Borrowing versus Code-Switching in West Tarangan (Indonesia)

SIL International
Publications in Sociolinguistics

Publication 8

Publications in Sociolinguistics is a serial publication of SIL International. The series is a venue for works covering a broad range of topics in sociolinguistics. While most volumes are authored by members of SIL suitable works by others will also form part of the series.

Series Editor

M. Paul Lewis

Associate Editor

Gloria Kindell

Volume Editor

Marilyn Mayers

Production Staff

Bonnie Brown, Managing Editor
Judy Benjamin, Compositor
Hazel Shorey, Graphic Artist

Borrowing versus Code-Switching in West Tarangan (Indonesia)

Richard J. Nivens

SIL International
Dallas, Texas

Library of Congress Catalog No: 2002109892
ISBN: 1-55671-134-4
ISSN: 1091-9074

Printed in the United States of America

10 09 08 07 06 05 04 03 02 10 9 8 7 6 5 4 3 2 1

Copies of this and other publications of SIL International may be obtained from

International Academic Bookstore
SIL International
7500 W. Camp Wisdom Road
Dallas, TX 75236-5699

Voice: 972-708-7404
Fax: 972-708-7363
Email: academic_books@sil.org
Internet: http://www.ethnologue.com

Dedicated to Yeshua,

the source of Truth

Contents

List of Tables

List of Figures

Abbreviations and Symbols

1s	1st person singular
2s	2nd person singular
3s	3rd person singular
3sa	3rd person singular animate
3sn	3rd person singular inanimate
1p	1st person plural
1pi	1st person plural inclusive
1pe	1st person plural exclusive
2p	2nd person plural
3p	3rd person plural
3pa	3rd person plural animate
3pn	3rd person plural inanimate
AM	Ambonese Malay
AN	animate
CA	conversation analysis
CAUS	causative
CS	code-switching
DEF	definite
DM	Dobo Malay
DQ	direct quote
DUP	reduplication
EL	embedded language
FUT	future
GPM	Gereja Protestan Maluku (Protestant Church of Maluku)

INAN	inanimate
IP	inalienably possessed nouns
LCP	language contact phenomenon
LOC	locative
ML	matrix language
NZR	nominalizer
PL	plural
POSS	possessive word
PSV	passive
R	r- prefix
RECIP	reciprocal
REL	relativizer
S	sentence (or clause)
SG	singular
SIM	Standard Indonesian Malay
WT	West Tarangan
WTA	West Tarangan Variety A
WTB	West Tarangan Variety B
=	clitic boundary
•	word boundary within a lexical unit
(...) or ( )	unintelligible or semi-intelligible
á, é, í, ó, ú	final stress
ô, ê	tense mid vowels (o and e with no circumflex: lax mid vowels)
[ ]	portions of utterances excluded from lexical analysis
:	length
÷	speaker interrupted by another, but continues
—	abruptly unfinished sentence
-	abruptly unfinished word
[3]	3-second pause, [2] 2-second pause, etc.
[.]	micro-pause
{}	author's comments on the situation or manner of speech
[ok ma]	false start or repaired stretch
« »	direct quote
underlining	Malay mode

Superscript Codes in Examples

A	Malay items for which some crucial semantic component is absent from the WT near-equivalent
C	miscellaneous Malay cultural imports
D	Malay displacements (other than kin terms) of WT lexical units which are equivalent in every way
F	high-frequency Malay items
H	high register of Malay
J	Malay job titles
K	Malay kin terms
L	Malay items with lengthy WT near-equivalents
M	Malay items used metalinguistically
N	Malay items with no WT equivalent (other than cultural imports)
P	Malay proper nouns and NPs
R	discourse-induced repetitions of Malay items
S	Malay idioms and items with figurative senses
W	items of Malay origin which interviewees insisted were WT, not Malay
X	Malay exclative interjections
Y	lone non-default Malay items
YC	imported Malay concepts for which a WT term has been coined
YN	Malay terms (mainly causative constructions) for which a WT wording is readily available but would involve a syntactic restructuring of the sentence

z	Malay items for which it was impossible to investigate any possible motivation for using Malay, often due to ambiguities or unintelligible words in the context
^	speakers overlap
#	Malay numbering (following one of the other codes)
+	words with a pluralizing reduplication
/	mixed phrases
//	Malay item integrated by using dummy verb -m
%	mixed words

Acknowledgements

I would like to express first of all my thanks to Jesus, my king. From him come the skill, the motivation, the good ideas, and the clarity of thought. All good things are gifts from above.

As for purely human sources of ideas and inspiration, I would first of all like to thank Jim Collins, who persistently encouraged me to enter the Ph.D. program at the University of Hawai'i, which resulted in the dissertation upon which this book is based. I have gained great insights into the nature of language in general from Mike Forman and George Grace, and into the nature of Austronesian languages from Jim Collins and Bob Blust. To the late Stan Starosta I am grateful for helping me understand two contrasting theories of syntax; Ken Rehg and Patricia Donegan have deepened my knowledge of phonology. To the fifth member of my committee, Barbara Andaya, I am grateful for stepping in to fill in an empty slot at the last minute.

I also owe a large debt of gratitude to my most excellent wife, Susan, and to our sons Kristofer, Michael, and David. They have been a great source of support and comfort to me, despite my being so often absent from them in body or mind.

To my Indonesian friends and colleagues, I would like to say:

> Dalam kesempatan ini, penulis menyampaikan terima kasih yang sedalam-dalamnya kepada sahabat-sahabat dari Tarangan Barat yang telah menolong kami dalam penelitian bahasa maupun hal-hal yang lain. Penulis juga menghaturkan terima kasih kepada Pusat Pembinaan dan Pengembangan Bahasa, terutama Bapak Dr. Hasan Alwi, dan Lembaga Ilmu

Pengetahuan Indonesia (LIPI) yang telah mensponsori penelitian penulis sejak 1995 sampai 1997. Juga kepada Universitas Pattimura (UNPATTI), khususunya Bapak Drs. M. S. Darakay, yang telah memberi bantuan sejak 1985 sampai tahun 1991, dan ucapan terima kasih yang sama kepada pejabat-pejabat pemerintah Republik Indonesia yang telah memberikan ijin kepada penulis untuk mengadakan penelitian di pulau-pulau Aru. Semoga Tuhan memberkati budi baik mereka semua.

I am indebted to my friends in SIL International (SIL), especially my colleagues in Ambon, who have been a source of great encouragement to me. I owe a particular debt of gratitude to the SIL software developers who gave us the Text Analysis programs and Shoebox, without which this book would have been a lot more work. When I was far from an adequate library, Liz Storkey and her team of researchers, along with the Dallas SIL librarians, were a great help to me. Dr. Jerome D. Thayer of Andrews University showed me how to apply the chi-square test to my data.

I would like to give special thanks to our financial supporters, who have been very patient while some extremely important things have been left on the back burner for a while. The East-West Center, the University of Hawai'i, and SIL have also provided financial support in addition to academic training, for which I am grateful.

1

Introduction

Pluralitas non est ponenda sine necessitas.

William of Occam

Make everything as simple as possible, but no simpler.

Albert Einstein

The present work has three main goals. The first is merely to describe the language contact phenomena (LCP)[1] that characterize the speech of urban West Tarangan (WT) speakers of multi-ethnic Dobo town in the Aru islands in southeast Maluku, Indonesia. In presenting data from a corpus involving Malay and one of the many indigenous languages of Indonesia, this study contributes one more language pair to the growing field of LCP research. Although language mixing is a common phenomenon throughout the archipelago, there have been very few[2] studies of LCP in Indonesia; the current study therefore is a move toward filling an obvious geographic gap in LCP research. In addition, there is a different typological configuration of languages included here from those most often studied in LCP research, namely studies which involve two standardized national languages (e.g., French/English, Spanish/English). Due to fairly

[1]WT speakers themselves use the term *campur bahasa* 'language mixing' for the kind of phenomena described here. However, there are some language contact phenomena that would probably not be considered *campur bahasa* by WT speakers (e.g., foreign accent, calquing, and convergence).

[2]I know of only five: Collins (1981), Mustakim (1994), Denes et al. (1994), Syahdan (1996), and van Staden (1998).

recent changes in WT society (especially for WT speakers residing in Dobo), Malay is lexically dominant over WT for many conversational topics relevant to modern WT culture. As is seen in chapter 4, the corpus is very different from many other bilingual corpora as a result of this lexical inequality.

The second goal of the present work is to determine the effect of idiolectal differences, discourse context, and the availability of equivalent lexical units on the occurrence of embedded-language elements in a bilingual[3] corpus. This requires a lexical analysis of a corpus of recorded conversations, focusing especially on particular words, speakers, and discourse contexts. This is the focus of chapter 4.

This study's third main goal is to propose a psycholinguistically realistic accounting of the longer stretches of Malay occurring in the corpus as presented in chapter 5.

In this introductory chapter, I first provide a brief overview of LCP research. Then in §1.2 I discuss certain premises which I hold about the nature of language, contrasting them with premises which seem to underlie some other studies. In §1.3, I describe the micro-analytical perspective from which the present study examines the corpus.

1.1 An overview of research on code-switching and other LCP

1.1.1 Three waves of research

The history of LCP research has been adequately reviewed by Clyne (1988), Myers-Scotton (1993a:45–74), Milroy and Muysken (1995b), and others. In more specific surveys of the literature, Martin-Jones (1995) has reviewed LCP research in the educational context, while Myers-Scotton (1993b:19–45) summarized a number of syntactic constraints on code-switching (CS) that have been proposed, and Muysken (1995) gave an overview of different approaches and assumptions regarding syntactic analysis of CS.

LCP research began in earnest with Weinreich (1953) and Haugen (1950, 1953). With the publication of Blom and Gumperz (1972), CS became a well-known concept, among sociolinguists if not among linguists in general. Blom and Gumperz proposed that there are two types of CS, based on different social or communicative functions. The first, called SITUATIONAL CS, refers to a switch in language corresponding to a change in the speech situation, e.g., a change in topic, or in the constellation of

[3]Throughout the present work, the term bilingual(ism) should be understood to include multilingual(ism).

participants. A simple example would be where two participants are conversing in language A, and a third participant joins them whose presence (for whatever reason) induces them to switch to language B. The second, called metaphorical CS by Blom and Gumperz but later renamed CONVERSATIONAL CS by Gumperz (1982:59), refers to the use of another language because of the tone or other information conveyed by the very use of that language—in other words, the medium is part of the message. For example, someone may switch to the language of political authorities in an attempt to assume a position of authority himself.

Building on the pioneering work of Blom and Gumperz, the first wave of CS research continued to be sociolinguistic in nature, and this first wave continues to the present day. Major works in this first wave included Gal (1979) and Gardner-Chloros (1991). Scotton (1988) proposed a typology of social motivations underlying CS which adds two types of CS to Gumperz's initial two. She proposed that a speaker's communicative competence includes the knowledge "that for particular conventionalized exchanges in their community, a certain code choice indexes an expected set of rights and obligations. This code is the unmarked choice for that exchange" (p. 63). Thus, she referred to situational CS as SEQUENTIAL UNMARKED CHOICE since its social function is to maintain the unmarked character of the interaction; conversational CS was renamed SWITCHING AS A MARKED CHOICE. SWITCHING AS AN UNMARKED CHOICE referred to a discourse where participants code-switch repeatedly throughout as a normal mode of interaction, "symbolizing simultaneous identity in two rights and obligations balances" (p. 68). Scotton (pp. 76–78) proposed that this type of switching occurs only among in-group members in communities where (a) both languages are unmarked for the current interaction, and (b) both languages are markers of social identities which are appropriate in the current interaction. Finally, SWITCHING AS AN EXPLORATORY CHOICE can occur in nonconvention-alized exchanges, in which participants are attempting to negotiate what is to be the unmarked code choice in that particular interaction.

Myers-Scotton's model has not been universally accepted. In fact, there is still no general agreement on terminology or even on how LCP should be divided up into discrete categories—and indeed, whether such division is even possible (Gardner-Chloros 1995). In particular, the distinction between CS and borrowing has generated a great deal of discussion. This question is taken up again in §1.1.2 and is a major theme of chapter 4.

A second wave of CS research began when syntacticians began proposing constraints on which junctures within the sentence could allow a switch from one language to another. Several constraints were proposed; some were immediately found to be peculiar to the language pair in question or even to the

particular corpus studied,[4] while others were retained and modified in subsequent papers, e.g., the EQUIVALENCE CONSTRAINT (Poplack 1980), the FREE MORPHEME CONSTRAINT (Poplack 1980), and the GOVERNMENT CONSTRAINT (DiSciullo, Muysken, and Singh 1986). Some CS researchers have claimed that counter-examples to all of these constraints have been found, and have concluded that the constraints are therefore invalid. Others maintain these constraints as statistical tendencies if not as absolute rules of grammaticality. Myers-Scotton (1993b) presented a syntactic model of constraints on CS, while Poplack and her associates (e.g., Poplack and Meechan 1995) disagreed strongly with some of her conclusions. In the present work, I use Myers-Scotton's terms MATRIX LANGUAGE (ML) for the language which "sets the morphosyntactic frame of sentences" (1993b:3) and EMBEDDED LANGUAGE (EL) for the language which contributes items (varying in length from morphemes to stretches longer than a sentence) inserted into that ML.[5] The ML is a large-scale language choice which is generally maintained throughout a conversation between bilingual speakers, though a turnover in the ML is possible.

According to Martin-Jones (1995:94), in the late 1970s to early 1980s, the approach in LCP research was still "static, taxonomic and quantitative. The focus [was] still on individual acts rather than on the sequential flow of...discourse." This changed with the third wave of LCP research, which grew out of the CONVERSATION ANALYSIS (CA) tradition, in which researchers focus on how speakers' social identities and meanings are related to the flow of discourse. Auer, a major proponent of this third wave of LCP research, has argued (1995:116) that we must pay attention to the sequential environment (preceding and following utterances) of a code alternation in order to understand what it meant to the participants. Auer (1998) is a collection of papers representative of this third wave of LCP research. Gal (1988:290) stated: "Clearly we need to understand considerably more about the relationship between the social-historical and the social-interactional pressures on language as these interact with cognitive and purely grammatical constraints." The present work is in part an attempt to contribute something to this understanding.

All of these perspectives are valid and necessary, and I refer to them later in this work when describing the WT/Malay corpus. In particular, I have found the discourse-flow perspective to be essential to understanding the language contact phenomena found in the corpus.

[4]For example, the claim by Timm (1975:477) and Gumperz (1976) that a switch could not occur between subject and verb if the subject was pronominal. Other notable constraints include the SUBCATEGORIZATION CONSTRAINT (Bentahila and Davies 1983) and the CLOSED-CLASS CONSTRAINT (Joshi 1985).

[5]Nortier (1990:158) differentiates between matrix language and BASE LANGUAGE, a term also used by Poplack (1980), Grosjean (1995), and others, with varying definitions.

