mitufala i karem</i> | <i>mitrifala i karem</i> | <i>mifala i karem</i>   |
| 2         | <i>yu karem</i>    | <i>yutufala i karem</i> | <i>yutrifala i karem</i> | <i>yufala i karem</i>   |
| 3         | <i>hem i karem</i> | <i>tufala i karem</i>   | <i>trifala i karem</i>   | <i>olgeta oli karem</i> |

Table 1: Paradigm for finite verbs in Bislama, *karem* 'carry, bring'.

Like many of the substrate languages that continue to be spoken in Vanuatu, Bislama makes four distinctions in number, and in the first person non-singular forms also distinguishes a 'we' that is inclusive of the addressee, and a 'we' that excludes the addressee. In each cell, *karem* is preceded by first the pronoun and then a morpheme most commonly realized as *i*. At two points in the paradigm the pattern of *i* marking breaks down: one is in the third person plural, where the morpheme *oli* intervenes between the verb and the pronoun. Like *i*, it is clear that *oli* is cliticized to the main verb. Example (1) shows that although there is some flexibility in the placement of adverbials, exemplified in (1b) by the adverb *bae* that indicates irrealis mood, (2)

---

<sup>1</sup> This paper revises material from chapters 5 and 6 of my 1997 dissertation. I am grateful to the Wenner-Gren Foundation for their support (grant #5742) and to Sharon Morrie Tabi in Santo. I also thank the following individuals for comments on earlier versions: Terry Crowley, Gillian Sankoff, Bill Labov, Naomi Nagy. For discussions on split *pro*-drop systems I thank David Heap and Richard Kayne. None of the above necessarily agree with the analysis presented here.

shows that it is ungrammatical when placed between the verb and *oli*. The same facts hold for *i*, and for other adverbs, e.g. those with focus functions.

- (1) a. *Ol pikinini oli krae from kek*  
The children cried out for cake. (M-94-3, Lolan)  
b. *(Bae) ol pikinini (bae) oli krae from kek.*  
The children will/would cry out for cake.
- (2) \**Ol pikinini oli (bae) krae from kek.*

The second place the paradigm in Table 1 breaks down is the first and second person singular and the first person inclusive non-singular forms (highlighted in italics). We turn now to a consideration of what the proper analysis of these cells is.

## 1 Prior Accounts of the Data

Prior accounts of Bislama have analyzed these forms in three ways. Tryon's (1987) analysis is the most economical. Referring to these morphemes in his pedagogic grammar as 'predicate markers' (a rather meaningless descriptor, which nevertheless is widely used in discussing Bislama and the related language Tok Pisin), Tryon analyzes Bislama as having a predicate marker, *oli* in 3p, a zero predicate marker in 1s, 2s and 1incl. forms, and the predicate marker *i* elsewhere.

Crowley (1990), analyzes the system as having an underlying *i* in all persons and numbers (except 3p). He suggests that the *i* is assimilated to a preceding high vowel. That is, Crowley argues that (3b) is the underlying form for the surface form (3a).

- (3) a. *Mi harem wan noes long naet*  
I heard a noise in the night  
b. *Mi i harem wan noes long naet (/mi + i/ > /mi/)*

This analysis would seem to be much influenced by the (to my knowledge, undisputed) analysis for the Tok Pisin. Tok Pisin, while closely related to Bislama, differs in some crucial ways from Bislama in the finite verb system. Tok Pisin has *i* marking in careful speech with 3p, and wide-spread deletion of *i* marking in all persons and numbers in casual speech. Neither of these properties hold for Bislama. Moreover, Crowley's analysis is somewhat at odds with some of his other data. For instance, in (4a) he records that when

adverbial modification of a 1s verb does not result in *i* surfacing, instead we find doubling of the pronoun. As (4b) shows, this results in a pattern is not analogous to pronoun doubling in other persons, rather it seems to suggest that the second *mi* in (4a) is functioning as the *i* in (4b) does.

