

*Constraints on null subjects in
Bislama (Vanuatu):
social and linguistic factors*

Also in Pacific Linguistics

Levinsohn, Stephen H., ed. 1991. *Thematic continuity and development in languages of Sabah*.

Corston, S.H., 1996. *Ergativity in Roviana, Solomon Islands*.

Clark, Ross, 1998. *A dictionary of the Mele language (Atara Imere), Vanuatu*.

Tryon, Darrell, 1987. *Bislama: an introduction to the national language of Vanuatu*.

Crowley, Terry, 1999. *Ura: a disappearing language of Southern Vanuatu*.

Vamarasi, Marit, 1999. *Grammatical relations in Bahasa Indonesia*.

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, the Philippines, Indonesia, Southeast and South Asia, and Australia.

Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The Editorial Board of Pacific Linguistics is made up of the academic staff of the School's Department of Linguistics. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise who are usually not members of the editorial board.

The 470 publications before Publication 501 were numbered in four series, A, B, C and D, but the series were merged from the beginning of the year 2000 because the rationale for them had gradually disappeared.

FOUNDING EDITOR: Stephen A. Wurm

EDITORIAL BOARD: Malcolm D. Ross and Darrell T. Tryon (Managing Editors),
John Bowden, Andrew Pawley

Pacific Linguistics 506

Constraints on null subjects in Bislama (Vanuatu):

Social and linguistic factors

Miriam Meyerhoff



**Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University**

Published by Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University
Canberra ACT 0200
Australia

Copyright © The author

First published 2000

National Library of Australia Cataloguing-in-Publication entry:

Miriam Meyerhoff

Constraints on null subjects in Bislama (Vanuatu): social and linguistic factors

Bibliography.

ISBN 0 85883 522 3

1. Sociolinguistics – Vanuatu. 2. Bislama language –
Grammaticalisation. 3. Bislama language – Variation.
I. The Australian National University. Research School of
Pacific and Asian Studies. II. Title. (Pacific Linguistics 506).

427.99595

Copyedited by Becky Quick

Typeset by Jeanette Coombes

Maps redrawn by The Cartography Unit, RSPAS, ANU

Printed by ANU Printing Service, Canberra

Bound by F & M Perfect Bookbinding, Canberra

Table of Contents

<i>Ol Tankyu: Acknowledgements</i>	ix
<i>List of abbreviations</i>	x
1 On the nature of linguistic variables	1
1 Why the study of language variation is of interest	1
1.1 Towards a typology of variables: inherent and derived	5
1.1.1 Inherent and derived phonological variables	5
1.1.2 Inherent and derived syntactic variables	7
1.1.3 Sociolinguistic variables: never inherent and always optional	10
1.2 Implications of this typology of variables	11
1.3 Optional but underived variables	15
1.3.1 Interactions between discourse and the grammar	17
1.4 Conclusion	17
2 <i>Lukluk ples</i>: social and historical overview	19
2 Introduction	19
2.1 Brief history focusing on the region	21
2.1.1 First settlement to European contact	21
2.1.2 European contact	23
The labour trade	26
Twentieth century developments	28
2.2 Sketch of contemporary social life	31
2.2.1 Life in Santo (1994–1995)	31
2.2.2 Life on Malo (1994–1995)	33
2.3 Current linguistic situation	35
2.3.1 Bislama: the one national language	35
2.3.2 Use of English and French	37
2.3.3 Diversity and variation in Bislama today	39
2.3.4 Variation in Tamambo	40
2.4 Conclusion	41

3	<i>Tingting blong lanwris long Malo: language use, and some attitudes to language on Malo</i>	42
3	Introduction	42
3.1	Choosing a language: indexing non-linguistic information	42
3.2	Methods of investigating language use: intersections with social psychology and anthropology	44
3.3	Language use on Malo: <i>man Malo i havem tu sot</i>	46
3.3.1	A theoretical and practical disclaimer	47
3.4	Summary: sources of data on language use and preference	47
3.5	Self-reports and observations: use of Bislama and Tamambo	48
3.6	What are people reporting in their self-reports?	51
3.6.1	On the meaning of 'English' and its use outside of school	51
3.6.2	On the use of French outside of school	52
3.6.3	Languages used by <i>ol woman nara aelan</i>	54
3.6.4	Language choices in religion	55
3.6.5	Children's exposure to and control of Tamambo	56
3.6.6	Women's and men's different experience of 'public events'	58
3.7	Identifying subgroups: shared norms for choosing a language	60
3.7.1	Group A: The Malo hardcore	62
3.7.2	Group B: Malo mainlanders	63
3.7.3	Group C: Ties that bind (off-island)	64
3.7.4	Group D: <i>Woman nara aelan</i>	64
3.7.5	Summary of Malo subgroups	66
3.8	Motivating language choice: the instrumental <i>is</i> affective	66
3.8.1	With us or against us: collective or group factors affecting code choice	68
3.8.2	Trading on language: dos-and-don'ts at school	69
3.8.3	Windows on souls: personal and affective factors affecting code choice	72
3.9	Summary	74
4	<i>Wanem i stap yet: remaining matters of data collection and methodology</i>	76
4	Introduction	76
4.1	Sources of the data	76
4.2	Equipment	77
4.3	Speaker demographics	77
4.4	Transcription and coding of the corpus	79
4.5	Database management: the Goldsearch program	80
4.6	Non-linguistic factors	80

4.6.1	Marital status	81
4.6.2	Education	82
4.6.3	Origin of speakers	82
4.7	Summary	84
5	Verb inflection and subject pronouns	85
5	Introduction	85
5.1	Defining the ‘predicate marker’	85
5.2	<i>Mi mi</i> and <i>yu yu</i> : Subject focusing or subject agreement?	87
5.2.1	Review of previous analyses	90
5.2.2	Discourse role of subject referents as a diagnostic	92
5.2.3	<i>Mi</i> and <i>yu</i> are subject pronouns	95
5.3	Elision or absence: predicate marking with 1s/2s subjects	103
5.4	<i>I</i> still means ‘he’: a functional motivation for zero marking	106
5.5	Conclusion	108
6	<i>From wanem nao yumi no nidim wan jenis? Constraints on phonetically null subjects in Bislama</i>	110
6	Introduction	110
6.1	Generative accounts of phonetically null subjects	110
6.1.1	Split <i>pro</i> -drop systems	112
6.2	Pragmatic or discourse-based accounts	114
6.2.1	Reference tracking and algorithms of saliency	115
6.2.2	Switch- and echo-reference marking	117
6.2.3	Affective information expressed through pronoun choice	118
6.3	Establishing that Bislama is a <i>pro</i> -drop language	119
6.3.1	Fitting Bislama into the typology of null subject systems	120
6.4	Syntactic and semantic factors investigated	121
6.5	Discourse factors investigated	121
6.6	Social factors investigated	121
6.6.1	Origin of speaker	121
6.6.2	Education – language medium and level attained	122
6.6.3	Age of speaker: comments on the concept of ‘apparent time’	123
6.6.4	Sex of speaker: comments on the use of <i>sex qua sex</i> as a variable	125
6.6.5	Village versus urban residence	127
6.6.6	Social and attitudinal factors grouping Malo speakers	127
6.7	Defining the linguistic envelope of variation	128
6.7.1	Categorically null subjects	128
6.7.2	Optional subject-verb agreement in Bislama	129
6.7.3	Summary of tokens excluded from the corpus	130

6.8	Outline of method: statistical tools used	131
6.9	Results	132
6.10	Linguistic variables	134
6.10.1	Person of subject referent: an attempt to unify the analysis with other split systems	135
6.10.2	Information status of the referent	140
6.11	Social factors	142
6.11.1	Speaker sex	142
6.11.2	Origin and <i>lanwis</i>	143
6.11.3	Village and urban residence	145
6.11.4	Level of education	146
6.11.5	Age of speaker	148
6.11.6	Caregivers and <i>ol smol papa</i> : other possible models for children	151
6.11.7	Malo speakers' linguistic networks	155
6.12	Conclusion	158
7	Serial verbs in Bislama	160
7	Introduction	160
7.1	Defining serial verb constructions (SVCs)	160
7.1.1	Generative analyses of SVCs	162
7.1.2	Functional accounts of SVCs in Oceanic languages	163
7.2	Structural variants of SVCs	164
7.3	The structure of Bislama SVCs	165
7.3.1	Constraints on SVCs in Bislama	165
7.3.2	The difficulty of testing for unaccusativity in Bislama	165
7.3.3	Evidence of the structural dependence of the lower clause	166
7.3.4	Substrate parallels	168
7.4	Defining SVCs in terms of finiteness	174
7.4.1	Finiteness as a temporal relation	176
8	<i>Toktok Biaen; tumora blong wok ya</i>: conclusions and future directions	182
	<i>Appendix</i> : Glossary of Bislama terms	186
	<i>References</i>	189

Ol tankyu: Acknowledgments

Wan bigfala tangkyu i goaot long olgeta man mo woman we oli bin givhan long mi long Vanuatu. Long Vila: Ralph Regenvanu, John Lynch, Robert Early. Long Santo: Sharon Tabi, we hadwok blong hem yumi save luksave long evri pej blong buk ya. Tangkyu long George Hinge, Joyce Graham, Ben Sands, Emmie Tari, Rachel Mape, Sale Vira.

Ol wod blong mi oli no naf blong talemaot tangkyu long ol Natamambo. Mi no save putum nem blong ol man wanwan we oli bin givhan long mi, be sam nem oli hae tumas. Hemia: Makali Baniuri wetem famle blong hem; ol “woman nara aelan”: Jacqueline Jonas, Annemarie Winji, Lezbet Tom, Ngarinet Thomsen, Delin Jean. Afta ol Natamambo: Vira Leo, Litu wetem Marie Baniuri, Orry Seni, Rinette Seni, Sera John, Alona Vuti, Vui Toa. Nambawan tangkyu i go long Vanua, Loreen, Ian Baniuri wetem Sera Toa.

Thank you, too, to my dissertation committee – extraordinary scholars and wonderful individuals: Gillian Sankoff, William Labov, Howard Giles and Terry Crowley.

I also gratefully acknowledge the help of the Wenner-Gren Foundation, whose financial support made my fieldwork possible, the outstanding community of scholars at the University of Pennsylvania, and members of my extended family. This work is a revision of my 1997 University of Pennsylvania, Department of Linguistics dissertation. Thanks to the linguists at Cornell University, whose hospitality and intellectual company made those revisions possible.

The biggest thank you goes to my harshest critic and my staunchest supporter. Andrew Beach sharpened my wits and smoothed off my edges, and I’m very glad he stayed for the ride.

For my mother, Mary Morris Howard, who would have said it with greater insight, brevity and wit, and for my father, Hans Meyerhoff, who I wish had been here to talk it over with.

Long nem blong mama blong mi, Mary Morris Howard, we kleva mo fani blong hem i bitim blong mi. Mo papa blong mi, Hans Meyerhoff, we mi stap sore long hem yet.

List of abbreviations

The following abbreviations are used by the author in this work. Also included are abbreviations used by other authors cited in this work.

1	first person
2	second person
3	third person
3DU	third person dual
3NON-SG	third person non-singular
3SG	third person singular
ADV	nominal or single word adverbs (not included in the PP class)
AGR/Agr	agreement
ArgP	agreement phrase
A-movement	argument movement
AuxP	auxiliary phrase
Cb	backward looking centre
Cf	forward looking centre
CNJ	clausal (not nominal conjunctions)
COMP	complementisers
CONT	continuous
Cp	preferred centre
DIST.PAST	distant past
DU	dual
DC	discourse particles
DO	direct object
E	event time
ERG	ergative
EXCL	exclusive
F	focused objects (direct and oblique)
FA	focused adverbials
FS	focused subjects

FUT	future
HABIT	habitual
INCL	inclusive
INDEF	indefinite
INFL	inflectional
INST	instrument
IP	inflectional phrase
IRR	irrealis
LOC	location
NEG	negative
NegP	negation phrase
NP	noun phrase
OBL	oblique objects
PL	plural
POSS	possessive
PP	prepositional phrases
PRED.MKR	predicate marker
PREP	preposition
PRES	present
R	reference time
SG	singular
S	speech time
SBJ	subject
SPEC	specificity marker
SUBJ	subject
SVC	serial verb construction
TR	trial
TMA	tense-mood-aspect
TP	tense phrase
VP	verb phrase (including DO but not obliques)

Introduction	1
1.1. The book	1
1.2. The author	1
1.3. The title	1
1.4. The subtitle	1
1.5. The foreword	1
1.6. The preface	1
1.7. The acknowledgements	1
1.8. The list of abbreviations	1
1.9. The list of figures	1
1.10. The list of tables	1
1.11. The list of references	1
1.12. The list of footnotes	1
1.13. The list of appendices	1
1.14. The list of indexes	1
1.15. The list of glosses	1
1.16. The list of maps	1
1.17. The list of diagrams	1
1.18. The list of equations	1
1.19. The list of formulas	1
1.20. The list of symbols	1
1.21. The list of acronyms	1
1.22. The list of initialisms	1
1.23. The list of abbreviations	1
1.24. The list of abbreviations	1
1.25. The list of abbreviations	1
1.26. The list of abbreviations	1
1.27. The list of abbreviations	1
1.28. The list of abbreviations	1
1.29. The list of abbreviations	1
1.30. The list of abbreviations	1
1.31. The list of abbreviations	1
1.32. The list of abbreviations	1
1.33. The list of abbreviations	1
1.34. The list of abbreviations	1
1.35. The list of abbreviations	1
1.36. The list of abbreviations	1
1.37. The list of abbreviations	1
1.38. The list of abbreviations	1
1.39. The list of abbreviations	1
1.40. The list of abbreviations	1
1.41. The list of abbreviations	1
1.42. The list of abbreviations	1
1.43. The list of abbreviations	1
1.44. The list of abbreviations	1
1.45. The list of abbreviations	1
1.46. The list of abbreviations	1
1.47. The list of abbreviations	1
1.48. The list of abbreviations	1
1.49. The list of abbreviations	1
1.50. The list of abbreviations	1
1.51. The list of abbreviations	1
1.52. The list of abbreviations	1
1.53. The list of abbreviations	1
1.54. The list of abbreviations	1
1.55. The list of abbreviations	1
1.56. The list of abbreviations	1
1.57. The list of abbreviations	1
1.58. The list of abbreviations	1
1.59. The list of abbreviations	1
1.60. The list of abbreviations	1
1.61. The list of abbreviations	1
1.62. The list of abbreviations	1
1.63. The list of abbreviations	1
1.64. The list of abbreviations	1
1.65. The list of abbreviations	1
1.66. The list of abbreviations	1
1.67. The list of abbreviations	1
1.68. The list of abbreviations	1
1.69. The list of abbreviations	1
1.70. The list of abbreviations	1
1.71. The list of abbreviations	1
1.72. The list of abbreviations	1
1.73. The list of abbreviations	1
1.74. The list of abbreviations	1
1.75. The list of abbreviations	1
1.76. The list of abbreviations	1
1.77. The list of abbreviations	1
1.78. The list of abbreviations	1
1.79. The list of abbreviations	1
1.80. The list of abbreviations	1
1.81. The list of abbreviations	1
1.82. The list of abbreviations	1
1.83. The list of abbreviations	1
1.84. The list of abbreviations	1
1.85. The list of abbreviations	1
1.86. The list of abbreviations	1
1.87. The list of abbreviations	1
1.88. The list of abbreviations	1
1.89. The list of abbreviations	1
1.90. The list of abbreviations	1
1.91. The list of abbreviations	1
1.92. The list of abbreviations	1
1.93. The list of abbreviations	1
1.94. The list of abbreviations	1
1.95. The list of abbreviations	1
1.96. The list of abbreviations	1
1.97. The list of abbreviations	1
1.98. The list of abbreviations	1
1.99. The list of abbreviations	1
1.100. The list of abbreviations	1

1 *On the nature of linguistic variables*

1 Why the study of language variation is of interest

Let me claim that the study of variation is the soul of linguistics.

To do so is to place the work that follows squarely within one particular intellectual and methodological tradition of linguistics. Readers might expect that the generalisations to be made will be probabilistic rather than categorical. They might expect that the diachronic and the synchronic state of the language and its speakers will be afforded equal prominence. And finally, they would expect that the significance of different variants in a language will be explored as a function of both linguistic and non-linguistic factors. Linguistic factors alone may challenge our theories of language and the mind, and social factors alone may generate acute observations about the organisation and negotiation of personal and group factors. But only together do we actually approach an adequate theory or even an adequate description of language. If, as Aphra Behn wrote, “variety is the soul of pleasure”, then that would explain why, for many linguists, the study of language variation bestows the most pleasure and the most reward.

This work explores some of the relationships between language and society in northern Vanuatu, a nation in the south-west Pacific where multilingualism and multiculturalism are a long-standing norm, but where discourses of unity and nationhood are also important facts of modern Vanuatu life. Its focus is on Bislama, the English-lexified creole which serves as a lingua franca in Vanuatu and which is an integral part of this picture of multilingualism and national unity. The work starts from the premise that language is fundamentally a social construct, and proceeds to examine the social and linguistic constraints on one variable: the alternation between phonetically null and pronominal subjects. This structural variable is examined within the affective and instrumental frames of increased interpersonal and intergroup contact across linguistic boundaries. These are consequences of intermarriage, and internal migration to towns and are a significant fact of life in Vanuatu today.

Since linguistics is about describing and understanding the structure of different languages, variation is a puzzle that linguists cannot ignore. Crowley’s (1990) historical grammar of Bislama is an invaluable resource, and it is one to which I owe a great structural and motivational debt. He provides extensive information about the structure of Bislama, its relation to related languages in the region, and in addition he directs attention to the need for a more detailed study of the structure and use of Bislama in daily conversation, and the need

2 Chapter 1

for a more detailed social profile of the community of Bislama speakers. He points out the need for a more elaborated picture of how this data articulates than a historical grammar can attempt to provide.

It is worth remembering that structural and sociolinguistic description are, ultimately, in the same business. In both cases, the goal is to scrutinise apparent irregularities in a language in such a way that differences between tokens (the variation observed) can be demonstrated to be the result of a systematic, rule-governed grammar. Under this view, the main thing differentiating linguists is what they consider to be germane or interesting data. The work presented here would be classified as sociolinguistics because facts about speakers and facts about the organisation of the speech community are treated as being just as significant for an understanding of the human language faculty as facts about the purely linguistic environment of a variable are.

In addition to the synchronic tease that language variation provides, variation also stands at the heart of the study of diachronic change. It seems unlikely that there ever has been or ever will be a case of language change that is not preceded by a period in which there is some variability in norms. This must surely be true of even very rapid changes. For various reasons, it may be impossible to know retrospectively how the period of variability was structured or how long it lasted, but we reasonably assume that it was there.

In the case of historical changes, the uncertainty may be due to a lack of relevant or sufficient data, but even in the case of on-going changes, it may not be possible to know the entire history of a variable. Recent work by Labov (1994, 1996) has suggested where we might best look for the primary vectors of the spread of a phonological change, e.g. in certain speech styles (exchanges with peers), and in members of the speech community that match certain personality profiles and play certain roles in the community (Labov suggests individualists and people who are in some sense community activists are most likely to spearhead phonological change). It is not clear to what extent speakers matching these profiles should be considered the source of a change (its point of actuation), or the principal purveyors of a change (crucial for its systematic transmission in the larger speech community).

Of course, not all variation leads ineluctably to change. Some variation in the linguistic environment is just 'noise' (slips of the tongue, mishearings of another's utterance) and this has to be distinguished from systematic variation. But it has to also be recognised that linguistic noise does often get interpreted as systematic variation. Children, for instance, must be able to impose structure upon noise in the linguistic stream, interpreting and generalising from the input. Saffran et al. (1996) show that one of the ways they achieve this is by being highly sensitive to the overall probabilities of phonological combinations in the phonetic stream they are exposed to. This sensitivity, Saffran et al. suggest, is what allows the child to begin the task of parsing noise into words.

Other work by Labov (1989) and Sankoff (1994) has shown that children have a remarkable ability to match the probabilities of their parents' use of different variants (while often increasing the overall frequency of, for example, an innovative form). This ability to impose order and structure in novel ways does not abandon us as we grow older. Hudson and Newport (1999) have shown that adults retain an ability to attend to quite subtle variations in the overall probability of the linguistic input they are exposed to, and that adults can reliably apply the information they have extracted through passive exposure to subsequent tasks requiring grammaticality judgements. The chief difference may be that they exercise the ability less productively (especially outside of experimental settings). It is possible that adults'

continued ability to attend to on-going changes and adapt their grammar accordingly explain the patterns of diachronic probability matching Kroch (1989) observed, and which motivated his constant rate hypothesis for language change (This hypothesis is discussed in greater detail in §1.2).

Finally, it follows that if the study of language variation is of importance for understanding principles of language change, then there is a good chance that the study of language variation will show connections with other instances of social diffusion and social change. One of the more irresistible contributions that social psychology has made to the study of language use is the observation that distinctiveness and sameness are, paradoxically, both traits that people value. A great deal of interpersonal and intergroup behaviour has been expressed in terms of the tension and resolution of these goals. Consequently, a description and analysis of the different linguistic and non-linguistic means by which members of different communities balance these desires (whether consciously held or not) is fundamentally tied to an understanding of social and interpersonal relations.

One of the principal pleasures that the study of language variation offers is the ability to demonstrate that what superficially seemed to be irregularities and free alternations between forms are in fact the reflexes of more subtle regularities than had previously been imagined. For this reason, the study of linguistic variables in the social and structural context has traditionally been associated with ethnographic methods of data collection and analysis.

A good deal is said in sociolinguistics about the so-called observer's paradox (Labov 1972:209). Labov noted the methodological dilemma of anyone attempting to analyse natural language, i.e. the act of recording people's interactions necessarily changes the nature of those interactions. What the researcher seeks is, as a consequence of the act of research, impossible to get. A great deal less is said in linguistics about a more basic research dilemma in the social sciences, namely the problem that the formulation of hypotheses is essential towards efficient and cumulative progress in a field. Yet somehow hypotheses must be formed in such a specific way that the researcher is not biased towards only seeing and understanding what fits with the hypotheses.

This matter has been explored at greater length in other social sciences, principally anthropology, where progress and success depend on balancing two research tasks.¹ On the one hand, a community must be understood on its own terms in order for the ethnography of that community to make a meaningful contribution to the broader field of study. But on the other hand, the meaning of the contribution cannot be expressed without recourse to external parallels and contrasts. The balancing act between these tasks is not always easy. Geertz (1995) gives a personal account of his attempt to manage the tension between these goals in his fieldwork in rural Indonesia and urban Morocco. The value of an account like Geertz's lies in the frank discussion that reveals his own confusion and dissatisfaction with the way, over four decades, he has tried to make his work on Indonesian and Moroccan society relevant to others, without losing what is unique about his field work sites, what we might call their subjective reality.

In some ways, the observer's paradox follows straightforwardly from this more fundamental observational dilemma, and the centrality of the former in sociolinguistics reminds us of the field's roots in schools of ethnographic and anthropological research. This

¹ Though, in fact, the basic dilemma applies to the practice of all science, as Karl Popper pointed out.

is true even for the most quantitative approaches to sociolinguistics, a fact that is sometimes forgotten in the drive to reinvent the field.

So, having claimed that variation is the soul and pleasure of linguistics, what precisely is meant by 'variation'? One answer might be to say that the study of variation is about the study of linguistic variables. In fact, as fatuous as this answer seems, it raises some intriguing questions about the nature of variables, because either the term 'variable' or the concept is found in all branches of linguistics. But not everything called (or behaving like) a variable shares the synchronic and diachronic properties of the probabilistic, sociolinguistic variables mentioned above. In the remainder of this chapter, I will attempt to outline the salient differences between variables operating at different levels of linguistic analysis.

In doing so, I take up the question put by William Labov in an unpublished 1993 address, namely, *what can and cannot be a sociolinguistic variable?* That is, why do variables involved in some on-going change regularly demonstrate social stratification, while other variables show sensitivity to linguistic constraints, but not social constraints? Labov proposed a functional constraint on sociolinguistic variables and concluded that phonological changes are more often socially stratified than syntactic changes because the phonological variables can (in one sense) be more easily observed by speakers than syntactic ones can be.

I pose a slightly different question, the answer to which I believe must ultimately be related to the full answer still awaiting Labov's question. I propose a more basic question: *what distinguishes variables in quantitative sociolinguistics from other forms of linguistic variables?*

We know that in many cases, what superficially appears to be random variation is in fact constrained by linguistic factors or is correlated significantly enough with particular social factors that some kind of intimate or even causal relation may be inferred. In this way, this variation is distinguished from the highly unconstrained variation in performance of, for example, false starts and slips of the tongue.² At the other end of the spectrum, it is obviously to be distinguished from the kind of parametric variation examined in formal syntax and typological linguistics, which (ideally) applies uniformly within a language. For example, English is analysed as a head initial language; Japanese as head final. This typological variation accounts for the fact that English has prepositions and Japanese postpositions.

In other words, the current state of the art in the study of variation and change seems unable to answer the question I put in anything more than the following, purely quantitative, way: sociolinguistic variables are more constrained than performance errors, and less constrained than typological variation.

It seems to me that this is a fairly unsatisfactory and endocentric response to the question, one which the field of variationist studies is ready to outgrow. I propose that a more informative and principled basis for typologising variables exists. The response presented here reviews a wide range of variables, not just those involved in language change, and on the basis of this review proposes a more comprehensive typology of variables. It is suggested that this provides a fuller and more principled description of the differences and similarities between all forms of linguistic variable. It stresses the connections, rather than the divide, between the study of language variation in use and other branches of linguistics.

² Though as Hay et al. 1999 (among others) have shown, even slips of the tongue are not entirely random. Their form reflects strong linguistic constraints.

1.1 Towards a typology of variables: inherent and derived

I will suggest that there are essentially two types of linguistic variable: those which might be called inherent variables, and those which might be called derived variables. I will suggest that understanding what distinguishes inherent from derived variables contributes to an understanding of how the study of socially stratified variables fits into the larger enterprise of the study of the nature of language.

First, examples of what is meant by inherent and derived variables will be discussed, showing the relevance of these concepts in phonology and syntax. Then socially stratified variables, those integral to theories of the relationship between language variation and change, will be discussed. It will be seen that these variables are never inherent, and always derived. However, it will be argued that they are qualitatively different from derived variables in the other domains, for the simple reason that derived variables may involve obligatory or optional processes. Socially stratified variables can *a priori* never be obligatory; their social significance and utility is dependent on their optionality. This naturally implies that all inherent variables, and all obligatorily derived variables cannot acquire social significance, and will not exhibit patterns of social stratification. Finally, the place of discourse and lexical variables in the more general study of language change will be discussed. The question of why these variables are so seldom implicated in systemic change will be put and a first attempt at an answer will be forwarded.

1.1.1 Inherent and derived phonological variables

After 30 years of sociolinguistic fieldwork, we are more than familiar with the notion of socially stratified phonological variables. But for much longer than that, researchers have been endeavouring to describe and explain other kinds of phonological variables in structurally consistent ways.

Perhaps the most commonplace exemplars of variables in the phonological component of the grammar are those I would like to call inherent variables. Consider a phenomenon such as vowel (or consonant) harmony. Constant to all analyses of vowel harmony is the notion that underlyingly there is a segment which is underspecified for some phonological feature. This principle remains unchanged, regardless of whether we analyse harmony systems as the spreading of features across tiers (e.g. Ringen's 1988 analysis of Hungarian vowel harmony, and Shaw's 1991 analysis of consonant harmony) or as a ranking of optimal constraints (e.g. Embick 1994).

For example, the underlying representation of a vowel might include a specification for height, but lack a specification for front or back. The underspecified feature functions as a variable, i.e. its surface form is determined once a relationship is established between the featural variable and some other phoneme, often another vowel, and usually within the domain of the prosodic word.

For example, consider vowel harmony in Bislama. Intransitive predicates and nouns can be transformed into transitive predicates in Bislama by the addition of a transitive suffix. This process is quite productive and is quite general (with only a few exceptions). The suffix

6 Chapter 1

has the form *-Vm*, where *-V* indicates a vowel that harmonises with the stressed vowel in the stem. Examples are given in example 1.1.³

1.1.a.	/wip/ 'whip'	/wipim/ 'to beat up'
b.	/stil/ 'to steal; burgle'	/stilim/ 'to steal (something)'
c.	/brum/ 'broom' (n.)	/brumum/ 'to sweep (something)'
d.	/huk/ 'to fish'	/hukum/ 'to catch (a fish)'
e.	/sa'pɔt/ 'support'	/sa'pɔtɛm/ 'support (something)'
f.	/kat/ 'to cut'	/katɛm/ 'to cut (something)'

As can be seen the *-V* harmonises with the high vowels /i/ and /u/, and elsewhere surfaces as /ɛ/. Effectively what this means is that the vowel in the transitive suffix is a kind of variable; one of the *inherent* properties of the suffix vowel is that it is underspecified and this underspecification results in variability in its surface representation. In this case, its form happens to vary according to properties of another (prosodically salient) vowel in the stem. It is part of the very nature of this suffix that its form varies; there is no optionality. Some vowel must be realised in the suffix and a non-harmonic vowel is ungrammatical (**rusem*).

On the other hand, there are also instances in phonology where the variable constituent is not part of the underlying representation. Instead, it is a constituent that has been created in the course of the derivation from the underlying representation to the surface representation. Buckley's (1994) analysis of Kashaya stress assignment provides an example of this.

Buckley argues that extrametricality must be able to apply repeatedly in Kashaya, i.e. at successive stages in a derivation. This results in apparent violations of two principles, one, that extrametricality only applies at one level of the derivation, and two, that the domain of extrametricality is strictly peripheral. He shows that stress patterns in Kashaya can be most economically accounted for if two modifications to the theory are made. First, segments must be able to be marked as extrametrical at different levels of the derivation, and second, their extrametrical status must be able to persist through the derivation. So, for example, at an early stage in its derivation, a word may satisfy the structural requirements for extrametricality. But subsequent processes occurring later in a derivation, e.g. some types of affixes, and vowel lengthening, may again create a phonological form that satisfies the structural requirements to be considered extrametrical. If this happens, a word may have several domains of extrametricality laminated over earlier ones, and stress assignment may appear to 'skip' over non-peripheral syllables.

³ Crowley (1990:300-302) provides a detailed outline of the factors conditioning the form of the transitive allomorph in modern Bislama. Note that the forms here are represented phonemically, in speech unstressed /i/ and /u/ are lax [ɪ] and [ʊ].

Now, it seems reasonable to consider the property a syllable might have of counting or non-counting for the purposes of stress assignment to be a kind of variable (in some contexts the same phonological string counts for stress assignment; in other contexts, it does not). However, the variable of extrametricality in Kashaya is qualitatively different from the variable underlying the Bislama transitive suffix.

The difference lies in how the variables arise. In Bislama vowel harmony, the variable is there from inception, i.e. it is an *inherent* component of the morpheme; in the case of Kashaya extrametricality, the variable only emerges after a certain amount of analysis has taken place.⁴ For this reason, such a variable might be considered a *derived* variable.

Obviously, inherent and derived variables are similar because they both result in surface variation. The harmonic vowel varies between a surface realisation as a front or back vowel (and also high or non-high). The suffix vowel receives a full featural specification once it has combined with the stem. In Kashaya, main word stress varies depending on whether the left-most syllable is extrametrical. But extrametricality can be calculated at several stages of the derivation and can be bleb by other phonological processes.

But the two types of variable also differ as to whether the variability is a basic property of the constituent or not. A harmonic V slot is inherently variable, while a syllable derives its extrametricality.

1.2.2 Inherent and derived syntactic variables

Turning from phonology to syntax, we find that the very term 'variable' is widely used in this module of the grammar. The term is used to describe both inherent variables and derived variables. Quantifier bound anaphora are an obvious example of an inherent variable, and since any syntactic theory must account for the particular relationship between them and their antecedent, they provide a non-theory dependent place to start.

The facts about the distribution and interpretation of quantifier bound anaphora are very familiar, and will not be dwelt on at length. It is well-known that in an utterance like 1.2 the pronoun *her* has a variable reading.

1.2 *Every girl_i petted her_i cat.*

That is, because the anaphor, *her*, falls within the scope of a quantificational noun phrase (*every girl*) it is not co-indexed with any single referent (the way a third person singular pronoun is in its unmarked uses). Instead, its interpretation ranges over all the members of the set of girls denoted by *every girl*, that is to say, it varies, sometimes picking out *Girl A*, sometimes *Girl B* and so forth. Since this property of variable reference is a fundamental syntactic fact in such examples, it seems reasonable to define quantifier bound anaphora as instances of an *inherent* syntactic variable.

The question then arises as to whether there are derived variables in syntax as well. In the discussion that follows, I focus on derived variables in the (post-) P&P model of syntax (Chomsky 1995), but I believe that as a phenomenon they are not restricted to this syntactic formalism. Derived variables seem to be a necessary component of any syntactic theory – for example, even in a radically non-derivational framework such as Lexicase, some

⁴ And what is nice about the Kashaya example is that this variable keeps emerging, or being derived, as the word is built up.

dependencies between lexemes are derived, via so-called *tobi-isi* [lit. 'stepping stone'] features (Starosta forthcoming).

In generative theories of syntax, the class of variables is integral to the typology of empty categories; I want to explore the possibility that it is not merely an accident that syntactic variables share the same designation as the core elements of sociolinguistic inquiry. However, I will suggest that what differentiates them as classes of variables is that the variables of formal syntax are obligatorily derived and the sociolinguistic variables are optionally derived.

What is the nature of these syntactic variables? Conventional practice in generative theory is to distinguish the trace left behind after movement of a *wh*-phrase from the traces left behind by other movement. Only *wh*-traces are known as variables. Traces left by movement of core arguments to argument positions, so-called *A-movement* of passivisation or the raising of subjects of verbs like *seem*, or to a non-argument position, so-called *A'-movement* such as to a focus position, are not known as variables.

A clear terminological distinction between traces of *wh*-movement and other movement is very useful within these syntactic frameworks,⁵ but it does have the effect of masking similarities between the traces of all movement – similarities that may be useful in developing the broader typology of variables sought here.

The standard generative typology of empty categories focuses on the nature of the constituent that is moved, the hypothesised motivation for the movement, or the well-formedness conditions that license the movement taking place. These are all valid bases for comparison, and they highlight interesting differences between types of movement. If we consider *wh*-phrases, the constituent moved differs from other constituents in that the reference of a *wh*-phrase ranges variably over a set of possible candidates, rather like the bound anaphora discussed above. Indeed, *wh*-phrases can be bound by quantifiers in the same way that a pronoun can be, as shown in 1.3.

1.3 Which canary_i did each cat harass t_i?

The reference of *which canary* (the interpretive object of *harass*) ranges over a set determined by *each cat* in the same way that the pronoun *her* did in 1.2: my cat, Pink, harassed my canary. Alice's cat, Patches, harassed Rebecca's canary, etc. But whenever a *wh*-question is put, a set of possible answers is opened, some of which constitute true answers and some of which constitute false answers, e.g. 1.4. In that sense, all *wh*-phrases share the property of selecting their reference from a set of entities.

1.4a. When did Jeremy move out?

(A: *December*, true; *January (etc.)*, false.)

b. What colour clothes does Kirsty always wear?

(A: *Black and gold*, true; *Turquoise (etc.)*, false.)

When seen in this light, the reason these empty categories are known as variables is because of the nature of the *wh*-phrase. Being a variable is not a property inherent to the trace itself, rather it is a property that it possesses by virtue of being co-indexed with the constituent that is the head of the chain.

⁵ For example, they also may satisfy the requirement for government differently: traces of *A*-movement are properly head governed; variables are antecedent governed (Rizzi 1990).

But let me suggest another way of looking at traces, a way that highlights similarities between the traces that are left by different kinds of movement and is also, I believe, a way that draws attention to similarities between traces and the quantifier bound anaphora, which I have proposed are inherent syntactic variables.

Consider the quantifier bound anaphor in 1.2. We might say that the pronoun *her* marks a position in the clause where an interpretive dependency is opened. In 1.3 and 1.4b it is also clear that the *wh*-phrase satisfies a dependency that is opened elsewhere in the clause, i.e. it is interpreted as being the patient of the transitive verb. This is also true in cases of focus movement, e.g. 1.5a subject focusing and 1.5b object focusing:

1.5a. *Blood sausage_i, they think t_i is great for breakfast.*

b. *A ghost_i, she claims her house has t_i.*

Although it is more debatable whether or not adjuncts are necessarily raised from lower in the clause, if we consider the interpretation of adjuncts in complex clauses then it becomes clear that a *wh*-adjunct, illustrated in 1.4a and 1.6, can also be seen as satisfying a dependency opened somewhere lower (or further on) in the clause.

1.6 *When_i do you believe [Jeremy moved out t_i]?*

In all these cases, the dependency opened (through movement, or inherently in the case of quantifier bound anaphora) can only be satisfied, or interpreted, by the higher constituent that it is co-indexed with. How that co-indexation arises is very different in the case of quantifier bound anaphora, movement of arguments, and movement of adjuncts, but the end result is that they share a fundamental property. Traces and quantifier bound anaphors all mark a dependency whose interpretation is only satisfied by co-indexation with a higher constituent.

Within the typology of variables being built up here, a major difference between the bound anaphora and traces/variables arising from movement is that the latter are not part of the basic, underlying structure of the clause. Instead, they arise as a consequence of other hypothesised requirements of the clause (which might be stated as, e.g., the requirement that all nouns have one and only one case assigned to them, or that a finite clause have a subject, or that a feature marking a clause as interrogative be checked overtly in English). In other words, within a minimalist vocabulary, these are variables that are *derived* in the course of exhaustive feature checking; in other generative theories (such as Lexicase, and some forms of categorial grammar, e.g. O'Grady 1998), it could be said that the need to exhaust or satisfy all open lexical dependencies *derives* interpretive relations between constituents in a clause.

The discussion here has been couched in the terminology of generative syntax, but it might be noted that functional syntax also relies on a notion of derived variables. Even though functionalist paradigms posit nothing like movement within a clause, they do, nevertheless, *derive* the interpretation of pronominal anaphors. A significant difference between generative and functional approaches to anaphor resolution are that functionalists propose that the interpretation of anaphors is derived at the level of discourse rather than proposing mechanisms that resolve (some) anaphor interpretation clause internally (some discourse-based accounts of anaphor resolution are discussed in §6.2).

To sum up, then, it has been suggested that it is possible to distinguish between inherent and derived syntactic variables. The distinction between the two types of variables in syntax

parallels the distinction drawn in phonology where variables may be variables *sui generis*, and also may be the consequence of the application of other processes in that domain of the grammar.

1.1.3 Sociolinguistic variables: never inherent and always optional

It is well-known that there is nothing equivalent to inherent variables in sociolinguistic. The fact that a particular linguistic variable, or even a class of linguistic variables, has social significance, as demonstrated by a non-random distribution of its variants in non-linguistic domains, is not an inherent part of the meaning of that variable.

As it has often been pointed out (most recently by Labov 1993; Holmes 1997), there is nothing about, e.g. the English short *a* phoneme that means its realisation must vary over a range of phonetic forms, some low, some mid and some markedly raised (in US English). No inherent property of that phoneme requires that the raised variants be evaluated as regional or social or stylistic markers. Social meaning and phonetic quality are assumed to be independent factors, and the statistical tests undertaken in quantitative sociolinguistics actually require this assumption to be made.

Work on attitudes to language and subjective reactions to linguistic variables also crucially makes this assumption. Consider the alternate pronunciations of 'man' in regional varieties of United States English, and 'air' in New Zealand English. If a raised variant of the short *a* in 'man' is realised (i.e. [me^ən] or [mi^ən]), or if 'ear' and 'air' are both realised as [i^ə], and if these realisations trigger negative reactions in a listener, this is not because there is anything about the sound [i^ə] nor about the underlying phoneme itself that inherently requires an evaluation to be made. We can reject the notion that there might be mysterious iconic processes at work, since exactly the same sound has a neutral evaluation when it realises the final syllable of 'idea'. A major contribution of sociolinguistics to the study of language in general has been to demonstrate that often the non-linguistic significance of a variant is a striking analogue to the evaluations made of the prototypical users of that variant.

Since the social meaning of a phonetic variant depends on the underlying phoneme it is taken to be a realisation of, it follows that sociolinguistic variables are derived variables. The significance that [i^ə] as a realisation of the nucleus in 'man' has acquired is a consequence of it being perceived to be derived from the phoneme /æ/, and the additional perception of this derivation as being strongly correlated with certain speakers, styles or situations about which a listener has (independently formulated) opinions and evaluations.

But there is a crucial way in which this variable differs from the derived variables illustrated in the last two sections. Raising of the nucleus of 'man' or 'air' is not obligatory. (The variants are also not discrete, i.e. raising is a gradient property, and it is not at all clear that all and any raising need be evaluated socially. This point will be elaborated below.) Sociolinguistic variables are derived optionally, and I want to argue that it is this optionality that frees them to index non-linguistic factors, such as formal written style, or a socially coherent group in the speech community. Note that claiming they are optionally derived is not to claim that they are consciously derived. More frequent use of different variants in

different styles is undoubtedly often an automatic or sub-conscious process, especially when the variables are phonological.⁶

The claim that sociolinguistic variables are optionally derived variables is an alternate way of stating the sociolinguistic facts, namely that a single underlying form has multiple possible realisations within and among groups of speakers. But the reason for restating it in these terms is to make the connection with other kinds of linguistic variables more transparent.

It follows, therefore, that inherent variables, and variables resulting from obligatory linguistic derivations do not acquire social significance. This is obvious, since these variables are always assigned meaning in highly constrained and uniform ways. This being so, there is no way in which they could be socially or stylistically stratified.

As noted earlier, some variation is simply noise. However, just as the child learns to perceive patterns and structure in the noise it is exposed to,⁷ sociolinguistic variables presumably emerge from the noise of day-to-day communication. (The problem of determining precisely whose 'noise' is attended to and transformed into the basis for stable patterns is a question that is still far from being answered.) Naturally, this has implications for theories of language change and should also inform theories of grammaticisation.

Once speakers perceive variation in normal interactions to be non-random, communicative routines are available as candidates for reanalysis in the core grammar. Phonological variation in the frequency and extent of vowel raising, for instance, may result in a reanalysis of the perceptual boundaries of that phoneme (and perhaps the realisation of the whole phonemic system). This was the basis of Martinet's (1952) claim that perturbations in the phonetic realisation of a phoneme will be tolerated within an existing system so long as what he called the phoneme's margin of security is not disturbed. If the phonetic changes begin to encroach on this, the result will be reanalysis of the inventory, often resulting in a larger push or pull chain of phonological change.

Likewise, it has been shown that variation at the level of discourse may be reinterpreted as meaningful variation rather than noise, and as a consequence of this restructuring there may be quite marked effects on the syntax or morphology of a language. Discourse and lexical variables will be discussed in §1.3, where I will argue that these examples are to be distinguished from sociolinguistic variables because the variables in these cases are not derived.

1.2 Implications of this typology of variables

There are ample attestations in the sociolinguistics literature of phonological variables indexing non-linguistic factors. If the typology being forwarded here is correct, it obviously makes the prediction that when they are optionally derived, syntactic variables might acquire social significance too. In this section I return to the question posed in Labov's 1993 paper, why do some stochastic linguistic variables exhibit social stratification and not others, and will reconsider it in light of the analysis of variables being developed here.

⁶ Whether styles are considered to be a function of situational factors, interpersonal accommodating or self-monitoring does not change this point.

⁷ Perhaps even (re)analysing input quite lacking in structure as regularities. Sankoff and Brown (1976) discuss such a case, and see §1.3.1 below.

The literature on syntactic variation and change is by no means as rich as it is for phonological variation and change. This is partly because tokens of syntactic variables are not as abundant in natural speech as phonological variables are, and syntactic variables are therefore only amenable for study where the available corpora are very large indeed. This is more often the case with historical studies of variation and change than it is with synchronic studies, because the latter are usually based on smaller corpora of naturally occurring speech.

In the 1993 paper, Labov observed the asymmetry in the distribution of stochastic (probabilistic) variables and suggested that the generalisation to be made from this was that only structural variation that is observable (from speakers' point of view) may be socially stratified. He concluded that the reason syntactic variables seldom exhibit social stratification was because their structure is seldom transparent to language users.

Here, I have put the case slightly differently. I have observed that stochastic variables are distinct from other types of linguistic variable in that stochastic variables are always derived, and specifically, that they are always optionally derived. Based on this analysis, one might conclude that syntactic variables are seldom socially stratified because syntactic variables are seldom optionally derived.

Under this hypothesis, there are two criteria that syntactic variables might fail to satisfy. First, they might not be optional. This surely cannot be the explanation. It is seldom the case, even when the factors constraining a syntactic variable are broken down, that we are able to reduce the variation to the application of an obligatory rule. For instance, Oliveira e Silva (1982, discussed in Kroch 1989) considers the variation between possessive NPs with and without determiners in Portuguese over approximately 200 years. During this time, possessives consisting of the possessive pronouns and the head noun increasingly alternate with forms that are introduced by a determiner. That is, forms like 1.7 variably give way to forms like 1.8.

1.7 *Maria conhece meu irmão.*
Maria knows my brother.

1.8 *Maria conhece o meu irmão.*
Maria knows the my brother.
'Maria knows my brother.' [Kroch 1989, example 7a.]

As it happens, kinship terms such as 'my brother' very strongly disfavour the new form of possessive, i.e. they seldom occur with the determiner. But even by breaking down the variation according to NP type and isolating kinship terms, it would be impossible to say that there was no optionality. Around 10% of the time, even kinship terms occur with a determiner in a possessive NP.

The second criterion that syntactic variables could fail to satisfy would be their nature. They could fail to be derived variables. This is, in fact, essentially the position argued by Kroch (1989). After examining the relative frequency of conservative and innovative variants in a number of cases of syntactic change over two to three centuries, Kroch concludes that there is a constancy in the rate of change which is compatible with the idea that speakers are consistently alternating between quite different grammars during the entire history of a change. That is to say, his case amounts to the claim that variants of a syntactic variable are not derived from a single underlying form (as phonological variables are derived from a single underlyingly phoneme). Instead, two competing variants represent the input of two completely distinct grammatical systems.

As persuasively as this argument is made in Kroch's paper, there are two reasons we should be cautious about accepting it without reservation. The first reason relates to the limits of our information about historical variation. The second relates to what we are prepared to assume triggers a switch between different grammars.

Although historical corpora provide us with the sheer quantity of data that is needed in order to track syntactic variation, they generally lack the kind of detail about the writers contributing to it that would enable us to determine whether or not there really is a lack of social variation.

Consider Pintzuk's (1995) analysis of the loss of V2 in English, or Santorini's (1993) analysis of the shift from INFL-final to INFL-medial in Yiddish. Their work suggests that at some point there was variation within a speech community between two quite different underlying representations of tensed clauses. Unfortunately, we do not have the data that would tell us whether the people using more of the innovative INFL-medial clause structure were the 'burnouts' (Eckert 1989) of the 16th century Yiddish-speaking stetl. Nor do we know whether they were assimilationists with the most frequent social contacts outside the Yiddish-speaking community's networks (similar to the leaders of language change discussed in Gal 1979; Milroy 1980). It certainly wouldn't surprise us if ethnographic material indicating that they fulfilled these roles in the speech community suddenly came to light. Nor, given what we know about language variation at this point in time, would we be surprised if we found out that the association between these social roles and the variation had existed entirely below the level of consciousness and manifested itself only as a statistical correlation. But the fact is, ethnographic studies being thin on the ground, this information is entirely lacking in these cases.

In fact, we should be careful about assuming too readily that syntactic variables do not show social stratification. Nevalainen and Raumolin-Brunberg, eds (1996) and Nevalainen (1998) show that some of the syntactic and morphosyntactic changes Early Modern English underwent (subject-verb agreement and negative concord) were subject to class, sex, and regional stratification. Since this is about the limit to the non-linguistic information that can be derived from their historical corpora, we cannot afford to be too picky about questions such as whether such groupings of the corpus contributors are essentialistic. We may believe that there was probably a good deal more individual variation than these generalisations suggest, and we may want to be careful not to attribute deterministic qualities to correlations with sex, class and region of origin, but we also cannot ignore these results. They need to be taken as serious indications that social indexing is possible with (morpho)syntactic variables.

Synchronic evidence, too, suggests that such indexing is possible. Kroch and Small (1978) also show that variation in the use of complementisers and in the placement of verb particles in written and spoken modern English suggests that these variables are indices of stylistic factors.

Although data on intraspeaker syntactic variation is limited, the data we do have also gives pause for thought. Rickford and McNair-Knox (1994) compare the frequency of AAVE features in the speech of one woman, Foxy Boston, with different addressees in different conversations, and showed statistically significant differences in the frequency of some syntactic variables (absence of 3SG present verb inflection, absence of copula, invariant *be* as a marker of habitual aspect). Rickford and McNair-Knox show that features of both AAVE and standard American English are always present to a greater or lesser extent in Foxy Boston's speech, but that Foxy Boston adjusts (whether consciously or unconsciously) the relative frequency with which they occur depending on who she is talking to.

One might object, however, that AAVE and standard American English clearly do have different grammars, and therefore that the data from Foxy Boston is not a counter-instance to Kroch’s dual grammar hypothesis. However, in work on variation in Bislama (Meyerhoff forthcoming b), it has been found that even a variable which shows no evidence whatsoever of being socially stratified at the level of the speech community may nevertheless be constrained by non-linguistic factors at a very local level. The overall frequency of phonetically null subjects and overt pronominal subjects (the subject of detailed examination in Chapter 6 of this work) varies depending on who the speaker is talking to. So, for instance, an elderly speaker of Bislama recorded telling the same story to his family and then retelling the story six months later to me alone showed a significant decrease in the frequency of phonetically null subjects. This is shown in Table 1.1.

Table 1.1: Difference in frequency of null subjects in Bislama depending on addressee (from Meyerhoff forthcoming b).

Addressee(s)					
Extended family		MM only		Total	
\emptyset subject	% all clauses	\emptyset subject	% all clauses	\emptyset subject	% all clauses
N = 50	71 %	N = 40	62 %	N = 90	67 %

Some conceptual problems are raised if we attempt to assimilate the data on intraindividual variation to the model of two grammars that Kroch proposed as part of the constant rate hypothesis. The variation Kroch and Santorini and Pintzuk examined is linguistically constrained. In these cases, where all aspects of the variation (conditioning factors and variants) is system internal, it is perhaps plausible to propose a theory where a specific linguistic environment triggers the speaker’s use of a specific grammar. The task then becomes to specify the lines along which these grammars differ for speakers, preferably in ways that contribute to a larger understanding of the organisation of the grammar of natural languages. Indeed, this is the focus of much work in historical syntax today.

In short, it might well be the case that when syntactic variation is entirely system-internal it is correct to say that syntactic variables are not derived, and for this reason they fail to be socially indexed.

But cases of socially stratified syntactic variation, and particularly cases of stylistic or intraindividual variation, are qualitatively unlike the examples of syntactic change that appear to be only linguistically constrained. If we were to adopt Kroch’s analysis for the speaker in Table 1.1, we would be proposing that entirely different grammars are triggered in a speaker by a change of situation, addressee or goals and intentions. The number of potential grammars any one speaker might possess immediately explodes as a function of the infinite number of contexts that she might find herself in.

This creates an unwieldy descriptive situation, and an impossible theoretical one. It is precisely the situation that the notion of variable rules was intended to provide a principled alternative to. Positing an infinite number of grammars in every speaker precludes the ability to generalise, which is the principal goal of linguistics. More specifically, it jeopardises the ability to demonstrate the underlying systematicity of variation and to directly address the

evaluative judgements that are made of non-standard varieties, which is the principal goal of sociolinguistics.

Returning to the initial question in this section, it seems that the definitive answer as to why phonological variables are more frequently socially stratified than syntactic ones are remains open for further research. Labov's hypothesis that the extent to which a variable is a surface phenomenon remains a possible explanation. It has been shown, however, that the lack of social stratification in syntactic variables might have been overstated, possibly because of limitations to the corpora. It has been suggested that although some syntactic variables may not be derived variables (and in fact may not be true variables at all in the sense that phonological variables are), Kroch's hypothesis that syntactic variation represents multiple underlying grammars has some undesirable implications, particularly for those cases where there is evidence of intraindividual variation in syntax.

A further factor relevant to the nature of syntactic variables has also been ignored up till now. This is the fact that although phonological variables are continuous, non-discrete entities, it is generally assumed that syntactic variables are discrete (e.g. a possessive NP has a determiner or does not have a determiner; a subject is phonetically null or it is an overt pronoun). While it is true that syntactic variants surface as discrete options (unlike phonological variants which surface as a range along a phonetic continuum), it might be worth exploring the extent to which this surface fact necessarily reflects how grammars are organised underlyingly. Having for some time now worked on syntactic reanalysis and grammatical shift in Bislama, a language that many people speak as a first language (i.e. not their sole first language) and which is spoken against a backdrop of continued extensive multilingualism, the notion of what one might call a compromise or variable grammar seems an extremely pragmatic option for speakers to exercise.

Discussion of the implications of this typology of variables will have to leave open, therefore, the question of whether some types of syntactic variables are derived and some are not, and what the theoretical basis for this distinction might be. It also leaves open for further research the possibility that there is an interaction between the derived status of a variable and the cognitive constraints proposed by Labov or other factors in the ordering of information (such as have been argued for by DuBois 1987, §1.3.1).

1.3 Optional but underived variables

Up to this point nothing has been said about lexical variation and variation in discourse patterns or conversational routines. These variables do, of course, operate at levels of the grammar that speakers are highly aware of and are frequently the core of linguistic stereotypes, used to satirise or imitate a particular group of speakers.⁸ These variables are clearly optional, rather than inherent, as a look at just a few will show.

Schieffelin (1984) discusses the constraints on the use of two referentially identical terms in Kaluli. A sibling may refer to a younger brother either as *nao*, lit. 'my brother' or as *adε* – a socially more restricted term. It is used reciprocally during childhood, prototypically between an older sister and a younger brother. The *adε* relationship is intimately related to the socialisation of caring routines and might be considered a positive politeness strategy (Brown

⁸ This need not be restricted to outgroup satire. Schilling-Estes (1998) discusses cases of self-satire.

& Levinson 1987). Although there are some constraints on the use of *ade* (*ade* can only be used in some speech acts – it requires an appeal or request and the invocation of pity), as with all politeness routines, the decision to engage in that speech act and to use *ade* is entirely optional. That is, even though orders cannot be given using *ade*, requests need not be made with it. An older sister chooses to use *ade* (and the whiny tone of voice that goes with it) in order to achieve certain interpersonal effects.

Likewise, the negative politeness routines associated with making requests in English are optional, and there is intraspeaker variation in the way requests are phrased. The constraints include conventions relating to the interlocutor, the place of utterance and the desired effect. This means that a modal request form, such as ‘Could I have a word with you for a second?’, will be preferred over a bald imperative request, such as ‘Listen to me’, depending on the pre-existing social relationship between the speaker and the addressee, or the relationship that the speaker wants to create between them.

Now, if these forms are optional, and since it is clear that they are socially marked, how do they fit into the schema outlined here? We are not accustomed to seeing politeness strategies or the alternations between lexical items based on the formality of a situation stratified in the manner of more canonical sociolinguistic variables. Nor are we accustomed to seeing them lead language change. The reason for this, I believe, is that they lack one crucial property. Clearly, they are not part of the core grammar and cannot be considered inherent, but even though these variables are optional, they are not generally derived.⁹

It may be inappropriate to use one variant in a particular place or to a particular addressee, but there is no underlying constituent that requires a speaker to insert a marker indexing addressee, context etc. Another way of looking at it would be to say that although discourse and lexical variables are ubiquitous markers of interpersonal and intergroup boundaries, there are no routines or lexical items that derive or inherently require an index of personal or group factors.

Bell (1984) makes the strong claim that variables that show stratification in different speech styles only if they show social stratification. Since he states the relationship in causal terms, this amounts to a claim that style variables are, in our terms, derived. But this claim is not supported empirically by the distribution of lexical and discourse variables. As Labov (1993) has pointed out, these may indeed show fine stylistic stratification (any native speaker of English could easily generate any number of stylistic variants for the request given above), but they do not show evidence of equally fine-tuned social stratification.

⁹ A possible derived politeness variable might be found in languages where formality, distance and other social factors are consistently marked in pronouns and nouns (and on verbs in, e.g., Japanese). An interesting investigation might be to examine the extent to which derived variables are implicated in language change, the account developed here would predict that they are. In general, though, we do not at present have anywhere near the evidence that would be required in order to argue that politeness markers, for example, or slang words are derived from some inherently underspecified semantic or psychological constituent, nor what the nature of that constituent might be.

1.3.1 Interactions between discourse and the grammar

This is not to say that elements of the discourse are never linked to changes in grammar. Indeed, in some cases the increasing frequency of certain discourse patterns is a strong candidate as the causal factor in the diachronic reanalysis of syntactic structure.

Sankoff and Brown (1976) provided evidence suggesting that the frequency of certain patterns or routines in the discourse of Tok Pisin was reanalysed by speakers as a syntactic innovation. As a result, a novel means of marking relative clause boundaries was incorporated into the syntax of Tok Pisin. DuBois (1987) showed that an examination of the distribution of given and new information in Sacapultec Mayan discourse strongly suggests that the synchronic ergative-absolutive case marking system is a grammaticisation of preferences for the ordering of given and new information in speech. Similarly, it is possible that the very common cross-linguistic reanalysis of verbs of speaking (or other reported discourse) as complementisers may be the consequence of speakers striving to mark evidentiality (Güldemann & von Roncador, eds, forthcoming). In such a case, too, it is clear that interpersonal and discourse factors may shape syntactic change.

In the typology of variables presented here, it has been suggested that the variables involved most intimately in language change are only those variables that are both optional and derived. Although discourse and lexical variables are optional, they are not derived. Consequently, although it cannot be ruled out that speakers may overgeneralise or reanalyse optional lexical and discourse variables, they are not so likely to be involved in syntactic change. However, as mentioned above, the way in which cognitive or interpersonal factors, such as discourse ordering and politeness strategies, and structural constraints on the nature of variables interact in the instantiation and progress of language change remains a profitable avenue for future research.

1.4 Conclusion

The significance of variation and the typology of variables that have been discussed in this chapter will form the backdrop for a detailed examination of variation in conversational Bislama. In Chapters 2–4, the social dynamics and history of the fieldwork site are discussed and the language histories and social and linguistic attitudes of the specific speakers recorded for this study will also be sketched.

This in turn readies the ground for a detailed examination of variation in the Bislama verb phrase. The interrelationships between and influence of different levels of the grammar on each other emerge in the course of this study. In Chapter 5, data is presented that indicates Bislama has developed a full system of subject-verb agreement. Though this system derived historically from English pronouns, the forms have been reanalysed in Bislama and are now part of the verb morphology. Chapter 7 will show that these forms have grammaticised sufficiently that there is a genuine sense in which we can speak of there being a default form of subject-verb agreement in Bislama.

However, in Chapter 6 we will see that the distribution of phonetically null subjects in finite clauses suggests that the extent to which different agreement markers remain referentially transparent appears to have fed the development of a restricted system of *pro*-drop, where third person singular and plural pronouns can be phonetically null; first and second person subjects, on the other hand, tend to be realised as pronouns. This is shown to

correlate directly with the extent to which the agreement morphology associated with these referents is overt and distinct in singular and plural. Children appear to be sensitive to the distribution of pronouns and null subjects in their parents' and immediate caregivers' speech, and there is evidence that what is a general tendency in the community is becoming fully grammaticised among speakers under 12 years.

2 Lukluk ples: *social and linguistic overview*

2 Introduction

I carried out my fieldwork and data collection in two places. One was the urban centre in northern Vanuatu known as Santo, and the other was in villages on Malo island, a short boat ride away from Santo. Both fall within the 'central zone' of the archipelago, where 64% of the population of the islands lives and 77% of the copra is produced, i.e. the zone of greatest economic wealth and best served by various means of transport (Bonnemaïson 1986:485-486, 489). In this chapter, I will give some of the social and historical background to the region, finishing with a description of the social life of both the village and urban communities as they were in 1994–95 when I tape recorded spoken Bislama there.

Santo township is one of the two municipalities in Vanuatu and the second largest population centre (after the capital Port Vila), with a largely Ni-Vanuatu¹ population of approximately 8,000 (calculated on the basis of the 1989 census, population 6,965 with population growth rate of 2.17% per annum according to the Central Intelligence Agency (1997). If anything, it should be noted that this probably underestimates the Santo population as these growth statistics are for natural population increase only and do not take into account increasing urbanisation). I selected it as the locus of my fieldwork for two reasons.

The first was that having visited Port Vila in 1991, I was aware of how easy it is for Europeans to get around that town with only English or French. I wanted to work in a town that was less dependent on tourism and had a smaller expatriate population, and where Bislama would be a more important medium for all residents. As I will discuss later, Santo satisfied these requirements. The second reason for deciding to work in the Santo region was that I had a personal invitation there. My research plans had been encouraged and facilitated greatly by a Ni-Vanuatu woman living on Malo island who had been at university with me in New Zealand.

Santo is located on the sheltered south-eastern side of the island Espiritu Santo (also usually known simply as Santo – its naming will be discussed in §2.1.2) (Figure 2.1). The township is sometimes also referred to as Luganville or Kanal – Luganville is the French name for the township and is still used for administrative purposes and by some Francophone

¹ Ni-Vanuatu is the adjective derived from Vanuatu. A glossary of the Bislama terms used in this work can be found as Appendix A.

residents. Kanal is the Bislama name derived from Canal du Segond, the French name for the deep water channel that runs between Santo and Aore islands and which creates the sheltered waters that form the Santo harbour (Figure 2.2). Bislama speakers usually use Kanal or Santo to refer to the township. To avoid confusion, I will refer to the island as Espiritu Santo, and the township as Santo. As well, the term South Santo is used to refer to the region on the southern coast of Espiritu Santo, west of Santo town and immediately north of Malo.

Thus, in the fluidity and variation of its name alone, Santo evokes the multilingualism and multiculturalism of the area's history.

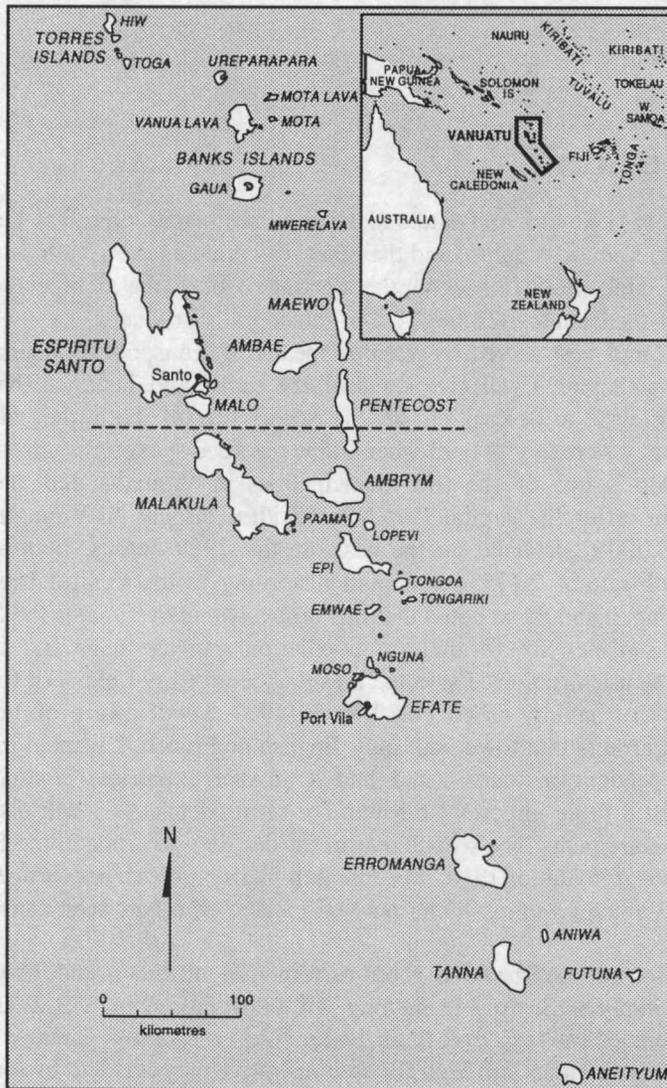


Figure 2.1: Major islands of Vanuatu, showing towns of Santo and Port Vila, with SW Pacific insert.

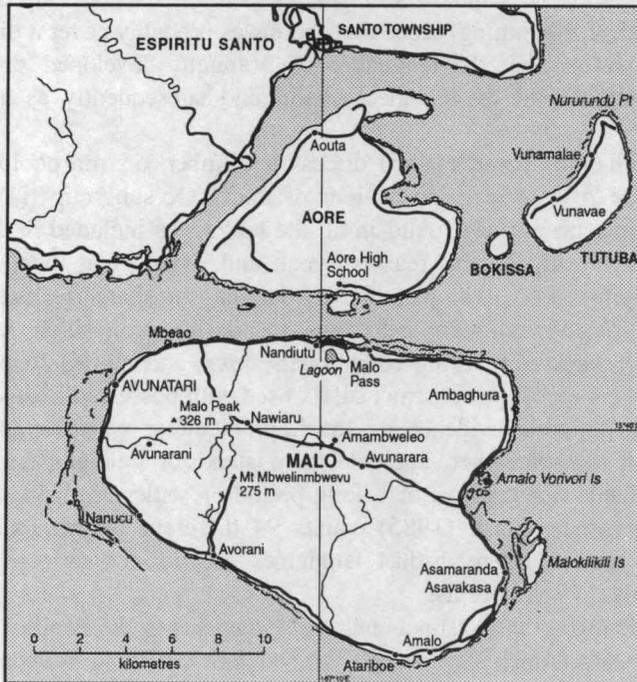


Figure 2.2: Map of Malo Island, and other islands off-shore of Santo township, showing major villages.

2.1 Brief history focusing on the region

2.1.1 First settlement to European contact ²

Vanuatu has been peopled for at least 3000 years (Nile & Clerk 1996; Encyclopaedia Britannica 1997). Archaeological evidence from the northern islands indicates that the area was settled by members of the Lapita culture from c.1500–1200 BCE (i.e. before current era). This culture is identified through and with its etched and impressed fired earthenware and obsidian trading pieces. Traces of this culture have been found from the Bismarck Archipelago in New Guinea to Fiji, Samoa and Tonga with Vanuatu obviously an intermediate point. (Pottery making is no longer widely practised in Vanuatu, in fact, only one region is known nowadays for its pottery tradition, Wusi on the west coast of Espiritu Santo.)

² Bonnemaision (1986, Chapters 1, 12-15) provides a much more detailed, and highly readable, portrait of the history of Vanuatu since European contact. What follows in this chapter hardly compares with Bonnemaision, and interested readers are urged to seek his work out directly.

The peoples associated with the Lapita culture also spread the Central/Eastern Oceanic (CEOc) language that developed into the 105 languages (Tryon 1976) spoken in Vanuatu today. All languages in the Vanuatu archipelago seem to have developed from a common ancestor. Ross et al. (forthcoming) argue that there was probably never a distinct Proto CEOc language *per se*, rather that the languages of Vanuatu developed directly from Proto Oceanic itself, first as part of a dialect chain, and subsequently as a series of linked languages.

However, Lynch and Tryon (1985) discuss a number of morphological and lexical features that can be “regarded as innovations of the CEOc subgroup” (1985:34). Although these features are not necessarily found in all the languages included in the subgroup, they are unique to this set and for this reason Lynch and Tryon treat them as a discrete and coherent subgrouping. For example, CEOc languages share reflexes of the CEOc prepositions **po[ŋ]koto/poto[ŋ]ko* ‘near’ and the dative preposition **muni*. They share reflexes of the verb particles marking conditionals, **ŋke*, and marking ‘spontaneous’, **tapa*, and they all have a reflex of the construct suffix used with possessed nouns, **-[k]i*.