1.1.2 Previous attempts at differentiating code-switching and borrowing

Poplack, Wheeler, and Westwood (1989:132) stated: "The bilingual behavior which has provoked the most controversy in linguistics is undoubtedly intrasentential code-switching (CS)." More recently, Muysken (1995:190) similarly referred to "the problem that has dominated the field in recent years: the demarcation between borrowing and code-switching." As Nortier (1990:183) has pointed out, neither CS nor borrowing can adequately be defined without reference to the other. Unfortunately, after many years of LCP research, and despite Muysken's assurance (pp. 188–189) that "it appears that at present the general contours of the phenomenon of lexical borrowing are becoming firmly established," there is still no consensus on exactly what these two terms represent or how to distinguish them in theory or practice (Milroy and Muysken 1995a:12). This points out the most basic problem in this field of study, one which it shares with linguistics as a whole: defining the object of study. In a discipline where there is no general consensus on the answer to the question "What is language?", it is no surprise that we find it doubly difficult to define the combination of two or more "languages". The remainder of the present section will reveal the morass of mutually contradictory terminology plaguing LCP research today.

Borrowing and CS are differentiated by some researchers based on the occurrence (or degree of occurrence) of linguistic integration, whether phonological, morphological, or syntactic. Haugen (1950:212), for example, used phonological assimilation as a distinguishing criterion. But Grosjean (1982:129) and Bokamba (1988) claimed that both phonological and morphological integration are necessary characteristics of borrowing; Köppe and Meisel (1995:277–278) seemed to agree, although they admitted that the distinction may not always be clear-cut. Hyltenstam (1995:307) claimed that borrowed forms are morphologically and syntactically, but not necessarily phonologically, integrated. Nortier (1990:185) pointed out that Pfaff (1979), Scotton (1988), and Bentahila and Davies (1983) do not consider morphological integration as a criterion in distinguishing borrowing from code switching, but that Berk-Seligson (1986) only applies the term "borrowing" to single items which are integrated on all linguistic levels, with all other EL insertions considered to be code switches. Similarly, Poplack (1980) considered a single word EL insertion to be CS even if it was assimilated phonologically, morphologically, or syntactically, as long as it was not assimilated in all three ways.

In contrast to the criterion of integration, which is a matter of linguistic output, Pfaff (1979:297) differentiated CS and borrowing based on considerations related to input and motivation. She asked, Does the speaker

consider the word to belong to ML or EL? If he considers it an EL word, does he know an ML equivalent? Is that ML equivalent frequently used in the community? The difference between Pfaff's conception and those based on integration is the difference between the actual definition of an object and an operational definition or heuristic. What the field desperately needs is not more labels, but a set of formally defined entities on both the psycholinguistic (input) and surface (output) levels which can be demonstrated to correlate with each other on a statistically significant level.

Poplack and Sankoff (1984:103–104) reviewed previous studies and listed criteria used by other researchers to identify loan words. In addition to integration, these include frequency in the speech community, native synonym displacement, and acceptability by native speakers. Poplack, Sankoff, and Miller (1988:52) defined LEXICAL BORROWING as "the incorporation of individual L_2 words (or compounds [sic] functioning as single words) into discourse of L_1, the host or recipient language, usually phonologically and morphologically adapted to conform with the patterns of that language, and occupying a sentence slot dictated by its syntax." They also defined a variety of types of borrowing with regard to the particular corpus they were studying:

- nonce borrowings[6]—words that occur only once in our corpus
- widespread loans—words used by many speakers[7]
- loan word—a borrowing that has achieved some degree of currency
- idiosyncratic borrowings—frequently used words in the speech of a single speaker
- recurrent borrowings—borrowings said more than ten times, though not necessarily by as many speakers

They also defined CODE-SWITCHES as "multi-word L_2 sentence fragments [sic] which remain morphologically and syntactically unadapted to recipient-language patterns" though they also admitted that single-word code-switches are theoretically possible; in a footnote they explained: "In previous studies (Poplack 1980; Sankoff and Poplack 1981), single-word code-switches of grammatical categories not susceptible to borrowing were detected, such as single L_2 pronouns or articles in L_1 discourse."

Muysken (1995:190) provided a more general definition of nonce loans as "elements [which] are borrowed on the spur of the moment, without yet having any status in the receiving speech community." Such a

[6]The term nonce borrowing was first introduced by Weinreich (1953:11).

[7]In their particular study, more than ten speakers.

definition assumes that the bilingual speakers under study represent only a portion of the total number of speakers of the ML; however, for WT (as for many other languages in Indonesia), nearly all speakers are bilingual in Malay. The question arises, then, whether Malay words which occur only once in the WT/Malay corpus should be considered nonce loans or not, since the criterion of speech community acceptance is not meaningful or relevant. This question of frequency of occurrence is treated in chapter 4.

Nonce loans are a subtype of what Mackey (1988:704) called INTERFERENCE, which he contrasted with both CS and borrowing.[8] According to Mackey, interference may consist of individual EL sounds, intonation patterns, morphemes, expressions, or words, for example "the transient words that accidentally appear in the stream of speech of bilinguals which only a few bilinguals" unconsciously allow to enter their speech. He further claimed: "The only difference between interference and codeswitching seems to be in the level of linguistic analysis. Both have been included under the general phenomenon of alternation." Unfortunately, Mackey's use of the term alternation differs from that of Auer (1995), who used code alternation as a cover term for both CS and transfer (i.e., insertion of EL elements).

Still another usage of the term alternation was proposed by Muysken (1995:180). In his scheme, INSERTIONAL CS refers to EL material being inserted into an ML linguistic context (which is rather similar to borrowing), while ALTERNATIONAL CS refers to a definite switch from language A to language B, without necessarily an immediate return to language A. Muysken argued that there is no clear boundary between these two types, but gave various criteria for determining for any given case whether insertion or alternation is the more likely analysis.

Myers-Scotton (1993b:204–205) outlined the two heuristic criteria by which she separated CS from borrowings. First, in a 20-hour corpus, any lexeme which occurs more than three times is considered a borrowing. Second, of the remaining forms, any lexeme which represents an object or concept new to ML culture is also considered a borrowing. All remaining lexemes, whether assimilated into ML phonology and morphology or not, are considered CS.

In contrast to the heuristic criteria by which she attempted to identify CS in a corpus, Myers-Scotton (1993b:3) actually defined CS as "the selection by bilinguals or multilinguals of forms from an embedded variety (or varieties) in utterances of a matrix variety during the same conversation…CS may take place on any level of linguistic differentiation (languages, styles, or dialects/registers)." This explicitly insertional definition is in contrast to Poplack's

[8]Mackey's term "borrowing" corresponds to Poplack, Sankoff, and Miller's term "loan words".

(1980:583) definition of CS as "the alternation of two languages within a single discourse, sentence, or constituent." Although Poplack's definition is perhaps too broad to be useful, other researchers decline to define the term at all, except in a very vague sense.[9]

Although there is little agreement on where to locate the boundaries between CS and other LCP, there is general agreement on the distinction between intersentential CS and intrasentential CS. However, one might argue that a switch between clauses of a multiclausal sentence bears more affinity to an intersentential switch than to an intraclausal switch. Furthermore, the proliferation of sentence fragments, one-word sentences, and overlapping speech in live discourse renders the distinction somewhat less than pristine.[10]

Auer (1995:117) argued that in attempting to understand LCP, "it is necessary to draw a very basic distinction: that between contact phenomena classified as such by the linguist, and contact phenomena seen and used as such by the bilingual participants themselves." He called for a move away from structural approaches toward an interpretive approach. In addition to this distinction between participants and analysts, I claim that we need to keep in mind the psycholinguistic differences between the speaker's production and the hearers' interpretations of LCP. That distinction yields two corresponding analytical perspectives, one in which the analyst interprets the data just as a hearer would, and the other (much more difficult) in which the analyst makes use of data unavailable to the hearer which bear on the question of what brought about a particular instance of LCP.[11] It may be that part of the disagreement among linguists as to the distinction between CS and borrowing has stemmed from a failure to recognize the distinction between these two perspectives. In any event, it may be that the distinction between CS and borrowing is fuzzy from both analytical perspectives, but for different reasons. My rationale for this is presented in §1.2.

Given the great variation in terminology, and the informal nature of the competing definitions, and the zeal with which various theoreticians

[9]For example, Heller (1995:159) stated, "I use code-switching in an extended sense, to refer not only to instances of inter-sentential switching, but also to less-structurally integrated instances of language alternation or language choice"; and Köppe and Meisel (1995:277) defined CS as "a certain skill of the bilingual speaker that requires pragmatic and grammatical competence in both languages."

[10]Some LCP researchers (e.g., Poplack 1980:602, Nortier 1990:1) use the term extrasentential CS to refer to both the insertion of EL interjections and tags as well as to the insertion of EL direct quotes; others use that term to refer only to the insertion of interjections and tags; still others use the terms tag switching or emblematic switching for the latter case (Poplack 1980).

[11]At present, the latter perspective involves a good deal of guesswork, but as neurolinguistics advances we may hope that some of our present guesses will be confirmed or rejected.

argue for their own definitions of CS and borrowing, I will not risk confusing the reader (or myself) by choosing one of these definitions over any others, or by creating yet another definition. In this work, these two terms are used only informally, as little better than folk terms. Instead of focusing on finding the best definition of CS and borrowing, I attempt to determine whether any given instance of Malay in the WT/Malay corpus is an instance of language choice or lexical choice (chapter 4), and whether instances of language choice involve a change of language mode (chapter 5). Before discussing further the distinction between language choice and lexical choice, in §1.2 I make explicit for the reader some of my own beliefs and assumptions about language and LCP which may be different from those of other LCP researchers. I attempt here to provide a realistic view of differentiating LCP which future studies can use and to present the data explicitly enough that researchers who disagree with my assumptions and definitions can still make use of this work.

1.2 Beliefs and assumptions

Gardner-Chloros (1995) critiqued the field of LCP research, claiming that much of it has been based on simplistic assumptions about the nature of language. She cited four such assumptions (pp. 68–71):

1. that a bilingual speaker has two or more discrete codes that he switches in a binary fashion
2. that average group behavior is representative of individuals' CS patterns
3. that CS is something distinct from other types of LCP
4. that there is always a base language an individual switches out of (into a donor language)

In §§1.2.1 and 1.2.2 I discuss the first and third assumptions mentioned by Gardner-Chloros; the second is discussed in §1.2.3. The fourth, which is a direct challenge to Myers-Scotton's (1993b, 1995) Matrix Language-Frame Model, is not taken up here.[12]

Gardner-Chloros's views on the nature of language, like my own, have been greatly influenced by the work of Robert LePage (e.g., LePage 1978, 1989, LePage and Tabouret-Keller 1985). LePage (1978) proposed a continuum from focused languages to diffuse languages. Standardized languages are examples of focused languages, in which there is great uniformity among idiolects. But LePage's work with Caribbean creoles led him to propose that

[12]As seen in chapters 4 and 5, the WT/Malay corpus is one which is consistent with Myers-Scotton's claim that a matrix language is (nearly) always identifiable in a bilingual corpus.

many languages are more diffuse, in the sense that there is much greater variation among speakers with regard to grammaticality judgments and other linguistic features.

LePage and Tabouret-Keller (1985:194) argued that the notion of grammaticality assumes that language is a closed system, whereas language is actually an open system; acceptability, in contrast, is "a much wider...concept which takes account of creative, innovative, analogical, inventive and tolerant capacities of the human mind ignored by the closed systems of many grammarians." It is therefore a mistake, they argued, for linguists to formulate theories appropriate only to (hypothetical?) focused languages and then apply those theories to diffuse languages.

Gardner-Chloros (1995) claimed that previous CS studies have examined (or perhaps merely assumed) discrete and focused language pairs and stable bilingual communities. However, she argued that stable bilingual communities do not exist, because all language ecologies are in the process of change. Furthermore, "distinct and focused" is just one end of a continuum, and does not represent the majority of bilingual situations.[13] She concluded that CS is "an analyst construct rather than an observable fact. It is a product of our conceptualisations about language contact and language mixing, and it is not separable, either ideologically or in practice, from borrowing, interference or pidginisation."

1.2.1 Language and languages

In order to describe language contact phenomena, one must first consider the nature of language itself. In this section, I discuss the validity of various labels which have been used to describe linguistic behavior and other human behaviors, and then focus on the nature of linguistic competence.

Reliance on folk terms in LCP research: domain, language, culture, society, speech community. I begin with Gardner-Chloros's argument against the assumption that a bilingual speaker makes a binary choice between one language and another. Mackey (1988:699–700) considered different definitions of bilingualism and concluded that the most inclusive definition is "the knowledge and use of two or more languages". The standard definition of CS, reiterated by Milroy and Muysken (1995b:7), is "the alternative use by bilinguals of two or more languages

[13]For example, WT itself is less focused than Dobo Malay, which is in turn less focused than standard Indonesian; and as the present work demonstrates, the LCP of WT speakers is somewhat diffuse as well.

in the same conversation". Similarly, Muysken (1995:189) defined CS as "the use of two languages in one clause or utterance", describing the phenomenon in more technical terms as "words with different language indices...inserted into a phrase structure." He further defined lexical borrowing as "the incorporation of lexical elements from one language in the lexicon of another language."

The conceptual problem with such definitions referring to two languages is that they are based on the folk belief that languages are discrete, countable, real entities. Hudson (1980:21) argued against the folk belief that the world's population is divided up into a discrete number of languages, and that each language is divided up into a discrete number of dialects. Similarly, LePage and Tabouret-Keller (1985:7) disagreed with "the underlying assumption [made by Labov, Trudgill, and others] that there is a language called 'English', and that the people they are dealing with are speakers of that language; variation in their behaviour, therefore, must be accounted for in terms of variation in units in that underlying system, a system which all speakers of the language share." Further, they stated (p. 9): "We do not...need to put a boundary around any group of speakers and say 'These are the speakers of Language A, different from Language B', except to the extent that the people think of themselves in that way, and identify with or distance themselves from others by their behaviour."

Despite widespread recognition by linguists of variation right down to the level of the idiolect, LCP researchers continue to frame research questions in terms of folk categories like language and dialect. For example, Moravcsik (1978:119) asked, "under what conditions does the bilingual start *mixing* properties of his two languages and under what conditions do such mixed-in properties become accepted by the total population of a language including the nonbilingual members?" Such a question assumes that there exists a speech community (a discrete number of individuals) which speaks a language (speech behavior which is uniform on many parameters and distinct from other languages on those same parameters).

The definitional problem cannot be solved, however, by analyzing a single idiolect, as many prominent linguists have proposed. LePage and Tabouret-Keller argued that focusing and diffusion also occur at the level of the individual speaker. "To the extent that [a person's speech] is reinforced [by feedback from others], his behaviour in that particular context may become more regular, more focussed; to the extent that he modifies his behaviour to accommodate others it may for a time become more variable, more diffuse" (1985:181–182).

Over a century ago, Whitney (1875:154) warned his fellow scholars, "We must be careful not to overrate the uniformity of language."

Sociolinguists have followed Whitney's advice and divided societies up in various ways (class, age-grading, gender, dialect, and so on) in order to find the primitive units by which linguistic variation can be accurately described. On closer examination, however, these units have all turned out themselves to be artifacts of analysis, not real entities; that is, not only is there no internal homogeneity, but the entities are not bounded sets, since there is no feature they all have in common which also distinguishes them from others. LePage and Tabouret-Keller (1985:207–249) argued that just as there is no real and discrete entity called a language, so there is no real and discrete entity called a culture; rather, individual people have unique habits and beliefs which are more or less similar to those people who have had a great degree of influence on them. But as with language, people hold a folk belief that they are members of a culture. Similarly, the term speech community is merely an analytical construct which does not exist in the real world. There are no discrete societies in the real world; these only exist as constructs in peoples' minds. And as with language, conformity to a culture and membership in a society are dynamic, not static; so even an individualized account of behavior should not be taken as an adequate description of a real individual, but only a snapshot of behavior by that individual over a certain brief period of time.

