

The emergence of creole subject–verb agreement and the licensing of null subjects

MIRIAM MEYERHOFF
University of Edinburgh

ABSTRACT

A corpus of conversational Bislama (a Melanesian creole spoken in Vanuatu, related to Tok Pisin and Solomon Islands Pijin) suggests that during the 20th century the creole has developed a set of regular inflectional morphemes on the verb that agree in person and number with the subject of the finite clause. It is shown that, where the agreement paradigm is referentially richest, the language is also beginning to grammaticize a tendency towards phonetically null subjects (pro-drop). Three possible analyses of the Bislama verb phrase are evaluated; consistent support for only one is found in the spoken Bislama corpus. The resulting paradigm of subject–verb agreement (*i*, *oli*, and \emptyset) is analyzed in terms of the historical development of Bislama. It is argued that the synchronic agreement marking reflects properties derived from (i) the lexifier (English), (ii) the substrate languages, and (iii) universal grammar. No one component fully accounts for the patterns of agreement marking observed. Instead, a synthesis of all three is required, as previously observed by, for example, G. Sankoff (1984) and Mufwene (1996). Substrate languages provide a model for subject agreement prefixing on the verb; the person features associated with the lexifier ‘he’ continue to be reflected in the distribution of Bislama *i*; and phonetically null subjects are emerging as the norm where the agreement paradigm best serves to identify the subject referent. This is consonant with generative accounts of null subject systems. Parallels with other languages (e.g., Italian, Franco-Provençal, Hebrew, Finnish) are examined.

This article undertakes three tasks: first, it attempts to resolve an outstanding question as to the most appropriate structural description of the relationship between subject and verb in Bislama; second, it discusses what the implications of this analysis might be for a creole ontogeny; and third, it attempts to unify this analysis of the verb system with the distribution of phonetically null subjects in Bislama. In this way, it shows that resolving patterns of variation in one domain of the grammar contributes to a meaningful account of what underlies variation in

My very great thanks to the Wenner-Gren Foundation for their support in the field (grant #5742) and to Sharon Morrie Tabi in Santo. I am also grateful to William Labov and two anonymous LVC reviewers for thoughtful commentary and constructive comments. The following individuals commented on earlier versions of this work: Terry Crowley, Michel DeGraff, David Heap, Richard Kayne, Naomi Nagy, Gillian Sankoff, and Jeff Siegel. None of them necessarily agree with my analysis.

Most of this article was written during a postdoctoral fellowship in the Department of Linguistics at Cornell University. The department provided me with a stimulating, challenging, and supportive research environment, and I extend my warmest thanks to everyone there.

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

	Singular	Dual	Trial	Plural
1 (incl.)	_____	<i>yumitu</i> karem	<i>yumitri</i> karem	<i>yumi</i> karem
1 (excl.)	<i>mi</i> karem	mitufala i karem	mitrifala i karem	mifala i karem
2	<i>yu</i> karem	yutufala i karem	yutrifala i karem	yufala i karem
3	hem i karem	tufala i karem	trifala i karem	olgeta oli karem

another domain. The article proposes an account of Bislama subject–verb agreement that synthesizes properties of the substrate, of the lexifier, and of language universals, in the spirit of, for example, G. Sankoff (1984) and Mufwene (1996).

We first turn to the data. Table 1 shows the paradigm for finite verbs in Bislama (the English-lexified creole¹ spoken by approx. 180,000 people in Vanuatu, SW Pacific) in all persons and number with the verb *karem* ‘to carry, to bring’. Points of the paradigm that are of special interest appear in italics.

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 includes the addressee and a ‘we’ that excludes the addressee. In most cells, *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. It is clear that *oli*, like *i*, is cliticized to the main verb. As shown in (1), there is some flexibility in the placement of adverbials; (1b) and (1c) illustrate this with the adverb *bae* indicating irrealis for utterances with full NP and pronoun subjects. But as (2) illustrates, it is ungrammatical for an adverbial to be 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 (M-94-3, Lolan)²
 PL child *oli* cry because cake
 ‘The children cried out for cake.’
 b. Bae ol pikinini oli krae from kek.
 IRR PL child *oli* cry because cake
 ‘The children will/would cry out for cake.’
 b’. Ol pikinini bae oli krae from kek.
 c. Bae olgeta oli krae from kek.
 they
 c’. Olgeta bae oli krae from kek.
- (2) *Ol pikinini oli bae krae from kek.

The second place the paradigm in Table 1 shows irregularity is the first and second person singular and the first person inclusive non-singular forms. We now turn to a consideration of what an appropriate analysis of these cells might be.

PRIOR ACCOUNTS OF THE DATA

Three possible analyses present themselves. The first, and most economical, is Tryon's (1987) analysis. Referring to these morphemes in his pedagogic grammar as "predicate markers," Tryon analyzed Bislama as having a predicate marker with the form *oli* in the third person plural, zero in the first and second person singular and in the first person inclusive, and *i* elsewhere (1987:18).

The second is the analysis put forward independently by Guy (1974), Camden (1977), and Charpentier (1979). They noted the comparative frequency of utterances with the form of (3), as well as strong structural parallels between the Bislama verb phrase and the patterns found in many of the substrate languages in Vanuatu. In Tangoa, shown in (4), the order of morphemes is pronoun, subject-verb agreement, and then main verb. On the basis of such parallels, they argued that what appears to be a pronoun (based on its function in the lexifier) in the cells in Table 1 for first and second person singular and first person inclusive is no pronoun at all. It is an agreement marker (in Charpentier's terms, "des modalités personnelles"; 1979:307); it has also been described in the literature as a resumptive pronoun (Baker & Mühlhäusler, 1996; Keesing, 1988). Consequently, the analysis of (3) is, as shown in (5a), directly analogous to the third singular form in (5b). Utterances with a single *mi*, these authors argued, are cases of pro-drop leaving only an agreement marker, which happens to be homophonous with the pronoun, and the main verb.

- (3) Mi mi kakae
1s 1s eat

- (4) ena na eri xaxau
1s.pronoun 1s.pred.mkr capability.aspect.mkr walk
'I am able to walk.' (Camden, 1979:58)³

- (5) a. mi mi kakae
 /je/mod.pers./manager/
 'je mange'
- b. em i go
 /il/mod.pers./aller
 'il va' (Charpentier, 1979:307)

Additional support for this analysis comes from an intriguing asymmetry in the distribution of adverbs. Consider the contrast between (6a) and (6b). The adverb *bae* may intervene between a third person singular pronoun and the verb in a straightforward way, as in (6a). But when the subject is first person singular, the pronoun must be doubled after the adverb *bae*, as in (6b). As (6c) shows, similar doubling with a third person singular pronoun is not normal.⁴ To some analysts, this has suggested that the second *mi* in (6b) fills the same slot that *i* does in (6a).

- (6) a. Hem bae i kam.
 3s IRR *i* come.
 'S/he will come.'
- b. Mi bae mi kam
 1s IRR 1s come
 'I will come.'
- c. *Hem bae hem i kam.
 3s IRR 3s *i* come

Crowley's (1990:230–252) historical grammar of Bislama described a range of patterns for subjects and verbs in all persons and numbers. He recorded variation between zero and *mi* and *yu* for the predicate marker when the subject is first or second person singular and between zero and *i* when it is *yumi(tu)* (1990:234–235).

For completeness, it should be noted that a third possible analysis exists. It has been argued for Tok Pisin (closely related to Bislama) that there is an underlying *i* in all persons and numbers which assimilates immediately following a high vowel. This analysis deserves mention, as the Tok Pisin facts are better known than the Bislama facts, and it is not uncommon (I have found) for people presented with the Bislama data to attempt such a unified analysis.⁵ Such an analysis for Bislama would suggest that the expected *i* would fail to surface before the verb following *mi*, *yu*, *yumi*, and *yumitu* because it has coalesced phonetically with the final vowel of the subject pronoun. That is, (7a) would have the underlying form of (7b).