Aside from the tangible evidence of the Lapita culture which suggests the heart of Vanuatu settlement was somewhere in the northern islands of Vanuatu, the linguistic diversity of the northern islands is suggestive of a long period of settlement with subsequent cultural and linguistic divergence. Clark (1985) counts 94 different languages in the North and Central Vanuatu region, where distinct languages are defined as those sharing $\leq 70\%$ cognates on Tryon’s (1976) word list.

Clark (1985) identifies three distinct chains of languages in the Espiritu Santo region, one in south Espiritu Santo (which includes all the southern off-lying islands including Malo), one a west Espiritu Santo chain, and a series of ‘clusters’ of ‘relatively isolated’ languages in east Espiritu Santo (1985:201). These chains do not overlap with other language chains in the northern and central islands. By far the largest of the Espiritu Santo chains is the southern chain, which can be further subdivided into four subchains: a Western chain (made up of Wusi, Malmariv, Lametin, Navut, Roria), a South-Western chain (Akei, Fortsenal, Wailapa, Araki, Tangoa), a South Central chain (Morouas, Amblong, Narango) and a south-eastern chain (Mavea, Tutuba, Aore, Tambotalo, Malo).

Thus, the picture we derive of south and west Espiritu Santo from this linguistic profile is one of probable lengthy settlement, perhaps involving repeated waves of settlement creating the linguistic diversity found in the area today (Ross et al. forthcoming). On the other hand, the east Espiritu Santo chain, which is a completely separate subgroup made up of languages that are less closely related to the rest of the Espiritu Santo languages than any of the other North-Central Vanuatu languages are (Clark 1985:202, 207), suggests a separate period of settlement for eastern Espiritu Santo during the dispersal of CEOc peoples and greater subsequent isolation.

Clark (1985) also notes that cultural boundaries between the northern region and the central islands converge with linguistic ones (this cultural and linguistic boundary is marked with a dotted line in Figure 2.1). Just as there is a clear linguistic boundary between the languages of Espiritu Santo (including Malo), Ambae and Maewo and the northern third of Pentecost, and Malakula and everywhere south of the Raga-speaking region in Pentecost to the south based on the levels of shared vocabulary and morphosyntactic structure, there are two other significant social boundaries at approximately the same point. Clark tells us that the linguistic boundary occurs roughly where the northern, traditionally matrilineal, dual

organisation social structures give way to the southern societies of patrilineal social organisation. In the north, the graded society is known by some reflex of the (Proto NCV) term **subwe*, while in the south the grading system is known by some reflex of the term **maqi*.

Rubinstein's (1981) picture of Malo island indicates that it should perhaps be considered a transitional region. The graded society is definitely affiliated with others of the northern region linguistically (Tamambo *sumbue*) and culturally (it is a closed, essentially secret society). However, according to Rubinstein, Malo is (and was traditionally 1981:142) a patrilineal society. In his 1978 thesis he discusses matrilineal kinship ties that people on Malo continue to keep track of, but these are of secondary importance to patrilineal rights in matters such as land claims. Allen (1981:19) suggests that the secret men's societies generally developed in Vanuatu as a response to dominant matrilineal systems, thus the situation on Malo today may indicate that at some time in the past its social structure was closer to the northern patterns than it is today.

In addition to this cultural boundary, there is a musicological boundary at a similar point. In the southern regions of Vanuatu the wooden slit gongs (hollowed trunks used as drums with a slit opening for resonance) are placed upright, while in the north slit gongs are placed horizontally (and there is a transition zone where the gongs are set at a 45° angle).

These ethnolinguistic facts will have some practical application in my analysis of variation in Bislama. When I discuss variation in the Santo-Malo speech communities in the following chapters, I will consider the influence a speaker's place of origin may have on the variation being investigated. When it is necessary to combine speakers into larger groups (if, for example, there is insufficient data to make statistically sound comparisons), I will use the linguistic and ethnographic subgroupings discussed by Clark, Lynch and Tryon, and Ross et al. as the basis for forming larger groups of speakers from linguistically similar areas.

2.1.2 *European contact*

European contact with Espiritu Santo is believed to have occurred for the first time on 30 April 1606, when Pedro Fernández de Quirós, a Portuguese navigator sailing under a Spanish flag, sent a boat ashore on Sakau island off the north east coast. De Quirós was in search of the great southern continent, the Terra Australis incognita, that Europeans at the time presumed must exist to balance the weight of the northern continents. On sighting the large land mass of Espiritu Santo he rather hopefully assumed that he had made the discovery he was after.

His contacts with the people living on the northern edges of Espiritu Santo and in the area now known as Big Bay were brief and uncordial. This appears to have been largely due to the Europeans' own consistently distrustful and aggressive behaviour. Representatives of the local population of several thousand (Kelly 1966:229) first approached the ship in Big Bay in canoes on 1 May, displaying 'bright red feathers on a stick' (Kelly 1966:204). Red colouring may have indicated that the feathers were in some way *tapu* ('sacred'), and Fray Martín de Munilla (a Franciscan travelling with De Quirós) records that the locals were upset when the Spaniards reached out to pick them up. De Quirós' crew fired a pre-emptive 'few volleys with their arquebuses' (Kelly 1966:204). This *modus operandi* continued in subsequent encounters on 9 May, 14 May and 17 May, with the Spaniards firing on various

(often unarmed) gatherings. From the 19th onwards, word seems to have got round and local Ni-Vanuatu began to actively fight back.

When De Quirós left Big Bay on 12 June, he was certain that he had in fact begun to chart a continent (Markham 1904:271, 478), despite the information he elicited from one Ni-Vanuatu informant identified as Tumay, who tried to explain that yes, the island was very large, but it was matched by a similarly-sized island further south, which De Quirós records as ‘Manicolo’ (Markham 1904:488). However, De Quirós was unable to prove he had found a continent since extensive exploration of the coast was prevented by unfavourable winds and (by then) active hostility on shore. Munilla, the Franciscan, and Juan de Iturbe, the voyage’s overseer and accountant, were less convinced they had been moored off a continent (Kelly 1966:91). Nonetheless, the name De Quirós gave the island, *La Austrialia del Espíritu Santo*, continues in its abbreviated form of *Espíritu Santo* as the official cartographic name for the island, though most people shorten it even further to *Santo*.

He also left behind the basis for some local myth-making. During conversation after dinner one night, I listened to a young man telling his younger cousin about a group of remote villages in northern *Espíritu Santo* where a large number of the inhabitants have straight, light-coloured hair. His explanation was that when De Quirós came, some of the sailors had stayed behind or had relationships with Ni-Vanuatu women and their descendants lived in these villages. There are no records of any of De Quirós’ crew staying for even a little R&R in Vanuatu, much less jumping ship entirely, and even if they had, it is unlikely they were fair in skin or hair, so other, more recent European contact seems a more likely source of this gene pool. In addition, the high rate of albinism in the Banks island group (further north of *Espíritu Santo*) would also seem to be a possible source for the light-colouring, rather than a handful of (presumably non-blonde) Iberian sailors three hundred years ago. Nonetheless, whether or not this story accurately represents the genetic source of the light-coloured people in northern *Espíritu Santo*, it performs a useful socio-historical function. It is a medium for passing on some key facts of the history of the region and indicates that the initial contact with De Quirós in *Espíritu Santo* continues to be perceived as a significant moment for at least some residents of the northern islands.

De Quirós’ trip was followed in 1769 by Louis-Antoine de Bougainville’s voyage. De Bougainville spent most of his time around the north eastern islands of Pentecost, Maewo and Ambae, and he, too, had some less than cordial encounters with locals. He appears to have experienced some difficulties on Pentecost and Ambae (“Ce départ déranga sans doute le projet des insulaires qui n’avaient pas encore tout disposé nous attaquer...une décharge mieux nourrie ralentit aussitôt leur attaque”³ 1992 [1771]:181-182, 185). However, de Bougainville did explore enough to realise that *Espíritu Santo* was in fact only a large island and was shocked by De Quirós’ assertions of a continent:

³ “Our departure undoubtedly upset the plans of the islanders who no longer felt disinclined to attack us...a more fortified volley deflected their attack quickly enough” [my translation].

Il est bien singulier que, précisément par la même latitude et la même longitude où Quiros place sa grande baie... nous ayons trouvé un passage de largeur égale à celle qu'il donne à l'ouverture de sa baie. Le navigateur espagnol a-t-il mal vu?⁴

(de Bougainville 1992 [1771]:187-188)

De Bougainville sailed around the southern coast of Espiritu Santo without charting the area, and renamed the islands "l'archipel des grandes Cyclades" (de Bougainville 1992 [1771]:187).

James Cook had intended to visit and chart De Quirós' La Austrialia on his remarkable voyage in the *Endeavour* but was stymied by losing time in repairing damage to the ship sustained off the Great Barrier Reef (Beaglehole 1974:259-260). He made up for the omission on his first voyage with the *Resolution* in 1772-75.

In a tremendous feat of marine cartography, Cook charted most of the Vanuatu archipelago in six weeks between 18 July and 31 August, 1772. His encounters in the north of Vanuatu were generally civil, though his men did shoot and kill at least one inhabitant in the southern islands (Beaglehole 1974:407). Cook arrived off Espiritu Santo on 24 August, resolved the discrepancy between De Quirós and Bougainville's descriptions of the island, and was out of sight of Vanuatu by 1 September. He left behind a bitch and a dog (as was his practice) and took away the previous European names for the area, though he gave a nod to his precursors' achievements when he renamed the north-east cape of Espiritu Santo, Cape Quirós. Beaglehole (1974:409) considers it unlikely that Cook settled on the name New Hebrides until some time after leaving the archipelago. It was, however, the name by which the islands were known to Europeans until 1980.

Really significant contact between Europeans and Ni-Vanuatu occurred during the 19th century when mission work drew European attention to the region, and trading in the region increased quite markedly. It all started rather badly. John Williams, a London-based missionary, was killed immediately on his arrival to Erromango in southern Vanuatu in November 1839 (Robertson 1902:53). One can't help wondering whether the memory of Cook's visit might have lingered though we do not know that the Ni-Vanuatu who Cook killed came from precisely the same area. Williams' death was followed by more aggressive missionising in the region through 1839-1842 (Laracy 1976), and it was probably a London Missionary Society ship that brought news to Sydney in September 1853 that sandalwood was plentiful in Espiritu Santo (Shineberg 1967:129; Laracy 1976). A trading rush quickly began. Until the sandalwood was exhausted in 1865, there was regular contact with traders and the Ni-Vanuatu of Espiritu Santo. Shineberg (1967:130) reports that in early 1857, for example, there were six Sydney ships anchored in the same port in Espiritu Santo at the same time.

During this period of economic opportunity in Espiritu Santo there was fairly significant internal migration to the area. A work force of at least 310 men from other islands is recorded as working at a sandalwood station on Espiritu Santo in 1860 (Shineberg 1967:191). However, Laracy (1976:132) believes that contact in the Santo region during the

4 "It is truly curious that at precisely the same latitude and precisely the same longitude that de Quirós located his large bay [on a continent: MM] we found a passage of exactly the same size as he gave to the mouth of his bay. Was our Spanish navigator blind?" [my translation]

early and mid 1800s was mainly with whalers and China traders including those interested in the *bêche de mer* 'sea cucumber' trade.⁵

In some areas of Vanuatu, a mission and a sandalwood trading depot were set up within a month of each other (Laracy 1976:132). The distinction between missionaries and traders was presumably sometimes rather opaque from a Ni-Vanuatu perspective. Hilliard (1976:183) notes that the Anglican missionaries came with the same trappings of wealth and exchange that the traders did, offering them in exchange for food or friendship, though Paton (1889) makes it clear that relations between some missionaries and traders were minimal or even antipathetical.

Mission work in this region began in earnest with the expansion of the presence of the Church of England from 1851–60 (Hilliard 1976:181-182), but permanent missions (and therefore the permanent presence of Anglo missionaries) were not established until c.1870 (Hilliard 1976:183). Jesuit missionaries began to arrive in the northern islands from 1887 onwards, though some of the first Jesuit missions were staffed only periodically into the 20th century (Monnier 1996). From 1909, however, there was a permanent mission in Santo (Saint Michel) (Monnier 1996:235). The first mission on West Malo was established by the Presbyterian church in 1888 (Rubinstein 1976:24).

Ni-Vanuatu contact with Europeans also became of quantitative and qualitative significance with the establishment of plantations in the Santo/Malo region in the late 19th century. Fortuné Lachaise, originally from Réunion, established a plantation on East Malo in 1880 (Monnier 1996:25) and George De Latour, a famously brutal and hated British settler, maintained a plantation on the island of Aore from c.1885 (Rubinstein 1976:23-24, Rannie 1912). The expropriation of the highly fertile land on Malo was pursued aggressively into the 20th century (Harrison 1935:48). We know that planters and their trading companies were well established in the Santo environs by 1910 because the records show that in a November 1910 earthquake they sustained considerable property damage (Monnier 1996:235). Most of this land was not, of course, returned to Ni-Vanuatu *kastom* owners until after independence in 1980.

The labour trade

By this time, another group of European traders had become a presence in the region. By the end of the 1860s, ships recruiting labourers for plantation work in the south-west Pacific (mainly what is now Australia, Fiji, New Caledonia and Samoa) were frequenting the northern islands of Vanuatu. The recruits were mainly young men between 15 and 35 (Scarr 1976:231), though a number of women also signed on for three years' work (Jolly 1987a). In the early years, the recruiting process was one of dubious honour. According to Scarr (1976:226, 246-247), blackbirding, as the recruiting was known, initially often involved kidnapping any Melanesians the ships came across.

Wawn (1973:255-257) gives an eye-witness report of the practice continuing into the 1880s. Even when contracts were negotiated early on, it is a matter of conjecture how

⁵ The *bêche de mer* trade is again the source of significant revenue for some of the Asian expatriates in Santo. The business has become sufficiently large that some observers have questioned the impact of the harvest on lagoons where large quantities of *bêche de mer* are being removed, since the sea cucumber plays an essential role in the processing of waste and the aerobicing of lagoon waters.

thoroughly the recruits understood what they were signing up for and for how long (Kay 1872; Hilliard 1976:197; Sankoff 1985). The claims of murder and kidnapping were finally investigated after being voiced in the British Parliament in 1871, and three Royal Commissions of Inquiry into the labour trade were held in Australia in the 1880s. The trade continued, somewhat more regulated by shipboard government agents and more characterised by voluntary recruits, into the 20th century, but conflicts between crews and recruits, crews and agents, and recruits and their elders in their village continued to be a feature throughout (Scarr 1976).

Generally, ships were not supposed to sign on single women, though for various reasons a number of women did sign up on their own. Jolly (1991a) puts the number at 6–10% of all labour recruits from Melanesia, and notes (1987a) that despite the common characterisation of them as sexual adventurers (e.g. Scarr 1976:243), there is evidence to suggest that they were not signing on out of lubriciousness. For example, they kept separate quarters from the men and treated the men's hold as they would a men's long house, i.e. they would not walk over it when they were on deck. Nor is it clear that they signed on as a means of eloping, as Scarr also suggests. Many shipboard marriages seem to have been performed at the behest of Victorian ship masters who were uncomfortable with the propriety of having single women aboard. Independent reports of the immigration agent in Brisbane and missionaries on various islands during the 1870s (Kay 1872) show that a number of couples signed on together (Hope 1872:44 notes that it was sometimes seen as a useful strategy to try and recruit a woman directly in the anticipation of her husband signing on with her). The motivations of the women signing on probably had as much to do with boredom and adventure, escape and profit as the motivations of the men.

The labour trade transformed Vanuatu society. Emigration to Queensland resulted in substantial depopulation in especially the smaller communities. Captain Wawn describes the resistance of elders on Epi and Lamén Islands to the repeated loss of active members of the community, at one point saying, 'the Lammen chiefs [must] be furious at my snapping up a lot of their young men... I should not have cared to have shown myself on that islet for at least a month to come' (Wawn 1973:96). The total number of individuals involved is a little unclear partly because on the expiration of their first contract, some recruits signed on again (either in Queensland or Vanuatu). Crowley (1990:88-89, citing Dutton 1980) shows that between 1863 and 1907 nearly 40,000 Ni-Vanuatu were working on Queensland plantations (over 15,000 were working on plantations elsewhere in the Pacific). Between 1870 and 1880 alone 25,358 young men and women left mainly for Queensland and Fiji. This, out of a population at the time of between 70,000 and 100,000 (Condominium 1967:20; Rice 1974:64).

However, it needs to be remembered that the population loss associated with the labour trade was staggered over a period of 40 years and off-set by repatriation at the end of a contract. Based on repatriation figures, the 1967 Census calculates that the net loss to the population of Vanuatu during this 40 year period was 10,000 (Condominium 1967:16). Notable as this is as a percentage of the whole population, it pales into insignificance by comparison with the profound effect introduced diseases had on the population.

The two can hardly have been unrelated. Scarr (1976:230, 250) notes that some contemporary doctors and missionaries attributed the devastating epidemics of the period to the contact with labour traders, though the Presbyterian minister John Paton's own records show (1889:255) that the Ni-Vanuatu and the traders considered the missionaries at least as

significant a medium in the introduction of diseases such as measles. Paton (1889:372) estimated that by the end of the 19th century the population of Vanuatu had sunk to 70,000 and that this was at most a mere quarter of what it had been before contact with Europeans. Epidemics of influenza and measles were capable of carrying off between 10% and 50% of a community (Rice 1974:246, citing Deacon 1934). Rice (1974:64-65) gives several estimates of pre-contact population figures, and it seems likely that at around 1800 Vanuatu's population was at least half a million and may have even been as high as three million (this compares with a total population of c.177,000 in 1996).

In addition, the labour trade had a lasting impact on Vanuatu in the changes it wrought in returning recruits' social behaviour. First-hand experience of European cultural norms in the plantations reinforced some of the social changes the missionaries were introducing in sex roles and social hierarchy and organisation. Colonists introduced their own value systems, redefining appropriate and inappropriate sex-role divisions of labour (Paton 1889; Jolly 1987b, 1991b), and the Patons' letters and journals make it clear that missionaries and traders had similar ideas about the most appropriate divisions of labour according to race.

Jolly (1987b) notes that although pre-colonial Vanuatu society was based on asymmetric social roles, e.g. between graded and non-graded members of the community, men and women, this asymmetry was underpinned by a fundamental belief in the intrinsic worth of individuals. She argues that post-colonial contact with European cultural norms has complicated the social climate considerably by introducing a European notion of equality which defines equality in terms of equal opportunity and equal access to resources and skills, and that this is in competition with the pre-existing conceptualisation of intrinsic equality (1987b:179). Jolly argues (1987b, 1989) that the post-colonial social scene has, if anything, introduced greater divisions between the roles of men and women in Vanuatu as producers and guardians of tradition, and encouraged the identification of men with the public domain and the identification of women with the private/domestic domain.

Paradoxically, the European sex roles and European concepts such as equal access and equal opportunity that were introduced together were in conflict with one another, in a way that Jolly suggests customary notions of equality based on intrinsic worth and customary asymmetric social roles were not. Jolly (1991a) points out that many Ni-Vanuatu women do not consider the discourse of European feminism as an appropriate response, for precisely these sorts of reasons. With its emphasis on individual autonomy rather than 'values of community and kinship' (1991a:57), European feminist discourse does not address the importance of the family in Vanuatu as the locus of cultural survival and resistance to white dominance.

However, the labour trade of the 19th century had one further consequence for Vanuatu society, one that is of greatest significance to this study. The indentured labour system brought many thousands of Ni-Vanuatu into contact with the Melanesian Pidgin English (MPE) spoken on plantations in several parts of the south-west Pacific, and it was the MPE spoken by these recruits on their return to Vanuatu that formed the basis for modern Bislama.

Twentieth century developments

In 1906 the islands (as the New Hebrides) were united administratively by the British and French governments, in what was to be known as the Condominium. This arrangement,

supposed to be a stop-gap until the French and British could resolve their respective claims on the islands as a colony, in fact continued until independence in 1980.

During the first half of the century, a number of immigrants from French Indochina came to work in Vanuatu as indentured labourers (designated 'coolies' on the official papers, Tonkinese Work Permits n.d.; in Vanuatu they are still sometimes referred to as 'Tonkinese', though the Gulf of Tonkin region from which they emigrated is now part of Vietnam). The first convoy arrived in Port Vila in August 1920, and "from 1925 nearly all the important plantations in the archipelago [were using] Asian labour".⁶ On Santo, this included plantations in Port Olry, Surunda, Palikula and in the immediate environs of Saint Michel (Monnier 1996:317). In total, there were nearly 6,000 Vietnamese workers in Vanuatu in 1929, 3,700 men, 1,200 women and 1,000 children (Monnier 1996:316).

The Vietnamese were hired on five year contracts, and appear to have kept to themselves during their residence in Vanuatu. There were the inevitable class distinctions between them and their Francophone employers, but it seems that there was relatively little mixing with Ni-Vanuatu, too. The contracts of the Compagnie Coloniale de Vaté, for instance, guarantee the Vietnamese (a) will be kept together as family units and provided with their own garden space, and (b) will be accommodated separately from the *indigènes noirs*. (Consequently, it seems unlikely that their presence had a significant effect on the linguistic situation that obtains today.)

In addition, the majority of the Vietnamese labourers returned to what is now Vietnam once their contracts were up. Some chose to remain in Vanuatu, and are now extremely successful and powerful families economically and politically. Anecdotally, I was told by a man in his late 50s in Santo that the Vietnamese around Santo often found the business of paying off debts and securing passage home to be a slow one. Added to which, repatriation appears to have slowed or stopped during the French war in Indochina (late 1940s through 1950s), and only resumed in 1960. The last group of workers and their descendants achieved their longed-for desire of a return to their homeland in 1963 and 1964 (Pisier 1980). Which is a sobering thought.

Based on the perceived success of the Vietnamese workers, recruits were also sought in China by some of the Francophone companies. By the mid-20th century, there were more residents of Chinese extraction than Vietnamese origin living in Santo (3.3% versus 1.4% of Santo's urban and periurban population respectively, Condominium 1967:30).

There was, of course, a period of intense Anglophone dominance in the Santo region during the Second World War, though this was not British. The US military established bases first on Efate and subsequently on Espiritu Santo where they used the shelter of Segond Channel as a northern port. There were also watch posts on Espiritu Santo's off-shore islands, and some of the older men in Malo in the community I stayed in had worked as water carriers and manual labourers for the American and New Zealand troops stationed on Malo when they were young.

During the war, over 250,000 troops from the US armed forces alone (along with smaller numbers of New Zealanders and other Allies) passed through or were stationed in Vanuatu. The main US bases were in northern Efate and Santo (Van Trease 1987:131), and the social and economic impact of this temporary invasion was profound.

⁶ "A partir de 1925, presque toutes les plantations importantes de l'Archipel utilisent la main-d'oeuvre asiatique" (Monnier 1996:187).

Much of the military's infrastructure remains in and around Santo and military detritus is the single largest attraction for recreation and tourism in the region today. In particular, the wrecks of US ships (most famously the SS President Coolidge in Segond Channel) are popular for divers, but tourists also visit plane wrecks and the point off which the US military dumped vast quantities of their surplus equipment (from armoured vehicles to soda bottles) rather than leaving it for local use. Army Quonset huts are a distinctive feature of Santo urban architecture and many in Santo are still used as churches, meeting halls or workshops. They are also found in various stages of disrepair outside the township and on the west coast of Malo. The roads which the military laid down are still the sole vehicular thoroughfare to some parts of Espiritu Santo and Malo.

And naturally, the money which such a large military presence funnelled through Santo is remembered fondly. As is the role model of African-American servicemen. The father in the family I lived with remembered that seeing African-American servicemen in positions of responsibility, driving heavy vehicles and giving orders made a strong impression on him and other Ni-Vanuatu (Van Trease 1987 also notes these sorts of reactions).

However, with the end of the war, the US and New Zealand presence in Santo evaporated, again leaving just the French and British administrations in the Santo region. It was the Condominium authorities that declined to buy the excess US equipment mentioned above. Presumably they hoped the US forces would abandon it rather than ship it back to the United States, and reasoned that they could get it for free if they just waited. However, this was precisely what the massive dumping exercise in Segond Channel was designed to foil.

One consequence of this period of joint, and rather competitive, Anglo-French administration was considerable parallel development of the islands' infrastructure and social services. In what seems to have been an unwillingness to allow either side to get one up or too influential in a region, schools, hospitals and legal services were sometimes duplicated in the same area by the British and French colonial authorities. However, despite the concerted efforts of the Presbyterian Mission to settle Australians in the Santo region during the 1890s (Van Trease 1987:32-33), and despite the institutional duplications, the main colonial presence, both economically and culturally, around Santo before 1980 was French. Competition between French and British interests in the Santo region last came forcefully to a head when both the (by then former) colonial powers supported opposing sides in the secessionist movement on Espiritu Santo after independence.

In the 1970s, a nationalist movement began to gain social and political prominence in the Santo region. The movement known as Nagriamel had as original cornerstones the reappropriation of alienated land and a re-emphasis on *kastom* 'customary practices', and was led by Jimmy Stevens, a man of English, Tongan and Banks Islands descent (Rice 1974:117-118). In the 1979 elections which established the composition for the first independent Ni-Vanuatu government to follow, Stevens' Espiritu Santo-based Nagriamel party failed to win a place in the national government, nor did Nagriamel gain control of the regional councils in the areas of its largest followings, Espiritu Santo and Tanna (in the south), as they had expected.

In the wake of Nagriamel's unexpected losses in national and regional elections, Stevens attempted a secession of Espiritu Santo from Vanuatu. With the financial and military help of French supporters and some US libertarians he was able to mount a rebellion in Espiritu

Santo (Lindstrom 1990).⁷ There was a massive evacuation of all expatriates, civil servants and known supporters of the government-elect from Espiritu Santo. The island was blockaded, and life became rather ugly for many of the people who remained. The government-elect responded by sending in the newly-formed military, the Vanuatu Mobile Force (VMF), who, with practical back-up from the Papua New Guinea Mobile Force and diplomatic support from Australia, went in to Espiritu Santo to subdue the rebellion as soon as Vanuatu's independence had been proclaimed. Before normalcy was restored in the latter half of 1980, both nationalists and secessionists had made the most of the lack of order to rob, beat or harass supporters of the opposition. According to one of my Santo interviewees, the rebellion presented an opportunity to settle some old scores, and that some people were killed. However, Van Trease (1987) makes no reference to any deaths.

The failure of the French-backed secession attempt saw a major shift in the demographics of the expatriate community in the Santo area. Before independence, the town had been socially and economically dominated by a French-speaking expatriate community (c.200 French families throughout Santo in 1954, according to Wilson 1994:174). In 1994–95, the expatriate community had shrunk to approximately 450 people total. Most of the Europeans were Anglophone (c.100, versus c.40 Francophone expatriates) and most of these were Australians (at this time there were still a number of trained Australian teachers at Matevulu College nearby. These seem to be now being gradually replaced by untrained members of the US Peace Corps: it remains to be seen what effects the new US migrants may have on the area). The bulk of all expatriates (c.300) were ethnic Vietnamese or Chinese business-owners and their families, and a small number of Japanese.

One major legacy of the Condominium were the two colonial languages of administration. Both French and English were official languages in the New Hebrides, and they continue to be official languages in Vanuatu and the only official languages for schooling. Since independence, Bislama has also had the status of an official language, but in addition, Bislama alone is recognised as the national language of Vanuatu.

2.2 Sketch of contemporary social life

2.1.1 *Life in Santo (1994–95)*

The population of Santo is primarily drawn from the northern islands, with some close-knit communities defined on the basis of people's customary island affiliation. Thus, even a person born and raised in Santo will consider themselves to be *man Pentecost* or *woman Ambae* etc. if that is where their family's main ties and land rights are. In Santo there are communities of significant size from Paama, Malakula, Pentecost and Ambrym, as well as some significant regional groupings from elsewhere on Espiritu Santo, e.g. people from the Big Bay region, Hog Harbour and Port Olry (a.k.a. *Pot Lori* in Bislama) – all north and generally east of Santo town. There is a small amount of business and industry in the area: the wharf, a recently opened coconut products factory, fisheries research and development, two sawmills, several bakeries, a cattle abattoir and recreational diving and tourism. The central and regional governments are reasonably large employers; Santo has a regional

⁷ Van Trease (1987) has a detailed discussion of the rebellion, the Nagriamel movement and the internal and international political interests underlying it.

hospital, a VMF barracks, a magistrate's and supreme court, a prison and a public works department.

Santo is a fairly young town, and predominantly male. Official census figures for a town like Santo, where there is significant internal migration, quickly get outdated. The 1989 census tells us that 42% of Santo's population was 14 years or under, 55% between 15 and 59 years (about the same as for Vanuatu as a whole, Central Intelligence Agency 1997).

From discussions with Santo residents, including a local magistrate and a health care practitioner, an educated guess about the demographics of the Santo community in 1994–95 is that the average age was between 25 and 30, and that men outnumbered women by about 60:40 (the official census figures are 53:47 for Santo overall, and 55:45 in the newest parts of town). These demographic trends are particularly marked in some sections of town which are basically shanty towns for migrant workers. A number of young men migrate to the municipality looking for work and some are successful in finding jobs, but under- and unemployment are high. This pattern of internal migration has a strong effect on the make-up of the Santo community. However, it is by no means novel. It has been the case for some time that young men or newly married men in Vanuatu will travel internally to seek work on plantations (Jolly 1989). Thus, the migration to the municipalities is an extension of a pre-existing routine. The fact that the destinations are generally Santo or Vila now is simply a reflection of changes in the source of economic opportunity in Vanuatu over the last decades.

Shortly before my arrival in 1994, the employment situation in Santo had been severely affected by repercussions from a national strike of public servants called in November 1993 in support of a 16% wage claim. The central government stuck with a 5% offer, withdrew from negotiations and dismissed all strikers from their jobs. A number of people returned or migrated to the Santo area after losing their jobs elsewhere in Vanuatu and many locally employed public servants lost their jobs. The effects and ill-feelings associated with this continued to be apparent throughout the period of my fieldwork, and many professionals, especially in the teaching or health care professions (where there are few or no options for private sector employment) had returned to a *kastom* lifestyle (e.g. gardening on their family lands, cooking without electricity or gas) or had found employment in service industries, e.g. bus driving, house cleaning.

People's social lives are very much self-driven in Santo. There are no regular cinemas, no theatres, few bars and when I first arrived no night clubs (the two previous ones had been closed down because of repeated violence. One bar added dancing while I was there). In 1993 a betting shop opened and this had become a popular place for people to spend their spare time. In the evenings, small groups of young men fill the streets, and during the day, small groups of young women also form. Most socialising happens in people's homes, where people usually just come calling. From about 7–11pm, many men and some women may go to a *nakamal* to drink *kava* for a while. This is a quiet occupation and contrasts with the bi-monthly roistering (which is largely male) at the local bar on the Thursday and Friday nights after payday.

The chief medium of communication in Santo is Bislama. This is true between Ni-Vanuatu and also between expatriates and Ni-Vanuatu and between most of the Asians and Ni-Vanuatu. Most people speak *lanwis* (the Melanesian language spoken in the area they come from) with (*wo*)*man ples* (i.e. people from the same place) and this is as likely to have instrumental motivations as it is to have affective ones. Language maintenance and the

desire to stress a speaker's family roots and regional affiliations are one set of reasons for using *lanwis*. Another is the obvious fact that using *lanwis* can be a useful way of speaking privately around outgroup members.

But in public domains, whether business, social or religious, Bislama is the norm. It, too, serves instrumental and affective purposes since not only is Bislama useful as a *lingua franca*, as I have noted, it is positively evaluated by many Ni-Vanuatu as a symbol of shared national identity.

2.2.2 *Life on Malo (1994–95)*

Malo, along with several other smaller off-shore islands, is part of the same administrative province as Espiritu Santo and there is fairly steady traffic between the two islands, either on private outboard dinghies or, more recently, on privately owned 12m fibreglass outboard boats that act as water taxis.

The community I came to know best on Malo provided a view of a more *kastom* lifestyle than the lifestyle of most residents of Santo. As is customary in fieldwork, in order to respect the privacy of the people who spoke freely with me on my tape-recordings, I have masked their names and identifying names of localities. I will give the area I stayed in on Malo the pseudonym Avunaples.

Avunaples proper, consists of a mission church, a small clinic, an English language school which goes to Grade 6, and a small number of immediately surrounding houses. There are a number of smaller, patrilocal villages in the neighbourhood of Avunaples. It should perhaps be noted that even in the matrilineal regions of Vanuatu such as East Ambae, patri- or virilocal residence is normal (Rodman 1981; Blackwood 1981). Often these are situated on the *kastom* land belonging to the oldest man in the family, but this is not invariably the case. The family that hosted me, for instance, lived on land which was purchased after the father of the family was forced off his *kastom* land by a French planter about 40 years ago. This land was restored to the family after independence when all *kastom* land rights were reasserted as part of the Republic of Vanuatu Constitution (except for land in Vila and Santo which is considered public land and can therefore be bought and sold [Van Trease 1987:260]). But three generations of the family continue to live in houses on the newer land around a clearing near a shallow portage.

The home of the mother and father of the family is basically in the centre of the village, with houses for their sons and their families spread out around them. Even some of the sons who do not live full-time on Malo keep a house in their family village. Most houses are still built using local materials, walls of woven bamboo and *natanggura* thatch, with a broken coral or poured concrete floor. Usually a house is built for a young man by the time he reaches his early twenties regardless of whether he has married. Until he has a wife, though, he may not cook for himself, but continues to eat with his parents. A schematic diagram of the layout of this village is given in Figure 2.3.

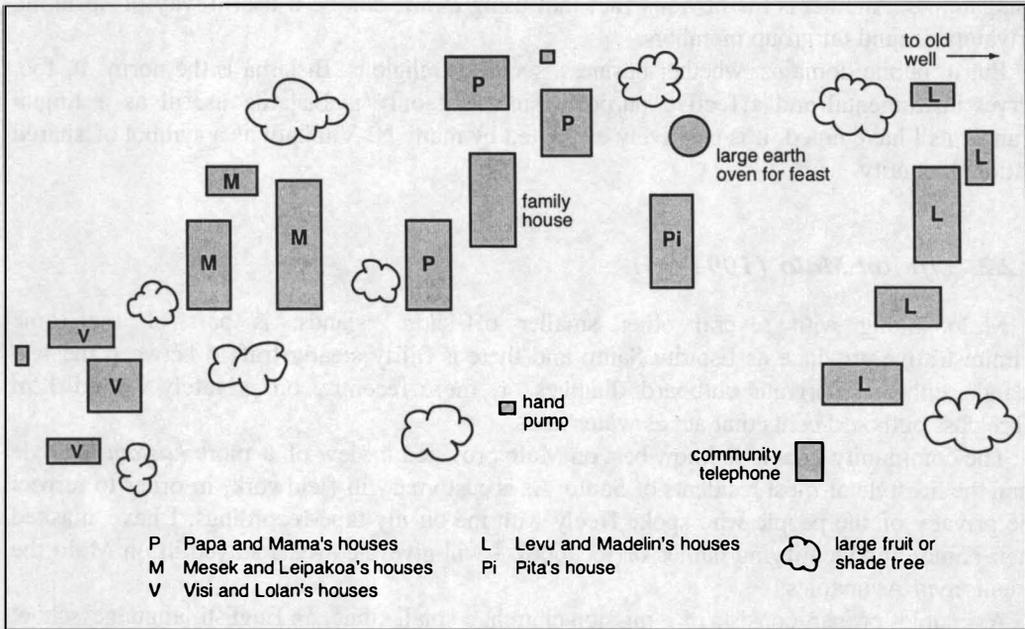


Figure 2.3: Layout of buildings and major features in a patrilineal village on Malo.

The family villages often look to one or other of the churches on the island and the school (or schools) that they send their children to as the locus for public social activities. Part of the legacy of the Condominium is that on Malo there is usually one English-language school and one French-language school located near each other. Some families educate all their children at one or other of the schools, but the family I lived with was not unusual in having educated three sons in French and three sons and one daughter in English.

Few children in Vanuatu progress past Grade 6 since schooling is neither free nor compulsory at any level in Vanuatu. The young people in Avunaples generally have one year of grace (when they count as school leavers) before they are expected to take a full role in adult society. This generally entails working the family gardens, though some girls contract out with family or friends as a *haosgel* 'housegirl'. This means living with another family (often she will do double duty in gardens, assisting in her parent's garden on Saturday), and helping the woman in the house with washing, pumping water, food preparation and childcare. Sometimes these arrangements provide young women with a means of seeing something of the rest of the country, as they may have employment arranged with a family member who lives in Santo or Vila.

Generally, young people are married by their early twenties, at which time socialising with their young, single, friends around the volleyball court and/or soccer field ceases and they start to mainly socialise around people's homes. People regularly drop in on family in the afternoons or evenings. As a rule of thumb, if the door to someone's kitchen house is open, people are 'at home' to visitors. Sleeping quarters and a *haos blong swim* ('bathing house') are separate structures near to a family's kitchen and dining house, and a *smol haos* ('toilet') somewhat further away.

Women and girls make trips to Santo to sell produce at market approximately every fortnight according to a roster drawn up by the market organisers and the women on Malo themselves. They are sometimes accompanied by men who enjoy the entertainment of a couple of days in town, but the market work itself is entirely conducted by women. Sometimes men will help out with provisioning in the Santo stores, but generally this is the women's job too.

During the period of my fieldwork there was a change in the way market was administered. When I first arrived, markets would start at 4pm and run for 24 hours, throughout the night. Most buying, of course, happened early in the evening and first thing in the morning when supplies were topped up. But regulations were eased after the 1995 local council election and women could start selling at market from whenever they arrived on Mondays, Wednesdays or Fridays.

The effect of this is to make markets last about 36 hours instead of 24, and while it allowed people to (in theory) make more money every market day, it has also made the whole event something of an endurance feat for the women involved, and several women commented on how much more tiring the new regime was. In order to get good places at the tables, Malo women try to arrive early on market days, which means a very early start to package produce (sometimes even to harvest it) and transport it to Santo.

Transport involves first catching a boat for a ride of between 30 and 60 minutes (depending on how far down the west coast of Malo you live), then catching a truck to Santo (another ride of about 40 minutes). Women then spend the whole day, the night and most of the next day at market, before catching trucks and boats back to Malo, usually arriving home at around 6pm.

However, long hours do not guarantee a steady stream of customers, as most buying happens at lunch time or early evening. What this means is that the market experience is predominantly one of sitting around, gossiping and trying to snatch some sleep under the trestle tables. Generally, socialising in and around the market is gender-segregated, much like other public socialising, but for the younger women, market days also provided a chance to get to know men.

The market is not the only connection between the people living on the villages of Malo and Santo. West Malo is an easy and (for people on salaries) relatively cheap commute from Santo. In 1994–95, it was possible to go one way for VT250 per person (freight extra). At that time, 100 Vatu converted to approximately US\$1, but a better indication of the actual value of VT100 is that in 1994–95 it was the price of two loaves of bread. Because West Malo is so accessible, many *man Malo* who work in Santo spend weekends or days off on the island. In a sense, the Avunaples community, though relatively distant from Santo, had sufficient ties with Santo township that one might consider Avunaples and its inhabitants a periurban extension of Santo.

2.3 Current linguistic situation

2.3.1 Bislama: the one national language

The constitution of the Republic of Vanuatu gives Bislama sole status as the national language, though not necessarily in perpetuity. The government may (subject to national referendum, as required for all changes to the constitution) replace it with some other

indigenous language. Bislama, English and French all have the status of official languages. French and English are specified as the principal languages of education, and the government is required to pay due attention to the preservation and maintenance of indigenous languages.

Until the 1960s, Bislama was still primarily a language spoken by men (Jolly 1991a:71). Its usefulness as a lingua franca meant it was not only used among Ni-Vanuatu. Bislama was often used between French and British administrators on Santo before the Second World War since they often could not speak each others' languages (Marshall 1937). It is tempting to think that Bislama provided French and British administrators with a 'neutral' medium for communication: neither party would have a linguistic advantage; and the question of the relative importance or status of the colonial languages could be avoided.

In the 1990s, however, Bislama is considerably more than a pidgin used purely for work or business. An increasing number of families, both in villages and in the towns use it on a day-to-day basis, and Tonkinson (1982) remarks that it is one of the least problematic vehicles for invoking national identity in Vanuatu since it is equally a part of the cultural heritage of all Ni-Vanuatu. This was also a recurring theme in conversations I had with Ni-Vanuatu about the role of Bislama and its function in their lives.

These attitudes were given a very public airing in 1995. The editors of the national newspaper, *The Vanuatu Weekly/Hebdomadaire*, announced that the paper would henceforth cease using Bislama and would only appear in English and French. In the ensuing row, a number of people explicitly stated their support of Bislama in terms of its function as a symbol and instantiation of national identity (see for example, letters to the editor in *VWH* 17 and 24 June 1995; 1, 15 and 22 July 1995). Also, as Tonkinson (1982) notes, the vast majority of the distinctive, Vanuatu string band music is sung to Bislama lyrics. The Ni-Vanuatu theatre group based in Vila, *Wan Smolbag Theatre* 'A Pocket Theatre', stages productions in Bislama, and some of its Bislama-language dramas garner a nation-wide audience (through regional tours and national television broadcasts). Clearly, for many Ni-Vanuatu, Bislama is the medium of choice for expressing developments in art and culture at a national and individual level.

Naturally, this attitude is not universal. Some Ni-Vanuatu consider Bislama a hindrance to the learning of "proper English" or a simplified language not suitable for modern life (these attitudes also emerge in the letters to the editor in the *VWH*). Some people also worry that knowing Bislama makes it harder to learn English (the importance of this perception must be seen in terms of the economic opportunities associated with having a good command of English). One Francophone Ni-Vanuatu in Santo was rather distressed to learn I was studying Bislama, and it became clear that his basic opposition to the use of Bislama was due to his perception that the language was essentially English. In §3.6.1, I will discuss further the apparent haziness that exists between Bislama and English.

In general, though, most of the Ni-Vanuatu I spoke to expressed positive feelings about Bislama at the supra-local level. The prevailing opinion among Ni-Vanuatu is that it is an extremely important marker of shared national identity, perhaps more so than any one other social marker. At the local level, however, perceptions of and attitudes towards Bislama become more complicated. Locally, some people see Bislama as a threat to the survival of local languages (and hence local culture). This creates a situation in which a single person may express support for Bislama as a marker of national identity, but rue its use instead of a local *lanwis* in their community. This paradox will be returned to in §3.8.1.

The land one comes from, or that one's family comes from, continues to be a very important marker of personal identity even in urban areas such as Santo and all Ni-Vanuatu identify as (*wo*)*man ples* (a man or woman from a particular island). Even though I was told that a woman acquires the same island affiliation as her husband if she marries off island, in practice most of the women I met continued to identify themselves and be identified in their husband's community with the island they were born on. *Woman Maewo* 'a woman from Maewo' might marry *man Malo* 'a man from Malo' and their children would certainly be *pikinini Malo* 'Malo children', but her friendship and social networks were very likely to be with *ol woman nara aelan* 'women from other islands' (the social significance of this ingroup label in the Malo community is discussed in Chapter 3).

Women in both Santo and Malo seemed to be happy to hold on to their ties to home. They were understandably eager to talk about where they grew up and to compare life in their home village with life where they had settled down. And other members of the community would continue to often identify a woman by her birth island rather than by her husband's island. In a place like Malo, where the group of *woman nara aelan* was still fairly small (as compared with Santo, for instance), a woman's place of origin could stand in for her name in conversation, i.e. *woman Pentekos ya* 'the Pentecost woman' could be used rather than 'Anita'.

However, even as intergroup distinctions between islands remain strong, it is increasingly the case that there is one thing Ni-Vanuatu share, and that is Bislama. The national language cuts across different islands of origin and different *kastom*. The use of Bislama among and to *woman nara aelan* on Malo (to be discussed in Chapter 3) and the widespread use of Bislama in Santo, serves to underline the significance of Bislama as a unifying social code in situations of the greatest intergroup diversity.

2.3.2 Use of English and French

Although there are three official languages, people's emotional attachment to them differs somewhat. In the communities in which I conducted my fieldwork, English was widely seen as a highly instrumental language, something it is useful to be familiar with. This was especially true in Santo where, as I noted, there are now more Anglophone than Francophone expatriates.

Although some of the families who educated their children in English sometimes used the language in the home, this was not generally the case. None of the Anglophone educated speakers claimed to use English with their children, and indeed, I very seldom heard it used. When pressed, some speakers, e.g. Alis, reported that they might use a few words of English here and there, but really only for joking around. There were only a few occasions when someone would toss out self-conscious quotes in English, e.g. on M-95-10, Bretian refers to Sylvester Stallone in *Demolition Man* as "cop of the future, *wantaem*" (where *wantaem* is a marker of emphasis), and on M-94-2, when Iawia tells his mother not to put the coconut milk too close to the fire, Lolani replies "I remember".

Several of the people I talked to commented that it was interesting that Francophone Ni-Vanuatu families were more likely to use French among themselves than Anglophone families were to use English (this point is developed further in §3.6.2, where I discuss the case of French-educated Ni-Vanuatu who use French outside of school). At present we have only fairly sketchy information suggesting why and how often Francophone Ni-Vanuatu use

French as compared with Anglophone Ni-Vanuatus' use of English. For example, Crowley (1990:22) notes his impression that this is true, and suggests that it is perhaps because the French-education system stigmatises Bislama more than the English system does (cf. the comments by the Francophone Ni-Vanuatu mentioned above).

This may be part of the reason (but if so, it remains to be seen whether this is because Bislama is perceived to be "too English" in the post-colonial competitiveness of the dual administration systems, or whether it is due to better teaching of French in French language schools). However, it may be even more evanescent and more over-arching than that.

As is the case in much of the Francophone world, the French language appears to be a very important marker of shared identity, and competence in French is much more affectively charged than is competence in English. (Nagy (1996) outlines the historical development of the significance of standard French as a unifying feature of French colonialism and French culture). More than just competence in a skill, knowing French invokes membership in an international community. *La francophonie* simply has no sociolinguistic parallel in English. The ideology of a unified culture and politics embedded in a single language continues to colour the rhetoric of the colonial and post-colonial Francophone world. This is not to say that the cultures of speakers of French are not as diverse as the cultures of speakers of English – clearly they are. But there is no socio-political framework in which use of English is embedded that is equivalent to the situation for French. In contrast to the centring of *la francophonie* on standard European French, English in Vanuatu has diffuse targets of prestige and status including British English, American English and Australian English in different contexts and for different functions.

In addition, there are regional and local reasons why French speakers might feel more inclined to stress their French identity through continued use of the language. Although the press associated with the continued colonial and military presence of the French in the Pacific is by no means all good, it must be remembered that the French have maintained a strong fiscal and cultural presence in the South Pacific. This reinforces and helps reify a pan-regional notion of *la francophonie* and reminds people of the benefits membership in it can accrue.

At a local level, in achieving independence Vanuatu by no means left behind all the tensions associated with the Condominium. Political allegiances continue to often be defined with linguistic hooks. During the first years of Vanuatu's nationhood, governments were mainly Anglophone, and there is some complaint that Francophone Ni-Vanuatu did not get fair access to resources during the Anglophone administrations. With the last two governments at the time of writing (coalitions led by the (Francophone) Union of Moderate Parties), there has been more than a little sense of score-settling.

Moreover, in addition to the historical (and continued) association of certain national political parties as 'French' or 'English', the contacts between Vanuatu and the Anglophone and Francophone world are qualitatively different today. France and the European Union continue to sponsor quite a lot of development and aid programs in Vanuatu, while Anglophone countries (particularly the UK but also Australia) have been decreasing their aid to and involvement in the country. These social and economic facts cannot be perceived in isolation from the language situation in Vanuatu. They are factors that play a role in determining the relative intergroup salience of individuals' educational and political allegiances, and to the prestige and symbolic value of individuals' language of education.

2.3.3 Diversity and variation in Bislama today

As has been noted, a major social function of Bislama is its role as a symbol of national unity and as an index of collective identity. This is not to say that Bislama is exactly the same throughout Vanuatu. Just as Bislama has developed separately from the closely related languages Tok Pisin and Solomon Islands Pijin over the last century, internal differences have emerged within Vanuatu.

Perhaps the most obvious level of variation is phonological. There are differences in the voicing of consonants, use of fricatives versus stops, presence or absence of /h/ as a phoneme, and differences in the extent of vowel epenthesis and in the phonological nature of the epenthetic vowel used. The Santo/Malo recordings show the pronunciation of *hem* '3SG' to most frequently be /hem/, as opposed to /ɛm/ which appears to be the norm in and around Vila. The presence of the phoneme /h/ in general is noticeable in northern speakers; forms such as /hek ~ heg/ for 'egg' are recorded for a number of speakers (e.g. M-95-19).

The /h/ variable seems to be more of a linguistic marker (Labov 1972), but there are also a number of linguistic stereotypes which speakers access readily. Speakers from Pentecost were much stereotyped on Malo for the frequent breaking up of consonant clusters with epenthetic vowels, e.g. realising *kopra* 'copra' as /ka'burɑ:/; *spun* 'spoon' as /si'pun/. The duration of long vowels in the Bislama of Tanna speakers was also the subject of jokes, e.g. M-95-18, (*wan*) *olfala* '(this) old guy' realised as /ɔ:lfa:l/ by a speaker imitating a Tanna accent.

The extent to which these phonological variations are a direct consequence of the transfer of substrate features into Bislama, and the extent to which they are arbitrary lexical developments remains open for systematic study.

There is some lexical variation in Bislama, mostly it seems due to borrowing from local languages, e.g. speakers in Santo identified *go long wolwol* [wulwu] 'go shopping', or *wos* [wus] 'to paddle; hammer in a nail' (following vernacular spellings) as tokens of Bislama that one might hear in Santo but not elsewhere. As it turns out, *wolwol* seems to be a borrowing from Tamambo (spoken on Malo), and *wos* from the Banks, thus it seems clear that one of the most salient features of the regional variety spoken in Santo is the blend of lexical borrowings from languages in contact in the town.

There seems to be some variation in function words. The variation between *from we* and *from se* (both meaning 'because') found in the Bislama spoken in Santo and Malo appears to strike people familiar with the Bislama spoken in the southern and central islands as odd. It remains to be seen whether these sorts of differences are systematic regional variations or whether the person in Vila who registered surprise at this particular variable was expressing an idiosyncratic opinion.

Syntactic variation in Bislama is less obvious than phonological and lexical variation, or at any rate, it is much less stereotyped among Bislama speakers. This is worth bearing in mind when we come to examine the patterns associated with variation between phonetically null and pronominal subjects in later chapters. As we will see, there is little evidence that this variable differs significantly according to speakers' place of origin.

2.3.4 Variation in Tamambo

Until quite recently, Tamambo was even less well-described than Bislama. Recent work by Dorothy Jauncey (1997) provides a comprehensive description of the variety of Tamambo spoken on West Malo. She notes some aspects of social, especially age-graded, variation in Tamambo, and also outlines phonological (1997:481) and syntactic differences between Tamambo and Tamapo, the variety spoken on the eastern side of the island, specifically the left- or right-headedness of NPs (pers.comm.).

In §3.6.5, I discuss some of the on-going changes in Tamambo spoken on West Malo, such as a loss of some fricatives. But these appear to be changes in progress (some of which the members of the community are well aware of) rather than regional variation. To date, no quantitative data on the social factors constraining variation in Tamambo has been gathered. The following outline of the variation is much indebted to Jauncey's work (1997:26-28), and to her astute observations in the field.

Tamambo syllable structure is (C)V(N), with main stress on the penultimate syllable (Jauncey's description suggests what in metrical terms would be described as bimoriac, left-headed feet, 1997:33). Older, especially male, speakers have two allophones of the velar fricative /x/: a voiced allophone [ɣ] that follows a stressed syllable, and a voiceless allophone [x] elsewhere. Thus,

2.1 /xaraxa/ 'crawl'
 → [xɑ'rɑɣɑ]

She notes that some of the older men who are particularly socially conservative (defined as those who have seldom if ever left Malo), alternate between the voiced velar variant and a voiced uvular fricative, i.e. [ʁ].

The most frequent pattern (typical of women, younger speakers, and more socially mobile men) is slightly different: /x/ is realised as [ɣ] following a stressed [-back] vowel, and varies between [x] and [h] elsewhere, for example:

2.2 /saxasaxa/ 'work' → [sɑɣɑ'sɑɣɑ]

Jauncey believes that there is a tendency for well-educated speakers in this age group, again especially women, to vary between [x] and [h] in all environments.

Children, teenagers and less socially mobile young adults (defined as those who have chosen to stay on Malo and work in the gardens) exhibit a variation on this second pattern: /x/ is realised as [g] following a stressed [-back] vowel, and varies between [x] and [h] elsewhere, as in 2.3, and occasionally, usually in children's speech but also sometimes in the speech of *woman nara aelan*, the stop loses its voicing, as in 2.4.

2.3 /saxasaxa/ 'work' → [sɑgɑ'sɑgɑ]
 /ni xo/ free 2s pronoun → ['ni go]
2.4 /bulaxi/ 'throw' → ['blak^hi]

The latter two pronunciations are socially marked in the Tamambo speaking community and subject to meta-linguistic comment by speakers. In §3.6.5 I will suggest some of the effects these and other changes may be having on the use of Tamambo in the Malo community.

2.4 Conclusion

In reviewing some of the social and historical developments in Vanuatu, focusing on the northern islands of Espiritu Santo and Malo where I conducted my fieldwork, I have tried to give the reader some idea of the full context of the study. The story of synchronic variation in any language cannot be completely dislocated from its diachronic foreword. This is particularly the case with a contact language like Bislama, a language that only has a few generations of native speakers and continues to function in a highly multilingual environment.

In the next chapter, how people choose between Bislama and other languages (when they are available to them) is examined in more detail as well as speakers' attitudes to the different languages they command. This will provide a more detailed picture of the social background that frames the linguistic variation to be discussed in Chapters 5 and 6.

3 Tingting blong lanwis long Malo: *language use, and some attitudes to language on Malo*

3 Introduction

Now that we have seen something of the linguistic and social history of the areas in which I conducted my fieldwork, in this chapter, I present a more detailed examination of the patterns of language use on Malo. The information provided in this chapter combines information provided by members of the Malo speech community and observations made during the course of my fieldwork. The chapter is structured as follows. In §3.1 and §3.2, the contributions other fields in the social sciences have made to the analysis and methodologies of studying language in use are discussed. Next, the reported and observed language choices made by individual speakers in my corpus are presented. I discuss in detail some of the more interesting findings with respect to addressee and social domain. I then discuss what we can infer or learn from the community's norms and from people's expressed opinions about how competing motivations may be managed on a case by case basis.

As a consequence, this chapter and Chapter 2 should provide the background which will enable readers to understand the significance of the social groupings that are used as the basis for quantitative analyses of variation in later chapters.

3.1 Choosing a language: indexing non-linguistic information

Although the Santo speech community uses Bislama as the chief mode of communication for everyone, Malo is a community of nearly 3,000 most of whom speak the local *lanwis*, Tamambo, as well as Bislama. This means that speakers have the option of choosing whether to speak Bislama or Tamambo. In addition, *ol woman nara aelan* will speak at least one other language, and some of them may, therefore, choose between at least three languages (some *woman nara aelan* form what amount to linguistic isolates, but others, e.g. from Ambae and the Shepherds, have others they can talk to in *lanwis*).

As in all multilingual communities, the choice of a particular language or languages in a given social domain is based on a number of linguistic, psychological and social factors. Carol Myers-Scotton's matrix language frame model (1993a) has contributed to our

understanding of the formal constraints on intrasentential code-switching by multilingual speakers. There is also a considerable literature on the affective and intraindividual factors important in situational and metaphorical code-switching (e.g. Poplack 1980; Gumperz 1982; Eastman 1992 and articles therein; Myers-Scotton 1993b. Some of these also deal with apparent formal constraints on code-switching).

The decision to use one language rather than another may be made on personal, political, affective, or pragmatic grounds. In many domains, it is not clear that any one consideration alone determines language choice, and a speaker's selection of linguistic code may be determined by some combination or resolution of several of the above factors. In turn, members of a speech community, who share essentially the same norms for code selection, can attribute psychological or social motivations to a speaker's choice of a particular language. Different languages may be taken to index a speaker's perceptions of self and their perceptions of the relationships with others in ways that are subject to quite conscious evaluation and comment. While the processes and decision-making that drive code-switching may not be consciously controlled, a simultaneous or *post hoc* interpretation of the code-switching may be something that members of the speech community can discuss critically and coherently.

I follow Ochs (1992) in using the term 'index' to refer to the relation between linguistic tokens and variables, speakers, speakers' stance and social categories. She uses 'index' rather than, e.g. 'reflect', because she argues the latter masks key aspects of the relation between tokens and a particular social category (age, class, gender, etc.) by implying one dimension is more actively involved in the construction of sociolinguistic meaning than the other. In fact, linguistic variables and social categories are much more interdependent. Acts maintaining, problematising and constructing social identities and interpersonal stances contribute to the social meaning of a variant, and uses of a variant feed or challenge existing social schemas.

This point has, of course, been stressed repeatedly in sociolinguistics, particularly by those working on language variation from the traditions of sociology or psychology. For example, Hymes suggests that language is as 'constructive' of perceptions as it is an indicator of them and that linguistic relativism should incorporate both dimensions (1974:192), and Gumperz (1977) points out that conversation is a process by which interlocutors' may alter their roles and status. Smith et al. (1980) also stress the importance of studying language and social categories as flexible and interdependent. They argue that 'people approach each other with fixed meanings...but as the interaction progresses, both mutually influence one another, ...[through] "compromise and accommodation"' (1980:291). This is reminiscent of Le Page and Tabouret-Keller's (1985) analysis of variation between codes within the creole continuum; in all cases the writers draw a picture of language variation as a strategic tool used to signal changes in social role or cultural identity. One disadvantage with much of the earlier work is that, whether intentionally or not, the models presented often suggested a rather more teleological process than is often the case in individual communicative events.

Ochs's use of the term 'indexing' is both more comprehensive and more agnostic with respect to the agency of a communicative event, i.e. it allows both for social roles to affect the selection of some forms, but also for the selection of forms or routines to affect the social roles the speaker is associated with, without telos or a particular direction. Eckert and McConnell-Ginet (1992), Morgan (1994) and Cameron (1996) (to mention but a few) have likewise made the case that sociolinguistic theory and especially sociolinguistic practice should treat linguistic and non-linguistic categories as being reflexively related. As I have

noted, it is a concern with a respectable lineage in sociology, and has recent parallels in anthropology and philosophy (e.g. Gal 1992; Bourdieu 1991; Butler 1990).

3.2 Intersections with social psychology and anthropology

We have come to expect that speakers' attitudes to their language(s), addressees and situation can interact dynamically with their linguistic system. In some cases, attitudes appear to be at least as strong a factor affecting linguistic variation as any internal linguistic factors. Ferguson's (1964) description of diglossia focused some attention on these factors, 'any attempt to [use the H form for ordinary conversation] is felt to be either pedantic...or in some sense disloyal to the community' (1964:435). But the breadth and depth of the interaction between speakers' attitudes to non-linguistic factors and their linguistic performance has been investigated more consistently in fields other than linguistics.

In social psychology, Tajfel and Turner's (1979) work has been highly influential on the modern research program, insofar as it is generally accepted now that the analysis of communication must take into account the fact that interlocutors both have strong roles that are defined primarily in relation to groups and have roles that are defined primarily in relation to other individuals. It seems clear that this range of social and personal affiliations plays a complicated part in people's linguistic behaviour. Recognising this, though, is only part of the task. The challenge then lies in appropriately analysing people's behaviour in light of this, and in developing methods that enable researchers to explore the impact of the full range of social psychological motivations interlocutors employ strategically (whether consciously or unconsciously).

Collections on the social psychology of language and intercultural communication starting with, e.g. Scherer and Giles, eds (1979), St Clair and Giles (1980), have showcased a number of different approaches, some of which have been of more enduring significance because of their practical relevance to and potential application to issues of language variation, miscommunication and access to opportunities in everyday life. Initially, the greatest efforts were put into developing methods for evaluating intergroup factors, through techniques such as matched guise (Lambert et al. 1960) and subjective reaction tests (Labov 1972), and metrics for evaluating ethnolinguistic vitality (Giles et al. 1977). These continue to form the backbone of much research in language and social psychology.

But there is also a growing body of resources using methods and analysis designed to investigate more interpersonal dimensions. For example, work adapting Brown and Levinson's (1987) politeness theory to experimental frames has focused more closely on interpersonal communication (Tracy 1990; Lim & Bowers 1991; Wood & Kroger 1994), and recent work on processes of equivocation shows both that deceptive behaviour is particularised to the person the speaker wants to deceive (Buller et al. 1994) and is evaluated differently by the people it is tailored to than by third-parties (Burgoon 1994). Work on uncertainty management (e.g. Berger & Calabrese 1975; Gudykunst 1995) also specifically addresses the communicative behaviour of individuals, examining the rationale for and mechanics of the strategies people use to establish and solidify (generally new) relationships.

There are clear differences in both methodology and what is accepted as common ground depending on whether the researchers approach the interaction from the psychological tradition, or the linguistic one. So in the social psychological literature we find greater use

than we do in linguistics of questionnaires and evaluation protocols to investigate linguistic behaviour within and across language boundaries.

For instance, Lanca et al. (1994) use these methods to show that Portuguese immigrants to Canada demonstrate major differences in their perceptions of their relationships with other language and ethnic groups depending on the immigrants' dominant language. Sachdev and Wright (1996) used questionnaire protocols to identify correlations between an individual's social patterns and their choice of L2 in school, and Noels et al. (1996) use questionnaires about social contact between groups to examine differences in Chinese immigrants' acquisition of English. Even Bourhis' (1984) foray outside the lab uses an experimental design to test language behaviour (actual and professed) in the wake of legislative changes to the status of French and English in Quebec. In same language contexts, experimental scenarios and guided evaluations were used by Jones et al. (1994) to investigate differences in how student-teacher interactions were evaluated by Asians and Anglo-Australians. Millar (1994) uses a questionnaire to try and get to the heart of attitude differences towards regional accents by teachers in Catholic and state schools in Northern Ireland.

Similarly, Ryan's and Giles' on-going work on intralanguage variation in speech to and with the elderly (e.g. Ryan et al. 1991; Ryan et al. 1995; Coupland et al. 1991; Giles et al. 1993) makes use of experimental and response-oriented techniques. They have shown that even quite picayune linguistic distinctions are interpreted as encoding an enormous amount of attitudinal information by speakers and hearers, and can affect the quality of the interaction markedly. Ross and Berwick (1992) show that an interlocutor's attitude towards their addressee, and the amount to which they accommodate linguistically to their interlocutor can have palpable effects on not just the form of a conversation, but also its success.

The linguistic tradition of investigating language attitudes has historically been, and I believe continues to be, influenced more directly by field methods in anthropology and sociology. Hymes (1974) points to a long and close tradition linking anthropologists and linguists. Moreover, he points out that insofar as sociolinguistics continues to have at its heart the goal of enabling people 'to understand their lives adequately in terms of the true determinants of them' (1974:86), this interdisciplinary tradition continues to be highly relevant. Although much of current anthropology has shifted its emphasis from defining cross-cultural commonalities and attempting to identify cross-cultural universals, a good deal of sociolinguistics is still predicated on the assumption that aspects of human linguistic and non-linguistic behaviour are cross-culturally and cross-linguistically salient.

It is assumed, for instance, that even though linguistic signs may originally emerge in a Saussurean state of pristine arbitrariness, words, routines and even highly abstract grammatical or phonological variables quickly become sullied – acquiring and generating sometimes non-arbitrary meanings once they serve as the tools of a speech community (a point that has been well-explored in theories of signification in some current philosophical literature). Sankoff (1972) pointed out that a speech community's attitudes to and associations with Tok Pisin could be exploited by a skillful code-switcher to increase his personal standing in the community and others' evaluation of the truth of what he said. Jackson (1989 [1974]) documented the powerful significance of multilingualism as a marker of identity among the Vaupés. The case study she presents shows that even quite a small community may guard against erosion of linguistic diversity. Among the Vaupés, linguistic boundaries are an important way of delimiting acceptable and proscribed marriage partners (essentially, you cannot marry someone you share a language with). Sociolinguistics deals

constantly with notions of markedness, but markedness must now be defined as either meaning marked according to our understanding of linguistic structure, or marked according to our understanding of social structure.

Research in sociolinguistics indicates its debt to anthropology in its underlying fascination with the possibility of cross-linguistic generalisations. The search for generalisations across cultures may be out of fashion now in anthropology, and it may be under siege in sociolinguistics, but I believe that the reason sociolinguists persist in examining certain social categories for statistically significant correlations with language is not because they are too unimaginative to see the complexity in social structures, but rather because at heart they are looking for the largest possible picture. Nagy (1996), for instance, provides a particularly explicit articulation of this. She urges that the investigation of contact-induced language change be systematised through the investigation of factors found to be common to the largest number of previous studies.

This tendency is also revealed in most introductory sociolinguistic texts, which usually have at least one chapter each on the interactions between speaker gender, speaker age and the organisation of social classes with both stable linguistic variants and language change in progress (e.g. Holmes 1992; Chambers 1995). Sociolinguists who have recently attempted to distance themselves from this unspoken research program and who have tried determinedly to take a less macro-analytic approach to variation in the speech community (e.g. MacElhinny 1992; Mendoza-Denton 1994; Hall 1996; Schilling-Estes & Schrider 1996; Bucholtz 1997; Fought 1997; Wolfram & Hazen 1997), merely serve to emphasise the continued connections between sociolinguistics and anthropology. The analytical trends in this work have been greatly influenced by recent shifts in anthropology. Gellner (1995) and Geertz (1995) discuss how in anthropology, the theoretical goals have shifted far from the universalist perspectives of Frazer and Lévi-Strauss and even from the intrasocietal, but nonetheless, generalising goals of Malinowski, to the individual relativism exemplified in Geertz (1988), Bauman and Sherzer, eds (1989:xi), Duranti and Goodwin, eds (1992), Besnier (1996), Leap (1996).

There are, of course, examples of linguists using more experimental techniques to investigate the relationship between perceptions of variables in the speech community, and their transmission (e.g. Graff et al. 1986; Niedzielski 1996, 1999). Likewise, some researchers in social psychology use natural data and ethnographic methods of analysis to develop and test models of the interaction between language and interpersonal and intergroup relations (e.g. Applegate & Sypher 1988; Cronen et al. 1988; Potter & Wetherell 1987; Carbaugh 1990). We must hope that further collaborations across the fields and continued interest in each other's work will only increase the extent to which methodologies and research programs develop and grow.

3.3 Language use on Malo: *man Malo i havem tu sot*

Let us consider, then, in more detail the patterns of language use on Malo, particularly with respect to what observed patterns and people's stated opinions tell us about the roles of Bislama and Tamambo in the community. The analysis of variation I subsequently undertake will benefit from this in that, as a consequence, I will be able to separate the Malo speakers of the corpus into four groups. These sub-groups will differ from each other in the degree to which members are active members in the Tamambo-speaking community. We will also see

that other social and attitudinal facts unite the different subgroups, in addition to their (non-) use of Tamambo.