LCP research has in the past referred to domain-specific use of different languages (Ferguson 1959). Preston (1988:690) described domain as "roughly a conglomerate of situational facts (setting, topic, personnel, relations, and purposes)". He then (pp. 692–693) argued that the analytical category domain is superfluous, since it is actually the components of domain that individually influence language, and these components do not always occur in the same constellations. Approaching the question from the practical side, Gardner-Chloros (1995:72) claimed that strict diglossia does not actually occur; she pointed out, for example, that Blom and Gumperz's (1972) classic account of CS involving standard and nonstandard varieties of Norwegian has been reanalyzed by Mæhlum (1990), who claimed that in that study the two speech varieties were idealized by the researchers.

Replacing these models of bounded sets nested neatly within one another are models of prototypes and networks. Historical linguists have long known that the tree model of language change is a simplistic distortion of reality; the wave model offers a more precise account of how changes spread. Similarly, in sociolinguistics it has become clear that analyses based on groups (e.g., age, gender, occupation, and education) do not yield as accurate an accounting of linguistic variation as do analyses based on social networks (Milroy 1987, Milroy and Li 1995:155). In both cases, the idea of simple relationships among a small number of

discrete entities (e.g., languages as nodes of a tree, or social classes as homogeneous groupings) has given way to a network of complex relationships among real individual entities. The difficulty in defining various types of LCP derives from the use of the word language in the definitions. An imprecise conception of language cannot be the basis of precise definitions of language contact phenomena.

Thus, languages and dialects are merely mental constructs which make it easier for us to talk about the ways people speak. Language is an abstraction; the only objects linguists actually study are speakers, not languages. People speak, and each person has his own unique and patterned way of speaking which changes across situations and over time. When two people share a sufficient number of patterns, they are able to communicate to a certain degree, and we can say informally that they share the same language. But when attempting to define linguistic terms like borrowing and code-switching and bilingualism, we cannot make use of such informal labels.

The consistency and stability of linguistic competence. Syntacticians interested in describing language competence realize the problems of variation between individuals and have focused on describing idiolects rather than languages. But Hockett (1958:321) defined idiolect as the totality of speech habits of a single person at a given time; thus, a person's idiolect is not static but dynamic. A speaker's internal competence changes over time, based on input from other speakers as well as input from his own creativity and even from his own performance errors. There is thus no absolute distinction between competence and performance. As a result, the question "Do CS data constitute competence data or performance data?" is unanswerable.

Is language a thing or a process? LePage and Tabouret-Keller (1985:234) argued that the reification of language is part of a general folk-theory process of linguistic nationism, by which the ways of speaking of a certain group of people are first given a label (e.g., English) and then the label is assumed to attach to some autonomous object from which a grammar and lexicon can be extracted by analysis. However, language is an abstract noun, representing an activity rather than an object; linguistic theory and linguistic definitions ought to be formally based on that premise rather than on the terms and concepts of folk theory.

If competence is variable, then it follows that speakers are inconsistent—not just at the level of performance, but even at the level of competence. While inconsistency in linguistic performance may be referred to as performance error, inconsistency at the level of linguistic competence is no error at all; rather, inconsistency, like redundancy, is a necessary characteristic of the human mind.

What kind of inconsistency is relevant to the analysis of LCP? As will be seen in §1.2.3, the difference between language choice and lexical choice hinges on the notion of equivalence. As discussed in chapter 4, equivalence of lexical items goes well beyond equivalence of denotation; it includes equivalence of connotation, frequency, expected impact on the hearer, and other types of equivalence. At any given point, a bilingual speaker in bilingual mode must decide whether he actually has a choice between sufficiently synonymous lexemes. The notion of inconsistency (in both competence and performance) indicates that at one point a speaker may consider two lexemes to be equivalent, while at another point the same speaker may consider them not equivalent—even if the discourse contexts are all but identical. Since people are inconsistent creatures, these factors do not always have the same proportional weight. I have yet to see a study of CS that explicitly takes such things into account.

LePage and Tabouret-Keller (1985:189–190) argued that even linguists sometimes do not know whether the claims they make about language actually refer to language behavior itself or to static, reified descriptions of it. That is, we sometimes assume that linguistic behavior, observed over a limited span of time, actually represents a more permanent reality. Instead, such descriptions are extreme oversimplifications. What the analyst has observed is a system in flux, and the model the analyst uses should explicitly take into account that fact.

Linguists are all familiar with Saussure's view that languages are self-contained, internally self-consistent, mechanistic systems *òu tout se tient*. Muysken (1995:196) pointed out that under such an assumption, "code-switching is impossible in principle, but there are numerous ways that this fundamental impossibility can be circumvented." He listed four strategies which make CS acceptable:

1. paratactic switching—when there is no tight relation (e.g., of government) holding between two elements
2. if there is equivalence
3. if the alien element is morphologically integrated
4. if there is a neutral site—where a word could belong to either language

Actually, however, no such difficulty exists; CS is only impossible if human brains work like computers, but they do not. Mental categories are innately fuzzy, and people are inherently inconsistent; therefore, CS is not as theoretically impossible as Muysken makes it out to be, if we allow our theories to be as fuzzy as the objects they are attempting to represent.

Given that human behavior (including linguistic behavior and LCP) is inconsistent, theories of LCP should have such inconsistency built-in.

Indeed, Muysken's assessment of the current state of the field is one of "pluralism and the growing recognition that various mechanisms may play a role in different code-switching situations" (1995:196). I claim that various mechanisms are operative not only for different language pairs, but even for the same speaker at different points of the same utterance.

Occam's razor has been cited to justify the goal of finding a single explanation to account for any given linguistic phenomenon. But a realistic model of human behavior will allow for the inconsistency which is part of human nature, by allowing for more than one option for rationales, even allowing them to operate simultaneously, or allowing misunderstandings which result from one participant following one principle and another participant following a different principle.[14] The point here is that models of reality should be just as fuzzy (vague, ambiguous) as whatever aspect of reality they are purported to model. If we assume that the human mind is completely self-consistent, then we will wrongly construct self-consistent theories as well.

If language is merely an informal label for an individual's dynamic speech behavior, then where is it that languages are in contact? Only in the minds of human beings. LCP researchers have puzzled about the speaker's mental lexicon—is there one complex language system stored in the bilingual brain, or two separate systems? This kind of question assumes the Saussurean view of a language as *une système òu tout se tient*, i.e., a closed, self-contained, internally-consistent system, a thing which a number of speakers possess in common; however, as argued above, that kind of system is a fiction, a construct based on folk theories of language. Patterns there are, but no self-contained system. What is in the speaker's mental lexicon? Not languages, just words, and a complex and dynamic network of relationships among those words. A bilingual does not have two separate networks, but one larger, continually evolving network within which the two languages are merely large superclusters of network connections (Grosjean 1995:270–271). There is no boundary between the two.

Within each supercluster, there are a great many words indexed as to language membership; they are tagged. But these tags do not refer to actual entities in the real world; rather, they refer to the individual speaker's folk view of what that language is. In fact, there is a large area between

[14]This has implications for syntactic theory as well: it may be that sometimes syntactic patterns are best represented by tree structures, sometimes by relational networks, and sometimes by linear strings of frequent collocations as lexical routines; the latter analysis would explain the phenomenon of historical reanalysis. Similarly, the inability of syntacticians to agree on a general theory of syntax may be because speakers use more than one set of syntactic principles in organizing speech.

the cores of the two superclusters, where individual speakers may disagree about the tagging of various words.

Absolute and relative competence. The term competence has been used in linguistic literature in two different ways. The most familiar usage is in the Chomskyan sense of the internal ideal language of a speaker, which contrasts with performance—the speaker's actual linguistic output. But when Dabène and Moore (1995:29) referred to different speakers having various degrees of competence in a language, they were using the term in the popular sense of how a speaker's language behavior is evaluated by others, i.e., those members of society who set standards (whether by formal educational standards or other less formal means of promoting uniformity). Thus, we may distinguish between ABSOLUTE COMPETENCE (Chomsky's competence) and RELATIVE COMPETENCE, of which Hymes's (1972) COMMUNICATIVE COMPETENCE is one type.[15]

As a reaction against the notion that CS is a random, unconstrained behavior, sociolinguists have proposed that it is instead a skilled behavior.[16] However, it is possible for that pendulum to swing too far. The notion of the inconsistency of (absolute) competence requires that we admit some degree of seemingly random choices by a speaker; some of these may be misinterpreted by a hearer as meaningful choices, but what speakers signal is not always intentional; all too often, a hearer erroneously infers a speaker's intentions. For example, a speaker's internal competence may include a foreign accent, which is erroneously perceived by a hearer as being a statement of ethnic loyalty. Or, a speaker may select lexical items perceived by hearers as being from a high register, and deemed by them inappropriate to the low conversational context. In such cases, we may speak not of performance errors but of competence flaws, since the speaker's absolute competence is evaluated as inadequate by the community. Such social judgment can motivate a change in the individual's absolute competence to match the expectations of those in his social network; or, if not sanctioned in some way (e.g., by ridicule), it can spread to other speakers and change the network's standards of relative competence-one speaker at a time. The latter scenario would of course lead to linguistic divergence with other social networks.

As I have reexamined the validity of notions such as domain, language, and speech community, I propose also a reexamination the notion of

[15]Gumperz (1982:209) defined communicative competence as "the knowledge of linguistic and related communicative conventions that speakers must have to create and sustain conversational cooperation."

[16]For example, Poplack (1980) demonstrated that intersentential CS correlated with balanced bilingualism in her sample.

societal conventions which allegedly give meaning to CS. The strong form of this hypothesis assumes a great degree of uniformity throughout the speech community (or social network), by which all members of the community or network have the same internal grammar of the meaning of CS. I assume a much weaker form of the hypothesis, one which allows more interpersonal (and even intrapersonal) variation.[17]

For example, Gal (1988:287–288) stated:

> Individuals in an interacting group need not share all varieties but the tacit conventions for using varieties and the interpretive strategies associated with speaking are expected to be known, if not necessarily used, by those who participate together in significant interaction.

I argue that such a conception is too neat. In the real world, each person has his own conventions and only an imperfect understanding of the conventions of others. A person's conventions may include noncommunicative idiosyncracies, judged as competence flaws by others. Such stylistic clashes between individuals bear the seeds of what has been called dialect variation, class variation, and so on, to the point of mutually-unintelligible languages.

In summary, just as Heisenberg's Uncertainty Principle prevents physicists from describing both the position and the momentum of an electron at any given point in time, so linguists are prevented from fully describing language by the fact that language is not an object, but a process through time, and by the inability of the human mind to fully grasp such an ephemeral entity. Furthermore, analytical categories which linguists have made use of in attempting to describe language—society, community, culture, domain, idiolect, competence—have themselves turned out to be as hard to pin down as the electron.

This does not mean, however, that language is unstudiable. It merely means that when linguists (and, most relevant here, LCP researchers) formulate definitions, they ought to take the squishiness of their field of study into account. Any study of human behavior must assume a certain degree of random behavior. On the other hand, we should not dismiss all (or even most) variation as mere inconsistencies. The conversation analysis approach will be extremely valuable in years to come as LCP researchers attempt to find explanations for variation in the dynamic discourse context, including variation in speaker roles and statuses. Neurolinguistic

[17]My perspective, then, contrasts with that of many sociolinguists in a way similar to the position of the Neolinguists (or idealists) of a century ago: "counter to the mechanical conceptions of the Junggrammatiker they put forward not a social principal [sic], but an individual-creative origin that was held responsible for all language change and variation" (Hagen 1988:403–404).

research would also be valuable if instruments are developed to read individual neurons firing (especially without the speakers' knowledge), and if theories are developed which explain the relationships between such neural activity and bilingual phenomena.

1.2.2 Types of LCP

Part of the confusion surrounding LCP was explained by Muysken (1995:188): "In many situations of intense language contact, a number of phenomena involving 'mixing' are going on at the same time: lexical borrowing, code-switching, interference, calquing, relexification, semantic borrowing, L_1 transfer in L_2 learning, possibly convergence." Thus, when confronted with a bilingual corpus, the analyst must attempt to isolate the particular types of LCP he wishes to focus on. The question is, are there boundaries between types of LCP? Or are CS, borrowing, and other types of LCP better defined as prototypes with no sharp boundaries? Furthermore, are there different answers to this question depending on whether we approach it from the speaker's perspective or from the hearer's perspective? And is the analyst's perspective different still?

Muysken (1995:189) defined CS as "the use of two languages in one clause or utterance", and lexical borrowing as "the incorporation of lexical elements from one language in the lexicon of another language". In order to make finer distinctions, he drew upon the notion of LISTEDNESS (DiSciullo and Williams 1989:3–21): "the degree to which a particular element or structure is part of a memorised list, which has gained acceptance within a particular speech community." Although Muysken first claims that listedness is a scalar quality,[18] he then uses it as a binary feature to neatly subdivide both CS and borrowing into two further subcategories. Although listedness is a valuable concept, it is not static; it may vary due to the dynamic nature of competence.

Many other researchers have attempted to draw sharp distinctions between CS and borrowing, resulting in a variety of definitions for these phenomena. But Gardner-Chloros (1995) disagreed with binary criteria by which we can supposedly distinguish CS and borrowing. She listed (pp. 73–74) three criteria which (she claims) fail to distinguish the two:

1. *morphophonemic integration with the surrounding language:* both loans and code-switches can be morphologically and phonologically integrated or unintegrated with the

[18] "We can arrange linguistic elements on a scale running from essentially creative to essentially reproductive" (Muysken 1995:189).

surrounding language, depending on a wide variety of personal and linguistic factors.
2. *native synonym displacement:* there are examples of both loans and code-switches filling "lexical gaps" in the surrounding language and of them adding themselves as a further option to the "native" equivalent.
3. *grammatical category:* although loans are often nouns, all grammatical categories are potentially borrowable; conversely, in many, though not in all, contexts, noun code-switching is statistically the most common kind to occur in the data.

Unfortunately, Gardner-Chloros did not back up these claims with data. In particular, we would like to know (a) what definition of CS would include morphologically-integrated forms; (b) how she would determine whether two words are equivalent; and (c) what definition of CS was she assuming when she claimed that noun CS is the most common kind? Without such specific information, there is unfortunately the danger of circularity in her claims. Moreover, there are other criteria which she did not argue against, such as frequency and equivalence.

Auer (1995:124–126) described four types of CODE-ALTERNATION, including three types of CS and an additional pattern called TRANSFER:

1. **discourse-related CS** occurs when two speakers begin by both using language A, then one of them decides to switch to language B and they both continue using language B.
2. **preference-related CS** occurs when one speaker uses language A and the other uses language B; usually one of the speakers eventually acquiesces and both use the other speaker's preferred language (in which case the label *language negotiation is appropriate*).
3. **an unnamed type of CS** in which bilingual speakers switch between languages within a turn in a way that makes it impossible to decide if language A or B is the "base language".
4. **transfer** (or *insertion*) occurs when a speaker is using language A and inserts a small amount of language B into the middle of his turn, without affecting language choice for the interaction at all. Transfer usually involves a word or another structure from language B.