- (7) a. Mi fas smol long ples ya.
 1s fast small in place SPEC
 'I got a bit stuck there.'
- b. Mi i fas smol long ples ya. (/mi + i/ > /mi/)

But Tok Pisin differs in crucial ways from Bislama in the finite verb system. First, Tok Pisin has *i* marking in both third person singular and plural. Second, when first or second singular forms are followed by an adverb, *i* surfaces in Tok Pisin (Verhaar, 1995:70), which makes it plausible to analyze Tok Pisin as having *i* underlyingly in these persons even when it is not realized phonetically. Finally, there is widespread deletion of *i* marking in all persons and numbers in Tok Pisin narratives (G. Sankoff, 1994). None of these are characteristics of Bislama, which suggests that the two systems are underlyingly distinct in this respect. There are also language-internal reasons, to be presented later on, for doubting an analysis of an underlying *i* in all persons and numbers in Bislama.

In sum, the difference between the analyses is profound. According to Tryon, *mi* and *yu* in Table 1 function only as subject pronouns in Bislama, and the analogy with Tok Pisin would also assume this. But, according to the analysis of Guy/Camden/Charpentier, *mi* and *yu* function as both subject pronoun and as subject-verb agreement. We would like to have a principled basis for favoring one analysis over the other. This article presents evidence that is consistent with an analysis of *i* and *oli* as subject-verb agreement, like the subject-verb agreement found in

other Vanuatu languages. The evidence also indicates that the agreement in the first and second person singular is a zero morpheme.

DATA SOURCES

The case will be made using quantitative data taken from a corpus of conversational Bislama recorded in northern Vanuatu in 1994–1995. A total of 42 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 recorded, though two of the children under 12 years were recorded only telling stories. Tapes were made in the northern township Santo and in a village on Malo, a nearby island. Roughly equal numbers of women and men were recorded. Segments of conversation were transcribed and tagged syntactically, yielding a corpus of about 30,000 words; all examples that follow are drawn from the corpus, unless noted otherwise. The references in the examples identify place of recording, year, tape number, and speaker’s pseudonym.

DESCRIBING THE BISLAMA VERB PHRASE

In this section, I review quantitative evidence bearing on the three analyses outlined earlier. It should be noted that the analysis from this point on is primarily based on data for first and second singular forms. There were very few of the first person inclusive forms in the corpus examined (61 tokens).

Tryon’s analysis and the analogy with Tok Pisin differ only in detail—in both, *mi* and *yu* are considered pronouns and never agreement markers—so I first evaluate the relative merits of these positions. Ultimately, the weight of the evidence favors Tryon’s proposal. Having determined which of the two pronominal analyses is the more appropriate, I then evaluate the third analysis (*mi* and *yu* as both pronoun and agreement morpheme) against Tryon’s.

Evaluating the pronoun-only analyses

If there were an underlying *i* present with first and second singular subjects, there are three contexts in which it might be expected to surface. These three contexts are now examined in turn.

First, let us note that Bislama clearly allows phonetically null subjects. Since this is true in both main and subordinate clauses and since there is also no overt expletive subject, we can refer to Bislama as a pro-drop language. Examples (8) and (9) show null subjects when the subject is interpreted as third and first singular, respectively; in fact, null forms are allowed with all persons and numbers.

- (8) Denis hem i kam, Ø **i blokem** hem (S-95-7, Sevi)
 3s *i* come, Ø *i* block 3s
 ‘Denis came [and he] stopped her.’

- (9) [Yu talem se wanem taem we yu kambak bakegen?]
 O, Ø **talem** se tangkiu tumas long hadwok blong hem (S-94-1, Timoti)
 O, say COMP thank.you too.much for labor of 3s
 [What did you say when you got back?]
 ‘Oh, [I] said thank you very much for all her hard work.’

Example (9) shows that, when the subject is interpreted as first (or second) singular but is phonetically null, no underlying *i* morpheme surfaces. This alone is not conclusive evidence against *i* assimilation; it simply requires that any putative *i* assimilation must happen immediately whenever the morpheme finds itself adjacent to a high vowel: that is, the process operates as in (10a), not (10b).

- (10) a. mi + i + karem
 (mi + i) + karem assimilation
 Ø karem pro-drop
 Ø karem
 ‘[I] carried it.’
- b. mi + i + karem
 Ø + i + karem pro-drop
 — assimilation
 *Ø i karem

However, this analysis fails to square well with two other facts. First, the data in (11) show that adverbials modifying the verb phrase (*bae*) or the subject (*wan nomo*) can intervene between a third person singular subject and the complex *i* + V, without any change to the unmarked inventory and order of morphemes in the clauses. However, example (12) shows that this is not the case when the subject is first or second person singular.

- (11) a. Hem bae i kam slip long hem? (M-95-19, Elise)
 3s IRR *i* come sleep on 3s
 ‘Will she come nest on it?’
- b. Be hem wan nomo i karem nao. (M-94-2, Lolan)
 but 3s one only *i* carry now.
 ‘And he carried [it] all by himself.’
- (12) a. Wan di, bae *Ø/mi mekem long hem (S-95-13, Sikal)
 one day IRR 1s make to 3s
 ‘One day, I’ll do [that] to her.’
- b. Mi wan *Ø/mi stap. (S-94-1, Naomi)
 1s one 1s stop.
 ‘I was here on my own.’

Again, no underlying *i* surfaces. Instead, as (12) shows, an apparent copy of the pronoun is obligatory in these cases. As we saw in (6), this favors the Guy/Camden/Charpentier analysis, which we return to momentarily.

Second, if there were an underlying *i* in all persons and numbers in Bislama and if this *i* assimilated with a preceding high vowel, then we would expect the *i* morpheme to assimilate with any preceding subject with a final high vowel. An analysis of such tokens in the corpus does not provide robust support for this. All

TABLE 2. *Presence and absence of i when immediately preceded by a high vowel in complex NPs, simple Ns, and proper names*

Subjects with Final /i/ or /u/	Form of Predicate		Total
	<i>i</i> + V	∅ + V	
NP + <i>blong mi</i>	9	3	12
NP + <i>blong yu</i>	6	0	6
N (final /i/)	12	0	12
N (final /u/)	7	0	7
Proper name	12	14	26
Total	46	17	63

complex NPs where the final segment is the high vowel of a pronoun (13–14) and all simplex nouns (15) and proper names ending in high vowels (16) in the corpus were examined.

- (13) a. Afta mi pusum [hed blong mi] i go. (M-95-19, Mesek)
 after 1s push head of 1s *i* go
 ‘So then I pushed my head through.’
 b. *Afta mi pusum [hed blong mi] go
- (14) a. [Woman blong yu] i kam long wea? (S-94-3, Simeon)
 woman of 2s *i* come from where
 ‘Where does your wife come from?’
 b. *[Woman blong yu] kam long wea?
- (15) a. Tu i stap stanap olsem. (M-95-, Saemon)
 two *i* CONT stand.up like
 ‘Two [other people] were standing up like [this].’
 b. Nambatri i kam bihaen. (M-95-19, Alis)
 number.three *i* come behind
 ‘The third one follows.’
- (16) a. (Tammy/Wili/Lili) i no save. (S-95-11, Juliet)
i NEG know
 ‘Tammy/Wili/Lili didn’t realize it.’
 b. #Tammy no save

Table 2 summarizes the quantitative results. It shows that an overt *i* almost always occurs following complex NP subjects ending in a high vowel and likewise with simplex nouns. The data with proper names ending in high vowels is mixed, showing a slight preference for assimilation.⁶

In sum, the bulk of the evidence in examples (11) through (16) suggests that Tryon’s analysis of a zero predicate marker in the first and second person singular fits the data better than does an analysis of an underlying *i*. In addition, an anonymous reviewer pointed out that the data in (11) suggests that *i* is a clitic, and that

the direction of cliticization is rightwards onto the verb. If *i* has this property, then assimilation with a (leftward) subject would be blocked, and we would find *i* occurring with first and second singular subjects if it really were there. Since it does not, this assumption strengthens the analysis that it is not present in these persons and numbers.

In the next section Tryon's analysis of *mi* and *yu* as pronouns is compared with the Guy/Camden/Charpentier account, under which *mi* and *yu* have distinct (but homophonous) functions as pronoun and agreement morpheme.

Evaluating the subject agreement analysis

In order to weigh the goodness of fit of Tryon's analysis against the Guy/Camden/Charpentier analysis, data from another domain of the grammar (discourse–syntax interface) are required. In this section, the distribution of third person singular pronouns and *i* in a range of discourse conditions is compared with the distribution of *mi* and *yu* in those same conditions.