3.3.1 A theoretical and practical disclaimer

It is important to bear in mind the instrumental purpose of this discussion of how people on Malo use Bislama and Tamambo for the work. I do not seek to provide an account of the norms and motivations governing code-switching in the Tamambo speaking community. I do not have the Tamambo skills necessary to even begin to attempt this. Naturally, many of my tapes with Malo speakers have tokens of code-switching, but I cannot exploit this data. My purpose, therefore, in recounting and examining people's reported patterns of language use in the community and my own observations of the same, is to attempt to reach a more richly textured understanding of the make-up of the speech community. This becomes part of the non-linguistic information relevant to the structure of the larger Malo community which can be used to look more closely for trends in the diffusion and transmission of linguistic factors varying in this milieu.

I also see this mode of scrutiny of the Malo speech community as but one preliminary step towards a genuine understanding of the way linguistic choices (including choices between and within languages) interrelate with speakers' personal and group identities on Malo. To pre-empt the obvious criticism of this entire chapter: it is true that the account I will give is rather one-dimensional. To report on patterns of language use and use this as a basis of forming subgroups might not seem to be true to the spirit of my comments about the reflexive and dynamic nature of communication just outlined (the spirit of which will raise its head again in the discussion of social variables, such as speaker sex, in §6.6.4). It is true that the subgroups I will finish up with are not always recognised as unifiable subgroups by the community itself. But I will attempt to show, in my discussions of the individual members of the groups, that the linguistic subgroupings make sense in other (social) terms too.

These subgroups are primarily a tool for a researcher's benefit. They shed a different light on the community's dynamics. It is a light that has linguistic and (as I will show) social hues the community itself recognises, and is one that I hypothesised might interact with or constrain patterns of linguistic variation. Further work in Malo might show that the subgroupings are not the right ones, being too small or too large, or held together by factors which I was mistaken in seeing as related. But this work is hardly the final word in the study of the Malo speech community, no more so than it is in the study of Bislama.

3.4 Summary: sources of data on language use and preference

The data I will use is drawn from two sources: it includes information that members of the community volunteered to me in conversation or that they gave as responses to direct questioning, and it includes observations and inferences that I made during the course of my fieldwork. That is, a combination of speakers' considered, on-record opinions and inferences made on the basis of speakers' less monitored behaviour.

Between three and six months after I made most of my recordings, I asked all the Malo speakers who contributed to my recordings a series of questions. This was not done as a formal questionnaire, but rather took the form of visits to people's homes that I later made notes from. I inquired about what language or languages they prefer using in different social domains and with different addressees and why. The other data consists of informal observations of speakers' language choice and the comments I sometimes heard speakers make about others in the speech community.

Gal (1979) found that when she questioned speakers directly about their domains of preference for using German and Hungarian in Oberwart she received a tidier view of the bilingualism in the community than her own observations suggested was actually the case. Similarly, I found that informal observations of actual behaviour on Malo revealed greater variability than speakers' own reports of their language use indicated. Bourhis (1984) noted discrepancies between what language (French or English) Canadians claimed they would use if asked for help by a stranger, and the language they actually did use.

Whether these differences in reported and observed language choices are conscious attempts to present an idealised official picture to others or not is unclear. It can certainly be construed as helpful to the questioner to have respondents pre-filter the data before responding, but there is evidence that a lot of communicative behaviour is controlled well below the level of consciousness, and cannot be retrieved even on direct questioning before or after an exchange between interlocutors (Ebesu & Miller 1994).

In the next section, I will first provide the fullest picture for the record, but then, like the speakers themselves, I will abstract away from most of the variability in code selection and deal with dominant languages for the speakers in the given domains.

3.5 Self-reports and observations: use of Bislama and Tamambo

Table 3.1 provides the full results from my direct inquiries about speakers' language use and preferred languages in different domains, and includes in brackets my observations about their performance when this added to the information they provided. Table 3.2 reduces the information in Table 3.1 to the primary language(s) used. This shows clearly how I have arranged the speakers and domains to form an implicational scale.

Only the language use and preferences of the Malo speakers over the age of 12 are reported here. The children under 12 years that I recorded were excluded from this summary. The children do not negotiate the language to be used in different domains and with different addressees as freely as adults do. For instance, when speaking to an adult, they reply in the language they are addressed in (to the extent of their competence in that language). All of the children in the families I lived with and got to know had lived in this area of Malo most of their lives. The vast majority were competent, though not balanced, bilinguals.¹ The children's patterns of language use are by no means uninteresting or uninformative, and they will be discussed at different points of this chapter, but it did not seem appropriate to incorporate them into Table 3.1–3.2 with the adults.

¹ I knew only one family where parents talked to the children only in Bislama. But this family is not part of the corpus.

Table 3.1: Languages used in six social domains (different settings and addressees) on Malo. Adults (over 12 years) only. Capital letters indicate speaker reports using B[islama], T[amambo] or F[rench]. Lower case indicates speaker observed using b[islama] or t[amambo]. Parentheses indicate reported or observed language use occasionally.

Speaker	F/S	H	C	P	R	W	L1
Alis	B	B	B(T)	b	B	B	B
Anita	B	B(F)	B(F)	B	B	B	R
Janette	B	B	B	b	B	B	M
Leipakoa	B(T)	B(T)	B	B	b	B	K
Lolan	B	B	B	B	B	B	E
Dien	T,B	B,T	B,T	t,b	b	B	A
Nina	B(T)	T,B	B,T	b,t	B	B	Er
Visi	T	T,B	T,B	T,B	b	B	T
Susana	T	T	T,B	T,B	B	B	T
Mesek	T	T(B)	T,B	B	B(T)	B	T
Vosale	T	T	T(B)	T,B	B	B	T
Leikitah	T(B)	T(B)	T	b	B	B	Ta
Livai	T	T	T	B	b(t)	B	T
Papa	T(B)	T(B)	T	B,T	B(T)	B	T
Tarip	T	T	T(B)	B	B,T	B,T	T
Obed	T(B)	T	T	B	B,T	B,T	T
Atesolo	T	T	T	T,B	B	B	T
Madelin	T	T	T	T,b	b	B	T
Rovi	T	T	T(B)	T	b	B,T	T
Mama	T	T	T(b)	T	T, B	B (t)	T

Key to domains:

F/S	with friends/socialising	H	at home
C	with children	P	at public events, large gatherings
R	in religion, prayer	W	with <i>ol woman nara aelan</i>

Key to L1 (speaker's reported first language):

B:	Bislama	K:	Nakanamaga (Emae)
T:	Tamambo	Er:	Erakor (Efate)
Ta:	Tangoa (similar to T)	E:	English
A:	West Ambae		
R:	Raga (Pentecost)		
M:	North Maewo		

All speakers are identified by pseudonyms which were chosen in such a way as to retain the flavour of the person's real name wherever possible. That is, speakers who use French or English names have been given French or English pseudonyms; people who use *kastom* or locally-marked names have been given Ni-Vanuatu names (though in two or three cases, the Ni-Vanuatu names I have used are associated with areas in Vanuatu other than where the speaker really comes from). Women's names are bold in the tables.

Note that the tables provide an indication of language use. This does not mean that people do not sometimes address a person in Tamambo in a domain where the speaker reports using Bislama. Some of the *woman nara aelan*, for instance, have enough passive command of Tamambo that people may speak to them in it, but they do not use it productively in that domain for one reason or another. The constraints on their using it will be discussed shortly.

Table 3.2: Primary language used in different social domains (Malo speakers over 12 years). (+) = Bislama; (-) = Tamambo. (For L1 codes, see Table 3.1.)

Speaker	F/S	H	C	P	R	W	L1
Alis	+	+	+	+	+	+	B
Anita	+	+	+	+	+	+	R
Janette	+	+	+	+	+	+	M
Leipakoa	+	+	+	+	+	+	K
Lolan	+	+	+	+	+	+	E
Dien	-/+	+/-	+/-	-/+	+	+	A
Nina	+	-/+	+/-	+/-	+	+	Er
Visi	-	-/+	-/+	-/+	+	+	T
Susana	-	-	-/+	-/+	+	+	T
Mesek	-	-	-/+	+	+	+	T
Vosale	-	-	-	-/+	+	+	T
Leikitah	-	-	-	+	+	+	Ta
Livai	-	-	-	+	+	+	T
Papa	-	-	-	+/-	+	+	T
Tarip	-	-	-	+	+/-	+/-	T
Obed	-	-	-	+	+/-	+/-	T
Atesolo	-	-	-	-/+	+	+	T
Madelin	-	-	-	-/+	+	+	T
Rovi	-	-	-	-	+	+/-	T
Mama	-	-	-	-	-	+	T

In Table 3.2, we see only the main language(s) speakers used in a particular domain or with particular addressees. Since I am interested in speakers' use of Bislama, a plus (+) sign indicates Bislama is primarily used in this domain, a minus (-) sign indicates Tamambo. In general, this table abstracts away from my observations of speakers, too. The exception to this is some data for women at public events (I will return to this in §3.6.6 below). In reducing the data to speakers' professed preferences, we gain some idea of what language they believe is appropriate or aspire to in this domain. I will discuss some of the interview data relating to beliefs about appropriateness shortly. Again, women's names are given in bold.

3.6 What are people reporting in their self-reports?

Some points in these tables raise further questions. In the following sections, I will look in more depth at the role English and French have and discuss why they appear where they do in Table 3.1. I will also discuss in more detail the nature of the specific social domains people responded to, and account for why more women than men report using Tamambo in public. And I will give a more detailed account of the kinds of family-specific factors that differences between some *woman nara aelan's* reports of language use highlight. Finally, I will return to the question of children's competence in Tamambo (alluded to above) as a factor in determining what language a person chooses to use.

In short, a mixture of instrumental and affective motivations will be introduced and discussed. A kinship chart of the Malo speakers is provided in Figure 3.1. This will allow readers to track the relationships between some of the speakers as they are introduced in the following sections for discussion.

3.6.1 On the meaning of 'English' and its use outside of school

An interesting detail in Table 3.1 is Lolan's claim to have spoken English when she was growing up at home. It is difficult to know how to interpret this assertion. Technically, Lolan counts as *woman Aneityum*, but her family moved around southern and central Vanuatu a lot when she was younger. Her primary schooling was on Tanna. She and her siblings learnt Anejom, the language of Aneityum, with differing degrees of fluency. Since she would not have learnt Anejom as a community language when she was a child on Tanna, we must conclude that her parents used Anejom to some extent in the home.

My conclusion was that what her assertion means is that her family spoke a lot of Bislama at home, but a Bislama that consciously targeted standard English. When I first met Lolan in New Zealand in 1990 her English was good, but by no means native.

There is a reasonable amount of anecdotal evidence which suggests that perceptions of Bislama and English are rather blurry. *Pace* my earlier comments about the generally positive esteem in which Bislama is held, one sometimes hears Bislama referred to as *rabis Inglis* 'rubbish English', or even just *Inglis*. Crowley (1994) notes that one of the problems in interpreting the results from the 1990 Vanuatu Census questions on language proficiency and language use is that a person may claim to speak *Inglis* and they may truly believe that they are speaking English, but no native speaker of English could immediately understand the language that person is producing. Crowley (pers.comm.) says that some years after

conducting fieldwork on Paama, a Paamese friend mentioned to him that his Bislama was so bad when he arrived, they had to speak English to him. This is not at all what Crowley himself remembers. He thought he was speaking Bislama, and was certainly not speaking English, but apparently his Paamese interlocutors were equally sure that they weren't speaking Bislama, and concluded that they were speaking English (even if Crowley could not recognise it as such).

We find other evidence that Bislama and English have fuzzy perceptual boundaries. The speaker identified as Madelin told me that if she hears her children use a Bislama word, she tries to correct them by giving the word in Tamambo. She told me she explains to the children that that is how it is said in *lanwis blong yumi* 'our (INCL) language', and that the Bislama variant is *lanwis blong Dorothy* 'Dorothy's language'. Dorothy Jauncey, who she was referring to, was an Anglophone researcher working on a grammar of Tamambo. This seemed to indicate that in Madelin's mind the boundary between Bislama and English is loose, or perhaps non-existent.²

I provide these examples as external support for my suggestion that when Lolani reported her home language was 'English', this report can be taken to indicate a Bislama targeted at English was spoken.

3.6.2 On the use of French outside of school

The second fascinating point that emerges from Table 3.1 is the use of French as a home language. In contrast to English, which none of the Anglophone Ni-Vanuatu reported using outside of the domain of formal schooling, we notice that Anita reported using French in the home. While only one person in the corpus, her use of French outside of work is indicative of a wider pattern in the community. If people had been schooled in French, they seemed much more likely to use it at home than English-educated speakers were to use English.

In Anita's case, this was perhaps unsurprising since she had been a teacher at a local French language primary school. However, to my knowledge none of the teachers at the English language school used English with their children, all used exclusively Bislama or both Bislama and Tamambo. Moreover, Anita was not alone among the Francophones in the corpus in this respect. Atesolo had been schooled in French and he also said that his father sometimes used French with him and his brothers (but that he personally preferred to answer in Tamambo). In addition, I sometimes heard Obed and Rovi's family, who had mainly been schooled in French, joking in French, especially in town. Shortly before I left, Obed took a Monday-Saturday job in a bakery in Santo where he said that he and his employer, a Vietnamese, spoke in French.

² Terry Crowley points out that Madelin might believe Dorothy has two languages, Bislama and English. This might be the case, the ambiguity lies in the Bislama *lanwis*, which refers both to the abstract concept of 'language' and also has the more specific meaning of 'vernacular language'. Since the conversation pivoted on the use of *lanwis* to mean 'vernacular language', i.e. Tamambo, I believe the more likely interpretation is that *lanwis blong Doroti* meant 'Dorothy's (vernacular) language', i.e. English. Yet, it seems most implausible that when she hears her children using an English-derived Bislama word, she perceives them to be speaking English. Consequently, I believe my interpretation holds, i.e. that this utterance shows a blurring of the categories between English and Bislama.

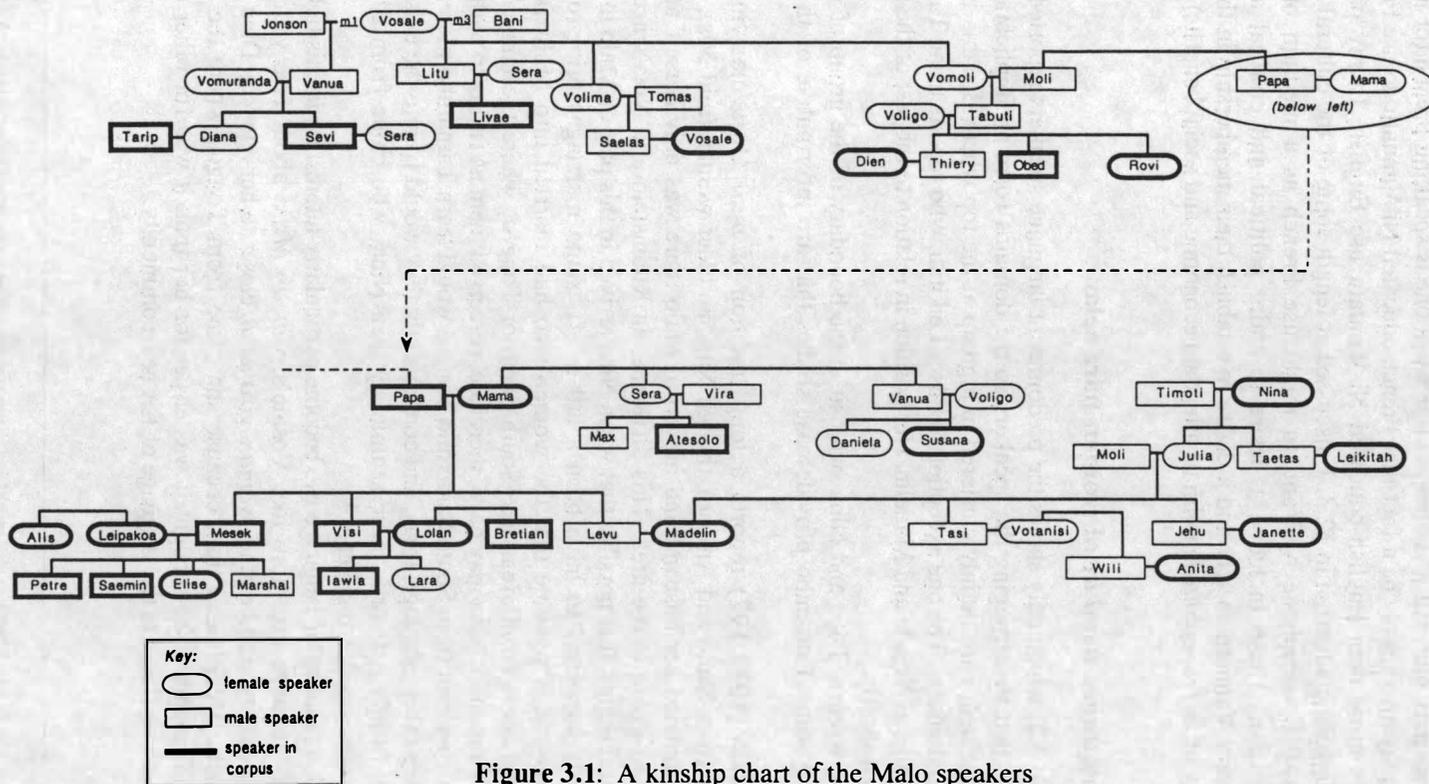


Figure 3.1: A kinship chart of the Malo speakers

There are several reasons why people may be more likely to report using French at home than English. The first is perceptual. As I noted, the perceived difference between English and Bislama may blur, but it is always clear when one is speaking French (or trying to). Of course, it may also simply be a fact that French-educated Ni-Vanuatu do use French among themselves more than English-educated Ni-Vanuatu use English. In my profile of the Vanuatu linguistic situation in §2.3, I discussed at length some of the cultural and political reasons why Francophone Ni-Vanuatu might use French as a medium of intergroup communication. These included a desire to claim political and cultural autonomy in contemporary Vanuatu society and a desire to establish their membership in the pan-global community of *la francophonie* (with its attendant economic and social benefits).

3.6.3 Languages used by ol woman nara aelan

In Table 3.2, which only shows the predominant language speakers reported, it becomes fairly clear that by arranging the speakers so the domains form an implication scale, the *woman nara aelan* are mainly clustered in a group at the top where Bislama is used in the most social domains. The one exception to this is Leikitah, who reports using Tamambo in as many domains as Vosale and Madelin. Her position in relation to the rest of the speakers will be discussed shortly.

The two women, Dien and Nina, who sit at the boundary of the group of *woman nara aelan* using some Tamambo provide case studies that are informative of the rest of the community.

Rubinstein (1978:197) reports a long tradition of *man Malo* marrying off-island (especially from Santo and Ambae), though Nina, an older woman (mid 50s), claimed that when she married her husband and moved to Malo there was only herself and one other *woman nara aelan* in the area. This difference in Rubinstein and Nina's accounts might mean several things. It is possible that when Nina arrived in this part of Malo in about 1960, there simply happened to have been a lull in off-island marriages prior to her arrival. Alternatively, it is possible that the women who had married into Malo families most recently had come from areas like South Santo or Tongoa, where the languages are very similar to Tamambo. Nina may have, therefore, perceived them as fitting in with the locals.

Certainly women from South Santo and Tongoa would learn Tamambo more quickly than women from other language areas, and consequently they would fit into the existing (on- and off-island) family networks more seamlessly than Nina, who came from Efate, and the friend she mentioned, from Emae.³

Leikitah's patterns of language use become particularly interesting in this light. Leikitah's choices of language are more like Vosale's, *woman Malo* about five years older than Leikitah, than they are like other *woman nara aelan* closer to her age (e.g. Dien and Janette). Leikitah herself explained that, because she came from Tongoa where the L1 is rather similar to Tamambo, she thought it was easier for her than it was for other *woman nara aelan* to become fluent in the language of her new community.

³ Which is not to say I believe she might have not realised they were from other islands. As noted in §2.3.1. after marriage women continue to be strongly identified with the island they come from.

At any rate, whether Nina genuinely was (by historical accident) only one of two *woman nara aelan* in this part of Malo when she moved there or not is probably not relevant. Even if she was not particularly isolated linguistically and even if there were more people (women and men) in the area she lived who spoke Bislama, other social facts make it extremely likely that she would have become fluent in Tamambo. Her age locates her firmly in the generation of women marrying *man Malo* who were expected to learn Tamambo. Other *woman nara aelan* mentioned that community expectations about this have been changing. For instance, the contrast between Leipakoa and Lolan, who married brothers, illustrates nicely the changes that have occurred in people's expectations of women's language learning. Leipakoa, who is in her late 30s, learnt some Tamambo (and can use it when necessary with older women who speak no Bislama). But Lolan, who is in her early 30s, never speaks Tamambo (except for the odd words or jests). Her passive command is good, but she has been able to live on Malo without speaking it.

Dien is much younger than Nina (in her late 20s) and has only been married to *man Malo* for two years, but uses Tamambo more than other *woman nara aelan* who have been on Malo for considerably longer. Like Leikitah, Dien's L1 gives her a bit of a head start (52% shared cognate vocabulary between her L1 Nduindui and Tamambo, Tryon 1976:68), but Dien's frequent use of Tamambo probably has more to do with non-linguistic factors.

The family village she lives in, which includes Obed and Rovi and neighbours Vosale and Livai's family village, uses Tamambo frequently in daily interactions. The people I interviewed from these two villages all spoke strongly about the importance of maintaining the local language. These families have a sufficiently large number of pre-school children between them that they have built a separate building to serve as a kindergarten. Most of the childcare there is conducted by young women like Rovi in both Tamambo and Bislama. Dien told me she was trying hard to learn and to use Tamambo (even though she can't always say what she wants to in it) because her two baby daughters have to know it. In the family Dien married into, we would say that Tamambo is treated as a core cultural value (Smolicz 1981).

Dien and Janette make an interesting contrast, since both have lived on Malo only briefly (two years or less), but they exhibit quite different patterns of language use. The main feature that distinguishes them does not seem to be any property or feature that is inherent to themselves. Rather, they are distinguished by the attitudes of their family group towards Tamambo and Bislama. The family Janette married into put less pressure on her to learn and use Tamambo with her children. Janette's husband, Jehu, was a teacher (in English), and teachers are subject to postings outside of their home area. Thus, Bislama is very likely to be their community language wherever they end up teaching. It is possible that Jehu and his family felt they were being practical in not placing undue pressure on Janette to speak Tamambo. Indeed, in 1996 Jehu and his family were posted to northern Espiritu Santo.

3.6.4 Language choices in religion

The social domains in which speakers choose between Tamambo and Bislama could also be arranged to form an implicational scale. In general, the convention is that people use Bislama whenever there is someone in the audience who may not understand Tamambo. This shapes patterns of language use for all speakers and is a major reason why the domain of religion and prayer is predominantly a Bislama domain, since church gatherings in the community I stayed in almost invariably included people from off-island.

In this Malo community, church services were sometimes held in Tamambo and there were some sacred materials (psalms and hymns) composed and printed in Tamambo. Generally, though, even when it was pre-advertised that the service on a particular Sunday would be in *lanwis*, the main *lanwis* component was the sermon and/or the children's address. Readings and hymns would still be in Bislama.

At the moment, the Bislama *Nyutesteman wetem ol Sam* 'New Testament and Book of Psalms' and church hymnals are the only substantial and widely-read publications in Bislama, and they are used throughout Vanuatu.⁴ In addition, the emphasis on inclusion in a religious domain makes the use of Bislama in church almost inevitable, though the norms here also differ denominationally. Protestant churches (both established denominations like the Presbyterians and Anglicans, and the newer ones, including evangelical ones such as Church of Christ and the Latter Day Saints) tend to use Bislama; Catholic churches use more French.

Virtually any gathering of more than twenty adults on west Malo would include at least one off-islander, and in some church settings this might even be the officiating minister (the local Protestant church had one pastor from Pentecost and one from Malo). Consequently, the notion that one should not exclude addressees or hearers by using a language they are not competent in, favours the use of Bislama in and around church services or events.

3.6.5 Children's exposure to and control of Tamambo

Two domains seem to bound subgroups within the Malo community. One of these is talk to children in the community, and the other is talk at public events or large public gatherings. These domains mark a transition between individuals who clearly prefer to use Bislama as an interpersonal code, and individuals who prefer Tamambo in interpersonal situations. In this section I will discuss what the language people choose to speak to children tells us about individuals, and in §3.6.6 I will discuss the data relating to language choice at public events.

A strong motivation people on Malo expressed for addressing children in Tamambo is that in doing so, they guarantee the language is safely passed on for the future. Moreover, it is believed that a child born on Malo needs Tamambo as a means of establishing themselves as a *bona fide* member of the community. However, despite this, a growing number of children in this area are either monolingual in Bislama or are heavily Bislama dominant.

In some cases, parents on Malo believe that Bislama will be more useful to their children than Tamambo (either directly, or indirectly by providing a head start to learning English). This was the rationale expressed by one of the local teachers, *man Malo*, who with his wife, *woman Malo*, had made the decision to use only Bislama with their three children at home. There was another couple from Malo who lived and worked in Santo at the time I was doing my fieldwork, who also only used Bislama with their children, despite the fact that they spent regular weekends on Malo with the children's grandparents and extended family. Again in this case, the mother felt that Bislama was more useful to the children, especially living in Santo, than Tamambo would be.

⁴ Since 1995 the Old Testament in Bislama has been published.

However, even the children who live in the village or family groups where a number of people continue to use Tamambo do not necessarily acquire full competence in the language. Aside from the social or attitudinal factors that may be contributing to the shift towards Bislama among younger speakers, the situation is not helped by some internal factors of Tamambo. Younger speakers' lack of fluency in the language may be partly due to fairly high levels of insecurity and self-consciousness about their linguistic abilities. And this insecurity may, in turn, be influenced by ongoing changes to the Tamambo spoken on west Malo today.

Rovi, for instance, said to me once rather sadly that she and other young people on Malo don't know how to speak Tamambo properly anymore. When asked what she meant by this, she explained she meant that her Tamambo vocabulary wasn't as large as older speakers' was, and that even when she and her friends knew the Tamambo word or phrase, they would often use a Bislama one instead. And she mentioned the different way young people pronounce some words from older speakers. This had also been commented on jokingly once in the market in Santo when I had asked some of the women from Malo the name of a particular fruit. An older woman replied [naɣaviɣa], and Rovi replied [naɣaviga].

As noted in §2.3.4, fortition of /x/ to [g] is widespread among children, teenagers and young adults who lead more traditional Malo lives (e.g. have stayed to work in the gardens) on west Malo (Jauncey 1997). Jauncey has also noted children using [k^h] where older speakers would use [g]. She informally observed that some of the more conservative speakers of Tamambo commented disfavouredly on the variants of [g] used by younger speakers, and she reports (pers.comm.) that the voiceless aspirated stop (apparently the most recent innovation) was evaluated more negatively than the voiced stop. It is quite clear, then, from the comments Jauncey and I have noted that this sound change is something the members of the speech community consciously evaluate, and that the marked status of the stop variants is the cause of some discomfort and linguistic insecurity for some younger speakers such as Rovi.

Other phonological changes are occurring in Tamambo, but are subject to less conscious awareness among speakers.⁵ I would suggest that combined with many children's limited exposure to Tamambo (for societal reasons) and their therefore limited command of basic Tamambo vocabulary, the uncertainty fostered by these changes conspires to make it all the more likely that they will be Bislama dominant. My observations of talk between and among children without adults present was, perforce, limited to chance eavesdropping, but one conversation that I overheard between some of the children in my village may illustrate what I mean by this.

During one visit, I was sleeping in a room next to a tree swing and I was able to overhear conversations among the children playing on the swing. One afternoon I heard Madelin's two children, whose parents deliberately tried to speak Tamambo to them, playing with the younger children in Mesek and Leipakoa's Bislama-dominant household, Saemin and Elise. The conversation turned into a lesson on Tamambo vocabulary.

⁵ I once listened to several Tamambo speakers in their thirties argue over whether there even was a distinction between the nasals that I will shortly discuss.

In addition to the fortition of /x/, the language is also undergoing a merger of /m^w/ and /m/. This merger has created homophones out of some fairly high frequency lexical items, e.g. /m^wata/ 'snake' and /mata/ 'eye'. The conversation between the four children playing on the swing involved a fair amount of switching between Bislama and Tamambo, but at one point they began to discuss people's names. Saemin, at 10 years, was the oldest of the four, and he explained to the younger three that the name of one of the pastors was Tamambo for 'snake'. This was greeted with much horror and amazement. Elise in particular said that she thought that was a terrible name for a pastor to have, but Saemin assured her it was absolutely true. In fact, the pastor's name meant 'eye', and the confusion would only arise for speakers of Tamambo who have lost the /m^w/ and /m/ distinction. Since this merger also appears to be age-graded (though closer to completion than the shift of /x/ > [ɣ] > [g]), naturally any confusion, when it does arise, is most likely to be a problem for younger speakers.

This exchange between the four children tells us a great deal. It indicates that the children in this village are growing up bilingual, but that they are not balanced bilinguals. Any child of eight, like Elise, growing up in a balanced bilingual community should know high frequency vocabulary like 'snake' and 'eye'.⁶ And Elise was by no means the most limited speaker of Tamambo among her age peers. The younger children's misunderstanding may have been exacerbated by the ongoing merger between /m^w/ and /m/, but the root of their problem was a lack of basic proficiency in Tamambo. It is easy to trace a path between Elise's reduced language skills, her surprise at the gap in her knowledge, and Rovi's insecurity (eight years older) in her overall competence in Tamambo.

The question is, then, what speakers are to do faced with both the objectively poor Tamambo of some children in the community and the low subjective evaluations even a more fluent speaker like Rovi has of her Tamambo abilities. Even while they recognise the practical implications of younger speakers' restricted Tamambo, they may have strong affective reasons for wanting to use it with the children of the community.

Table 3.2 suggests that a third of the adults who use both Tamambo and Bislama (5 out of 15) are unsure about what to do. They resolve the dilemma by vacillating between the two. Thus, the language that speakers report choosing with children reflects a tension between the speaker's own competence and the competence of their addressee. It also reflects tensions between their perceptions about the importance of local *kastom* and local identity, and the increasingly pervasive pan-Vanuatuan identity transmitted and negotiated through Bislama.

3.6.6 Women's and men's different experience of 'public events'

A perhaps surprising result in Table 3.2 is that more women, including some *woman nara aelan*, reported using Tamambo in public events or large gatherings, while *man Malo* reported using Bislama. Examples of public events or gatherings are weddings, village meetings, the Independence Day celebrations. Table 3.2. shows that of the Malo speakers who used both Bislama and Tamambo, seven out of eight women reported using Tamambo at public events and six out of the same eight reported using Bislama. But only three out of

⁶ Measuring high and low frequency words in Tamambo is tricky, since there is no corpus of spoken Tamambo by which to quantify this. Because brown tree snakes are found on Malo, and rather disliked, I assume all these children would have come across the concept of them if not the fact.

the seven adult men in the corpus reported using Tamambo at large gatherings, and all seven said they would use Bislama.

Interpreting this result is not entirely straightforward, though, since the role of women and men is so different in public gatherings. Women are almost entirely excluded from agentive roles in the public domain and seldom address a public group of listeners. Rubinstein notes that women on Malo 'do not and did not [in the past] enter into the public, political, and economic, hierarchically-ranked realm of male competition and interaction' (1978:315). Local perceptions are of men 'as activists, as doers, as owners, as social creators and activists, and as transactors' (1978:319), while women's activities are considered 'more simple, more clearly defined [and] politically private' (1978:320).

Given this difference in men's and women's experiences of the public domain, I fear that the way I posed my questions about language use at public events was not sensitive enough to draw out this role distinction. Generally, I asked people something like, '*Sapos i gat fulap man i stap, olsem long saed blong ol bigfala samting, bae yu toktok long wanem? Bae yu toktok lanwis o Bislama o bae yu miksim?*' ('Say there were a whole lot of people around, like at some big event, what would you talk in? Would you speak Tamambo or Bislama or a mixture?').

Obviously, the differences in women's and men's social roles in public mean that this question would have evoked quite different images of addressees and talk type for a woman than it would have for a man. Women's addressees and talk types at large gatherings would have more in common with their addressees in more domestic domains, e.g. when talking to friends or socialising. For many women, the public gathering they would be most familiar with would be market on Malo or in Santo where business is conducted with and surrounded by friends (though, of course, in Santo Bislama is necessary for communicating with purchasers and sellers from other areas).

This is a rather different social situation than the ones men would have had in mind when asked essentially the same question. The roles they would have positioned themselves in to answer the question would have been speaking to the gathered community at celebrations for Independence Day, or making announcements after church services, during a *bringanbae* 'bring and buy', or reporting discussions on the proper ownership of local land. I only saw a woman make public announcements of this nature twice. Once was when the woman leading a group that was fundraising for Christmas spoke after church to ask people not to go to a dance up-island on the same night as her group was holding a *bringanbae*, and the second time was when the same woman, in her capacity then as school principal, reminded the community after church that a parent-teacher meeting was planned for that week.

This difference in the nature of women's and men's public gatherings may account for why seven women and only three men reported using Tamambo at these events, and why so many *man Malo* reported using Bislama in large gatherings. I believe it is also why in Table 3.1 there are only observations about the languages used at public events for five women in the corpus. I realised, looking at my notes after talking to them, that their replies had been framed in terms of whether they were talking to friends and relations from Malo or *woman nara aelan*, and often examples were given of what they would do at market on Malo or if a large group were walking a long way together. Women's reported linguistic behaviour in public gatherings is more like their reported linguistic behaviour in the domestic domains of talking to friends than men's behaviour is. This is for the very good reason that usually only men have negotiated a separate, truly public identity.

3.7 Identifying subgroups: shared norms for choosing a language

Let us return to the information in Table 3.2 (reprinted here for convenience). I have added divisions at the points at which there seems to be a clear shift between speakers' preferred or target language in different social domains. I have also reproduced the family tree in Figure 3.1 for ease of reference.

This method of dividing the speakers into subgroups is similar to the principal components analysis of data pioneered in sociolinguistics by Horvath (1985). There has been a recent return to using these methods (also known as cluster analysis) by other researchers, e.g. Garrett et al. (1997). The point of approaching the data in this manner is that the researcher avoids trying to squeeze the data into a fit with their preconceptions about individual speakers in the speech community. In this case, not only do we find that subgrouping the speakers is possible, we find that the subgroups consist of people who form something akin to an attitudinal or behavioural natural class.

Table 3.2: Primary language used in different social domains (Malo speakers over 12 years). (+) = Bislama; (-) = Tamambo. (For L1 codes, see Table 3.1.)

Speaker	Age	F/S	H	C	P	R	W	L1
Alis	20	+	+	+	+	+	+	B
Anita	26	+	+	+	+	+	+	R
Janette	30	+	+	+	+	+	+	M
Lolan	31	+	+	+	+	+	+	En
Leipakoa	38	+	+	+	+	+	+	K
Dien	24	-/+	+/-	+/-	-/+	+	+	A
Nina	60	+	-/+	+/-	+/-	+	+	Er
Visi	34	-	-/+	-/+	-/+	+	+	T
Susana	14	-	-	-/+	-/+	+	+	T
Mesek	45	-	-	-/+	+	+	+	T
Obed	18	-	-	-	+	+/-	+/-	T
Livai	24	-	-	-	+	+	+	T
Leikitah	26	-	-	-	+	+	+	Ta
Tarip	31	-	-	-	+	+/-	+/-	T
Papa	65	-	-	-	+/-	+	+	T
Atesolo	18	-	-	-	-/+	+	+	T
Madelin	26	-	-	-	-/+	+	+	T
Vosale	31	-	-	-	-/+	+	+	T
Rovi	16	-	-	-	-	+	+/-	T
Mama	60	-	-	-	-	-/+	+/-	T

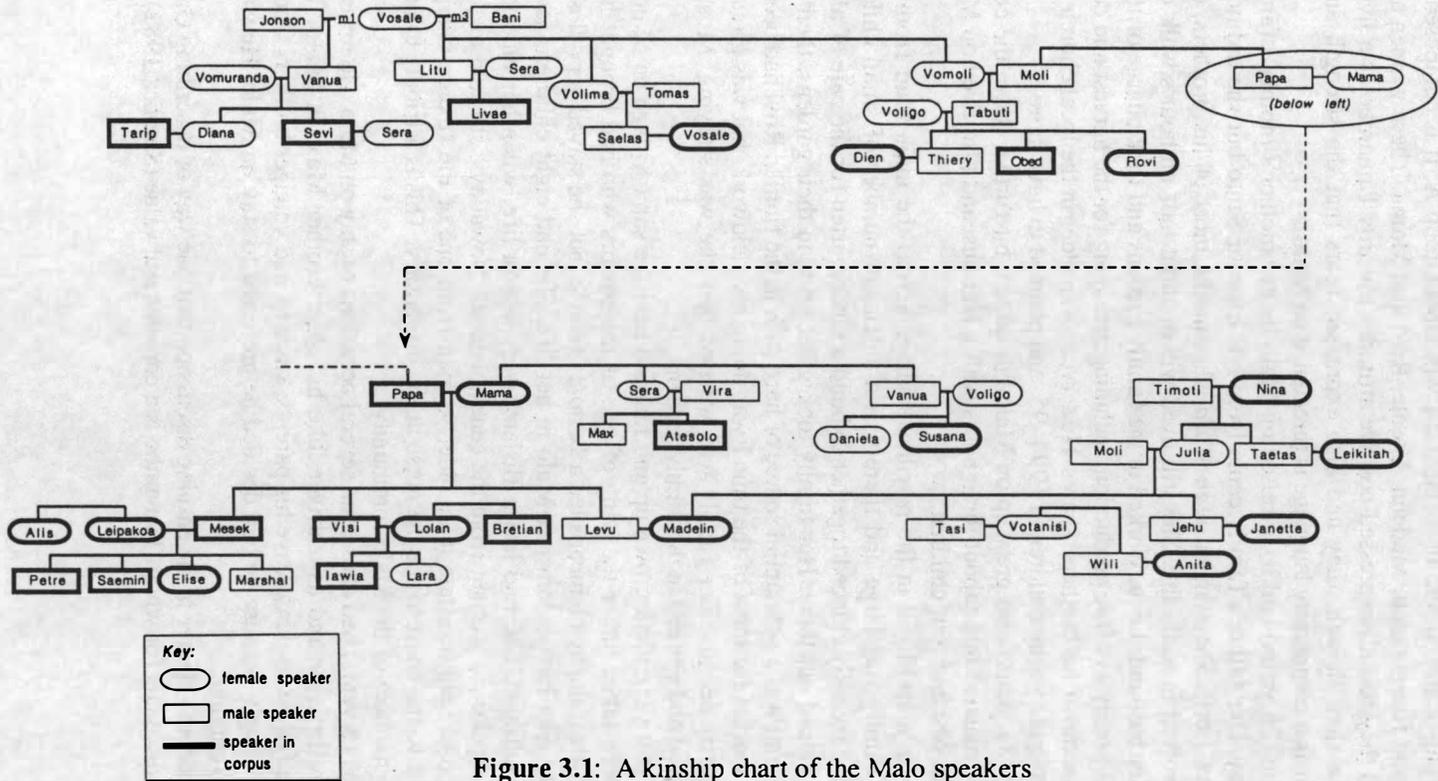


Figure 3.1: A kinship chart of the Malo speakers

3.7.1 Group A: the Malo hardcore

The group at the bottom of Table 3.2 I will label Group A. It is composed of one man, Atesolo, and four women, Madelin, Vosale, Rovi and Mama. These women are all *woman Malo* and they had all expressed positive attitudes towards Tamambo as a living language. More than that, though, they had also expressed fears that the strength and vitality of Tamambo as a community language is threatened on Malo.

Madelin (26 years) had been raised on Malo by her mother's mother's family since she was a baby. Her father's family came from NE Espiritu Santo but she did not speak their *lanwis* (Pot Lori). She married *man Malo* who, unlike most of his brothers, chose to stay firmly involved in Malo life. She still lived within about half an hour's walk of her family village. Her husband, Levu, worked in the family gardens and took charge of the drying and sale of the family's coffee production, including arranging for the harvest and completing the drying of some of his brothers' shares. He served as an elder in the local church. Madelin and he had two quite young children in 1994–95, and planned to have more.

Vosale (31 years) also grew up on Malo but quite a bit further down the coast from the village she married into (about three hours walk). Her husband also lived on Malo, working his gardens. She had four children.

Rovi (16 years) lived in the patrilocal village across the main road from Vosale. Her extended family was large and there was a sufficient number of small children that the brothers had recently clubbed together to build a kindergarten for the care of all the children and Rovi helped out there. Her family took great pride in their gardens, their produce was wonderful and was a substantial source of hard cash in the family. Rovi had been a regular at the Santo market for most of the time I was doing my fieldwork, but was grounded towards the end partly because her family was worried that she was straying. Most of her *stret* brothers lived and worked in the family gardens.

Mama (60 years) had grown up and married near the same village on Malo. As a child, she had been taken under the wing of one of the pastors who had brought her up. Partly because of this, she by no means led a *kastom* lifestyle, but she stayed very close to the local community. She has seldom left Malo in her life. She had eight children, seven of whom lived to adulthood. She had learnt Bislama very late in life, when her oldest son, Mesek, married Leipakoa, a woman from the central islands. Nowadays, she told me, she might be called upon to help translate if an older woman from one of the *kastom* villages up-island came down to the coast to see the nurse at the dispensary. This is a tribute to her exemplary status as a member of the Malo community.

Atesolo (18 years) had chosen on several occasions to stay on Malo. He refused job offers as a football player and as a driver, like his older brother Max, in Santo. This was not because he felt he couldn't leave his parents, since he had younger brothers who were still at home, but simply because he had decided he preferred to stay on Malo. He was unmarried and had no children.

It will become clearer in the ensuing discussion that one way of describing Group A is that they are individuals for whom Tamambo is a core cultural value (Smolicz 1981).

3.7.2 Group B: Malo mainlanders

The next group up, Group B, is composed of four men, Obed, Livae, Tarip and Papa, and one woman, Leikitah. Two of these men have travelled within Vanuatu and worked off Malo for short periods, and Leikitah is a *woman nara aelan*. The people in this group are different from Group A, in that they generally say they use Bislama at large public gatherings (though since Group B is mainly men this is not surprising, see the discussion above). They are distinct, though, from the next group of speakers above them (Group C) in that they profess to use Tamambo with children in the community. They also signal a greater commitment to the use of Tamambo than Group C in that some of Group B say that they use Tamambo in their silent prayers sometimes and try to use it with *woman nara aelan* after they have been on Malo for a while.

Obed (18 years) was Rovi's *stret* brother. At the start of my fieldwork, he lived and worked the gardens on Malo but he aspired to see a bit of life, and took a job in a Santo bakery in mid-1995. He was unmarried and without children, so the main reason he came back to Malo on the weekend was to spend time with his parents and siblings. While he maintained quite active ties with members of his extended family who lived in more *kastom* villages up-island (regularly going up to visit them), he was also obviously very curious about city-life and asked me many questions about New Zealand and the customs there.

Livae (24 years) was unmarried, but a bit of a player. He sometimes worked for one of the Protestant churches as a sort of lay counsellor, and would be sent with materials to distribute and to organise meeting groups in various parts of Espiritu Santo. He had very much hoped to leave Malo to do a church training program either in Fiji or Australia in 1995, but plans fell through.

Tarip (30 years) had worked as a builder in the northern islands, sometimes as a subcontractor to expatriate builders. However, since marrying a *woman Malo*, whose father was a respected leader in the Malo community, he had stayed on Malo and was living off his family's land. His crops were fairly traditional, the main cash crop still being copra. He also served as a church elder and had two small children.

Papa (65 years) had lived in two places on Malo. He had been alienated (as in: run off at gun-point when he tried to contest ownership) from his *kastom* land by a French planter during the Condominium and bought the land he and his wife and sons now lived on the west coast of Malo. He had spent most of life working on Malo, but had learnt his Bislama as a young man during WWII when he carried water up the hill for the Americans who kept a watch station on the hill of Malo. These days, he works every second week as mate on the ferry service his son Mesek runs.

Leikitah (26 years) was a second generation *woman nara aelan* in the family she married into. The mother of her husband, Nina, was also *woman nara aelan*, and it should be noted that Nina was one of the few *woman nara aelan* that people from Malo said they could and did speak Tamambo with. As discussed in §3.6.3, Leikitah's L1 was similar to Tamambo, and she thought that it had been easier for her to learn Tamambo as a consequence.

3.7.3 Group C: ties that bind (off-island)

The third group, Group C, consists of three women, Dien, Nina, and Susana, and two *stret* brothers Visi, and Mesek. As a generalisation, these are people who are even more oriented to lifestyles or work or family off Malo than members of Group B. This group differs from Group B in that they use Bislama when talking to their own and other people's children.

Dien (24 years) is a first generation *woman nara aelan*, in her family. She is married to one of Rovi's and Obed's *stret* brothers. She has only lived on Malo for two years, but is learning Tamambo and uses it at home and with her children where she can. Since her Tamambo is not that fluent yet, she says she sometimes simply doesn't know how to say what she wants in Tamambo. She had two very little children, and claimed to still desire to pursue employment in Santo at some stage. At this time, her work was limited to the gardens and the markets.

Nina (56 years) was the *woman nara aelan* who was Leikitah's mother-in-law. She had worked as a nurse in her youth, but since her marriage had stayed on Malo leading a more *kastom* lifestyle. She was sometimes called on to act as midwife in the larger community, and was active in social and women's groups. This included an all-woman string band on Malo, which in itself was a little iconoclastic (generally the bands are male), who played on a mixture of traditional string band instruments, such as a string bass, and ukulele, as well as non-traditional instruments like the spoons and coconut shells. She also periodically went to stay in Santo or even Vila for meetings of the women's groups she was involved in. Nina had been widowed for some time and her two sons were married.

Susana (14 years) lived with Visi and Lolana as their *haosgel*, since Visi worked in Santo during the week and Lolana taught at the Malo school. Because of this arrangement, and because Visi and Lolana did not do a lot of work on the gardens they had rights over, her work in the gardens was limited to giving a hand in her own family's gardens during the weekend sometimes.

Visi (34 years) and Mesek are brothers and both have regular work commitments that take them off Malo to Santo. Visi works full-time during the week in Santo and has two children who live with his wife, Lolana, full-time on Malo.

Mesek (40 years) runs a 12m fibreglass launch as a ferry between west Malo and Santo. A common routine for him is to take the ferry to Santo in the morning, spend the day in town and then bring the boat back in the late afternoon. Mesek also serves on the Santo-Malo regional council, which requires him to be in Santo for meetings. He and Leipakoa have three children of their own and have adopted three children of one of Leipakoa's *stret* sister's. Both Mesek and Visi have travelled overseas at some period in their lives.

3.7.4 Group D: woman nara aelan

The fourth group is very homogeneous even in their diversity. Alis, Anita, Janette, Lolana, Leipakoa are all first generation *woman nara aelan* in their families. Some of them have learnt some Tamambo, but none use it regularly and not as the language of choice with their children. They are essentially speakers of Bislama.

Alis (21 years) is Leipakoa's *stret* sister. She came to live on Malo when she was still in primary school after her mother died when they were living in Vila. Her first language was Bislama and though she learnt some Tamambo she said that she actually tried not to. She meant this as an act of resistance, as an indication that she did not intend to stay on Malo forever. Now, she seldom uses Tamambo, except with Mama and some of the other close members of the family who, she says, won't laugh at her. More than ten years after moving to Malo, she fluctuates between being resigned to her life helping her sister out in the home and working the family gardens, and telling me that she really had to get out of Malo and wanted to go work in Vila. She is not married, and expressed little desire to be responsible to (or for) someone else.

Leipakoa (38 years) was married to Mesek. When she arrived on Malo, she started to learn some Tamambo but since Lolan arrived in the same patrilocal village and since a number of other *woman nara aelan* have married in to the general area, there has been less social or pragmatic pressure on her to use Tamambo. Like Alis (but on a separate occasion), she said that a major disincentive to using Tamambo was having people laugh at your mistakes. She would use it with Mama and Papa, though, around the village.

Lolan (31 years) is married to Visi. She never uses Tamambo: even when it is clear she is following a conversation in Tamambo between Visi and one of his uncles, she makes comments in Bislama. She is a teacher at the local primary school, and told me that sometimes she would use a little Tamambo with her class if she wanted to try and explain something they were having difficulty with. Again, for her, the matter of saving face seemed relevant. The children in her class would not ridicule or correct her if she occasionally used (incorrect or accented) Tamambo there.⁷ Lolan's family were quite active in politics nationally (both in government and in the church). Lolan was the person I first met in New Zealand and who acted as the primary broker for me in the Malo community. Lolan is an active member of the west Malo community, both as a teacher and in the local church. She socialises primarily with other *woman nara aelan*.

Anita (28 years) was a teacher in the French language system (though for much of the period of my fieldwork she was suspended on account of the strike and was working her husband's gardens again). She had four children who she wanted to educate well. Although she lived on Malo and her husband kept to traditional activities like growing crops, they were diversifying considerably. In addition to the copra which everyone continued to produce, and the coffee which some other families had moved into, she and her husband were growing vanilla, a very labour-intensive but also very lucrative crop.

During the period she was unemployed, Anita built her own concrete wood-fired oven and learnt how to make bread from an Australian Presbyterian women's auxiliary pamphlet. She then contracted with the local *nakamal* to provide them with bread every day and had brisk sales of her daily excess.

⁷ As Canagarajah (1995) shows in a Tamil classroom, this sort of code-switching, particularly when it involves a situationally proscribed language, is an important cue that children use to learn what the significance of code-switching is in their community and the associations with a particular language and the relative position of each language.

Janette (30 years) was married to Jehu, a teacher who was a brother of Madelin's. During the strike he lost his job and he and Janette had come back to Malo with their children to resume a *kastom* lifestyle. During this period, they lived in his family village and Janette worked the gardens and raised their three children (they now have more). Janette's ties with her family on Maewo were still very strong. She was very close to her mother and her grandmother and had gone home to her mother to give birth to her first child (which caused all sorts of problems with Jehu's family). She found her position as an outsider in her Malo family difficult and socialised a lot with other *woman nara aelan*, though she was also a frequent visitor at Madelin's.

3.7.5 Summary of Malo subgroups

The four groups just outlined do not directly correspond to any single social subgrouping on Malo. The social group *ol woman nara aelan* has independent status on Malo, but not all the *woman nara aelan* show the same patterns of language use. Nonetheless, the four groups have some internal coherence despite some individual heterogeneity. They will be included as a non-linguistic factor when evaluating variation in the use of null subjects in Bislama.

3.8 Motivating language choice: the instrumental *is* affective

Gardner and Lambert's (1972) longitudinal study of the performance of language learners in Canada provided an influential model of the role individuals' motivations have to play in their ability and fluency in a second language. They define motivations as being an interaction of a number of attitudes held by the individual language learner, which they grouped into two categories. The first were instrumental motivations (reflecting how useful the learner perceived the language would be to them); the second were integrative motivations (reflecting the learner's desire to become part of the society or culture that they perceived the language to be a gateway to). These rough distinctions between pragmatic and affective attitudes have surfaced in others' work investigating the hallmarks of successful language acquisition, e.g. Spolsky (1969) and Schumann (1976, 1994) and form an important basis for empirical investigations into code selection among bilinguals (Yamamoto 1995).

Affective and instrumental motivations continue to be relevant even if fluent levels of bilingualism have been achieved. Where research such as Gardner and Lambert (1972) focused on the success or failure of learners to acquire a second language, as this chapter has shown, bilingual speakers must still consciously or subconsciously choose which language they will use when. As Canagarajah (1995) neatly points out, the relationship between code choice and psychological motivations is commutative. Affective and instrumental needs may motivate the acquisition and exploitation of code-switching as a linguistic strategy. But in Canagarajah's case, since switching between Tamil and English in Sri Lanka is so expressive, it is equally true that in the process of acquiring and trying out code-switching young speakers learn which personal and social motivations are acceptable to express and what the appropriate means are for expressing them.

Myers-Scotton (1993b:91) suggests that the social significance of code-switching lies in the variation between the marked and unmarked norms of a speech community. However, her attempt to formalise code-switching is rather less successful with the non-linguistic factors than the linguistic ones. She frames a speaker's choice of code as an index of their rights and obligations (1993b:84), but she also recognises that social rights and obligations may sometimes be at odds with personal rights and desires. Her attempt to chart a more rigorous path through the social motivations of code-switching is welcome. But it starts shipping water when the core maxim has to be relativised to include interpersonal motivations that may be at odds with the norms of the speech community (1993b:114). The social and psychological insights her framework provides really take us no further than principles which have been long agreed upon or established.

But with respect to the discreteness of affective and instrumental motivations, Goffman (1974:519ff.) points out that these categories often overlap. He suggests that, at the same time as speakers are conveying straightforward referential information, they convey considerable information about their personal attitudes and state of mind. Moreover, he suggests that speakers actively exploit this possibility. For instance, the instrumental function of a deictic (locating the utterance with respect to space, time and causality) can be, in his terms, 'laminated' with information about the speaker's willingness to take responsibility for the content of their utterance. (This lamination of instrumental and affective motivations is particularly apparent in some of the work on obstacle avoidance, e.g. Roloff and Janiszewski 1989 define an experimental obstacle as either an addressee's unwillingness or inability to comply with a request.)

This is highly relevant to the ways speakers decide which language code to use, and with whom. It is certainly true that we can classify motivations that a bilingual speaker may have as those which are largely practical or pragmatic, and those which are more personal or sentimental. Some of these have been explicitly mentioned in the discussions of individuals above.

A person on Malo may decide to use Bislama or Tamambo after asking themselves: 'Will my interlocutor understand me?', or 'Do I know the vocabulary to express what I want to say?'. Or they may wonder 'Which language is more likely to get me access to the resources I desire?'.

But conversely, the speaker's internal processes may reflect dilemmas more like: 'Do I want to stress the commonalities or emphasise the differences between me and my interlocutor?', or 'Which language is the more beautiful or appropriate medium for conveying my information?'. Or they may make their decision based more on something like: 'Am I especially proud of one of my (personal or group) identities? How best can I show this?'.

In practice, though, instrumental and affective motivations hardly form a rigid dichotomy. For many people, it is hard to separate even something that seems to be a highly practical consideration, such as: 'Will my addressee understand me?' from an affective consideration, such as: 'Do I care if my addressee (or anyone else listening) feels excluded because they can't understand me?'.

Having acknowledged the interdependence of these factors, I will continue to use the terms instrumental and affective as discrete terms. This results in some themes recurring in the discussion, but I believe that this degree of circularity is an accurate reflection of the complexity of the local and supra-local factors influencing the linguistic situation on Malo, and the complexity of the individual's task.

3.8.1 With us or against us: collective or group factors affecting code choice

Some opinions and attitudes about language use in the community were relatively easy to solicit and evaluate, especially people's attitudes and opinions about the linguistic behaviour appropriate to the domains of school and community affairs. In the next sections, I will present some case studies as examples of how some of the feelings about Bislama and Tamambo were demonstrated in the Malo community and how tensions centring on language could become intergroup and interpersonal issues.

Sometimes people made their opinions about language issues apparent through comments obliquely referring to language. These sorts of conversational clues indicated some of the factors that weighed heavily with them in choosing a language in different social situations. But occasionally there were public discussions specifically concerned with the most appropriate language choices in different domains.

One such debate played out in a meeting after church. A number of people had gathered for the week's events to be announced. Usually these announcements were made in Bislama – following the general principle that if anyone in your audience might not understand Tamambo, it is nicer for a speaker to use the lingua franca – and on this occasion the announcements started in Bislama, too. One message reminded people that a church committee was soliciting new verses and musical settings for a revision of the hymn and song book. The speaker noted, in Bislama, that submissions were sought in Bislama and Tamambo. Following this, two members of the committee stood up in turn to speak. Both were respected elders in the church and community. The first to take the floor was a man who is equally fluent in Tamambo, Bislama and English (having worked in Australia for some years when he was younger). He spoke about the committee's work a little in Bislama, but mainly in Tamambo, and he said they especially welcomed submissions in Tamambo.

The second speaker then got up and spoke vigorously and at some length, entirely in Tamambo. He also spoke in support of the committee's work, but his decision to use only *lanwis* in this social situation was invested with considerable meaning in itself. I interpreted his point as being that the church group and the community work should be primarily conducted in and in support of Tamambo. There was a clear, if unspoken, implication that speaking and working in Bislama was at least less appropriate to this enterprise and perhaps was even antithetical to the more important community goals. Indeed, this interpretation was confirmed on the walk home, when people said that the gist of this elder's message had been the significance of Tamambo for all *Natamambo* 'people from Malo'.

In other words, although speakers agree that community affairs are most appropriately conducted in Bislama (and some people, as we will see, said that even when a group was made up entirely of *man Malo*, they would still prefer to use Bislama for public business), there are times when local concerns and the salience of island identity are more important than the norm which includes all listeners, and are more important than any urbanity or social standing that the speaker might accrue by using Bislama.

It is possible that there were frequently cases where speakers were obliged to negotiate or contest their code selection based on the different priorities they gave to social and psychological factors. But as far as I know, these sorts of mismatches in communicative behaviour were seldom commented on or directly criticised in public. However, there are two reasons why I may have been unaware of such discussions.

The first empirical caveat is that the family I lived with went to some pains not to air dirty laundry in my view. Generally, criticisms of other people were phrased indirectly around me both in private and in public; I sometimes had to get the unexpurgated version of an event from less protective friends. This included criticisms of other people's use of language, as well as other moral judgements.

Secondly, since I was competent in Bislama but not in Tamambo, any comments made in Tamambo went past me completely. I could be sensitive to the points at which speakers code-switched into *lanwis*, and was aware of the symbolic effect of such a shift, but I had to rely on other people to gloss it for me afterwards. However, since some of my closest friends were *woman nara aelan*, they were not always in a position to help.

In short, this example shows us that even in domains where it seems to be generally agreed upon that instrumental needs, such as making oneself understood to the entire audience, members of the community can insist on a debate. Affective factors, such as community pride and respect for the local culture, can outweigh communicative efficiency for some members of the speech community. Metalinguistic comments on the appropriateness of a particular language are not common, but the underlying issues are the subject of lively community interest.

3.8.2 Trading on language: dos-and-don'ts at school

School had a rather complicated role to play in the linguistic marketplace of this community. The government sanctions and supports teaching in French or English in Vanuatu, but the extent to which the school domain is actually devoted to these languages varies. The local school in this area on Malo was in fact trilingual: English was the official language, but Bislama and Tamambo were contraband actively traded among all members of the school community, from students to principal. Of course, the three languages have rather different significance for the different members of the school community, since only the person with the most social power, i.e. the principal, can shift between all three languages relatively cost free.

In this section, I will try to give some idea of the complex function the school plays as a multilingual domain, both in terms of how it is perceived by members of the wider community, and in terms of how it socialises children. I will give an example showing that, through school, children learn that different languages available in the community have different values and different norms of appropriateness, and also how they make the most of the fact that code-switching can be the source of social censure as well as a means of binding the community together through shared resources.

The proscription against Bislama in school was not absolute. It could, for instance, be lifted for functional reasons. When parents came to school for parent-teacher meetings, or for meetings to arrange fundraisers etc., the English-only rules had to be relaxed for the benefit of parents or helpers. In addition, since most of the teachers were more comfortable in Bislama than in English, a lot of meetings between the staff were held in Bislama, too.

Moreover, though officially proscribed, Bislama was not completely lacking in status in the school. Virtually everyone agreed that an important and valuable aspect of formal schooling was that school was where you learn Bislama, regardless of whether your family spoke it at home. Both Vosale and Madelin independently told me that one of the reasons they and their husbands were very firm about speaking Tamambo at home was that children learn Bislama at school anyway.

However, the status of Tamambo at school was very low. Vosale and Livae reported that when they were in primary school they remembered being punished for speaking Tamambo. While I was doing my fieldwork, I was aware of more cases where children were punished for speaking Tamambo than for speaking Bislama, though technically, both students and teachers could be punished for using either. It may be worth noting that two of the teachers in the school were well-known (and the subject of some negative comment among *Natamambo*) for using only Bislama in their own families. As mentioned earlier, one teacher, *man Malo* married to *woman Malo*, was particularly notable for speaking to his children only in Bislama.

If, as we have evidence to believe, maintenance of *lanwis* is highly valued by the community, it seems curious that it does not censure more actively members whose behaviour is detrimental to the goal of language maintenance. French and English, as the state-mandated languages of education, are in some sense above any community censure, but neither Bislama nor Tamambo is officially sanctioned in school. Yet, as I have shown, the community tolerates and even expects Bislama in the semi-public domain of school. Why the community does not more forcefully discourage people from using Bislama with children seems again to be related to the tension between the affective and instrumental functions of the different languages. Comments by Atesolo, a young French-educated man, when I asked him about this give us some indication of this.

As we saw in §3.7.1., Atesolo showed very strong language and cultural loyalty to Malo. He was very critical of children growing up on Malo without being fluent in Tamambo. He mitigated his criticism of some families who speak Bislama to their children saying that when the mother is *woman nara aelan*, '*i gud leleber*' ('it's sort of alright').⁸ When I asked him why he thinks some *Natamambo* choose to speak Bislama at home, he suggested that maybe they thought it would help the children learn English later.

⁸ This may have been a nod to the fact that the family I lived with used mostly Bislama at home, and most of the mothers were *woman nara aelan*.

This reply is worth noting for two reasons. Firstly, Atesolo is not alone in believing this. It is one of the arguments put forward for encouraging the use of Bislama in education and in favour of using it as a medium for the promotion of national literacy. Secondly, because it indicates what he thinks the medium of personal and economic opportunity is most likely to be in the future. We must give this opinion added weight since it comes from someone who has himself chosen to pursue his living on Malo.

At the moment, a person's status as true *man Malo* is closely tied up with his competence in Tamambo. Tarip was one of several people in the community who defined linguistic competence as a criterion by which a person could establish their rights to land on Malo. How could someone claim ground on Malo, he put to me, if they don't speak Tamambo? Given this, the growing (if grudging on some people's part) acceptance of Bislama as a community language, has much wider social implications. Since knowing Tamambo is a major way of establishing one's rights and access to traditional ways of making a living on Malo (such as control of gardens), parents whose children hardly speak Tamambo must envisage their children's opportunities being more national rather than strictly local.

Thus, the pull between people's desire to see children have the most opportunities open to them even if that involves leaving Malo, and their desire to maintain the local culture and language is what lies beneath the community's rather ambivalent relationship with Bislama in the school: proscribed, but essential.

The children quickly become quite clear about the relative prestige of the three languages, both in and out of the school domain. We can see this in the way children trade on the languages, using them for social effect. Two examples serve to make the point.

One evening after dinner, three generations of the family had gathered, talking together in Tamambo and Bislama. The youngest boy, Iawia, who was 10 years old and is fluent in both Bislama and Tamambo, wanted to tell a funny story. He called to a *sista*, also fluent in Tamambo, while the conversation carried on around him. When he got her attention, he told the story in Bislama rather than Tamambo, which was actually what the other conversation was in at the time and the language the two of them usually used with their grandparents (also present). But by using Bislama, he got the attention and laughter of the entire room, monolinguals and bilinguals alike, even though he was nominally just addressing her.

On another occasion, Petre reported Saemon and Iawia for using Tamambo at school. The two younger boys were kept behind at the end of the day and punished for the infraction. It was very clear that Petre had been settling some kind of score. When I asked why he had reported them, he just shrugged – clearly it was just 'because' – and neither of the other two boys wanted to explain to me what they had done to make Petre want to get even with them.

Both the joke and the vendetta are significant. They show that even under-12s are aware of, and actively exploit, the capital associated with the different languages of the community in order to score points with and over their peers.

Thus, in looking at the school, we find that we are not seeing a domain with a uniform linguistic code, nor with unequivocal linguistic norms. Understanding how Bislama and Tamambo are used at school involves understanding the functions they serve outside the school yard, with the added fillip of an external standard. Children not only learn Bislama at school, they learn that languages in a multilingual community are more than referential tools.

They learn that a person's freedom to choose between languages is not independent of their behaviour and status outside the school, and they are given a domain in which they can practice using language to outmanoeuvre or charm others, skills that will be useful to them outside of school.

3.8.3 *Windows on souls: personal and affective factors affecting code choice*

Members of the Malo speech community also mentioned that more personal and private factors can affect their choice of language, although the extent to which they presented these factors as personal might depend on the image they wished to present to their interlocutor.

For instance, Mesek said that even though he would like to use Tamambo more on public occasions, he would usually use Bislama. In this case, though, it was not just because he wanted the whole audience to be able to understand. It turned out that he worried his formal Tamambo was not good enough: he said he simply did not always know the words or phrases appropriate to the formal registers. In other words, pragmatic concerns, such as comprehensibility, were but one facet of his decision. It was also shaped by more fundamental interpersonal issues to do with the respect due to the traditions of the language and his own face.

Similarly, close scrutiny of the way people dealt with the language issues raised by the presence of the *woman nara aelan* showed that ostensibly instrumental or pragmatic decisions had interpersonal dimensions to them as well. Both the *woman nara aelan* and *man Malo* gave the former's lack of competence in Tamambo as a reason for using Bislama rather than Tamambo. But speakers often contextualised the comprehension criterion in interpersonal terms. Susana, for instance, was one of several *Natamambo* who said that it would be rude to use a language which not everyone in the audience understood. She gave the example of a *woman nara aelan* joining a group speaking Tamambo. Susana said she would switch to Bislama, because 'you don't want her to think you're talking about her'.

However, while it was accepted that a woman new to Malo would have to be addressed in Bislama because of her lack of Tamambo skills, a woman could not keep trading on this without incurring some social costs. Some *man Malo*, such as Livae, Atesolo and Obed, explicitly said that although they would initially speak Bislama to *woman nara aelan*, after a while they felt it was fair to start addressing them in Tamambo. They were vague about how long the appropriate grace period was, though I had some indication of this myself. Four months after I first started working on Malo, I began to have people address me in Tamambo on Malo. Six months after I arrived, I was addressed in Tamambo by *man Malo* in a street in Santo (and was told off for not being able to do basic chit-chat in Tamambo yet, when I'd spent so much time there). In other words, while everyone on Malo accepts an inability to speak Tamambo as a reason to use Bislama with someone, this is tempered by attitudes about how long it should take a new arrival to speak Tamambo.

The *woman nara aelan* were aware of this. When questioned directly, both Leipakoa Alis expressed some shame about not using Tamambo at home (Leipakoa said, '*I nogud, a?*' 'It's terrible, eh?'), thus acknowledging that they were not serving the community goals of language maintenance. But Leipakoa sought some justification of her use of Bislama by pointing out that even her husband didn't consistently use Tamambo at home. Notice that this

implies a household order in which Leipakoa sees the role of passing on the *lanwis* as the task of her husband other *man Malo*, not necessarily her own.

When I asked Leipakoa whether there really wasn't anybody on Malo who she ever used Tamambo with, her reply was even more revealing. She said perhaps she did use Tamambo sometimes, but only with other people's children, old women (who speak no Bislama), or Nina. In other words, she said, people who can't or won't laugh if she makes mistakes. Anywhere she is worried about being ashamed, she prefers to use Bislama. Obviously Leipakoa's linguistic fluency accuracy is an important factor contributing to her self-esteem. In situations where she perceives a potential threat to her self-esteem, she prefers to use a language she is confident in. Apparently, other people's laughter is of more concern to her than someone's perception of her as a lazy language learner.

Nor was Leipakoa the only woman who let her competence in Tamambo be overridden by interpersonal factors. Janette Lolan also both mentioned that their fear of being laughed at constrained who they would attempt to use Tamambo around. Consequently, Janette's main use for Tamambo was as a way of occasionally conducting private conversations with other women from Malo when they were at the market in Santo.