Taken as a whole, Auer (p. 116) defined code-alternation as "a relationship of contiguous juxtaposition of semiotic systems, such that the appropriate recipients of the resulting complex sign are in a position to interpret this juxtaposition as such" (1995:116). He went on to discuss

key terms in that definition, including contiguous, juxtaposition, semiotic systems, and interpret. By stipulating **contiguous**, Auer ruled out divergent language choices from different parts of a conversation, or from different conversations; he did not mention, however, that contiguous also rules out instances of noncontiguous lexical collocations, such as Weinreich's (1953:50) Yiddish/English example *er hot ge**céjn**dt zajn **majnd*** 'he changed his mind', which Weinreich referred to as a "transfer of an analyzed compound".

It is in his use of the terms juxtaposition and semiotic systems that Auer comes into conflict with Gardner-Chloros. By stipulating **juxtaposition**, Auer intentionally excluded gradual transitions from one code into the other; similarly, by referring to semiotic systems he was excluding the possibility of single-parameter changes being analyzed as code-alternation. He would presumably also exclude multi-parameter changes as well. So, foreign accent, alien morphology, alien syntactic patterns, alien lexical items, or anything short of a complete changeover from one language to the other, do not qualify as code-alternation for Auer.

Gardner-Chloros (1995:77) claimed, to the contrary, that "at its most fluid level, code-switching involves shifting at particular linguistic levels rather than a wholesale transition from one discrete code to another." She urged LCP researchers to study prosodic and paralinguistic levels, not just lexemes (p. 85). With regard to creole speech analyzed by LePage and Tabouret-Keller (1985) as diffuse languages, Gardner-Chloros pointed out that such a complete switch may not even be possible, since there is no such thing as pure code for those speakers. Thus, in contrast to Auer's definition, Gardner-Chloros would apply the label code-switching even to gradual or partial changes in language choice.[19]

To summarize, Auer and Mackey have used the term code-switching in reference to the extreme end of a continuum of LCP, at which particular instances of CS involve a complete changeover from one code to another. Gardner-Chloros, in contrast, has used the term code-switching in reference to a much wider variety of phenomena, and openly wondered whether such a complete and *instantaneous* change of pure codes ever actually occurs. Rather than prolonging the debate by choosing one definition of CS over another, in §1.2.3 I frame the discussion in terms of what kinds of choices lie behind a bilingual's insertion of embedded language (EL) items in matrix language (ML) discourse.

[19]Recall, however, that Mackey (1965) would call this alternation, not code-switching.

1.2.3 Language choice and lexical choice

LCP researchers typically claim that CS among bilinguals conveys the same kind of information as does style-shifting in monolingual speech. For example, Gal (1988:286) stated that "in many multilingual communities the choice of one language over another has the same social function or significance as the selection among lexical variants in monolingual societies." Gumperz (1982:98) likewise claimed: "Code switching signals contextual information equivalent to what in monolingual settings is conveyed through prosody or other syntactic or lexical processes." He thus sees CS as one of many CONTEXTUALIZATION CUES, which he describes as:

> constellations of surface features of message form...by which speakers signal and listeners interpret what the activity is, how semantic content is to be understood and *how* each sentence relates to what precedes or follows....Roughly speaking, a contextualization cue is any feature of linguistic form that contributes to the signalling of contextual presuppositions.

Auer (1984, 1995) continued this line of research, pointing out (1995:123) that like other contextualization cues such as intonation, rhythm, gesture, and posture code alternation is a means by which speakers "make relevant/maintain/revise/cancel some aspects of context which, in turn, is responsible for the interpretation of an utterance in its particular locus of occurrence." Such aspects of context include:

1. speech genre: the larger activity participants are engaged in
2. speech act: the small-scale activity
3. key: the mood...in which this activity is performed
4. topic
5. participants' roles: the participant constellation, comprising speaker, recipient, bystander, etc.
6. the social relationship between participants
7. modality: the relationship between a speaker and the information being conveyed via language

Auer argued that since code alternation works like other contextualization cues, and often along with them, they need a uniform analysis. He also pointed out that since contextualization cues often bundle together, the resulting redundancy enables analysts (and, for that matter, other participants in the conversation) to determine the conversational functions of one cue (e.g., code alternation) based on their understanding of the functions of the other cues.

Scotton (1988) proposed that in cases of CS as an unmarked choice, i.e., the kind of rapid-fire alternation for which Auer (1995:126) claimed that no base language is identifiable, "the speaker wishes more than one social identity to be salient in the current exchange." Auer (1995:130) further claimed, however, that with this type of CS "the individual switches, although they are not 'socially meaningful', nevertheless may have individual discourse-related functions." I do not deny that small-scale language choice is used as a contextualization cue. Nor do I deny that CS as an unmarked choice can be used to index dual social identities. However, in analyzing any particular bilingual corpus, researchers are obliged to demonstrate that what they take to be a meaningful language choice is not merely an instance of expanded lexical choice. The present work should not be seen as an attempt to supplant such lines of research; rather, it represents a necessary methodological precursor to such research.

I assume that in most cases, the transfer of referential meaning from speaker to hearer is of paramount importance. Methodologically, then, it is essential to first exclude from the CS versus borrowing debate those EL words which were used by speakers for reasons of lexical semantics alone. However, it is unclear whether all LCP research in the literature to date has made this distinction. Some LCP researchers seem to have tacitly assumed that most EL items in their corpora are instances of language choice, when in fact for many EL items language choice is not an issue; the speaker is merely choosing the most convenient lexical package for the concept he wishes to convey.

On a large scale, language choice involves the question "Who speaks how to whom, and when, and why?" (Fishman 1965). That is, large-scale language choice involves diglossia or situational CS which correlates with change of topic, participants, and so on. It also involves the decisions bilinguals make to carry on whole conversations in one language or another. Thus, Köppe and Meisel (1995:278) defined language choice as "the ability to select the appropriate language as base language of the conversation according to the interlocutor, the topic of the conversation, or the situational context." Similarly, Hyltenstam (1995:307) used the term language choice to mean "a bilingual speaker's choice of one of his/her languages, with or without in-mixing of the other, for a specific discourse (see Grosjean 1982)."

Choosing a word because of its perceived[20] language membership is an instance of small-scale language choice. Choosing a word or phrase for any other reason I refer to as LEXICAL CHOICE. Small-scale language choice may involve a sentence or two, a phrase, a string of words which do not

[20]Perceived by the speakers, that is, not by linguists focusing only on historical reconstruction.

constitute a syntactic constituent, a single word, or even a bound morpheme. Despite all the disagreement about where to draw the line between CS and borrowing, I believe there is a general consensus that borrowing represents lexical choice while CS constitutes language choice.

Grosjean (1995:261–263) has proposed that any given bilingual speaker fluctuates along a continuum from monolingual mode to bilingual mode depending on the situation. In monolingual mode, bilingual speakers "deactivate, as best they can, the other language". In bilingual mode, however, both languages are activated. When speakers in bilingual mode produce stretches of speech exhibiting EL morphosyntax (as in Myers-Scotton's EL Islands), we may say they have switched to EL MODE; otherwise, they are in ML MODE. I believe that by using this more precise set of terms, it is possible to avoid fruitless disputes about where to draw the line between code-switching and other language contact phenomena such as borrowing.

Returning to Gal's (1988:286) claim that the choice of one language over another has a certain social function, there are two points to be made. First, the statement assumes that there is such a thing as one language versus another language in the bilingual speaker's brain. As I have stated previously, I disagree with such a claim. Second, what CS researchers (at least, in the first and third waves of research) really want to know is, what makes speakers choose to use an EL item **as a member of EL**, not for other reasons (e.g., denotation, connotation, and collocation, and the aesthetics of its phonological form). That is, a word has many features which may make it a desirable choice; etymology is only one. In order for CS research to determine why speakers make etymology-based choices, it is first necessary to exclude from the data all words chosen for other reasons.[21] This is the focus of chapter 4.

In trying to determine whether a single EL word, or root, or even a string of EL words, represents language choice or lexical choice, phonological integration is an indicator which must be examined carefully. Among the many English words which are etymologically French, some are more phonologically assimilated than others. Thus, even established loans may remain unassimilated to some degree. Similarly, depending on a bilingual speaker's EL competence, a stretch of EL words marked by ML phonological features may still be considered a complete change of language choice if that speaker (in monolingual mode) speaks EL with what monolingual EL speakers call a foreign accent. While some speakers may maintain two distinct phonological systems, others may allow the two to converge.

[21]By choice, I do not necessarily mean conscious choice, or intentional choice, since CS may result from language choice performance errors.

Similarly, a bilingual's competence in his L_2 with regard to syntax may be different from the syntactic patterns of monolingual L_2 speakers—that is, his (idiosyncratic) L_2 syntax may be partially converged with that of his L_1. If that is the case, then EL words in ML syntactic patterns are not necessarily evidence against a CS analysis; syntactic patterns in L_2 stretches must be compared not with the syntactic patterns of monolingual L_2 speakers, but with patterns of the same speaker when speaking L_2 in monolingual mode. For example, in an interview conducted completely in English (or rather, Englishes), the German figure skater Mandy Woetzel complained about her partner, *He has always everything to explain!* (for standard American English, *He always has to explain everything!*). If we hold to the simplistic assumption that discrete languages exist, we might say this sentence has English lexemes filling a German syntactic structure, with the auxiliary in second position and the direct object preceding the final verb. But without further investigation of Ms. Woetzel's idiolectal variety of English, we cannot determine whether this sentence was a performance error on her part (i.e., her English competence is really not so dissimilar from standard English) or whether this structure is a feature of her English competence (which most English speakers would consider flawed competence).

Furthermore, if a theory of syntax is assumed in which syntactic structures are projected from features of the lexemes themselves, then L_2 lexemes can be borrowed with syntactic structures in tow. Therefore, a brief sequence of EL words in EL (and not ML) order could still be considered lexical choice rather than language choice; some examples of this are examined in chapter 5. Conversely, an EL word chosen because the speaker considers it a member of EL is still an instance of language choice even if the surrounding ML words show no evidence of ML syntax. I, therefore, disagree with Poplack and Meechan's (1995) assumption that EL words in a completely ML syntactic frame are borrowings (if, that is, borrowing is the same as lexical choice).

I have argued that both phonological assimilation and syntactic integration fail as indicators of language choice versus lexical choice. What of morphological integration? Poplack (1980) posited the FREE MORPHEME CONSTRAINT, which states that a code-switch cannot occur between a bound morpheme and an affix; but I argue in chapter 4 that EL morphemes morphologically integrated into ML may be instances of either language choice or lexical choice. As for EL items exhibiting EL inflectional morphology in place of ML morphology, I follow Myers-Scotton (1993b) in assuming that such EL items (even as short as a single word) are instances of what she calls EL ISLANDS (and what I will call EL MODE). Note, however, that in order to make such a determination there

must be morphological elements in the two languages which are equivalent (or at least overlapping) in function.

The model I am presenting here, therefore, is a continuum with lexical choice (borrowing?) at one end and a complete change of language mode (code-switching?) at the other. In between these two extremes are instances of incomplete language choice, which do not involve a change of language mode.[22] Just as lexical choices can fail to be completely integrated, so also language choices may consist of one or more markers (e.g., EL verb roots, tags, or phonological ethnicity markers) rather than a complete change of mode.

If, as argued previously, a borrowed word may project a syntactic structure into which other borrowed words may be placed, the question arises: how does the analyst differentiate between this and instances of EL mode, in which EL morphosyntactic procedures are completely activated? Aside from EL morphology, which provides no help for the WT/Malay corpus, and the obvious indicator of how long the EL stretch is, I show in chapter 5 that relative frequency can indicate whether an EL word in an EL syntactic structure should be viewed as projecting that structure (i.e., the structure is borrowed) or inserted into it (i.e., the structure represents a change to EL mode).

The present study, because of the nature of the WT/Malay corpus, focuses on small-scale language choices. Many recent sociolinguistic studies have demonstrated the benefits of microanalysis in accounting for CS patterns. But the same attention to details does not seem to be reflected in attempts to differentiate CS from borrowing. Instead, the data are reduced to counts of switched nouns, NPs, VPs, and so on, or to counts of different types of syntactic contexts in which lone EL nouns occur, and statistical methods are then applied. The present study, in contrast, examines each EL word or sequence of words one by one, under the assumption that each context is unique and, therefore, not reducible to a structural formula.

1.3 The individual speaker perspective

Given a set of assumptions about the nature of language and LCP such as those outlined in §1.2, there are still many approaches one can take to the study of a bilingual corpus. Different bilingual corpora are more or less suited to different research questions, whether due to the linguistic

[22]Muysken (1987) referred to this middle ground as various degrees of integration of borrowed elements, while Gardner-Chloros (1995) apparently thought that both borrowing and CS are found in this zone. As stated previously, I do not intend to quibble about terminology.

features of the languages involved, the situations in which the data were recorded, the technology used for recording the data, or other factors. In this section I describe the perspective I have taken in the present work, which may be different from other equally valid perspectives.

Geerts's (1988) review of LCP research distinguished studies of macro-contact (on the level of speech communities) and studies of microcontact (on the level of individual speakers). As noted in §1.2, Gardner-Chloros (1995:71) complained that LCP researchers have drawn invalid conclusions about CS because they have failed to keep the two types of studies conceptually distinct; specifically, they have assumed that average group behavior is representative of individuals' CS patterns (cf. LePage and Tabouret-Keller 1985:9). For example, Scotton (1988) defined four different types of CS which occur in various types of speech communities and social contexts, but Gardner-Chloros (1995:84) claimed that there is also variation within these communities and types; in particular, she claims that CS as an unmarked choice covers a range of patterns which vary widely both in quantity and quality, from violating grammatical constraints, to switching very frequently, to just using tag switching; and that "CS is a highly individualistic phenomenon". This cannot mean, however, that an individual's CS patterns (or any other linguistic or behavioral patterns) are unrelated to the patterns of the people he has regular contact with or desires to emulate or be accepted by.

In the present work, I focus on microanalysis (in terms of individual speakers and individual lexemes) and all but ignore macroanalysis. In doing so, I am not implying that macroanalysis of CS patterns (e.g., on the level of the speech community) is invalid. Rather, I claim that macroanalysis with no regard for microanalysis is invalid. In order to inform a microanalysis, an informal assessment on the macrolevel is essential; and the microanalysis is incomplete if not followed up by macroanalysis (and, as long as funding holds out, more microanalysis, and so on). My informal macrolevel assessment of the interplay of WT and Malay is presented in chapter 2.

In §1.2.2, I referred to Muysken's (1995:189) usage of the notion of listedness. I am revising his notion to refer to an individual's (dynamic) competence (PSYCHOLINGUISTIC LISTEDNESS), rather than speech community acceptance (SOCIOLINGUISTIC LISTEDNESS). Psycholinguistic listedness refers to entries in the mental lexicon, including words, lexical phrases, idioms, and collocations. As with sociolinguistic listedness, psycholinguistic listedness refers to a scalar quality, even though in practice I use it to make a binary distinction between default and non-default Malay items.