The hypothesis is as follows. If *mi* and *yu* only function as pronouns in Bislama (Tryon's contention), then we would expect to find them distributed in different discourse contexts with the same relative frequency as other, indubitable pronouns, such as the third person singular *hem*. We should also find, mutatis mutandis, that the frequency of phonetically null third person singular subjects (with the form $\emptyset i V$) is much the same as bare verbs, where the referent that is interpreted as the subject is first or second person singular (with form $\emptyset V$). And we should also expect to find that the frequency of focused third person singular subjects (with the form $NP_i, hem_i i V$) in each discourse context is much the same as the frequency of forms like *mi mi V*. This hypothesis may be illustrated as follows (forms expected to be distributed similarly are in boldface).

$$\frac{\mathbf{mi mi V}}{NP, \mathbf{hem i V}} \text{ vs. } \frac{\mathbf{mi V}}{\mathbf{hem i V}} \text{ and } \frac{\emptyset V}{\emptyset i V}$$

However, if the Guy/Camden/Charpentier analysis is correct, and if *mi* and *yu* may be subject–verb agreement markers homophonous with the pronouns, then we would expect their distribution across different discourse contexts to be different. Specifically, we would expect that a singleton occurrence of *mi* would be distributed in a manner comparable to third person singular forms in which the pronoun has been dropped and only the agreement marker *i* remains. Utterances with a sequence of *mi mi + V* should be found with comparable frequency in different discourse contexts as focused third person singular subjects (i.e., $NP_i, hem_i i V$). This hypothesis may be illustrated as follows (forms expected to be distributed similarly are in boldface).

$$\frac{\mathbf{mi mi V}}{\mathbf{hem i V}} \text{ vs. } \frac{\mathbf{mi V}}{\emptyset i V}$$

It is important to note that it is not predicted that the relative frequency of these forms would be the same between groups (above and below the line). That is, it

is not predicted that first and second singular subjects would be phonetically null as often overall or even as often in a specific interclausal context as third singular subjects are. The similarity that is required is similarity within groups. This means that, if a particular interclausal discourse context favors phonetically null third singular subjects more than another, it is predicted that the same context would be ranked in the same order relative to the others when the subject is first or second singular.

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 subject⁷ 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.

“Clause” was defined as containing a finite verb, whether main or subordinate. Nonfinite clauses (including imperatives) and the second verb in serial verbs expressing motion or location, neither of which ever 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, whether overtly realized or phonetically null, giving *A-o* “overt subject,” *A* “phonetically null subject,” *P-o* “overt other argument,” and *P* “phonetically null other argument.”

Other research into the distribution of subject anaphors has found that the degree to which the subject is salient or given information sometimes acts as a constraint on the form that speakers select. Distinctions between clause chains where the subject referent is the same across clauses and those where it switches have been found to be important constraints on the distribution of subjects in varieties of Spanish (Cameron, 1993), and the position of intra-clausal salience has been found to be a salient constraint on the distribution of anaphora in Centering Theory (Grosz, Joshi & Weinstein, 1995). Both these considerations motivated the decision to distinguish between the *A* and *A-o* conditions (same subject chain; prior referent in a position of greater intraclausal salience) and the *P* and *P-o* conditions (switch-subject chains; prior referent in a position of lesser intraclausal salience).

In theory, there might be intersecting constraints on information structure that mean that focusing applies differently to subjects of different persons and numbers. For instance, first and second person are generally hearer- and discourse-old information, but third person may be new information. It is possible, therefore, that for pragmatic reasons speakers might want to focus third person subjects more often than first and second person subjects. If this were the case, it would invalidate the comparisons I have proposed, since I am assuming that focused forms and pronouns are distributed in a roughly equivalent manner for all persons

TABLE 3. *Number of tokens for each subject form in the corpus in five relations with identical referents in the preceding clause*

Subject Form	Discourse Factors					Total
	A-o	A	P-o	P	N	
<i>mi mi/you</i>	62	8*	20*	1*	73	164
<i>mi/you</i>	725	69	95	37	438	1,364
∅ (first/second singular)	45	35	4*	1*	33	118
NP, <i>hem i</i>	20*	4*	8*	2*	62	96
<i>hem i</i>	302	75	72	28*	187	664
∅ <i>i</i>	301	432	102	33	142	1,010
Total	1,455	623	301	102	935	3,416

Note: Asterisks indicate cells with fewer than 30 tokens.

TABLE 4. *Weighting of mi mi and NP, hem i subjects compared (Tryon's analysis predicts isomorphism), Pearson correlation: .449*

	A-o	A	P-o	P	N
<i>mi mi</i>	0.428	0.49	0.72	0.2	0.6
NP, <i>hem i</i>	0.366	0.29	0.48	0.39	0.75

and numbers. But as the data show, the distribution of focused subjects, pronouns, and null subjects is a unitary process across persons in Bislama.

Table 3 summarizes the frequency with which the different subject forms were found in the five interclausal conditions. An analysis of this data using VARBRUL (D. Sankoff, Rousseau, Hindle, & Pintzuk, 1992) was conducted. Guy (1980) showed that logistic regressions must be used cautiously when the number of tokens in cells falls below 30. The cells in Table 3 marked with an asterisk are ones with fewer than 30 tokens. These cells are clustered in the *P* condition, suggesting that this interclausal relation may not make a significant contribution to the data overall. This possibility was checked by conducting a multivariate analysis of the data omitting clauses where the subject had been a phonetically null non-subject argument in the prior clause (the *P* condition). There was a significant difference between the two runs, indicating that these tokens make a significant contribution to the overall analysis. Consequently, this factor was retained for the within-groups comparison.⁸

The results of the VARBRUL analysis are given in Tables 4 through 8. Tables 4, 5, and 6 test the hypothesis that *mi* and *you* are pronouns only. Tables 7 and 8 test the hypothesis that *mi* and *you* are both pronouns and subject-verb agreement markers.

TABLE 5. *Weighting of mi and hem i subjects compared (Tryon's analysis predicts isomorphism), Pearson correlation: .448*

	A-o	A	P-o	P	N
<i>mi</i>	0.572	0.52	0.29	0.8	0.41
<i>hem i</i>	0.634	0.71	0.52	0.61	0.25

TABLE 6. *Weighting of ∅ (first/second person singular) and ∅ i (third person singular) subjects compared (Tryon's analysis predicts isomorphism), Pearson correlation: .772*

	A-o	A	P-o	P	N
∅	0.457	0.86	0.43	0.24	0.5
∅ <i>i</i>	0.389	0.77	0.47	0.41	0.29

The degree to which there is comparable within-group behavior of the different subject forms was evaluated using a Pearson correlation to make each pairwise comparison. The correlation for Table 6 is very good, showing that null first and second singular subjects are distributed across the different discourse conditions in a manner very similar to null third singular subjects. The correlations for Tables 4 and 5 are lower, but still positive. This compares favorably with the tests of the Guy/Camden/Charpentier hypothesis, as shown in Tables 7 and 8.

The Pearson correlations for both Tables 7 and 8 return negative values, indicating that the forms being compared in these cases are not distributed in the same way across the different discourse conditions. Subjects with the form *mi* or *yu* alone do not pattern like third singular subjects that are phonetically null (leaving only the agreement marker *i*). I conclude, therefore, that the singleton forms *mi* and *yu* are not subject agreement markers on the verb. Although we saw good evidence in (11) and (12) that *mi* in these clauses might be comparable with the third singular *i*, and although Table 3 bears out the suggestion that doubled subjects are relatively more frequent in the first and second singular than they are in the third singular (approx. 10% of all clauses with first and second singular subjects vs. approx. 5% of all clauses with third singular subjects), a detailed examination of how these forms are used indicates that *mi mi* clusters are not comparable with *hem i* clusters.