In §3.7.4 I noted that Alis' language history indicates how important self-image can be to the shape of a speaker's linguistic repertoire. When she first came to Malo, she didn't want to believe she would be staying for long. She deliberately chose not to learn Tamambo properly, essentially as an act of resistance.

Does this mean that the *woman nara aelan* in Groups B C care less about whether people laugh at them? We have no reason for necessarily thinking this is the case. It seems more likely that in their cases (Nina, Dien Leikitah) there were other factors conducive to their attempting to use Tamambo.

When Nina was introspecting about her language patterns, she realised something she found curious. Her impression was that she used to use far more Tamambo with her children (now grown married themselves), but that now she used as much Bislama as Tamambo. This is probably not as curious as Nina suggests. It seems to indicate that Nina's patterns of language use were tied to her role as a mother. While she was actively involved in the routines as caregiver in bringing up her children, she used Tamambo more. This would be consonant with Rubinstein's (1978:315) observation that an important way in which a woman acquires social status on Malo is by comparing herself against other women in their roles as mothers wives, i.e. by commodifying the maternal identity. In Nina's case, as her maternal identity was eclipsed by other more salient ones, there are fewer direct benefits to her in demonstrating that she is successfully constructing her sons' identities as *man Malo* through their acquisition of Tamambo. there are fewer costs to her in using Bislama.

Dien, on the other hand, is still actively involved in constructing a social identity for herself as a mother. Using Tamambo with her daughters has integrative functions for her in the community, since it is one of the social acts that establishes her as a good wife mother in the (public) opinion of even the women like Leipakoa who do not use Tamambo with their children. It also has integrative functions for her daughters will be particularly important for any sons she may have. As noted in the last section, competence in Tamambo remains an index of a person's group identity on Malo. Many people, both *Natamambo* and some *woman nara aelan*, closely associate loyalty to and a sense of belonging to the place with being able to speak *lanwis*.

There was one other affective motivation that several people expressed for using Tamambo. Some of the *man* and *woman Malo* said that they felt an obligation or necessity to pass on Tamambo as a way of keeping indigenous science and philosophy alive. Vosale said, 'if we forget our *lanwis* we won't know important things like names for trees and birds'. The scenario she invokes is an existential, as much as a referential crisis. If a person learns the *lanwis* name of a tree or bird, there is a good chance they will learn something about its *kastom* uses and symbolism as well. Vosale is worried at the prospect of a community of Natamambo who, ignorant of the *kastom* names for their flora and fauna, will be ignorant of how Natamambo relate to and can use the plants and animals on their island.

Vosale's fear springs of a sense of collective good: maintenance of her *lanwis* is important for the well-being of the culture and the community, not just the individuals in it. And this evaluation of collective good can take on a pan-Vanuatu cast, too. Vosale mentioned to me twice, several months apart, that a government spokesperson had been on the radio saying that all Ni-Vanuatu had a moral and social responsibility to teach their children their *lanwis*, in order that the languages of Vanuatu not die out. Thus, even at levels of national discourse, instrumental and affective pressures blur or may even come into conflict. The government super-imposes language norms for education that favour French and English, presumably for the long-term benefits and opportunities those languages represent. But it also encourages the maintenance of *lanwis* in the community since in their number and diversity the languages of Vanuatu are symbols of national strength and regional loyalty.

3.9 Summary

In conclusion, then, we now possess a rather detailed description of the attitudes towards language that are expressed and demonstrated in the multilingual community on Malo. What is the purpose of this? In studying variation, we want to make sure that we have an informed basis for stratifying or grouping our data. In the case of syntactic variables, we want cross-linguistic and internal evidence for the most salient structural properties of the grammar to form the basis for organising our inquiry into the most significant factors shaping the variation observed. In the case of non-linguistic variables, we look to the community and the individuals for indications as to the most salient factors and properties used by the speakers to organise themselves and we inquire whether these factors interact significantly with the linguistic variation that interests us.

In this chapter, I have outlined the main attitudinal factors which play a role in shaping people's perceptions of themselves and other members of the community with respect to choice of language code. We have seen that social power (by which I mean the ability to effect a social change (Radke & Stam 1994)) and personal beliefs and fears all play a part in determining when different members of the speech community will use either the local *lanwis*, Tamambo, or the national language, Bislama.

By exploring speakers' reported and observed language use in different social domains, we were able to separate them into four groups, ranging from the speakers who identified positively and sometimes defiantly with the local language and what the language represents about their traditional culture, through to a group composed mainly of woman from other islands, who for personal as well as practical reasons only use Tamambo in highly restricted domains.

The transition zones, as it were, in the community were filled by men who worked off island or women whose families expressed strong positive feelings about language maintenance. The language chosen for speaking to other people's children distinguished these two groups further, the members of Group B preferring to use Tamambo to all children on Malo, and the members of Group C preferring Bislama or using the two languages equally. This domain, in which all speakers have greater social status than their addressee, is particularly important in our groupings of members of the speech community. Of all the domains investigated, speaking to other people's children constitutes the social setting in which speakers are all roughly equal in their ability to make language choices based on what they want to do, rather than what their addressee or the setting conventionally requires them to choose.

But a detailed understanding of people's perceptions and evaluations of the languages in the community is of more than biographical or prurient interest. The four groups that I have separated the members of the Malo community into on the basis of their linguistic behaviour will be put to use in the following chapters. When looking at the linguistic and non-linguistic bases for variation in argument deletion, variation between these groups will be one of the subjective factors against which I match the distribution of null and pronominal subjects.

4 Wanem i stap yet: *remaining matters of data collection and methodology*

4 Introduction

In general, I have raised matters related to the methodology of the data collection and analysis when and where they are most germane in the text. Therefore, this section of the work is provided essentially as a summary for quick reference. It provides

- (i) an outline of the methods by which I collected the corpus of spoken Bislama used for the analysis of variation between phonetically null and pronominal subjects in the following chapters,
- (ii) the methods of transforming the corpus of transcripts into a manageable linguistic database, including the tagging conventions and the identification of non-linguistic factors for the analysis of variation, and
- (iii) it gives a summary of the demographics of the people whose speech was transcribed for the corpus. In this section, there is a brief discussion about the motivation for including these factors where this has not been covered already in the preceding two chapters.

4.1 Sources of the data

Data was gathered in two locales in northern Vanuatu between September 1994 and August 1995, one urban and one village. Tape recordings of a number of conversations contributed to the corpus. Three recordings were two party conversations between myself and the interviewee. The remainder were either multi-party conversations with myself as an active participant, or (in the village community) multi-party conversations in people's homes that I was present for but were not primarily directed at me (e.g. after dinner talk, troubles talk between women friends).

4.2 Equipment

Recordings were made on a Marantz PMD 430 three head tape recorder, using a Sony ECM-F9 (omnidirectional) condenser microphone. Two recordings were made using a tie-tack style Realistic 33-1063 microphone (M-94-5 with Anita, and S-94-2 with Lili). However, I dispensed with the pin-on microphone after these recordings because the speakers were very self-conscious and because it was unsuitable for the multi-participant conversations I preferred to record.

4.3 Speaker demographics

The speech of 43 speakers was transcribed and makes up the corpus. The speakers' demographic profiles are summarised in Tables 4.1 to 4.4 below.

Table 4.1: Women from Santo (urban) in corpus – demographic summary.

Name	Age	Demographic description (occupation, education, marital status, origin)
Naomi	33	Midwife. Secondary (Eng.). Married. East Santo.
Lili	24	Waitress. Class 6 (Eng.). Married. Big Bay, Santo.
Elsina	32	Clerk. Secondary (Eng.). Married. West Ambae.
Adeline	20	Postal worker. Secondary (Eng.). Married. Banks.
Lisa	30	Postal worker. Secondary (Eng.). Married. Efate.
Seman	16	Student, secondary (Eng.). Unmarried. Efate.
Rebeka	21	Housegirl. Class 6 (Eng.). Married. Malo.
Juliet	40	Cook. Primary (French). Married. Ambae.
Rinette	36	Magistrate. Secondary (French). Married. Malakula.
N = 9		

Table 4.2: Men from Santo (urban) in corpus – demographic summary

Name	Age	Description (occupation, education, marital status, origin)
Timoti	34	Health administrator. Secondary (Eng.). Married. Banks.
Jonas	42	Farm worker. Primary (French). Married. Mavea Is.
Stivin	68	Area chief. No education. Married. Mavea Is.
Simeon	35	Lawyer. Secondary (Eng.). Married. Pentecost.
Ezra	50	Truck driver. Secondary (Eng.). Unmarried. South Santo.
Wingston	33	Bank clerk. Secondary (Eng.). Married. West Espiritu Santo.
Sevi	24	Bartender. Secondary (Eng.). Married. Malo.
Sikal	29	Army. Secondary (French). Unmarried. Banks.
Bretian	16	Student. Secondary (Eng.). Unmarried. Aneityum.
N = 9		

Table 4.3: Women from Malo (village) in corpus – demographic summary

Name	Age	Description (occupation, education, marital status, origin)
Alis	21	Housegirl. Primary (Eng.). Unmarried. Efate.
Lepakoa	39	Mother. Primary (Eng.). Married. Efate.
Lolan	31	Teacher. Secondary (Eng.). Married. Aneityum.
Susana	14	Housegirl. Primary (Eng.). Unmarried. Malo.
Janette	30	Mother. Primary (Eng.). Married. Maewo.
Madelin	24	Mother. Primary (French). Married. Malo.
Anita	28	Teacher. Secondary (French). Married. Pentecost.
Vosale	30	Mother. Primary (Eng.). Married. Malo.
Dien	24	Mother. Secondary (Eng.). Married. East Ambae.
Rovi	16	At home. Primary (French/Eng.). Unmarried. Malo.
Leikitah	28	Mother. Primary (Eng.). Married. South Santo.
Nina	56	Retired nurse. Primary (Eng.). Widowed. Efate.
Mama	60	Mother. No education. Married. Malo.
Elise	8	Primary (Eng.). Unmarried. Malo.
N = 14		

Table 4.4: Men from Malo (village) in corpus – demographic summary

Name	Age	Description (occupation, education, marital status, origin)
Mesek	40	Boat/ferry driver. Primary (French). Married. Malo.
Visi	34	Army medic. Secondary (French). Married. Malo.
Tarip	30	Builder. Secondary (Eng.). Married. Malo.
Livae	24	Church counsellor. Secondary (Eng.). Unmarried. Malo.
Papa	65	At home. No education. Married. Malo.
Jackson	21	Teacher. Secondary (French). Unmarried. Malo.
Obed	20	At home. Primary (French). Unmarried. Malo.
Atesolo	18	Truck driver. Primary (French). Unmarried. Malo.
Iawia	11	Child. Primary (Eng.). Unmarried. Malo.
Petre	12	Child. Primary (Eng.). Unmarried. Malo.
Saemon	10	Child. Primary (Eng.). Unmarried. Malo.
N = 11		

4.4 Transcription and coding of the corpus

Excerpts from the tapes were transcribed and translated with the assistance of Sharon (Morrie) Tabi. Sharon Tabi was a teacher at one of the East Santo English language primary schools. Ms Tabi was born on Pentecost, but has spoken almost nothing but Bislama at home since she was five. Her language skills in her L1, Raga, continue to be good, but she considers herself (and told me her parents would concur) to be dominant in Bislama.

I am considerably indebted to Sharon Tabi for her help, her native speaker ear was able to decipher considerably more when there was overlapping speech and where conversations were recorded with lots of background noise than I was. Not only did she help me to understand the words on my tapes, but she also helped me immeasurably in understanding the sense of what was being said. This included making explicit for me culturally determined presuppositions underlying some of the conversations recorded, and I have relied on her intuitions in resolving ambiguities in pronominal and null subject reference in the corpus. This research and my own experience in Vanuatu would have been much the poorer without her.

The transcripts were then tagged (by myself, based on the translations completed jointly) for parts of speech. This was done very roughly, since I did not want to presuppose too much structure in the parsing conventions. One of the purposes of working from a conversational database is that we may find that in use, certain forms which seem ambiguous between different grammatical classes are disambiguated, and conversely, words which may seem straightforward realisations of one grammatical category, are in fact ambiguous between more than one.

For instance, Crowley (1995) classes *olsem* as an adverb 'thus', a preposition 'like' and a complementiser 'that'. I have shown (Meyerhoff 1998, forthcoming a) that the distribution of *olsem* in writing and speech suggests that it is not an adverb, and an analysis based on natural occurrences in discourse illustrates a number of other functions that *olsem* can serve as well, i.e. hedge, conjunction and demonstrative.

The tagged corpus annotated the text for the following syntactic and lexical categories:

SBJ	subject
FS	focused subjects
VP	verb phrase (including DO but not obliques)
DO	direct object
OBL	oblique objects
F	focused objects (direct and oblique)
PP	prepositional phrases
ADV	nominal or single word adverbs (not included in the PP class)
FA	focused adverbials
COMP	complementisers
CNJ	clausal (not nominal) conjunctions
DC	discourse particles (e.g. <i>ale</i> 'well, so'; <i>ya</i> clause finally 'eh'; expletives <i>kas</i> , <i>fak</i>)

Speakers were coded for the various social factors outlined in Tables 4.1–4.4, and a file that contained this information about the individual speakers was established. This was then automatically related to the transcript files through the program Goldsearch.

4.5 Database management: the Goldsearch program

None of the existing database programs available for use on the Macintosh (Excel, Filemaker, even Nisus which has a powerful search language) would perform exactly the kind of iterative searches through the text that I wanted. I sought a program that would not only find matches of a search string, but would then relate each match of the search string to the speakers' social information held in a separate file. I, therefore, enlisted the aid of some like-minded people in devising a program that would do this.

The program we ended up producing runs on a Unix system and has been dubbed Goldsearch (in part because the output is ready to be used with the multivariate analysis program Goldvarb (Sankoff et al. 1992). The scripting of Goldsearch was done by David Boas (Harvard University/Massachusetts General Hospital) and we were assisted in the program design by Naomi Nagy (University of New Hampshire).

Essentially, Goldsearch does the following. It relates two files: one data file, and one file with information about the composition of the data that remains constant and is predictable. In this case, they correspond to the corpus of transcripts in which each turn is attributed to one and only one speaker, and the file with speaker demographics. The search options for working on the corpus file are defined to allow easy retrieval of parsed text. For instance, the program automatically matches brackets in the search (although the search can also be restricted to find simple text matches). It also has three wild cards which allow variables in the search string to be restricted to (i) within words, (ii) within constituent boundaries, or (iii) within a speaker's turn (but across constituent boundaries). The program allows the user to search for matching tokens of a search string and, through a 'not' annotation, complementary tokens of a search string.

The search automatically produces two output files. The first output file is a list of all matches found. This is the text output and is useful for further linguistic analysis. The second output file is a list made up of one entry of the speaker code for each of the matched tokens found. That is, if six matches of the search string are found that have been uttered by Speaker A, six copies of the coded demographic information on Speaker A appear on separate lines of the second output file. This output file is designed so as to be able to be copied as is and be used as a token file for multivariate analysis with, for example, Goldvarb.

The program is freely available and can be downloaded from the World Wide Web at:

<http://english-1.unh.edu/nagy/Goldsearch/Goldsearch.html>

4.6 Non-linguistic factors

Each speaker was coded for the following social and demographic factors:

- * speaker sex (discussion §2.2.2 and §6.6.4)
- * age (in five year groupings) (discussion §6.6.3)
- * marital status (never married, married or widowed)
- * level of formal schooling attained (primary, secondary or beyond, none)
- * language of formal schooling (French, English, not applicable)
- * place of origin (discussion §3.6.3 and §6.6.1)

The social salience of speaker sex and age have already been discussed. In the next sections, I discuss why it was hypothesised that marital status and language and level of schooling might be sociolinguistically salient in the communities of my fieldwork. Some further remarks concerning the potential linguistic salience of a person's place of origin are provided, locating the discussion in larger questions related to the part substrate languages play in the synchronic shape of a creole.

4.6.1 Marital status

Marital status was included as a factor because my observations of the communities involved suggested that marriage (or entering a *de facto* relationship) had a marked effect on a young person's social roles and social activities. As I have noted, young women in particular have major changes in lifestyle if they marry or move in with a man. They are increasingly less likely to socialise with other young women and (on Malo) are less likely to spend their afternoons playing volleyball and hanging around the local co-operative store. Early on in my fieldwork, Rebeka, one of my Santo speakers, started a relationship with a young *man Malakula*. I noticed that shortly after they started seriously going out with each other and she was moving in with him at his house where he worked, she started grooming herself differently. For example, she began tying her hair back, and she stopped wearing her volleyball shorts even occasionally to her work in town.

The main reason a young woman's life is more likely to be significantly altered by entering into a relationship than a young man's, is that she usually starts having children. Vanuatu's population growth is a hefty 2.17% per annum. Pharmacological contraception is not widely-used and the Vanuatu constitution makes abortion illegal without exception. One *woman Malo* who lived in Santo and who I occasionally travelled with from the island back into town was involved in family planning programs in the Sanma region. According to her, the pill is the contraceptive of choice for women in Vanuatu. IUDs are sometimes used, but the risks associated with any complications meant that counsellors do not recommend IUDs for women outside of areas serviced by hospitals, i.e. effectively anywhere outside of Santo and Vila. This woman was one of several Ni-Vanuatu women engaged in education programs directed at women in the region.

Social factors seem to be the main reason condoms are not widely used as a method of family planning. They were available commercially in Santo, and so cheaply (VT20 each) that I assume they were subsidised by the government, however men do not generally seem to take responsibility for contraception within marriage. Part of the public health program being carried out by the main British aid agency active in northern Vanuatu during my fieldwork was family planning education. A British man in his late 20s had been appointed to work on these programs with Ni-Vanuatu men in the northern region in order to encourage greater awareness among men of family planning issues. Unlike family planning resources for women, I never heard of any Ni-Vanuatu men who were employed specifically as education officers in this area (though naturally some of the men involved in health care in the region as doctors, medics or administrators were involved in this from time to time). Vanuatu is officially AIDS-free, and official encouragement of condom use for reasons other than prophylaxis is muted. Terry Crowley reports (*pers.comm.*) that where they are free (as through the health centre on Erromango), condoms were more widely used. This may indicate a broad behavioural shift taking shape in Vanuatu.

Having said all that, though, family planning in Vanuatu still generally seems to focus on responsible plans for having children, e.g. encouraging appropriate spacing of children. Keeping family sizes down seemed a secondary goal. And although Ni-Vanuatu men do do some child-minding and child-rearing, the main responsibility for care and training of children lies with women.

Thus, child-bearing and marriage have a significant effect on the social roles and opportunities available to Ni-Vanuatu women. Especially in public or large gatherings, women are more likely to be responsible for the children, and women with children tend to form a group or community of their own at festivities and communal events.

I was made particularly aware of the salience of marriage and maternity during my fieldwork because I fell in a fairly empty social space, being married but childless (and husbandless for the duration of the fieldwork). Friends who I felt I had the most in common with and who recognised me as being their peer usually had several children, some as old as eleven. My anomalousness apparently made it difficult to categorise me in the Malo community. Especially when I first arrived and everyone was trying to work out how I would fit into the community, I was sent off to sit with or complete chores with teenage girls as often as I was sent off with young mothers.

In light of all this, I felt I could not assume that chronological age would be a more important social identity than a person's social age (pre-marital, post-marital). For this reason I coded speakers' marital status as a separate factor. Obviously, chronological age is still important as this may indicate changes in progress in the speech community. As we will see in the next chapter, speaker's marital status did not seem to be a particularly salient factor with respect to the syntactic variables investigated.

4.6.2 Education

Speakers were also coded for their language of education and level of education on the off-chance that the structure of their Bislama was being affected by their fluency and exposure to English or French. It was also theoretically possible that the dual education systems might have given rise to distinct dialects of Bislama (though it must be said, this perception did not seem to be held by anyone in Vanuatu). As it transpired, the language of education showed absolutely no interaction with the linguistic variables investigated in this study.

Speakers fell into three groups, those who had no formal education, those who had some primary education (as noted earlier, most people leave school after Class 6 at age 12), and those who had some secondary education (including speakers who had undergone post-secondary education, e.g. some university or vocational training).

4.6.3 Origin of speakers

Earlier sections have discussed the interpersonal and intergroup salience of people's home island affiliations. It was also considered important to code data for the origin of the speaker since it is always possible that variation observed in a creole is the consequence of speakers calquing constructions or discourse patterns from their L1. The influence of substrate patterns on the structure of creoles is often positioned as if in opposition to the influence of innate

linguistic universals (Schroeder 1990). There have been sometimes vehement debates of this between Bickerton (1981, 1990) on the one hand, and Keesing (1988) and Corne (e.g. Corne et al. 1996) on the other.

Bickerton has repeatedly argued that inherent cognitive schema determine the development of creoles under rapid creolisation. As a consequence, he argues that the structures found in creoles are clear indications of the features of a program that biologically predisposes humans to the acquisition of language.

Keesing (1985) argued equally strenuously that the grammar of Solomon Islands Pijin was a direct, comprehensive and obvious calque of syntactic patterns found in substrate languages of the region (the significance of West African substrate patterns has been argued for some structures in Atlantic creoles, e.g. Alleyne 1980). A similarly strong view is forwarded by Siegel, Sandeman and Corne (1999) for the New Caledonia French-lexified creole, Tayo.

In recent years, there has been a welcome trend towards acknowledging the fact that substrate patterns and universals often provide equally good models for a particular grammatical or pragmatic structure, e.g. Mufwene (1986, 1991, 1996), Byrne (1990), Sankoff (1996) for Atlantic and Pacific creoles, and Prince (1992) for Yiddish.

Amidst the substrate versus universals debate, Singh's (1996) analysis of Trinidadian English provides an important reminder that the role of non-standard varieties of English in shaping the structure of creoles cannot be ignored either. Work by Smith (1998) is concerned with providing a more detailed picture of non-standard negation patterns that are candidates for input in Caribbean creoles, e.g. Jamaican Patwa. Mufwene (1996:88-90) reviews how the importance of regional and non-standard varieties of the lexifier has waxed and waned in the history of creole studies.

Crowley (1989) shows that the development of a lexical copula in Bislama may have derived from several sources. The range of analyses speakers may impose on a variable or developing structure will, he argues, lend strength to the form and provide it with an added advantage for survival. Similarly, Sankoff (1993, 1996) showed that the grammaticisation of *yet* in Tok Pisin could not possibly be attributed to either universals or the substrate alone. A lexeme drawn from the English lexicon initially retained the functions associated with English *yet*. But the superficial similarity with a substrate lexeme, Tolai *iat* facilitated its spread as a focus marker. Sankoff's data shows that Tok Pisin *yet* is now specialised as a marker of focus with pronouns (as opposed to nouns). This pattern suggests that speakers have co-opted *yet* as a means of grammaticising a more universal distinction between the referential and discourse salience of pronouns and full NPs. As in the case of the Bislama copula, all three possible influences – universals, substrate and the lexifier – contribute to the restructuring of the creole. Byrne's (1990) survey of verb serialisation in several creoles, concludes that both substrate patterns and language universals must be responsible for the range and the type of serial verb constructions found. I believe the data he presents suggests that some evanescent property or feature of the lexifier grammar may also have a role to play in determining the typology of creole serial verbs.

However, despite the empirical difficulties in determining whether a property of a creole is shaped by universals or the substrate, it is worth keeping track of where our speakers of Bislama come from. Speakers are grouped in linguistically and ethnographically sound subgroups, based on the data discussed in Chapters 2 and 3.

As I have already noted, even if differences in the origin of speakers correlates positively with linguistic difference, we may not necessarily be able to conclude that substrate patterns are shaping the structure of Bislama grammar. This is because many of the regional languages spoken by people in the corpus are not well described. Thus, it may be impossible to verify whether a pattern observed in Bislama is a substrate calque. This is especially difficult when the variation we are looking at is constrained by information structure or discourse patterns, since these facts are seldom dealt with in enough detail in traditional grammars to allow any meaningful comparison. Prince (1996) has shown that the influences of substrate grammars may show up in subtle ways in Yiddish. Understanding the Yiddish variation, therefore, requires a detailed understanding of the organisation of the relevant constraints in the substrate grammars.

4.7 Summary

This chapter outlined the mechanics of the data collection and data analysis, including linguistic and non-linguistic factors relevant to the corpus. Some of the social factors discussed in this section have already been introduced in previous chapters, in the analyses of the pronominal variables which follow, both the linguistic and social contexts in which Bislama is embedded will be needed to shed light on different aspects of variation.

5 *Verb inflection and subject pronouns*

5 Introduction

The subject of a lexical verb in Bislama may be realised overtly or it may be phonetically null. When overt, i.e. a proper noun, a noun phrase or a pronoun, the interpretation of the subject is a fairly straightforward process. However, when the subject is phonetically null, it is less clear how interlocutors interpret the referent. In Chapter 6, the distribution and function of overt and phonetically null subjects in conversational Bislama will be examined in detail.

However, before we are able to consider the interpretation of phonetically null subjects, we must first determine the most appropriate analysis of the verb phrase. This is necessary because there are competing analyses of the Bislama verb phrase and these differ precisely over the correct analysis of pronouns and of the so-called predicate markers. This chapter investigates the distribution of focused and unfocused pronouns in order to resolve the differences over how pronouns and subject-verb agreement in Bislama should be analysed. I will show that the morpheme sometimes called the predicate marker in Bislama (and related languages) has three forms in Bislama, *i*, *oli* and a zero. I will also show that the distribution of these morphemes is clearly that of subject-verb agreement; this means that the descriptive term, predicate marker, can be dispensed with for Bislama.

5.1 Defining the ‘predicate marker’

Bislama, with the closely related varieties Tok Pisin and Solomon Islands Pijin, has a morpheme that delimits the left edge or start of the verb phrase. This has come to be referred to by the descriptive term ‘predicate marker’, e.g. Crowley (1995) for Bislama, Mihalic (1971), through to Verhaar (1995) for Tok Pisin. Keesing (1988) prefers the term ‘subject-referencing pronoun’ for Solomons Pijin. The latter term draws attention to both the referential independence of the forms, and to the relationship between subject and verb that they index.

The form of this marker varies in the different languages; in Bislama the form surfaces as *i* with most persons and numbers, but as *oli* for prototypical 3PL subjects.¹ By prototypical, I mean 3PL subjects where the subject NP is marked with *ol* ‘the (PL)’ or *olgeta* ‘they; all the’.

5.1 [Hem] *i no talem long ol sambodi*
 3SG PRED.MKR NEG tell PREP PL somebody
 ‘She doesn’t tell a soul’ (S-95-10, Rebeka)

5.2 *Olgeta oli pas long intaviu yestedei*
 3PL PRED.MKR pass PREP interview yesterday
 ‘They had all interviewed the day before’ (M-95-12, Dien)

Other non-singular third person subjects, such as *tufala* ‘they (5.2); those two’ show some inter- and even intraspeaker variation in the form of predicate marking considered appropriate, as shown in 5.3:

5.3 *I gat tu devel oli kam nao...*
 PRED.MKR have two devil PRED.MKR come now

tufala i go, oli luk
 3DU PRED.MKR go PRED.MKR look
 ‘Two devils came along...the two of them went and looked [for him]’
 (M-95-17, Leikitah)

Though the written convention is to represent *i* and *oli* as separate words, it is clear that they are clitics on the verb. It is impossible to interpose any of the modifiers such as *nao* ‘now’, *nomo* ‘only’, *tu* ‘also’ or *bakegen* ‘again’ between *i* and *oli* and the core elements of the verb phrase (the lexical head and any aspect or mood markers). Crowley (1987a, 1990) notes that this property is criterial in distinguishing the fused form *oli* in Bislama from the Tok Pisin *ol i* (3PL pronoun plus predicate marker). In Tok Pisin, the analogous forms *nau*, *tasol*, *tu* and *yet* can occur between *ol* and *i* (Verhaar 1995:357), but *nao*, *nomo*, *tu* and *bakegen* cannot split *oli* in Bislama. In a similar vein, we can note that it is impossible in Bislama to insert a modifier of the verb between *i* or *oli* and the core elements of the verb phrase.

The predicate marker is subject to a great deal less optionality in Bislama than has been documented for the related form found in Tok Pisin (Woolford 1977; Romaine 1988; Sankoff 1994), but it is not clear what factors contribute to its occasional absence in Bislama. It may be a stylistic strategy, analogous to English subjectless sentences in co-ordinated constructions, e.g. ‘Jeremy brought the car round and Ø parked it in front’, that is, not part of the core syntax. However, it is also possible that the optionality indicates Bislama

¹ I will use the following abbreviations for person and number of subjects:

1	first person	SG	singular
2	second person	DU	dual
3	third person	TR	trial
INCL	inclusive	PL	plural
EXCL	exclusive		

Until the correct analysis of *i* and *oli* in Bislama is resolved, all pronoun forms will be glossed with their person and number features, and *i* and *oli* will be glossed ‘PRED.MKR’, for ‘predicate marker’.

is at the early stages of a major reanalysis of the predicate marker such as that documented for urban Tok Pisin by Sankoff (1994).

In short, the appearance and form of the predicate marker for all persons and numbers possible in Bislama is as shown in Table 5.1 (using the verb *pleple* 'play' as an example). Five cells in the table do not have any overt marker on the predicate, i.e. 1SG, 2SG, 1DU(INCL), 1TR(INCL) and 1PL(INCL) subjects. These italicised entries are those where opinion diverges as to the correct analysis of the predicate.

Table 5.1: Verb paradigm in Bislama: pronoun + so-called predicate marker + lexical verb. (*Italicised entries have disputed analyses.*)

	singular	dual	trial	plural
1(INCL)	–	<i>yumitu pleple</i>	<i>yumitri pleple</i>	<i>yumi pleple</i>
1(EXCL)	<i>mi pleple</i>	mitufala i pleple	mitrifala i pleple	mifala i pleple
2	<i>yu pleple</i>	yutufala i pleple	yutrifala i pleple	yufala i pleple
3	hem i pleple	tufala i pleple	trifala i pleple	olgeta oli pleple

Churchill (1911:21f) tentatively suggested that *i* itself is the predicate (by analogy with Samoan, an approach of dubious merit), but to be fair Churchill admits that his grammatical sketch of Bislama is far from adequate. It is now generally accepted that *i* and *oli* are part of the verb phrase and have some functional relation with the lexical head. There is much less consensus, though, as to whether there is a predicate marker and, if so, what its form is for the five italicised cells in Table 5.1. In the next section, I will attempt to resolve the matter by undertaking a systematic investigation of pronouns and predicate markers in use. This will clear the way for a full analysis of variation between phonetically null and overt subjects in Chapter 6.

5.2 *Mi mi* and *yu yu*: subject focusing or subject agreement?

The problem researchers have been faced with is whether 1SG and 2SG have an analogue of the 3SG *i* or not. The question arises because of the apparent frequency with which 1SG and 2SG subjects appear in doubled sequences, i.e. *mi mi* and *yu yu*. Furthermore, two possible analyses of the phonological relationship between singleton *mi* and *yu* and the lexical verb have been proposed.

I first review the evidence for and against treating doubled subjects such as *mi mi* as if they were formally equivalent to the 3SG form *hem i*. I will show that based on their distribution in different discourse environments doubled forms of 1SG and 2SG should be considered focus constructions, and not equivalent to the 3SG pattern *hem i*.

I then review the arguments for and against the presence of an underlying *i* marking the relationship between 1SG and 2SG subjects and the verb. I will conclude that the evidence suggests that there is no underlying phonological segment.

The notion that doubled *mi mi* is analogous to *hem i* is by no means outlandish. Doubling of the 1SG and 2SG pronouns is found in situations parallel to the 3SG pronoun and the following predicate marker *i*. This is the case, for instance, if a speaker wishes to use one of

the modifiers mentioned above, or if the speaker wants to insert *bae*, marking the verb phrase as irrealis.² Compare examples 5.4 and 5.5:

- 5.4a. *Mi nomo mi nyuhebridian long taem ya.*
 1SG only 1SG new.hebridean PREP time specific
 'In those days, I was the only Ni-Vanuatu.' (S-95-15, Rinette)
- a'. **Mi nomo nyuhebridian long taem ya*
 **Nomo mi nyuhebridian long taem ya*
- b. *Yu bae yu gobak wataem?*
 2SG IRR 2SG go.back what.time
 'When are you going back?' (M-94-1, Alis)
- b'. **Yu bae gobak wataem?*
- c. *Bae mi go spel long Ambae.*
 IRR 1SG go holiday PREP Ambae
 'I should go spend the holidays on Ambae.' (M-95-15, Jackson)
- 5.5a. *Hem nomo i olsem.*
 3SG only PRED.MKR like
 'He's the only one like that.' (M-94-2, Leipakoa)
- a'. * \emptyset *nomo i olsem*
- b. *Hem bae i kam slip long hem?*
 3SG IRR PRED.MKR come lie PREP 3SG
 'Will she come (and) nest on it?' (M-95-19, Elise)

As 5.4a, b show, an adverb such as *nomo* (modifying the subject) or *bae* (modifying the verb phrase) may not intervene directly between a 1SG or 2SG pronoun and the lexical verb. Either the adverb must be placed at the left edge of the verb phrase, e.g. *bae* in 5.4c, or 1SG and 2SG must double. Other pronouns, exemplified by the 3SG in 5.5a-b allow *nomo* or *bae* to occur between the subject pronoun without having to double. Example 5.5a' shows that a focus marker such as *nomo* must modify a pronoun, it cannot modify a phonetically null subject, nor the predicate marker *i*. So there are good reasons for thinking that the clitic *i* and the doubled *mi* and *yu* hold the same position in the verb phrase.

In addition, in secondary predications with 1SG and 2SG subjects we find *mi* and *yu* occur where with a 3SG subject we find the predicate marker. This is shown in the (constructed) examples in 5.6. So there appears to be a distributional isomorphism between *mi* and *yu* and the 3SG predicate marker *i* in secondary predications, as shown in 5.6a and 5.6b (intuitions of Sharon Tabi):

- 5.6a. *Hem i kamtru i kros.*
 3SG PRED.MKR arrive PRED.MKR cross
 'S/he arrived angry.'

² By irrealis I mean a (i) future event, (ii) a desire or wish, (iii) the consequent of a conditional, and (iv) (sometimes) negative clauses. The analytical or even ontological usefulness of the term 'irrealis' is critiqued by Bybee (1998), and a reappraisal of the use of the term is perhaps warranted in Bislama since it is not obligatorily marked even in (i)–(iii).

5.6a'. *Hem_i i kamtru hem_{uvj} i kros.*

5.6b. *Mi kamtru mi kros.*
 1SG arrive 1SG cross
 'I arrived angry.'

As (5.6a') shows, repeating the pronoun with 3SG blocks an interpretation as a secondary predication. Instead, the pronouns are interpreted as having disjoint reference.

The formal facts relating to the 1DU, 1TR and 1PL(INCL) pronouns are rather muddier than the formal facts associated with *mi* and *yu*. Tokens of the dual, trial, and plural inclusive are, for one thing, much more infrequent in conversation. Consequently, tokens co-occurring with the crucial indicators *bae* or *nao* etc. are also harder to find in naturally occurring data. My own data, for instance has exactly one such token.

5.7 *Be yumi bae i no save stopem maot blong yu.*
 but 1PL.INCL IRR PRED.MKR NEG can stop mouth of 2SG
 'But we can't close off your mouth.' (M-94-2, Leipakoa)

Written examples are easier to find, e.g. in the Bislama *Nyutesteman* Bible

5.8 *Yumi evriwan i kakae bred ya.*
 1PL.INCL everyone PRED.MKR eat bread SPEC
 'We (INCL) all eat this bread.' (Fas Korin/1st Corinthians 10:17)

It is possible that the *i* surfaces in sentences like this because *evriwan* requires the pronoun to be interpreted distributively, but the evidence suggests that this is not what motivates its appearance in 5.8. We also find the *i* surfacing in 5.9 and 5.10 where the interpretation of the pronoun is not intended to be individuated, on the contrary, the context stresses a collective reading:

5.9 *Taswe yumi evriwan i olsemwan bodi nomo.*
 thats.why 1PL.INCL everyone PRED.MKR like one body only
 'For that reason we are all as one.' (Fas Korin 10:17)

5.10 *Olgeta evriwan oli kakae sem kakae ya nomo we*
 3PL everyone PRED.MKR eat same food SPEC only COMP

God i givim long olgeta.

God PRED.MKR give PREP 3PL

'They all eat the same food that God provides for them.' (Fas Korin 10:3)

Example 5.7 from my corpus is in line with the patterns described in Crowley (1990) for *yumi*. Crowley characterises the predicate marker with 1PL(INCL) and 1DU(INCL) *yumitu* as alternating between a zero form when the pronoun is adjacent to the verb, and an *i* when there is some intervening constituent, e.g. an adverb (1990:234, 235). The behaviour of 1DU and 1PL(INCL) with respect to adverbs differs from the behaviour of 1SG and 2SG. This has compelled researchers to present different analyses of the relationship between the verb and *yumi* or *yumitu*, than the relationship between the verb and *mi* or *yu*. Given the paucity of data available that is on the 1DU and 1PL(INCL), I will focus primarily on the distribution of 1SG and 2SG. However, I will return to the implications of 5.7 and Crowley's observations about 1DU(INCL) later in this chapter.

In summary, two formal facts have led some researchers (Guy 1974a; Camden 1977; Charpentier 1979) to analyse doubled 1SG and 2SG sequences as equivalent to 3SG pronoun + *i* sequences. These facts are: (i) *mi* and *yu* cannot be separated from the verb by adjectives or adverbs; (ii) in secondary predications *mi* or *yu* is obligatory with the subordinate verb. In both respects, *mi* and *yu* behave like the clitic *i* that occurs with most other subjects.

Finally, there is a perceptual reason why some observers have analysed doubled 1SG and 2SG sequences as being like *hem i*. The doubled 1SG and 2SG have seemed to some people (including myself) to be more frequent than doubled referents in other persons and numbers. Since no quantitative analyses of Bislama have been undertaken before now, such assessments of frequency cannot be distinguished from perceptual salience. I shall use the corpus of spoken Bislama that I gathered to test the validity of this perception.

5.2.1 Review of previous analyses

The divergent analyses of these doubled strings can be summed up in the following way. Guy (1974a), Camden (1977) and Charpentier (1979) analyse the first pronoun as the subject and the second as the predicate marker, exactly comparable with the pronoun followed by *i* found elsewhere in the verb paradigm, e.g. in 3SG *hem i pleple*. Thus, Guy, Camden and Charpentier (GCC) all argue that where we find clauses with a single *mi* or *yu* preverbally, these are the result of the subject having been dropped. It is worth noting that all three of these researchers worked extensively in northern Vanuatu.

This is not the only possible conclusion. As Crowley (1990:239) points out, it could equally be the case that in singleton cases of *mi* and *yu* the predicate marker has been dropped. Crowley (1990) and Tryon (1987) analyse the doubled sequences as ‘focused’ or ‘topicalised’ (respectively). In other words, they argue they are exactly comparable with a focused 3SG subject, *NP, hem i*, as in:

- 5.11 *Denis, hem i plan blong mekem wan trik.*
 Denis 3SG PRED.MKR plan COMP make one trick
 ‘Denis, he planned a trick.’ (S-95-7, Sevi)

Tryon analyses 1SG, 1PL(INCL) and 2SG as having no predicate marker at all (1987:20). Tryon does not state what he thinks the situation is for *yumitu* and *yumitri*, except to note that he believes they only occur in rapid speech (1987:20) and more properly surface as *yumitufala* and *yumitrifala*. In that case, the *-fala* suffix demands an *i* predicate marker. My own data indicates that this analysis of variation between *yumitu* and *yumitufala* etc. is not quite correct. I have equal numbers of *yumitu* and *yumitufala* in conversation.³

³ The choice of one over the other does not seem to be a dialect difference. *Yumitu* occurs in the speech of a man from the Banks, a woman from Efate, a woman from the south, a man from Malo and a girl from Malo. *Yumitufala* occurs in the speech of a woman from Efate and the same woman from the south, Lolan. The basis for the variation between these forms remains a matter for future research. It is possible it relates to grammatical role or information status, e.g.:

- (i) *Susana, olsem wanem yumitufala kalarem laplap blong yumitu bakegen?*
 Susana like what 1DU(INCL) colour laplap of 1DU(INCL) again
 ‘Susana, what say we dye our laplap again?’ (M-94-3, Lolan)

Crowley's description of Bislama characterises the predicate marking with *mi* and *yu* as alternating between a "zero" (1990:234) and "predicate markers which are identical...to the pronouns themselves" (1990:235). The alternation is attributed to different linear configurations in the clause. As noted already, the marking with *yumi* and *yumitu* is characterised as a similar alternation between zero and *i*.

Another possible analysis is that the predicate marker *i* is present underlyingly in all these cases but that it assimilates with the final high vowel in *mi*, *yu* and *yumi(tu)*. This analysis has been proposed sufficiently often to me (informally) by people who work on or are familiar with varieties of Melanesian English, that this chapter will also consider evidence for and against it. In support of this analysis, one might point to the fact that *i* appears with 1DU and 1PL(INCL.) forms when they are separated from the verb, as noted in 5.7–5.9 above. This possible analysis will be referred to as the *generalised i analysis* in the discussion that follows.

Finally, Tryon (1987) also suggests that speakers of different varieties of Bislama may analyse and therefore use the doubled pronominal forms differently. He suggests that the GCC analysis may hold for some regional varieties of Bislama. In other words, he suggests that in some varieties of Bislama the grammar of the verb paradigm might differ from the one he gives, and doubled 1SG or 2SG would not be focus constructions. We cannot rule out this possibility, since as I noted at the start of this section, different researchers have worked on Bislama in different regions. Crowley's main Bislama ties are from the central region, while Tryon (like GCC) has long-standing ties in northern Vanuatu.

The analyses forwarded by GCC and Tryon and the generalised *i* analysis are summarised in Table 5.2.

Table 5.2: Summary of competing analyses of 'doubled' 1SG and 2SG pronominal subjects in Bislama and distribution of the so-called 'predicate marker' in Bislama.

Guy/Camden/Charpentier			
	opt. pronoun	pred. marker	V
	mi	mi	pleple
	yu	yu	pleple
	hem	i	pleple

Tryon			
focus	pronoun	pred. marker	V
mi	mi	--	pleple
yu	yu	--	pleple
Denis	hem	i	pleple

Generalised <i>i</i> analysis				
focus	pronoun	pred. marker	V	phonological rule
mi	mi	i	pleple	$i \rightarrow \emptyset / V [_{\text{VP}} \text{---}]$
yu	yu	i	pleple	[+hi]
Denis	hem	i	pleple	

Thus, there are two questions that need to be answered. The first is, is the second pronoun in the doubled sequences, *mi mi* and *yu yu* a pronoun or a predicate marker? The second is, if a pronoun, what happens to the predicate marker in these cases?

The question is one that is ideally suited to quantitative evaluation. We need to determine whether the distribution of doubled pronoun sequences in discourse is more like the distribution of focused NPs, or more like the distribution of a pronoun with a predicate marker.

Assuming Tryon and the generalised *i* analysis are correct, we would expect to find that the relative frequency of doubled 1SG and 2SG subjects across different clause types resembles the relative frequency of focused 3SG subjects more than it resembles the relative frequency of *hem i + V* subject sequences. On the other hand, if the GCC analysis is correct, we would expect the distribution and relative frequency of doubled 1SG and 2SG subjects to be more similar to the distribution and relative frequency of *hem i + V* subject sequences than to the distribution and relative frequency of focused 3SG subjects.

5.2.2 Discourse role of subject referents as a diagnostic

All clauses in my corpus with singular subjects were extracted for comparison across a number of syntactic, semantic and pragmatic factors. In addition, the social information about speakers was used to test whether there are demographic factors interacting with speakers' production of and, I assume, their own underlying analysis of, singular pronominal subjects. The hypotheses and method can be summarised as follows:

HYPOTHESIS A:

Mi mi and *yu yu* sequences are a pronoun plus predicate marking (cf. GCC).
They are equivalent to *hem i* sequences.

HYPOTHESIS B:

Mi mi and *yu yu* sequences are focused 1SG and 2SG forms (cf. Crowley/Tryon).
They are equivalent to NP, *hem i* sequences.

METHOD: Compare distribution of doubled and singleton tokens of 1SG and 2SG subjects with:

- (a) *hem i*
- (b) NP, *hem i*
- (c) 3SG null subjects, $\emptyset i$

EVALUATE these distributions with respect to

- (i) pragmatic or discourse factors (information status of referent)
- (ii) social or demographic factors

The corpus was, therefore, coded for a number of linguistic and non-linguistic factors. Some of the linguistic factors (relating to the syntactic and semantic status of the clause) that were part of the coding scheme will not be discussed here. They are reviewed in the discussion of variation between phonetically null and pronominal subjects in Chapter 6 (§6.5.1). However, they are not relevant to the specific hypotheses being tested at this point.

The hypotheses that we are testing are that the status of the subject referent (focused or not) in the discourse is salient, and that there are non-linguistic (specifically, regional) factors which may also be interacting with the distribution of doubled and singleton 1SG and 2SG pronouns.

Thus, the corpus was coded for the following conditions:

- (i) the grammatical role of the current clause's subject in the immediately preceding clause (prior subject, prior patient or oblique, or not present in preceding clause)
- (ii) the form of the referent in the previous clause (phonetically overt or phonetically null).

These criteria were selected on the basis of previous cross-linguistic work on anaphor resolution and the distribution of phonetically null subjects, such as Centring Theory and switch-reference marking. These frameworks are discussed in detail in §6.2.1 and §6.2.2 respectively. The various approaches uniformly assume that interlocutors track the grammatical or linear status of referents from clause to clause. Therefore, since we are interested here in referents that are always the subject of the current clause, we want to know whether the same referent was present in the preceding clause, and if it was what its thematic or linear prominence was. In the vast majority of clauses, the order of elements in Bislama is SVO.

Crossing these conditions resulted in five different discourse states: tokens where the current subject was the subject of the preceding clause and the referent was realised overtly (pronoun or NP) in the preceding clause; tokens where the current subject was a phonetically null subject in the preceding clause; tokens where the subject of the current clause was a phonetically overt, non-subject argument in the preceding clause; tokens where the current subject was a phonetically null, non-subject argument in the prior clause; and tokens where the referent of the current clause's subject was something not even present in the preceding clause. Examples of interclausal transitions in all five conditions are shown below, where the clause taken to be the 'current clause' for the purposes of the calculation is highlighted by an arrow.

5.12 Current subject was (overt) subject in preceding clause

Afta mi singaotem hem
after 1SG sing.out 3SG

→ *mi toktok long hem*
1SG talk PREP 3SG

mi talem se
1SG tell COMP

"*Mi mi gobak long wok.*"
1SG 1SG return PREP work

Hem i laf we i laf.
3SG PRED.MKR laugh COMP PRED.MKR laugh
'So I called for him [and] I said to him [crossly], I said, "I'm going back to work." He was laughing and laughing.' (S-95-6, Elsina)

5.13 Current subject was (null) subject in preceding clause

Smolwan i stap
small.one PRED.MKR stay

Ø i stap glad nomo
Ø PRED.MKR stay glad only

→ *Ø i laf*
Ø PRED.MKR laugh

Ø i lukluk
Ø PRED.MKR look.look

from haos i ben olsem
because house PRED.MKR bend like

Ø i kam klosap long hem.
Ø PRED.MKR come close.up PREP 3SG

‘There the baby is, happy as anything, [it]’s laughing, [it]’s looking around, because the house has bent like [this], [it]’s fallen towards the baby.’
(S-95-9, Lisa)

5.14 Current subject was (overt) DO or OBL in preceding clause

Be from Visi i talem se
but because Visi PRED.MKR tell COMP

sapos wan i dring yis alkohol
if one PRED.MKR drink yeast alcohol

mi kasem hem
1SG catch 3SG

→ *bae Ø i pulum ol grasnil long misin.*
IRR Ø PRED.MKR pull PL needle.grass PREP mission
‘But because Visi said if anyone drinks yeast alcohol, [and] I catch them, [they]’ll be weeding at the mission.’ (M-95-15, Livae)

5.15 Current subject was (null) DO or OBL in preceding clause

Ating hem i fasin blong hem olsem
maybe 3SG PRED.MKR fashion of 3SG like

se taem yumi maredem Ø long san
COMP time 1PL.INCL marry Ø PREP sun

→ *Ø i hang*
Ø PRED.MKR hang

bambae Ø i foldaon nomo.
IRR Ø PRED.MKR fall.down only

‘I think that’s just the way it is, when you pollinate [the bloom] during the day, [it]wilts, [it]’ll just fall off.’ (M-94-5, Anita)

5.16 Current subject not present in preceding clause

Olgeta oli pas long intevyu yestede
 3PL PRED.MKR pass PREP interview yesterday

→ *tede nao mi mi harem.*
 today now 1SG 1SG hear

Mi kros we
 1SG cross COMP

man, mi harem nogud we.
 man 1SG hear no.good COMP

'They interviewed the day before, but I only heard the next day.
 I was so angry, man, I felt really bad.' (M-95-12, Dien)

The rationale motivating the selection of the social and demographic factors investigated are discussed in detail elsewhere (§3.7, §4.6 and §6.6). They included speaker's sex, their age, their marital status, level of formal schooling attained and the language of instruction, and – of greatest significance to Tryon's suggestion – place of origin.

5.2.3 Results: *mi* and *yu* are subject pronouns

The data were subjected to a multivariate analysis using the Goldvarb statistical analysis program (Sankoff et al. 1992) designed to evaluate the significance of factors which may significantly influence linguistic variables. The program analyses the different (independent) factor groups, but allows the researcher to evaluate the significance of individual factor groups by weighting the distribution of factors with respect to their relative frequency in the data and with respect to the frequency of factors in other groups.

Since, at this stage of the investigation, I was interested in comparing the relative frequency of 1SG/2SG and 3SG subjects in the focused and simple pronoun conditions, the form of the subject was used as the dependent variable for Goldvarb's step-up/step-down multivariate analysis of variation.

This revealed that non-linguistic factors do not significantly influence the form of the 1SG and 2SG subjects selected by speakers. Separate analyses of the variation for speakers from Malo (village) and speakers from Santo (urban) showed that in no case was the speaker's language of education, nor their marital status a significant factor. Since there was some slight indication that speaker sex, age, level of education and place of origin might significantly influence the linguistic variable, I adopted a conservative policy, and retained these factors in the analysis.

However, more revealing results could be obtained from the discourse factors investigated. Our task is to determine which of two proposed structural parallelisms is supported most by the data. We are trying to do this by comparing the distribution of the structures that are hypothesised to be analogous to each other across five discourse conditions. A multivariate analysis can tell us that the factor group coding tokens for the discourse status of the subject referent has a significant effect on the data being compared, but it cannot tell us anything further. Goldvarb is designed to test data where the dependent variable is binomial. But if we want to test whether the structures we are interested in are distributed in essentially the same way across five different discourse conditions, this entails a

multinomial dependent variable. Some other test of the distribution of the different forms is required. In this case, a rank order correlation statistic was performed.

The corpus of tokens to be compared consisted of 3,961 clauses. This was all the clauses in the corpus which had a 1SG or 2SG subject, either singleton or doubled. It also included all clauses with 3SG subjects, and those with a 3SG subject that was emphasised or focused by having the referent doubled before the pronoun in the form of a full NP. It also included all clauses where the subject was interpretively 3SG, but was phonetically null, leaving only *i* followed by the lexical verb. This condition was, of course, necessary to fully test the GCC hypotheses. If doubled 1SG and 2SG subjects are analogous to *hem i*, then *ipso facto* singleton 1SG and 2SG subjects are analogous to 3SG subjects where the pronoun is phonetically null, and only the predicate marker remains.

Chi-square tests were conducted on the observed Ns for each of the conditions to be compared, and these showed that the differences between the different linguistic forms in the five different discourse conditions were in all cases highly significant. A comparison of the absolute numbers of each subject type across all discourse states indicated that the data we are dealing with is highly robust:

- doubled 1SG and 2SG with focus 3SG, $\chi^2 = 62.8$, with 4df $p < 0.001$.
- singleton 1SG and 2SG with pronoun 3SG, $\chi^2 = 131$, with 4df $p < 0.001$.
- doubled 1SG and 2SG with pronoun 3SG, $\chi^2 = 31.5$, with 4df $p < 0.001$.
- singleton 1SG and 2SG subjects with null 3SG, $\chi^2 = 648.9$, with 4df $p < 0.001$.

Having determined that the data we are examining is sound and highly unlikely to have been drawn from the same sample, a test of the relative frequency of the data in the different discourse conditions was performed.

The following hypotheses were tested: 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 3SG *hem*. We should also find, *mutatis mutandis*, that the frequency of phonetically null 3SG subjects (with the form $\emptyset i V$) is much the same as bare verbs where the interpretive subject is 1SG/2SG (the form $\emptyset V$). And we should also expect to find that the frequency of focused 3SG 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 is summarised in Figure 5.1.

$$\begin{array}{ccc} \underline{mi mi V} \dots\dots & \text{versus} & \dots \underline{mi V} \dots\dots \text{and} & \dots \underline{\emptyset V} \\ \underline{NP, hem i V} & & \underline{hem i V} & \underline{\emptyset i V} \end{array}$$

Figure 5.1: Within group similarities expected in the distribution of 3SG and 1SG/2SG 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 3SG forms *i* which 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 3SG subjects, i.e. $NP_i, hem_i i V$. This is summarised in Figure 5.2.

$\frac{mi\ mi\ V}{hem\ i\ V}$ versus $\frac{mi\ V}{\emptyset\ i\ V}$

Figure 5.2: Within group similarities expected in the distribution of 3SG and 1SG/2SG 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 1SG/2SG subjects will be phonetically null as often overall, or even as often in a specific interclausal context as 3SG subjects are. The similarity that is required is a **within** groups similarity. This means that if an interclausal discourse context favours phonetically null 3SG 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 1SG/2SG.

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

- A* the subject of the (current) clause had also been the subject⁴ of the preceding clause;
- P* the subject of the (current) clause was some other argument in the preceding clause;
- N* the subject referent 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 realised versus 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.

Table 5.3 summarises the frequency with which the different subject forms were found in the five interclausal conditions.

	<i>A</i>	<i>P</i>	<i>N</i>	<i>A-o</i>	<i>P-o</i>
3SG	100%	100%	100%	100%	100%
1SG/2SG	100%	100%	100%	100%	100%

⁴ Bislama has no passive, so this condition generally reflects a continuity between clauses in thematic role as well.

Table 5.3: Number of tokens for each subject form in the corpus in five relations with identical referents in the preceding clause (*mi mi* represents both 1SG and 2SG data) [cells marked * those where tokens < 30, cf. Guy 1988].

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

An analysis of this data using Goldvarb 2.1 (Sankoff et al. 1992) was conducted. The logistic regression performed by Goldvarb weights the significance of the frequencies shown in Table 3, allowing, e.g. for the relative overall frequency of each interclausal context and for the relative frequency of each subject form in the corpus as a whole. The weightings provide a more accurate basis for comparing the within group frequencies than comparing simple percentages provides.

Logistic regressions of this type have provided an analytic framework for much variationist analysis, but the methods have some limits. The cells in Table 3 marked with an asterisk are cells in which there are fewer than 30 tokens. Guy (1988) shows that use of logistic regressions must be cautious when the number of tokens in some cells falls below 30. In this data, the cells of greatest concern are clustered in the *P* condition, which suggests 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 leaving out clauses where the subject had been a phonetically null non-subject argument in the prior clause (the *P* condition). But, as shown in Table 5.4, it was found that there was a significant difference between the two runs. This indicates that these tokens make a significant contribution to the overall analysis. Consequently, this factor is retained for the within-groups comparison.

Table 5.4: Testing the significance of the contributions of the tokens in *P* condition to the model's fit in all discourse conditions (all cdn).

LL = log likelihood

	LL (all cdn)	LL (no P)	df	p
<i>mi mi/ yu yu</i>	-486.591	-483.004	1	<0.025
\emptyset (1SG/2SG)	-361.077	-357.032	2	<0.025
<i>NP, hem i</i>	-246.193	-238.75	1	<0.001
<i>hem i</i>	-965.824	-922.704	1	<0.001

Subsequent runs evaluated the probabilities of *mi mi/you you* and \emptyset (1SG/2SG) 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 conclusive results obtain omitting the *P-o* condition. Naturally, the *NP, hem i* tokens remained in the analysis because they are central to the hypotheses.

The results of the logistic regressions are given in Tables 5.5–5.9 and Figures 5.3–5.7. The frequency of the different forms of subject in different discourse conditions is converted into probability weightings in the tables, expressing the degree to which each condition favours or disfavors the different forms of subject; the pair-wise comparisons laid out in Figures 5.1–5.2 are shown graphically in Figures 5.3–5.7 to facilitate comparison. The first set of tables and figures (Tables 5.5–5.7 and Figures 5.3–5.5) tests the hypothesis that *mi* and *you* are pronouns only; the second set (Tables 5.8–5.9 and Figures 5.6–5.7) tests the hypothesis that *mi* and *you* are both pronouns and subject-verb agreement markers.

Table 5.5: Weighting of *mi mi* and *NP, hem i* subjects compared (Analysis B predicts isomorphism).

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

Table 5.6: Weighting of *mi* and *hem i* subjects compared (Analysis B predicts isomorphism).

	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 5.7: Weighting of \emptyset (1SG/2SG) and $\emptyset i$ (3SG) subjects compared (Analysis B predicts isomorphism).

	A-o	A	P-o	P	N
\emptyset (1SG/2SG)	0.457	0.86	0.43	0.24	0.5
$\emptyset i$	0.389	0.77	0.47	0.41	0.29

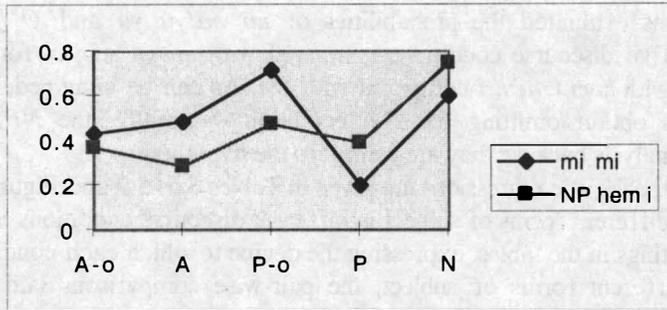


Figure 5.3: Weighting of *mi mi* and *NP, hem i* subjects compared. LINEST = 0.4908.

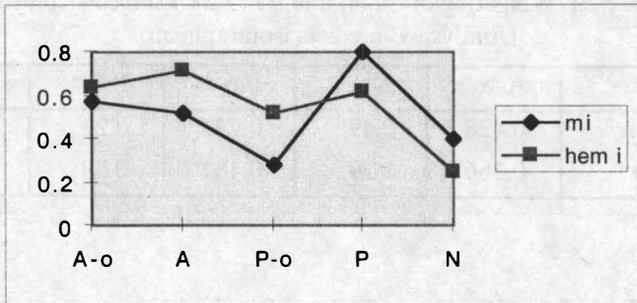


Figure 5.4: Weighting of *mi* and *hem i* subjects compared. LINEST = 0.4908.

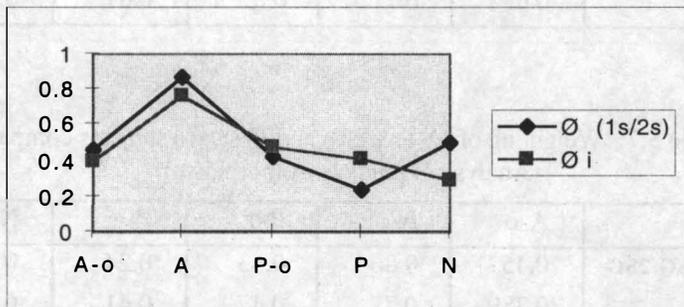


Figure 5.5: Weighting of \emptyset (1s/2s) and $\emptyset i$ (3SG) subjects compared. LINEST = 0.9711.

The probability weightings should be interpreted in the following way. A probability close to zero indicates that a particular subject form is strongly disfavoured in that interclausal context. A probability near one indicates that that form of a subject is strongly favoured in that interclausal environment. The degree to which there is comparable within group behaviour 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 5.3–5.5 show LINEST values that are consistently positive, and in the case of the hypothesised null subject condition (Figure 5.5) behaviour of 3SG and 1SG/2SG 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 3SG 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 3SG referent.

This is in marked contrast to the picture painted when testing the GCC analysis, as shown in Tables 5.8–5.9 and Figures 5.6–5.7.

Table 5.8: Weighting of *mi mi* and *hem i* subjects compared (Analysis C predicts isomorphism).

	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 5.9: Weighting of *mi* and $\emptyset i$ (3SG) subjects compared (Analysis C predicts isomorphism).

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

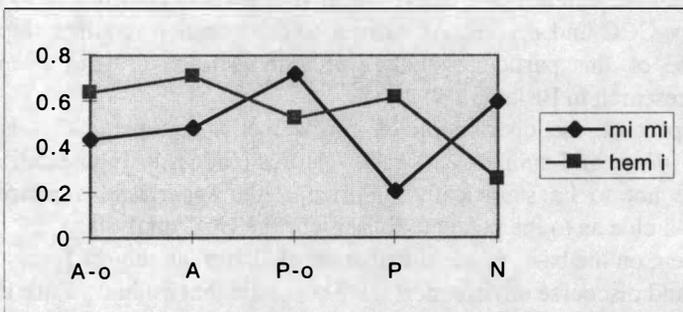


Figure 5.6: Weighting of *mi mi* and *hem i* subjects compared. LINEST = -0.491.

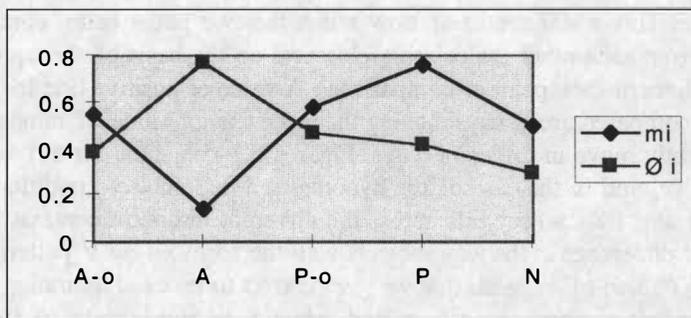


Figure 5.7: Weighting of *mi* and $\emptyset i$ (3SG) subjects compared.
LINEST = -0.971.

Figure 5.6 shows that 1SG/2SG 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 3SG subjects, while Figure 5.7 shows even more clearly that subjects with the form *mi* or *yu* alone do not pattern like 3SG 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. Although we saw good evidence in 5.11 and 5.12 to suggest that *mi* in these clauses might be comparable with the *i* in 3SG, and although Table 5.3 bears out the suggestion that doubled subjects are relatively more frequent in 1SG/2SG than they are in 3SG (c.10% of all clauses with 1SG/2SG subjects versus c.5% of all clauses with 3SG subjects), a detailed examination of how these forms are used indicates quite clearly that *mi mi* clusters are not the same as *hem i* clusters.

This data raises an interesting question about whether the doubling of 1SG and 2SG is actually interpreted by speakers of Bislama as a focus construction (i.e. the extent to which these forms are functionally analogous to *NP*, *hem i* utterances). Manessy (1995:126-127) points out that such doubling need not be interpreted as drawing attentional focus or giving emphasis. Regarding syntactically focused constructions in Abijan French, he notes

Un procédé fréquent, mais non exclusif, dans la langue parlée [français populaire d'Abijan] est la thématisation du substantif sujet ou objet, apparemment sans valeur d'insistance particulière, alors que celle du pronom est manifestement emphatique.

It is possible that the absence of a clear 'valeur d'insistance' contributed to the misanalysis of these forms by GCC and myself. An answer to this question requires empirical evidence beyond the scope of this particular study, yet it indicates a fruitful avenue for future psycholinguistic research in Bislama.

One other aspect of the distribution of forms merits comment. This is the similarity between doubled *mi mi* and simplex 3SG *hem i* clauses (shown in Figure 5.6). Although this similarity proved not to be statistically significant, the superficial similarity found here perhaps provides a clue as to the perceptual basis for the GCC analysis.

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 3SG and 3PL *oli* in these forms is a zero morpheme. The balance of evidence from several domains of the linguistic system supports the analysis of subject-verb agreement proposed by Tryon. Consequently,

single tokens of *mi* and *yu* can be analysed as pronouns. Having established this, this paradigm will form the basis in the next chapter for an examination of the circumstances under which subject pronouns surface overtly and are phonetically null.

As a final exercise, though, the unresolved question about the precise nature of the predicate marking with 1SG and 2SG subjects must be addressed.

5.3 Elision or absence: predicate marking with 1SG/2SG subjects

We have established that the distribution and use of doubled *mi mi* and *yu yu* sequences in everyday speech in Bislama is the same as focused 3SG subjects. Consequently, we reject the GCC analysis of the second *mi* and *yu* in these cases being the predicate marker. This section will evaluate the merits of Tryon's analysis (no, i.e. zero, predicate marker with 1SG and 2SG subjects) with the generalised *i* analysis, under which the predicate marker *i*, found elsewhere in the system, is present underlying in 1SG and 2SG, but is deleted in the environment of a preceding high vowel.

In the overview of the system we noted that the deletion hypothesis is supported by the reappearance of *i* with the 1PL(INCL.) *yumi* when an adverb (post)modifying the subject intervenes between the pronoun and the lexical verb, i.e. *yumi V*, *yumi evriwan i V*, but not **yumi evriwan V*. However, the deletion hypothesis is not supported by the behaviour of *mi* and *yu* when there is an intervening postmodifying adverb such as *nomo* or *wan*.

- 5.17 *Mi nomo mi nyuhebridian long taem ya.*
 1SG only 1SG new.hebridean PREP time SPEC
 'I was the only Ni-Vanuatu [there] at that time.' (S-95-15, Rinette)

- 5.18 *Bae olgeta oli stap, mi wan mi go.*
 IRR 3PL PRED.MKR stay 1SG one 1SG go
 'They could all stay here, I'd go by myself.' (M-95-12, Dien)

We can reject the possibility that the repetition of the subject pronoun here is an obligatory consequence of *nomo* and *wan* modification creating focus constructions. It is true that *mi wan* and *mi nomo* have the effect of making the speaker's reference to self highly salient in the discourse, but *nomo* and *wan* do not of themselves necessarily generate focus constructions and can occur following an NP that has been fronted for focus, for example:

- 5.19 *Hem nomo i olsem.*
 3SG only PRED.MKR like
 'He's the only one who's like that.' (M-94-2, Leipakoa)

- 5.20 *Tammy hem wan i go.*
 Tammy 3SG one PRED.MKR go
 'Tammy went all on her own.' (S-95-11, Juliet)

Furthermore, if the absence of *i* following *mi* and *yu* is due to its assimilation with a preceding high vowel, we might expect the rule to apply generally, i.e. for the predicate marker to delete whenever it is immediately preceded by a high vowel. However, this is not the case in Bislama (nor in Tok Pisin, as Verhaar [1995:80] notes, disputing an analysis of phonological assimilation for Tok Pisin, too). It is relatively easy to find examples where non-subject *mi* and *yu* occur phrase finally, immediately before the predicate marker. In examples 5.21–5.22 the complex NPs under investigation are enclosed in square brackets.