What is it, then, that we need to know about individual speakers? First, we need to study a speaker's lexical habits. Auer (1995:121) stated: "In

order to pinpoint the conversational meaning of such a case of code-alternation, we need to know about the 'episode-external' preferences of speakers for one language or the other, or about the community norms for that particular kind of interaction." In the present work, however, the focus is not on norms for a particular kind of interaction, but norms for the lexical expression of a particular concept-what may be termed LEXICAL PREFERENCES.

Second, in attempting to explain a speaker's language choices we need to focus not on what is true of the situation, participants, and so on, but what the speaker believes to be true.[23] Thus, rather than saying that some types of CS are based on the hearer's linguistic preferences or competences, I say that CS may be based on what the speaker **believes** about the hearer.

Third, we need to allow (or even expect) that the speaker's language behavior will vary during a conversation due to many factors, including the dynamicity of absolute competence, the well-known phenomenon of performance errors, the fluidity of participant statuses and roles,[24] feedback (positive or negative) from other participants, and more. In any individual's language behavior, there is a complex constellation of causative (or rather, persuasive) factors involved.

Should we then abandon all hope of describing an individual's competence and focus instead on statistical tendencies within a larger group or network of speakers? While this is a valid approach taken by many LCP researchers, it is not the approach of the present work. Rather, since language only exists in the minds of individuals, I focus here on the (absolute) bilingual competence of individual speakers, realizing all the while that such competence is dynamic.

1.4 Conclusions

As previously stated, Muysken's (1995:196) assessment of the current state of the field was one of "pluralism and the growing recognition that various mechanisms may play a role in different code-switching situations." For example, Auer (1995:117–124) argued against the extreme, mutually-exclusive notions that (a) CS is meaningful only by virtue of the contrastive nature of the switch, without regard to the direction of the switch (i.e., language A to language B or the other way around), and that (b) CS is **never** meaningful by virtue of the contrast alone, but only by virtue of the social meaning attached to the languages involved. He presents

[23]Starosta (1991) has already shown this to be true in syntax.
[24]See Mioni 1988:173 and Breitborde 1983.

data demonstrating that the meaning of CS sometimes derives from one principle, sometimes from the other.

Similarly, I do not have the goal of finding the one best explanation for any particular linguistic phenomenon; I do not assume that a good argument for one hypothesis (e.g., of syntactic processing) necessarily implies that an alternative hypothesis is wrong, especially when attempting to analyze human behavior. It is possible that the human brain is capable of accessing a variety of principles simultaneously (e.g., tree structures, relational networks, left-to-right sequences, memorized strings of words). Occam's razor is only valid if we take all the data into account-which is beyond our ability, and probably always will be. Therefore, although LCP researchers (including myself) have drawn a wide variety of conclusions from a wide variety of language pairs and corpus types, I think we are still far from a comprehensive theory of language contact phenomena. The present work is an admittedly limited addition to our knowledge of the subject.

In this chapter, I have outlined the main goals of the study, and made explicit my own beliefs and assumptions about the nature of language and language contact phenomena (LCP). I rejected the structuralist view of LCP as representing deviations from the pure structure of a language. Research in the area of language variation has made it clear that there is no such pure norm from which to deviate; rather, there is always variation from speaker to speaker and even within a single speaker over time. I have described the confusion in the literature surrounding the use of such terms as code-switching (CS) and borrowing, and the many attempts at differentiating these two phenomena. I believe that much of the confusion derives from basing definitions of these terms on a structuralist view of language, whereas language contact is one area where structuralism most obviously fails. I concluded that attempting to redefine these terms would only breed more confusion, and decided to frame my own discussion in terms of language choice versus lexical choice, and in terms of language mode.

In chapter 2 of this work, I provide essential background to the research to be presented in chapters 4 and 5, by examining the language ecology of West Tarangan as well as giving an overview of linguistic differences between WT and Malay. In chapter 3, I explicate the methodology I used as well as describe some characteristics of the resulting corpus. In chapter 4, I present the results of extensive lexical analysis of the corpus and propose a distinction between default and non-default Malay, as categories dividing a continuum whose endpoints are necessary Malay and gratuitous Malay. In chapter 5, I follow a three-cycle approach toward the analysis of the Malay stretches involving non-default Malay lexical units.

The main goals of the present work are to describe the language mixing which occurs in my corpus of fifteen hours of recorded conversation and to demonstrate the need to examine particular words, particular speakers, and particular discourse contexts in order to explain that mixing in a psycholinguistically realistic way.

2

West Tarangan: An Island in a Sea of Malay

3rd Fisherman. Master, I marvel how the fishes live in the sea.
1st Fisherman. Why, as men do a-land: the great ones eat up the little ones.

William Shakespeare, *Pericles, Prince of Tyre*.

...these wild, inhospitable regions, doomed for ages yet to come to hopeless barbarism....

Alfred R. Wallace, *The Malay Archipelago*
(writing about the Aru Islands)

Before proceeding with an analysis of the data which form the primary focus of this work, in §2.1 I examine the context from which the data were extracted, making use of Haugen's (1971) framework of language ecology. Then in §2.2 I discuss linguistic similarities and differences between West Tarangan and Dobo Malay which are relevant to the present analysis of language contact phenomena.

No extensive research had been carried out on the West Tarangan language and culture until July 1987, when I began work there as a staff member of Pattimura University (UNPATTI) in Ambon, as part of the UNPATTI-SIL cooperative agreement. Between stays in Ambon, I resided with my wife and children in the villages of Doka Timur and Kalar-Kalar, and then in Dobo, until July 1991, except for a six-month hiatus in

America. A number of papers, published and unpublished, resulted from those first four years of fieldwork (Nivens 1989, 1991a, 1991b, 1992a, 1992b, 1992c, 1993a, 1993b). The present work is based on fieldwork conducted between September 1995 and June 1997.

2.1 Language ecology of West Tarangan

Haugen (1971) defined language ecology as "the study of interactions between any given language and its environment", and defined that environment as "the society that uses it as one of its codes". Haugen concluded his paper with a set of ten questions to be answered for every language in a description of its language ecology as follows.

1. What is its classification in relation to other languages?
2. Who are its users?...with respect to locale, class, religion or any other relevant grouping
3. What are its domains of use?
4. What concurrent languages are employed by its users?
5. What internal varieties does the language show?...not only regional, but also social and contactual dialects
6. What is the nature of its written traditions?
7. To what degree has its written form been standardized, i.e., unified and codified?
8. What kind of institutional support has it won, either in government, education, or private organizations, either to regulate its form or propagate it?
9. What are the attitudes of its users toward the language, in terms of intimacy and status, leading to personal identification?
10. [What is] its status in a typology of ecological classification[?] ...where the language stands and where it is going in comparison with the other languages of the world?

In this section I selectively discuss those questions which are particularly relevant for the present work, focusing primarily on the West Tarangan language (WT), but also touching on the varieties of Malay spoken by West Tarangan speakers.

2.1.1 Concurrent languages and domains

A number of languages other than West Tarangan are included in the repertoires of many WT speakers, the most prominent of these being Dobo

Malay (DM). This is the language of wider communication in Aru, reportedly having supplanted West Tarangan itself in that role at some time in the past 100 years. It is my impression that the majority of West Tarangan speakers are fairly fluent in Dobo Malay, though apparently in the past (and for the current oldest generation) this has been more true of men than of women, both in the sense of a higher number of bilinguals and in the sense of the levels of bilingual proficiency. It is also more true for the youngest generation than for the oldest, though I know of only one village (and a small one at that) where children do not speak West Tarangan at all. My informal observation has been that most WT speakers either use or avoid DM based on who is present. One purpose of the present work was to test and either reject or refine that observation.

Standard Indonesian Malay (SIM)—or a variety of Malay approximating SIM—is used by WT speakers in well-defined domains of life, and thus is in a diglossic relationship with West Tarangan and Dobo Malay. It is used throughout nearly all Christian worship services, and in the sermon portion of Islamic worship services. SIM is also propagated in schools; there are primary schools (grades 1–6) in every Tarangan village, a junior high school (grades 7–9) in two villages, and grades K–12 in Dobo. Bilingualism in this standard form of Malay seems to be merely receptive for most WT speakers, and even then may be at a rather low level of proficiency, depending on the topic. WT speakers refer to both DM and SIM as *Bahasa Indonesia* when speaking Malay, and as *Maláy* when speaking WT, apparently viewing the two varieties as high and low registers of the same language.

Arabic is used in Muslim worship services, Quran recitals, and other Islamic events. Other languages may be known to some degree but not actually used; for example, English is taught as a required subject in schools. Other languages of Maluku (e.g., Kei, Fordata, Selaru, Luang, Kisar) may be known to certain WT speakers because of friendships or intermarriage.

West Tarangan is the common language of daily life in the home and community. It may be heard in the final portion of church services, and sometimes even in the sermon. It is sometimes used in personal letters and other informal writing, such as graffiti, and even in official meetings of village-level officers of the Indonesian government system. Visitors from East Tarangan attempt to use West Tarangan when visiting. In all these domains, Malay varieties may also be used.

There are a few domains, however, which are exclusive to West Tarangan, generally those closely associated with traditional WT culture as they see it. For example, traditional songs maintain what are believed to be the original words, resulting in archaic, obscure language; songs composed more recently, whether in a traditional form of music or a more modern pan-Indonesian style, may have a word or two of Malay but are

overwhelmingly West Tarangan. Events referred to by the Malay word adat (traditional religion, rituals, ceremonies, law, intended to maintain order between humans and between humans and spirits) are conducted solely in West Tarangan, and in an archaic form of the language, if one is speaking to ancestor spirits. In addition, many WT speakers have expressed a purist attitude against inclusion of Malay words in published WT literature.

Thus, it seems that DM can be used in any of the casual, daily-life domains in which WT is used-but speakers range along a continuum from avoiding Malay insertions to speaking entire discourses in Malay. It is my hope that the present work can be used as a benchmark to measure (at some future time) whether the use of DM has increased at the expense of WT.

2.1.2 Users of WT and DM

Nearly all WT speakers are users of West Tarangan by direct inheritance (i.e., belong to indigenous kin groups). In addition, however, there is some intermarrying with other Aruese groups or even non-Aruese Indonesians; some merchants from elsewhere in Indonesia learn the language; and some kin groups claim to have come from elsewhere long ago, but are now fully incorporated into West Tarangan society and speak West Tarangan natively, retaining no social ties with their ancestral homes.

According to Wallace (1869:327), Dobo in 1857 was the trading settlement of the Bugis and Chinese, who annually visited the Aru Islands. Nearly all of these traders arrived in January with the west wind and left in July with the east wind; Wallace estimated the population of Dobo at its peak that year to be somewhat over 500, mostly "Chinese, Bugis, Ceramese, and half-caste Javanese, with a sprinkling of half-wild Papuans from Timor, Babber, and other islands" (p. 335). He observed fifteen large boats from Macassar, and perhaps a hundred small boats from Seram, Goram, and Kei (p. 368). Malay was the lingua franca of this multi-ethnic community. In a study of bilingualism among the Dobel people of Aru, Hughes (1995) has summarized the spread of Malay in Aru as follows:

> At the time of the visit of the British naturalist, Alfred Wallace, in the 1850s, the inhabitants were nearly all monolingual in their native tongues (Wallace, 1869). Most contact with non-Aru languages has been since the late nineteenth century, through (a) trade, (b) Christian and Muslim proselytism, and (c) government administration, initially Dutch, then Japanese (1942–45), then Indonesian. All such contact has been through the medium of Malay/Indonesian.

Although many Aruese people who never traveled to Dobo were insulated from contact with Malay, Wallace (1869:367–68) mentioned that the Bugis and Seramese traders often settled in their villages and married native women. Even without intermarriage, daily contact with immigrant traders made some small degree of Malay use more common in many villages. At present, according to Hughes (1995), all Dobel speakers except preschool children are bilingual to some extent in DM. I believe the same to be true of WT speakers.

Haugen (1971:25) suggested locale, class, and religion as possibilities for relevant grouping of users. Other variables used by various sociolinguistic researchers include age, educational level, occupation, and gender. Of these, locale, age, occupation, and educational level seem most relevant to the present study of bilingual behavior in WT society.

Locale. West Tarangan has roughly 8,000–9,000 speakers, which is a typical size for the Maluku region of Indonesia.[25] It is the largest language group in the Aru Islands, and has more speakers resident in Dobo town (the largest town and administrative capital of the islands) than any other Aruese language group. Most WT speakers live in the western part of Tarangan island, about 50 miles south of Dobo. Transportation between Dobo and Tarangan island is by boat and is often dangerous. Emigrants (and their descendants) include about 1,000 in Dobo;[26] fewer in Ambon, the provincial capital; and fewer still in other parts of Indonesia—in particular, Sorong, Jayapura, and Surabaya (see figure 2.1). In general, most West Taranganese in Dobo and Ambon marry other West Taranganese and continue to use the language. However, I have no data on language maintenance elsewhere. Reportedly the language of Aruese ethnic identity for Aruese elsewhere (e.g., Jayapura) is West Tarangan—perhaps because more West Taranganese have emigrated than other Aruese groups have. Some Aruese claim that WT has been used in the past as a language of wider communication within Aru, and that is still the case on Tarangan island itself: East Taranganese speak (or attempt to speak) WT, but the reverse is not the case. Even on Tarangan island, differences in locale have relevance for language use: some villages are near the site of an Indonesian Navy base established in 1991, some are near to other dialects of

[25]At the time of this study, the Maluku region was a single province. It has since been divided into two provinces, Maluku and North Maluku. Half of the approximately 130 living languages of Maluku region have between 1,000 and 12,000 speakers; all but four have fewer than 50,000, and only Ambonese Malay has over 100,000. One-third of the languages of this region have fewer than 1,000 speakers; the median language group size for the region is approximately 2,500 speakers (Grimes 2000).

[26]The total population of Dobo is about 13,000; the population of Aru is about 63,000 (Pulau Pulau Aru Dalam Angka 1996).