To recapitulate, on the basis of the distribution of different subject forms across a range of phonological and discourse environments, we can conclude that *mi* and *yu* are not subject–verb agreement markers, and that the agreement marker anal-

TABLE 7. *Weighting of mi mi and hem i subjects compared (Guy/Camden/Charpentier analysis predicts isomorphism), Pearson correlation: -.449*

	A-o	A	P-o	P	N
<i>mi mi</i>	0.428	0.49	0.72	0.2	0.6
<i>hem i</i>	0.634	0.71	0.52	0.61	0.25

TABLE 8. *Weighting of mi and Ø i (third person singular) subjects compared (Guy/Camden/Charpentier analysis predicts isomorphism), Pearson correlation: -.773*

	A-o	A	P-o	P	N
<i>mi</i>	0.543	0.14	0.57	0.76	0.5
<i>Ø i</i>	0.389	0.77	0.47	0.41	0.29

ogous to third singular *i* and third plural *oli* in these forms is a zero morpheme. Thus, we can revise Table 1 to reflect these findings, as in Table 9.

Since we lack the information to test agreement with the first person inclusive forms in a like manner, we can only tentatively suggest that, for these forms, it is likely that the agreement marker is zero. In fact, if this were the case, it would make a good deal of sense in terms of the historical origins of the *i* agreement marker. In the next section I locate the quantitative findings in the context of the development of the Bislama verb system, outlining their implications for our understanding of a creole's ontogeny.

IMPLICATIONS FOR A CREOLE ONTOGENY

The historical record for Bislama is comparatively good, and this allows us to trace the development of the agreement system identified in the preceding section with some surety.⁹ In the manner we are familiar with from other studies of language contact, an earlier predecessor of Bislama made use of vocabulary from the lexifier(s), while failing to adopt what little morphological redundancy or complexity existed in the lexifier system. In the case of Bislama, the lexifier was chiefly English (though some influence of French is evident). By the mid-19th century, this system had become relatively stable, and a finite clause required only two basic elements: a full NP or pronominal subject and the main verb. This is seen most clearly with third person singular and plural subjects. Examples (17) and (18) contrast the simple pattern of pronoun or full NP subject followed by

TABLE 9. *Basic verb paradigm for Bislama (revised), with boldface indicating person and number agreement*

	Singular	Plural
1	mi Ø karem	mifala i karem
2	yu Ø karem	yufala i karem
3	hem i karem	olgeta oli karem

bare verb attested in Vanuatu in the earliest records of the language with the modern Bislama equivalents.

- (17) a. Missi make him bokis sing.
missionary make 3s box sing
‘The missionary plays the piano.’ (1867)
cf. modern Bislama: *Misi i mekem bokis* . . .
- b. He no same black fellow belong Caledonia!
3s NEG same black fellow of
‘He’s not like the people from New Caledonia!’ (1875)
cf. modern Bislama: *Hem i no olsem* . . .

- (18) a. All man go away.
all man go away
‘Everyone/all the men left.’ (1851)
cf. modern Bislama: *Ol man oli go(we)*.
- b. Plenty man come.
plenty man come
‘Lots of people came.’ (1867)
cf. modern Bislama: *Plante man oli kam*.

The pronouns in the Melanesian English spoken in Vanuatu at this time looked more like the standard English pronominal system than do the pronoun systems used in any of the descendants of Melanesian English today. For example, in the first person singular, there was variation between *ae* and *mi* until the 20th century in Vanuatu; in modern Bislama, Tok Pisin, and Solomon Islands Pijin only *mi* remains. In the third person singular, there was variability between *i* and *hem* until the 1930s; all three of the modern languages only have *hem* or *em* (Crowley, 1990:231). In fact, Crowley showed that the usual form of the third singular pronoun in Vanuatu throughout the 19th century was *i* (from English ‘he’), so the modern standards reflect a significant reanalysis of the system in the last 100 years.

Two other facts documented by Crowley (1990) are relevant here. At the same time as we see the gradual reallocation of *hem* for *i* in the third person singular, he showed that the distribution of *i* was generalized, so that it occurred more widely between a nominal subject and the verb. By the time *hem* became the categorical form of the third singular pronoun in the 1930s, *i* became categorically used with third singular nominal subjects and the verb (Crowley, 1990:243).¹⁰ Crowley

(1990:231, 244–245) also showed that, during the 1920s and 1930s, the pronoun system was fleshed out to include the full range of persons and numbers used in modern Bislama, and that there was a simultaneous expansion in the distribution of *i* for use with non-singular third person pronouns and subsequently to first and second person non-singular pronouns.

During this period (even as now) extensive contact with the substrate languages of Vanuatu continued to be the norm. In a number of these languages, morphemes that are variously described as subject agreement or pronominal copies occur as prefixes on the verb. The parallelism between one substrate exemplar of this, Tangoa, and Bislama was noted in (4). Thus, the potential for *i* to be reanalyzed along the lines of substrate models is obvious. However, this should not blind us to other influential factors. A synthetic approach best explains how contact languages converge on the structures they do.

G. Sankoff (1984) showed that diachronic variation in the Tok Pisin verb phrase reflects the interdependence of the substrate(s) and the lexifier and also probably the effects of even more ineffable, universal factors. She argued that changes in the distribution of *i* with main verbs reflect the influence of cognitive factors (specifically, processing constraints), structural constraints associated with the superstrate input, and substrate patterns (in this case, from both syntax and discourse). Crowley concurred, suggesting that substrate features are more likely to surface in a contact language when they converge with a model in the lexifier and speakers' cognitive requirements (1990:252). Both G. Sankoff's and Crowley's accounts of the grammaticization of elements in Melanesian Englishes and the account proposed here are much in accord with Mufwene's (1996) Founder Principle. Mufwene discussed the likely origins of a number of features of Caribbean and Atlantic creoles and concluded that the greatest likelihood that a form will survive, out of several competing variants, is usually when it is salient in the lexifier and compatible with lexifier and substrate models (1996:103). He argued that the role for universal grammar in language restructuring is to determine which of the competing forms will be selected (1996:114).

The analysis of subject–verb agreement in Bislama presented here adopts a similar synthetic perspective, although I propose a somewhat greater role for universal constraints on syntactic structure than Mufwene endorses. Naturally, the substrate is important. There is continued widespread bilingualism between Bislama and the numerous indigenous Oceanic languages spoken in Vanuatu in which subject agreement is marked by verbal prefixes. These models must have provided a tempting and uncomplicated way of reanalyzing and eliminating competition between pronominal variants.

But properties of subject pronouns in the superstrate may have also played a role in the development of this system. It is worth remembering that English subject pronouns are relatively free morphemes, but not entirely so. They cannot, for instance, stand alone, as object pronouns can:¹¹

(19) Q: Who was already waiting for them?

A: Me/ Him/I was/ He was/ *I/ *He.

In other words, speakers confronted with variability in the form of the third singular pronoun in earlier stages of Bislama would have been getting mixed input about the status of 'he' in the lexifier English. Although it can receive phrasal stress, in other respects it shows clitic-like properties. It seems likely that speakers resolved the situation by transferring substrate features that converge with features of the superstrate system to the creole. *Hem* (from the free object pronoun 'him'¹²) was reallocated as the invariant third person singular pronoun form and *i* (from the partly bound subject pronoun 'he') as third person singular subject-verb agreement. Since the object pronoun 'him' is a free, stress-bearing morpheme, it would have been more salient than the subject pronoun 'he', and as Mufwene (1996:111) showed, the more salient of competing forms often wins out. Keesing (1988) came to a similar conclusion regarding the convergence of substrate models with lexifier forms in his account of similar forms in Solomon Islands Pijin.

But a system in which agreement is marked in only some persons and numbers seems to have been inherently unstable. Either substrate models (languages in which agreement is marked in all persons and numbers), or universal tendencies (such as towards paradigm regularization), or most likely the convergence of both set the stage for a subsequent generalization of agreement marking to other persons and numbers.¹³ In some cases the resulting agreement morpheme is phonologically null, but what is more important is that the ensuing system makes a regular distinction between singular and non-singular in all persons.

With the exception of G. Sankoff's work on variation in Tok Pisin, little research has fully incorporated the variation that existed in early stages of Melanesian English into the analysis of synchronic patterns and the emergence of structure.¹⁴ Instead, the study of Pacific creoles has been characterized by a tendency to abstract away from the variability of the input. The effect of this has been to emphasize the structural stability and the degree of systematicity that existed even in early forms of these languages. This can certainly be a worthwhile strategy, particularly in view of the continued widespread perception (both in and out of Melanesia) that these languages are nothing more than broken English. In addition, by positing discrete invariant stages of the grammar it has been possible to apply tools and methodologies drawn from historical linguistics to the analysis of these varieties. This has had the beneficial effect of emphasizing the relevance of creoles to issues of broader linguistic significance.