- 5.21a. *Afta mi pusum [hed blong mi] i go.*
 after 1SG push head of 1SG PRED.MKR go
 ‘So then I pushed my head through.’ (M-95-19, Mesek)
- b. *Mo gud sapos yu bin harem [mama blong mi] i*
 more good if 2SG ANT hear mother of 1SG PRED.MKR
talemaot storian ya we.
 tell.out story spec COMP
 ‘It would have been much better if you’d heard my mother tell that story.’
 (S-94-5, Ezra)
- 5.22a. *[Woman blong yu] i kam long wea?*
 woman of 2SG PRED.MKR come PREP where
 ‘Where does your wife come from?’ (S-94-3, Simeon)
- b. *Taem yu wokabaot [leg blong yu] i draon.*
 time 2SG walk.about leg of 2SG PRED.MKR drown
 ‘When you walk across it, your legs sink in’ (M-94-7, Tarip)

And we find similar patterns when a proper name ends with a high vowel, e.g.

- 5.23a. *Tammy i no save.*
 Tammy PRED.MKR NEG know
 ‘Tammy didn’t realise it.’ (S-95-11, Juliet)
- b. *Ale Lili i putum top blong rop long han blong hem.*
 go Lili PRED.MKR put top of rope PREP hand of 3SG
 ‘So Lily put the end of the vine on her hand.’ (S-95-10, Rebeka)
- c. *Wili i talem se bae mi aplae.*
 Wili PRED.MKR tell COMP IRR 1SG apply
 ‘Willy said I should apply.’ (S-95-6, Elsina)

In the entire corpus, subjects that have non-subject *mi* and *yu* finally, are followed by the predicate marker *i* 15 out of a total 18 times (83%). As Table 5.10 shows, proper names ending with a high vowel show a slightly different pattern. These are almost equally split between having *i* or null predicate marker following (12 versus 14 respectively). That is, proper names apparently favour deletion more than pronouns.

Table 5.10: Presence and absence of /i/ predicate marker when immediately preceded by a high vowel (non-subject *mi* and *yu* and Proper Names). All speakers.

Subjects with /i/ or /u/ final	Form of predicate marking		Total
	<i>i</i> + V	Ø + V	
NP + <i>blong mi</i>	9	3	12
NP + <i>blong yu</i>	6	0	6
Proper name	12	14	26
Total	27	17	44

The data with non-subject *mi* and *yu* shows that the predicate marker is absent only occasionally (16% of the time) and we do not know whether this is significantly different from the frequency with which the *i* predicate marker is omitted in casual speech in Bislama with all subjects. Deletion of the predicate marker does occur in Bislama (cf. fn.3 above), though with considerably less frequency than is found in urban varieties of Tok Pisin, e.g. Sankoff (1994). (Smith (1995) shows that the urban dialect differs from others in this respect. His corpus indicates that the predicate marker continues to be used frequently in Highlands Tok Pisin.) It remains to be determined whether *i* deletion in Bislama is a stylistic device associated with particular registers or speech events, or whether it is an early indicator of an ongoing reanalysis of the grammar similar to that occurring in urban Tok Pisin. However, I will not attempt to resolve this question here. I will, instead, turn to the question of how and why the grammatical class of a word affects the application of a phonological rule.

Nespor and Vogel (1986) present data from Bantu which seems similar to the data presented here. They illustrate how phonological rules are often sensitive to the structure (or lack of syntactic structure) of the constituents involved. The general principle underlying processes of optional restructuring is essentially that a simplex constituent, particularly one that is phonologically light, may attach to an adjacent constituent especially if this too lacks phonological weight. Extending this account to the data here, if we assume a simple, non-branching structure associated with pronouns such as *mi* and *yu*, this might explain why the *i* of the verb phrase does not delete in examples 5.21–5.22. Both constituents are phonologically light but because they are adjacent to one another, they are able to restructure, generating a derived constituent of more optimal weight.

The distribution of *i* with proper names, on the other hand, suggests they only optionally restructure. The names are all two syllables, which in itself means that they have more phonological weight than the pronouns *mi* and *yu* do. Moreover, it is not unreasonable to assume that proper names are associated with a more complex syntactic and semantic structure than pronouns. Since they are unique referring expressions, they may be associated with some kind of null determiner indicating the specificity of their reference.

But note that under this account we would expect to also find *i* surfacing following tokens of subject *mi* and *yu*, yet this is simply unattested. Even Crowley's heterogeneous description of predicates with 1SG and 2SG subjects gives the alternation as being between zero and markers identical to the pronouns (1990:235). Consequently, the data reviewed in this section suggests that we can with some degree of confidence reject the generalised *i* analysis.

In conclusion, then, the synchronic distribution of subject pronouns favours Tryon's analysis of the subject-verb agreement system as being one in which the agreement marker with 1SG and 2SG subjects is uniformly phonologically null, i.e. a zero. It is more economical to analyse utterances with the form of example 5.4 as having a focused pronoun followed by an adverb, an unfocused subject and zero subject-verb agreement (as Tryon does), than to describe agreement as having alternative forms (Crowley 1990:234–235). That is:

5.24	<i>Mi</i>	<i>nomo</i>	<i>mi</i>	Ø	<i>nyuhebridian</i>
	FS	ADV	SBJ	AGR	new.hebridean
	*SBJ	ADV	AGR		new.hebridean

Under this account, an utterance with the form of 5.4c *bae mi go spel* would be analysed as having the irrealis adverb *bae* preceding the subject and the verb phrase.

This account still leaves the reemergence of *i* in examples 5.7–5.9 unexplained. The next section presents some arguments intended to motivate the appearance of *i* as an agreement marker in the system as a whole.

5.4 *I* still means ‘he’: a functional motivation for zero marking

In addition to the synchronic evidence which suggests that Tryon’s analysis of a zero predicate marker is correct, it is worth noting that this analysis enables us to tell a coherent story about the diachronic development of the agreement system in Bislama. This in turn will help us account for the occasional appearance of *i* with the 1PL(INCL.) *yumi*. Consider the historical source of the predicate marker *i*.

Both Sankoff (1977) and Crowley (1990) show that the predicate marker was originally derived from the English 3SG subject ‘he’, and that it occurred with 3SG subject referents before generalising as a predicate marker which could occur with subjects of other persons and number (perhaps as early as the 1890s, according to Keesing 1991). Clearly, this meaning has been bleached by now, since it is the default agreement marker occurring not only with 3SG, but also with 1PL(EXCL.), 2PL, and some 3PL subjects (I will have more to say about default predicate marking in the discussion of serial verb constructions in Chapter 7).

But notice that these are all pronouns whose reference is drawn from sets which do not specifically exclude the third person. In the case of 3SG, the set is only a non-participant in the conversation, i.e. the set of referents is {3}. The exclusive 1PL *mifala* takes its reference from the set of {1, 3}, while 2PL *yufala* takes its reference from a set which excludes the speaker, includes the addressee(s) and may include, but crucially does not exclude, third parties, i.e. {2, (3)}. And the 3PL pronouns which occur with *i* predicate marking, e.g. *tufala* ‘they (2)’, *trifala* ‘they (3)’, take their reference from a set excluding speaker and addressee, and including third parties, i.e. {3}.

On the other hand, both 1SG and 2SG draw their reference from highly restricted sets which, by virtue of their singularity, explicitly exclude non-participants, i.e. consist of {1} and {2} respectively. We know that the generalisation of predicate marking in Bislama occurred gradually, and its patterns would seem to accord with a gradual process of semantic bleaching of *i*. Crowley (1990:244) shows that in Bislama it had been reanalysed by speakers as a predicate marker with 3SG and 3PL by 1920. The synchronic distribution of *i* as a predicate marker suggests that this process of semantic bleaching has been arrested before completion. In other words, *i* is not fully grammaticised in Bislama, instead it retains vestiges of its original meaning of 3SG (< ‘he’). Thus, there is a referential clash between the predicate marker *i* and the subjects *mi* and *yu*. Thus, in addition to the descriptive reasons just adduced, there is a semantic motivation for a null predicate marker in 1SG/2SG.

Earlier, I set aside discussion of the 1PL(INCL.) subject *yumi*. This, and the 1PL(DU) *yumitu*, also do not mark the predicate with subject-verb agreement using *i* when the pronoun occurs unmodified. However, in examples 5.8–5.9 we saw that *i* did appear following *yumi evriwan* in the Bible. This contrast between *yumi* (‘speaker and addressee(s)’) and *yumi evriwan* (‘we (include.) everyone’) can be neatly accounted for by the analysis proposed here. *Yumi evriwan* includes third parties (i.e. more than just the writer and reader of the Bible). If *i* has not been entirely bleached of its original semantics, and continues to designate non-participants, we would predict precisely this contrast in Bislama. *Yumi* (and likewise *yumitu*, *yumitri*) draws its referents from the restricted set of {1, 2} unless non-participants are

specifically included as with *evriwan*, in which case the set extends to {1, 2, 3}. Thus, it is not the case that all *yumi* subjects have a predicate marker *i* underlyingly, rather the predicate marker *i* is inserted in just those cases when the field of reference is extended to include third parties.

This analysis may not fully account for all the data found with *yumi*. In example 5.7, repeated here, we saw that an adverb, such as the irrealis marker *bae*, when inserted between *yumi* and the verb, is also an environment in which *i* surfaces. It has been claimed that this also holds for *yumitu*.

5.7 *Be yumi bae i no save stopem maot blong yu.*
'But we can't close off your mouth.' (M-94-2, Leipakoa)

Assuming there is indeed a robust and systematic correlation between the presence of an adverb and the appearance of *i* agreement (and this remains to be tested, examples of this kind are rare in the corpus), then this fact is not predicted by the analysis here. If it is correct that the synchronic distribution of *i* is constrained by its historical association with 3SG reference, then there is no reason why it should be dependent on word order. This indicates clearly that the current distribution of *i* is not determined entirely by its historical derivation. Decades of semantic bleaching have led to some reanalysis. It will be seen in Chapter 7 that other syntactic facts suggest that the synchronic status of *i* is also the default subject-verb agreement. That is, *i* surfaces when intervening lexical material prevents the head of the subject pronoun from directly c-commanding the head of INFL.

The analysis of 5.7 would then be that *bae* is a barrier and prevents the subject entering into the requisite specifier-head relationship that is required for full and proper subject-verb agreement to surface. Indeed, since *bae* behaves far more like an adverb than like verb morphology in Bislama (in terms of freedom of placement), it is quite possible that this is in fact what has happened here. However, a good deal more work needs to be done in order to determine precisely what the relationship between *bae* and the rest of the clause is.

Let me be clear about the strength of my claim. I am suggesting that some synchronic aspects of the pronoun-predicate marking system in Bislama are reflexes of a more compositional, historical state of the grammar. However, once *i* had undergone a certain amount of semantic bleaching, and was on the way to becoming the default predicate marker for Bislama, there is nothing to prevent it being generalised with either new pronouns that might be added to the system, or in syntactic contexts in which subject and verb are sufficiently distant that some form of default agreement might be required.

I am definitely not suggesting that speakers of Bislama in the 1990s actively perceive *i* to have third person reference and that they therefore actively perceive this to be semantically anomalous with 1SG, 1PL(INCL) and 2SG pronouns. Rather, I am suggesting that this was, at an earlier stage of Bislama when *i* was still more or less interpreted as being a 3SG pronoun, an active consideration. Since then, *i* has become sufficiently semantically bleached that morphosyntactic considerations play a much larger role in determining which clauses or verbs will be marked with *i*. Nor does the historical derivation of this form mean that in the future speakers might not continue to generalise the use of *i* with other pronouns. The reported synchronic variation between *yumi(tu) i* and *yumi(tu) Ø*, for instance, might be the consequence of a clash between older, semantic factors governing the use of *i*, and newer, structural norms.

The forms of subject-verb agreement in Bislama seem to provide us with an example of the way in which synchronic variation in the grammar may reflect obsolete grammatical

distinctions. Labov (1989) discusses the case of variation between /ɪn/ and /ɪŋ/ in English, and shows that the relative frequency of the two forms continues to be partly constrained by the historical derivation of the word. The morphology that differentiated the two word classes (adjectives versus verbs) has long since merged in standard English, so that ‘a *shining* star’ and ‘a star is *shining*’ are morphophonemically identical. Nonetheless, a small but significant factor in the synchronic variation is whether or not the token corresponds to what was historically realised as *-inde* or *-inge*.

5.5 Conclusion

The subject-verb agreement paradigm in Bislama is essentially that proposed by Tryon. There are three subject-verb agreement markers:

Table 5.5: Subject-verb agreement paradigm in Bislama.

Form	Context
zero	1SG, 2SG and 1PL(INCL) subjects
<i>oli</i>	3PL subjects
<i>i</i>	elsewhere

The synchronic stability of the system can also be explained by the regularity this situation generates. At first it might seem that the predicate marking system labours with some highly salient exceptions. It has been argued that 1SG and 2SG are conceptually salient (Mithun 1991:174), and that from this follows their discourse salience in all languages (Chafe 1994) and their morphosyntactic salience in some languages (Rohrbacher 1995; Vainikka & Levy 1995). It might seem odd, then, that of all persons and numbers, Bislama marks first and second person subjects with the phonetically least salient of agreement inflections and with the paradigmatically most marked form.

However on reflection, it becomes clear that the Bislama predicate marking system is synchronically quite balanced, and the balance it has achieved is interpretively salient. For all persons, predicate marking distinguishes between singular and non-singular referents (unambiguously identifying the singular subject for each person). Though this subject-verb agreement is much more impoverished than the agreement found in other languages that allow phonetically null subjects, the interpretation of null subjects in Bislama must certainly be boot-strapped to some extent by this number distinction.

This chapter has examined and evaluated the distribution of 1SG, 2SG and 3SG subjects in Bislama with respect to a number of independent linguistic and non-linguistic factors. I showed that the most salient dimension was the information status of the subject referent. An analysis of 3,961 clauses occurring in natural conversation showed that the distribution of doubled 1SG and 2SG, *mi mi* and *yu yu*, is directly comparable with the distribution of the clearly focused 3SG subjects with the form *NP, hem i*. I concluded, therefore, that *mi* and *yu* are genuine pronouns, and not predicate markers. Consequently, it is clear that the so-called predicate marker is in fact subject-verb agreement morphology. In the following chapters, *i* and *oli* will cease to be glossed as ‘PRED.MRK’. Interlinear glosses will instead use the abbreviation AGR for subject-verb agreement.

I then considered the evidence for 1SG and 2SG subjects having the underlying form of the default subject-verb agreement *i*. The behaviour of *i* with other non-1SG and 2SG subjects showed no evidence for a general phonological rule of *i* assimilation that might be applying with 1SG and 2SG subjects. I concluded that the predicate marker with 1SG and 2SG subjects is zero. This work has set the groundwork for an analysis of clauses with phonetically overt and phonetically null subjects. In the next chapter, I will turn to defining the conditions and factors that influence this.

6 From wanem nao yumi no nidim wan jenis? *Constraints on phonetically null subjects in Bislama*

6 Introduction

In this chapter I describe the conditions under which we find phonetically null subjects in Bislama. I will distinguish between clauses where the subject is obligatorily null, and clauses where there is variation between null and overt pronominal subjects. In clauses where the subject variably occurs overtly, I will describe the conditions under which subjects are phonetically null and under which subjects are realised as pronouns, and I will account for the factors that govern the selection of either option. I will show that phonetically null subjects are primarily a language specific morphosyntactic phenomenon, but that they are also the consequence of the application of more general constraints on information tracking. Social factors are not by and large found to be significant. This indicates that with respect to this variable, we are dealing with a relatively homogeneous and stable community grammar.

6.1 Generative accounts of phonetically null subjects

A considerable amount of energy has been given over to the analysis of languages such as Spanish and Japanese that do not require the subjects of finite clauses to be overtly realised, and to the specification of how these languages differ from languages such as English, which do require such subjects to be overtly realised. Generative analyses of the so-called *pro*-drop languages have observed that several syntactic properties are associated with the facility to have phonetically null subjects in languages like Spanish, and that there are marked qualitative differences between *pro*-drop languages with rich inflectional subject-verb agreement morphology and those with an absence of such agreement. I do not intend to provide a detailed survey of the history of generative analyses of *pro*-drop languages, but in this section I will review some of the more important generalisations and trends that have emerged in the literature that are germane to my analysis of null subjects in Bislama, paying particular attention to the analysis of split systems of *pro*-drop.

It was initially hypothesised that an account of the so-called 'pro-drop parameter' would have to explain the co-occurrence of a number of syntactic properties. Comparisons of English and Italian, for instance, noted that the two languages differ not only in whether they allow a phonetically null subject (both in main clauses and in subordinate clauses), but also in a number of apparently related phenomena. These included the fact that Italian allows postposing of subjects (without, it was claimed, any stylistic marking), and the fact that Italian allows subjects to undergo long distance movement, and movement which apparently violates the *that-trace* filter noted for English (Rizzi 1982). Naturally, these facts posed rather a problem for a unified cross-linguistic account of grammar based on constraints on movement and the proper status of the empty categories left after movement.

The solution proposed to this theory-internal quandary made use of another obvious difference between Italian and English. Rizzi (1982) suggested that in Italian the subject is actually base-generated in a postverbal position (thereby arguing that the movement of subjects in Italian is chimerical), and the preverbal subject position is an empty category. The claim was that this empty category, given the name *pro*, was governed and given the necessary interpretation for person and number by the rich subject-verb agreement paradigm of Italian. The inherent link between phonetically null subjects and subject post-posing has subsequently been disputed (Roberge 1990:26 for summary), as has the inherent significance of subject-verb morphology that uniquely distinguishes between all persons and numbers. Huang (1989) and Borer (1989), for instance, questioned the need for two empty argument categories, *pro* and PRO, and pursued possible unifications of these. Clearly, the interpretation of PRO (under Control by a verb such as 'promise', or as PRO_{arb}) need not be by means of identification through the verbal morphology.

It was quickly pointed out, too, that languages with absolutely no subject-verb agreement, such as Chinese and Japanese, also have phonetically null subjects (Huang 1984). The proposed parameter was revised to a requirement for 'morphological uniformity', i.e. null subjects were posited to be syntactically licit when a language has an all or nothing inflectional paradigm, either unique identification of all persons and numbers, or no inflections at all (Jaeggli & Safir 1989:29ff.). While there was a pleasing categorical sound to this generalisation, the two types of languages remained a disjunctive pair. Huang (1989) suggested that in languages with no subject-verb agreement inflection, the identification of the subject referent was through discourse operators. That is, even while he attempted to locate the relevant operators in the syntactic structure, he was obliged to shunt the identification of the null arguments in languages without verb inflection out of core syntax and into the discourse component.

Morphological uniformity was pretty much state of the (generative) art when the articles in Jaeggli and Safir (1989) were collected. So much so that Jaeggli and Safir risked putting morphological uniformity forward as a maxim. But even in this collection, problems with the generalisation were being raised. Borer (1989) drew attention to languages in which phonetically null subjects are grammatical in some persons and numbers, and in some tenses, but not in others – something we will discuss in detail shortly, as it is of direct relevance to Bislama.

In truth, the promise that the theory of morphological uniformity held out was fleeting. It encompassed more *pro*-drop systems than any stipulation about 'rich' agreement did (such as Rizzi 1982), but in addition to problems in accounting for split systems, morphological uniformity fails to account for morphologically uniform systems that are not *pro*-drop, such as Danish. Nor does it explain why languages that make the same number of distinctions in

their subject-verb agreement paradigm may be parametrised as either *pro*-drop (Portuguese) or non-*pro*-drop (German) (Rohrbacher 1994:260).

Speas (1996) pursued the apparent correlation between morphological richness and phonetically null arguments. Ultimately, she derived *pro*-drop from constraints on structure building. She argued that an XP will only be projected if the head or specifier of the XP has phonological or semantic content (her Economy of Projection, 1996:203). The rich agreement morphology associated with a *pro*-drop language, she argues, is base generated in the head of AGR. This alone provides the substance required for an AgrP; consequently there is no need for an overt subject on structural grounds. Interpretively, too, the rich agreement is able to identify the person and number of the referent. The requirement for an overt subject in a language like English lies then in its weak agreement subject-verb morphology. Because this is not generated in the head of AGR, an overt subject must move to its specifier.

For Speas' analysis to work, though, we still need some way of determining whether an agreement paradigm is sufficiently rich, or strong, for the agreement to be generated in the head of ARG and to identify the subject referent. As Jaeggli and Safir (1989:29) point out, an adequate definition of 'rich' has always proved elusive. Rohrbacher (1994, 1995) provides a serious attempt to remedy this, giving a precise specification of the minimal referential distinctions that must be made in the morphology. Ultimately, he grounds this in principles of real world salience. He believes that if first and second person are uniquely marked by distinct agreement affixes, this may have further ramifications with respect to Germanic V2 and the possibility of *pro*-drop. I will return to the functional component of his analysis in §6.10.1.

6.1.1 Split *pro*-drop systems

As I have already intimated, languages that allow phonetically null subjects in only a restricted set of syntactic domains are something of a problem for most analyses. The question is what motivates splits at particular points in a system, and how best to analyse split *pro*-drop systems syntactically.

Referential versus non-referential subjects

Frequently, the only subjects that a language will allow to be null are non-referential 3SG subjects, i.e. the non-referential subjects found with 'weather' verbs, or those known as 'expletive' subjects in generative analyses of syntax, or 'dummy' subjects in the functionalist literature (Givón 1984:89).

These are analogous to English 'there', and perhaps English 'it'. 'It' differs from 'there' in that 'it' can be a sentential or propositional anaphor, while 'there' is completely non-referential. Iatridou and Embick (1997) note that in this respect 'it' differs from *pro* in *pro*-drop languages like modern Greek and Italian. They suggest the difference arises from the fact that for full interpretation *pro* must be co-indexed with a referent that is fully specified for person, number or gender, whereas 'it' is underspecified for number, and thus can be used felicitously with antecedents such as clauses or propositions.

A well-known example of a language that allows non-referential subjects to be phonetically null is German. German allows non-referential subjects to be phonetically null

when they occur in subordinate clauses and with dative experiencer verbs. Rohrbacher (1994:258ff.) reviews evidence that non-referential null subjects often pattern in ways that are quite distinct from referential null subjects, and suggests that null non-referential subjects may not project structure in the syntax. He argues that our theory of subjects should not, therefore, automatically treat expletive and referential null subjects as a single class. It would be an empirical question whether for any given language they are the same or different.

This view of subjects thus emphasises the difference between subjects which are thematic arguments of a lexical predicate and those that appear to be simply structural place holders. However, Heycock (1994) argues that predication is a syntactic primitive, and that every [+V] category is a syntactic predicate. Because every predicate requires an argument in a position of mutual c-command with the predicate, she argues that expletive subjects are syntactically no different from referential, theta-marked subjects. A split between phonetically null expletives and overt referential subjects in Heycock's analysis reduces to a lack of a theta role assigned to expletives, not a structural difference.

Person and number distinctions

However, there are also languages where referential null subjects are restricted to a certain subset of persons or number, or are restricted to specific persons in different tenses or modalities. Languages as typologically diverse as Hebrew and Finnish (Borer 1989; Vainikka & Levy 1995) have such a restriction on null subjects. In both Finnish and Hebrew, phonetically null subjects are attested to only be grammatical with first and second person subjects, and in Hebrew the distribution is further restricted to only first and second persons in the past and future tenses. As we will see, in §6.10.1, these authors argue that the possibilities for *pro*-drop are causally related to the extent of morphological transparency in these persons and numbers.

This kind of split system (with first and second person favouring null subjects and third person favouring independent NP subjects) may be even more widespread than this. The variation in Samoan between independent pronouns (which occur in post-verbal position and receive ergative case) and clitic pronouns (which occur preverbally), seems to be a closely analogous case (Cook 1994).

- 6.1a. *Na tutu e a'u le mōli.*
 past light ERG I the lamp
 'I lit the lamp.'
- b. *Na 'ou tutu-ina le mōli.*
 past I light-Cia the lamp
 'I lit the lamp.'

(Examples from Milner 1966:209; Cook's 1994, 5a-b)

Note that 'ou has the same form with both transitive and intransitive verbs, that is, there is no ergative-absolutive case distinction as there is with the independent pronouns *a'u* which is marked with the ergative in 6.1a. According to Cook (1994:60), the *-Cia* suffix (often described as a transitive) functions to focus the patient, or defocus the agent, or both. It is sometimes called the passive.

Cook attests (without quantification) that 1SG and 2SG clitic pronouns pattern rather differently from 3SG (1994:67). He suggests that the post-verbal position is associated with

'highly focused' referents (1994:61); the preverbal clitic position with more predictable information (1994:68). The situation Cook describes looks very similar to a split *pro*-drop system where third person subjects are more likely to be realised with an overt pronoun than first and second person subjects (which are more likely to be realised as verbal inflections). Since Cook's account of the distribution of clitic pronouns and the suffix *-Cia* is a bit circular – he suggests both *-Cia* and the clitic subjects are indicators of agent defocusing, and that they therefore predict the appearance of each other – a promising alternative might be to attempt a reanalysis of the Samoan data, assuming Samoan to be a split *pro*-drop system.

Cameron (1993), reports that Kameyama (1985) gives four languages which she claims only allow *pro*-drop in the third person, Rapanui, Tagalog, Khalka and Mota. However, Kameyama's account for at least two of these languages is suspect. Cameron (1993:328) shows that the Tagalog data is dubious evidence of *pro*-drop, and it seems that Kameyama misanalysed Rapanui. Du Feu (1996:65, 140) states that Rapanui pronouns are optional for all persons and numbers, both in main clauses and subordinate (finite and non-finite) clauses. That these are genuine null referential subjects is shown by the morphological contrast between the null subject in 6.2 and the agentless passive form in 6.3.

6.2 *He heriki i te 'ana hai mauku.*
 ACT strew RLT SPE cave INST tufted.grass
 '(We) used to strew the cave with tufted grass.'

6.3 *He heriki te 'ana hai mauku.*
 ACT strew SPE cave INST tufted.grass
 'The cave was strewn with tufted grass.'

(Du Feu's abbreviations: ACT = action [i.e. verb unspecified for tense/mood, MM]; RLT = relational particle [cf. accusative, MM]; SPE = specific; INST = instrumental)

Prince (1996) notes that traditional grammars of Yiddish have claimed that null subjects can only be used when the referent is 2SG, but she shows that this is in fact a misperception. She suggests that the illusion of a split *pro*-drop system may arise from the somewhat different discourse constraints that govern the felicitous use of phonetically null 2SG subjects and null subjects of other persons and number.

In sum, languages that split genuine, i.e. referential, *pro*-drop systems are not abundant, but they are sufficiently common that they pose a very real and interesting problem for the structural treatment of *pro*-drop. They raise questions about the uniform generation of inflections in the verb paradigm, and the way different inflections are interpreted. The place at which systems split also invites attention. Are the restrictions on which subjects may be phonetically null predictable cross-linguistically, or are the constraints always going to be idiosyncratic to each language? The distribution of phonetically null subjects in Bislama will provide data that goes some way towards resolving these issues.

6.2 Pragmatic or discourse-based accounts

Close investigation of even supposedly canonical *pro*-drop languages such as Italian show that there is more to the interpretation of phonetically null subjects than just the interpretation of referential information encoded in the morphology. Phonetically null subjects may be licensed by structural factors, but their fullest interpretation depends on interlocutors keeping

track of the information status of the referent as well. Even in languages like Italian and Spanish, archetypes of *pro*-drop, the speaker cannot launch into a new topic of conversation with a phonetically null subject and expect the hearer to accurately retrieve the referent unless it is given by the context (e.g. first and second person).

6.2.1 Reference tracking and algorithms of salience

There appears to be very general agreement about the nature of the interclausal dependencies required for the felicitous use and interpretation of pronominal or zero anaphora, namely the antecedent referent cannot have occurred too far back in the discourse. Immediately, of course, the more challenging question of measuring distance in discourse is raised. Some approaches have assumed that recency or distance are essentially linear, or surface level metrics. Others have taken discourse to have more structure, and assumed that the hierarchical prominence of a referent in the discourse structure is more important than linear order in resolving anaphoric reference. I will review some exemplars of both approaches as an indication of how the measures I have used in this work fit into the overall tradition.

Work within the functionalist paradigm tends to emphasise the importance of recency of mention in the selection and interpretation of anaphors. The underlying, and sometimes clearly articulated, assumption is that limitations on human memory radically constrain the number of referents that anyone can keep track of at one time. For example, Givón's discussion of the formal realisation of what he calls 'topics' deals in part with constraints on anaphoric reference. He argues that cross-linguistically a topic is more likely to be referred to anaphorically than anything else in the sentence, but that locality is a major factor in licensing anaphors, 'zero anaphora requires the shortest referential distance to the left – one clause – plus most commonly subjecthood' (1984:403). That is, he proposes the main constraint is linear locality, but he recognises that structural prominence plays a role too. In fact, it is a little hard to pin down precisely what Givón thinks are the most salient factors and what their relationship to each other may be, as at various points in his work he highlights the importance of thematic role, linear precedence and sheer temporal recency (1984:907-909).

Chafe (1994) also builds his analysis of referent tracking in discourse around the premise that 'the number of different referents that can be active [in the interlocutors' consciousness] at the same time is very small' (1994:79). But he concedes that how many and the processes by which interlocutors ensure a referent stays active or allow others to become backgrounded are still poorly understood. With third person referents, he also finds that the referent has to be frequently refreshed through use, in order for it to remain active, and for speakers to be able to refer to it with an anaphor.

Fox (1987) and Webber (1991) propose different models of discourse structure which nonetheless share the notion of discourse being constructed in terms of dominance and adjunction relations much like the relations that exist between intrasentential elements. Anaphor use, therefore, is constrained by the location of the antecedent referent within the discourse structure. There must be a unidirectional link between the anaphor and the referent. Thus, when speakers use an anaphor to felicitously refer back to a referent that may have occurred a considerable distance back in the discourse (as they often do), the felicity derives from the structure provided by the discourse itself.

The idea that anaphor resolution is calculated on the basis of intersentential relations, some of which are structurally more salient than others, is also the basis for Centring Theory. Grosz, Joshi and Weinstein (1995) outline a computational algorithm for the interpretation of pronouns that distinguishes three units of computation, and two intersentential relations. The mechanism for anaphor interpretation that they propose is, therefore, one of considerable computational economy. The basic units required for the computation are a list of the referents in the current utterance, known as a list of 'centres', the 'backward looking centre' and the 'preferred centre'.

Each referent in the list of centres (C_f list) is a logically possible candidate for subsequent anaphoric reference, but they are not all equally likely candidates. The preferred centre (C_p) is the first element in the current list of centres and is the element that is most likely to be carried over into the next utterance (Brennan et al. 1987). The backward looking centre (C_b) in any given utterance is that referent that was ranked most highly in the list of centres in the previous utterance (thus, there is only ever one C_b). The theory holds that if anything in the current utterance is realised as an anaphor, the backward looking centre will be (of course, more than just the C_b may be realised as an anaphor). Examples are shown below using a text from the Wellington Corpus of Spoken New Zealand English.¹

- 6.4a. *My grandfather took over Abraham and*
 b. *one of the grandsons went back and took that name*
 c. *but he he made it ah Abram.* (JJ8, Wai)

line 6.4b: C_f list = {one of the grandsons; that name},
 C_p = one of the grandsons,
 C_b = that name

line 6.4c: C_f list = {he, it, Abram}, C_p = he, C_b = he

As the theory would predict, if anything in the C_f list is an anaphor, the C_b is. In line 6.4b, the anaphoric element is *that (name)*, referring back to 'Abraham' in line 6.4a. In line 6.4c, both the C_b , *he* and another element on the C_f list are anaphors.

The relations between the C_p and C_b of successive utterances are used as a measure of the smoothness and computational complexity of the intersentential transitions, which account for the perceived coherence of the discourse. Tests of Centring Theory on natural language indicate that there is a positive, though not perfect, correlation between the smoothness of the intersentential transition and the form of a referent. In many cases, the smoothest and least computationally complex transitions are most likely to be realised by zero anaphors (in languages allowing these) or pronouns than a full NP.

Much work has been done to investigate the different ways languages rank entities on the C_f list, because the assignment of C_p status is crucial to the perceived smoothness of subsequent transitions. Linear order seems to be an important factor. In an SVO language like English, the C_p will often be the subject (as indicated in the C_f lists given for 6.4b and 6.4c above), but it need not be. Rambow (1993) found that in German, leftward scrambling

¹ Thanks to the School of Linguistics and Applied Language Study, Victoria University of Wellington for kindly allowing me to use this corpus.

could make a referent the C_p . Increasingly, it seems that linear order can be overridden by syntactic, or morphosyntactic factors. Walker et al. (1994) show that in Japanese *wa* marking of a topic, or morphosyntactic marking of speaker empathy, puts a referent at the head of the C_f list regardless of its position in the sentence. Di Eugenio (1990) shows that in Italian, referential inflection marking on verbs interacts with the C_f list, and Turan (1996) and Cote (1996) argue that thematic role assignment can override linear order in both Turkish and English in determining the order of an utterance's C_f list.

Centring may well be one of several strategies available in natural languages for the encoding and interpretation of anaphors. However, even in a single language it may not be the only strategy used. Prince (1996) shows Yiddish seems to be a mixed system. Variation between pronominal and zero anaphors is sensitive to the smoothness of the intersentential transition for most persons and numbers, but transitions involving 2SG are blind to Centring constraints.

6.2.2 Switch- and echo-reference marking

Cameron (1992) observed across two varieties of spoken Spanish that the relation between the grammatical subjects of consecutive clauses had a strong interaction with the use of a phonetically null subject. Where the subjects in two consecutive clauses had the same referent, a null subject was more likely; where the subjects in two consecutive clauses switched referents, a pronoun subject was more likely.

There are some sound reasons, based on the structure of the Melanesian substrate languages, for hypothesising that switch- versus same-subject reference might be a factor in Bislama. It would clearly be a candidate for serious investigation in Tok Pisin, since switch-versus echo-subjects are marked morphologically on the verb in some Papuan languages. The substrate models are less widespread in the case of Bislama, but it is possible that variation between phonetically null and overt subjects might have been grammaticised in Bislama under the influence of morphological patterns found in substrate languages.

Languages in southern Vanuatu, such as Sye and Ura of Erromango, distinguish switch-from echo-subject relations between clauses through the use of verb prefixes (Crowley 1998: 248ff.). However, we must ask how likely it is that these languages would have been models for the subject marking system in modern Bislama.

Crowley (1990:166ff.) notes that the languages of central and NE Vanuatu contributed the majority of the substrate vocabulary items found in Bislama today. He observes that this is a fair reflection of the number of people from the central and NE islands involved in internal and pan-Melanesian work-related migrations of the 19th and 20th centuries, so this is by no means unexpected. We might, likewise, expect that the most significant substrate influences on the grammar of Bislama would be the patterns of the central and NE languages, e.g. Nakanamanga or Paamese, and Raga or the language of Maewo. On the other hand, we cannot rule out the possibility that a substrate distinction between switch- and echo-subject reference might have helped determine the directions in which Bislama speakers have grammaticised the variation between phonetically null and pronominal subjects.

6.2.3 *Affective information expressed through pronoun choice*

In addition to the interclausal relations that may interact with the variation between the use of pronouns and phonetically null subjects, as Walker et al. (1994) noted there are non-linguistic factors relating to speakers' attitudes and perceptions that may have an effect on where and how speakers choose to use a pronoun or a null subject.

Duranti (1984) has argued that variation between pronouns and phonetically null third person subjects conveys a great deal more than referential information. He argues that the use of overt 3SG pronouns in Italian encodes information about the speaker's attitude to the referent.

Certainly it is well-attested that variation between different pronouns may convey important information about the speaker's affect or attitudes towards the referent. Sankoff and Laberge (1980) showed this to be the case for *on* in Montréal French, and Mühlhäusler and Harré (1990) discuss numerous cases of pronoun variation functioning in this way, including, for instance, the different affective information that can be conveyed by the use of 1PL instead of 1SG in English (1990:174).

Several studies have shown that the use of pronouns to express affective or attitudinal information can have repercussions in the shape of the pronominal system as a whole over time. Zilliacus (1953) draws our attention to the non-referential uses of 1PL in Classical Greek. He observes that in the literature of the period, individuals use the plural to refer to themselves as a way of co-opting the strength of a larger group they are associated with, 'weak and defenceless women [and other relatively unempowered groups such as shepherds, messengers, and country folk] use the plural conspicuously often, undoubtedly in order to identify themselves with the whole female sex' (1953:24, my translation).² In social psychological terms, we might say that the use of 1PL is a strategy by which speakers position themselves in a stronger position relative to their interlocutors by invoking the ethnolinguistic vitality of a group identity they can claim membership to.

Brown and Gilman's (1968) influential study of the development of honorifics in European languages shows that speakers' use of (second person or third person) plural pronouns to convey non-referential, interpersonal information motivated wider changes to the pronoun inventories of the Modern European languages and also to Latin. Haugen (1975) argues convincingly that the loss of the dual as a distinct number in Icelandic was caused by the increasing use of the plural as a respectful address form. This led to an across the board reanalysis of the first and second person pronouns, resulting in the current system where the historical 1PL and 2PL are now reserved for use as honorifics, the historical 1DU and 2DU are used for all non-singular forms, and the dual/plural distinction has been lost entirely.

While these sorts of developments may take place slowly over a long period, they do not only creep in. In Bislama, pronouns have quickly been recruited to serve non-referential purposes. For instance, speakers sometimes use the 1PL(INCL.) *yumi* in contexts where literal inclusion of the addressee is impossible. I have argued (Meyerhoff 1997) that this use of *yumi* is constrained by intergroup factors, and that it is more felicitous when the linguistic or social context serves to highlight the interlocutors' shared group membership.

² "So lässt sich beobachten, dass schwache und wehrlose Frauen auffallend oft den Plural heranziehen, um sich gewissermassen mit dem ganzen weiblichen Geschlechte zu identifizieren" (Zilliacus 1953:24). He notes the same use of the plural in the speech of "Hirten, Boten und Landmänner".

Thus, given speakers' wide-spread tendency to use pronominal variation to express more than referential information, it would be unsurprising to find that variation between phonetically null and overt subjects are put to the same use, as Duranti suggests is the case for Italian.

6.3 Establishing that Bislama is a pro-drop language

It seems clear that Bislama should be classified as a *pro*-drop language. Non-referential subjects, that is, expletive subjects or subjects with weather verbs, are phonetically null in Bislama as shown in the short piece of discourse in 6.5 where a weather verb is followed by a clause with an expletive subject.

- 6.5 \emptyset *i kol*
 \emptyset AGR cold
- \emptyset *i gat tri nyusilan ami, oli sik, oli gobak.*
 \emptyset AGR have three New.Zealand army AGR sick AGR return
 'It was cold, [and] there were three New Zealand army [guys] who got sick
 and had to go back [to camp].' (M-94-6, Visi)

In 6.6, we see that the Bislama equivalent of raising verbs like 'seem' also occur with a phonetically null subject.

- 6.6 \emptyset *i luk olsem* \emptyset *i stap jam nomo.*
 \emptyset AGR look COMP \emptyset AGR HABIT³ jump only
 'It's like he just jumps.' (M-95-11, Bretian)
 (lit: 'it looks like...', 'it seems that...')

Moreover, we also find null subjects in embedded clauses in Bislama. Even in English, co-ordination of clauses with the same subject may create situations in which a finite verb surfaces without an overt subject, as shown in example 6.7.

- 6.7 *Kirsty is fond of black and gold and \emptyset wears those colours all the time.*

But in Bislama we find examples of co-ordinated null subjects (6.8a), as well as examples of switch-subject finite verb sequences where there is no overt subject in the second clause (6.8b). Note that clausal co-ordination in Bislama is often null. We even find phonetically null subjects in subordinate clauses (6.8c), something that is only possible when phonetically null subjects are generated by syntactic processes.

- 6.8a. *Denis hem i kam, \emptyset i blokem hem.*
 Denis 3SG AGR come \emptyset AGR block 3SG
 'Denis came and stopped her.' (S-95-7, Sevi)

³ Bretian is talking about the way someone he knows tends to play basketball, hence the translation of *stap* as habitual rather than continuous or ongoing activity.

- b. *Taem we mifala i givim siks handred tausen*
 time COMP 1PL AGR give six hundred thousand

Ø oli karem Ø i go.

Ø AGR carry Ø AGR go

'When we gave [them] 600,000 [vatu], [they] took [it] away.'

(S-94-4, Stivin)

- c. *Hem i stap sore long mi*
 3SG AGR CONT sorry PREP 1SG

from we Ø i harem we mi toktok we difren.

because COMP Ø AGR hear COMP 1SG talk COMP different

'She was feeling sorry for me, because she heard that my voice was all funny.'

(M-95-16, Nina)

Assuming that children crucially depend on the presence or absence of expletive subjects in sentences like 6.5 and 6.6 to learn whether or not their language is *pro*-drop (Hyams 1987), and that null subjects in embedded or switch reference clauses are not generated in the same way as co-ordinated null subjects as in 6.7, we can say that Bislama is typologically a *pro*-drop language.

6.3.1 Fitting Bislama into the typology of null subject systems

Chapter 5 showed that Bislama verbs in finite clauses are inflected to a limited extent in agreement with the person and number of the subject. It was noted that for all persons, the inflection on the verb distinguishes singular and non-singular subjects. The consequence of this rather limited morphology is that the theory would predict phonetically null subjects to be licensed in Bislama. However, since the interpretation of them will not follow straightforwardly in all cases from the syntax, we might well expect there to be interactions with salient constraints operating at levels above the morphosyntax, such as a metric for keeping track of referents at the level of discourse.

Although generative discussions of canonical *pro*-drop languages like Spanish or Italian might suggest that phonetically null subjects are overwhelmingly the norm, there is, in fact, a certain amount of variation. Even in Spanish, rates of *pro*-drop vary across different regional varieties. Cameron (1992) found an overall rate of 55% null subjects in spoken Puerto Rican Spanish, while Silva-Corvalán (1994) reports 70% null subjects in spoken East Los Angeles Spanish, and 62% null subjects in spoken Santiago Spanish. Conversely, even in English, a language that does not have syntactic null subjects, discourse variables allow 1% of main clauses to occur with a phonetically null subject (see §6.7). The overall frequency of phonetically null subjects in Bislama finite clauses (44%) falls between the rates for spoken Spanish and English (the process by which this figure was derived is discussed in detail in §6.7 forward).

However, as I have pointed out, Bislama presents rather an interesting situation for the typology of null subject languages. It cannot be assimilated directly into the morphological uniformity paradigm, since it is neither a null subject language with rich subject-verb agreement, nor a null subject language without subject-verb agreement. A reasonable hypothesis, then, is that constraints on phonetically null subjects in Bislama may be

morphosyntactic, but need not necessarily be restricted to the morphosyntax. In order to place Bislama within the typology of *pro*-drop languages, the influence of a number of possible factors will need to be tested.

In the following sections I will review the linguistic and non-linguistic factors that I have used as a basis for evaluating the variation between phonetically null subjects and pronouns in Bislama. In selecting the linguistic variables, I have been motivated by open questions in linguistic theory about the nature of *pro*-drop and claims that are on record already about the grammar of Bislama. In selecting and coding for social factors, I have been motivated by my subjective evaluation of social and intergroup categorisations that are important to the people I worked with, and by objective factors such as the usefulness of a category for our understanding of language change. I provide a lengthy discussion of the use of social categories based on speakers' chronological age and their gender.

6.4 Syntactic and semantic factors investigated

The correlation between several linguistic factors, relating to the tense, aspect and modality of the clause, will be tested in the data.

The possibility of subjects being phonetically null is largely undescribed in the literature on Bislama and related languages. Crowley (1990:241) notes that an overt subject may be absent with third person referents, especially 3PL (Tryon 1987:22 also notes subjects are optional with 3PL). Crowley suggests that this is constrained by an animacy hierarchy that has been grammaticised on the pattern of some substrate languages. He notes his impression that null subjects are more common when the referent is non-human (or inanimate) than when it is human or anthropomorphised (Crowley 1990:241). I will test this hypothesis.

6.5 Discourse factors investigated

In addition to these syntactic and semantic variables, the correlation between use of a phonetically null subject and several pragmatic variables will be tested. In Meyerhoff (1998), I argue that in Bislama the argument of a preposition (specifically *olsem* 'like') can be phonetically null if the referent has already been introduced in the discourse. It seems plausible that all constraints on null arguments might be linked in the speaker's grammar, at least by analogy, if not in actual kind. Consequently, because constraints on non-subject argument deletion in Bislama are sensitive to the information status of a referent, the distribution of phonetically null and pronominal subjects will be tested against the discourse status of the referent. This will essentially be cast as a test of switch-reference, the discourse factor identified in Cameron's and Silva-Corvalán's studies of spoken Spanish.

6.6 Social factors investigated

6.6.1 Origin of speaker

As we saw in Chapter 5, Tryon suggested that the analysis of doubled 1SG and 2SG pronouns (therefore also singleton pronouns) may be subject to regional variation (1987:23). I will test whether speakers from any particular region show different patterns of pronoun

and null subject variation than speakers from other regions. The influence of substrate language grammars on the structure of creoles has been noted since the very earliest descriptions of these languages (Schroeder 1990:134 notes that Schuchardt 1883 comments on substrate transfer into the precursor of Bislama).⁴ We would assume that if we find systematic differences in speakers' Bislama that correlate with the place of origin, they may be the reflex of patterns found in speakers' L1, or the local *lanwis*.

However, there are two constraints on the extent to which it will be possible to draw these inferences. First, some of the speakers in my corpus come from L1 regions for which there is no grammar of their *lanwis*. This makes it impossible for me to compare variation between pronominal and null subjects in their substrate with the patterns of variation found in Bislama. Where a grammar of *lanwis* exists, the situation is not much better. It is difficult to find studies where discourse factors are considered in anything like the detail needed to compare them to the Bislama data I will be discussing here.

Second, given the design of my sample, speakers from Santo and Malo predominate. The majority of the *lanwis* spoken in these areas fall into the Santo subgrouping that Tryon (1968) calls 'Oceanic Melanesian' (only one of the Santo speakers came from the distinct East Espiritu Santo subgroup around Hog Harbour⁵). I have formed subgroups of the other speakers on the basis of typological similarity between the vernaculars. In doing so, I have attempted to create subgroups that will be large enough to enable a comparison of possible vernacular/substrate influences. For convenience, I have assigned the subgroup names corresponding to the regional government names (since they do approximate these regions), though this was a post-hoc decision and was not the initial rationale for constituting the subgroups.

Thus, in the analysis 'Sanma' refers to speakers from Espiritu Santo and Malo, 'Banks' to speakers from the Banks Islands, 'Penama' to speakers from Ambae, Maewo and the Raga-speaking areas of Pentecost, 'Malampa' refers to speakers from Malakula, 'Shefa' to speakers from the Shepherds group and from Efate, and 'Tafea' refers to speakers whose vernacular is the language of Aneityum (with some residence on Tanna).

6.6.2 Education – language medium and level attained

If the use of phonetically null subjects is due to vernacular or substrate influences, we might expect to find that they are disfavoured among speakers with high levels of fluency in English and French. As non-*pro*-drop languages, the patterns learnt here might favour use of pronominal subjects. I will, therefore, test level of education.

⁴ It may be worth noting, with respect to the ongoing debate over the role of substrate versus universals in the shaping of creole structure, that Schroeder also notes (1990:142) equally early arguments that the structure of creoles provides evidence of innate universal (syntactic and communicative) tendencies (von Coelho 1880-1886).

⁵ This speaker, Naomi, was probably a speaker of Lore diakarkar. When I asked her what her L1 was she told me it was the language of "Hog Haba" and I regret not having asked further. However, one of her classificatory nephews in Santo did tell me his *lanwis* was Lore diakarkar.

6.6.3 Age of speaker: comments on the concept of 'apparent time'

It is possible that phonetically null subjects have not been commented on much because they are a relatively novel aspect of Bislama grammar (Charpentier 1979:353 suggests null subjects are a neologism). Following the method commonly used in social dialectology (since Labov 1966), change across time in the Bislama speech community will be inferred or discounted by examining the synchronic distribution of phonetically null subjects across a number of age groups in the community. This method builds on the principle, long central to dialectology, that a speaker's linguistic system changes relatively little as they age. Bailey et al. (1991) review the empirical justifications for using the apparent time model.

It is not entirely clear the extent to which the apparent time methodology (sampling across age groups at a single time) can be taken as an indication of the actual rate of a change in progress. It is assumed that speakers carry their vernacular phonology through their lives essentially unchanged (Chambers 1995:194), and thus monotonic change over age groups is taken as an indicator that some kind of change is taking place. But the extent to which we can use the same data to predict the rate at which a change will continue to progress through the community is empirically more muddy. The best way to examine the methodology's predictive powers and limitations is to compare the trends in apparent time data with the trends observed over real time. Joy Fowler's retest of Labov's Department Store Study (sampling use of post-vocalic (r) in essentially the same populations in New York City) provides one such opportunity. Her results suggest that within a single speech community the rate of change may progress at different rates in different social classes (Fowler 1986, discussed in Labov 1994:89-94).

A comparison of Fowler's 1986 results and Labov's from 1962 shows that the percentages of the Saks speakers that were fully (r)-ful actually dropped, while the percentages of fully (r)-ful speakers in Macy's increased between 1962 and 1986. Given the real time lag between the two studies (and barring any drastic changes in the clients and employment pool of the two stores), the 15-30 year olds in the 1962 study should have been essentially the same people as Fowler's 35-50 year olds. But although approximately the same percentage of the 35-50 year old speakers at Saks in 1986 used some (r) as 15-30 year olds did in 1962, the percentage of categorical (r) usage drops dramatically from approximately 68% to approximately 38%. Similarly, a smaller percentage of the 55-70 year old speakers in 1986 than the 35-50 year olds in 1962 are partly (r)-ful (however in this group, the percentage of categorically (r)-ful speakers remains the same). On the other hand, in Macy's, there were increases in the percentage of 55-70 year old and 35-50 year old speakers who were both partly and fully (r)-ful in 1986, compared to their 1962 counterparts.

We have to wonder why there seem to be two distinct patterns of language change being mapped out in the same speech community. There are at least two possibilities. The differences between the Saks speakers and the Macy's speakers over real time may indicate that two distinct linguistic processes of change are at work. Or the different patterns of change may be socially, not linguistically, motivated, i.e. the patterns might be a further reflection of the groups' social distinctiveness. If the latter is the case, this has interesting implications for how freely we can generalise from apparent time data. The apparent time methodology is reliable only in so far as speakers carry their phonology with them, essentially unchanged, throughout their life span.

The data for the lower middle class speakers in Macy's provide support for this assumption. To be sure, the behaviour of these speakers is not static over time. Fowler shows that the percentages of (r)-ful and partly (r)-ful speakers among this group increase slightly, but this is consistent with Labov's initial hypothesis that LMC hypercorrection was a precondition to a sudden increase in (r) usage across the NYC speech community. Fowler's data indicates that hypercorrection did not play as strong a role as Labov initially suggested (across classes and with respect to the degree of the change in the LMC). But the hypercorrection of LMC speakers in 1962 does seem to have been an indicator of increases in *their* rates of (r) usage over real time.

Labov (1994:92) clearly shows that across the speech community the spread of (r) has continued as predicted. This is true even for the upper middle class speakers at Saks when they are considered as a group. Thus, Fowler's replication of the Department Store Study supports the notion that linguistic change can be observed through apparent time measures, as Labov originally proposed. But I am interested in exploring further the implications of the decreases in (r) usage among age-adjusted speaker samples within the UMC. As noted, the hypothesised function of LMC hypercorrection can account for increases in age-adjusted peers among the LMC speakers at Macy's, but a decrease in the use of (r) in UMC speakers as they age is intriguing.

One possibility is that as soon as the LMC speakers increased their use of the prestige variant, the UMC changed their direction of change. This strategy of 'shifting the goal posts' is an important one for theories of language change – less powerful or prestigious social groups may consciously or unconsciously target a goal of a prestige variant in the speech of a more statusful or powerful social group, but as they approach a critical (though undetermined) level of the statusful variant, the prestige group begins to use some other variant (Kroch 1996). This new variant becomes the new prestige by association with these speakers. However, this cannot be what is responsible for the patterns observed among the UMC in NYC. As noted, the overall level of (r)-fulness continues to increase, even among the UMC. From this, we infer that the prestige variant has not changed for the UMC.

It is also possible that the differences in the UMC speakers can be dismissed as differences in the sampling and linguistic analysis of Labov and Fowler. However, Labov (1994:87) specifically rejects this possibility.

A further possibility is that the 1986 UMC speakers have lower percentages of partial or categorical (r) usage than their 1962 age-adjusted counterparts because the speakers have reduced their levels of (r) as they aged. It is unclear what intergroup factors might motivate this, perhaps some notion of age-appropriate behaviour within the UMC, perhaps some nostalgic desire to use variants formerly associated with being a native New Yorker. In the absence of more detailed information about the internal organisation and belief systems of this subset of speakers, such explanations must remain conjectural.

And for the purposes of this discussion, they are neither here nor there. No matter how obscure the interpersonal or intergroup motivations for speakers' continued monitoring and change of their linguistic systems might be, the point is that they do. Sankoff and Brown's (1976) analysis of the development of *ia* as a strategy for delimiting relative clauses in Tok Pisin found that the first speakers to grammaticise *ia* in this way must have been adult speakers in the Tok Pisin community, from whom it spread to other adults and children. They suggested that linguistic evidence like this may offer useful insights into how powerful the pressures towards intergroup cohesiveness may be for speakers of any age

Wald (1986) also provides data which he argues shows that speakers continue to adjust their syntax through adolescence. He concedes that there is some difficulty in distinguishing between genuine reanalyses or extensions of the speaker's syntax and what are merely expansions of the speakers' stylistic repertoire. For instance, some of the examples he gives of the acquisition of modals in adolescence simply look like the acquisition of new lexical items. It is particularly hard to evaluate whether his data on *mighta* followed by a verb in the simple past constitutes a substantial reanalysis of the speaker's grammar, since we do not know what the distribution of past particles and simple past forms are in general for his speakers. However, his data on the loss of 'but...though' constructions (1986:164ff.) is a comprehensive reanalysis of 'though' between adolescence and adulthood.

Labov (1994:84, 112) suggests that there may be a criterial distinction between phonological and syntactic changes, i.e. that syntactic variation can effect restructuring of speakers' grammars into adulthood, but that speakers' vernacular phonology is not subject to such restructuring. Note that this in turn requires that we distinguish vernacular changes in progress from stable age-graded changes. This is because the latter do of course assume that during their lives speakers restructure their phonology in response to social or affective factors such as their (desired) integration into work domains. Fowler's data suggests that the distinction between the two levels of grammar may not be as clean cut as this, and that some restructuring of speakers' vernacular phonology may occur in maturity.

What we conclude, then, is that apparent time observations across a large community are reliable indicators of on-going change, and probably of the direction of a change (Labov 1994 summarises other real time surveys which support these claims, e.g. Cedergren's resampling of speakers in Panama, Trudgill's resampling in Norwich). We should be careful, however, about extrapolating too much about what the grammar or even the phonology of the speech community was at some earlier point in time, and about the unidirectionality of the change. Fowler's data shows that sub-groups within the community may alter their direction or rate of change, just as Strassel and Boberg (1996) showed that a change may reverse across a region.

Bearing these caveats in mind, the distribution of null subjects will be examined in the discourse of speakers in several age groups for indications of on-going change.

6.6.4 *Sex of speaker: comments on the use of sex qua sex as a variable*

Chapter 2 presented ethnographic data showing that sex is a highly salient social category in Vanuatu society. The social spaces women and men negotiate social power within are clearly demarcated, often mutually exclusive and some of them are themselves determined by roles that are dependent on an individual's sex, e.g. childbearing and child rearing. In other words, sex roles are considerably less open to negotiation than they are in, for instance, white middle class New Zealand communities.

Yumi Wok (1996) gives a picture of how strong some contemporary ideas about the roles appropriate for women are in much of Vanuatu society. As the authors of *Yumi Wok* note, prejudices and expectations based purely on an individual's sex have a serious and real impact on the opportunities open to Ni-Vanuatu women, and they have a marked effect on the nature of the lives they lead. On the whole, for instance, Ni-Vanuatu women have less access to education and health care, and are threatened more by domestic violence than men (see also Jolly 1996). Since other forms of social behaviour and social practice are rigidly

stratified according to sex, it seemed reasonable to hypothesise that the Santo and/or Malo speech communities might show sex stratification of linguistic variants.

Thus, despite some recent and quite trenchant debate in the field of sociolinguistics about the political and methodological ramifications of essentialising speakers on the basis of an external category such as sex, it has been included as a variable in this study.

In addition to the local reasons for using speaker sex or gender as a basis for analysing variation, it is worth remembering that the history of variationist linguistics has shown that change can sometimes be highlighted in very useful ways by differences in men's and women's use of linguistic variants. Stratification of linguistic variables by speaker gender has been demonstrated now in a range of cultures, from the urban East Coast of the United States (Labov 1972 [1966]), the Gullah speaking islands of North Carolina (Nichols 1983), to Arabic communities in Tunisia (Trabelsi 1991) and Cairo (Haeri 1996). It has been demonstrated to be a salient dimension for stable variables, e.g. variation between alveolar and velar nasals – the IN/ING variation (Fischer 1958; Wald & Shopen 1981), use of non-standard past tense forms (Cheshire 1982; Eisikovits 1987). It has also usefully highlighted the trends of changes in progress, e.g. those affecting the whole short vowel system in the Northern Cities region (Eckert 1989), various vowel shifts in Philadelphia (Labov 1990), and the variation between alveolar and glottal stops in New Zealand (Holmes 1995).

A number of researchers (e.g. Lazreg 1988; Troemel-Ploetz 1991; Freed 1992; Uchida 1992) have argued that as long as there are social and economic inequalities between women and men, any reification of the boundaries between the sexes in the study of language is a political act. Consequently, they argue, using sex alone as a measure of variation should be employed with caution. It can obfuscate interactions with other social factors and perpetuate stereotypes about both men and women. Any conclusions or explanations about linguistic differences observed between women and men need to be expressed particularly thoughtfully for these reasons.

I agree that essentialising individuals as archetypes of a group may encourage the ongoing perception and treatment of individual women in stereotypic frames. In addition, I am aware that increasingly longitudinal or in depth studies of speakers have shown that generalising group means to individuals' performance is unwarranted. It may miss important details about how individuals organise their social and linguistic resources. For example, work by Janet Holmes and her colleagues (e.g. Holmes et al. 1991; Holmes & Bell 1992; Britain 1992; Meyerhoff 1994) on variation in New Zealand English has shown over and over that although women on average appear to be leaders of change, the use of innovative forms is not actually sex-specific. The leaders of vernacular changes are a subset of women who are associated with a number of social factors: sex, ethnicity and particular domestic and social networks.

Bearing all this in mind, I have employed speaker sex primarily as a diagnostic of language change, mindful of the fact that if a statistical correlation with gender is found, this should only be the first step in a full examination of the locus of variation in the speech community. Insofar as I have used gender as a non-linguistic variable in this study, I accept that I am doing nothing to challenge some pervasive assumptions in Ni-Vanuatu society, such as the idea that biological differences can and should map onto differences in social status, rights and opportunity. But neither does it mean that I personally buy into qualitative evaluations about the role of women and men in Vanuatu society. I think this is made clear in the accounts of some gender-preferential variables that I have presented elsewhere (Meyerhoff 1997, 1999). Frankly, I am not convinced that it is for me to encourage the

deconstruction of the Ni-Vanuatu view of gender as a social category. Women in the communities I stayed in might well find this an appropriate goal, but I do not believe I have the authority to engage in the work here.

Cameron et al. (1992) and Rickford (1997) have recently produced eloquent challenges for the linguistic community, arguing that our work should be located in methods that are both consultative and empowering. But empowerment works both ways. I have tried to make clear that I feel I learnt a great deal in conducting my fieldwork. I learnt a considerable amount about Bislama and I learnt a lot about its social milieu. Even though members of the communities I stayed in gave freely of their knowledge, attitudes and opinions (about language and many other things), I do not believe I was empowered by them to pursue a socio-political agenda.

Thus, I have focused my efforts in this work on exploring and questioning aspects of the Bislama speech community that I felt I was empowered to deal with. I used them as the basis for delimiting intergroup boundaries that I hypothesised might be theoretically useful. I am confident that the group identities I have chosen are well-motivated within the community. To the extent that they help me understand the nature of and direction of language variation in Bislama, I am confident that they are well-motivated linguistically.

6.6.5 Village versus urban residence

As was also noted in Chapters 2 and 3, the sociolinguistic situation of the speakers sampled in the village (Malo) and the urban area (Santo) differed considerably in terms of the norms for Bislama use and attitudes towards the language. Thus, the data will be analysed for differences in the distribution of variants in the speech of urban and village residents.

In the light of my previous discussion about the social significance of divisions based on speaker gender, it is worth noting that the urban-rural dimension is also not a value-free continuum (indeed, what intergroup boundaries are?). A number of the Ni-Vanuatu I knew and worked with in Santo and Malo operated with a clear scale: urban-village-bush (with some decrease in social prestige at each point). It would probably not be fair to say that people living in the village on Malo saw themselves as inferior to the people in Santo (as I noted in §3.7.1, some people actively chose a village life over work in town), but they certainly saw themselves as leading more modern lives than *man bus* 'people from the bush'. I hope I need not stress that I hardly endorse or hope to encourage prejudices against *man bus* or their *kastom*.

6.6.6 Social and attitudinal factors grouping Malo speakers

Chapter 3 presented information on the language attitudes and reported bilingualism among the Malo residents. This information provided the basis for dividing the Malo residents into four sub-groups (plus the children), and these groups will also be tested for any significant correlations with individuals' linguistic behaviour. Ethnographic information about the social categorisations important in the Malo community will also be used to augment tests of possible change in progress. Children's use of the variable will be contrasted with both their immediate caregivers' and their classificatory mothers' and fathers' use of the variable.

In sum, a wide range of linguistic and non-linguistic factors will be tested against speakers' use of pronouns or phonetically null subjects in Bislama. In each case, the decision to include the variable was either motivated by

- (i) claims made in existing descriptions of Bislama,
- (ii) the demonstrable cross-linguistic salience of social categories as the basis for the stratification of language, or
- (iii) descriptions of the factors influencing the optionality or required presence of pronoun subjects in other *pro*-drop languages.

6.7 Defining the linguistic envelope of variation

In this section, I gradually eliminate all environments in which speakers use phonetically null subjects categorically or at near categorical levels, until we are left with all and only those environments that show genuine variability between the two forms.

6.7.1 Categorically null subjects

Since we are interested in investigating the variation between pronominal subjects and phonetically null subjects, it is important to exclude from the database subject positions that are categorically one or the other. In Bislama, a pronoun subject always co-varies with phonetically null alternates, but there are some structures where an overt pronominal subject is never found.

Environments that never have pronominal subjects are the second verb in a serial verb construction (SVC) (6.9), existential sentences (6.10), i.e. in generative terms, expletive subjects, and subjects of weather verbs (6.11) (which were coded with expletive subjects).

6.9 \emptyset *i ron* \emptyset *i go long haos.*
 \emptyset AGR run \emptyset AGR go PREP house
 'She ran home.' (M-94-4, Janette)

6.10 *Yes, \emptyset i gat sam tu.*
 yes \emptyset AGR have some too
 'Yes, there are some [here] too.' (M-95-15, Livai)

6.11 \emptyset *oli fraet from* \emptyset *i kolkol tumas.*
 \emptyset AGR fright because \emptyset AGR cold very
 'They were scared because it was too cold.' (M-94-6, Visi)

It is worth distinguishing expletive subjects that correspond to English 'there' from ones that correspond to English 'it' or 'that' (which are potentially pro-clausal, i.e. refer to clauses rather than individual referents). The latter subjects are generally phonetically null in Bislama (6.12a), but may occur with the 3SG pronoun *hem* (6.12b).

6.12a. *Sapos olsem, \emptyset i gud.*
 if like \emptyset AGR good
 'if [they] do, that would be good.' (M-95-13, Rovi)

- b. *Hem i gud... yu go longwe ples olsem blong stadi.*
 3SG AGR good... 2SG go there place like COMP study
 'it's good (that they are strict)...you go there, like, to study.' (S-95-9, Adelin)

Phonetically null subjects also occur when bare infinitives are used as deverbal nominals, cases where the subject is PRO_{arb}. These tokens are, however, rare in speech.

Verbs following *wantem* 'want' and *traem* 'try' generally occurred with phonetically null subjects, though because this was not categorically true (i.e. 4% pronominal subjects), they were retained at this point in the investigation. Fourteen percent of the clauses occurring below the complementiser *blong* '(in order) to' occurred with pronominal subjects, and of the clauses apparently functioning as imperatives, 14% also had pronominal subjects.

Verbs that occur as the complement of *wantem*, *traem* and *blong* are syntactically distinct from *kam*, *go*, *stap*, etc. when the latter occur as the second verb in a SVC. The second verb in an SVC is obligatorily marked with the predicate marker *i*, while the complement of *wantem*, *traem* and *blong* occurs either with no inflection whatsoever, or occurs with the full range of subject-verb agreement, *i/oli/∅*. Although the fully inflected option is proportionally disfavoured, it is by no means impossible.

6.7.2 Optional subject-verb agreement in Bislama

It has been well documented that in urban Tok Pisin (Sankoff 1993), the morphological equivalents of Bislama *i/oli/∅* subject-verb agreement are frequently omitted (Smith 1995 notes, however, that Highlands Tok Pisin diverges from urban varieties. Highlands speakers continue to use *i* frequently). Christine Jourdan (pers.comm.) also notes that subject agreement on verbs in Solomon Islands Pijin appears to be being lost.

One of the noticeable features of Bislama among the Melanesian varieties of contact English is the high frequency of subject-verb agreement morphology.⁶ Agreement is sometimes omitted in my spoken corpus, but native speaker consultants generally saw this as 'errors' (i.e. not *brava Bislama* 'real/proper Bislama').

- 6.13 *Lili, i no tektaem i no wet blong ∅, i singaotem samwan.*
 Lili AGR NEG take.time AGR NEG wait COMP ∅ call someone
 'Lili didn't stop, she didn't wait to call out for someone.' (S-95-10, Rebeka)

When discussing 6.13 above, Sharon Tabi noted the following as acceptable variants with the same meaning:

- 6.13' *Lili, i no tektaem i no wet blong ∅, i singaotem samwan.*
 6.13" *Lili, i no tektaem i no wet blong hem, i singaotem samwan.*

In addition, she reports no difference in meaning between the attested example 6.14, and example 6.14', which is the more frequent pattern following *go*:

⁶ This does not seem to be a stereotype commented on by speakers of Bislama and Tok Pisin though. A more salient marker of Bislama for native speakers of these languages seems to be the complementiser *we*, which is not found in Tok Pisin and Pijin. It is more frequent in conversational Bislama, as a means of expressing affect or intensification. An example is shown in 6.8c.

6.14 *Lili i go i lukim rop ya.*
 Lili AGR go AGR look rope SPEC
 'Lili went and found the vine.' (Rebeka, S-95-10)

6.14' *Lili i go lukim rop ya.*

I suspect that the variation between 6.14 and 6.14' is a consequence of the fact that *go* and *kam* are well on the way to being grammaticised as auxiliary verbs (Charpentier 1979:352-353 notes this for *go*). *Stap* 'stay' and *save* 'know; be able' have already been grammaticised in this way (without losing their lexical meanings) and are used as auxiliaries marking continuous or habitual aspect, and ability or habitual aspect, respectively.

Whether *go* and *kam* will fully grammaticise in Bislama must remain an empirical question. We need not necessarily expect them to progress any further, since 'go' and 'come' seem to have halted on the grammaticisation path as (semi-)auxiliaries in English. There are some substrate models for preverbal inceptive mood marking in the verb phrase, though, and these might encourage the continued process of grammaticisation. Lonwolwol, for example, marks the 'immediate [future]' (Paton 1971:52), Paamese marks an 'immediate anticipatory' (Crowley 1982:136), and Big Nambas a 'proximity aspect' (Fox 1979:64), by which Fox means temporal proximity.