- (4) a. *Mi bae mi no kam*  
I won't come. (Crowley 1990:235)
- b. *Hem bae i no kam/ \*Hem bae hem i no kam.*  
S/he won't come.

Finally, Guy (1974), Camden (1977) and Charpentier (1979) independently forward a radically different analysis. They note the comparative frequency of utterances with the form of (5). They also note strong structural parallels between the Bislama verb phrase and the patterns found in many of the substrate languages in Vanuatu. On the basis of such parallels, they argue, what appears (on the basis of its function in the lexifier) to be a pronoun in the cells for 1s, 2s and 1 incl. in Table 1 is no pronoun at all. It is an agreement marker, or, in Charpentier's terms "des modalités personnelles" (1979:307). Utterances in which there is a singleton *mi* are, they argue, cases of *pro*-drop leaving only an agreement marker that happens to be homophonous with the pronoun and the main verb.

- (5) *Mi mi kakae*  
1s 1s eat 'I ate/eat'

In sum, the difference between the three analyses is this: Tryon and Crowley believe *mi* and *yu* in Table 1 function only as subject pronouns in Bislama; Guy, Camden, and Charpentier (GCC) believe *mi* and *yu* function as both subject pronoun and as subject-verb agreement.

Clearly, these are fundamentally different analyses. In addition, they fail to adequately define what *i* and *oli* actually are in Bislama, resorting to terms like 'predicate marker' or 'marker of person modality'. In the next two sections, I will argue that Tryon and GCC are both right. The morphemes *i* and *oli* are bona fide subject-verb agreement, as is found in other Vanuatu languages, but the agreement in 1s and 2s is a zero morpheme.

I will make my case using quantitative data taken from a corpus of conversational Bislama recorded in northern Vanuatu in 1994-5. Forty-two speakers, ranging in age from 8 to approximately 65 years were recorded in their own homes or when stopping to visit with the researcher. In all cases, day-to-day conversations with friends, family and the researcher were re-

corded, though two of the children under 12 years were recorded telling stories. Tapes were made both in the northern township, Santo, and in a village on a nearby island. Roughly equal numbers of women and men were recorded. Segments of conversation were transcribed to generate a corpus of c. 30,000 words; all examples that follow are drawn from the corpus, unless noted otherwise. Speakers are identified with pseudonyms.

## 2 The Proper Description of the Bislama Verb Phrase

In this section, I review quantitative evidence for and against the three analyses outlined above. It should be noted that the analysis is based primarily on data on 1s and 2s forms. There were very few of the 1incl. forms in the corpus examined. Since Crowley's and Tryon's analyses differ only in the detail (in principle agreeing that *mi* and *yu* are simply pronouns), I will evaluate first the evidence for each of these positions, ultimately showing that the weight of the evidence is in favor of Tryon's proposal. Having determined which of the two pronominal analyses is the more appropriate, I will evaluate how well the pronominal and subject-verb agreement analyses of *mi* and *yu* fit the data.

### 2.1 Evaluating the "Pronoun Only" Analyses

In evaluating Tryon's and Crowley's analyses, we will consider all evidence that might indicate there is indeed an underlying *i* in 1s and 2s. We will examine the strength of the evidence in three domains that would support Crowley's analysis of *i* assimilation to a preceding high vowel.

First, let us note that it is clear Bislama allows phonetically null subjects (since this is true in both main and subordinate clauses, and there is also no overt expletive subject, we can refer to Bislama using the short-hand, calling it a *pro-drop* language). However when the subject is interpreted as 1s (or 2s) but is phonetically null, no underlying *i* morpheme resurfaces. Since Crowley made no explicit claims for the ordering of the *i* assimilation rule in Bislama, it would be possible for his analysis of *i* assimilation to hold if its application were ordered after some rule of subject deletion applied.

But Crowley's analysis also fails to square well with two other facts. Let us consider in more detail the facts alluded to in (4a) and (4b). The data shows that adverbials modifying the verb phrase (*bae*) or the subject (*wan nomo*) can intervene between a 3s subject and the complex *i* + *V*, without any change to the unmarked inventory and order of morphemes in the clauses. However, this is not the case when the subject is 1s or 2s.