Nevertheless, this idealization risks obscuring important details about how speakers identify and ultimately resolve problems in the analysis of their language. In this case, by examining the variable distribution of pronominal subjects we can reveal how morphosyntactic complexity, lost at an earlier stage in Bislama, is reintroduced. I have (following G. Sankoff and Crowley) adopted a more catholic approach to the ontogeny of a creole and have shown that elements of the lexifier were modified in such a way as to satisfy needs that may have ultimately derived from the substrate or from more general cognitive or linguistic constraints.

The speed with which this occurred is remarkable. The developments outlined in this section essentially took place within three generations of speakers, while

TABLE 10. *Frequency of overt and phonetically null subjects in Bislama: first and second person singular compared with third person singular*

Subject Form	<i>mi/yu</i>	<i>hem</i>
Null	118	1,010
Overt	1,364	664

Bislama was still principally acquired as a second (at least) language, in contact with numerous substrate languages, and while the pool of speakers was still largely composed of a minority of speakers in the wider speech community (typically, men in paid employment).

It was suggested, too, that the reanalysis observed may have been motivated by imperatives even more abstract than the transfer of substrate models. In the next section, I present further evidence to support the claim that the synchronic distribution of pronouns and agreement marking indicate that the reanalysis of Bislama is at least partly constrained by crosslinguistic principles of universal grammar.

THE DEVELOPMENT OF A GRAMMAR LICENSING PRO-DROP

This claim is supported by looking more closely at the patterns of phonetically null and overt pronominal subjects found in conversational Bislama. It is argued that the distribution of these forms is best explained in terms of the morphosyntactic informativeness of the inflectional system. In doing so, we see that the account of first and second person singular subject–verb agreement put forward earlier is more than descriptively adequate. It provides the basis for a unified account of the distribution of phonetically null subjects as well.

In Table 3, the reader may have noticed a marked difference in the frequency of overt/null first and second singular subjects and the frequency of overt/null third singular subjects. The relevant frequencies are repeated in Table 10.

Table 11 shows that this distinction between the persons generalizes across singular and plural. Table 11 shows the probability with which subjects in each person and number occur with a phonetically null subject (complete co-occurrence = 1) or a pronominal subject (complete co-occurrence = 0). The probabilities are shown as weightings derived by means of logistic regressions calculated using VARBRUL (D. Sankoff et al., 1992).

These calculations take into account the relative frequency of each subject type in the corpus as a whole (shown in the far right column of Table 11) as well as their frequency relative to the other linguistic and nonlinguistic variables investigated. This controls for the possibility, for example, that a single speaker or

TABLE 11. VARBRUL weightings for all speakers for person and number of subject referent (0 = consistently overt subject; 1 = consistently null subject), input probability for all speakers = 0.418

Person and Number of Subject Referent	VARBRUL Weighting	Number of Clauses
Third singular	0.709	1,719
Third plural	0.869	846
First singular	0.147	1,054
First plural	0.268	375
Second singular	0.206	397
Second plural	0.197	71

small group of speakers might be responsible for an overwhelming preponderance of the tokens in one category. The results show that third person subjects pattern alike in the singular and the plural; in these conditions speakers favor a phonetically null subject. First and second person singular and plural subjects pattern alike as well, but in these conditions speakers favor the use of an overt pronoun in subject position.

Unfortunately, the historical record concerning the use and distribution of phonetically null subjects in Bislama is sketchy. Crowley (1990:241) noted the possibility of zero anaphora, but only for third person referents. He suggested that the use of zero anaphors in these cases is constrained by the animacy of the referent (with human nouns favoring a pronominal anaphor). But the examples he provided suggested that animacy is not the principal constraint, since human subjects are frequently null when they have nonspecific reference. Since first and second person subjects are more likely to be realized with a pronoun and since first and second subjects are de facto high on an animacy hierarchy, it is possible that Crowley’s conclusion for third person is an artefact of this.

Charpentier provided our best historical perspective, noting that “In older pidgin [Bislama], the verb was always preceded by a noun or subject pronoun Nowadays, the personal pronoun tends to be left out, leaving utterances that consist only of subject–verb agreement and the verb stem” (1979:353; translation David Heap).¹⁵ This suggests that the possibility of omitting pronominal subjects is a grammatical innovation that emerged during the 1960s 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, created a comparatively large pool of first language speakers of only Bislama.¹⁶ Since these are precisely the conditions under which we might expect to see the effects of more general or universally endowed constraints on language become stronger than the effects of the substrate, the development of what appears to be a pro-drop grammar is especially noteworthy.

Moreover, the fact that Bislama appears to have grammaticized a split pro-drop system is of even greater significance.

For reasons of space, it is clearly impossible to review the details of accounts (both generative and functional) that have been provided for grammars in which null subjects (or zero anaphora) occur. A brief sketch must suffice. I suggest that the data from Bislama shows that a grammar of pro-drop requires recourse to principles from both schools of thought.

A generative account of pro-drop requires a licensing mechanism for the omission of the subject and some means by which the subject can be identified. Licensing, being a structural property, is an attempt to capture how phonetically null subjects and overt subjects could equally well satisfy requirements to assign a verb's theta-roles and case assignment. Identification is more of an interpretive property, intended to capture the insight that, despite their lack of phonetic substance, interlocutors make reliable and correct associations concerning person, number, and salience of the subject referent.

Initially, based on the European languages that license phonetically null subjects, it was supposed that identification was satisfied by the rich inflectional morphology found on a finite verb uniquely distinguishing subjects in all persons and numbers. However, it was quickly noted that languages without any agreement between subject and verb nevertheless license null subjects, and the theory was revised accordingly. Jaeggli and Safir (1989) proposed that a system of morphological uniformity was the necessary condition (i.e., all persons marked or no persons marked), but this account, too, rapidly foundered on empirical rocks. Grammars in which null subjects are licensed in some, but not all, persons and tenses were noted (Borer, 1989), and somehow licensing had to be defined in such a way that it could apply to only parts of the grammar. Different derivations for different parts of the morphosyntactic system were proposed (Poletto, 1996; Vainikka & Levy, 1995), and Rohrbacher (1995) suggested that the crucial trigger required to license null subjects might be an overt morphosyntactic distinction between first and second person subjects, since these two persons have an inherent discourse salience.

Functional accounts of pro-drop have argued that the phenomenon is constrained by the salience of the subject. Different measures or conceptualizations of salience have been proposed. Chafe (1994) suggested that a limited number of referents can be active in interlocutors' consciousness, but conceded that the number of referents that can be active at once and the processes by which interlocutors keep some referents (especially third person referents) active and allow others to fall into the background are still poorly understood. Fox (1987) and Webber (1991) engaged with these questions using more specific analytical tools. Although their models of discourse structure differ, they share a perspective in which discourse relations are organized according to principles of dominance and adjunction, analogous to the kinds of intersentential relations posited in systems of formal grammar. It is suggested that, although speakers may use anaphors to refer to referents of propositions that occurred some time back in the discourse, their ability to do this is constrained to cases in which there is a unidirectional chain linking the anaphor and the referent/proposition in the discourse structure.

Grosz et al. (1995) posited a particularly constrained metric for anaphor resolution, which they called Centering. This is based on measures of a referent's prior salience in the discourse and its planned persistence in the next utterance. These constraints render Centering considerably more tractable and plausible as a model of natural language processing and production than, say, Givón's (1984) measures of salience and persistence. Prince (1996) presented interesting results from Yiddish texts showing that Centering provides an excellent model for anaphora resolution in all persons and numbers except second person singular. This suggests that, even if the Centering algorithm accurately represents natural tendencies towards anaphor resolution, in some languages social or cultural factors may mark specific subject referents (such as second singular in Yiddish) as conversationally salient, overriding Centering constraints and allowing a subset of pronouns to behave somewhat differently than other subjects.

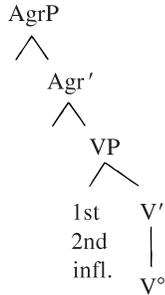
However, despite differences in how different discourse-based theories propose that interlocutors resolve anaphors, there is general agreement that some cline of salience correlates directly with a subject's form. Phonetically null subjects (in languages that allow them) are uniformly seen as occurring only when the referent has high (or highest) saliency. Thus, the chief difference between discourse-based and generative accounts of pro-drop is that discourse-based theories require identification of a null subject (via some attentional algorithm), but make no predictions about what kinds of grammars will license null subjects. Generative theories, on the other hand, frame pro-drop as a reflex of particular intrasentential structures that license it and assume identification to also be resolved sentence-internally.