The constraints governing the deletion of subject agreement morphology remain to be determined, and will not be attempted in this work. I believe that an examination of the variation between presence and absence of subject agreement in finite clauses should be undertaken in conjunction with an examination of the variation between the full subject agreement paradigm found with finite verbs and the restricted agreement, *i*, found with the second verb in SVCs. I will propose an analysis of the restricted agreement found in SVCs in Chapter 7.

6.7.3 Summary of tokens excluded from the corpus

Three environments of categorical, or near categorical, absence of pronouns were, therefore, excluded from the investigation. Table 6.1 gives the environments excluded from the analysis of variation with the numbers of tokens that occurred as phonetically null subjects and as pronouns for each of these three conditions. Clearly the second verb in a serial verb construction does not vary between pronominal and null subjects. It, therefore, must be excluded. A verb immediately following *go* or *kam*, likewise, approaches near categorical use of null subject, and also merits exclusion from the database.

Table 6.1: Summary of linguistic environments excluded from analysis of subject variation. Categorical or near categorical null subject environments (see text for discussion of 'expletive subjects').

Environment	Null subject	Pronoun subject	Total
V2 in SVC	173 100%	0	173
non-referential or expletive subject	389 88%	53 12%	442
<i>go</i> or <i>kam</i> + V	167 99%	2 1%	169

There is somewhat greater variation with the non-referential or expletive subjects, but these, too, have been excluded from the analysis. This may seem curious at first, particularly as the percentage of subjects realised as a pronoun in this condition is very similar to the percentage of subjects realised as pronouns with verbs following the complementiser *blong* 'to; in order to', and with verbs having imperative illocutionary force. Following *blong* and in clauses with imperative force, subjects are also realised as pronouns 12–14% of the time.

But there is a qualitative difference between the subjects with *blong* and imperatives, compared to the expletive or non-referential subjects. With *blong* and imperatives, the subject usually refers to a specific person or thing. In cases where the subject following *blong* has non-specific reference (i.e. ranges over all possible referents of a set, PRO_{arb}), it nonetheless picks out some agent as the referent. As 6.12b showed, non-referential subjects may be realised as the 3SG pronoun *hem* when *hem* refers to a proposition, a sequence of clauses or a state of affairs, rather than an individual or thing. All the pronoun expletive or non-referential subjects indicated in Table 6.1 refer to single propositions or larger discourse units, e.g. in summing up a story *hem* acts like the English demonstrative 'that', as in: 'And that's how the elephant got its trunk'.⁷

These proclausal (as opposed to pronominal) anaphors are often lumped in with true expletives, e.g. Iatridou and Embick (1997), though they should probably be analysed separately more consistently. Cross-linguistically, true expletive and proclausal anaphors do not pattern alike (as Iatridou and Embick (1997) ultimately find). However, in coding my corpus I followed the convention of grouping them together. Thus, the possibility of using a 3SG pronoun for a clausal antecedent explains why the percentage of pronominal expletive or non-referential subjects appears to be quite high. Nevertheless, their qualitative difference from referential subjects justifies holding them aside from the analysis of variance we are interested in here. Being non-nominal, we do not necessarily expect them to be constrained by the same factors, such as referential persistence or referential salience, that nominal anaphors are.

Finally, clauses with two other kinds of subjects were also excluded from the analysis of variance. These were clauses with focused subjects (NP + pronoun + *i/oli* + V, see Chapter 5, N = 295), and clauses with full NP subjects (N = 496). These were excluded because they differ qualitatively from pronouns and null subjects. Pronouns and null subjects are always anaphoric, whereas full NPs may be new referents. Given the relatively small number of clauses with full NP subjects in the entire corpus, it is unlikely that adding them to the investigation would much have altered the analysis.

The envelope of variation is defined, therefore, as finite clauses with nominal subjects (excluding clausal NPs).

6.8 Outline of method: statistical tools used

As an initial step, a multivariate analysis of variation was conducted on the whole corpus. This gave an indication of which of the factors interacted significantly with the dependent linguistic variable, presence or absence of a pronoun subject. Individual social and linguistic

⁷ The converse is not always the case. Proclausal anaphors need not be realised with a 3SG pronoun.

factors were examined more closely using both the capabilities of Goldvarb 2.0 (Sankoff et al. 1992), and the statistical tests available with Datadesk 4.1 (Velleman et al. 1993).

The multivariate analysis of the entire corpus indicated that linguistic, but not social, factors had a significant correlation with speakers' rates of null versus pronominal subjects. Specifically, the step-up, step-down test that Goldvarb conducts indicated that gender and origin of the speaker did not exert an influence on the variation, but that age, level of education and urban versus village residence showed a weak, positive correlation with the dependent variable. The step-up, step-down procedure evaluates changes to the data's goodness of fit to the model. These changes are tested as each factor group is added (logistically, not arithmetically) one by one in all possible combinations with the other factor groups, and is subsequently removed in the same fashion. In this way, the program provides a measure of the strength of interaction between individual factor groups and the dependent variable, as well as providing an indication of whether there are interactions between the factor groups being treated as independent variables.

The preliminary multivariate analysis selected the syntactic factors as having the greatest influence on the variation, that is, whether the subject variable was associated with the main (finite) verb of the clause, or whether it was the subject of a complement of *wantem* or *traem* or the complementiser *blong*. The discourse variable was also identified as exerting a significant effect on the variation between null and overt subjects. That is, the form and grammatical role of the current clause's subject in the preceding clause (exemplified and discussed in detail in §5.2.2.).

But a step-up, step-down analysis operates on the corpus *in toto*, treating each of the clauses as individual data points. In the case of social factors we are concerned with the speakers who represent the different social dimensions on which the community is divided. For this reason, a multivariate analysis of the whole corpus is not the most informative diagnostic of whether the dependent variable is constrained by social factors. To test the influence of social factors, it is necessary to convert individuals' scores for pronominal and null subjects into means, and then to test the significance of the variance between these means on the distribution of the dependent linguistic variable. In this way, we can compare speakers, rather than comparing tokens. Clearly, with social factors, this is a more meaningful approach to take.

6.9 Results

The corpus consisted of 4,627 clauses. Of these, 2,055 occurred with a phonetically null subject (44%), and 2,572 occurred with a pronominal subject (56%). This places Bislama somewhere between the extremes of languages strongly inclined to omit pronoun subjects and languages which strongly favour full realisation of a pronoun subject. Earlier I noted that Silva-Corvalán reports 70% phonetically null subjects in spoken Los Angeles Spanish. Even in a supposedly non *pro*-drop language like English discourse factors allow speakers to felicitously omit subjects in finite clauses. Cote (1996) examines the discourse conditions under which null subjects occur in spoken English but does not quantify them as a percentage of all finite clauses in her data of switch board conversations. Her impression is that null

subjects occur in non-co-ordinated finite clauses less than 1% of the time in spoken English (Cote pers.comm.).⁸

In order to obtain some kind of empirical basis for comparison with spoken English, I went to the Wellington Corpus of Spoken New Zealand English. I quantified the number of phonetically null subjects in the speech of two Pakeha women, i.e. New Zealanders of European ancestry, Kylie and Lisa. I compared the number of null and the number of overt subjects used with finite verbs, excluding expletive subjects. Unlike Cote, I counted both unequivocal cases of null subjects (6.15) and also cases where a null subject occurred in a co-ordinated construction (6.16).

6.15 I think I was shocked so I mean didn't feel too much of it,
Ø lugged us in the ambulance. (RF 16, Kylie)

6.16 we flew over the top like Batman and Robin
and Ø landed on the road. (RF 16, Kylie)

In a (admittedly small) sample of 225 finite verbs, Kylie and Lisa had 24 null subjects, i.e. 11%. If we exclude all tokens of null subjects in co-ordinated constructions as in 6.16, the total drops to four, i.e. 2%.

On the basis of this data (and Cote's report), we can conclude that English speakers drop referential pronouns in something like one out of a hundred clauses. This gives us some idea of where Bislama fits in relation to archetypal *pro*-drop and non-*pro*-drop languages.

Tables 6.2 and 6.3 give the mean scores for number of null subjects as a percentage of all utterances for each of the speakers investigated. The range is between 10% and 100% phonetically null subjects. Mama, who used 100% null subjects, actually contributed very little to the corpus (her contributions were side comments in larger family conversations), and therefore, very little store should be set by her performance alone. Consequently, the influence reported for social factors where she is one of only a small number of representatives must also be treated with caution. But because Mama comes close to Nina demographically, and because Mama's utterances had been transcribed as part of a larger family dinner conversation, I decided to include her in the corpus. I felt the small number of tokens she produced would augment the contributions of Nina. Similarly, Seman's speech was transcribed as part of the same family conversation. In all respects (except level of education) she matches Rebeka, so I retained her speech in the corpus. The average of the individual speaker means is higher than 44% because of rounding off for individuals to create whole percentages.

In the following sections, I will discuss the effects of individual social and linguistic factors on speakers' variable use of phonetically null and pronominal subjects.

⁸ Cote is now engaged in an investigation of null subjects in written dialogue. Her impression is that their frequency is higher in writing than in natural speech. She hypothesises that this may be because dropping subjects is strongly associated with speech, and writers may exaggerate null subjects in their dialogues to create an enhanced semblance of naturalness. It remains to be tested whether writers observe the same constraints as speakers but simply with greater frequency (Cote pers.comm.).

Table 6.2: Percentage of phonetically null subjects in speech of women, with the amount of speech per individual. Means for Malo (village) and Santo (town) speakers.

Speaker (women)	null subject mean as %	Sample size N clauses
Malo		
Alis	46	129
Anita	42	99
Dien	40	130
Elise	67	48
Janette	34	198
Leikitah	59	78
Leipakoa	43	131
Lolan	32	378
Madelin	26	47
Mama	100	6
Nina	48	125
Rovi	53	78
Susana	58	177
Vosale	57	98
<i>Malo mean</i>	<i>50</i>	<i>123</i>
Santo		
Adelin	56	90
Elsina	32	165
Juliet	47	110
Lili	44	89
Lisa	49	79
Naomi	44	73
Rebeka	48	170
Rinette	43	77
Seman	89	9
<i>Santo mean</i>	<i>50</i>	<i>96</i>

Table 6.3: Percentage of phonetically null subjects in speech of men, with the amount of speech per individual. Means of Malo (village) and Santo (town) speakers.

Speaker (men)	null subject mean as %	Sample size N clauses
Malo		
Atesolo	53	75
Iawia	77	122
Jackson	53	75
Livai	29	122
Mesek	47	138
Obed	34	77
Papa	71	124
Petre	77	84
Saemon	77	76
Tarip	26	89
Visi	44	155
<i>Malo mean</i>	<i>53</i>	<i>103</i>
Santo		
Bretian	45	239
Ezra	45	66
Jonas	10	41
Sevi	37	87
Sikal	32	224
Simeon	18	127
Stiven	42	59
Timoti	26	73
Wingston	31	59
<i>Santo mean</i>	<i>32</i>	<i>108</i>

Overall mean of null subjects in the corpus: 44%

6.10 Linguistic variables

The results of the distribution of phonetically null subjects for the linguistic variables investigated showed stronger effects than social factors associated with the speakers. Goldvarb selected the following variables as having significant and independent effects on

the variation between phonetically null subjects and pronominal subjects: person and number of the referent, the information status of the referent in the preceding clause, and the markedness of the verb group (what has been distinguished as tensed, indicative verbs versus most non-indicative clauses in the preceding discussions) (χ^2 per cell = 1.9).⁹

6.10.1 Person of subject referent: an attempt to unify the analysis with other split systems

A step-up, step-down analysis of subject form using the multiple analysis of variance in the Goldvarb program indicated that the person and number of the subject referent had the strongest conditioning effect on the variation observed in the corpus of all speakers.

The results present an interesting extension to the typology of split *pro*-drop systems. It has been claimed that *pro*-drop systems may split along person and number, such that a phonetically null subject is grammatical with 1SG and 2SG but a pronominal subject is required for third person (it remains to be determined how categorically speakers of these languages actually observe this distinction, or under what circumstances the restrictions on first and second or third person can be relaxed).

However, we have repeatedly seen that in Bislama the system maintains the same split between first and second person and third, but that in Bislama the third person strongly favours phonetically null subjects, while first and second person favour pronominal subjects.

Table 6.4: Goldvarb weightings for all speakers for person and number of subject referent (factor selected by multivariate analysis as having the most significant effect on occurrence of phonetically null subjects). (0 = pronoun subject; 1 = phonetically null subject). Input probability for all speakers = .418.

Person and number of subject referent	Goldvarb weighting	N clauses
3SG	0.709	1,719
3PL	0.869	846
1SG	0.147	1,054
1PL	0.268	375
2SG	0.206	397
2PL	0.197	71

There are some surprises in these results. That the system splits along these lines, and that first and second person subjects pattern together, and that third person subjects pattern another way, is not in itself surprising. Other mixed *pro*-drop systems have been attested: in

⁹ At the end of a Goldvarb run, a high χ^2 per cell, e.g. >2, usually indicates some kind of large-scale interaction between the variables and suggests closer examination of the distribution of the dependent variable is needed. A relatively low χ^2 , e.g. < 1, suggests there is no significant interaction between variables. The value shown here, χ^2 per cell = 1.9 lies at about the threshold. The rest of the chapter attempts to tease apart the extent and locus of interaction between variables by undertaking analyses of several variables one by one, and by examining various social variables in relation to each other.

the generative literature, Hebrew and Finnish have been discussed. Both these languages have been described as having a split in exactly the same place. Rules for *pro*-drop with first and second person subjects are the same; and rules for *pro*-drop with third person are different from first and second person. But what is interesting is that Finnish and Hebrew manifest this split in completely the opposite way.

Table 6.5 shows that in Bislama, speakers favour null subjects with third person referents and they tend to use pronouns with first and second person subjects. In Finnish and Hebrew, it is attested that speakers use null subjects with first and second person subjects, and that they must use pronoun subjects with third person referents.

Table 6.5: Summary of preferences for pronominal versus phonetically null subjects in different persons for split *pro*-drop systems.

	Bislama	Finnish/Hebrew
1st	pronoun	null subject
2nd	pronoun	null subject
3rd	null subject	pronoun

Most of the literature on *pro*-drop has not dealt with mixed systems like these. Those researchers who have made previous attempts to analyse the Finnish and Hebrew systems have presented both functional and derivational arguments to support their analyses. They have tried to motivate the requirement that speakers use a pronoun with third person referents by comparing the relative amounts of information available in the third person verbal inflection to the amount of information present in the inflection found on the verb with first and second person subjects.

Vainikka and Levy (1995) noted that both Hebrew and Finnish have the option of null first and second person pronouns in precisely the tenses where the subject agreement inflection is fairly transparently derived from the independent first and second person pronouns, as shown in 6.17 and 6.18. They distinguish between affixes they call 'highly referential' – the first and second person affixes – and those which are not highly referential – the third person affix.

- 6.17 1PL-*mme* cf. independent 1PL pronoun *me*
 2PL-*tte* cf. independent 2PL pronoun *te*
 Finnish (from Vainikka and Levy 1995)

- 6.18 1PL-*nu* cf. independent 1PL pronoun *anaxnu*
 2PL-*tem* cf. independent 2PL pronoun *atem*
 Hebrew (past tense) (from Vainikka and Levy 1995)

Vainikka and Levy go on to argue that where the affix is 'highly referential' like this, it functions as a structural subject, and that when the affix is low in referentiality, it is generated as an agreement feature only. Similar findings showing a correlation between the licensing of phonetically null subjects and the presence of morphologically transparent or uniquely identifying subject-verb agreement are reported in other languages (see Poletto (1996, forthcoming) and Heap (1997) for northern varieties of Italian; even more local cases of 'splits' in a system have been reported, e.g. null subjects in 1PL present in Old High German, and for the first person only in the conditional in Gallic, Wayne Harbert pers.comm.).

How can this insight into the interpretability of inflections help us with the cross-linguistic facts demonstrated by Bislama? Remember that the Bislama agreement marker *i* originally derived from English 'he'. Its expansion to the other persons and numbers occurred relatively late (in the 20th century). The subject-agreement morphology associated with third person plural, *oli*, derives from a fusion of English 'all' and 'he'. *Ol* from 'all' continues to exist as a marker of plurality in modern Bislama. Moreover, *oli* is uniquely identified within the agreement-marking system – it only ever appears with a 3PL subject.

Thus, we could well argue that the 3SG and 3PL agreement on Bislama verbs is maximally referential. We are still not in a position to predict whether a language will opt for a split *pro*-drop system. But once we find a language in which there is a split, we can predict the direction in which the split will go. All that is required is that the subject agreement affixes on the verb continue to be morphologically transparent. The distribution of *i* throughout the subject-verb agreement paradigm gives us reason to think this is the case for *i* (as discussed in Chapter 5), and the strong resemblance *oli* bears to the plural marker *ol* suggests that it too satisfies the requirement for transparency. My findings on the way speakers use null subjects in conversational Bislama certainly seem to support the idea that these elements are not so completely grammaticised that they have been bleached of all meaning.

I am aware of one exception to this generalisation. Hirschbühler's (1995) study of *pro*-drop in Middle French (texts from between 1505 and 1515) shows that there was a split *pro*-drop system at this point in the development of French. Essentially, 3SG and 3PL favour phonetically null subjects, and 1SG and 2SG do not. The intriguing split, however, lies with respect to 1PL and 2PL. Although 2PL favours phonetically null subjects (2PL = 105 null subjects out of a total of 337 in embedded clauses, i.e. 31%), 1PL does not (1PL = 4 out of 88 in embedded clauses, i.e. 5%) (Hirschbühler 1995:265-266). Since 1PL and 2PL are both unique morphological identifiers of the referent, the theory that split *pro*-drop systems are based on morphological transparency should take account of this example.

The answer, I believe lies in the marginal status of *pro*-drop as a syntactic property in Middle French. Platzack (1995), in discussing the loss of V2 between Old French and Modern French, notes that Middle French was characterised by a number of syntactic changes. Around this period, subject pronouns ceased to be independent pronouns and became verbal clitics (opinion differs as to whether the cliticisation was phonological, e.g. Roberts 1992, or syntactic, e.g. Adams 1987). Platzack (1995) argues that this created a situation of sufficient structural ambiguity for the language learner that it contributed to the loss of V2 in French. He argues that if the clitics were syntactic, this would have led to a situation in which the vast majority of utterances in a V2 grammar were of the form XP cl-V. This structure is no longer clearly a V2 construction, leading to instability and the eventual loss of V2.

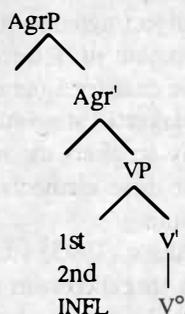
This should have led to an increase in *pro*-drop under Platzack's account. Indeed, he cites Vance (1989) as showing a general increase in the number of phonetically null subjects from Old French to Middle French. But paradoxically, at exactly this time the rich subject-verb agreement morphology of Old French was eroding in Middle French (Vance 1989, cited Platzack 1995). Obviously, Modern French is not a *pro*-drop language: only 1PL and 2PL agreement remain phonetically unique.

In other words, the asymmetry between 1PL and 2PL that runs counter to our expectations regarding *pro*-drop is probably a consequence of the fact that Middle French was undergoing two substantial changes pulling speakers in two directions with respect to *pro*-drop. The cliticisation of pronouns and the loss of V2 created an environment in which

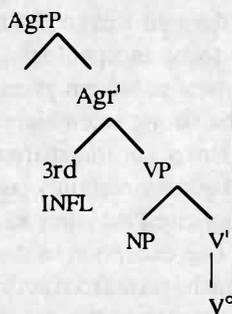
pro-drop was favoured, while simultaneously the loss of unique morphological inflection on the verb created an environment in which *pro*-drop was disfavoured.

Let us turn to the specifics of the structural analysis proposed by Vainikka and Levy (1995). They suggest that there is a fundamental difference between the way first and second subjects are base-generated in Hebrew and Finnish and the way third person subjects are base-generated. They propose that highly referential inflection is base generated in subject position as in 6.19a, while agreement that is low in referentiality is ordinary agreement, i.e. is generated in the functional head, Agr as in 6.19b.

6.19a. 'highly referential' inflection



6.19b. inflection low in referentiality



The argument for Hebrew and Finnish is essentially the following: because the first and second person inflections are highly referential, because they have a relatively transparent connection to the independent first and second person pronouns, it is therefore highly plausible that speakers might treat them as if they were subjects within the VP. On what is asserted to be the relatively infrequent instances of first and second person co-occurring with a pronoun¹⁰ in these languages, the pronoun is (presumably) considered adjoined in a focus position in the clause. On the other hand, since the subject-verb agreement is base-generated as an affix in the third person, and since the verb does not assign argument roles (agent, experiencer, etc.) to Agr, a phonetically overt subject is required in order to avoid a violation of the theta-criterion.

In the case of Bislama, though, it is harder to bend the bow. Although third person generally favours a null subject, there is still a sizeable minority of clauses where the third person subjects are pronominal. There is no readily discernible stylistic distinction in Bislama between cases where the third person pronoun is null or overt.¹¹ Adopting Vainikka and Levy's proposal would therefore seem to commit us to claiming that in Bislama third person agreement is generated as agreement 20–40% of the time, while 60–80% of the time it is base-generated as a subject.

¹⁰ A problem with all these discussions is there is seldom any quantitative data to enrich the writers claims (Heap's work is an exception) – one might think Hebrew and Finnish never use pronouns with first and second person and always use pronouns with third. Cameron's and Duranti's work on Spanish and Italian clearly shows that there is a good deal of variability even in the prototypical null subject languages.

¹¹ This question merits further attention. My corpus provides ample opportunity for comparing the speech of individuals with me and with family or friends. It would be interesting to see whether comparing individuals in this way would show addressee, or monitoring, effects. Patterns of null and pronominal subjects might also differ in written Bislama than in face-to-face interaction.

The other possibility is that third agreement is consistently base-generated in the subject position of a clause. In that case, the 20–40% of utterances where third person subjects are overt (*hem/Sale i V*; *olgeta/ol pikinini oli V*) would be clauses where the subject is focused. However, we have already seen that there is a well-established strategy for focusing third person subjects in Bislama (the *NP, hem i V* pattern illustrated in Chapter 5). In other words, a difficulty with this analysis of third person clauses is that it hypothesises a language in which there are two levels of focus in the third person, and only one in first and second.

Perhaps a more economical account would be that there is no structural distinction between the way first and second subjects, and third person subjects are generated (as entire classes), nor in the way their agreement morphology is generated. All null subjects might have the same underlying structure, along the lines proposed by Rohrbacher (1995) and Speas (1996), i.e. the subject-verb agreement is base-generated as the head of Agr. Third person subjects in Bislama would, in my view, simply be more amenable to deletion after the fact because the agreement itself facilitates reconstruction of a subject referent.

The Bislama data also seems to indicate that the hypothesis that first and second person subjects are *a priori* more referential than third person must be rejected. This hypothesis is widely assumed in functionalist syntax. For instance, Chafe (1994:79) refers to “the usually active status of the ideas of the speaker and listener”, but the same assumption also finds its way into some generative accounts of typological variation.

Rohrbacher’s (1994, 1995) account of parametrisation of V2 in the Germanic languages rests on morphological distinctions between speaker (first person) and hearer (second person). He motivates the significance of these distinctions by invoking the ubiquity and hence cognitive salience of the speaker and listener. He hypothesises that in addition to being crucial for the development of V2, the inherently high referentiality of first and second person may be linked to possibilities for *pro*-drop, too. Vainikka and Levy’s (1995) article is an attempt to follow through on Rohrbacher’s prediction.

The idea is that because first and second person are related to the identification of the speaker and hearer, and because the speaker and hearer are communicative constants, reference to the speaker and hearer can be assumed by the speaker, and is maximally retrievable by their interlocutors. Third person, under this view, is defined negatively. It is not [speaker] and not [hearer]. The way the Bislama (and Faetar and Francoprovençal, Nagy and Heap 1998) system has split its *pro*-drop system indicates that cognitive or common sense factors, e.g. the fact that you can count on knowing who the speaker is and (usually) who the hearer is, do not necessarily form the basis on which speakers organise their subject reference system. If we are looking for cross-linguistic generalisations, morphological and interpretive transparency seems to be more salient than some kind of cognitive hierarchy of referential salience among pronouns.

So to sum up the discussion of the data on use of null subjects with different persons and numbers: the significance of the findings in Bislama is two-fold. First, this Bislama data provides clear evidence of the need to extend our typology of mixed *pro*-drop languages. Second, it suggests that interlocutors exploit and rely on the morphological transparency of the system they are working with to constrain and interpret patterns of subject deletion. However, given the frequency of overt third person subjects in spoken Bislama, it seems undesirable to argue that first and second, and third person subjects are base-generated in different positions.

6.10.2 Information status of the referent

Table 6.6 shows the results of investigating the discourse status of the subject referent. Again, the Goldvarb weightings indicate the extent to which that particular discourse condition favours or disfavors a null subject. We see that there is a clear preference for speakers to follow a null subject with another coindexed null subject. All other discourse conditions favour the use of a pronominal subject.

Table 6.6: Goldvarb weightings for all speakers for form and grammatical role of subject referent in prior clause (factor selected by multivariate analysis as having the second most significant effect on occurrence of phonetically null subjects). (0 = pronoun subject; 1 = phonetically null subject).

Information status of subject referent	Goldvarb weighting	N clauses
null subject in prior clause	0.817	1,011
pronoun subject in prior clause	0.406	1,799
null other argument in prior clause	0.385	116
pronoun other argument in prior clause	0.458	358
referent not in prior clause	0.370	1,343

Table 6.6 shows that a phonetically null subject is strongly favoured if both the subject is coreferential with the subject of the preceding clause, and the subject was also null in the preceding clause (difference in form of referent in preceding clause is highly significant, $\chi^2 = 702.49$ with 1 d.f., $p < 0.001$).

Looking at the last line of Table 6.6 we can see that when the referent of the current clause's subject was not even present in the preceding clause, speakers are most likely to use a pronoun. This makes sense of course. If the referent was not in the immediate context, it is less likely to be easily retrievable or identifiable for the hearer.

However, it is more of a surprise that speakers also favoured use of an overt pronoun when the subject of the current clause was also the subject of the preceding clause, but where the referent had been an overt element in the preceding clause. We would not expect this if the constraints operating on the deletion of subjects are primarily constraints related to the discourse salience or identifiability of a referent. If ease of retrievability or identification was the main criterion, we would expect there to be no difference between the weightings for the first two conditions in Table 6.6. That is, we would expect the form of the subject referent in the preceding clause to have no effect.

So while the first linguistic variable we looked at, the person and number of the subject, showed what seemed to be a functional effect on the distribution of null subjects, the discourse variable measuring subject perseverance seems to be showing a counter-functional effect. What we appear to be observing is something like the phenomenon to which Scherre and Naro (1991, 1992) gave the epithet 'birds of a feather', encapsulating the way agreement morphology appears within phrases in Brazilian Portuguese.

While the use of a pronoun is associated with a switch in subjects (the last three lines in Table 6.6), the use of a null subject is not correlated with all cases of subject persistence. So we cannot say that the variation between phonetically null and pronominal subjects in Bislama has been grammaticised as a morphological device for distinguishing switch versus echo-subject reference (a morphosyntactic strategy used in some substrate languages as discussed earlier). Instead, Bislama speakers exhibit a stronger priming effect where like begets like.

Notice, too, what the four discourse conditions that are found in the lower portion of Table 6.6 indicate about the nature of this priming effect. We notice that all four of these conditions have very similar weightings indicating that a pronominal subject is found in these cases significantly more often than a phonetically null subject. The priming principle would lead us to expect this in the cases where we have perseverance of the referent in subject position, and in the prior clause the referent was realised as a pronoun.

However, the same priming effect is not found when the referent switches thematic roles between the clauses. The weightings for the conditions where the referent was a null non-subject argument and a pronominal non-subject argument in the preceding clause are very similar. If anything, they run counter to the priming effect of same subject conditions. Yet, the distribution of tokens in these discourse conditions was within the levels of significance ($\chi^2 = 4.83$, 1 df, $p < 0.05$), indicating that this difference in the weightings should be treated seriously.

This provides us with a clear indication of how straightforwardly priming operates as a surface level phenomenon. Although I have been discussing the tracking of referents, Table 6.6 indicates quite clearly that speakers are not primarily tracking the referent itself. That they are sensitive to switches in subject referent is supported by their tendency to use pronouns when the subject referent was not even present in the preceding clause. But the very similar (low) weightings for pronoun and null forms of non-subject argument in the prior clause indicate that speakers' primary concern is with tracking subjects. We might wonder about the extent to which the priming is sensitive to subjecthood or whether linear order is more important. Generally, subjects in Bislama are the first argument in the clause, but objects can be fronted. An empirical investigation of the salience of first position will have to be left for the future.

6.11 Social factors

6.11.1 Speaker gender

As indicated by the step-up, step-down analysis, there are no significant differences in the way men and women use pronominal and null subjects.

This was verified using the individual means shown in Tables 6.2 and 6.3. A t-test was performed on the means for female and male speakers. The null hypothesis assumes that the means of both samples are drawn from the same underlying population. The t-test returned a t-ratio value of 1.18 ($p > 0.2$). Very little of the variation in the whole corpus ($r^2 = 3.3\%$) was accounted for by speakers' gender.

It should be noted, though, that a t-test is most appropriate as a measure for comparing the distribution of means between the two groups only if the populations being compared have roughly the same internal variance (Woods et al. 1986:182). In this case, the F-ratio of the standard deviations of the samples of men and women was very low ($F_{1,41} = 1.39$, $p > 0.5$). This suggests that the t-test may not be the most appropriate measure for comparing the two groups.

In addition, to these tests, I made a detailed multivariate analysis comparing the variation for female speakers and male speakers. There were no significant differences in the weightings of the linguistic factors for women or men.

Table 6.7: Goldvarb weightings of linguistic factors showing similarity between women's and men's norms for use of phonetically null subjects. (0 = pronoun subject condition favoured; 1 = null subject favoured).

Linguistic environment	Goldvarb weighting	
	Women	Men
null subject in prior clause	0.821	0.811
pronoun subject in prior clause	0.395	0.403
null other argument in prior clause	0.490	0.277
pronoun other argument in prior clause	0.443	0.461
not present in prior clause	0.371	0.399
1SG referent	0.135	0.154
2SG referent (specific)	0.221	0.210
2SG referent (non-specific <i>yu</i>)	0.088	0.246
3SG referent	0.694	0.748
1PL referent	0.285	0.252
2PL referent	0.195	0.178
3PL referent	0.861	0.866
subject of main verb (finite clause)	0.438	0.453
<i>blong</i> + V	0.972	0.952
<i>wantem/traem</i> + V	0.991	0.990
imperative	0.970	0.973
<i>input probability</i>	.41	.41

The very similar patterns of male and female speakers is highlighted in Figure 6.1, which plots regression lines for both sexes' frequencies of null subjects over a regression for age. The significance of age as a factor will be discussed in §6.11.5.

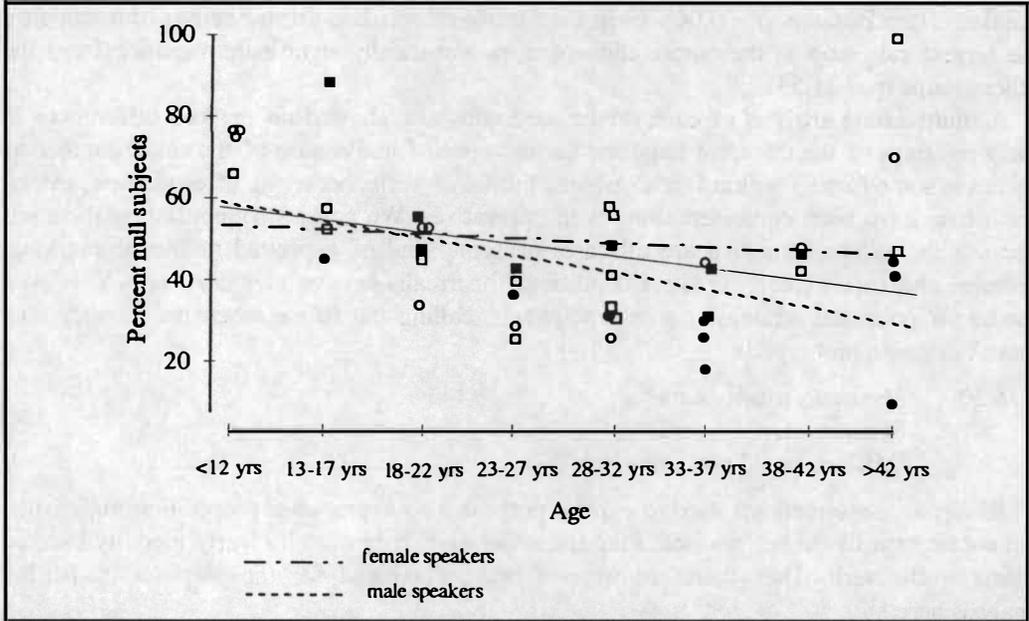


Figure 6.1: squares = female, circles = male speakers. Regression lines plotted for speaker gender overlaid on speaker age, showing very little difference between the female and male speakers with respect to individual mean null subject frequencies. Solid icons indicate urban speakers.

6.11.2 Origin and lanwis

A t-test performed on speakers by their sub-group of origin (the measure intended to highlight any possible substrate influences on their Bislama) indicated this variable was also not significant ($p > 0.1$; $r^2 = 5.3\%$). But again, it should be noted that the extent to which the groups can be reliably compared using a t-test is arguable. The variance between the groups (as measured by the differences in their standard deviations) was marked (returning an F-ratio of $F_{1,41} = 2.28$).

Malakula has not been tested for substrate influences. Only one speaker in the corpus came from (northern) Malakula, so the sample was too small to generalise from. It did not seem prudent to combine this speaker with any other groups, though, as Tryon’s (1968) classification of Vanuatu languages indicates that the languages of northern Malakula constitute a coherent linguistic subgroup of their own, i.e. they are distinct from the languages of both Espiritu Santo and Malo, and the languages of Ambae, Maewo and Pentecost (the two groups of greatest regional proximity).¹²

T-tests indicate that the remaining linguistic subgroups, Sanma, Banks, Penama, Shefa and Tafea, are not significantly different from each other. In each case, the hypothesis that speakers’ mean use of phonetically null subjects represent samples drawn from different underlying populations was not supported (for each of Shefa, Tafea and Banks, $p > 0.3$; for

¹² In fact, recent work by Jauncey (1997, not available to me when conducting these analyses) on the marking of realis/irrealis in northern Vanuatu languages suggests that the subject prefixing system of the northern Malakula language might pattern with the languages of south and east Santo.

speakers from Penama, $p = 0.06$). Even the sample of speakers from Sanma, who made up the largest subgroup in the corpus did not show statistically significant variance from the other groups ($p = 0.053$).

A multivariate analysis of each of the subgroups also showed no marked differences in the weightings of the different linguistic factors coded for. Because of the small number of tokens in some factors within factor groups, tokens of verbs occurring after *wantem*, *traem*, and *blong* have been combined along with imperatives. We can justify conflating these sets because the tokens combined are all cases of some kind of aspectual or modal marking. *Wantem* and *traem* clearly create a conditional or irrealis environment. *Traem + V* is even used as a politeness strategy, e.g. my neighbour calling out to me when her laundry sink wasn't draining (not taped):

6.20 *Meriam, traem kam?*
 Meriam try come
 'Miriam, could you come here?'

Blong, a complementiser used to express purpose also expresses a proposition anticipated but not necessarily (to be) realised. Like the imperative, they are all clearly modally marked forms of the verb. They therefore warrant being combined for the purposes of further analysis here.¹³

Table 6.8: Goldvarb weightings of linguistic factors by speakers' origins showing lack of significant differences, rejecting the possibility of substrate influences on variation between pronominal and phonetically null subjects. (0 = pronoun subject condition favoured; 1 = null subject favoured).

Linguistic environment	Goldvarb factor weight				
	Sanma	Penama	Banks	Shefa	Tafea
null subject in prior clause	0.778	0.882	0.849	0.786	0.822
pronoun subject in prior clause	0.392	0.424	0.440	0.459	0.394
null other argument in prior clause	0.358	0.511	0.410	0.369	0.448
pronoun other argument in prior clause	0.466	0.375	0.430	0.301	0.629
not present in prior clause	0.384	0.352	0.355	0.341	0.434
3SG referent	0.685	0.716	0.692	0.739	0.775
3PL referent	0.890	0.925	0.697	0.780	0.857
1SG referent	0.088	0.192	0.393	0.162	0.156
2SG referent (specific)	0.185	0.251	0.324	0.169	0.306
2SG referent (non-specific <i>yu</i>)	0.146	0.158	0.271	0.152	0.024
1PL referent	0.234	0.227	0.353	0.207	0.408
2PL referent	0.146	0.159	0.226	0.418	n.d.
subject of main verb (finite clause)	0.451	0.451	0.406	0.442	0.435
modally marked	0.976	0.951	0.976	0.975	0.988
<i>input probability</i>	.496	.291	.342	.444	.298

¹³ But note that not all modally marked clauses fall into this group. Clauses marked irrealis with the sentential adverbial *bae* (i.e. most conditionals, futures, and some hortatives) behave like the main (finite) verbs. In other words, with respect to obligatory pronoun and subject-verb agreement marking irrealis *bae* clauses behave the same as realis clauses.

6.11.3 Village and urban residence

Although the multivariate analysis of the whole corpus that was conducted using Goldvarb indicated that the village/urban variable might be significant, further tests did not support this hypothesis. The F-ratio between these two groups was good ($F_{1,41} = 3.79$), but a t-test comparing the means of speakers living in the urban centre of Santo and the villages on Malo, returned a t-ratio value of 1.95 ($p = 0.058$), only at the margins of statistical significance. The amount of the variation accounted for by this variable was somewhat higher than the other social variables we have looked at, but still very low ($r^2 = 8.5\%$).

But there is a strong interaction between the residence variable and the speaker's age, and between the residence variable and the speaker's level of education. Together, residence and education accounted for approximately one eighth of the variation observed ($r^2 = 13.8\%$). Residence and age together accounted for 16% of the variation. However, we know that the strongest interaction is between education and residence. A multiple regression with both residence and education (secondary versus less than secondary, see §6.11.4) reduced the significance of residence greatly ($p = 0.28$, cf. residence and age, residence $p = 0.14$).

The fact that the significance of residence may be dependent on some other factor seems to be borne out by closer examination of the phonetically null subjects in different linguistic environments for both groups. As Table 6.9 shows, a multivariate analysis did not reveal any consistent differences between the two populations in their behaviour with respect to one factor group or another. Goldvarb weightings for different person and number factors showed some variation between the urban and village speakers, with urban speakers using more null 1SG and 1PL subjects than village speakers, but village speakers used more null subjects in 3PL than urban speakers did.

Table 6.9: Goldvarb weightings of linguistic factors for village and urban speakers' use of phonetically null subjects showing slightly higher tendency to use null subjects among village speakers (0 = nonnull subject condition favoured; 1 = null subject favoured).

Linguistic environment	Goldvarb factor weight	
	Village	Urban
null subject in prior clause	0.805	0.824
pronoun subject in prior clause	0.399	0.410
null other argument in prior clause	0.401	0.407
pronoun other argument in prior clause	0.464	0.446
not present in prior clause	0.371	0.398
3SG referent	0.724	0.699
3PL referent	0.903	0.793
1SG referent	0.095	0.226
2SG referent (specific)	0.219	0.185
2SG referent (non-specific <i>yu</i>)	0.156	0.206
1PL referent	0.223	0.356
2PL referent	0.158	0.203
subject of main verb (finite clause)	0.460	0.441
<i>blong + V</i>	0.963	0.968
<i>wantem/traem + V</i>	1.000	0.973
imperative	0.956	0.990
<i>input probability</i>	.437	.361

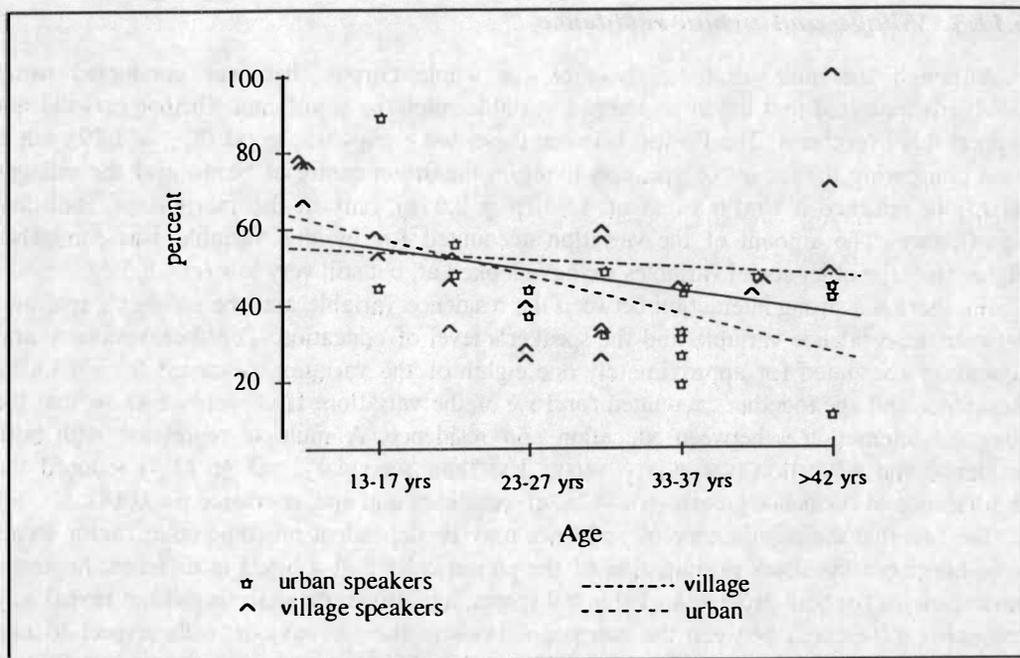


Figure 6.2: Regression lines plotted for speaker place of residence shown against a regression for age.

6.11.4 Level of education

Speakers were divided into two groups: those with (some) secondary education, and those with only primary education or no education. There were two reasons for making the division this way.

Firstly, there is a qualitative difference between primary and secondary education. Most Ni-Vanuatu have some primary education (a number complete primary school), but access to secondary education is less common and much more expensive. There are always more applicants for places in secondary schools than there are places to go around, and the main way of gate-keeping is through a state-administered exam at the end of Class 6. The exam has very high fail rates; this is partly because the government ensures that only enough students pass as there are places for in the state secondary schools, and also because the level of French or English required to pass is quite high. Thus, if speakers' education level is going to exert a significant (superstrate) effects on the patterns of variation in their Bislama, we would expect the transition to secondary schooling to be the most important dividing line.

Secondly, as noted in §6.11.3, level of education, along with age and residence, are not wholly independent of each other. The speakers with no education at all are all in the very oldest age group, and they are also more likely to be village residents.

A t-test comparing speakers with secondary education and speakers with primary or no education returned a value of 2.30 ($p = 0.26$). A regression analysis using level of education as the independent variable accounted for over eleven percent of the variation observed ($r^2 = 11.4\%$)

But a multivariate analysis of the distribution of phonetically null subjects with respect to this variable provides a more nuanced picture of the variation. Table 6.10 shows the Goldvarb weightings for each of the linguistic factors, contrasting speakers with secondary education in one column and speakers with primary or no education in the second column. We can see that the main difference between speakers with some secondary education and those without is the frequency with which 1SG and 2SG subjects are realised as phonetically null. (As in Table 6.8, several verb environments have been combined in a class of modally marked clauses because of low Ns in some of the factor groups.)

Table 6.10: Goldvarb weightings of linguistic factors contrasting speakers with some secondary with speaker who had less than secondary education. (0 = pronoun subject condition favoured; 1 = null subject favoured).

Linguistic environment	Goldvarb factor weight	
	Secondary	Not sec.
null subject in prior clause	0.842	0.786
pronoun subject in prior clause	0.406	0.389
null other argument in prior clause	0.408	0.376
pronoun other argument in prior clause	0.535	0.386
not present in prior clause	0.401	0.373
3SG referent	0.745	0.665
3PL referent	0.887	0.838
1SG referent	0.203	0.100
2SG referent (specific)	0.322	0.120
2SG referent (non-specific <i>yu</i>)	0.124	0.188
1PL referent	0.290	0.241
2PL referent	0.216	0.164
subject of main verb (finite clause)	0.433	0.457
modally marked	0.976	0.968
<i>input probability</i>	.320	.516

Speakers with less than secondary education more strongly favour pronoun subjects when the referent is 1SG or 2SG than speakers with secondary education are. As we will see in the next section, this pattern of exaggerating community standards is characteristic of the youngest speakers, who, by definition, have less than secondary education. Again, the significance of level of education as a variable must be presented cautiously. It is not completely independent of another social variable, age.

6.11.5 Age of speaker

As an initial basis for comparison, speakers were divided into four age groups, Group 1 = 17 years and under, Group 2 = 18–27 years, Group 3 = 28–37 years, Group 4 = over 38 years. Individual speakers were coded in eight different age groups, but these were combined into the four groups shown here to allow comparison of the full range of linguistic factors at each age.

These sub-divisions are rather younger and smaller than those often used in studies of variation across larger communities, but they correspond roughly to generational divisions in the communities of Santo and Malo. Members of Group 2, for instance were at the age where they were settling down and having children, members of Group 4 were of an age where their children were getting old enough to start settling down themselves.

Regression analyses and t-tests indicated that the variation in use of phonetically null subjects across these four age groups was significant (t-ratio = 2.11, p = 0.04; r² = 9.8%).

Figure 6.3 shows that the speakers who use most null subjects are those in the youngest and oldest age groups. A chi-square test of the significance of the variation between the youngest, oldest and middle age groups' rates of pronominal and null subjects found the differences to be highly significant ($\chi^2 = 144.36$ with 3 d.f., p < 0.001). But as the regression line shows, overall the trend is downwards with age.

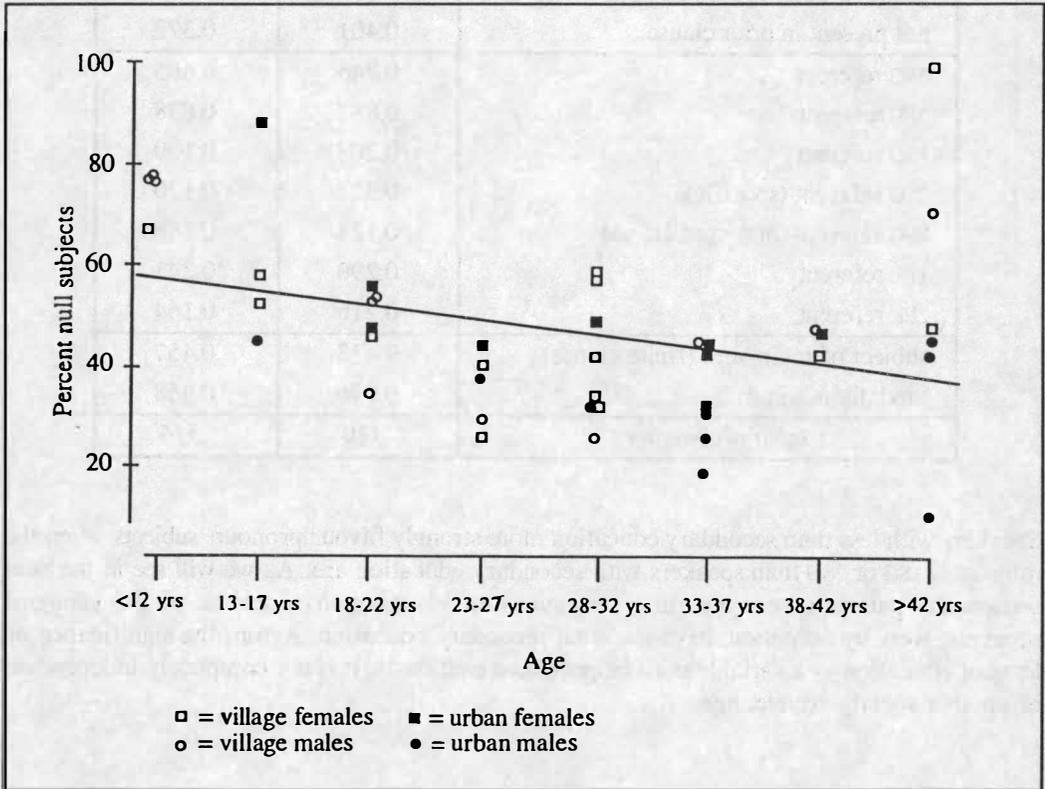


Figure 6.3: Age regression for all speakers showing gradual overall increase in use of null subjects in younger age groups.

Again, though, these data reward closer examination. A detailed analysis of the variation in the four age groups showed considerable fluctuations among the older and younger speakers in the frequency of null subjects within the linguistic factor groups. The weightings for individual factors are shown in Table 6.11, and Figures 6.4 and 6.5 present the information graphically.

Table 6.11: Goldvarb weightings of linguistic factors for four age groups, showing no curvilinear pattern for age, only an increased tendency to use phonetically null subjects among youngest speakers. Gp 1 = under 18 yrs, Gp 2 = 18–27 yrs, Gp 3 = 28–37 yrs, Gp 4 = over 38 yrs. (0 = pronoun subject condition favoured; 1 = null subject favoured).

Linguistic environment	Goldvarb factor weight			
	Gp 1	Gp 2	Gp 3	Gp 4
null subject in prior clause	0.755	0.805	0.849	0.799
pronoun subject in prior clause	0.366	0.409	0.407	0.448
null other argument in prior clause	0.237	0.389	0.521	0.394
pronoun other argument in prior clause	0.463	0.393	0.539	0.375
not present in prior clause	0.378	0.392	0.375	0.355
3SG referent	0.703	0.704	0.724	0.641
3PL referent	0.812	0.909	0.870	0.869
1SG referent	0.057	0.123	0.222	0.142
2SG referent (specific)	0.057	0.305	0.255	0.221
2SG referent (non-specific <i>yu</i>)	0.179	0.114	0.126	0.198
1PL referent	0.193	0.297	0.296	0.231
2PL referent	0.105	0.273	0.175	0.346
subject of main verb (finite clause)	0.465	0.436	0.433	0.452
modally marked	0.973	0.970	0.979	0.977
<i>input probability</i>	.631	.382	.313	.489

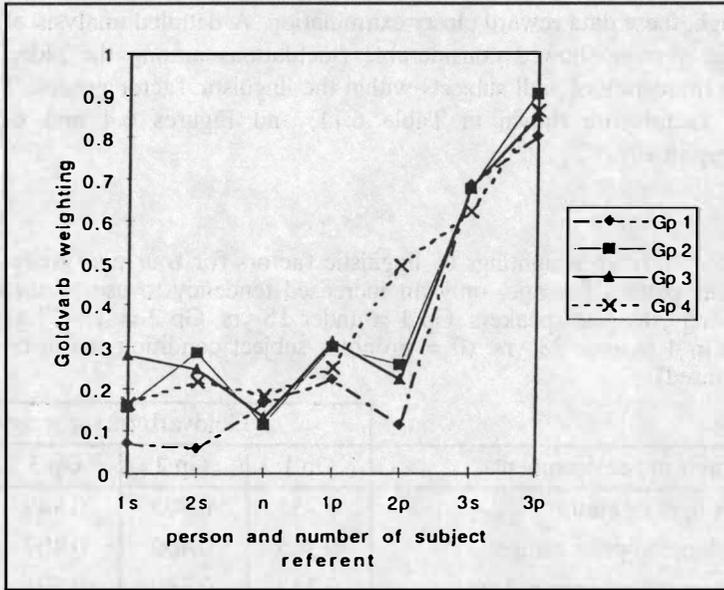


Figure 6.4: Goldvarb weightings showing influence of morphosyntactic factors on null subjects in four age groups of speakers. Note no consistent pattern of Groups 1 and 4 leading Groups 2 and 3 within the factor group. Along x-axis: person and number variants, n = non-specific subject referent

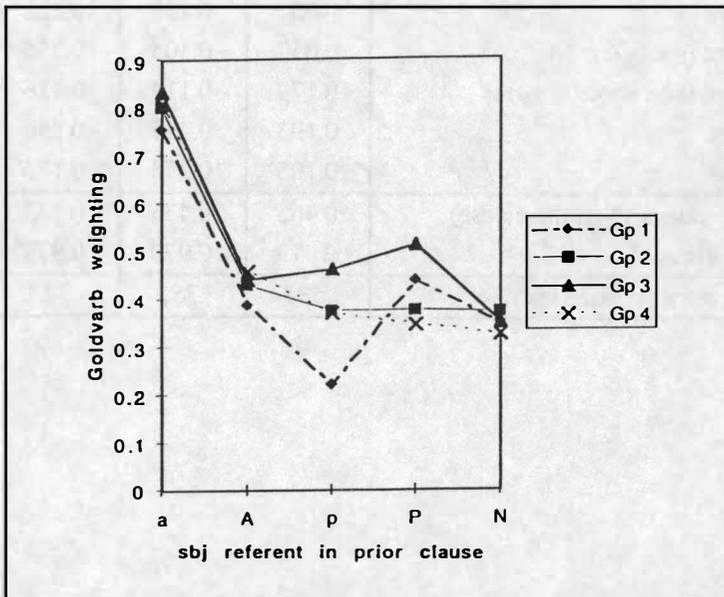


Figure 6.5: Goldvarb weightings showing influence of discourse factors on null subjects in four age groups of speakers. Note no consistent pattern of Groups 1 and 4 leading Groups 2 and 3 within the factor group. Reading along x-axis: a = null subject in prior clause, A = pronoun subject in prior clause, p = null other argument, P = pronoun other argument, N = not in prior clause.

Any differences with respect to Group 4 (the oldest age group) should be interpreted conservatively. As I noted in §6.11.4, there is a correlation between older age and no education. This interaction is intimated by details of the multivariate analysis. Goldvarb calculates the chi-square per cell for Group 4 to be quite high ($\chi^2/\text{cell} = 2.9$). This statistic is a good indicator that there is an interaction between several of the factors tested.

In addition, all the youngest speakers (under 12 years) in the corpus were recorded on Malo. The (statistical) analysis of variation in Group 1 speakers did not indicate a strong interaction between younger speakers' age and other factors (the chi-square per cell was acceptable, $\chi^2/\text{cell} = 1.7$). But given that we know Group 1 breaks down into further socially cohesive subgroups, I shall examine the behaviour of the very youngest speakers in the sample in the next section.

6.11.6 Caregivers and ol smol papa: other possible models for children

The children's use of phonetically null subjects was compared with two other groups in the Malo community: their immediate caregivers and the younger classificatory parents of the children.

Immediate caregivers were the children's mother and father and the *haosgel* living with them, in one case a sister of the mother and in another case the niece of the father. The younger classificatory parents were those people the children called *smol papa* 'small father'. The term is a derived one: *papa* is the classificatory term for all men their parents call *brata* 'brother',¹⁴ and *smol* indicates their relative youth.

Obviously, the children spend most time in contact with their immediate caregivers. In terms of quantity of talk, these are their chief models and socialisers. However, the quality of the relationship between children and their *smol papa* is also meaningful. In the families I lived with on Malo, the children see a lot of their *smol papa*, and because these people are closer in age to the children than the children's parents are, their linguistic and social patterns might be a more salient model than the children's caregivers' patterns.

Table 6.12 provides the Goldvarb weightings of the linguistic factors for each of the three groups of speakers, children, immediate caregivers and people in the corpus who count as the children's *smol papa*. Figures 6.6 and 6.7 represent the same information graphically. The data is presented as a line graph in order to highlight parallelisms and differences between the different groups' behaviour (I do not claim that the factors form a continuum within groups).

There are two things to notice about the data in Table 6.12 and illustrated in Figures 6.6 and 6.7. The first is the great similarity in the behaviour of the children's immediate caregivers and the *smol papa*. Though it was hypothesised that in social terms the two groups might exert different strength as linguistic role models for the children, the distinction proves irrelevant with respect to this variable. The behaviour of the caregivers is not to be distinguished from the behaviour of the *smol papa*. The Pearson correlation statistic comparing the weightings for the caregivers and *smol papa* given below Table 6.12 is highly

¹⁴ By definition, on their father's side. On Malo, their mother's brothers are called *tawean* by the father (and his brothers), and the children call their mother's brothers *angel*. See Rubinstein 1976, Chapter 3 for more detail on traditional kinship terminology in Tamambo.

significant ($p < 0.001$). The Pearson correlation is an appropriate statistic to use under these circumstances because it tests the relatively strong hypothesis that the two sets being compared are interdependent.

Table 6.12: Goldvarb weightings of linguistic factors comparing pronominal versus null subjects in speech of children (speakers under 12 years: Elise, Iawia, Petre, Saemon), their immediate caregivers (Leipakoa, Lolan, Alis, Susana, Visi and Mesek), and people the children call *smol papa* (Obed, Atesolo, Rovi, Susana). (0 = pronoun subject; 1 = null subject). Note increase in input probability for under 12 year speakers (compared to under 18 years, Table 6.11), and lack of evidence that children are matching the speech of their caregivers.

Linguistic environment	Goldvarb factor weight		
	Children	Caregivers	Smol papa
null subject in prior clause	0.637	0.811	0.829
pronoun subject in prior clause	0.299	0.413	0.343
null other argument in prior clause	0.138	0.535	0.549
pronoun other argument in prior clause	0.579	0.526	0.362
not present in prior clause	0.486	0.363	0.389
3SG referent	0.668	0.686	0.712
3PL referent	0.725	0.949	0.918
1SG referent	0.029	0.130	0.053
2SG referent (specific)	0.008	0.191	0.036
2SG referent (non-specific <i>yu</i>)	0.476	0.113	0.060
1PL referent	0.013	0.255	0.315
2PL referent	0.039	0.278	no data
subject of main verb (finite clause)	0.453	0.427	0.414
modally marked	0.990	0.991	0.996
<i>input probability</i>	.787	.414	.520

Pearson correlation between children and parents: 0.763, $p < 0.01$

Pearson correlation between children and *smol papa*: 0.777, $p < 0.01$

Pearson correlation between parents and *smol papa*: 0.984, $p < 0.001$

Pearson between children, and *smol papa* and parents' average: 0.797, $p < 0.001$

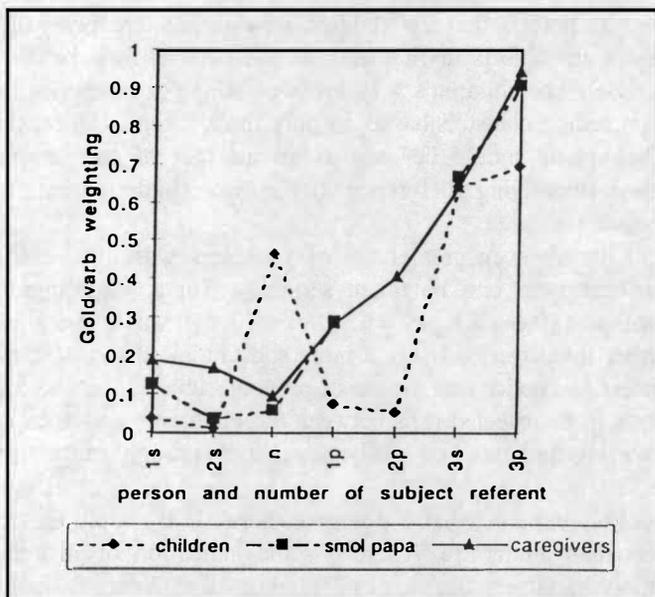


Figure 6.6: Goldvarb weightings showing influence of morphosyntactic factors on null subjects for children, their immediate caregivers and children’s *smol papa*. Note similarity of caregivers and *smol papa*, and children’s heightened use of pronominal subjects in first and second person (0 = pronoun subject; 1 = null subject).

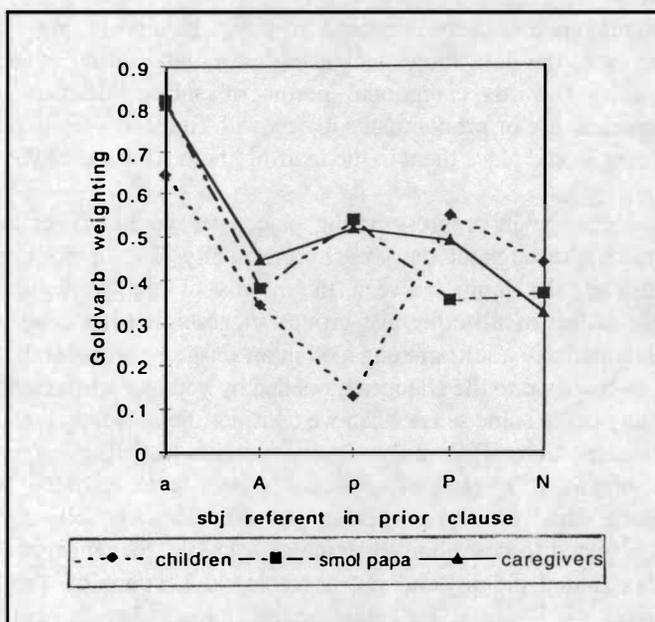


Figure 6.7: Goldvarb weightings showing influence of discourse factors on null subjects for children, their immediate caregivers and children’s *smol papa*. Note similarity of caregivers and *smol papa*. (0 = pronoun subject; 1 = null subject).

The second thing to note is that the children's weightings are generally parallel to their caregivers' but that with subjects in first and second persons their behaviour departs from their caregivers' model. The children's very low weighting for phonetically null non-subject arguments in the preceding clause is based on only three tokens. Hence, this departure from their caregivers' behaviour should be seen as an artefact of the sample, rather than as reflecting a genuine underlying difference in the way children treat these arguments in syntactic and discourse terms.

However, the children's even greater use of pronouns with 1SG, 1PL and 2SG referent subjects than their caregivers can be taken seriously. These weightings are derived from somewhat larger samples (1SG = 21, 1PL = 13, 2SG = 17; 2PL was a very small sample, N = 3), but obviously further investigation using a more substantial corpus of children's Bislama is desirable. Nonetheless, we notice that the use of pronoun subjects for 1SG and 2SG among the *smol papa* (speakers in their teens) falls between the children's and their caregivers' use of pronoun subjects for 1SG and 2SG, and this begins to look strongly suggestive of some kind of change over time.

What could possibly explain what is going on here? Is it simply that children are more egocentric speakers than adults are? There is some indication of this in very early child language acquisition literature. Slobin (1977) attests that Spanish speaking children go through periods when they use more pronouns than their parents. But the acquisition data does not tell us much about 8–11 year olds. If the relative lack of socialisation and greater egocentricity of the children is what is responsible for their greater use of first and second person pronouns, then this data suggests that the acquisition of non-egocentric, or more altruistic, conversational norms is a process of socialisation and acquisition that continues well into adolescence.

The other possibility is that there is indeed on-going change taking place in the speech community. In this case, the data shows increasing grammaticisation in the syntactic system, with speakers moving towards categorical norms of subject deletion with third person referents and categorical use of pronominal subjects with first and second person. In this case, the children's patterns would place them in the lead of the next stage in the grammaticisation process.

But if there is a change in progress taking place, we would expect to see a consistent pattern of age-grading throughout the speech community. This is not demonstrated in the Bislama data. Instead, the jump between the norms of the youngest members of the community and the norms of all other age groups suggests that if change is taking place in the community, then perhaps adults are engaged in an on-going process of restructuring their linguistic systems in response to the changes modelled by younger speakers in the community.

This proposal may make some sense when we consider the sociolinguistic characteristics of a language like Bislama. Given Bislama's primary function as a lingua franca, and given its strong affective role as a symbol of national unity, there may be strong social and instrumental pressures which encourage speakers to maintain a flexible grammar. Jeff Siegel (pers.comm.) has objected to describing Bislama as a creole. Siegel argues Bislama is better described as an 'expanded pidgin' and not a creole on account of the high numbers of multilingual speakers of Bislama. He believes this means that a continued transfer of substrate phonology and grammar is possible with Bislama (something that is not characteristic of, for instance, the Caribbean creoles) and which means there is likely to be considerable internal variability in Bislama. This is a valid observation but it ignores two facts about Bislama.

One is the early age at which speakers acquire Bislama (well within the critical period). This means that *a priori* we should expect no more transfer from the learner's other native language(s) than we find in other communities where children are multilingual from an early age. The second fact that is ignored is the wide range of social functions Bislama serves (everything from language of the home to theatre). Jourdan (1985) makes a good case for using the social expansion of a contact language as the basic criterion for distinguishing creoles from (expanded) pidgins.

The distinction is not pure sophism, since it is directly concerned with the autonomy of the linguistic system being described. Calling Bislama an expanded pidgin is to claim that the syntactic (and phonological) systems in Bislama are not autonomous, and that variation observed in Bislama is actually a reflex of interlanguage variation in Vanuatu. On the other hand, I take an assertion of creole status for Bislama to entail a claim that Bislama does indeed have an autonomous grammar.

The data on subject variation that has been presented here clearly indicates that, for this variable at least, internal pressures, not substrate factors, are principal constraints on the variation observed. In addition, it seems the only non-linguistic variable which has emerged as a potentially important constraint on the variation is not L1 or origin of the speaker, but rather younger speakers' reanalysis and grammaticisation of existing structural patterns.

What this means is that Siegel may be right in principle, i.e. that continued multilingualism in Vanuatu plays an important part in the shaping the profile of Bislama variation, but that he may be wrong in the details. If it is correct, as hypothesised here, that the relative uniformity of all adults versus the youngest speakers' distribution of null subjects has been caused by all adults continuing to level towards the models provided in children's speech, then this requires some explanation. Observations of change via apparent time (discussed in §6.6.3) rely on the fact that research into language variation and change has typically found speakers do not significantly alter their linguistic system as they age.¹⁵ However, because Bislama remains the lingua franca of a linguistically highly diverse society, this may establish the necessary social conditions for adults' continuing reorganisation of their grammar of Bislama.

6.11.7 Malo speakers' linguistic networks

There is another possible account of the children's behaviour that needs to be discussed and ruled out. This relates to the children's linguistic networks on Malo. In Chapter 3, I showed that it was possible to group the Malo speakers on the basis of their reported (and observed) use of Bislama and Tamambo. In that discussion, I omitted the children, because I pointed out that their networks and linguistic choices are more tightly circumscribed by others than adults are.

One possibility, however, might be that the children growing up in the villages on Malo might be demonstrating the same patterns of variation as the adult speakers who expressed the strongest self-identification with Malo culture. In this section I will look at the behaviour of the different Malo subgroups and the Malo children.

The Malo adults were divided into four groups that corresponded to the relative strength of their identification with Malo culture and the Tamambo language, and the frequency with

¹⁵ As noted above, this may be one major respect in which phonological and syntactic variables differ.

which they use Tamambo. Group A were the speakers who had either chosen not to leave Malo or who expressed the strongest concern for language maintenance and used the language most often with children and even *ol woman nara aelan*. The children's norms for subject deletion do not match the norms of the adults who had the most multiplex and dense social networks on Malo, and indeed, there is no clear stratification of the Malo speakers based on these attitude and language use subgroupings as Figures 6.8 and 6.9 show. Numbers of tokens for the discourse condition where the current subject was a null non-subject argument in the preceding clause are low. Divergences in the weightings between groups on this factor should be interpreted cautiously, with this in mind.

We can best interpret the data in Figure 6.8 in conjunction with the patterns we found in the detailed examination of the use of null subjects in the speech of children, their immediate caregivers and their *smol papa* in Figures 6.6 and 6.7, above. Figures 6.8 and 6.9 clearly indicate that the grading shown in Figures 6.6 and 6.7 can be interpreted as evidence of a degree of age-grading in the community. Figure 6.9 shows that the Malo community shares one grammar, i.e. the variation between null and pronominal subjects in different grammatical and discourse conditions is not stratified socially.