This data is extremely problematic for Crowley's analysis. Not only does the adverb fail to make the underlying *i* he posits appear, but even if the assimilation rule were triggered before adverb insertion (an ordering that would be unusual in itself), an apparent copy of the pronoun is obligatory in these cases. This evidence favors the GCC analysis, which we return to shortly.

The second fact that needs to be considered is that Crowley's analysis predicts that whenever a subject ends in a high vowel in other persons and numbers, the *i* morpheme should also assimilate. In fact, an analysis of such tokens in the corpus tends not to support this. All complex NPs where the final segment is a high vowel (6-7), and all proper names ending in high vowels (8) in the corpus were examined.

- (6) a. *Afta mi pusum [hed blong mi] i go*  
So then I pushed my head through. (M-95-19, Mesek)  
b. \**Afta mi pusum [hed blong mi] go*
- (7) a. *[Woman blong yu] i kam long wea?*  
Where does your wife come from? (S-94-3, Simeon)  
b. \**[Woman blong yu] kam long wea?*
- (8) a. *(Tammy/Wili/Lili) i no save*  
T/W/L didn't realize it. (S-95-11, Juliet)  
b. #*Tammy no save*

Table 2 summarizes the quantitative results. It shows that an overt *i* almost categorically follows complex NP subjects ending in a high vowel. The data with proper names ending in high vowels is mixed, showing a slight tendency for assimilation.<sup>2</sup>

| Subjects w.<br><i>/i/ or /u/ final</i> | Form of predicate |       | Total |
|--|-------------------|-------|-------|
|  | <i>i</i> + V      | Ø + V |       |
| (10) NP + <i>blong mi</i>              | 9                 | 3     | 12    |
| (11) NP + <i>blong yu</i>              | 6                 | 0     | 6     |
| (12) Proper Name                       | 12                | 14    | 26    |
| <b>Total</b>                           | 27                | 17    | 44    |

Table 2: Presence and absence of *i* when immediately preceded by a high vowel (non-subject *mi* and *yu* and Proper Names).

<sup>2</sup> The presence or absence of *i* in these contexts was verified by Sharon Tabi, a Bislama-dominant speaker from Pentecost.

In sum, although the data does not uniformly line up in favor of one analysis over another, the bulk of the evidence suggests that if the morphemes in question are best analyzed as pronouns, then Tryon's analysis of a zero 'predicate marker' in 1s and 2s fits the data better than Crowley's analysis of an underlying *i*.

## 2.2 Evaluating the 'Subject Agreement' Analysis

Just as the competition between Crowley's and Tryon's analyses had to be resolved by turning to evidence elsewhere in the grammar (phonology-syntax interface), weighing the goodness of fit of Tryon's analysis and the GCC analysis requires examining data from another domain of the grammar (discourse-syntax interface). In this section I will compare the distribution of 3s pronouns and *i* with the disputed forms *mi* and *yu* in a range of discourse conditions.

The hypothesis is as follows: If *mi* and *yu* only function as pronouns in Bislama (Tryon's contention), then we expect to find them distributed in different discourse contexts with the same relative frequency as we find other, indubitable, pronouns such as the 3s *hem*. We should also find, *mutatis mutandis*, that the frequency of phonetically null 3s subjects (with the form  $\emptyset$  *i* *V*) is much the same as bare verbs where the interpretive subject is 1s/2s (the form  $\emptyset$  *V*). And we should also expect to find that the frequency of focused 3s subjects (with the form *NP<sub>i</sub> hem<sub>i</sub> i V*) in each discourse context is much the same as the frequency of forms like *mi mi V*.

$$\frac{mi\ mi\ V}{NP, hem\ i\ V} \dots VS \dots \frac{mi\ V}{hem\ i\ V} \dots AND \dots \frac{\emptyset\ V}{\emptyset\ i\ V}$$

Figure 1: Within group similarities expected in the distribution of 3s and 1s/2s subjects if *mi* and *yu* are pronouns.