Although I am aware of no work in the functionalist tradition that specifically looks at split systems of pro-drop, Chafe's reference to "the usually active status of the ideas of the speaker and listener" (1994:79) would suggest that a functional analysis of Hebrew and Finnish might proceed along the following lines. In both these languages, the split is such that pro-drop is allowed in the first and second persons, but an overt pronoun is required in the third.¹⁷ It might be proposed that first and second person can be phonetically null by virtue of the referents' necessary presence in any speech event, since this would ease the processing burden on the hearer. However, such a functional hypothesis needs to be tested empirically. One way to do this would be to compare the relative rates of phonetically null first and second person pronouns in reported speech with the rates of null forms when the speaker and hearer are referred to. If salience in the immediate context determined how easily these forms could be omitted, we would expect null forms to be more common when referring to the actual interlocutors than in reported speech (in which a first or second person might index a third party). This presents an obvious avenue for future research.

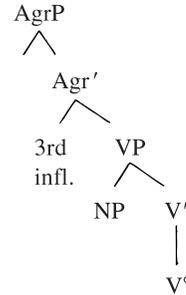
In fact, this kind of functional argument has been drawn on in motivating the direction in which a pronominal system splits in even the generative accounts of Hebrew and Finnish. Vainikka and Levy (1995) suggested that first and second person subjects may be phonetically null because they are highly referential. That is, the agreement suffix is transparently derived from the independent pronoun forms, and identification of the referent through a transfer of all necessary phi-features from the agreement suffix is possible.

It requires considerably more work to make a split pro-drop system tractable with the licensing requirement imposed by generative theories. Vainikka and Levy (1995) suggested that inflections for first and second person are base-generated in the subject position, thus licensing pro-drop in these persons. The less informative inflections of third person are, they suggested, base-generated in the head of Infl, as shown in (20).

(20) a. highly referential inflection



b. inflection low in referentiality



The subject inflections in (20a) act as arguments, absorbing a theta-role from the verb, and so a null subject is merged in the spec of Agr only in order for nominative case to be assigned felicitously. By contrast, in (20b), the inflection is not in a position to absorb the agent theta-role. Thus, for the sentence to be grammatical, some other overt subject must be generated in the subject position.

Poletto (1996) presented a similar analysis for the distribution of subject clitics and null subjects in Basso Polesano (Italy). She suggested that the partial licensing of null subjects in Basso Polesano (null subjects occur in first and second person; third person requires an overt clitic pronoun¹⁸) is due to the fact that first and second person subject–verb agreement is base-generated (or we might now say, merged) in the position of an argument of the verb, whereas the third person agreement is part of Infl, much as in (20b). Quasi-functional motivations for this situation can be offered: for example, because first and second person are always contextually salient, this maps into a kind of structural salience filling the syntactic position of subject (in this way, generative accounts find common ground with functional accounts).

Extending these accounts to the situation in Bislama would require an inversion of the mechanism. With third person subjects (which favor being null) agreement would be argument-like, and in the first and second the agreement *i* would be part of Infl. Note that this vitiates the functional account of a split system. Data from Faetar (a dialect of Franco-Provençal spoken in Italy) show that such a counterfunctional split in a pro-drop system is not peculiar to Bislama (Nagy & Heap, 1998).

The Bislama data, instead, suggest that Vainikka and Levy's notion of high referentiality is more appropriate. Like Finnish and Hebrew first and second person subject–verb agreement, third person agreement in Bislama is most transparently referential. The third singular agreement marker *i* derives from the lexifier pronoun 'he', and the third plural *oli* derives from the lexifier 'all'. Again we see

cause for pursuing the more unified G. Sankoff/Crowley approach to creole development, here stressing the interdependence of lexifier patterns and perhaps the universal principles governing the syntactic relations permitted in pro-drop languages.

However, even if these structural accounts are appropriate for Hebrew, Finnish, and Basso Polesano, once the extent of the variation in the split system of Bislama is taken into consideration (see Table 10), we have reason to pause. It is worth noting that none of the authors discussing these languages gives us any information on the actual frequency with which the different persons and numbers occur with or without an overt subject, so we have no empirical way of evaluating the appropriateness of their accounts. Heap's (1997) quantitative analysis of variation between pronouns and null anaphors in Southern French and Northern Italian dialects showed that the splits observed there are far from categorical.

In order to posit different derivations for clauses with a null subject and clauses with a pronominal subject, the Bislama data requires that a grammar be flexible. We cannot simply state that the grammar of Bislama generates third person agreement as arguments and first and second person agreement as inflections. As we have seen, this is not true approximately 20% of the time in each case. In about one in five cases, speakers merge third person subject–verb agreement as Infl, requiring a separate subject to fill argument position; conversely, first and second person agreement are treated as syntactic arguments at roughly the same frequency. Two questions of theoretical importance arise out of this descriptive fact. First, why do we see the rates of variability we do? Second, is the variation within persons incompatible with a structural account of the variation between persons?

With respect to the first question, we note that in both cases (third person and first/second persons) the variation falls within roughly the same bounds. This would be a remarkable coincidence if they were completely independent. However, if the structural analyses proposed for, say, Basso Polesano hold for Bislama too, then it follows that the different subject types are not independent variables. If there is some latitude or variability in where subject–verb agreement is base-generated (as argument or inflection), then it makes sense for the degree of variability to be consistent across all persons.

Another question arises as a consequence of the relationship between the notion of a grammar and variation. If we adopt the proposed structural analysis of Bislama, according to which clauses with a third person subject base-generate agreement in an argument position and those with first and second person subjects base-generate agreement in Infl, we do not want this to imply that the Bislama sentences that do not seem to be generated in this way (approx. 20%) are ill-formed or ungrammatical. This clearly would be undesirable. Instead, the Bislama data provides important evidence that speakers may analyze or implement core elements of the grammar in more than one way. The historical record provides some clues as to how a situation such as this might arise.

In the early 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* 'first singular', *yu* 'second singular', and (*h*)*i* 'third singular/plural') was

used. By the turn of the century, the system had been reanalyzed so that third person singular and plural pronouns were distinct (*hem* ‘third singular’ and *olgeta-(fala)* ‘third plural’), with *i* being reassigned as an agreement marker (Crowley, 1990:231, 243). But even while *hem* was crystallizing as the third singular pronoun, *i* continued to occur in what appeared to be subject position in some clauses—namely, when *hem* was absent. It cannot be determined whether these instances 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 third singular null subjects, but the fact of the variation alone is vital.

I hypothesize that the variation at this crucial stage might have created the possibility for speakers to interpret *i* either as a pronoun or as an agreement marker. This would have been compatible with both the immediate history of the forms (English *he* and pidgin (*h*)*i* as subject pronouns) and with the available substrate models (Oceanic languages with agreement prefixes on the verb). In other words, I suggest that convergence between the lexifier and the substrate is the proximate cause of creole variation.

If I am correct, this might have fed 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 emerges for first and second person as well. Independent factors, perhaps general (or universal) principles favoring regularity in an inflectional paradigm, result in the singular/plural contrast that is marked overtly in the third person being generalized further. The contrast between \emptyset and *i* agreement in the first and second persons distinguishes singular and plural, respectively. In short, the data again requires an account of creole variation and creole grammatical development that synthesizes the influence of factors from a variety of domains.

We lack the quantitative data from earlier parts of the 20th century that would conclusively support the diachronic account just outlined. Nevertheless, some of the data from the 1990s suggest that it is more than speculation and that ongoing developments in Bislama continue to follow this path. 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 suggests that there is indeed some change taking place with respect to the null subject variable.

When the comparison was restricted to the children in the corpus and their immediate caregivers and their uncles and aunts (13 individuals in total), it was found that the children produced a significantly different pattern than both older groups of speakers (a regression analysis comparing children aged 12 and under with caregivers and young aunts and uncles yields $r^2(\text{adjusted}) = 71.6\%$, $p < .001$).¹⁹ 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. The probabilities for children under 12 are shown in Table 12 and should be compared with the overall trends in the community shown in Table 11.