Figures 6.8 and 6.9 show that children are following the patterns of their caregivers, but appear to have grammaticised the distinction between first and second, and third person subjects even more than their parents have. Figure 6.8 shows that there was a better match between the children's caregivers (and *smol papa*) and the children's behaviour than there is between the children and the Group A Malo speakers, i.e. those who expressed the strongest positive affect towards Tamambo and local Malo culture, and who used Tamambo the most.

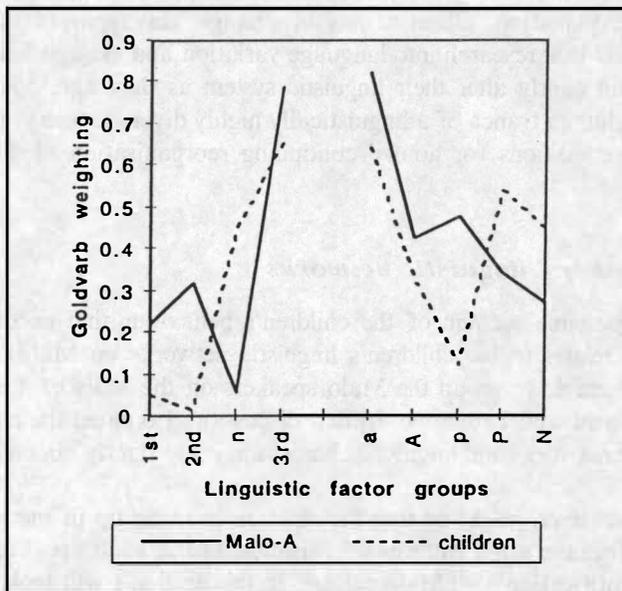


Figure 6.8: Comparison of weightings for children and members of Malo Group A, adults with strongest Malo ties, indicating that children are not matching the patterns of adult speakers with strongest local affiliations. (0 = pronoun subject; 1 = null subject).

Key to x-axis: a = null subject in prior clause, A = pronoun subject in prior clause, p = null other argument, P = pronoun other argument, N = not in prior clause, person and number variants, n = non-specific referent.

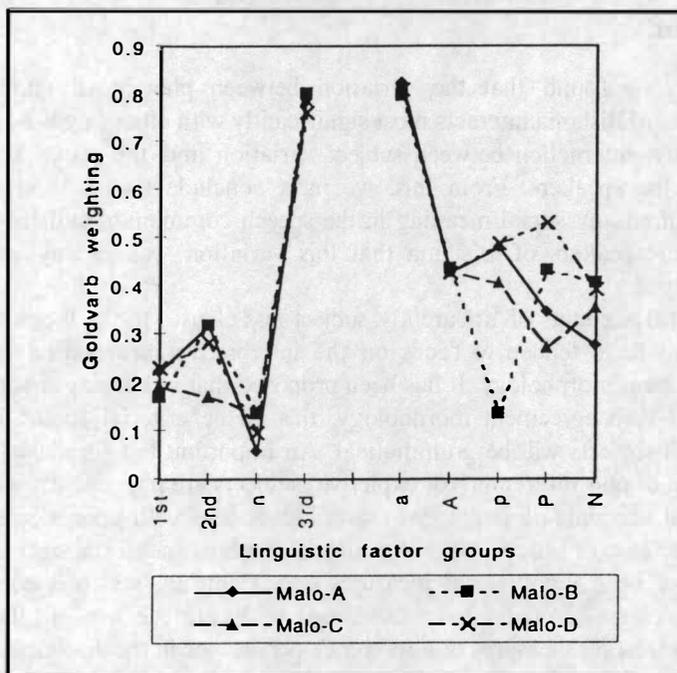


Figure 6.9: Comparison of weightings for the four subgroups of adult Malo speakers, showing strong consistency across the community grammar, irrespective of patterns of language use and strength of affiliations with Malo culture. (0 = pronoun subject; 1 = null subject).

Key to x-axis: a = null subject in prior clause, A = pronoun subject in prior clause, p = null other argument, P = pronoun other argument, N = not in prior clause, person and number variants, n = non-specific referent.

We can conclude, then, that the transmission and development of the community grammar is located primarily in the children’s immediate environment. Although children’s social networks are perforce strongly tied to Malo, and they lead a Malo-centric existence, their *pro*-drop norms do not match the norms of the adults who have chosen to maintain strong ties to Malo. Rather, as we would expect, there seems to be a transmission of *pro*-drop norms within families. Successive generations (caregiver to *smol papa* to caregivers’ children) seems to be further grammaticising the split *pro*-drop system in Bislama that already exists in the community grammar. If Charpentier (1979) was correct, and Bislama *pro*-drop is an innovation, it has spread remarkably quickly through the community, and shows signs of rapidly stabilising into a pattern of pronoun use with first and second person, and phonetically null subjects with third.

Note that this has implications for the syntactic analysis. If the children are pushing their parents’ patterns of *pro*-drop towards complete grammaticisation, Bislama speakers may yet end up with a system of split *pro*-drop that can be described in the terms Vainikka and Levy (1995) proposed. My chief objection to trying to fit Bislama into their model was that the frequency with which Bislama currently realises third person subjects with both pronoun and subject-verb agreement. As 6.19a, b showed, their model left no room for a separate pronominal subject to be generated. A fully grammaticised distinction between third person and other subjects might, though, merit re-evaluation in these terms.

6.12 Conclusion

In this chapter, we found that the variation between phonetically null subjects and pronominal subjects in Bislama interacts most significantly with other linguistic factors. There is very little, if any, interaction between subject variation and the social or demographic characteristics of the speakers. From this, we must conclude that it is unlikely that the variation has acquired any social meaning in the speech community, and indeed, there was no indication from speakers of Bislama that this variation was in any way socially or regionally stereotyped.

I reviewed several accounts of apparently subject-less clauses in the literature, noting that generative accounts have tended to focus on the interpretive information encoded in the subject-verb agreement morphology. It has been proposed that in languages with referentially transparent subject-verb agreement morphology, this is the criterial factor in determining whether or not null subjects will be grammatical. An important test for these so-called *pro*-drop languages is that non-referential (or expletive) subjects are phonetically null.

More functional accounts of languages that allow clauses with phonetically null subjects have focused on the ease of interpretation through algorithms measuring referential salience. Some of these have been highly local measures (e.g. Centring, which is concerned with a sequence of three clauses), others have attempted to locate the use of phonetically null subjects within much larger measures of a referent's persistence in the discourse.

I defined the envelope of variation between phonetically null subjects and pronominal subjects for Bislama as being referential subjects associated with lexical verbs in finite and non-finite clauses. Excluded, therefore, were the second verb in serial verb constructions, verbs immediately following *go* or *kam*, and expletive subjects. The second verb in SVCs and the verbs following *go* and *kam* provide aspectual information relevant to the main verb and never occur with a pronominal subject. Expletive subjects are also categorically null.

I then examined the variation between phonetically null and pronominal subjects in Bislama for significant differences with respect to a number of linguistic and social variables. The person and number of the subject referent showed the strongest correlation with null subjects. Third person subjects were very likely to be realised as null subjects, while first and second person subjects were more likely to be realised as pronouns. This was explained in terms of the morphological transparency of the 3SG and 3PL subject-verb agreement morphology. Bislama therefore provides clear evidence of a formal split in the *pro*-drop system in a direction which has received relatively little attention in the syntactic literature.

A surface level priming effect was also found. Phonetically null and pronominal subjects were not in complementary distribution with respect to the persistence of the subject referent across consecutive clauses, but I noted a tendency for phonetically null subjects to follow another (co-referential) phonetically null subject.

Social factors showed little interaction with the variation. There was some indication that the very youngest speakers in the corpus have generalised the use of pronominal subjects with first and second person referents more than adults have, but there was no pattern of steady age-grading. I suggested that this indicates that if a change in progress is actually taking place (towards full grammaticisation of a distinction between first and second versus third persons), the lack of steady age-grading indicates that adult speakers of Bislama are altering their norms even as they age. In a larger discussion of the use of apparent time to divine changes that might have taken place in real time, I noted that several sociolinguistic studies have suggested that there may be a qualitative difference in the rigidity of phonological and

syntactic variables. Speakers seem to carry their phonological vernacular norms with them as they age, while their vernacular syntactic norms may be more flexible. Since the bulk of the work on contact languages like Bislama and Tok Pisin has focused on syntactic variation, not phonological, we do not know whether the speakers' phonological systems remain as stable in these languages as they do in, for example, vernacular English. The acquisition and maintenance of phonological systems in these contact languages (where wide-spread multilingualism continues to be the norm) remains an avenue for future investigation.

7 *Serial verbs in Bislama*

7 Introduction

We have seen that in some clauses in Bislama, the subject may vary between being overt or phonetically null. However, some clauses never occurred with an overt subject. We noted that verbs occurring as the complement to the lexical verbs *wantem*, *traem*, *go* and *kam*, for instance, almost categorically occur with a phonetically null subject. This was also the case with complements of purposive *blong*. It was also noted that the second verb in a serial verb construction (SVC) never occurs with an overt subject. In this chapter, I will survey previous analyses of the structure of SVCs and define SVCs along strict semantic criteria. I will then look more closely at the form of SVCs in Bislama, and I will show that the invariant appearance of the subject-verb agreement marker *i* on the lower verb (V_2) in an SVC is a consequence of the underlying structure of the V_2 . I will conclude by arguing that the other constructions in which we categorically find phonetically null subjects are also a consequence of the restricted structure of the complement verb.

7.1 Defining serial verb constructions (SVCs)

The interesting theoretical question with respect to SVCs is whether cross-linguistically we can present a unified account of their syntactic structure just as we can provide a unified account of their semantic and interpretive properties. The latter seems to be a more tractable task cross-linguistically, while the syntactic structures associated with these semantic and interpretive properties vary. I will adopt a conservative definition of SVCs for which the necessary conditions are semantic and syntactic:

7.1 Definition of SVCs

a chain of two (or more) verbs with a simultaneous (but semantically transparent) event interpretation, lacking overt co-ordinators, i.e. (skeletal) $N_1 V_1 (N_2) V_2 \dots$

These properties are found in other definitions of SVCs in the literature, but other definitions often have disjunctive criteria (Jansen et al. 1978) or (as in the case of Sebba 1987) explicitly give primacy to structural properties. I believe this is a mistake. An emphasis on the semantic property of SVCs (multiple verbs, one event) provides a clear and discrete class of constructions. I believe the proper way to go about investigating the structural

properties of SVCs is to assume that their structural properties are a consequence of their semantic properties. In my opinion, structural definitions of SVCs tend to bleed into cases of simple clausal co-ordination (with non-overt co-ordinators), where the interclausal dependencies may be epiphenomenal or chance.

I will, therefore, use (1) to define the necessary conditions for categorising a concatenation of verbs as an SVC because I believe that these are the core serialisations. I believe that our understanding of the structure and semantics of less core serialisations (i.e. those which meet one but not both of these criteria) only confuse the issue at the moment. We simply do not understand enough about the structure of core serialisation strategies to merit extending our analysis to peripheral varieties where a number of variables (syntactic, semantic, functional) must be considered at the same time. I hope that by restricting my focus in this discussion to core serialisations as defined in 7.1, the relations obtaining between elements and referents in these kinds of SVCs will become clearer and we will be able to move towards a cross-linguistic account of the ways in which SVCs can vary parametrically in their structure.

Some examples, then, of the phenomenon in Bislama are shown in 7.2–7.4:

- 7.2 *Be wan big ston i ron i kam.*
 but one big stone AGR run AGR come
 ‘But this huge stone rolls down.’ (S-95-9, Lisa)
- 7.3a. *Afta nao oli ron i kam.*
 after now AGR run AGR come
 ‘And so they ran over.’ (M-94-2, Susana)
- b. *Afta mi stap wokbaot i go.*
 after 1SG PROG walk AGR go
 ‘And so I was walking off.’ (M-94-2, Lolan)
- c. *From taem yu pas i go aot...*
 because time 2SG pass AGR go out
 ‘Because when you left...’ (M-95-15, Jackson)
- 7.4a. *Tin biskit ya oli go putum i stap daon longwe.*
 tin biscuit SPEC AGR go put AGR stay down there
 ‘The biscuit tin was put over that way.’ (M-94-2, Leipakoa)
- b. *Afta oli sakem mi i go long solwota.*
 after AGR throw 1SG AGR go PREP saltwater
 ‘So they threw me in the ocean.’ (M-94-6, Visi)
- c. *[Yumi] karem ol ting i kam long haos.*
 1PL.INCL carry all thing AGR come PREP house
 ‘We bring everything in to the house.’ (M-95-8, Vosale)

In Bislama, the set of SVCs is fairly small (there are strict semantic restrictions on the type of verb that can appear as the V_2 discussed fully in §7.3.1), and they always meet the criteria set out in example 7.1. They are also restricted to the concatenation of two verbs (or as I shall argue, clauses), so henceforth in this discussion I will refer only to two verb SVCs. I have been unable to find examples where more than two verbs are concatenated without overt co-ordination and where they also express only one event. This contrasts with Tok Pisin, where SVCs may have more than two verbs concatenated with a single event interpretation

(Sankoff 1984:113-114; Verhaar 1995:112), and also excludes cases of multiple iterations of *i go* conveying duration (also found in Tok Pisin).

The issue that any comprehensive account of SVCs must deal with is how the semantics of SVCs are mapped differently into the syntax of various languages. Specifically, what is the structure associated with the second or lower of the two verb phrases in a two verb SVC? A number of people have attempted to account for the distribution of the functional categories marked in SVCs. Attention has focused almost exclusively on the distribution of TMA auxiliaries or inflection, and the presence of overt versus null subjects in the two clauses. Bickerton (1993) does also consider the distribution of subject-verb agreement marking.

7.1.1 *Generative analyses of SVCs*

Within the generative framework, Woolford's (1977) analysis of SVCs as a V with a VP sister, and Baker's (1989) analysis of them as two sister VPs, immediately dominated by one IP, has given way to attempts to describe the dependence of the lower clause on the higher one in terms of syntactic anaphoricity and arguments about the empirical adequacy of the accounts forwarded. A brief review will suffice to give the flavour of the debate.

Borer (1989) proposes that the relation between the lower verb and the higher verb in Saramaccan SVCs is a result of the lower clause having an anaphoric AGR, coindexed with the AGR of the higher clause. The co-indexation of the V_1 AGR and the co-indexation of the higher and lower AGR result (by commutativity) in the lower clause having the same subject as the higher clause. Obviously, this account will be inadequate for dealing with instances of SVCs where the lower clause subject is interpreted as being coindexed with the higher clause object.¹ Bickerton (1993) essentially follows Borer's proposal in his analysis of SVCs in a range of creoles, including Seselwa. However, Corne et al. (1996) have subsequently questioned the robustness of Bickerton's Seselwa data, which makes it difficult to evaluate Bickerton's (1993) conclusions.

Byrne (1990, 1991), responding specifically to Borer's analysis of the Saramaccan data, points out that his account greatly abstracts away from the variation found in Saramaccan. He argues that tense, not agreement, is the functional category responsible for the higher and lower clauses in Saramaccan having the same TMA interpretation, and often the same TMA auxiliaries. He argues though that the variation in where the TMA auxiliaries occur (with the V_1 and/or the V_2) indicate that the correct analysis of SVCs is to have a single tense operator raising to the head of the SVC. Once this operator has established its scope over both clauses by being in a position of structural dominance, Byrne argues that only a feature spreading account of TMA auxiliaries captures the full range of data. What this amounts to is a claim that the auxiliaries are base-generated as variables without semantic content.

With Hale's (1991) treatment of Misumalpan SVCs, the notion of anaphoricity returns, but Hale argues that the Tense of the lower clause is where the locus of anaphoricity lies, not AGR. Collins (1995) also analyses the interpretive identity of the two clauses in SVCs as being manifested in their sharing a single tense phrase (TP). Collins formulates his proposal in Minimalist terms, suggesting that what distinguishes serialising languages from non-serialising languages cross-linguistically is that the former allow multiple V features to be checked in the head of Tense, while in the latter, the V feature associated with the head of Tense can only be checked once.

¹ For Woolford (1977:176), these are the strict definition of SVCs.

The notion of multiple feature checking has gained some attention recently since it was incorporated into the Minimalist Program (Chomsky 1995). The possibility that languages may allow more than one complementiser position was forwarded by Reinhart (1981) as an account of how a *wh*-element and a complementiser may come to appear at the head of relative clauses in Hebrew, and where movement tests indicate that this is not a case of scrambling. Chomsky (1995:296) suggests multiple subject constructions are licensed in Icelandic because Icelandic allows the N feature in the spec of AGR to be checked by multiple NPs. Richards (1997) explores the possibilities of multiple feature checking further and suggests that if languages are parametrised so as to allow multiple feature checking, this would account for multiple *wh*-movement in Bulgarian and *wh*-scrambling in Japanese. Thus, Collins' proposal falls within a wider endeavour to explore the cross-linguistic implications of the notion of multiple feature checking.

7.1.2 Functional accounts of SVCs in Oceanic languages

A separate tradition of analysing SVCs in more functional terms is well-developed in the literature on Oceanic languages. Following Crowley's (1987b) influential typology of SVCs for Paamese, a whole issue of *Oceanic Linguistics* (1993) was given over to the description and discussion of SVCs in a number of languages. These articles largely followed Crowley in distinguishing four kinds of SVCs:

- (i) same-subject SVCs (interpretive subject of V_1 = interpretive subject of V_2);
- (ii) switch-subject SVCs (interpretive subject of V_2 = interpretive object of V_1);
- (iii) multiple object SVCs (sub-class of same subject and switch subject SVCs, both V_1 and V_2 have their own object); and
- (iv) ambient serialisations (V_2 is predicated of the event or proposition expressed in preceding clause).

Of course, not all serialising languages make use of all these strategies.

Sperlich (1993) shows that Namakir only has SVCs that are same-subjects and ambient serialisations. Early (1993) focuses on serialising strategies in Lewo and notes that there is inter- and intralinguistic evidence from Lewo (and other Oceanic languages) that SVCs may be a way-station on a grammaticisation path between clause co-ordination and verb compounding. In subsequent work (Early 1995), he notes that certain postures, specifically *sit*, *stand* and *lie*, which are found in many languages as parts of SVCs, provide aspectual information in Lewo, and may be the basis for the development of fully inflectional aspect morphemes.

The emphasis in this analytic tradition on the discourse function and semantic function of concatenated verb phrases or clauses is helpful in distinguishing like and unlike serialisation strategies. As I have noted, I will be using a conservative definition of SVCs, which is not to say that I do not think the structurally similar patterns of verb concatenation are uninteresting. On the contrary, I believe they will ultimately be part of the whole story about the grammaticisation of verbs to serve semantic functions, but I also believe a more focused examination of SVCs at this point may in the long run help us understand what the actual mechanisms are by which grammaticisation occurs.

7.2 Structural variants of SVCs

The following combinations of TMA auxiliary or inflection on the verbs of SVCs are attested cross-linguistically:

Table 7.1: A structural inventory of SVC languages
(*italics*, languages spoken in Vanuatu²).

TMA auxiliaries or inflection present	Examples
(i) V_1 and V_2	Akan, Ewe, Hindi, Saramaccan, <i>Namakir, Lewo, Baki (Epi), Big Bay</i>
(ii) V_1 but not V_2	Ewe, White Hmong, Sranan, <i>Nguna, Paamese, Lonwolwol, Kuliviu.</i> <i>Tangoa, Tasiriki, Nogugu, Big Bay (?)</i> , Akan, Saramaccan
(iii) V_2 but not V_1	Saramaccan, Akan

As indicated in the second column by the exemplary languages I have provided, languages need not fall exclusively into one category or another. Byrne's (1990) whole point with respect to Borer's (1989) analysis of Saramaccan is that Saramaccan uses all three strategies.

Seuren (1990:14) says that type (ii) is the 'most common' form of SVCs (though he gives no indication of how he measures frequency or commonness). The third type, exemplified only by Saramaccan and Akan as far as I have been able to find out, seems to be the most marginal, that is, it is not selected by a language as the exclusive strategy for forming SVCs. Both Akan and Saramaccan make use of the other two strategies as well. By comparison, it seems that languages may restrict themselves to using only strategy (i) or (ii).

There is another class which does not fit into this schema, but which is represented by a number of languages spoken in Melanesia. These are SVCs where the V_2 has verbal morphology attached to it, but where the inflectional morphology may be limited to subject-verb agreement. Moreover, this is often restricted to a sub-class of the subject-verb agreement paradigm.

As 7.2–7.4 showed, the V_2 in Bislama SVCs is not completely bare of inflectional marking. This rather suggests that Baker's (1989) analysis of SVCs as being sister VPs dominated by one IP will not universally capture the formal properties of SVCs. The V_2 in Bislama seems to be associated with more structure than simply a VP. But not all subject verb inflections, nor all TMA markers, are grammatical in the V_2 . This fact, in conjunction with the interpretation of a SVC as being a single event, suggests that whatever structure is associated with the V_2 , it is not the same as the structure associated with the V_1 .

Collins' (1995) Minimalist analysis of SVCs provides an up-dated architecture for deriving SVCs from syntactic structures in which there is only one tense node or operator. The mechanisms of feature checking inherent to Minimalist syntax can account for SVCs of

² Sources: Akan (Baker 1989; Byrne 1990); Ewe (Collins 1997); Lewo (Early 1993); Lonwolwol (Paton 1971); Hindi (Bhatt 1998); Namakir (Sperlich 1993); Nguna (Ray 1978, Schuetz 1969); Sranan (Sebba 1987); Saramaccan (Borer 1989; Byrne 1990, 1991); Paamese (Crowley 1982, 1987b); White Hmong (Riddle 1990). Baki, Big Bay, Kiliviu, Nogugu, Tangoa, Tasiriki (Ray 1978).

type (i) and type (ii) in Table 7.1 (V_1 and V_2 marking; V_1 but not V_2 marking). Where TMA marking occurs on V_1 but not on V_2 , Collins proposes that the checking of the V feature in TP occurs only at LF. In languages where the V_1 and V_2 of SVCs bear the same TMA marking we must presume that the functional categories appear with the lower verb as a consequence of some kind of downward percolation through co-indexation.

But Collins' account cannot explain the appearance of SVCs of type (iii) (V_2 marked but not V_1), or the kind of SVCs exemplified by Bislama (where the V_2 appears to be a kind of degraded V_1).

Moreover, the substrate Melanesian languages for Bislama show similar patterns of structurally degraded V_2 s. Thus, the Bislama pattern is by no means an idiosyncrasy of one language. After fully describing the Bislama pattern, I will attempt to provide a structural explanation of Bislama SVCs that will generalise across the structural patterns associated with SVCs in other languages.

I want to suggest that the most promising way to attempt to account for the range of structural patterns found in SVCs lies not in parametrisations of Tense *per se*, nor of AGR *per se*, but rather the semantic features associated with them that distinguish between finite and non-finite clauses. I am going to argue that Borer and Hale were right, insofar as there must be a second TP in SVCs, but that the interpretation of SVCs as a single event is a consequence of the lower TP being defective. I will show that for Bislama, at least, there are larger advantages to this analysis. If the lower clause of Bislama SVCs has a defective TP, we can present a unified description of the conditions under which default subject-verb agreement appears throughout the language.

7.3 The structure of Bislama SVCs

7.3.1 Constraints on SVCs in Bislama

We have noted that Bislama SVCs occur with a closed class of verbs in the V_2 position. The V_2 always indicates path or location, the archetypal verbs are *go* 'go; motion away from deictic centre', *kam* 'come; motion towards deictic centre', *stap* 'stay; be (in one place)'. Less common specifications for the manner of *stap* can be found: *le, slip* 'lie; horizontal posture relative to something else', *hang* 'hang; vertical posture relative to something else'. It has been noted that if a language is going to serialise with any verbs, verbs of motions like *go* and *kam* are 'virtually universal' in serialising languages (Sebba 1987:184).

We can only presume that these verbs are unaccusative by analogy with many other languages, in which verbs of motion, like *go* and *kam*, fall into the unaccusative class.

7.3.2 The difficulty of testing for unaccusativity in Bislama

Diagnostics commonly used to establish a homogenous class of unaccusatives (e.g. auxiliary selection) are virtually impossible to apply in Bislama. There are no auxiliary class distinctions analogous to the *essere/habere* distinction in Italian.

It is also hard to test with movement whether the subject of these verbs was base-generated as a complement or specifier. Bislama does not have movement in questions; relative clauses also do not seem to necessarily involve movement. Relative clauses may be formed with the complementiser *we* followed by a tensed clause. The argument or adjunct corresponding to the moved element in English may or may not be present in the *we* clause. Moreover, *we* or

any other overt complementiser is not obligatory. In these cases, it becomes impossible to make a formal distinction between a relative clause with a null complementiser and coordinated clauses with a zero conjunction.

The lack of overt movement makes it very difficult to establish conclusively what the base-generated position of the subject of *go* or *kam* might be in Bislama. That is to say, if there is a class of unaccusatives in Bislama, it must be defined primarily on semantic grounds (cf. Levin & Rappaport Hovav 1995), rather than syntactic ones. It is possible that there simply is no such class in Bislama. Massam and Roberge (1997) argue that Niuean makes no distinction between unaccusative and unergative intransitive verbs and it may be that Bislama is like Niuean in this respect. The point will not be crucial to the discussion or analysis of SVCs in Bislama – for the analysis I will propose, it will not matter if the subject of the second verb is base-generated in the spec of the lower VP, or as the complement of the V head.

7.3.3 Evidence of the dependence of the lower clause

All the verbs that occur as the V_2 of an SVC also continue to be able to be used as main verbs in addition to being the V_2 in SVCs. However, in SVCs they do not introduce a new event into the discourse, i.e. they are not functioning as the main verb in a clause, rather they are dominated by the V_1 of the SVC. This can be shown in a number of ways.

First, we saw in 7.2–7.4, repeated here for convenience, that the V_2 of an SVC always and only occurs with the 3SG subject-verb agreement marker, regardless of what the interpretive subject of the V_2 (subject or object of the V_1) is, and the lower verb can never have a phonetically overt subject of its own (as shown in 7.3b').

- 7.2 *Be wan big ston i ron i kam.*
 'But this huge stone rolls down.' (S-95-9, Lisa)
 [singular subject intransitive V_1]
- a. *Afta nao oli ron i kam.*
 'And so they ran over.' (M-94-2, Susana)
 [non-singular subject intransitive V_1]
- 7.3b. *Afta mi stap wokbaot i go.*
 'And I started to walk off.' (M-94-2, Lolan)
- b. **Mi stap wokabaot mi go* [with single event reading]
- c. *From taem yu pas i go aot...*
 'Because when you left...' (M-95-15, Jackson)
- 7.4a. *Tin biskit ya oli go putum i stap daon longwe.*
 'The biscuit tin was put over that way.' (M-94-2, Leipakoa)
 [singular object transitive V_1]
- b. *Afta oli sakem mi i go long solwota.*
 'So they threw me in the ocean.' (M-94-6, Visi)
- c. *[Yumi] karem ol ting i kam long haos.*
 'We bring everything in to the house.' (M-95-8, Vosale)
 [non-singular object transitive V_1]

There appears to be defective or default subject-verb agreement inflection on the V_2 , rather than the full paradigm of agreement marking found with the V_1 . Examples 7.4b and 7.4c show that the interpretive subject of the V_2 (direct object of V_1) does not cause 1SG or 3PL subject verb agreement to surface as it would if *go* and *kam* were the main verb. This indicates that the structures are not two independent clauses joined by a phonetically null co-ordinator. The only interpretation for 7.3b', for instance, would be that of two conjoined events, i.e. 'I was walking **and** I went', etc.³

It is also clear that the V_2 is not a predication of the entire first clause (Crowley's 1987b 'ambient serialisation'). It is always dependent on the TMA of the first clause (cf. 7.3b above), so that it is always ungrammatical for the V_2 to occur with any of the Bislama TMA markers, as shown in the (constructed) example 7.5:

7.5a. *Ol woman oli no/bin/stap ronem ol devel i go.*
 PL woman AGR NEG/ANT/(CONT/HABIT) chase PL devil AGR go
 'The women don't chase/have chased/chase the devils away'
 (or theoretically with all three: 'The women didn't always chase the devils away')

b. **Ol woman oli (no/bin/stap) ronem ol devel i no/bin/stap go.*

Third, a pronoun in the lower clause can be bound by a referent in the higher clause. Data verified for me by Robert Early with native speakers in Vila indicate that in 7.6 the pronoun may be interpreted as either bound or free.

7.6 *Ol pikinini ya oli ron i go long haos blong hem.*
 PL child SPEC AGR run AGR go PREP house of 3SG
 'The children ran away to their own homes.'
 'The children ran away to his/her house.'

If the bound reading of the lower pronoun is to be accounted for in terms of the Binding Theory (Chomsky 1986) (and it is debatable whether the Binding Theory is relevant for languages like Bislama where there is no formal distinction between free pronouns and anaphors), I take this to show that the PP is an adjunct of *go*, structurally dominated by the higher verb *ron* (a free pronoun reading is also possible, since the adverbial can also adjoin above *ron*).

Fourth, the V_2 is not easily defeasible from the V_1 and cannot be treated as a separate event. Thus, an attempt at gapping as in 7.7 is unacceptable.

7.7 **Olgeta oli ron i kam be hem i go.*
 3PL AGR run AGR come but 3SG AGR go
 'They ran towards us and she Ø away'⁴

³ Bislama co-ordinators are often phonetically null. This is illustrated abundantly in the narratives transcribed in my corpus.

⁴ J.R. Ross drew my attention to the implications of this example. He also helped focus my attention on other evidence of on-going grammaticisation of some verb serialisations in Bislama. As example 7.7 shows, the notion of 'going' is not really defeasible from the whole SVC when used to describe literal movement along a path, and this contrasts with directional adverbs in English where (i) is acceptable:

(i) Anna swam up, and Barbara down

The semantic relationship between V_1 and V_2 in SVCs in Bislama is therefore, more fully grammaticised than the relation between English verbs of motion and adverbs of path or motion. It is

The basic facts of 7.5–7.7 could all be predicted given the general property of SVCs as single events. What is not necessarily predicted is the consistent appearance of the subject-verb agreement marker *i* in the lower clause.

7.3.4 Substrate parallels

As noted earlier, the appearance of subject-verb agreement in the V_2 is not peculiar to Bislama. Similar examples of reduced inflection are found in the SVCs of some of the Melanesian languages that form the Bislama substrate, and I will look now more closely at the patterns shown in these languages. A number of the Vanuatu languages have SVCs where both the higher and the lower verb have full inflectional marking, e.g. Namakir (Sperlich 1993) and Lewo (Early 1993) (both Central Vanuatu languages). Early (1993:71) says the inflection is optional on the lower verb, but his examples in both his 1993 and 1995 papers show full person and number inflection. I will not discuss these SVCs in any detail, since presumably the only substrate influence such patterns will have exerted on the Bislama SVCs is a tendency to express verbal relations of manner or path through serialisation. They have apparently not contributed directly to the syntactic model used by Bislama. On the other hand, the Melanesian languages in which there are SVCs and where the SVCs show the same kind of reduced inflection in the lower verb, merit more detailed examination, since these languages are most likely to have fed both the semantics and the syntax of the Bislama SVCs.

In Port Sandwich (Charpentier 1979) and Sakao (Guy 1974b), the V_2 occurs with 3SG subject-verb agreement marking, as in Bislama, regardless of the person and number of the interpretive subject of the V_2 .

7.8 *U-mandr i paus.*
3DU-start 3SG paddle
'They start to paddle.' (Port Sandwich)

7.9 *Tim-hop ahal o-jan tot.*
INDEF.SBJ-follow road 3SG-go there
'They followed the road that goes there.' (Sakao)

more like the relationship between the adverbs grammaticised as verb particles which makes (ii) and (iii) unacceptable:

- (ii) *Anna got knocked up and Barbara down
- (iii) *Anna ran to and Barbara fro

The process of grammaticisation is seen even more clearly when Bislama SVCs are used to express more metaphorical notions of movement or path, for example:

- (iv) *Ol praes oli stap klaem i go antap.*
PL price AGR CONT climb AGR go on.top
'Prices are rising'

and temporal uses of SVCs, for example:

- (v) *Stat long 1980 i kam, ofis ya hem i muv seven taem...*
start PREP 1980 AGR come office SPEC 3SG AGR move seven time
'Since 1980, this office has moved seven times...' (Yumi Wok 1996:35)

There are emerging instances of SVCs where the relation between *stap* (as V_2) and the matrix V_1 may be somewhat semantically anomalous. Thus, Terry Crowley and I both find (vi) acceptable

- (vi) *Hem i ron(ron)i stap*
3SG AGR run AGR stay
'S/he ran on the spot' (and also 'S/he ran around in a small area' for Crowley)

Jauncey (1995) describes SVCs in Tamambo within Crowley's framework of core, nuclear and ambient serialisation. Like many corpora, her examples mainly consist of 3SG referents (subject and object). Since Bislama, Sakao and Port Sandwich all suggest that a default agreement marker is likely to be the 3SG form, this constraint on the corpora can have important implications. Where the subject and object are both 3SG it is impossible to distinguish between cases of switch subject reference, i.e. true 3SG subject agreement in the V_2 , and cases where this agreement marking occurs as a default.

Jauncey's historical reconstruction of the 3SG subject-verb agreement marker is that it derived from a realis marker used with subjects of all person and number (Jauncey 1997). This suggests that for language-internal reasons the synchronic 3SG marker might well be the default form if a default was required for some reason by the grammar (of course it is frequently the case cross-linguistically that the 3SG is the default form in a paradigm).⁵ Examples from her corpus suggest that in Tamambo, as in Bislama, Sakao and Port Sandwich, the 3SG can appear as a default subject-verb agreement marker on the V_2 of an SVC.

- 7.10 *Ku-te turu a mahere.*
 1SG-NEG stand 3SG straight
 'I didn't stand up straight.' (Tamambo)

Jauncey (1995) analyses (7.10) as an instance of ambient serialisation, i.e. the V_2 is predicated of the event of the V_1 . The closest match to a Bislama SVC in Jauncey's data shows a V_2 that agrees fully with its interpretive subject:

- 7.11 *Mo lai-ra na mai a-imo.*
 3SG take-3PL.OBJ 3PL come LOC-home
 'He brought them home.' (Tamambo)

Jauncey (pers.comm.) says that 3SG agreement on the lower verb in 7.11 is ungrammatical. Her analysis, then, of verb concatenations like 7.10 as ambient serialisations is based primarily on formal grounds, specifically, the contrast between agreement on the lower verb there and the lower verb in 7.11. Sperlich (1993) also gives the impression that the form of agreement morphology found on the lower verb is used to *define* ambient serialisation. Wherever 3SG agreement appears on the lower verb instead of the expected agreement, given the interpretive subject, the SVC is classified as ambient serialisation.

It would, of course, be interesting to ask native speakers whether their intuition is that the act of standing is straight, or whether it is the speaker that is straight in an example like 7.10. When I did seek these intuitions about the Bislama serial verbs, from several speakers (including Sharon Tabi, Alis, who speaks Bislama as her L1, and Nina), they reported

⁵ Jauncey analyses *mo-* as 3SG realis in modern standard Tamambo, but notes that in the speech of some very old speakers there are vestigial indications that *mo-* was a general realis marker as it can co-occur with 3PL subject-verb agreement markers, for example:

- (i) *na-mo-ate*
 3PL-*mo*-sit/stay
 'They sat'
- (ii) *na-mo-ololo*
 3PL-*mo*-be.respectful
 'They showed respect'

In Tamambo, like very many Vanuatu languages, realis is the marked form, and irrealis the unmarked (Jauncey 1997).

(without exception) that (i) separating the two verbs into sub-events is not something they felt comfortable doing, but (ii) to the extent that they were able to do this, the interpretation was not a clausal predication. Clearly, this is an important intuition to elicit from Tamambo (and Namakir) speakers. Until we find out whether they understand sentences like 7.10 to be a predication of the event in the first clause, I feel that we cannot rule out the possibility of a mixed system in Tamambo. That is, Tamambo may have both SVCs with full subject-verb inflection and SVCs with degraded V_2 inflection.

There are regional precedents for this. Other Vanuatu languages show evidence of mixed SVC systems. Atchin (Capell & Layard 1980) shows some SVCs with degraded subject-verb agreement morphology on the V_2 , and some where the agreement is absent. Atchin also compounds verbs (Crowley's [1987b] 'nuclear serialisation'), as many of the Vanuatu languages do, including Bislama. However, I want to treat these compounds or nuclear serialisations as quite distinct from the core serialisations under discussion here. I believe there are sound semantic and syntactic reasons for keeping the analysis of them separate. The verb compounds may express a single event, but often this is in idiosyncratic or unpredictable ways, e.g. Tamambo *tiu + suri* 'leave' + 'follow', meaning 'imitate' (Jauncey 1995). This contrasts with the semantic transparency of the Bislama SVCs.

Moreover, even where the compound is more like the core SVCs and both verbs feed the interpretation, the distinct morphosyntactic properties of core and nuclear serialisations indicate that they are not structurally equivalent. The compounded concatenations of verbs classified as nuclear serialisation are treated by the derivational morphology as single verbs, e.g. Atchin affixes the transitive morpheme after the V_1 - V_2 compound and does not distribute it throughout over the two of them (Capell & Layard 1980:167, text A57):

- 7.12 *Or tuloni luha ni.*
 3DU.INDEF lead return transitive
 'They took them back home.' (Atchin)

In Lewo, the two verbs are treated as a unit for nominalisations (Early 1993:77):

- 7.13 *Na-visa-yu-ena*
 NOM-say-extend-NOM
 'discussion' (Lewo)

These facts suggest to me that in these verb concatenations the V_2 is simply a VP complement of the higher verb, i.e. it would be impossible for the lower verb to have the same richness of functional categories marked as the higher verb does, simply because the structure associated with those functional categories is not available. There is no need to attempt a unified syntactic analysis of nuclear and core serialisations. Naturally, it would be desirable to be able to motivate the lack of structure associated with nuclear serialisations with independent facts. As I have noted, there seems to be a tendency for core serialisations to be semantically transparent and nuclear serialisations to result in less predictable meanings, but this correspondence is not universal, e.g. 7.12.

Capell and Layard, however, also provide examples from Atchin of the kind of SVCs we are interested in. Unlike most Melanesian languages, Atchin marks tense morphologically on the verb, distinguishing past, future and what Capell and Layard call an 'indefinite' (1980:36), interpretable as either past or present. In SVCs, person and number distinctions are obviated in the V_2 , and tense distinctions also disappear in favour of the indefinite tense (Capell & Layard 1980:34, text A38):

- 7.14 *Mwi lasiri iniri e lap.*
 3SG.PAST see 3NON-SG.OBJ 3SG.INDEF many
 ‘He saw they were numerous.’ (Atchin)
 (Capell & Layard (1980:34) say: lit. “He saw them, they are many”)

In Big Nambas (Fox 1979), it is unclear whether SVCs exist as a bona fide verb class. In the few examples Fox gives in which the two verbs would be interpreted as simultaneous events the V_1 and V_2 occur with an overt co-ordinator *ka-* ‘and’ (Fox 1979: 84):

- 7.15a. *N-m’at ka-da-lu.*
 1SG.real-be.feverish and-CONT-vomit
 ‘I was feverish and vomiting.’ (Big Nambas)
- b. *A-r-m’at ka-r-lu.*
 3PL.real-restricted.PL-be.feverish and-restricted.PL-vomit
 ‘They were both feverish and vomiting’

Since a lack of overt co-ordination is often taken to be criterial to defining SVCs, these tokens are marginal. It is possible that the subject prefix should be analysed syntactically as a subject argument cliticised to the verb. If that analysis is correct for Big Nambas, example 7.15 would not be an SVC, it would be a case of co-ordinated clauses with subject ellipsis.

However, if they are examples of SVCs, Big Nambas shows that our theory must be able to generate SVCs in which the V_2 is defective compared with the V_1 in lacking some, but not all, of the subject-verb relations marked inflectionally. In Big Nambas, the first verbal prefix does not occur with the V_2 . This prefix marks subject-verb agreement for person and number and modality (realis, irrealis and conditional) in finite clauses. The morpheme glossed restricted plural on the V_2 in 7.15b is not true subject number marking, because true subject number is restricted to the left edge of the verb phrase. Rather, it should probably be seen as a modifier or restricter on the subject prefix. In 7.15b, the number modifier happens to occur immediately contiguous to the subject-verb agreement marker, but the two morphemes may be separated by aspect morphemes and a negation morpheme, e.g. *-da-* ‘continuative aspect’ in 7.15a could precede *-r-lu* in 7.15b.

Evidence from Lonwolwol (Paton 1971) serialisation is difficult to interpret for several reasons. Firstly, Paton gives two forms of subject-verb agreement for 3SG: a zero (1971:17) and *mV-* (*-V-* indicates an ‘adaptable’ [i.e. harmonic] vowel, 1971:68). Furthermore, he gives examples of serialisation involving the same V_2 where it is sometimes inflected for person, number and tense, and sometimes is not:

- 7.16a. *ɔr me ren ɲa ru.*
 the place is.light there remain
 ‘It remains twilight.’ (Lonwolwol, Paton 1971:55)
- b. *tesigɔ... dɔ fan gate muru.*
 Tesigo... 3SG.sit down upon.it 3SG.remain
 ‘Tesigo...remains sitting on it.’ (Lonwolwol, Paton 1971:55)

It is clear in 7.16b that the verb *-ru* ‘remain’ has been inflected for third singular present tense subject, but at the same time *dɔ* ‘sit’ shows null 3SG agreement. In 7.16a, and in 7.17 – these examples provide the most direct parallels with the Bislama forms I am interested in – the V_2 occurs without overt agreement, like *dɔ*:

7.17a. *namdemelɔlɛ ca su faloh van.*
 1SG-PRES-wish COMP 1TR.INCL paddle go
 'I wish that we-few go in the canoe.' (lit. 'paddle go')
 (Lonwolwol, Paton 1971:98)⁶

b. *o aru barbar mɛ.*
 2SG bring pig come
 'Bring the pig here.' (Lonwolwol, Paton 1971:81)

c. *o aru sisɛ gɔli van.*
 2SG take this thing go
 'Take this thing away.' (Lonwolwol, Paton 1971:81)

Because of the confusion with respect to the form of 3SG agreement, it is impossible to decide for certain whether the forms in 7.16a and 7.17 are examples of Lonwolwol SVCs where the V_2 has zero 3SG subject-verb agreement inflection or whether they are examples of compounding, where the V_1 selects a bare VP as its complement.

In Namakir, it also seems that SVCs may occur either with full subject-verb agreement and functional inflections such as the intransitive marker on the lower as well as the higher verb (7.18a). There are also cases where the lower verb occurs unexpectedly with 3SG agreement (7.18b) (Sperlich 1993:104, 105, his examples 15 and 19):

7.18a. *Te ni betoŋ a-dah bikiak-ih ko pa row na-malal.*
 NEG 1SG tell 3SG-come to.2SG-NEG 2SG INTR go art-garden
 'I didn't call out to you to go to the garden.' (Namakir)

b. *Ko rah a-dah.*
 2SG IMP.come 3SG-come
 'Come here!'

Namakir will obviously reward more in depth examination. If 7.18a is indeed comparable to the Bislama SVCs, it is curious that the lower verb is not independently negated. The scope of the (discontinuous) negative marker extends over both verbs, and seems to treat them as a V-V compound. Yet the lower V appears to have some derivational and inflectional morphology associated with it (intransitive marker and 2SG, respectively). Unfortunately, since we do not know exactly what the status of *ko* (2SG) is, i.e. whether it is a free pronoun, clitic or 2SG agreement morpheme, we are not in a position to resolve the structural relationship between negation and the two verbs.

Ross et al. (forthcoming) distinguish the SVCs of the north and central Vanuatu languages from the echo- versus switch-subject morphosyntax developed in southern Vanuatu languages. However, there are some similarities between the SVC and the echo-subject systems. In Sye, Crowley (1998) notes that the echo-subject prefixes mark only a reduced set of the referential distinctions made with main verb subjects:

7.19a. *Yococ-tapmi m-aruvo.*
 1SG-tried echo.subject.SG-sing
 'I tried to sing.' (Sye)

⁶ Paton uses 'c' to represent the palato-alveolar affricate /dʒ/. I provide Paton's translation in 7.17a, the word for word gloss is drawn from various places in his grammar. Examples 7.17b-c follow Paton in gloss and translation.

- b. *Komli-tapmi ml-aruvo.*
 1PL.INCL-tried echo.SBJ.PL-sing
 'We (INCL.) tried to sing.'
- c. *Kamli-tapmi ml-aruvo.*
 1PL.EXCL-tried echo.SBJ.PL-sing
 'We (EXCL.) tried to sing.'

The examples show that echo-subject verb marking only distinguishes number; the prefix is the same for all persons.⁷ One might argue that 'try' and its complement do not satisfy the requirement that the two verbs express one event and therefore are not SVCs. Indeed, I am not trying to argue that these constructions are types of SVCs, the point of the Sye data is merely that there is some evidence indicating that a language may use serialisation or echo- versus switch-subject marking, but in both strategies the lower verb is often marked with reduced morphology. The extent to which this generalisation holds across other echo-subject marking systems would be an interesting avenue of further research.

To summarise the data presented in this section: Bislama and a number of the substrate Melanesian languages indicate that the possible serialisation patterns given in Table 7.1 need to be expanded in order to account for SVCs where the V_2 has some of the inflection characteristic of the matrix V_1 , but where the range of inflection is degraded. In Bislama, and other Melanesian languages with this degraded inflection, it tends to surface as 3SG. This is consonant with the observation that, cross-linguistically, 3SG is often the least marked or default form in the verb paradigm.

In Bislama, the dependent nature of the lower clause is manifest in a number of syntactic constraints, including a prohibition on having an overt subject pronoun. This restriction and the defective subject-verb agreement on the V_2 in Bislama and some of the languages forming its substrate show that the term 'switch subject serialisations' used to describe some of these SVCs rather overstates the SVCs' syntax. Switch subject may be the correct interpretive characterisation, but it predicts a richer structure than is actually observed.

How then should we amend the typology of SVCs shown in Table 7.1? One possibility would be to add the presence or absence of subject-verb agreement on V_1 and V_2 to the typology. This might be descriptively adequate, but I believe it would add an unnecessary degree of complexity to the typology. I believe a more economical way of amending the typology of SVCs is to redraft Table 7.1 in terms of the finiteness of V_1 and V_2 . This requires a definition of what it means for a clause to be finite or non-finite, and this is what I will turn to next.

⁷ Subject-verb agreement on main verbs also marks a rich array of tense distinctions in Sye; echo-subject marking fails to make such temporal distinctions too. For example (omitting some of the detail of Crowley's interlinear glosses:

- (i) *Kamli-touri ovon kuri mli-tantvinu.*
 1PL.EXCL.DIST.PAST-leadPL dog echo.SBJ.PL-cross river
 'We led the dogs and crossed the river.' (Crowley 1998:248)
- (ii) *Kokl-anwi ml-agku hogku hai narisac.*
 1PL.INCL.FUT-say echo.SBJ.PL-say SIM INDEF great.leader
 'We will say he was like a great leader.' (Crowley 1998:256)

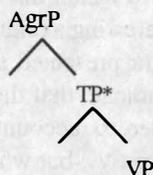
7.4 Defining SVCs in terms of finiteness

Let us return to the central property of SVCs. They are concatenations of verbs with a simultaneous event interpretation. This has been taken to indicate that the V_2 does not have its own TP. But if Bislama and its substrate languages have some form of subject-agreement morphology, this suggests that there is an AgrP associated with the V_2 of an SVC. This suggests two hypotheses: (i) the AgrP of V_2 immediately dominates a VP, i.e. there are no intervening functional projections corresponding to TP or AuxP, and NegP (7.20a), or (ii) the AgrP of the V_2 of an SVC is associated with the same functional projections as any independent clause (7.20b), but that the functional projections dominating the V_2 are in some way defective. A description of how they are defective will explain why the V_2 does not occur with the full range of verbal inflections or auxiliaries.

7.20a. ...



7.20b. ...



*In some way defective

I will not initially pursue the first hypothesis, as it promises to have some undesirable theoretical implications. While there are a number of structures which require us to posit reduced functional projections above a verb (e.g. gapping, co-ordination), these usually observe the order of a full clause. A hypothesis that SVCs have an AgrP immediately dominating (only) a VP requires structural ellipsis and a powerful notion of reconstruction. Such a theory of reconstruction is the less straight-forward of the two hypotheses proposed.

I will, therefore, examine the second hypothesis as the source of a possible solution to the problem of how to analyse SVCs. I assume that a solution making use of existing aspects of the syntax and semantics of natural languages is not only theoretically more desirable, but will also be less costly to the speakers of a language. I will suggest, then, that the correct analysis of SVCs cross-linguistically is whether or not a finite clause in the V_1 only, or the V_1 and the V_2 , is grammatical.

Tense and finiteness are not completely independent measures and sometimes co-occur in the form of a portmanteau morpheme, but it is possible and important, I believe, to distinguish between the two. Much of the literature conflates the two terms or uses them interchangeably. So, in a great deal of the generative treatments of language acquisition, research discusses children's acquisition of 'tense' and 'finite' morphology and the word order associated with tensed or finite clauses, as if the two were synonyms (e.g. Wexler 1994; Haegeman 1995). Discussions of the verb second phenomenon in the Germanic languages also conflates the terms (e.g. Holmberg & Platzack 1990; Schwartz & Vikner 1996),⁸ as do discussions of negation in Romance (e.g. Zanuttini 1996), and head movement of verbs (e.g. Pollock 1989). Chomsky (1995) alternates between labelling the same functional projection TP and FP when outlining a minimalist syntax.

⁸ Although some analyses of verb second in Germanic associate the finiteness feature with the head of Comp, rather than the head of Tense, or INFL, the nature of the clauses referred to as finite and/or tensed are the same.

Treating tense and finiteness effectively as the same thing is understandable with Indo-European languages, which make past/non-past distinctions (and often a great many more tense distinctions than that) obligatorily. For many European languages finiteness and tense may even be a port-manteau morpheme (Holmberg and Platzack (1990) suggest that finiteness is a feature associated with the head of Tense). But even in European languages there is not perfect isomorphism between the two notions; infinitives are treated in much of the work mentioned above as lacking tense, but of course we distinguish 'to cook' from 'to have cooked'.

The connection between the two becomes tenuous though when we turn our attention both to Oceanic languages, and to creoles as a class of languages. In many of the Vanuatu languages discussed in §7.3.4. no distinction is made between past and non-past tense. Where languages do make tense distinctions, we seem to also find an indefinite tense (ambiguous between past and present interpretations). All the evidence points therefore, to the conclusion that, for the languages of this region, the distinction between realis and irrealis is more a part of the core grammar than a distinction between past and non-past tense. The influence of this substrate norm is apparent in the grammar of Bislama. The distinction between realis and irrealis clauses is regularly marked in Bislama, though not obligatorily. That is, Bislama usually indicates irrealis contexts with the sentential adverb *bae*.⁹

The absence of tense distinctions in the grammar of the languages is by no means an areal peculiarity. Surveys of pidgin and creole grammar (Holm 1988; Bickerton 1990) note that the unmarked verb form in these languages may also have past or non-past interpretation, and adverbs or discourse context allow speakers to disambiguate utterances. Although a language may use the unmarked verb form for both past and non-past events, it will nonetheless be possible to specify temporal reference when necessary. For instance, an aspect marker or auxiliary can occur with the verb. This is sometimes referred to in the creole literature as anterior tense marking (Holm 1988). This functions to push the event to some point before the current temporal reference point, i.e. past, if the current reference point is present, or past-before-past (pluperfect), if the current reference point is past. This shows that the anterior marker is not a tense as such, but rather an aspect marker. These auxiliaries or inflections that serve this function in some of the Vanuatu languages discussed above are glossed as the perfect (e.g. Port Sandwich, *inongg* 'elle traduit donc un parfait' [Charpentier 1979:104], Lonwolwol, *bur* 'a 'perfect sense' [Paton 1971:51]).

Clearly, then, a cross-linguistic definition of or identification of finiteness cannot be piggy-backed on tense. Instead, we have to take a step back and ask what the intuitions are underlying finiteness, and how these apply to Bislama and its substrate languages. At a very basic level, finiteness seems to be a property characteristic of an independent clause. Sebba (1987) outlines the fundamental problem of analysing SVCs in Sranan as being a question of how to deal with 'clauses which apparently contain two finite verbs' (Sebba 1987:1). It becomes clear that what Sebba means is that both verbs seem to be associated with enough structure that the SVCs could be carved up into two grammatical independent clauses. As we have seen, this is not a property common to all SVCs (even in Sranan as Sebba goes on to discuss), but I believe Sebba was on the right track in investigating SVCs in terms of finiteness. I think we can derive some important properties of SVCs by typologising them in this way, and moreover that we can do so in a way that is not inconsistent with other discussions of non-finite clauses.

⁹ Meyerhoff (1996) found an unexpected correlation between transitive marking and irrealis clauses, reinforcing the notion that the realis-irrealis distinction is a major feature of the Bislama grammar.

7.4.1 Finiteness as a temporal relation

What, then, is it for a clause to be finite if it is not simply that it is specified for tense? I believe the answer lies in Reichenbach's (1947) now standard formalisation of tense as a set of relationships between the described event (E), the time of speech (S) and the point of reference (R).

I would like to suggest that what makes a clause finite is that the three-way relation between E, R, and S is established for interlocutors. However, different languages rely crucially on different mechanisms for establishing this three-way relation, and this difference is what surfaces morphosyntactically as either a past/non-past system or an anteriority system.

I hypothesise that in systems that grammaticalise a past/non-past distinction, like English, the relationship that it is crucial to establish is the relationship between E and S. This follows from the fact that what distinguishes the simple past from the simple present in such systems is only the relationship between E and S. In the simple past, E precedes S ($E < S$), while in the simple present, E falls at the same point as S ($E = S$).

Hornstein (1990) describes past/non-past systems slightly differently, emphasising the critical role R plays in our understanding of the temporal relations of a clause. Hornstein (1990:109-110) argues that the relationship between E and S is derived transitively from knowing the relationship between E and R, and knowing the relationship between R and S. Nothing in what I am proposing necessarily runs counter to this. In emphasising the importance of knowing the E, S relations in past/non-past systems, I am not dismissing R. In fact, my analysis of non-finite clauses will make use of the pivotal role Hornstein's system gives to R.

On the other hand, in systems where a past/non-past distinction is not grammaticised, and temporal ordering of events is specified more with respect to anteriority or perfect aspect, as it is in Bislama, it is crucial to establish the relationship between E and R, not the relationship between E and S. S is specified by interclausal (discourse) relations or by adverbs, but it is not grammaticised. The Bislama system follows straightforwardly then. The simple past and the simple present are isomorphic. In both cases, E and R are the same. Sentence-level operators determine S. Thus, an utterance will be interpreted as present in one context if S is specified as being the same as E and R; the same utterance will be interpreted as past in another context if S is specified as being after E and R. An auxiliary marking anteriority functions exactly like the perfect, separating E and R, so that E precedes R.

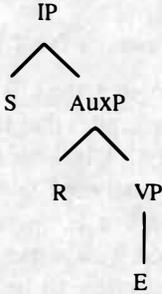
The difference between past/non-past systems and anteriority marking systems is summarised in Table 7.2.

Table 7.2: Requirements for a clause to be finite in two temporal marking systems (*derived from relation between R and E, R and S, Hornstein 1990).

System type	E, R, S relation required	
(i) past/non-past	$E \leq S^*$	
	simple past $E < S$	simple present $E = S$
(ii) anterior/unmarked	$E \leq R$	
	anterior $E < R$	unmarked $E = R$

If I am correct that what makes a clause finite is knowing the relationship between E and R in a Bislama-like system, and knowing the relationship between E and S in an English-like system, then it follows that some kind of disruption in the calculation of these relationships would make a clause non-finite. I assume Hornstein is basically correct in proposing that the E, R and S points of a finite clause have morphological reflexes. I take his proposal to amount to a mapping of E, R and S to VP, AuxP/TP and IP respectively as shown in example 7.21 (cf. his discussion 1990:109-110).

7.21 ...



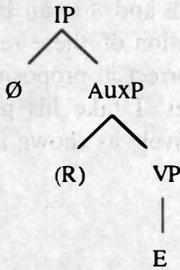
As long as there is a verb, E will be able to be determined, so it is hard to see how this could be the source of the disruption. I have already argued that S in languages like Bislama is not specified by the verb or the functional projections associated with it. Instead, I have argued that it is provided by relations between clauses and/or relations between the verb and adverbs. So it seems that an underspecification of S in the VP (and its associated functional projections) can not be the factor that cross-linguistically distinguishes non-finite from finite clauses.¹⁰

This leaves R, then, as a possible way of differentiating finite and non-finite clauses. I have distinguished the two verbal systems by suggesting that past/non-past systems rely on fixing an E, S relation, and the anteriority systems rely on fixing an E, R relation. But remember that Hornstein suggested R was the medium by which the E, S relation is established. Thus, a disruption in fixing R will prevent an anterior marking system from creating a finite clause because a coherent E, R relation will not be able to be established. And a disruption in fixing R will also prevent a clause from being finite in a past/non-past system, since an inability to fix the E, R relation and the R, S relation will make it impossible to establish the E, S relation. Consequently, I propose that what distinguishes finite from non-finite clauses cross-linguistically is that non-finite clauses lack a specification for R.

This differs from Hornstein's treatment of non-finite clauses. He suggests that infinitives lack a specification for S, not R (Hornstein 1990:147, 156). Hornstein distinguishes two types of non-finite verbs, infinitives and what he calls 'naked infinitives' (e.g. the second verb in 'Kirsty made Jeremy leave'). He suggests that the latter lack both a specification for S and R (1990:155-156).

¹⁰ If it was, then my analysis of anteriority marking systems would entail that all clauses in these languages are non-finite, and this is clearly not my desire nor my intention.

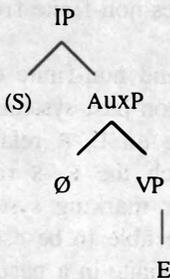
7.22 ...



[Hornstein's analysis of infinitives. No specification for S, sometimes none for R.]

This seems an odd claim to make since it amounts to a claim that there is no time of speech, yet obviously we can utter an infinitive. However, Hornstein argues that there is no S, and the R point of an infinitive is anchored to the matrix clause E. I believe this analysis misses an important generalisation. If R in an infinitive must be anchored outside its own clause, then this would seem to imply that the R point of an infinitive is, in and of itself, underspecified. Naturally, in any clause an R point may be anchored with respect to the E or R point of a preceding clause, but Hornstein wants the R of an infinitive to be obligatorily anchored to the matrix. In other words, our accounts of the role played by S in non-finite clauses are fundamentally different, but our accounts of the role R plays in non-finite clauses agree in spirit, though they differ in letter. We can proceed assuming that the underspecification (or lack of specification) of R is definitional of a non-finite clause.

7.23 ...



[Alternative analysis of non-finite clauses. Lack of, or underspecification of, R.]

There are also good empirical reasons for assuming Hornstein is mistaken about the specification of S in non-finite clauses. If he is correct, and the S specification has a fairly direct structural mapping onto AgrP, then the presence of any person or number agreement on non-finite verbs is problematic. Raposo (1987) shows that even some European languages have the option of person and number agreement morphology with infinitives. And Crowley's Sye data (examples 7.19a–c) show verbal complements of 'try', which it seems reasonable to assume are non-finite forms, appearing with subject referencing morphology, which I assume to be the surface reflex of an AgrP.

In some languages, I assume that there is also a difference between finite and non-finite clauses with respect to the ability of AgrP to attract a subject. If, in some languages, a further definitional property of what it is to be non-finite is an inability of AgrP to attract a nominal element to its specifier we will not see any subject-verb agreement morphology in non-finite forms. English is one such language. However, in other languages, if the impoverished

specification for R is the only way in which a non-finite clause differs from a finite clause, we would expect to find non-finite verbs in these languages bearing subject agreement morphology. Assuming Hornstein's basic mapping of R to the TP (or AuxP) is correct, an underspecification or lack of specification for R would amount to a defective TP. The AgrP of a clause dominating a defective TP might, nonetheless, have all the features associated with AgrP in a finite clause. I would like to suggest that the V_2 in Bislama SVCs have an AgrP, no different from that of the V_1 .

If we adopt this analysis of Bislama SVCs, we can propose a structure associated with the V_2 of an SVC which generalises cross-linguistically. We are able to generate the first two patterns in Table 7.1 (V_1 and V_2 , V_1 but not V_2) and also the patterns found in the Melanesian languages, including Bislama. It allows us to clearly distinguish Bislama SVCs from clauses that are linked by zero co-ordination. In the latter, two events are related to fully specified R and S points, that is, they have non-defective TP and AgrP.

It also allows us to discriminate SVCs from gapping, in which the second verb phrase is underspecified for E, for example, the constructed example 7.24, where the second clause independently sets S and R, but the E is dependent on reconstructing the E from the preceding clause.

- 7.24 *Sera i save kakae mit, Emi i no save.*
 Sera AGR HABIT eat meat Emi AGR NEG HABIT
 'Sarah eats meat, but Emmy is vegetarian.'

Moreover, in gapping, a coreferential subject in the V_2 is infelicitous (either as a zero or an overt pronoun), as indicated in 7.25 (also constructed).

- 7.25 **Sera, i bin wok long Kanal fastaem, naoia (hem/Ø) i no.*

The proposal that the lower clause of an SVC has a defective TP does not seem to predict the third SVC pattern shown in Table 7.1 (V_2 but not V_1). This requires an analysis under which the full TMA marking on the V_1 is either deleted late in the derivation, or the V_1 does not check all the features associated with its functional projections. If we assume the E point in both V_1 and V_2 are the same in SVCs, then the heads of the two verb phrases would be coindexed. Where the heads are coindexed, it is possible that a language might allow the relations between S, R and E to percolate from V_2 to V_1 . In that case actual movement in the V_1 to establish these relations would be redundant. Even if these temporal features do not require movement to unify, this proposal leaves open the question of why the other features we must assume are associated with the functional projections and verb of the higher clause do not require checking.

I noted at the outset that the V_2 but not V_1 strategy is rather marginal, in that (as far as we know) no language makes use of only this pattern for serialising, it always co-varies with V_1 and V_2 and V_1 but not V_2 . Byrne (1990, 1991), who discusses this pattern for SVCs most extensively, does not give enough data to determine whether this pattern is constrained in some way by interclausal dependencies. We can only assume that if there is no semantic basis for choosing between the three syntactic options, there are constraints operating at some other level of the grammar (e.g. information structure, diachronic change).

In languages where the V_1 and the V_2 both have full TMA marking and a simultaneous event interpretation I assume that something like the following process of interpretation and morphosyntactic spell-out occurs. Assuming the R for the V_2 in these languages is underspecified as in Bislama SVCs, and assuming the relation between E, R and S in the lower clause is established by co-indexation with the E, R and S relations established independently

for the higher clause, then there is nothing to stop a language from spelling out this (anaphorically derived) relation in the V_2 . A language's choice between the options of V_1 but not V_2 and V_1 and V_2 might be seen as two equally functional resolutions of competing motivations: one to avoid redundancy; the other to provide a surface realisation of all morphemes.

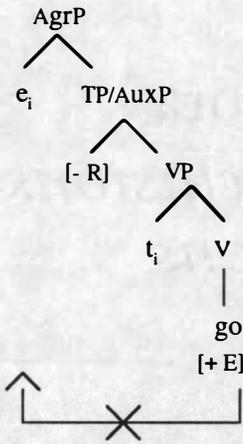
Again, the theoretical problem is how to explain a system using more than one of the possible patterns to construct SVCs, and again we are hampered by the fact that descriptions of SVCs tend to discuss the morphosyntactic and/or pronominal binding facts more thoroughly than the discourse constraints that may be interacting with the variation.

If the V_2 of SVCs are non-finite clauses, and if (*ex hypothesi*) these are distinguished from finite clauses by a lack of or underspecification for R, then we expect the V_2 to have all the functional projections of a normal clause, but the features associated with the functional projections will not be the same as those associated with a finite clause. Differences in which semantic or tense features are underspecified will account for cross-linguistic differences in the form of SVCs. In all languages, R will be underspecified, and in some languages, non-finite AgrP will differ from finite AgrP in that the nominal feature associated with it may be weak (instead of strong as in a finite clause). I am using 'weak' in the sense of Chomsky (1995), that is to say, a language with a weak nominal feature associated with Agr will not attract a nominal argument before Spell-out, whereas a language with a strong feature will. The theoretical apparatus of this is, however, irrelevant to the core of my analysis.

In Bislama, it seems that the non-finite AgrP in the V_2 of SVCs is essentially the same as it is in finite clauses, i.e. a strong nominal feature in the specifier attracts the subject of the V_2 causing it to move out of the specifier of VP. A specification for R, however, is lacking. Since Bislama is an anteriority marking system and not a past/non-past system the temporal relationship crucial in a finite clause is the relationship between E and R. Thus, an underspecification for R will have structural consequences. Without a specification for R, determining the E, R relationship will be impossible. Under this circumstance, we may suppose that movement of the lower V (the locus of the specification of E in the lower clause) to adjoin to the head of TP (the locus of R, when specified) is illicit. Thus, the V_2 cannot leave the specifier of the lower VP.

In Chapter 5, I determined that the subject-verb agreement is such that when the subject is 1SG, 2SG, 1PL(INCL.) or is the pronoun *olgeta*, specific forms of agreement appear on the verb (zero, and *oli* with *olgeta*). Elsewhere, therefore, the agreement was *i*. Since subject-verb agreement occurs when the subject and verb enter into a specifier-head relation, we might predict that default agreement will appear in the lower clause of SVCs. As 7.26 shows, if the verb in the lower clause of a SVC is prevented from raising out of the VP, but the subject nonetheless raises to check the nominal feature in the specifier of the lower AgrP, subject and verb in the lower clause will not be in a spec-head relation.

7.26 ...



Therefore, the only agreement morphology which will be able to surface is the default, *i*. Thus, we have a principled account of why the default agreement morphology appears in the non-finite clauses of SVCs.

It is not entirely clear (within a principles and parameters model) what the status of the subject of the lower clause is, since in SVCs with the form $N_1 V_1 V_2$, there is obligatory co-indexing of the subject of V_1 and V_2 . Yet in SVCs with the form $N_1 V_2 N_2 V_2$, the subject of V_2 cannot be interpreted as N_1 . I shall leave it notated as simply an empty category (*e*) for the time being.

The subject of the lower clause in a SVC in Bislama seems to have some properties generally associated with *pro* (the phonetically null subjects found in main clauses discussed in the last chapter). Assuming my analysis of Bislama SVCs is correct, the empty category in this clause is, for instance, governed by Agr. But it also seems to possess some of the characteristics of PRO in control constructions. It does not co-vary with an overt pronoun and it is obligatorily controlled by the next highest nominal (phonetically null or overt object of the higher clause if transitive, or the phonetically null or overt subject if intransitive).

This set of mixed properties was, of course, one of the reasons Borer (1989) felt motivated to collapse the PRO/*pro* distinction in favour of one empty pronominal category which she proposed was coindexed with an anaphoric Agr. Data collected in the field does not, unfortunately, provide sufficient detail on this; a precise investigation of the nature of this empty category, and the nature of the control relations between nominals in the higher clause and the subject of the lower clause, will have to stand as an avenue for future research.