However, if the Guy-Camden-Charpentier analysis is correct and *mi* and *yu* may be subject-verb agreement markers homophonous with the pronouns, then we expect to find their distribution across different discourse contexts to be different. Specifically, we expect that a singleton occurrence of *mi* will be distributed in a manner comparable to 3s forms where the pronoun has been dropped and only the agreement marker *i* remains. Utterances in which there is a sequence of *mi mi* + *V* should be found with comparable frequency in different discourse contexts as focused 3s subjects, i.e. *NP<sub>i</sub> hem<sub>i</sub> i V*.



$$\frac{mi\ mi\ V}{hem\ i\ V} \dots vs \dots \frac{mi\ V}{\emptyset\ i\ V}$$

Figure 2: within group similarities expected in the distribution of 3s and 1s/2s subjects if *mi* and *yu* are pronouns *and* (homophonous) subject-verb agreement.

It is important to note that the prediction is not that the relative frequency of these forms will be the same *between* groups. For example, the prediction is not that 1s/2s subjects will be phonetically null as often overall, or even as often in a specific interclausal context as 3s subjects are. The similarity that is required is a *within* groups similarity. This means that if an interclausal discourse context favors phonetically null 3s subjects more than another, it is predicted that the same contexts will be ranked in the same order relative to one another when the subject is 1s/2s.

Five interclausal relations were distinguished based on the form and the grammatical role of the referent. These were cases where:

- A the subject of the (current) clause had also been the subject<sup>3</sup> of the preceding clause;
- P the subject of the (current) clause was some other argument in the preceding clause; and
- N the subject of the (current) clause was not present in the preceding clause.

where 'clause' was defined as a finite verb, whether main or subordinate. Non-finite clauses (including imperatives) and the second verb in serial verbs expressing motion or location, both of which never have an overt subject, were excluded from the data. In the first two conditions, the antecedent referent was further discriminated on the basis of its form, overtly realized vs. phonetically null (giving *A-o* 'overt subject', *A* 'phonetically null subject', *P-o* 'overt other argument', and *P* 'phonetically null other argument'). There is no *a priori* reason to suspect that further constraints on the focusing of subjects apply when the referent is the speaker or addressee but do not apply when the subject is a third party, thus the experiment seems valid.

An analysis of this data using Goldvarb 2.1 (Sankoff et al. 1992) was conducted. Guy (1980) has shown that use of logistic regressions must be

---

<sup>3</sup> Bislama has no passive, so this condition generally reflects a continuity in the thematic role between clauses as well.

cautious when the data is highly skewed or very patchy, i.e. when the number of tokens in some cells falls below 30. In this data, the problem cells are clustered in the *P* and *P-o* conditions. A multivariate analysis of the data leaving out the data in these conditions was conducted but in each case this resulted in a model that was a significantly worse fit to the data than was obtained when the data from these conditions was included. Consequently, these factors are retained for the within groups comparison.

The results of the logistic regressions are represented graphically in Figures 3-7. The frequency of the different forms of subject in different discourse conditions is converted into probability weightings in the tables,

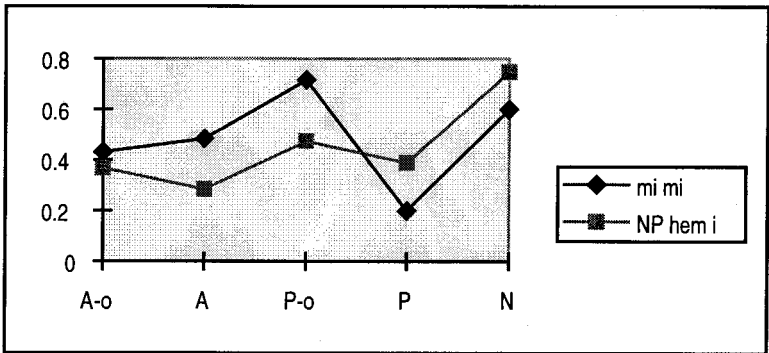


Figure 3: Weighting of *mi mi* and *NP, hem i* subjects compared (Analysis B predicts isomorphism). LINEST = 0.4908.