In short, the data from the children and their caregivers suggest that the synchronic variation found in the speech of individuals in Bislama reflects dia-

TABLE 12. VARBRUL weightings for children 12 years and under only (0 = consistently overt subject; 1 = consistently null subject)

Person and Number of Subject Referent	VARBRUL Weighting
Third singular	0.668
Third plural	0.725
First singular	0.029
First plural	0.013
Second singular	0.008
Second plural	0.039

chronic change in progress. While variation may persist over some generations, I believe that, because the children approach near-categoriality in the first and second person, this suggests that the system may be converging on one with less (though not necessarily no) variation; one that would be compatible with some of the structural analyses of split pro-drop systems observed in other language families.

Notice that this synthetic (lexifier, substrate, universalist) account of the Bislama constraints on pro-drop does not tell us anything about the extent to which these factors contribute to the observed facts. For example, we can only suggest that the substrate provides a morphological model for the development of subject–verb agreement. The analysis of Bislama subject–verb agreement outlined earlier is an argument that is a bit more complex syntactically. It posits that the morphological model is later reanalyzed syntactically, such that the morpheme is treated as an argument in the third person singular and plural (which is why the null pronominal condition is favored with these subjects) and as inflection elsewhere (which is why in first and second person speakers tend to use a subject pronoun). There may indeed be a direct syntactic model for this in substrate languages, but unfortunately as yet we are unable to identify it. We need to understand better what the most important substrate languages have been (and are) for Bislama, and we need more grammars that analyze the discourse–syntax interface and internal variation in those languages. Until we have this, we cannot claim that the substrate alone provides a syntactic model for this variable.

This distinction is worth making since there seems to be an inclination to overemphasize the influence of substrates. A comparison of word/morpheme order may support an analysis of the transfer of words or morphemes, but to suggest that the syntactic structure associated with those words has been transferred requires more than an analysis of surface word order. Winford (1993) discussed the necessity of matching not just forms but abstract relations in creole studies. Transfer of abstract syntactic elements from the lexifier can be ruled out since, although we assume speakers are targeting the lexifier, we do not assume that they have such intimate familiarity with it that they could map such highly

abstract aspects of the lexifier into a pidgin or creole. A useful contrast is provided in King's (forthcoming) work on contact-induced change in Acadian French, where speakers clearly do have the necessary familiarity with English to transfer aspects of English syntax that might be projected by a borrowed lexeme.

Considerable work is still required on the distribution of null and pronominal subjects in the substrate languages in Vanuatu. Assuming that the languages which have had the greatest input to the vocabulary of Bislama have likewise had the greatest influence on the emerging syntax (and this remains an empirical question), what is most needed is information on the distribution and constraints on null subjects in languages of central Vanuatu. Even fine grammars such as Early (1994) for Lewo tend not to have this kind of detail. It is hoped that future research will address this matter. However, even if patterns of phonetically null subjects are found in substrate languages which prove to be comparable to those found in Bislama, this will not necessarily obviate the role of universals entirely.

One final question needs to be answered as to the probable outcome of this ongoing reanalysis. It is unlikely that the outcome will be a time in which speakers of Bislama categorically use null anaphors in the third person and overt pronouns in the first and second. We see that 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 preceding clause play a part in constraining the distribution of null subjects in Bislama. An important consequence of their significance is to 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.

CONCLUSION

At the outset, the problem faced in this article 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 first and second person singular).

This analysis was then considered in light of the historical record concerning the development of the pronoun system and the generalization of *i* in finite clauses in Bislama, and it was seen that a mixture of resources (lexifier and substrate) were speedily exapted for use in the developing morphological system. To the

extent that this provides us with a clearer picture of the manner and rapidity with which one language has reorganized its morphosyntactic system, it contributes to a better understanding of more general features of language contact.

Finally, 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. Null subjects are emerging as the norm in the third person, but not in the first and second person in Bislama. This makes sense since subject–verb agreement is maximally informative in the third person. That is, as generative accounts of the null subject parameter would predict, the emerging Bislama system satisfies the requirement that a null subject be identified.

In the current climate of research into contact languages and with the recent emphasis on the possible transfer of substrate features in the study of Pacific creoles in particular, the story of the emergence of subject–verb agreement in Bislama provides a timely reminder. What surfaces in the form of variable or categorical rules in a contact language is often a form that is compatible with substrate(s), lexifier, and universals. It seems that, as in much of life, the whole story in language contact is a story of greed, compromise, and the path of least resistance.

NOTES

1. Calling Bislama a creole is not uncontroversial. Although increasing numbers of speakers acquire it as their first language, it remains an L2 for the majority. However, a great many of those speakers begin to acquire it under L1-like conditions in school playgrounds, etc. well within the critical period, say, at 5 to 6 years of age. Many such speakers go on to become Bislama-dominant adults, employing Bislama for all the social, discursive, and affective functions a monolingual speaker of Bislama would. See Jourdan (1985) for similar social and functional arguments regarding Solomons Pijin.
2. Examples are identified according to place of recording, year, tape number, and speaker.
3. Camden’s translation and gloss of morphemes. Camden used *c* for a voiceless velar fricative; he actually wrote the /r/ with a tilde to indicate a retroflex trill.
4. Example (6c) would have to be read with heavy stress on the first pronoun and a pause before *bae*. I take these facts to indicate that the first pronoun is adjoined outside the clause, similar to left-dislocated NPs in English. Such stress patterns are not associated with (6b).
5. Indeed, Crowley (1990:234) noted it as “an alternative generalization.” In recent personal correspondence, Crowley distanced himself from any implied endorsement of this alternative.
6. The presence or absence of *i* in these contexts was verified by Sharon Tabi, a Bislama-dominant speaker from Pentecost.
7. Bislama has no passive, so this condition generally reflects a continuity between clauses in thematic role as well.
8. Testing the significance of the contributions of the tokens in the *P* condition to the model’s fit in all discourse conditions yields the following:

	LL (all conditions)	LL (no P)	<i>df</i>	<i>p</i>
<i>mi mi/you you</i>	–486.591	–483.004	1	<0.025
∅ (first/second singular)	–361.077	–357.032	2	<0.025
NP, <i>hem i</i>	–246.193	–238.75	1	<0.001
<i>hem i</i>	–965.824	–922.704	1	<0.001

Runs evaluated the probabilities of *mi mi/you you* and ∅ (first/second singular) forms occurring in all five discourse conditions compared with *mi/you* subject forms; NP, *hem i* forms compared with *hem i*;

hem i compared with $\emptyset i$. As can be imagined, similar or more dramatic results obtain omitting the *P-o* condition. The NP, *hem i* tokens remained in the analysis because they are central to the hypotheses.

9. Much of this section is informed by Crowley (1990). I am also indebted to Terry Crowley for sharing with me a copy of his corpus of early Bislama. I have augmented his corpus with some materials I have also found. His support, of course, in no way entails that he agrees with my interpretation of the facts.

10. "Categorically" here is used in the sense of statistically approaching categoriality. Crowley did note some tokens of third person singular NP followed immediately by the verb in the 1920s and 1930s, but at levels so low that they are probably statistically not significant. He did not provide actual numbers, only percentages, so this cannot be tested. Note that G. Sankoff (1994) traced a similar specialization of *i* in Tok Pisin, occurring somewhat earlier than in Bislama (between the 1870s and 1900).

11. It is important to remember that we are interested in older norms for spoken colloquial English, which were the superstrate influences in Bislama. Subject pronouns may be beginning to be used in a wider range of domains in English (e.g., as a response to (18)), but this recent innovation is irrelevant to the point being made here. My thanks to Richard Kayne and David Heap for discussion of this.

12. Perhaps both 'him' and 'them', the third plural pronoun has dialectal variants of *em*.

13. Let me forestall objections to the idea that a system which marks agreement only in the third person singular is unstable, since English has such a system in the present. Two things can be noted about English. First, outside of contact situations, there is no substratum with which universal tendencies might converge. Second, there is actually a good deal of variability in the English system. Many English vernaculars generalize zero or *-s* inflections in the present. Insofar as we might cast speakers of vernaculars as speaking in a situation of language contact (between the vernacular and the standard), it might be argued that we *do* find the tendency towards restructuring.