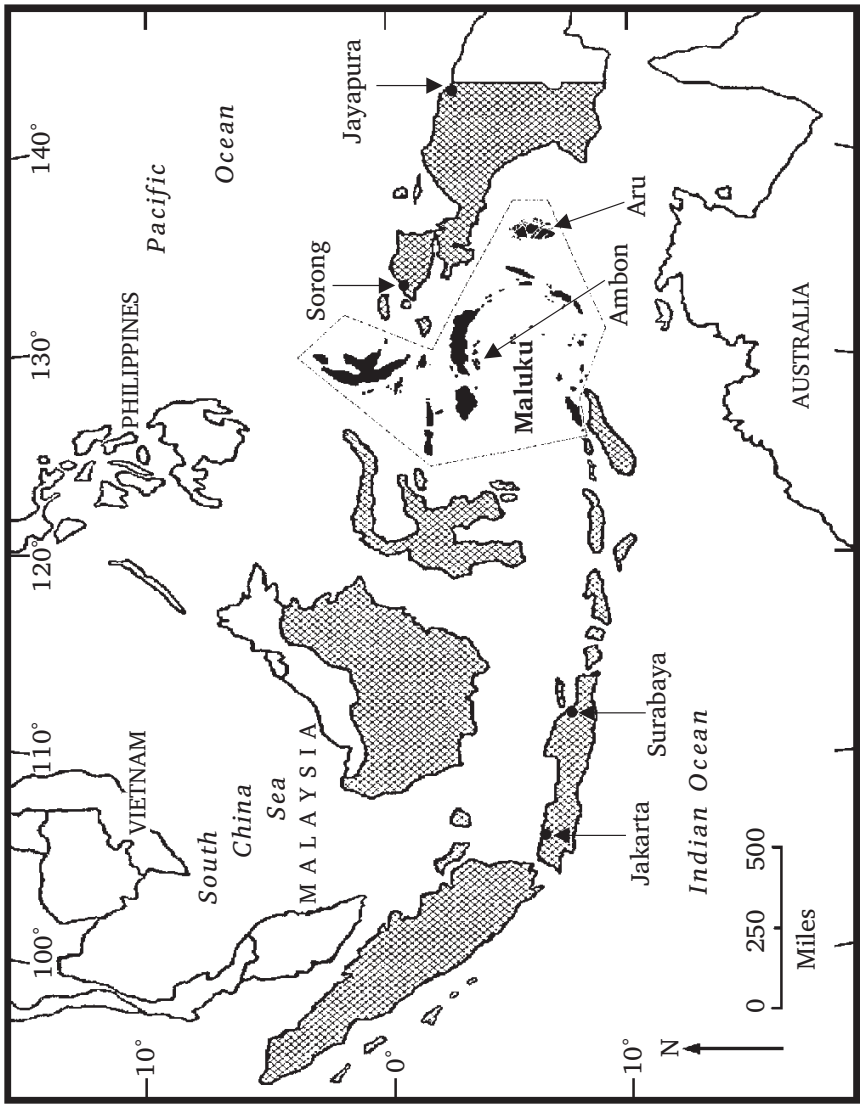

Figure 2.1. The Maluku region of Indonesia
(adapted from Taber 1996:vii)

West Tarangan, and some are near neighboring languages (East Tarangan, Karey, Manombai).

The Aru Islands are part of the Maluku province, whose capital city is Ambon. Aru is about 100 miles from north to south, and only about 100 miles from the coast of Irian Jaya. Although this is considerably closer than the 400 miles to Ambon, the Aruese consider the cultural ties to Ambon to be much stronger than any ties to Irian Jaya. Aru constitutes a single *kecamatan* (administrative district) within the *kabupaten* (regency) of Southeast Maluku, whose capital is Tual in the Kei islands. Aru is also a single klasis (church district) in the structure of the Protestant Church of Maluku (GPM). Since there are at least a dozen indigenous languages in Aru, neither government nor church officials in Dobo are strongly motivated to learn any of the local languages; this reinforces the use of Malay as the sole vehicle for inter-ethnic communication.

Speakers of Dobo Malay have the advantage of considering themselves speakers of Indonesian, even though linguistically the differences between DM and SIM (lexical, phonological, morphological, and syntactic) are so great that mutual intelligibility is far from automatic. Still, the users of DM feel a certain solidarity with other speakers of Indonesian from one end of the country to the other. Thus, the locale for DM can be seen as comprising the entire country, or the province of Maluku, or just the Aru Islands—depending on the attitude of the user.[27] In addition to Aruese, residents of Dobo are mainly Chinese, Buginese, Butonese, and people from other parts of southern Maluku.

Religion, education, occupation, and age. As WT culture has developed over the last century or two, the use of Malay has become ever more important. According to Kruger (1959:32), in 1692 there were 100 baptized Christian converts in Aru. But as recently as 1857, Wallace noted (1869:367) that there were only three or four villages in Aru "where schoolmasters from Amboyna reside, and the people are nominally Christians, and are to some extent educated and civilized". None of these were on Tarangan island. If Wallace had visited Tarangan, he would have found no signs of any imported religion, just as in the small village (Wanumbai) which he visited; the first Taranganese conversions to Christianity occurred in the 1880s, and some had no converts yet even in 1950. However, today there is a strong GPM (Protestant Church of Maluku) presence in every WT village; as mentioned in §2.1.1, SIM is the only language used in most GPM activities, even if only WT speakers are present.

[27]Similarly, WT speakers can consider themselves to speak the same language as East Tarangan, or southern Aru, or all of Aru, depending on their attitudes toward other Aruese.

Most WT villages are 100 percent Protestant. Only five of the eighteen WT villages contain Muslims; of these, Protestants still constitute an overwhelming majority in all but one—Lutur, which is almost evenly divided among Protestants and Muslims. Islam came late to West Tarangan, apparently making its first converts in the 1950s. In contrast, the DM-speaking population has a larger minority of Muslims. Wallace (1869:366) saw a small mosque in Dobo in 1857; today, Muslims constitute a sizable minority in Dobo. Roman Catholics are also present in Dobo as well as in two WT villages. In Aru as a whole, about 32 percent of the population are Muslims, 55 percent Protestants, and 12 percent Roman Catholics (Umat Beragama 1992:56).

Every WT village has an elementary school; most such schools are affiliated with the GPM. Like other elementary schools in Indonesia, SIM is both the medium of instruction and the primary subject of instruction. Many WT parents have adopted the policy of teaching their children Malay before teaching them WT, in order to increase their chances of success in school. The policy seems to be working; in 1996 there were currently many more WT students at Pattimura University in Ambon than there were in 1987, when I began my fieldwork. Interestingly, in contrast to the previous generation of university students, the current generation have maintained their ability to speak WT.[28]

Education is desired both for its prestige value and because it is seen as a means to obtaining cash income. A number of WT speakers have become schoolteachers or government employees, positions which necessitate the use of SIM. Other means of obtaining cash require the use of Dobo Malay, such as laboring for non-WT speakers in Dobo, the sale of crops and other items harvested from land or sea, or the sale of boats. The purchase of goods not produced in WT villages also requires DM. Occupations and roles which require the use of SIM include village-level government and church positions.

Age as a sociolinguistic variable is obviously related to education, since more WT children receive secondary education now than in previous generations. As in most minority-language societies around the world, young Taranganese are more educated and more bilingual in the national language than are the old. And as with many minority languages, some of the younger generation are abandoning the language of their ancestors. More is said about this in the Epilogue.

[28]Craig Marshall (personal communication) similarly reports what he calls a lost generation of Fordata speakers, between 15 and 25 years old, who have a limited proficiency in Fordata. He believes this is due to the vernacular being strongly prohibited by schoolteachers in previous years.

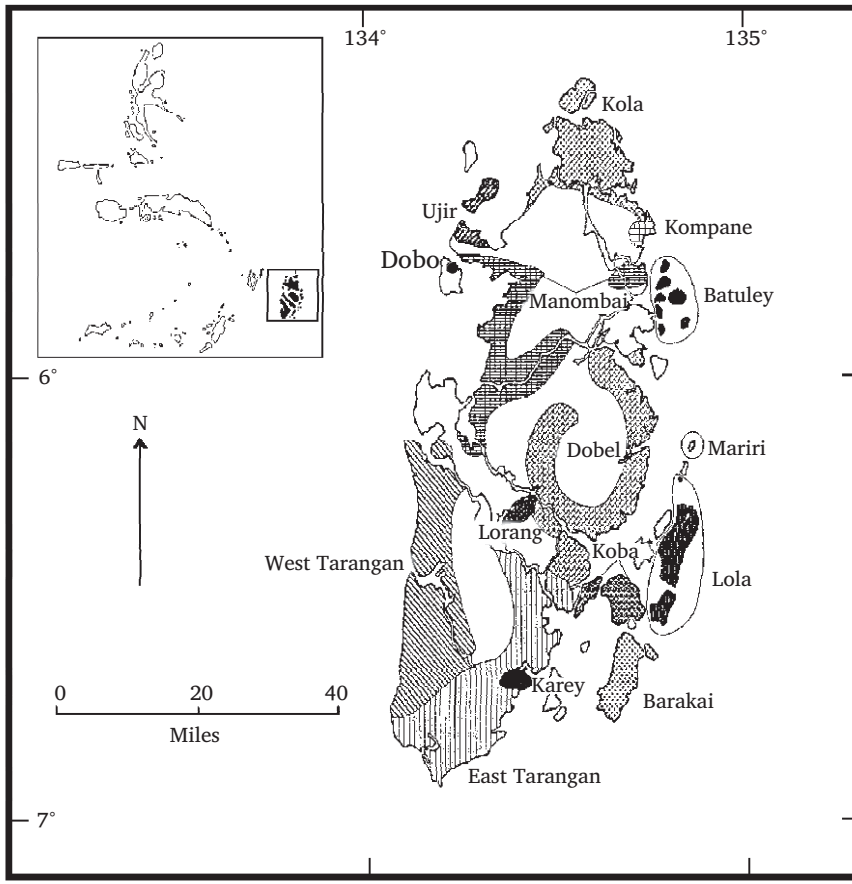

Figure 2.2. Indigenous languages of Aru (adapted from Taber 1996:93)

2.1.3 Linguistic classification of WT and DM

West Tarangan is one of at least twelve languages[29] in the Aru Islands, being spoken in 18 of 123 villages there (see figure 2.2). All languages indigenous to Aru are closely related to one another, and may well constitute a distinct subgroup within the putative Central Malayo-Polynesian subgroup of Austronesian languages. However, further research is necessary to determine whether the smallest subgroup containing all the languages indigenous to Aru contains other languages as well.

Dobo Malay has never been researched, but it is very similar to Ambonese Malay. Collins (personal communication) believes it may be even more closely related to Banda Malay, but unfortunately very little research has been carried out on Banda Malay either.[30] Ambonese Malay has been described from a number of different perspectives by quite a few researchers, including van Hoevell (1877), Collins (1980, 1983), van Minde (1997), Grimes (1991), Steinhauer (1991), Nivens (1994), and Tjia (1997). Malay is a Western Malayo-Polynesian language. Collins (1995) has suggested that dialects of Malay in eastern Indonesia show traces of a close relationship to Brunei Malay, with influences from Makassar Malay as well.

With regard to a general linguistic typology, both WT and DM have nominative-accusative syntax, SVO basic word order, little affixation, and are generally analytic (i.e., affixes are easily segmentable). More is said about the morphology and syntax of both languages in §2.2.

2.1.4 Internal varieties and standardization

As for regional varieties, there are eighteen or nineteen West Tarangan villages, each with its own idiosyncracies (mainly lexicon, also phonology, morphology). None of these regional dialects is considered a standard or prestige dialect; rather, each village considers its own way of speaking to most accurately reflect the speech of the ancestors. These village dialects can be grouped on linguistic grounds into two subgroups of nine villages each, which I refer to as WTA and WTB (see figure 2.3).[31] Between the two subgroups there are difficulties of mutual intelligibility, and it may be more accurate to call them separate languages rather than separate dialects of the same language.

---

[29]Languages, that is, as measured in terms of mutual intelligibility.

[30]A word list of Banda Malay was published by Stokhof (1982).

[31]WTA includes the villages of Kabalukin, Kalar-Kalar, Feruni, Ngaiguli, Fatural, Ngaibor, Marafenfen, Popjetur, and Gaimar. WTB includes Juring, Hokmar, Lutur, Rebi, Lor-Lor, Jeroil, Doka Barat, Laininir, and Doka Timur. Gaimar, however, is in the process of shifting from WTA to WTB (Nivens 1996). It is possible that the people of Jerukin village on Maikor island also speak a variety of WTB.

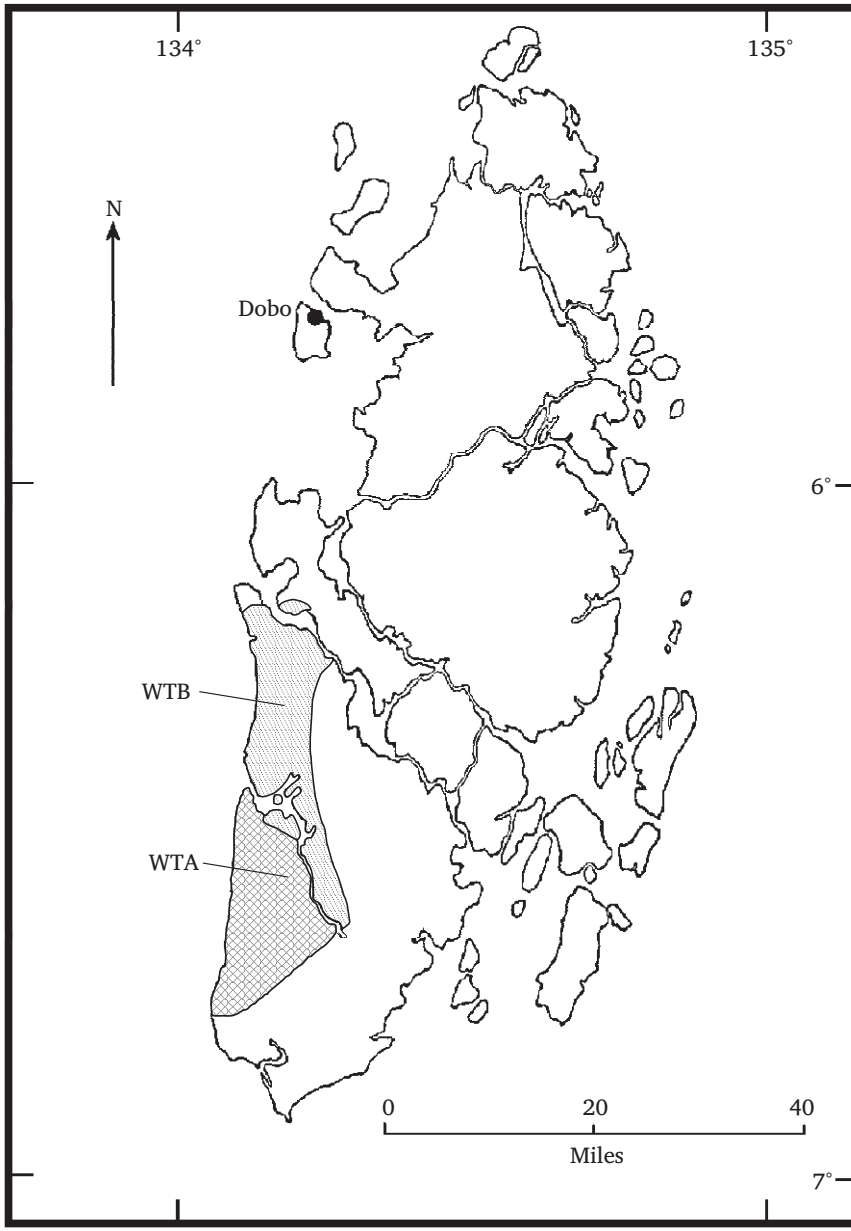

Figure 2.3. WTA and WTB (adapted from Taber 1996:102)

In the 1850s, most speakers of Malay in Dobo were traders who spent only six months of each year there; since they were not permanent residents of Dobo, the Malay of Dobo would have been a very diffuse variety. Today, however, DM has spread into all villages of the Aru Islands, and it is a much more focused variety-more focused, indeed, than either WTA or WTB. According to the WT speakers I asked, DM also has internal variation according to the language background of the DM speaker. Thus, there is reportedly a Chinese variety of DM, a Dobelese variety of DM, a West Taranganese variety of DM, and so on. Further research is needed to verify this claim and to determine what levels of language are involved; it is possible that the differences are merely phonetic and intonational.