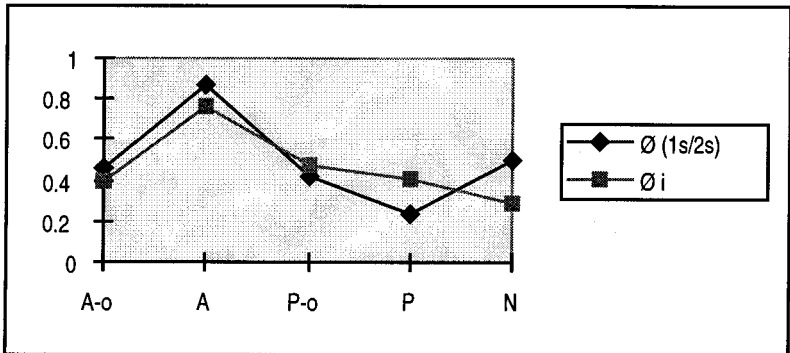


Figure 4: Weighting of *mi* and *hem i* subjects compared (Analysis B predicts isomorphism). LINEST = 0.4908.

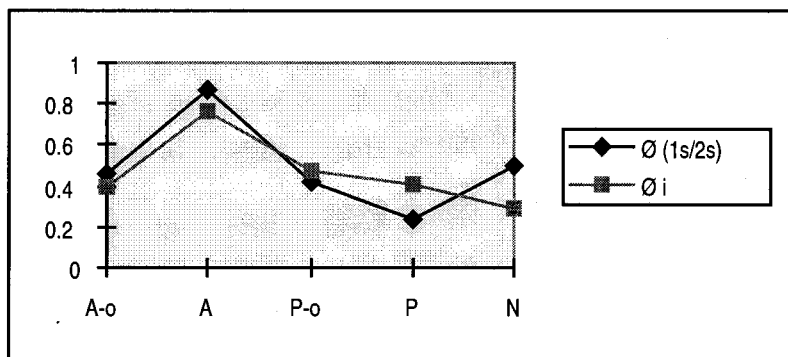


Figure 5: Weighting of  $\emptyset$  (1s/2s) and  $\emptyset i$  (3s) subjects compared (Analysis B predicts isomorphism). LINEST = 0.9711.

expressing the degree to which each condition favors or disfavors the different forms of subject. Figures 3-5 test the hypothesis that *mi* and *yu* are pronouns only; Figures 6-7 test the hypothesis that *mi* and *yu* are both pronouns and subject-verb agreement markers.

The degree to which there is comparable within group behavior of the different subject forms has been evaluated using the LINEST (linear estimate) function. This is a measure of how much the two paths being compared behave alike or diverge from each other (calculating closeness on the basis of the square difference between the two lines at each point of comparison). A value of positive one indicates that the two forms being compared are always moving the same way; a value of minus one indicates that they consistently move in different ways. Figures 3-5 show LINEST values that are consistently positive, and in the case of the hypothesized null subject condition (Figure 5) behavior of 3s and 1s/2s subjects across the different discourse contexts hardly differs at all. The modest difference in the way subjects with the form *mi mi V* pattern compared to the 3s subjects *NP, hem i V* suggests that we were correct to proceed assuming that there are no major independent constraints on how and when it is appropriate to focus a subject referring to the speaker or addressee compared to when it is appropriate to focus a 3s referent.

This is in marked contrast to the picture painted when testing the GCC analysis, as shown in Figures 6-7.