Of course, it remains to be demonstrated that contact between closely related dialects has the same linguistic outcomes as contact between unrelated or distantly related languages. It is clear, for instance, that the two situations often differ radically in terms of the social evaluations of the varieties in contact, and it would not be surprising to find that such qualitative differences are reflected in the linguistic outcomes.

14. Crowley, too, recorded not only the *fact* of variation at different stages of Bislama but also the relative *frequency* of forms, and in my opinion his 1990 book remains the next most sophisticated exemplar of variationist linguistics for Melanesian Englishes. The situation for Melanesia stands in marked contrast to the work done on Atlantic creoles, where connections between synchronic and diachronic variation have been attended to more closely; see Naro (1981), Rickford (1987), Rickford and Blake (1990), Patrick (1996), Scherre and Naro (1991), Singler (1987, 1989), and Winford (1990, 1993).

15. "Dans le pidgin ancien, le verbe était toujours précédé d'un nom ou d'un prénom sujet. ... Aujourd'hui, le prénom personnel tend à être abandonné, la modalité personnelle et le radicale verbal composant seuls un énoncé" (Charpentier, 1979:353).

16. 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.

17. This simplifies the system somewhat, since there are also interactions with tense in both languages (Borer, 1989; Vainikka & Levy, 1995).

18. I simplify Poletto's data somewhat. She shows that there is a clitic associated with first and second person in Basso Polesano, but proves that this is not analogous to the third person clitics.

19. All speakers, including the children, have some competence in at least one other vernacular besides Bislama. For most of them, and for all the children, this was Tamambo (the language spoken on western Malo). All the children concerned began to acquire Bislama along with Tamambo in the home from birth as at least one of their primary caregivers (mother or aunt) was a speaker of some language other than Tamambo.

REFERENCES

- Baker, Philip, & Mühlhäusler, Peter. (1996). The development and diffusion of pronouns in Pacific Pidgin English. In Stephen Wurm, Peter Mühlhäusler, & Darrell T. Tryon (eds.), *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*. Berlin: Mouton de Gruyter. 537–549.
- Borer, Hagit. (1989). Anaphoric AGR. In Osvaldo Jaeggli & Kenneth J. Safir (eds.), *The null subject parameter*. Dordrecht: Kluwer. 69–109.

- Camden, Pastor Bill [William G.]. (1977). *A descriptive dictionary: Bislama to English*. Maropa Bookshop: Vila.
- Camden, William G. (1979). Parallels in the structure and lexicon and syntax between New Hebrides Bislama and the South Santo language spoken at Tangoa. *Papers in Pidgin and Creole Linguistics*, No. 2 (Pacific Linguistics, A-57). Canberra: Australian National University. 51-117.
- Cameron, Richard. (1993). Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change* 5:305-334.
- Chafe, Wallace. (1994) *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: Chicago University Press.
- Charpentier, Jean-Michel. (1979). *Le Pidgin Bislama(n) et le multilinguisme aux Nouvelles-Hébrides* (Langues 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.
- Early, Robert. (1994). *A grammar of Lewo, Vanuatu*. Doctoral dissertation, Australian National University.
- Fox, Barbara. (1987). *Discourse structure and anaphora*. Cambridge: Cambridge University Press.
- Givón, Talmy. (1984). *Syntax: A functional-typological introduction* (Vol. 1). Amsterdam: Benjamins.
- Grosz, Barbara, Joshi, Aravind, & Weinstein, Scott. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21:203-225.
- 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. 1-36.
- Guy, J. B. M. (1974). *Handbook of Bichelamar/Manuel de Bichelamar* (Pacific Linguistics, C-34). Canberra: Australian National University.
- Heap, David John. (1997). *La variation grammaticale en géolinguistique: Les pronoms sujet en roman central*. Doctoral dissertation, Department of French Language and Literature, University of Toronto.
- Jaeggli, Osvaldo, & Safir, Kenneth J. (1989). The null subject parameter and parametric theory. In Osvaldo Jaeggli & Kenneth J. Safir (eds.), *The null subject parameter*. Dordrecht: Kluwer. 1-44.
- Jourdan, Christine. (1985). *Sapos iumi mitim iumi: The social context of creolization in the Solomon Islands*. Doctoral dissertation, Australian National University.
- Keesing, Roger M. (1988). *Melanesian Pidgin and the Oceanic substrate*. Stanford, CA: Stanford University Press.
- King, Ruth. (forthcoming). *The lexical basis of grammatical borrowing: A Prince Edward Island French case study*. Amsterdam: Benjamins.
- Mufwene, Salikoko S. (1996). The Founder Principle in creole genesis. *Diachronica* 13:83-134.
- Nagy, Naomi, & Heap, David. (1998). Subject pronoun variation in Faetar and Francoprovençal. *Papers in Sociolinguistics. NWAWE-26 à l'Université Laval*. Québec: Nuits Blanches.
- Naro, Anthony J. (1981). The social and structural dimensions of a syntactic change. *Language* 57:63-98.
- Patrick, Peter L. (1996). The urbanization of creole phonology: Variation and change in Jamaican. In Gregory R. Guy, Crawford Feagin, Deborah Schiffrin, & John Baugh (eds.), *Towards a social science of language: Papers in honor of William Labov, I: Variation and change in language and society*. Amsterdam: Benjamins. 329-355.
- Poletto, Cecilia. (1996). Three kinds of subject clitics in Basso Polesano and the theory of pro. In Adriana Belletti & Luigi Rizzi (eds.), *Parameters and functional heads: Essays in parametric syntax*. New York: Oxford University Press. 269-300.
- Prince, Ellen F. (1996). Subject pro-drop in Yiddish. Manuscript, University of Pennsylvania.
- Rickford, John Russell. (1987). *Dimensions of a creole continuum*. Stanford, CA: Stanford University Press.
- Rickford, John Russell, & Blake, Renée (1990). Copula contraction and absence in Barbadian English, Samana English, and Vernacular Black English. In Kira Hall, Jean-Pierre Koenig, Michael Meacham, Sondra Reinman, & Laurel A. Sutton (eds.), *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society. 257-268.
- Rohrbacher, Bernhard. (1995). Explaining the syntactic consequences of 'rich' agreement morphology: On the licensing of V-to-AgrS raising and pro. In Raul Aronovich, William Byrne, Susanne Preuss, & Martha Senturia (eds.), *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*. Palo Alto, CA: Stanford Linguistics Association/CSLI. 350-364.

- Sankoff, David, Rousseau, Pascale, Hindle, Don, & Pintzuk, Susan. (1992). GOLDVARB, version 2.1. Adapted for the Macintosh by David Rand, 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.
- _____. (1994). An historical and evolutionary approach to variation in the Tok Pisin verb phrase. In *Papers from the 30th Regional Meeting of the Chicago Linguistic Society. Vol. 2: The parasession on variation in linguistic theory*. Chicago: Chicago Linguistic Society. 293–320.
- Scherre, Maria Marta Pereira, & Naro, Anthony J. (1991). Marking in discourse: “Birds of a feather.” *Language Variation and Change* 3:23–32.
- Singler, John Victor. (1987). The city, the mesolect, and innovation. *Journal of Pidgin and Creole Languages* 2:119–147.
- _____. (1989). Plural marking in Liberian Settler English, 1820–1980. *American Speech* 64:40–64.
- Tryon, Darrell. (1987). *Bislama: An introduction to the national language of Vanuatu* (Pacific Linguistics, D-72). Canberra: Australian National University.
- Vainikka, Anne, & Levy, Yonata. (1995). *Empty subjects in Finnish and Hebrew* (IRCS Report series 95–31). Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania.
- Verhaar, John W. M. (1995). *Toward a reference grammar of Tok Pisin: An experiment in corpus linguistics* (Oceanic Linguistics Special Publication No. 26). Honolulu: University of Hawai‘i Press.
- Webber, Bonnie L. (1991). Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes* 6:107–135.
- Winford, Donald. (1990). Copula variability, accountability, and the concept of ‘polylectal’ grammars. *Journal of Pidgin and Creole Languages* 5:223–252.
- _____. (1993). Variability in the use of perfect Have in Trinidadian English: A problem of categorial and semantic mismatch. *Language Variation and Change* 5:141–187.-