Owing to the influence of church and government, Dobo Malay today also has a high register which approximates SIM in phonology, syntax, and morphology. In fact, a variety of Malay registers and styles are evident in the WT/Malay corpus; however, analyzing the interplay between these is beyond the scope of the present work. In general, I use the term Dobo Malay to refer to the low register which is far more frequent and more unlike SIM.

2.1.5 Written traditions

Prior to the arrival of Protestant schoolteachers in Aruese villages, literacy in any language was a foreign concept to most of the people of Aru. Wallace stated, "My very writing materials and books are to them [the men of Wanumbai village] weird things..." (1869:359). Because of the great similarity of the sounds of West Tarangan to the sounds of Malay (see §2.2.1), once people were taught how to write Malay in school they did not need to be taught how to write West Tarangan. For this reason, there has for several decades been a tradition of informally writing West Tarangan, including songs, personal letters, tattoos, boat names, and graffiti. In 1991, dozens of WT speakers (young and old) in Dobo proved their ability to write WT by presenting me with over 100 handwritten folktales in the hopes of having them published. Some of these were published, along with some other literature. However, literacy in WT is clearly not as strong a felt need for WT speakers as literacy in Indonesian, as evidenced by the lack of publications produced by WT speakers alone.

To my knowledge, Dobo Malay itself is seldom written. However, when it is thought of as a local variation of the national language Bahasa Indonesia rather than a language competing with it, it enjoys solidarity with a massive and long-standing written tradition. In many of the domains where WT speakers encounter Malay, they encounter it in both spoken and written modes.

2.1.6 Institutional support

The Indonesian government promotes SIM, claiming also to support preservation of vernaculars: Nababan (1985:6) pointed out that according to the Constitution, "those vernaculars that are properly maintained by their speakers" will be protected by the government. But as recently as twenty years ago, teachers in both government-owned and church-owned schools reportedly used to discipline children for speaking the vernacular. However, in recent years the Indonesian government has renewed its support for vernacular languages in its Muatan Lokal program, in which local cultures are taught in the elementary school curriculum.

As previously noted, the Protestant Church of Maluku (GPM) has traditionally propagated Malay—usually the SIM of western Indonesia, but at times Ambonese Malay. In the last few years, however, GPM leaders in Aru have expressed enthusiasm about using vernacular materials as a supplement to Malay materials.

Thus, there has been an increase in the official support for the maintenance of WT both by the government and by the GPM. Dobo Malay, however, has no such support, being viewed as a variety suitable only for informal settings.

2.1.7 Attitudes

The matched guise test (Lambert et al. 1960, Lambert 1967) typically used for measuring language attitudes is not practical for the West Tarangan situation because it requires a perfectly bilingual speaker of two language varieties, one whose identity is unknown to any of the test subjects. This may be possible for large metropolitan languages like French and English, for which the matched guise technique seems to have been used successfully, but not for small communities like West Tarangan villages.

One goal of the matched-guise technique is to determine which language is more associated with such qualities as leadership, prestige, success, intelligence, and the like; but for West Tarangan and other such communities it may be more appropriate to determine which language(s) are considered comfortable, i.e., which language(s) make the speaker or reader feel most at home.[32] In this regard, it is essential to consider at least two different functions of language, namely language as tool and language as identity marker. My informal opinion is that for many WT speakers, DM has changed from being simply a tool to being a language of

[32]See Showalter (2001) for a discussion of the matched-guise technique used in a less educated and socially integrated ethnic group.

home and family and close personal relationships. More is said on this topic in the Epilogue.

Many WT speakers have told me that when a Taranganese person uses only Malay when speaking to other WT speakers, it conveys a haughty attitude, especially if that person has lived elsewhere. Similar attitudes have been reported for other language groups of southeastern Maluku. For Dobel speakers (also of the Aru Islands), using only Malay is a sign of arrogance (Jock Hughes, personal communication). But according to Craig Marshall (personal communication), Fordata speakers who have lived away from Fordata Island and return speaking only Malay are considered both stupid and arrogant. In Selaru (David Coward, personal communication) and Luang (Mark Taber, personal communication), such people are mocked for being stupid. Luang speakers consider it arrogant only if the speaker seems to be using Malay to intentionally assert superiority.

Thus, it seems that most WT speakers tend to stay within a certain range of frequency of Malay. A more puristic avoidance of Malay would result in an awkward style, while a higher frequency of Malay (especially SIM) would give the impression of haughtiness. These reported attitudes are clearly parallel to the diglossic function of Malay as a higher variety than WT.

2.2 Linguistic differences between WT and DM

Studies of language contact phenomena typically make reference to differences and similarities between the languages involved. As mentioned in §1.1.2, some researchers have based a distinction between borrowing and code-switching on whether or not an EL word is phonologically, morphologically, or syntactically integrated into the ML. Although I am not relying on such criteria, it is appropriate here to consider what such integration looks like when it does occur. In this section, then, I survey just those morphological, syntactic, and phonological features which are relevant to this question.

2.2.1 Phonology and phonetics

Although the phonology of DM has never been studied, my informal observations indicate that it is all but identical to that of Ambonese Malay, which has been described by van Minde (1990, 1997). The phonologies of WTA and WTB have been described by Nivens (1992a, 1992c). Phonemes of WT, DM, and Standard Indonesian Malay (SIM) are shown in table 2.1.[33]

[33]Symbols in brackets represent phones rather than phonemes; they are included here to demonstrate phonetic similarities between the languages that would not be clear from a listing of phonemes alone.

Table 2.1. Phonemes of WT and two varieties of Malay

WTA	a	ɛ	ê	i	ɔ	ô	u	[p]	t	k	b	d	[g]	[dʒ]	ɸ	s	m	n	ŋ	r	l	w	y
WTB	a	ɛ	ê	i	ɔ	ô	u	[p]	t	k	b	d	g	[dʒ]	ɸ	s	m	n	ŋ	r	l	y	
DM	a		e	i	o		u	p	t	k	b	d	g	tʃ	dʒ	f	s	h	m	n	ñ	ŋ	r	l	w	y
SIM	a	e	i	o	u	ə		p	t	k	b	d	g	tʃ	dʒ	f	s	ʃ	h	m	n	ñ	ŋ	r	l	w	y

Of course, the distributions of these phonemes are different for each language, as are the phonological processes and morphophonemic rules. Of chief importance for the study of phonological (non-)assimilation of Malay words into WT discourse are those phonemes, prosodic structures, phonological processes, and morphophonemic rules of WT which are not shared by Malay. I forego further discussion of these until chapter 4, where I discuss evidence of phonological assimilation of Malay items in the WT/Malay corpus.

2.2.2 Morphology

WT has affixes for agreement and nominalization, a multipurpose prefix *r-*, and reduplication for various functions. WT morphology has been described by Nivens (1992a, 1992b, 1993a). Active verbs (transitives and active intransitives) take verb agreement prefixes, while stative intransitive verbs take agreement suffixes. All verbs can be nominalized by means of reduplication; active verbs can also be nominalized by means of prefixes which occur in place of the agreement prefix. The verb prefix *r-*, which occurs between the agreement prefix and the active verb root under certain conditions (including reflexivity and actor focus), has been described by Nivens (1991b).

Standard Indonesian Malay (SIM) has been described by Macdonald (1976), Dardjowidjojo (1978), Moeliono and Dardjowidjojo (1988), and Sneddon (1996, 1997). It is a language rich in productive derivational morphology, but poor in inflectional morphology; perhaps the only inflectional morphology in SIM is reduplication for marking plurality. Dobo Malay appears to have the same morphology as Ambonese Malay; in these dialects, most of the productive derivational affixes still evident in SIM are retained only as fossils or as submorphemic parts of words borrowed from SIM. According to Collins (1980:26), the only productive verbal affixes in AM are *ba-*, *ta-*, and *baku-*. Tjia (1995) claims that only *baku-* (reciprocal) is fully productive among the verbal affixes; *ba-* is rather unproductive, and *ta-* even less productive. As for nominal affixes, Tjia claims that only *paN-* is productive. Words which seem to be affixed with

SIM derivational morphology are either instances of high DM or else frozen forms borrowed into DM from other varieties of Malay. Thus, DM has even less productive affixation than WT.

Reduplication, however, is productive in both languages. Forms and functions of reduplication in WT have been described by Nivens (1993a); both forms and functions differ in minor ways from village to village, and radically from WTA to WTB. In addition to lexemes which are obligatorily reduplicated (e.g., *tuntún* 'mosquito'), any verb modifying a noun must be reduplicated in all WT dialects. In WTA only, reduplication also marks the scope of negation, as well as progressive aspect. None of these functions are shared by reduplication in Malay, which is mainly either lexically obligatory or else serves to mark plurality.[34]

The lack of overlap in the two morphological systems makes morphological integration of Malay verbs easy, since there are no affixes competing for the same position. However, as seen in chapter 5, the use of reduplication for different functions in WT and DM can (rarely) cause some confusion.

2.2.3 Syntax

The syntax of Ambonese Malay (which is virtually identical to that of Dobo Malay) has been described by Collins (1980), Tjia (1997), and van Minde (1997). Although linguists debate about whether SIM follows a nominative-accusative or ergative-absolutive pattern,[35] low Dobo Malay is clearly nominative-accusative (Tjia, personal communication). The syntax of WT has been described by Nivens (1989) and by Wattimury, Haulussy, and Pentry (1995); WT clause-level syntax is clearly a nominative-accusative system, at least within the lexicase framework (Nivens 1993b).

NP structure in DM and WT is similar but not identical. In both languages, a modifying stative verb follows the head noun; in DM it may be preceded by the relative pronoun *yang* as in (1), or without *yang* as in (2); in WT it must be reduplicated as in (3), usually without the loan *yang* inserted.

(1) *mesin yang baru* 'new engine'
mobil yang putih 'white car'

[34]WT marks plurality in the choice of determiner rather than by noun morphology. Malay reduplication may also mark variety in addition to plurality; it may also derive a noun denoting an object similar to the object referred to by the unreduplicated noun (Moeliono et al. 1988:166). Tjia (personal communication) claims that the functions of reduplication in AM are roughly the same as those of standard Indonesian.

[35]See for example Cartier (1979), Rafferty (1982), Hopper (1983), and Verhaar (1988).

(2) *ketua baru* ‘new chairperson’
rok hitam ‘black skirt’

(3) *mesin tubôybôy-ai* ‘new engine’
mobil e-lajírjír ‘white car’

A modifying active verb must be relativized (with yang) in DM as in (4), but must be reduplicated in WT (often with the loan yang) as in (5) and (6).

(4) *teman-teman* *yang* *kaluar*
DUP-friend REL go.out
friends who went out

(5) *surat* *yang* *nei* *i-jir-jir*
letter REL 3s 3s-DUP-write
the letter which he/she wrote

(6) *tamata* *yang* *da-bong-bongkar* *no*
person REL 3p-DUP-tear.apart that
the people who tore that apart

Cardinal numbers, when used for counting, generally precede the head noun in SIM; however, they may also follow the noun, except for nouns which serve as noun classifiers and certain nouns with which a following number has a labeling function[36] (Kaswanti Purwo 1988:79). In AM, counting numbers generally follow the noun, with the same exceptions as just noted (Tjia, personal communication); but in DM (at least, as evidenced in the WT/Malay corpus), counting numbers precede the noun, unless the number is part of a place name. In WT, all numbers must follow the head noun and any modifying verb.

In both languages, definite and indefinite articles generally occur in NP-final position.[37] WT nouns fall in two subcategories which roughly correspond to semantic animacy; this subcategorization is marked by the article and by verb agreement affixes rather than by noun morphology. Malay nouns must be assigned to one of these subcategories as well.[38] A study of how Malay nouns are thus integrated would make an interesting follow-up study to the present work; some speakers adhere to the criterion of semantic animacy, while others tend to put all Malay nouns in the animate subcategory.

[36]For example, *kelas satu* ‘first grade’, *tingkat dua* ‘second floor’.

[37]Tjia (personal communication) claims that a preposed demonstrative article in Malay indicates some kind of pragmatic focus. When the WT demonstrative determiner *nen* ‘this’ occurs with a personal name, it usually precedes the name.

[38]Poplack and Sankoff (1984) refer to such gender assignment of English insertions in Spanish as syntactic integration.

As for clause structure, both languages have the basic constituent order SVO, both have preverbal auxiliaries, both are prepositional. Unlike SIM, DM clearly exhibits nominative-accusative syntax, and there is no passive other than a construction *dapa* 'get' + verb which is nearly always used with animate patients undergoing adverse or unexpected events (Collins 1980:69, Tjia 1992).

(7) *kamong* *pung* *anjing* *dapa* *lempar* *tadi*
2p POSS dog PSV pelt a.while.ago
Your dog was pelted a while ago.

WT also has nominative-accusative syntax, and has no syntactic passive at all; note, however, that the form of the Indonesian short passive (i.e., with agent NP omitted) is similar to a WT sentence with a topicalized object and a generic third person agreement prefix on the verb, as in (8).[39]

(8) *bôt* *on* *aroka* *da-sai-sai* *=si.*
house this later 3p-DUP-tear.down already
This house will soon be torn down.

Another difference between DM and WT syntax is in their causative constructions. There are two causative constructions in AM involving the verb *kas(i)*; (9) has a permissive meaning, while (10) is more directly causal (Tjia 1997).

(9) *beta* *kas(i)* *dia* *pulang*
1s CAUS 3s go.home
I allowed him/her to go home.

(10) *beta* *kas(i)* *pulang* *dia*
1s CAUS go.home 3s
I sent him/her home.

It is unknown whether DM (and in particular, DM as spoken by the speakers in the WT/Malay corpus) has the same distinction. The pattern in (10) occurs many times in the corpus, but that in (9) is attested only once: in (11), the meaning is clearly causal, not merely permissive.

(11) \AD ᴰ***Tuak*** *no sakali ipopo* ᵞ***keuntungan*** *lebá•ia. [.]* ᴺ***Malahan,***
%ᵞ***ikasi*** *sêta [:]* ᴸ***miskin.***
That liquor doesn't bring any benefit at all. [.] On the contrary, it makes us [:] poor.

Interestingly, this rare Malay pattern is analogous to the WT pattern, which involves the dummy verb *-m* as seen in (12).

[39]The equal symbol preceding a vernacular form indicates that it is a clitic.

(12) *ok* *kom* *nei* *inaltúk*
1s 1s.CAUS 3s 3s.return
I sent him/her back.

Perhaps WT speakers have restricted their variety of Malay to precisely that pattern which is unlike WT. Another possibility is that all the occurrences of *kas(i)* analogous to (10) in the corpus are actually phrasal entries in the lexicon. Evidence for the latter hypothesis is found in the fact that causative *kas(i)* occurs sixty-three times[40] with a Malay verb complement, but never with a WT verb complement.