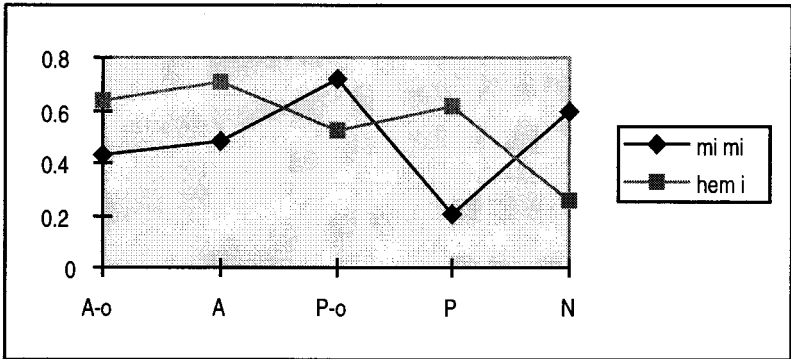


Figure 6: Weighting of *mi mi* and *hem i* subjects compared (Analysis C predicts isomorphism). LINEST = -0.491.

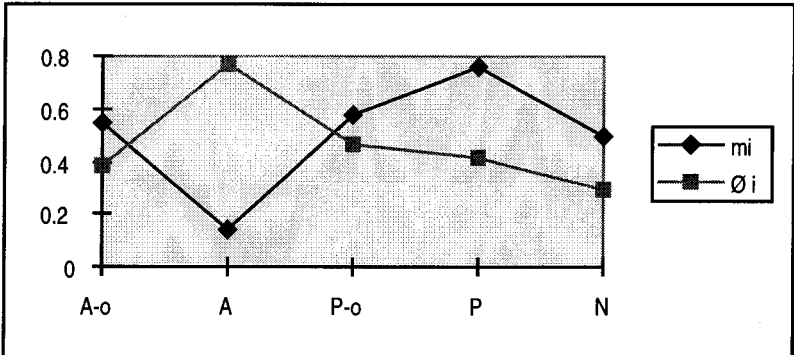


Figure 7: Weighting of *mi* and  $\emptyset i$  (3s) subjects compared (Analysis C predicts isomorphism). LINEST = -0.971.

Figure 6 shows that 1s/2s subjects with the form *mi mi* and *yu yu* are not distributed across the five interclausal environments in a manner comparable to the distribution of focused 3s subjects, while Figure 7 shows even more clearly that subjects with the form *mi* or *yu* alone do not pattern like 3s subjects that are phonetically null (leaving only the agreement marker *i*). We must conclude, then, that the singleton forms *mi* and *yu* are not subject agreement markers on the verb as GCC argue.

To sum up, then, on the basis of the distribution of different subject forms across a range of phonological and discourse environments, we conclude that *mi* and *yu* are not subject-verb agreement markers, and that the agreement marker analogous to 3s and 3p *oli* in these forms is a zero morpheme.

### 3 The Development of a Grammar Licensing *Pro*-drop

This claim will be supported by looking more closely at the patterns of phonetically null and overt pronominal subjects found in conversational Bislama. It will be argued that the distribution of these forms is best explained in terms of the morphosyntactic informativeness of the inflectional system outlined in section one. In doing so, we will see that the account of 1s/2s subject-verb agreement forwarded in this paper has more than mere descriptive adequacy. It provides the basis for a unified account of the distribution of phonetically null subjects as well.

There is a marked difference in the frequency of overt/null 1s/2s subjects and the frequency of overt/null 3s subjects. The relevant Ns are shown in Table 3:

| sbj form       | <i>mi &amp; yu</i> | <i>hem</i> |
|----------------|--------------------|------------|
| $\emptyset$    | 118                | 1010       |
| <i>pronoun</i> | 1364               | 664        |

Table 3: Frequency of pronoun and phonetically null subjects in Bislama (1s & 2s compared with 3s).

Table 4 shows that this distinction between the persons generalizes across singular and plural. Table 4 shows the probability with which subjects in each person and number occur with a phonetically null subject (complete co-

| Person and number of subject referent | Goldvarb weighting | N clauses |
|---------------------------------------|--------------------|-----------|
| 3s                                    | 0.709              | 1719      |
| 3p                                    | 0.869              | 846       |
| 1s                                    | 0.147              | 1054      |
| 1p                                    | 0.268              | 375       |
| 2s                                    | 0.206              | 397       |
| 2p                                    | 0.197              | 71        |

Table 4: Goldvarb weightings for all speakers for person and number of subject referent (0 = consistently pronoun subject; 1 = consistently null subject). Input probability for all speakers = 0.418

occurrence = 1) or a pronominal subject (complete co-occurrence = 0). The probabilities are shown as weightings derived by means of logistic regressions, again calculated using Goldvarb 2.1.