8 Toktok biae; tumora blong wok ya: *conclusions and future directions*

In early chapters, I devoted quite a lot of space to a thorough description of the speech communities in which I conducted my fieldwork. By giving an idea of the historical and contemporary setting for Bislama and the culture of the users, I intended to give some idea of the role Bislama plays as a medium of communication in Vanuatu, both nationally and at the very local levels. My methods of data collection, particularly my decision to undertake a reasonably lengthy period of fieldwork, were predicated on the conviction that a high level of ethnographic detail is required for an informed investigation of language variation. If we are to be in a position to learn the most from the patterns of language variation we may find in a speech community it is essential that our research be structured so that we are sure that the social and intergroup categories we use are those that are actually relevant to the community we are investigating.

That many of these categories are cross-culturally salient indirectly tells us something about the nature of most human societies, but in all speech communities there are certain to be groups defined in terms specific to their situation. In this study, four subgroups of the speakers recorded on Malo were defined based on factors uniquely relevant to the Malo community. Observations of speakers' use of *lanwis* and their use of Bislama, as well as individuals' reports about their attitudes to Bislama and its local and supra-local functions as a medium of communication were used to delimit four groups of adult speakers.

I was subsequently able to use these groupings to show that where the youngest speakers of Bislama in the Malo speech community differed from the general community norms, this was influenced by very local models. We were able to see that the children did not, for instance, acquire some highly local variety of Tamambo-influenced Bislama, since the pattern of variation observed in the children was not parallel to the patterns of the people I called 'the Malo hard core' (Group A). We saw that the children were matching, but in an exaggerated way, their immediate caregivers' patterns.

Thus, the study of language variation in Bislama was informed by properties of the social structure of one of the speech communities investigated. This also assisted us in interpreting the implications of the patterns of variation observed for linguistic theory, and the structure of Bislama.

The corpus of Bislama which provided the data for this work was drawn entirely from conversational interactions, a number of which I was not an active participant in. The use of a large body of discourse data proved to be helpful in the analysis of structural aspects of contemporary Bislama. The first question that this data was able to help resolve empirically was a dispute in the literature over the proper analysis of subject pronouns and subject-verb agreement morphology on verbs.

By comparing the distribution in discourse of 1SG and 2SG subjects with clearly focused forms of 3SG subjects and with clauses in which a 3SG subject is realised as a null pronoun, we saw that the distribution of doubled sequences of *mi mi* and *yu yu* are most like focused 3SG subjects, *NP hem i*. I concluded that there is a full paradigm of subject-verb agreement morphology in Bislama, and that the morphemes are realised as *oli* with 3PL subjects, zero with 1SG, 1PL(INCL.) and 2SG subjects, and *i* elsewhere.

With this issue resolved, I was able to undertake an analysis of the variation between phonetically null and pronoun subjects with main verbs in finite clauses. Previous work on Bislama had mentioned the possibility of null subjects only in passing, if at all. Consequently, this work has provided the first empirical data on how widespread this phenomenon is in Bislama (44% of finite clauses have null subjects), and the linguistic constraints affecting its distribution.

A multivariate analysis of the distribution of phonetically null subjects indicated that the linguistic variables that had the most significant interaction with the variation were the person and number of the subject, and the grammatical role and form of the referent in the immediately preceding clause.

We saw that Bislama is actually a split *pro*-drop system, where the split is between the different persons. First and second person subjects tend to be realised with pronouns, while third person subjects are usually null. I pointed out that this seems to be incompatible with Crowley's (1990) earlier suggestion that subject deletion is licensed when the subject referent is animate (possibly under the influence of the Melanesian substrate, he suggested). Since first and second person are perforce animate, this suggests that the salient constraints are to be found elsewhere.

Previous discussions of *pro*-drop systems split between the different persons are extremely limited, and some of the better analysed split systems go in the opposite direction – third person subjects require pronouns, while first and second person subjects allow null subjects. On the basis of these languages, some researchers had claimed that split *pro*-drop by the person of the referent is motivated by the higher referential or cognitive salience of first and second person. The Bislama data indicates that there are no hard and fast functional or cognitive principles that constrain the way in which a *pro*-drop system may split. If there is a cognitive dimension to split *pro*-drop systems, it is a tendency, not a maxim.

However, the morphology of Bislama does seem to play an important role in determining the parameters of *pro*-drop, as previous generative work on *pro*-drop would have predicted. The Bislama split has a positive correlation with the persons in which the subject-verb agreement has the greatest referential transparency. In other words, as is true of other split *pro*-drop systems attested, the interpretive richness of the verb morphology is a much more important factor in constraining null subjects than anything to do with the (hypothesised) real-world salience of different referents.

The other significant constraint on phonetically null subjects in Bislama was also a linguistic factor, i.e. the interclausal relationship between subject referents and their surface form. A chain between subjects in two consecutive clauses favoured the use of a null subject, but only when the referent had already been realised as a null subject. This finding was

somewhat thought-provoking, since it means that there is both a highly functional constraint on the use of null subjects in Bislama (the referential transparency of the verb morphology), but also a non-functional constraint (the surface form of the prior subject, when it is coreferential).

By examining the distribution of phonetically null subjects in a number of social groups, I was able to show that *pro*-drop is clearly part of a community grammar. There were no correlations with the social subgroupings that had earlier been shown to be salient in the Bislama speech community.

We did find some indication that the patterns observed are part of an on-going process of grammaticisation. The youngest speakers in the corpus essentially followed their immediate caregivers' patterns for *pro*-drop, but exaggerated the effect of the most important constraint. That is, in speakers under 12 years, the distinction between use of pronouns with first and second person subjects, and null third person subjects, showed signs of approaching categoricity.

In the final chapter, I returned to the question of when and where default subject-verb agreement occurs, looking at the obligatory use of null subjects and *i* in the lower clause of serial verb constructions. I argued that the invariant absence of a subject with the second verb of an SVC, and the invariant use of *i* on the verb can be explained if the lower verb in a Bislama SVC is non-finite.

I examined the temporal properties associated with finite clauses both in past/non-past systems and in realis/irrealis systems. I suggested that although both systems mark rather different temporal relations in the surface grammar of finite clauses, underlyingly both systems depend on an unambiguous specification of the point of temporal reference (R). I suggested, therefore, that an underspecification of R is what makes a clause non-finite in both past/non-past and realis/irrealis languages. Assuming that the dependent verb in an SVC is underspecified in this way, and assuming that this semantic underspecification maps into the syntax, the lower verb in an SVC will be prevented from entering into the structural relationship with its interpretive subject that is necessary to trigger the full range of subject-verb agreement morphology. Only the default form *i* may surface.

What then are some of the implications of this work and where might it lead us next?

It would be fascinating to examine the interpretation of null arguments elsewhere in Bislama, to see whether in all cases the primary constraint is the extent of overt morphological information that can be retrieved from the clause. Direct objects and objects of prepositions are optionally null in Bislama just as subjects are. In the case of obliques, it is clear that there is some interaction with the morphology (or apparent morphology) of the preposition. Those which appear to have the transitive suffix allow deletion (under circumstances still to be fully determined), those which do not end in [m], do not. The constraints on argument deletion are even more opaque. Presence of the transitive suffix on the verb seems to be a factor, but this by no means accounts for it all. As with null subjects, it seems likely that object deletion interacts with several factors, some morphosyntactic, and some related to information status.

It would also be worthwhile to investigate the role other markers of specificity have to play in the interpretation and deletability of arguments in Bislama. The distribution of the determiner *ya* and demonstratives were not examined in this study. Clearly, however, they are an important aspect of the interclausal environment, and they may well interact with the interclausal priming effect on null subjects that was observed.

I think the nature and interpretation of the default subject-verb agreement is a clear area in which more work could be undertaken. It would be interesting to examine further the

extent to which the structures of SVCs have become fully grammaticised. For example, it would be interesting to examine the experimental effects of altering the agreement marking on the lower verb. If the structure has not fully grammaticised, we should find respondents readily reanalyse such permutations, while if they are rejected as ungrammatical, this would be an indication that they have been reanalysed, perhaps as what are known as converbs in some of the literature (e.g. Haspelmath & König, eds 1995).

As I noted in the discussion in Chapter 5, the conclusions drawn about the nature of pronouns and agreement morphology in 1SG and 2SG were based on some subtle statistical differences in distribution. They may not, therefore, constitute the last word on the matter. Further work here would be worthwhile, particularly on the use of pronouns in secondary predications. This in turn, of course, relates to the structural and semantic analysis of SVCs.

Finally, the role of younger speakers in the continued grammaticisation of Bislama deserves further examination. In my discussion of the results in Chapter 6, I discussed the distribution of null subjects in the speech of under 12s as matching, but exaggerating the speech of their caregivers. But I also noted that since there is a lack of general age-grading in the community, this suggests that successive generations' patterns of grammaticisation feed into and alter the grammar of their elders. Thus, further examination of the interaction and the grammatical parallelisms between younger speakers and older speakers may tell us about much more than processes of grammaticisation in language acquisition. It may provide us with information on the diffusion of grammatical patterns throughout the speech community.

Appendix

Glossary of Bislama terms used

- bringanbae* A 'bring and buy' is a way of raising cash for a community group. Families gather together in a central place, bringing food they have prepared at home and which they then buy from each other for a small amount of money. For example in 1994–95 c.VT20 for a piece of *laplap* (cooked savoury pudding), a skewer of meat or a piece of fish (VT20 was equivalent to US\$0.20, or a little less than half the cost of a loaf of bread in town). The money is donated to whatever the occasion is in aid of, e.g. a school, the church, a sports team. A *bringanbae* would usually last all evening and were frequent at some times of the year, e.g. towards Christmas or during the football season when funds were needed for travel.
- haosgel* literally 'house girl', usually a teenager who lives with someone in her extended family to help out with domestic work. Most expatriates have a 'house girl', but in the towns this may be an older woman, working for wages, who does not live with the family.
- kastom* customary or traditional practice(s). *Kastom* forms the basis for one level of the judicial system and is supposed to guide interpretation of British and French-based law at the national level.
- lanwis* literally 'language' but usually refers more specifically to the Melanesian languages (but not Bislama) spoken in Vanuatu.
- man Malo/Ambae etc.* people self-identify with their place of origin, or where they have kastom land rights through parents. Bislama places the modifier after *man* or *woman*. *Man ples* is used generically to refer to people from the same area (usually people who speak the same *lanwis*).
- nakamal* a (generally) simple structure with forecourt area where *kava* is prepared and sold in the evenings. (*Kava* [*Piper methysticum*] when pounded and mixed with water gives a drink with a usually gentle narcotic effect.) Traditionally, the *nakamal* and the drinking of *kava* was supposed to be ceremonial, but since it is much cheaper than beer it is a popular recreational drug.

<i>Natamambo</i>	people from Malo
<i>natanggura</i>	sago palm (<i>Metroxylon warburgii</i>), the large leaves are sewn together in overlapping layers and the layers are used as a traditional roofing material.
<i>Ni-Vanuatu</i>	attributive adjective from <i>Vanuatu</i> . Used in Bislama and the English spoken in Vanuatu to refer to people, concepts or things particular to Vanuatu.
<i>stret</i>	'straight', of brothers and sisters this indicates people share the same birth parents. The term is used because <i>brata</i> 'brother' and <i>sista</i> 'sister' refer to both siblings and classificatory brothers and sisters, i.e. children of anyone a person calls <i>papa</i> 'father', which includes some Western (first and second) cousins.
<i>tapu</i>	'taboo', I have used the local Malo realisation of this term (familiar throughout the Pacific). It is true that a <i>tapu</i> on a place indicates it is good idea either to avoid it or observe respectful behaviour there, but 'forbidden' is not the best translation. With respect to <i>kastom</i> practices, it indicates an association with non-earthly powers. However, parents from Malo and Ambae were heard to use it with their children, telling them to leave something alone. I do not know if its use in this way is wide-spread in Vanuatu.
<i>woman Malo/Ambae</i>	people self-identify with their place of origin, or where they have <i>kastom</i> land rights through parents. Bislama places the modifier after <i>man</i> or <i>woman</i> .
<i>womannara aelan</i>	'woman from another island', the term used on Malo to refer to women from outside Malo who had married <i>man Malo</i> and moved there. In the village(s) I stayed in, these women self-identified in these terms and formed a cohesive group for the purposes of socialising and entertainment.

Glosses for the titles

<i>Lukluk ples</i>	To have a look around the area
<i>Tinging blong lanwis</i>	lit. 'idea(s) of language'
<i>Ol man Malo oli havem tu sot</i>	lit. 'people on Malo wear two shirts', i.e. 'people on Malo wear two hats'. <i>Mi no min se ol man Malo oli nawita. Mi min se oli save yusum lanwis wetem Bislama. From fasin ya man Malo i stap long tu wol. Wan wol hem i wol blong Bislama, wan wol hem i wol blong kastom, blong ol Natamambo.</i>
<i>Wanem i stap yet?</i>	What still remains?

From wanem nao yumi no nidim wan jenis?

lit. 'because what now we NEG need one change', i.e. When don't you need a pronoun?

Crowley's (1987a) Bislama language grammar uses *jenis* for pronoun or anaphor.

Terry Crowley hem i yusum jenis olsem Inglis 'pronoun' o Franis 'pronom'. Hem i talem se i bin jusum wod ya from yumi yusum ol smolsmol wod ya (olsem hem) taem we yumi jenis wan nem i go long wan 'pronoun/pronom'.

Toktok Biaen

Afterword (lit. 'talk behind')

Tumora blong wok ya

Avenues for future research (lit. 'tomorrow of work SPEC')

References

- Adams, Marianne, 1987, From Old French to the theory of prodrop. *Natural Language and Linguistic Theory* 5/1:1-32.
- Allen, Michael, 1981, Rethinking old problems: matriliney, secret societies and political evolution. In Michael Allen, ed. *Vanuatu: politics, economics and ritual in Island Melanesia*. 9-34. Sydney: Academic Press.
- Alleyne, Mervyn C., 1980, *Comparative Afro-American: an historical-comparative study of English-based Afro-American dialects of the New World*. Ann Arbor: Karoma.
- Applegate, James L. and Howard E. Sypher, 1988, A constructivist theory of communication and culture. In Young Yun Kim and William B. Gudykunst, eds *Theories in intercultural communication*, 41-65. Newbury Park, CA: Sage.
- Bailey, Guy, Tom Wilke, Jan Tillery and Lori Sand, 1991, The apparent time construct. *Language Variation and Change* 3/3:241-264.
- Baker, Mark C., 1989, Object sharing and projection in serial verb constructions. *Linguistic Inquiry* 20/4:513-553.
- Bauman, Richard and Joel Sherzer, eds, 1989, *Explorations in the ethnography of speaking* (2nd ed.) Cambridge: Cambridge University Press.
- Beaglehole, J.C., 1974, *The life of Captain James Cook*. Stanford, CA: Stanford University Press.
- Bell, Allan, 1984, Language style as audience design. *Language in Society* 13/2:145-204.
- Berger, Charles R. and Richard J. Calabrese, 1975, Some explorations in initial interaction and beyond: toward a developmental theory of interpersonal communication. *Human Communication Research* 1/2:99-112.
- Besnier, Niko, 1996, Authority and egalitarianism: discourses of leadership on Nukulaelae Atoll. In Richard Feinberg and Karen Watson-Gegeo, eds *Leadership and change in the Western Pacific: essays presented to Sir Raymond Firth on the occasion of his 90th birthday*. London School of Economics Monographs on Social Anthropology, 66. 93-128. London: Athlone Press.
- Bhatt, Rajesh, 1998, Complex predicates as polarity items. Paper presented at Complex Predicates Workshop, Deutsche Gesellschaft für Sprachwissenschaft, Universität Halle/Saale.
- Bickerton, Derek, 1981, *Roots of language*. Ann Arbor: Karoma.
- 1990, *Language and species*. Chicago: University of Chicago Press.
- 1993, Subject focus and pronouns. In Francis Byrne and Donald Winford, eds *Focus and grammatical relations in Creole languages*, 189-212. Amsterdam/Philadelphia: Benjamins.

- Blackwood, Peter, 1981, Rank, exchange and leadership in four Vanuatu societies. In Michael Allen, ed. *Vanuatu: politics, economics and ritual in Island Melanesia*, 35-84. Sydney: Academic Press.
- Borer, Hagit, 1989, Anaphoric AGR. In Osvaldo Jaeggli and Kenneth J. Safir, eds *The null subject parameter*, 69-109. Dordrecht: Kluwer.
- Bonnemaison, Joël, 1986, *L'Arbre et la Pirogue (Les fondements d'une identité territoriale, histoire et société dans l'archipel de Vanuatu (Mélanésie), livre 1*. Paris: Éditions de l'ORSTOM.
- Bougainville, Louis-Antoine de, 1992 [1771], *Voyage de la Frégate la Boudeuse et de la flute L'Étiolle Autour du Monde*. Paris: Editions la Découverte (1st edn 1771) Paris: Saillant et Nyon.
- Bourdieu, Pierre, 1991, *Language and symbolic power* (trans. Gino Raymond and Matthew Adamson). Cambridge, MA: Harvard University Press.
- Bourhis, Richard Y., 1984, Cross-cultural communication in Montréal: two field studies since Bill 101. *International Journal of the Sociology of Language*. 46:33-47.
- Brennan, Susan, Marilyn Walker Friedman, and Carl Pollard, 1987, A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of ACL*, Stanford.
- Britain, David, 1992, Linguistic change in intonation: the use of high rising terminals in New Zealand English. *Language Variation and Change* 4/1:77-104.
- Brown, H. Douglas, 1987, *Principles of language learning and teaching*. Englewood Cliffs, NJ: Prentice Hall.
- Brown, Penelope and Stephen C. Levinson, 1987, *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
- Brown, Roger and Albert Gilman, 1968, The pronouns of power and solidarity. In Joshua A. Fishman, ed. *Readings in the sociology of language*, 252-275. The Hague: Mouton.
- Bucholtz, Mary, 1997, Geek the girl: language, femininity, and female nerds. Paper presented at the 4th Berkeley Women and Language Conference. April 1996.
- Buckley, Eugene, 1994, Persistent and cumulative extrametricality in Kashaya. *Natural Language and Linguistic Theory* 12/3:423-464
- Buller, David B., Judee K. Burgoon, Cindy H. White and Amy S. Ebesu, 1994, Interpersonal deception VII: behavioral profiles of falsification, equivocation, and concealment. *Journal of Language and Social Psychology* 13/4:366-395.
- Burgoon, Judee K., 1994, Prologue. In Judee K. Burgoon, ed. *Journal of Language and Social Psychology: Special Issue – Interpersonal Deception* 13/4:357-365.
- Butler, Judith, 1990, *Gender trouble: gender and the subversion of identity*. New York: Routledge.
- Bybee, Joan, 1998, "Irrealis" as a grammatical category. *Anthropological Linguistics* 40/2:257-271.
- Byrne, Francis, 1990, Tense marking in serial structures. In Brian D. Joseph and Arnold M. Zwicky, eds *When Verbs Collide: Papers from the (1990) Ohio State Mini-Conference on Serial Verbs (OSUWPL 39)*, 149-177. Dept. of Linguistics: The Ohio State University.
- 1991, Approaches to "missing" internal (and external) arguments in serial structure: some presumed difficulties. In Francis Byrne and Thom Huebner, eds *Development and Structures of Creole Languages*, 207-222. Amsterdam/Philadelphia: Benjamins.

- Camden, Pastor Bill, 1977, *A descriptive dictionary: Bislama to English*. Maropa Bookshop: Vila.
- Cameron, Deborah, 1996, The language-gender interface: challenging co-optation. In Victoria L. Bergvall, Janet M. Bing and Alice F. Freed, eds *Rethinking language and gender research: theory and practice*, 31-53. London/NY: Longman.
- Cameron, Deborah, Elizabeth Frazer, Penelope Harvey, Ben Rampton and Kay Richardson, 1992, *Researching language: issues of power and method*. London: Routledge.
- Cameron, Richard, 1992, Pronominal and null subject variation in Spanish: constraints, dialects and functional compensation. PhD dissertation, University of Pennsylvania. [Also on-line as: IRCS Report 92-22, Philadelphia, PA: Institute for Research in Cognitive Science, University of Pennsylvania.]
- 1993, Ambiguous agreement, functional compensation, and non-specific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change* 5/3: 305-334.
- Canagarajah, A. Suresh, 1995, Functions of codeswitching in ESL classrooms: Socialising bilingualism in Jaffna. *Journal of Multilingual and Multicultural Development*. 6/3:173-195.
- Capell, Arthur and John Layard, 1980, *Materials in Atchin, Malekula: grammar, vocabulary and texts*. Canberra: Pacific Linguistics.
- Carbaugh, Donal, 1990, Intercultural communication. In Donal Carbaugh, ed. *Cultural communication and intercultural contact.*, 151-175. Hillsdale, NJ: Lawrence Erlbaum.
- Central Intelligence Agency, 1997, *Factbook*. www.odci.gov/cia/publications/nsolo/factbook
- Chafe, Wallace, 1994, *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing*. Chicago: Chicago University Press.
- Chambers, J.K., 1995, *Sociolinguistic theory: linguistic variation and its social significance*. Oxford: Basil Blackwell.
- 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].
- Cheshire, Jenny, 1982, Linguistic variation and social function. In Suzanne Romaine, ed. *Sociolinguistic variation in speech communities*, 153-166. London: Edward Arnold.
- Chomsky, Noam, 1986, *Knowledge of language: its nature, origins and use*. New York: Praeger.
- 1995, *The minimalist program*. Cambridge, MA: MIT Press.
- Churchill, William, 1911, *Beach-la-mar: the jargon or trade speech of the Western Pacific*. Washington, DC: The Carnegie Institution of Washington.
- Clark, Ross, 1985, Languages of north and central Vanuatu: groups, chains, clusters and waves. In Andrew K. Pawley and Lois Carrington, eds *Austronesian Linguistics at the 15th Pacific Science Congress*, 199-236. Canberra: Pacific Linguistics.
- Collins, Chris, 1995, Serial verb constructions and the theory of multiple feature checking. MS, Cornell University.
- 1997, Argument sharing in serial verb constructions. *Linguistic Inquiry* 28:461-497.
- Condominium of the New Hebrides, 1967, *A report on the first census of the population 1967*. Sydney[?]: Government Printer, New South Wales.

- Cook, Kenneth William, 1994, The empathy hierarchy and Samoan clitic pronouns. *Cognitive Linguistics* 5/1:57-75.
- Corne, Chris, Dierdre Coleman and Simon Curnow, 1996, Clause reduction in asyndetic coordination in Isle de France Creole: the "serial verbs" problem. In Philip Baker and Anand Syea, eds *Changing meaning, changing functions: papers relating to grammaticalization in contact languages*, 129-154. London: University of Westminster.
- Cote, Sharon Ann, 1996, Grammatical and discourse properties of null arguments in English. PhD dissertation, University of Pennsylvania.
- Coupland, Nikolas, Justine Coupland and Howard Giles, 1991, *Language, society and the elderly: discourse, identity and ageing*. Oxford: Blackwell.
- Cronen, Vernon E., Victoria Chen and W. Barnett Pearce, 1988, Coordinated management of meaning: a critical theory. In Young Yun Kim and William B. Gudykunst, eds *Theories in intercultural communication*, 66-98. Newbury Park, CA: Sage.
- Crowley, Terry, 1982, *The Paamese language of Vanuatu*. Canberra: Pacific Linguistics.
- 1987a, *Grama blong Bislama*. Suva, Fiji: University of the South Pacific, Extension Services.
- 1987b, Serial verbs in Paamese. *Studies in Language* 11/1:35-84.
- 1989, *Say, c'est*, and subordinate constructions in Melanesian pidgin. *Journal of Pidgin and Creole Languages* 4/2:185-210.
- 1990, *Beach-la-Mar to Bislama: the emergence of a national language in Vanuatu*. Oxford: Oxford University Press.
- 1994, Linguistic demography: interpreting the 1989 census results in Vanuatu. *Journal of Multilingual and Multicultural Development* 15/1:1-16.
- 1995, *A new dictionary of Bislama*. Suva: Institute of Pacific Studies, University of the South Pacific.
- 1998, *An Erromangan (Sye) grammar*. Honolulu: University of Hawai'i Press.
- Deacon, A. Bernard, 1934, *Malekula, a vanishing people in the New Hebrides*. London: George Routledge.
- Di Eugenio, Barbara, 1990, Centering theory and the Italian pronominal system. *Coling90: Proceedings of the 13th International Conference on Computational Linguistics*. Helsinki.
- DuBois, John W., 1987, The discourse basis of ergativity. *Language* 63/4:805-855.
- Du Feu, Veronica, 1996, *Rapanui*. London: Routledge.
- Duranti, Alessandro, 1984, The social meaning of subject pronouns in Italian conversation. *Text* 4/4:277-311.
- Duranti, Alessandro and Charles Goodwin, eds, 1992, *Rethinking context: language as an interactive phenomenon*. Cambridge: Cambridge University Press.
- Dutton, Tom, 1980, *Queensland canefields English of the late Nineteenth Century (a record of interview with two of the last surviving Kanakas in North Queensland, 1964)*. Canberra: Pacific Linguistics.
- Eastman, Carol M., 1992, Code-switching as an urban language-contact phenomenon. *Journal of Multilingual and Multicultural Development*. 13/1 and 2:1-17.
- Early, Robert, 1993, Nuclear layer serialization in Lewo. *Oceanic Linguistics* 32/1:65-93.

- 1995, *Sit, stand, lie*: posture verbs and imperfectives. Paper presented at the Second International Conference on Oceanic Linguistics. Suva, The University of the South Pacific. 3–7 July 1995.
- Ebesu, Amy S. and Michael D. Miller, 1994, Verbal and nonverbal behaviors as a function of deception type. *Journal of Language and Social Psychology* 13/4:418-442.
- Eckert, Penelope, 1989, The whole woman: sex and gender differences in variation. *Language Variation and Change* 1/3:245-267.
- Eckert, Penelope and Sally McConnell-Ginet, 1992, Think practically and look locally: language and gender as community-based practice. *Annual Review of Anthropology* 21:461-490.
- Eisikovits, Edina, 1987, Sex differences in inter-group and intra-group interaction among adolescents. In Anne Pauwels, ed. *Women and Language in Australian and New Zealand Society*, 45-58. Mosman, NSW: Australian Professional Pubs.
- Embick, David, 1994, Licensing and consonant harmony. *University of Pennsylvania Working Papers in Linguistics* 1:29-43.
- Encyclopedia Britannica*, 1994–1997. <http://www.eb.com>
- Ferguson, Charles A., 1964, Diglossia. In Dell Hymes, ed. *Language in culture and society*, 429-437. New York: Harper Row.
- Fischer, John L., 1958, Social influences on the choice of a linguistic variant. *Word* 14/1:47-56.
- Fought, Carmen, 1997, A majority sound change in a minority community. *University of Pennsylvania Working Papers in Linguistics* 4/1:39-56.
- Fowler, Joy, 1986, The social stratification of (r) in New York City department stores, 24 years after Labov. MS, New York University.
- Fox, Barbara, 1987, *Discourse structure and anaphora*. Cambridge: Cambridge University Press.
- Fox, Rev. G.J., 1979, *Big Nambas grammar*. Canberra: Pacific Linguistics.
- Freed, Alice F., 1992, We understand perfectly: a critique of Tannen's view of cross-sex communication. In Kira Hall, Mary Bucholtz and Birch Moonwomon, eds *Locating Power: proceedings of the Second Berkeley Women and Language Conference*, 144-152. Berkeley: Berkeley Women and Language Group.
- Gal, Susan, 1979, *Language shift: social determinants of linguistic change in bilingual Austria*. New York: Academic Press.
- 1992, Language, gender, and power: an anthropological view. In Kira Hall, Mary Bucholtz and Birch Moonwomon, eds *Locating Power: Proceedings of the Second Berkeley Women and Language Conference*, 153-161. Berkeley: Berkeley Women and Language Group.
- Gardner, Robert C. and Wallace E. Lambert, 1972, *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.
- Garrett, Peter, Nikolas Coupland and Angela Williams, 1997, Evaluating dialect in discourse: judgements of young Welsh English speakers. Paper presented at the 6th International Conference on Language and Social Psychology. University of Ottawa.
- Geertz, Clifford, 1988, *Works and lives: the anthropologist as author*. Oxford: Polity.
1995. *After the fact: two countries, four decades, one anthropologist*. Cambridge, MA: Harvard University Press.

- Gellner, Ernest, 1995, *Anthropology and politics: revolutions in the sacred grove*. Oxford: Basil Blackwell.
- Giles, Howard, Richard Y. Bourhis, Donald M. Taylor, 1977, Towards a theory of language in ethnic group relations. In Howard Giles, ed. *Language, ethnicity and intergroup relations*, 307-348. London: Academic Press.
- Giles, Howard, Susan Fox and Elisa Smith, 1993, Patronizing the elderly: intergenerational evaluation. *Research on Language and Social Interaction* 26/2:129-149.
- Givón, Talmy, 1984, *Syntax: A functional-typological introduction*. Volume I. Amsterdam/Philadelphia: John Benjamins.
- Goffman, Erving, 1974, *Frame analysis: an essay on the organization of experience*. Harmondsworth: Penguin.
- Graff, David, William Labov and Wendell A. Harris, 1986, Testing listeners' reactions to phonological markers of ethnic identity: A new method for sociolinguistic research. In David Sankoff, ed. *Diversity and diachrony*, 45-58. Amsterdam/Philadelphia: Benjamins.
- Grosz, Barbara, Aravind Joshi and Scott Weinstein, 1995, Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics* 21/2:203-225.
- Gudykunst, William B., 1995, Anxiety/Uncertainty Management (AUM) theory: current status. In Richard L. Wiseman, ed. *Intercultural communication theory*, 8-58. Thousand Oaks, CA: Sage.
- Güldemann, Tom and Manfred von Roncador, eds, forthcoming, *Reported discourse: a meeting ground for different linguistic domains* (Typological Studies in Language). Amsterdam: Benjamins.
- Gumperz, J.J., 1977, Sociocultural knowledge in conversational inference. In Muriel Saville-Troike, ed. *Twenty-eighth Annual Round Table Monographs on Language and Linguistics*, 191-211. Washington, DC: Georgetown University Press.
- 1982, *Discourse strategies*. Cambridge: Cambridge University Press.
- Guy, Gregory R., 1988, Advanced Varbrul analysis. In Kathleen Ferrera, Becky Brown, Keith Walters and John Baugh, eds *Linguistic change and contact: proceedings of the 16th annual conference on new ways of analyzing variation*, 124-136. Austin: University of Texas.
- Guy, J.B.M., 1974a, *Handbook of Bichelamar/Manuel de Bichelamar*. Canberra: Pacific Linguistics.
- 1974b, *A grammar of the Northern dialect of Sakao*. Canberra: Pacific Linguistics.
- Haegeman, Liliane, 1995, Root infinitives, tense, and truncated structures in Dutch. *Language Acquisition* 4/3:205-255.
- Haeri, Niloofar, 1996, *The sociolinguistic market of Cairo: gender, class and education* (Library of Arabic Linguistics). London/New York: Kegan Paul International.
- Hale, Ken, 1991, Misumalpan verb sequencing constructions. In Claire Lefebvre, ed. *Serial verbs: grammatical, comparative and cognitive approaches*, 1-35. Amsterdam/Philadelphia: John Benjamins.
- Hall, Kira, 1996, Hijra/Hijrin: language and gender identity. PhD dissertation. University of California at Berkeley.
- Haspelmath, Martin and Ekkehard König, eds, 1995, *Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds*. Berlin: Mouton de Gruyter.

- Haugen, Einar, 1975, Pronominal address in Icelandic: from you-two to you-all. *Language in Society* 4:323-339.
- Hay, Jennifer, Jane Faust West, Janet Pierrehumbert and Mary Beckman, 1999, Lexical frequency effects in speech errors. Paper presented at the annual meeting of the Linguistic Society of America, Los Angeles.
- Heap, David John, 1997, La variation grammaticale en géolinguistique: les pronoms sujet en roman central. PhD dissertation, University of Toronto.
- Heycock, Caroline, 1994, *Layers of predication: the non-lexical syntax of clauses*. New York: Garland.
- Hilliard, David, 1976, John Coleridge Patteson: missionary bishop of Melanesia. In J.W. Davidson and Deryck Scarr, eds *Pacific Islands portraits*, 177-200. Canberra: The Australian National University Press.
- Hirschbühler, Paul, 1995, Null subjects in verb-first embedded clauses in Philippe de Vigneulles' *Cent Nouvelles Nouvelles*. In Adrian Battye and Ian Roberts, eds *Clause structure and syntactic change*, 257-291. Oxford: Oxford University Press.
- Holm, Peter, 1988, *Pidgins and Creoles*. Cambridge: Cambridge University Press.
- Holmberg, Anders and Christer Platzack, 1990, On the role of inflection in Scandinavian syntax. In Werner Abraham, Wim Kosmeijer and Eric Reuland, eds *Issues in Germanic syntax*, 93-118. Berlin: Mouton de Gruyter.
- Holmes, Janet, 1992, *An introduction to sociolinguistics*. London: Longman.
- 1995, Glottal stops in New Zealand English: an analysis of variants of word-final /t/. *Linguistics* 33/3:433-463.
- 1997, Women, language and identity. *Journal of Sociolinguistics* 1/2:195-223.
- Holmes, Janet and Allan Bell, 1992, On shear markets and sharing sheep: the merger of EAR and AIR diphthongs in New Zealand English. *Language Variation and Change* 4/3:251-273.
- Holmes, Janet, Allan Bell and Mary Boyce, 1991, *Variation and change in New Zealand English: a social dialect investigation*. (Project Report to the Social Sciences Committee of the Foundation for Research, Science and Technology.) Wellington: Victoria University of Wellington.
- Hornstein, Norbert, 1990, *As time goes by: tense and universal grammar*. Cambridge, MA: MIT Press.
- Horvath, Barbara, 1985, *Variation in Australian English: the sociolects of Sydney*. Cambridge: Cambridge University Press.
- Huang, C.-T. James, 1984, On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15/4:531-574
- 1989, Pro-drop in Chinese: a generalized control theory. In Osvaldo Jaeggli and Kenneth J. Safir, eds *The null subject parameter*, 185-214. Dordrecht: Kluwer.
- Hudson, Carla L. and Elissa L. Newport, 1999, Adults, children, and creole genesis: an experimental approach to an old question. Paper presented at the annual meeting of the Society for Pidgin and Creole Linguistics, Los Angeles.
- Hyams, Nina M., 1987, Theory of parameters and syntactic development. In Thomas Roeper and Edwin Williams, eds *Parameter Setting*, 1-22. Dordrecht: Reidel.
- Hymes, Dell, 1974, *Foundations in sociolinguistics*. Philadelphia: University of Pennsylvania Press.

- Iatridou, Sabine and David Embick, 1997, *Apropos pro. Language* 73/1:58-78.
- Jackson, Jean, 1989, Language identity of the Colombian Vaupés Indians. In Richard Bauman and Joel Sherzer, eds *Explorations in the ethnography of speaking* (2nd edn), 50-64. Cambridge: Cambridge University Press.
- Jaeggli, Osvaldo and Kenneth J. Safir, 1989, The null subject parameter and parametric theory. In Osvaldo Jaeggli and Kenneth J. Safir, eds *The null subject parameter*, 1-44. Dordrecht: Kluwer.
- Jansen, Bert, Hilda Koopman and Pieter Muysken, 1978, Serial verbs in the creole languages. *Amsterdam Creole Studies II*. Amsterdam: University of Amsterdam Instituut voor Algemene Taalwetenschap.
- Jauncey, Dorothy, 1995, 'Verbiness' in Malo. MS, The Australian National University.
- 1997, A grammar of Tamambo, the language of western Malo. PhD dissertation. Canberra: The Australian National University.
- Jolly, Margaret, 1987a, The forgotten women: a history of migrant labour and gender relations in Vanuatu. *Oceania* 58/2:119-139.
- 1987b, The chimera of equality in Melanesia. *Mankind* 17/2:168-183.
- 1989, Sacred spaces: churches, men's houses and households in South Pentecost, Vanuatu. In Margaret Jolly and Martha Macintyre, eds *Family and gender in the Pacific*, 213-235. Cambridge: Cambridge University Press.
- 1991a, The politics of difference: feminism, colonialism and decolonisation in Vanuatu. In Gill Bottomley, Marie de Lepervanche and Jeannie Martin, eds *Intersexions: gender, class, culture, ethnicity*, 52-74. Sydney: Allen & Unwin.
- 1991b, "To save the girls for brighter and better lives": Presbyterian missions and women in the south of Vanuatu, 1848-1870. *The Journal of Pacific History* 26/1:27-48.
- 1996, Woman ikat raet long human raet o no? Woman's rights, human rights and domestic violence in Vanuatu. *Feminist Review* 52:169-190.
- Jones, Elizabeth, Cynthia Gallois, Michelle Barker and Victor Callan, 1994, Evaluations of interactions between students and academic staff: influence of communication accommodation, ethnic group, and status. *Journal of Language and Social Psychology* 13/2:158-191.
- Jourdan, Christine, 1985, Sapos Iumi Mitim Iumi: the social context of creolization in the Solomon Islands. PhD dissertation. Canberra: The Australian National University.
- Kameyama, Megumi, 1985, Zero anaphora: the case of Japanese. PhD dissertation. Stanford University.
- Keesing, Roger, 1985, Subject pronouns and tense-marking in Southeast Solomonic languages and Solomons Pijin: grounds for substratomania? In S.A. Wurm, ed. *Papers in Pidgin and Creole Linguistics* No.4, 97-132. Canberra: Pacific Linguistics.
- 1988, *Melanesian Pidgin and the Oceanic substrate*. Stanford, CA: Stanford University Press.
- 1991, The expansion of Melanesian Pidgin: further early evidence from the Solomons. *Journal of Pidgin and Creole Languages* 6/2:215-229.
- Kelly, Celsus (O.F.M.), (trans. and ed.) 1966, *La Australia del Espíritu Santo: the journal of Fray Martín de Munilla O.F.M. and other documents relating to The Voyage of Pedro Fernández de Quirós to the South Sea (1605-1606) and the Franciscan Missionary Plan (1617-1627)*, vol.1 & 2. Cambridge: Cambridge University Press (for the Hakluyt Society).

- Kroch, Anthony S., 1989, Reflexes of grammar in patterns of language change. *Language Variation and Change* 1/3:199-244.
- 1996, Dialect and style in the speech of Upper Class Philadelphia. In Gregory R. Guy, Crawford Feagin, Deborah Schiffrin and John Baugh, eds *Towards a social science of language: vol.1 Variation and change in language and society*, 23-45. Amsterdam/Philadelphia: John Benjamins.
- Kroch, Anthony S. and Cathy Small, 1978, Grammatical ideology and its effect on speech. In David Sankoff, ed. *Linguistic variation: models and methods*, 45-55. New York: Academic Press.
- Labov, William, 1966, *The social stratification of English in New York City*. Washington DC: Center for Applied Linguistics.
- 1972, *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- 1989, The child as linguistic historian. *Language variation and change* 1/1:85-97.
- 1990, The intersection of sex and social class in the course of linguistic change. *Language variation and change* 2/2:205-251.
- 1993, The unobservability of structure and its linguistic consequences. Paper presented at New Ways of Analyzing Variation 23, University of Ottawa.
- 1994, *Principles of linguistic change: volume 1: internal factors*. Oxford: Blackwell.
- 1996, The causes of linguistic change. Paper presented at New Ways of Analyzing Variation 25, University of Nevada, Las Vegas.
- Lambert, Wallace E., R. Hodgson, R.C. Gardner and S. Fillenbaum, 1960, Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60/1: 44-51.
- Lanca, Margaret, Christine Alksnis, Neil J. Roese and Robert C. Gardner, 1994, Effects of language choice on acculturation: a study of Portuguese immigrants in a multicultural setting. *Journal of Language and Social Psychology* 13/3:315-330.
- Laracy, Hugh M., 1976, Xavier Montrouzier: a missionary in Melanesia. In J.W. Davidson and Deryck Scarr, eds *Pacific Islands Portraits*, 127-145. Canberra: The Australian National University Press.
- Lazreg, Marnia, 1988, Feminism and difference: the perils of writing as a woman on women in Algeria. *Feminist Studies* 14/1:81-107.
- Leap, William, 1996, 'Slummin', 'fruit loops', 'trash' and 'naughty places'; How the language of *Gay City* reflects gay men's beliefs about gay gender(s). Paper presented at the Fourth Berkeley Women and Language Conference, University of California at Berkeley.
- Le Page, Robert B. and Andrée Tabouret-Keller, 1985, *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
- Levin, Beth and Malka Rappaport Hovav, 1995, *Unaccusativity: at the syntax-lexical semantics interface*. Cambridge, MA: MIT Press.
- Lim, Tae-Seop and John Waite Bowers, 1991, Facework: solidarity, approbation and tact. *Human Communication Research* 17/3:415-450.
- Lindstrom, Lamont, 1990, *Knowledge and power in a South Pacific society*. Washington: Smithsonian Institution Press.
- Lynch, John and Darrell T. Tryon, 1985, Central-Eastern Oceanic: a subgrouping hypothesis. In Andrew K. Pawley and Lois Carrington, eds *Austronesian Linguistics at the 15th Pacific Science Congress*, 31-52. Canberra: Pacific Linguistics.

- MacElhinny, Bonnie, 1992, 'I don't smile much anymore': Affect, gender, and the discourse of Pittsburgh police officers. In Kira Hall, Mary Bucholtz and Birch Moonwomon, eds *Locating Power: proceedings of the second Berkeley Women and Language Conference*, 386-403. Berkeley: Berkeley Women and Language Group.
- Manessy, Gabriel, 1995, *Créoles, pidgins, variétés véhiculaires: procès et genèse*. Paris: CNRS Editions.
- Markham, Clement, (trans. and ed.) 1904, *The voyages of Pedro Fernandez de Quiros, 1595 to 1606*, vol.1 & 2. The Hakluyt Society. (Reprinted 1967 Millwood, NY: Kraus Reprint).
- Marshall, A.J. (Jock), 1937, *The Black Musketeers: the work and adventures of a scientist on a South Sea island at war and in peace*. London: Heinemann.
- Martinet, André, 1952, Function, structure and sound change. *Word* 8/1:1-32.
- Massam, Diane and Yves Roberge, 1997, Operator-bound clitics and Niuean *ai*. In Jim Black and Virginia Motapanyane, eds *Clitics, checking and movement*, 273-299. Amsterdam/Philadelphia: John Benjamins.
- Mendoza-Denton, Norma, 1994, Language attitudes and gang affiliation among California Latina girls. In Mary Bucholz, A.C. Laing, Laurel A. Sutton and Caitlin Hines, eds *Cultural Performances: proceedings of the Third Berkeley Women and Language Conference*, 478-486. Berkeley, CA: Berkeley Women and Language Group.
- Meyerhoff, Miriam, 1994, 'Sounds pretty ethnic eh?' – a pragmatic particle in New Zealand English. *Language in Society* 23/3:367-388.
- 1996, Transitive marking in contact Englishes. *Australian Journal of Linguistics* 16/1:57-80.
- 1997, Engendering identities: Pronoun selection as an indicator of salient intergroup identities. *University of Pennsylvania Working Papers in Linguistics* 4/2:23-68.
- 1998, Comparing old and new information in Bislama: nominal deletion with *olsem*. Jan Tent and France Mugler, eds *SICOL: proceedings of the Second International Conference on Oceanic Linguistics: vol.1, Language Contact*, 85-93. Canberra: Pacific Linguistics.
- 1999, Being sorry in the Pacific. *Language in Society*. 28, 2. 225-238.
- forthcoming a, All the same? The emergence of complementisers in Bislama. In Tom Güldemann and Manfred von Roncador, eds *Reported Discourse: a meeting ground for different linguistic domains* (Typological Studies in Language). Amsterdam: Benjamins.
- forthcoming b, Dynamics of differentiation: On social psychology and cases of language variation. In Nikolas Coupland, Christopher Candlin and Srikant Sarangi eds *Sociolinguistics and social theory*. London: Longman.
- Mihalic, F., 1971, *The Jacaranda dictionary and grammar of Melanesian Pigin*. Milton, Queensland: Jacaranda Press.
- Millar, Sharon, 1994, Group identity, group strategies, and language in Northern Ireland. *Journal of Language and Social Psychology* 13/3:299-314.
- Milner, George B., 1966, *Samoan dictionary*. Oxford: Oxford University Press.
- Milroy, Lesley, 1980, *Observing and analysing natural language*. Oxford: Basil Blackwell.

- Mithun, Marianne, 1991, The role of motivation in the emergence of grammatical categories: the grammaticization of subjects. In Elizabeth Closs Traugott and Bernd Heine, eds *Approaches to grammaticalization: vol.1, Theoretical and methodological issues*, 159-184. Amsterdam/Philadelphia: John Benjamins.
- Monnier, Paul, 1996, *Apôtres des Nouvelles-Hébrides*. Nouméa, Nouvelle-Calédonie: Société d'études historiques de Nouvelle-Calédonie.
- Morgan, Marcyliena, 1994, The African-American speech community: reality and sociolinguists. In Marcyliena Morgan, ed. *Language and the social construction of identity in Creole situations*, 121-148. Los Angeles: Center for Afro-American Studies, UCLA.
- Mufwene, Salikoko S., 1986, The universalist and substrate hypotheses complement one another. In Pieter Muysken and Norval Smith, eds *Substrata versus Universals in creole genesis*, 129-162. Amsterdam/Philadelphia: Benjamins.
- 1991, Pidgins, creoles, typology and markedness. In Francis Byrne and Thom Huebner, eds *Development and structures of creole languages: essays in honor of Derek Bickerton*, 123-143. Amsterdam/Philadelphia: Benjamins.
- 1996, The founder principle in creole genesis. *Diachronica* 13/1:83-134.
- Mühlhäusler, Peter and Rom Harré, 1990, *Pronouns and people: the linguistic construction of social and personal identity*. Oxford: Basil Blackwell.
- Myers-Scotton, Carol, 1993a, *Duelling languages: grammatical structure in codeswitching*. Oxford: Oxford University Press.
- 1993b, *Social motivations for codeswitching: evidence from Africa*. Oxford: Oxford University Press.
- Nagy, Naomi, 1996, Language contact and language change in the Faetar speech community. PhD dissertation, University of Pennsylvania. [Also on-line as IRCS Report 96-08, Philadelphia, PA: University of Pennsylvania.]
- Nagy, Naomi and David Heap, 1998, Subject pronoun variation in Faetar and Francoprovençal. *Papers in Sociolinguistics: New Ways of Analyzing Variation-26 à l'Université Laval*, 291-300. Québec: Nuits Blanches.
- Nespor, Marina and Irene Vogel, 1986, *Prosodic phonology*. Dordrecht: Foris.
- Nevalainen, Terttu, 1998, Historical sociolinguistics: the bad-data problem and beyond. Paper presented at New Ways of Analyzing Variation 27, University of Athens, Georgia.
- Nevalainen, Terttu and Helena Raumolin-Brunberg, eds, 1996, *Sociolinguistics and language history: studies based on the corpus of early English correspondence*. Amsterdam: Rodopi.
- Nichols, Patricia C., 1983, Linguistic options and choices for Black women in the Rural South. In Barrie Thorne, Cheris Kramaræ and Nancy Henley, eds *Language, gender and society*, 54-68. Cambridge, MA: Newbury House.
- Niedzielski, Nancy, 1996, Acoustic analysis and language attitudes in Detroit. In (N)WAVES and MEANS: a selection of papers from New Ways of Analyzing Variation 24. *University of Pennsylvania Working Papers in Linguistics*, 3/1:73-86.
- 1999, The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18/1:62-85.
- Nile, Richard and Christian Clerk, 1996, *Cultural atlas of Australia, New Zealand and the South Pacific*. Abingdon: Andromeda Press.

- Noels, Kimberly A., Gordon Pon and Richard Clément, 1996, Language, identity, and adjustment: The role of linguistic self-confidence in the acculturation process. *Journal of Language and Social Psychology* 15/3:246-264.
- Nyutesteman wetem ol Sam long Bislama*, 1994, Suva: Bible Society in the South Pacific.
- Ochs, Elinor, 1992, Indexing gender. In Alessandro Duranti and Charles Goodwin, eds *Rethinking context: language as an interactive phenomenon*, 335-358. Cambridge: Cambridge University Press.
- O'Grady, William, 1998, Korean case: a computational approach. In B. Park and J. Yoon, eds *Selected Papers from the 11th International Conference on Korean Linguistics*, 1-19. Seoul: International Circle of Korean Linguistics.
- Oliveira e Silva, Giselle, 1982, Estudo da Regularidade na Variação dos Possessivos no Português do Rio de Janeiro. PhD dissertation, Universidade Federal do Rio de Janeiro.
- Paton, John G., 1889, *Missionary to the New Hebrides: an autobiography, edited by his brother* (James Paton). New York: Fleming H. Revell Co.
- Paton, W.F., 1971, *Ambrym (Lonwolwol) grammar*. Canberra: Pacific Linguistics.
- Pintzuk, Susan, 1995, Variation and change in Old English clause structure. *Language Variation and Change* 7/2:229-260.
- Pisier, Georges, 1980, Preface. In Jean Vanmai *Chân Dang: Les tonkinois de Calédonie au temps colonial*. Nouméa: Nouvelle-Calédonie: Société d'études historiques de la Nouvelles-Calédonie.
- Platzack, Christer, 1995, The loss of verb second in English and French. In Adrian Battye and Ian Roberts, eds *Clause structure and language change*, 200-226. Oxford: Oxford University Press.
- Poletto, Cecilia, 1996, Three kinds of subject clitics in Basso Polesano and the theory of *pro*. In Adriana Belletti and Luigi Rizzi, eds *Parameters and functional heads: essays in parametric syntax*, 269-300. Oxford/NY: Oxford University Press.
- forthcoming, *The higher functional field: evidence from the Northern Italian dialects*.
- Pollock, Jean-Yves, 1989, Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20/3:365-424.
- Poplack, Shana, 1980, Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18:581-618.
- Potter, Jonathan and Margaret Wetherell, 1987, *Discourse and social psychology: beyond attitudes and behaviour*. London: Sage.
- Prince, Ellen F., 1992, On syntax in discourse, in language contact situations. In Claire Kramsch and Sally McConnell-Ginet, eds *Test and context: cross-disciplinary perspectives on language study*, 98-112. Boston: C.D. Heath.
1996. Subject-prodrop in Yiddish. MS, University of Pennsylvania.
- Radtke, H. Lorraine and Henderikus J. Stam, 1994, Introduction. In H. Lorraine Radtke and Henderikus J. Stam, eds *Power/gender: social relations in theory and practice*, 1-14. London: Sage.
- Rambow, Owen, 1993, Pragmatic aspects of scrambling and topicalization in German. Paper presented at the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania.
- Raposo, Eduardo, 1987, Case theory and Infl-to-Comp: the inflected infinitive in European Portuguese. *Linguistic Inquiry*. 18/1:85-109.

- Ray, Sidney Herbert, 1978 [1936], *A comparative study of the Melanesian Island languages*. New York: AMS Press.
- Reichenbach, Hans, 1947, *Elements of symbolic logic*. New York: Macmillan.
- Reinhart, Tanya, 1981, A second comp position. In Adriana Belletti, Luciana Brandi and Luigi Rizzi, eds *Theory of markedness in generative grammar: proceedings of the 1979 GLOW conference*, 517-557. Pisa: Scuola Normale Superiore.
- Rice, Edward, 1974, *John Frum he come*. Garden City, NY: Doubleday & Co.
- Richards, Norvin, 1997, *What moves where when in which language?* (MIT Working Papers in Linguistics) Cambridge, MA: Massachusetts Institute of Technology.
- Rickford, John R., 1997, Unequal partnership: sociolinguistics and the African American speech community. *Language in Society* 26/2:161-197.
- Rickford, John R. and Faye McNair-Knox, 1994, Addressee- and topic-influenced style shift: a quantitative sociolinguistic study. In Douglas Biber and Edward Finegan, eds *Sociolinguistic perspectives on register*, 235-276. New York: Oxford University Press.
- Riddle, Elizabeth, 1990, Parataxis in White Hmong. In Brian D. Joseph and Arnold M. Zwicky, eds *When verbs collide: papers from the (1990) Ohio State Mini-Conference on Serial Verbs (OSUWPL 39)*, 65-76. Columbus, OH: The Ohio State University.
- Ringen, Catherine O., 1988, Transparency in Hungarian vowel harmony. *Phonology* 5/2:327-342.
- Rizzi, Luigi, 1982, *Issues in Italian syntax*. Dordrecht: Foris.
1990, *Relativized minimality*. Cambridge, MA: MIT Press.
- Roberge, Yves, 1990, *The syntactic recoverability of null arguments*. Montreal: McGill-Queen's University Press.
- Roberts, Ian, 1992, *Verbs and diachronic syntax*. Dordrecht: Kluwer.
- Robertson, H.A., 1902, *Erromanga: the martyr isle*. New York: Armstrong & Son.
- Rodman, Margaret, 1981, A boundary and a bridge: women's pig killing as a border-crossing between spheres of exchange in East Ambae. In Michael Allen, ed. *Vanuatu: Politics, Economics and Ritual in Island Melanesia*, 85-104. Sydney: Academic Press.
- Rohrbacher, Bernhard, 1994, The Germanic VO languages and the full paradigm: a theory of V to I raising. PhD dissertation, University of Massachusetts, Amherst.
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 and Martha Senturia, eds *The Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, 350-364. Palo Alto: Stanford Linguistics Associations/CSLI.
- Roloff, Michael E. and Chris A. Janiszewski, 1989, Overcoming obstacles to interpersonal compliance: a principle of message communication. *Human Communication Research* 16/1:33-61.
- Romaine, Suzanne, 1988, Variation and lexical change in Tok Pisin in Papua New Guinea. In *Georgetown University Round Table on Languages and Linguistics* 268-279.
- Ross, Malcolm, John Lynch and Terry Crowley, forthcoming, *The Oceanic languages*. London: Curzon Press.
- Ross, Steven and Richard Berwick, 1992, The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition* 14/2:159-176.

- Rubinstein, Robert L., 1978, Placing the self on Malo. PhD dissertation, Bryn Mawr University.
- 1981, Knowledge and political process on Malo. In Michael Allen, ed. *Vanuatu: Politics, Economics and Ritual in Island Melanesia*, 135-172. Sydney: Academic Press.
- Ryan, Ellen Bouchard, Richard Y. Bourhis and Uus Knops, 1991, Evaluative perceptions of patronizing speech addressed to elders. *Psychology and Aging* 6/3:442-450.
- Ryan, Ellen Bouchard, Mary Lee Hummert and Linda H. Boich, 1995, Communication predicaments of aging: patronizing behavior toward older adults. *Journal of Language and Social Psychology* 14/1 & 2:144-166.
- Sachdev, Itesh and Audrey Wright, 1996, Social influence and language learning: an experimental study. *Journal of Language and Social Psychology* 15/3:230-245.
- Saffran, Jennifer R., Richard N. Aslin and Elissa L. Newport, 1996, Statistical learning by eight-month-old infants. *Science* 274/5294:1926-1928.
- 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, 1972, Language use in multilingual societies: some alternative approaches. In J.B. Pride and Janet Holmes, eds *Sociolinguistics*, 33-51. Harmondsworth: Penguin.
- 1977, Variability and explanation in language and culture: cliticization in New Guinea Tok Pisin. In Muriel Saville-Troike, ed. *Georgetown University Round Table on Languages and Linguistics*, 59-73. Washington DC: Georgetown University Press.
- 1984, Substrate and universals in the Tok Pisin verb phrase. In Deborah Schiffarin, ed. *Meaning, form, and use in context: linguistic applications*, 104-119. Washington DC: Georgetown University Press.
- 1985, Contracts in Queensland: touching pen and marking paper. In Deborah Gewertz and Edward Schieffelin, eds *History and Ethnohistory in Papua New Guinea. (Oceania Monograph No.28)*, 100-126. Sydney: University of Sydney.
- 1993, Focus in Tok Pisin. In Francis Byrne and Donald Winford, eds *Focus and grammatical relations in Creole languages*, 117-140. Amsterdam/Philadelphia: John Benjamins.
- 1994, An historical and evolutionary approach to variation in the Tok Pisin verb phrase. *Parasession on Variation and Linguistic Theory: Chicago Linguistic Society* 30. 293-320.
- 1996, The Oceanic substrate in Melanesian pidgin/creole revisited: a tribute to Roger Keesing. In John Lynch and Fa'afu Pat, eds *Oceanic Studies: proceedings of the First International Conference on Oceanic Linguistics*, 421-450. Canberra: Pacific Linguistics.
- Sankoff, Gillian and Penelope Brown, 1976, The origins of syntax in discourse: a case study of Tok Pisin relatives. *Language* 52/3:631-666.
- Sankoff, Gillian and Suzanne Laberge, 1980, Anything *you* can do. In Gillian Sankoff, *The social life of language*, 271-293. Philadelphia: University of Pennsylvania Press.
- Santorini, Beatrice, 1993, The rate of phrase structure change in the history of Yiddish. *Language Variation and Change* 5/3:257-283.
- Scarr, Deryck, 1976, Recruits and recruiters: a portrait of the labour trade. In J.W. Davidson and Deryck Scarr, eds *Pacific Islands Portraits*, 225-251. Canberra: The Australian National University Press.

- Scherer, Klaus R. and Howard Giles, eds, 1979, *Social markers in speech*. Cambridge: Cambridge University Press.
- Scherre, Maria Marta Pereira and Anthony J. Naro, 1991, Marking in discourse: "Birds of a feather". *Language Variation and Change* 3/1:23-32.
1992. The serial effect on internal and external variables. *Language Variation and Change* 4/1:1-13.
- Schieffelin, Bambi B., 1984, *Ade: A sociolinguistic analysis of a relationship*. In John Baugh and Joel Sherzer, eds *Language in use: readings in sociolinguistics*, 229-243. Englewood Cliffs, NJ: Prentice Hall.
- Schilling-Estes, Natalie, 1998, Investigating 'self-conscious' speech: The performance register in Ocracoke English. *Language in Society* 27/1:53-83.
- Schilling-Estes, Natalie and Jessica Schrider, 1996, The symbolization of Islander identity: sex- and gender-based variation in Ocracoke English. Paper presented at New Ways of Analyzing Variation 25, University of Nevada, Las Vegas.
- Schroeder, Christoph, 1990, Pidgin- und Kreolsprachen: eine Forschungsübersicht. *Papiere zur Linguistik* 43/2:127-157.
- Schuetz, Albert J., 1969, *Nguna Grammar (Oceanic Linguistics Special publication No.5)*. Honolulu: University of Hawai'i Press.
- Schumann, John H., 1976, Social distance as a factor in second language acquisition. *Language Learning* 26:135-143.
- 1994, Where is cognition? Emotion and cognition in second language acquisition. *Studies in Second Language Acquisition* 16/2:231-242.
- Schwartz, Bonnie D. and Sten Vikner, 1996, The verb always leaves IP in V2 clauses. In Adriana Belletti and Luigi Rizzi, eds *Parameters and functional heads: essays in comparative syntax*, 11-62. Oxford: Oxford University Press.
- Sebba, Mark, 1987, *The syntax of serial verbs*. Amsterdam/Philadelphia: John Benjamins.
- Siegel, Jeff, Barbara Sandemann and Chris Come, 1999, Predicting substrate influence: Tense-modality-aspect marking in Tayo. Paper presented at the annual meeting of the Society for Pidgin and Creole Linguistics, Los Angeles.
- Seuren, Pieter A.M., 1990, Serial verb constructions. In Brian D. Joseph and Arnold M. Zwicky, eds *When verbs collide: papers from the (1990) Ohio State Mini-Conference on Serial Verbs (OSUWPL 39)*, 14-33. Dept. of Linguistics: The Ohio State University.
- Shaw, Patricia A., 1991, Consonant harmony systems: the special status of coronal harmony. In Carole Paradis and Jean-Francois Prunet, eds *The special status of coronals: internal and external evidence, II: phonetics and phonology*, 125-157. San Diego: Academic Press.
- Shineberg, Dorothy, 1967, *They came for sandalwood*. Carleton, Victoria: Melbourne University Press.
- Silva-Corvalán, Carmen, 1994, *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.
- Singh, Ishtla, 1996, Superstratal influence on the formation of Trinidad's English based creole. Paper presented at Sociolinguistic Symposium 11, University of Wales, Cardiff.
- Slobin, Dan, 1977, Language change in childhood and in history. In John Theodore Macnamara, ed. *Language learning and thought*, 185-214. New York: Academic Press.

- Smith, Geoff, 1995, Word frequency and regional language varieties in Tok Pisin. Paper presented at the Second International Conference on Oceanic Linguistics (SICOL). University of the South Pacific, Suva.
- Smith, Jennifer, 1998, 'I na think I could handle one. I really do na': Periphrastic *do* in Buckie, Scotland. Paper presented at New Ways of Analyzing Variation 27, Athens, Georgia.
- Smith, Philip M., Howard Giles and Miles Hewstone, 1980, Sociolinguistics: a social psychological perspective. In Robert N. St Clair and Howard Giles, eds *The social and psychological contexts of language*, 283-298. Hillsdale, NJ: Lawrence Erlbaum.
- Smolicz, J.J., 1981, Core values and cultural identity. *Ethnic and Racial Studies* 4/1:75-90.
- Speas, Margaret, 1996, Null objects in functional projections. In Johan Rooryck and Laurie Zaring, eds *Phrase structure and the lexicon*, 187-211. Dordrecht: Kluwer.
- Sperlich, Wolfgang B., 1993, Serial verb constructions in Namakir of Central Vanuatu. *Oceanic Linguistics* 32/1:95-110.
- Spolsky, Bernard, 1969, Attitudinal aspects of second language learning. *Language Learning* 19/3-4:271-283.
- St Clair, Robert N., 1980, The contexts of language. In Robert N. St Clair and Howard Giles, eds *The social and psychological contexts of language*, 11-38. Hillsdale, NJ: Lawrence Erlbaum.
- Starosta, Stanley, forthcoming, 41. Lexicase grammar. In Vilmos Igel, Ludwig, M. Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer, and Hening Lobin, eds *Dependency and valency – Dependenz und Valenz: an international handbook of contemporary research – Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin and New York: Walter de Gruyter.
- Strassel, Stephanie M. and Charles Boberg, 1996, The reversal of a sound change in Cincinnati. In (N)WAVES and MEANS: a selection of papers from New Ways of Analyzing Variation 24. *University of Pennsylvania Working Papers in Linguistics* 3/1:247-256.
- Tajfel, Henri and John C. Turner, 1979, An integrative theory of intergroup conflict. In William G. Austin and Stephen Worchel, eds *The social psychology of intergroup relations*, 33-53. Monterey, CA: Brooks-Cole.
- Tonkinese Work Permits, no date, Department of Manuscripts and University Archives, Cornell University Library (#4474).
- Tonkinson, Robert, 1982, National identity and the problem of kastom in Vanuatu. *Mankind* 13/4:306-315.
- Trabelsi, Chedia, 1991, De quelques aspects du langage des femmes de Tunis. *International Journal of the Sociology of Language* 87:87-98.
- Tracy, Karen, 1990, The many faces of facework. In Howard Giles and W. Peter Robinson, eds *Handbook of language and social psychology*, 209-226. Chichester: Wiley.
- Van Trease, Howard, 1987, *The politics of land in Vanuatu: from colony to independence*. Suva: Institute of Pacific Studies of the University of the South Pacific.
- Troemel-Plöetz, Senta, 1991, Selling the apolitical. *Discourse and Society* 2/4:489-502.
- Tryon, Darrell T., 1968, The languages of the New Hebrides: a checklist and general survey. In C. Beaumont, Darrell T. Tryon and S.A. Wurm, eds *Papers in Linguistics of Melanesia* No.3, 43-90. Canberra: Pacific Linguistics.
- 1976, *New Hebrides languages: an internal classification*. Canberra: Pacific Linguistics.

- 1987, *Bislama: an introduction to the national language of Vanuatu*. Canberra: Pacific Linguistics.
- Turan, Umit Deniz, 1996, Null vs. overt subjects in Turkish discourse: a centering analysis. PhD dissertation, University of Pennsylvania. [Also on-line as IRCS Report 96-13. Philadelphia, PA: University of Pennsylvania.]
- Uchida, Aki, 1992, When 'difference' is 'dominance': A critique of the 'anti-power-based' cultural approach to sex differences. *Language in Society* 21/4:547-568.
- Vainikka, Anne and Yonata Levy, 1995, Empty subjects in Finnish and Hebrew. MS, University of Pennsylvania and The Hebrew University. [On-line as IRCS Report 95-31. Philadelphia, PA: University of Pennsylvania.]
- Vance, Barbara Sue, 1989. Null subjects and syntactic change in Medieval French. PhD dissertation, Cornell University.
- Vanuatu Weekly/Hebdomadaire*. Port Vila: Vanuatu.
- Velleman, Paul, John Capehart and Paul Pratt, 1993. *Data desk 4.1*. Software for the Macintosh.
- Verhaar, John W. M., 1995, *Toward a reference grammar of Tok Pisin: an experiment in corpus linguistics*. Honolulu: University of Hawai'i Press.
- Wald, Benji, 1986, Syntactic development after childhood: beyond the vernacular? In David Sankoff, ed. *Diversity and diachrony*, 153-170. Amsterdam/Philadelphia: John Benjamins.
- Wald, Benji and Timothy Shopen, 1981, A researcher's guide to the sociolinguistic variable (ING). In Timothy Shopen and Joseph M. Williams, eds *Style and variables in English*, 219-249. Cambridge: Winthrop.
- Walker, Marilyn, Sharon Cote and Masayo Iida, 1994, Japanese discourse and the process of centering. *Computational Linguistics* 20/2:193-232.
- Wawn, William T., 1973, *The South Sea Islanders and the Queensland labour trade*. Peter Corris, ed. Honolulu: The University Press of Hawai'i.
- Webber, Bonnie L., 1991, Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes* 6/2:107-135.
- Wexler, Ken, 1994, Optional infinitives, head movement and the economy of derivations. In David Lightfoot and Norbert Horstein, eds *Verb movement*, 305-350. Cambridge: Cambridge University Press.
- Wilson, Edward O., 1994, *Naturalist*. Washington DC: Island Press.
- Wolfram, Walt and Kirk Hazen, 1997, Dialect accommodation and postinsular ethnolinguistic isolation. Paper presented at meeting of the Linguistic Society of America, Chicago.
- Wood, Linda A. and Rolf O. Kroger, 1994, The analysis of facework in discourse: review and proposal. *Journal of Language and Social Psychology* 13/3:248-277.
- Woods, Anthony, Paul Fletcher and Arthur Hughes, 1986, *Statistics in language studies*. Cambridge: Cambridge University Press.
- Woolford, Ellen, 1977, Aspects of Tok Pisin grammar. PhD dissertation, Duke University.
- Yamamoto, Masayo, 1995, Bilingualism in international families. *Journal of Multilingual and Multicultural Development*. 16/1 & 2:63-85.

- Yumi Wok Tugeta Blong Givhan Long Ol Woman: sot toktok blong Beijing Platfom blong Aksen blong Vanuatu*, 1996, Port Vila: Economic and Social Commission for Asia and the Pacific/Pacific Operations Centre.
- Zanuttini, Raffaella, 1996, On the relevance of tense for sentential negation. In Adriana Belletti and Luigi Rizzi, eds *Parameters and functional heads: essays in comparative syntax*, 181-207. Oxford: Oxford University Press.
- Zilliacus, Henrik, 1953, *Selbstgefühl und Servilität: Studien zum unregelmässigen Numerusgebrauch im Griechischen*. Helsingfors: Societas Scientiarum Fennica.