The results show that 3s and 3p subjects pattern alike, favoring a phonetically null subject, and that first and second person subjects singular and plural pattern alike, favoring the use of an overt pronoun in subject position. Unfortunately, the historical record on the use and distribution of phonetically null subjects in Bislama is sketchy at best. Charpentier notes "In older pidgin [Bislama], the verb was always preceded by a noun or pronominal subject... Nowadays, the personal pronoun tends to be left out, leaving utterances that consist only of subject-verb agreement and the verb stem" (1979: 353).<sup>4</sup> This suggests that the possibility of omitting pronominal subjects is a grammatical innovation that emerged during the 1960 or 1970s, following the crystallization of the pronoun and subject-verb agreement system in Table 1 and during the period in which growth of the urban centers, especially Port Vila, in Vanuatu created a comparatively large pool of first language speakers of only Bislama.<sup>5</sup> Since these are precisely the conditions under which we would expect to see the effects of UG become stronger than effects of the substrate, the development of what appears to be a grammar of *pro*-drop is especially noteworthy. Moreover, the fact that Bislama appears to have grammaticized a split *pro*-drop system is of even greater significance.

In the pidgin stages of Bislama (mid-19th century), there was no agreement between subject and verb in any person and only a simple pronominal system (*mi*, *ae* '1s', *yu* '2s' and *i* '3s/p') was used. Crowley's data strongly suggests that by the turn of the century, the system had been reanalyzed so that 3s and 3p pronouns were distinct (*hem* '3s' and *olgeta(fala)* '3p'), with *i* being reassigned as an agreement marker (1990: 231, 243). Since Crowley's data shows that even while *hem* was crystallizing as the 3s pronoun, *i* continued to occur in some clauses in what appears to be subject position (i.e. *hem* is absent). Whether these should be interpreted as holdovers from an earlier stage in the grammar (speakers using *i* as a pronoun), or whether they indicate the first examples of 3s null subjects cannot be determined. The fact of the variation is, however, vital.

---

<sup>4</sup> "Dans le pidgin ancien, le verbe était toujours précédé d'un nom ou d'un pronom sujet... Aujourd'hui, le pronom personnel tend à être abandonné, la modalité personnelle et le radical verbal composant seuls un énoncé" (Charpentier 1979: 353).

<sup>5</sup> For example, the oldest speaker I recorded in Santo/Malo who spoke Bislama as a sole first language was born in 1972 in Port Vila.

Let us assume that the variation at this crucial stage created the possibility for speakers to interpret *i* either as a pronoun or as an agreement marker. This is not implausible given the immediate history of the forms (English *he* and pidgin *i* as subject pronouns) and the substrate models available (Oceanic languages with agreement prefixes on the verb). In other words, convergence between the lexifier and the substrate results in creole variation. This necessitates further analysis and generalization on the part of the language users. As a solid pattern for the third person agreement begins to take shape, a set of agreement morphemes emerge for first and second person as well. Maintaining the singular/plural contrast marked overtly in the third person, first and second distinguish singular and plural—respectively,  $\emptyset$ , and *i*. In short, we find that the data requires a synthetic account of creole variation and creole grammatical development like those proposed by Sankoff (1984) and Crowley (1990).

We lack the amount of quantitative data from the 20th century that would be needed to conclusively support this account, nevertheless some of the 1990s data suggests that this is the path along which Bislama continues to develop. The analysis of the entire conversational corpus failed to reveal any significant apparent time effects for this variable. However, a closer examination of a subset of the corpus showed that there does indeed seem to be some change taking place with respect to the variable of null subjects.

When the comparison was restricted to the children in the corpus and their immediate caregivers and their uncles and aunts, a very strong effect for age was found.<sup>6</sup> The children appear to be generalizing the patterns in the wider speech community, approaching a categorical treatment of first and second person subjects in particular.

In short, the data from the children and their caregivers suggests that the synchronic variation found in each person in Bislama is indicative of a diachronic change in progress, during which speakers are moving towards a less variable system, one that would be compatible with some of the structural analyses of split *pro*-drop systems observed in other language families.

One final question needs to be answered as to the probable outcome of this ongoing reanalysis. It is unlikely that the outcome of this will be a time in which speakers of Bislama categorically use null anaphors in the third person and pronouns in the first and second. We see a degree of variability remains even in the first and second person for the children, and we noted earlier that intersentential variables (referent's form and grammatical role) in the

---

<sup>6</sup> Results of a regression analysis comparing children 12 and under with caregivers and young aunts and uncles,  $r^2(\text{adjusted}) = 71.6\%$ ,  $p < 0.001$ .

preceding clause play a part in constraining the distribution of null subjects in Bislama. An important consequence of their significance is that they remind us that speakers can bring numerous skills to bear in identifying the referent when the subject is phonetically null. While verbal inflections may be a preferred mode of identification, the fact that null subjects with first and second person referents are interpretable in Bislama (and more generally in languages with no verbal morphology to rely on) shows clearly that identification is by no means a unitary notion. The interaction between linguistic and subjective (or affective) variables will continue to be a factor, just as they are for prototypical null subject languages such as standard Italian. What I have shown here, however, is that they constrain a relatively small part of the variation observed in the form speakers use to express a subject in Bislama.

#### 4 Conclusion

At the outset, the problem faced in this paper was a descriptive one: which of three accounts of Bislama finite clauses best fits the data? Evidence from phonology, syntax and discourse was considered and it was concluded that the preverbal clitics in Bislama are subject-verb agreement, and that they have three variants, the overt forms *i* and *oli* and a null variant (in 1s and 2s).

Next, the distribution of null subjects in Bislama was examined and it was seen that variation emerges in the grammar at precisely the point where the verbal morphology is maximally informative. As generative accounts of the null subject parameter would predict, null subjects are emerging as the norm in the third person but not the norm in first and second person in Bislama. This makes sense since third person subject-verb agreement is maximally informative in the third person, i.e., it satisfies the requirement that a null subject be identified.

#### References

- Camden, Pastor Bill 1977. *A Descriptive Dictionary: Bislama to English*. Maropa Bookshop: Vila.
- Charpentier, Jean-Michel 1979. *Le Pidgin Bislama(n) et le multilinguisme aux Nouvelles-Hébrides*. *Language et Civilisations à Tradition Orale*, 35. Paris: Société d'Etudes Linguistiques et Anthropologiques de France [SELAF].
- Crowley, Terry 1990. *Beach-la-Mar to Bislama: The emergence of a national language in Vanuatu*. Oxford: Oxford University Press.



- Guy, Gregory R. 1980. Variation in the group and the individual: the case of final stop deletion. In William Labov (ed.) *Locating Language in Time and Space*. New York: Academic Press. 1-36.
- Guy, J.B.M. 1974. *Handbook of Bichelamar/Manuel de Bichelamar*. Pacific Linguistics, C-34. Canberra: The Australian National University.
- Sankoff, David, Pascale Rousseau, Don Hindle and Susan Pintzuk 1992. Goldvarb version 2.1. Adapted for the Macintosh by David Rand. Shareware, Université de Montréal.
- Sankoff, Gillian 1984. Substrate and universals in the Tok Pisin verb phrase. In Deborah Schiffrin (ed.) *Meaning, Form, and Use in Context: Linguistic applications*. Washington DC: Georgetown University Press. 104-119.

Department of Linguistics  
University of Hawai'i at Manoa  
1890 East-West Rd  
Honolulu, HI 96822  
[mhoff@hawaii.edu](mailto:mhoff@hawaii.edu)