Perhaps the most obvious difference between WT syntax and DM syntax is in the genitive constructions. In DM, the typical genitive construction is possessor-*pu(ng)*-possessed as in (13)–(15), though the medial word may (rarely) be omitted. The medial word takes the form *punya* in a higher (though still nonstandard) register of DM, as in (15).

(13) *dia* *pung* *laki*
3s POSS man
her husband

(14) *orang* *pu* *bini*
person POSS wife
somebody's wife

(15) *katong* *punya* *pekerjaan*
1p POSS work
our work

DM also has the SIM genitive construction possessed–possessor for certain collocations, the most common of these being locational nouns as possessed (e.g., *balakang* 'behind', *muka* 'front') preceding a noun as possessor.

(16) *balakang* *mesjid*
rear mosque
behind the mosque

(17) *air* *mata* *duyung*
water eye dugong
dugong tears

(18) *harga* *ikang*
price fish
the price of fish

In WT, inalienably possessed (IP) nouns—mostly body parts, kinship terms, and locational nouns—follow the possessor and most are inflected

[40] To be specific, *kas* occurs forty-three times and *kasi* twenty times. *Bikin(g)* also occurs three times as a causative verb in place of *kas(i)*, and *buat* occurs as a causative once.

to agree with it, as in (19)–(20). Non-IP nouns, in contrast, occur in genitive constructions similar to the typical DM genitive construction, the sole difference being that the medial possession word is inflected to agree with the possessor, as in (21)–(22).

(19) *nei* *ama-i*
3s father-3s
his/her father

(20) *ok* *mata-ng*
1s eye-1s
my eye(s)

(21) *nei* *kanei* *bôy*
3s POSS.3s father
his/her father

(22) *ok* *kanáng* *anakota* *ir*
1s POSS.1s plate DEF.PL.AN
my plates

In chapter 5, I examine the occurrence of these various genitive constructions in the WT/Malay corpus.

2.3 Summary

In this chapter, I have described the language ecology of West Tarangan and the linguistic differences between WT and Malay which are relevant to the present study. As WT culture has assimilated to the larger interethnic Malukan culture, Malay has made inroads on a number of fronts. As contact with other ethnic groups has been rendered generally peaceful by governing authorities, economic and educational opportunities have made knowledge of Malay (both DM and SIM) increasingly desirable. Malay is the language of domains which did not previously exist in WT culture, and is used to a large extent in most traditional domains as well, if only as a lexical resource to draw from. Linguistic similarities in phonology and syntax, and lack of overlap in morphology, have made the incorporation of Malay words into WT rather easy, as evidenced by the corpus examined in chapters 4 and 5.

3

Methodology and Corpus

You can't always get what you want
But if you try sometimes, you just might find
You get what you need

Mick Jagger and Keith Richards,
You Can't Always Get What You Want

One aim of the present work is to determine the effect of idiolectal differences, discourse context, and the availability of equivalent lexical units on the occurrence of EL elements in a bilingual corpus. In this chapter I describe the methodology I used in order to answer these questions and also provide information about the corpus, including the participants and situations from which the data were taken.

3.1 Methodology

In a nutshell, my methodology was as follows: first, I recorded WT speakers engaged in somewhat natural conversation. Second, my assistants and I transcribed the conversations, marking each word that was wholly or partly Malay. Third, I interviewed some of the speakers to determine what lexical motivations there may have been for choosing to use those particular Malay words in those particular contexts; based on these interviews, I tentatively posited a variety of categories of Malay words and morphemes occurring in the corpus. Finally, focusing only on those instances where Malay seemed to result from a free language choice

rather than merely a choice constrained by lexical or discourse factors, I examined the context to find evidence of any psycholinguistic or sociolinguistic motivation for such a switch in language mode. These methods are described in greater detail in the sections which follow.

3.1.1 Data collection methods

I did not attempt to obtain completely natural-occurring speech. To obtain such data, it would be necessary to record conversations without the knowledge of the speakers, whenever and wherever they happen to occur. Rather, I attempted to collect natural language in interaction situations which were as authentic as possible. The groups of participants were self-constituted with help from the recordists, but the situations were usually somewhat artificial because in all conversations but one, the participants knew they were being recorded; in that one, participants had given permission at an earlier date to be recorded without their knowledge.[41] I was not willing to record anyone without their consent, not only because of general ethical considerations but also because of my concern to maintain good relationships with WT speakers.

Twelve conversations (8 hours) were recorded in the homes of one or more of the participants, while four conversations (7 hours) were recorded in our kitchen; the participants in the latter four were close personal friends of ours who were already in the habit of conversing in our kitchen. Thus, I am fairly confident that the settings were familiar enough to promote natural conversational style. Still, I will not over-extrapolate and assume that the corpus examined here is representative of WT language behavior in general. The present study reveals patterns of language mixing which occur when WT speakers know they are being recorded for my research; it does not necessarily reveal patterns which occur in other contexts.

All of the conversations were recorded by one of two young men employed by me; both of them are fluent WT speakers. In most of the conversations the recordists did not take an active part in the conversations, although in one conversation the recordist was one of two major participants. I was never present during the recording, as I did not want my presence to influence language usage.

I did not tell anyone that I was studying patterns of language mixing. At first, this created a problem, because many WT speakers felt they were incapable of conversing in pure WT. I therefore instructed recordists to tell potential participants that it was alright if they

[41]Other recordings of this group of participants were made, but the recordings were not only natural, they were untranscribable.

mixed Malay in their speech, because I wanted to get recordings of how people normally talk. After that was made clear, a number of people were willing to be recorded. Even so, there are instances in the conversations where speakers began to use Malay, then edited their speech by replacing the Malay with WT; there are also instances of one speaker editing another speaker's speech in the same way. There is no way for us to know how many of those replacements were motivated by the presence of the tape recorder. However, as a speaker of both languages myself, and having lived in the WT language area for a few years, I can verify that the kind of language mixing in this corpus is representative of speech both directed to me and overheard by me.[42] Most of the time the participants were paying very little attention to the tape recorder, being caught up in the stories they were telling or hearing.

For purposes of analysis, I would prefer to have videotapes rather than audio tapes. However, unless the equipment were well-hidden, my previous experience with videotaping WT speakers indicates the data obtained would likely be far less natural. Tape recorders are already common objects to WT speakers, but video cameras make some feel very uncomfortable. I found three problems with attempting to videotape conversations. First, some WT speakers felt pressure to perform, resulting in either unwillingness to be recorded or, if recorded, did so in an awkward style. Second, in order for the recordist to capture the non-audio data I wanted to see, the participants had to be seated in a configuration which may not be optimal for natural speech-unless, of course, several cameras were used. Finally, the audio data on a videotape would likely be of lesser quality than that obtained from a good tape recorder.

Milroy and Muysken (1995a:9) claimed that obtrusive recording techniques will give rise to less code-switching and code-mixing. In the present study that may be particularly true. I am well-known among WT speakers as one who has been among them off and on since 1987, studying the WT language. Ordinarily, I have been interested in a purer form of WT than that which occurs in their everyday conversation. Therefore, it is possible that speakers were at times guarding their language behavior more closely than usual; possible evidence of this is discussed in chapter 4. Even so, the resulting corpus still contains far more EL insertions than corpora studied by other researchers.

[42]I did not play the recordings for other WT speakers to see if they concurred with this opinion.

3.1.2 Transcription methods

The recordists transcribed the conversations, after which I listened to the tapes myself and carefully edited the transcriptions, while also marking words (for later lexical analysis) which were either partly or completely Malay, and making notes to myself in between the lines about various instances of LCP as they occurred. Each speaker was identified by a two-letter code (e.g., AD, LL, OK).

Some stretches of speech were completely unintelligible; these are marked as (...). Other stretches of speech were semi-intelligible; these were also placed between parentheses and excluded from analysis. False starts were placed in square brackets, and likewise excluded from the lexical analysis.[43] A circumflex character was used to indicate overlapping speech. A complete list of transcription conventions may be found in the List of Abbreviations and Symbols in the front matter of this volume.

I have transcribed WT and SIM morphemes according to their respective orthographies, rather than adhering to a strictly phonological transcription.[44] Dobo Malay has no established orthography, so DM morphemes were often transcribed in such a way as to reflect DM phonology. But I have not taken pains to transcribe Malay morphemes in the corpus in a phonologically accurate way; instead, standard Indonesian spellings were often used even when the actual pronunciation may have been nonstandard. There are two reasons for this: first, the present work is primarily a lexical analysis, not a phonological analysis. Second, the recordings were unfortunately inadequate to determine exact pronunciations in many cases. It was often impossible to determine whether a speaker had used [ə] or [a], or whether the final nasal was alveolar or velar, or whether [h] was present or not. So, for example, even if the transcription reads *tahun*, the speaker may actually have said *taung*; however, wherever the transcription uses a nonstandard spelling, the pronunciation it reflects was clearly audible.

Aside from that caveat, I have taken great care to ensure that all transcribed data analyzed in this study are accurate morphemic representations of the original speech. In many instances (especially in quote

[43] There are about 1,600 such false starts in the corpus, ranging from a speaker stumbling over the first phoneme of a word to a speaker replacing a stretch longer than a clause with an improved version of the same stretch. Although the words used in false starts have not been included in the overall lexical analysis, in §4.6.1 I examine those false starts which involve a speaker replacing a Malay word with a WT word.

[44] In both standard Indonesian and WT orthographies, final glides /y/ and /w/ are spelled *i* and *u*, respectively, and the velar nasal is spelled *ng*. I have also followed these conventions in transcribing DM.

formulas) although I am certain of what the speaker said, any words which were not clear on the tape were excluded as unintelligible. As a non-native speaker researcher, I did not transcribe without the help of a native speaker; but I also found that a native speaker research assistant's work could not be assumed to be completely reliable. I found a great many significant errors in the original transcriptions. The analyst himself is ultimately responsible to guarantee the reliability of the transcription. If the reader is not given sufficient reason to trust the transcription, then the conclusions of the analyst must be questioned.[45]

3.1.3 Analysis methods

Interviewing. After ensuring that the transcription was reliable, I conducted two series of interviews to discuss with several of the speakers the Malay words used by them and others in the corpus. Ideally, I should have interviewed each speaker about his own use of Malay, given my comments in chapter 1 about individual language competences; but there was simply no time for such a fine-grained approach due to the vast number of Malay insertions in the corpus. Instead, I interviewed OK individually about her own use of Malay, since she was the only WTB speaker in the corpus (and also because she uses more Malay than most of the other speakers in the corpus; see section 3.2.1); and I interviewed AD, LL, and TN as a group regarding the WTA speakers' use of Malay.

I used the SIL Word List program to produce two lists (one for WTA speakers and one for WTB speakers) of Malay morphemes found in the corpus. In the interviews, I discussed each word in the lists, checking the actual occurrences of the words in the corpus to allow the context to determine the exact sense intended. For each instance of each word, then, the interviewees either attempted to provide a WT (near-)equivalent, or else indicated that there was none. We also discussed any differences in usage between WT and DM (near-)equivalents which might make a DM word more appropriate for a particular context than a WT word.

Briggs (1986) discussed the difficulties in successfully carrying out interviews cross-culturally. Since the subjects' understanding of the communication event is based on their own cultural repertoire of communication events, rather than that of the researcher, the information they provide may not actually be correct answers to the questions the researcher is intending to ask. According to Briggs, it is essential for the researcher to

[45] It is possible that other LCP researchers have followed the same kinds of procedures and standards as I describe here, but without a clear statement from them it is impossible to know.

not simply take answers at face value, but to interpret the information in accordance with previously observed communication styles in the culture under study. In the current study, the interviewees first attempted very obligingly to fill in the word list by giving simple WT equivalents to the Malay words without regard to the actual context in the corpus. For example, *perái* 'crowd' was given as an equivalent for *tamu* 'guest', since in WT culture guests tend to come in groups. However, I had had enough experience with the interviewees to realize what they were doing, and insisted that they give equivalents appropriate to the context or else admit there were none. Before long they often asked to see the context first, before giving any answer at all.

Still, there are other problems with this approach. Even after looking at the context, the interviewees sometimes offered a word which would fit the context but was not semantically equivalent to what the original speaker actually chose to say. Typically, the Malay word was semantically more specific than the nearest WT equivalent. In such cases, I considered the WT and DM words not to be equivalent.

In addition, I could not assume that the kind of psycholinguistic processing that takes place in an interview (translating word lists, a metalinguistic process) is the same as the mental processing that occurs during conversation (actually **using** language). In a live conversation, there is not time for the speaker to stop and search his mental lexicon for the best WT word; if a Malay word comes to mind, and it fits, it will be used by the speaker, to avoid being guilty of delay of game.

Not only are there differences in mental processing between interviews and conversations, but there are inevitably differences in the same person's conver-sational abilities from one day to the next. Therefore, even if I had taken the time to interview each speaker separately, it would still be impossible to arrive at 100 percent truth with regard to whether the speaker actually chose to use Malay or whether other factors (e.g., mental block) were at work. I must admit, then, that the codes assigned to Malay words in the corpus as a result of these interviews are only an approximation; but they are the best approximation possible, given the nature of the human language faculty.

Identifying lexical units. The lexical units upon which the analyses of chapters 4 and 5 are based are in many cases not single words, but strings of words. In some instances, such a word string is clearly a phrasal entry in the speaker's mental lexicon since the meaning of the whole is not equal to the sum of its parts. These include proper NPs as in (23) and (24) and euphemisms or idioms as in (25) and (26).

(23) WT proper phrasal lexical entries

Nata•Tanín	'Original Village [a boat name]'
Roprop•Jala	'February' [literally 'covering paths']
Taper•Têrbôt	'Mr. Têrbôt'
Tôr•Dauk	'Roosters Crow [a place name]'

(24) Malay proper phrasal lexical entries

Pulau•Babi	'Pig Island [a place name]'
Tahun•Baru	'New Year [name of event]'
SD•Anam	'Elementary School Number 6 [a school name]'

(25) WT idioms and euphemisms

nam•tapôran	'pregnant' [literally 'do body']
i-pan•ela	'3s-sink/drown' [literally '3s-fall go']
bôt•abil	'parlor' [literally 'inside house']
gul-ang•kala-i	'my head/hair' [literally 'head-1s shell-3s']

(26) Malay idioms and euphemisms

buang•air	'defecate/urinate' [literally 'throw out water']
muka•doit	'greedy' [literally 'money face']
pekerjaan•rumah	'homework' [literally 'house work']
orang•tua	'parent(s)' [literally 'old person(s)']

In some instances, Malay phrasal lexical entries are morphologically integrated into WT:

(27)

i-buka•jalan	'3s-make a way' [literally 'open a path']
i-tunju•jago	'3s-show off' [literally 'indicate/demonstrate strength']
da-masu•minta	'3p-propose marriage' [literally 'enter request']
da-dudu•adat	'3p-negotiate according to custom' [literally 'sit custom']
ku-cari•tahu	'1s-find out' [literally 'seek know']

Another kind of mixed phrasal lexical entry involves Malay verbs integrated into WT not by adding an agreement affix but by making the verb a complement of the dummy verb *-m* 'do'. This phenomenon is discussed in chapter 4. To facilitate the lexical analysis, such phrasal incorporations were linked with an underscore character in the transcription, as in *kom•bataria* '1s-do shout', to form phrasal lexical units.

In many cases, however, lexical units are larger than a lexical entry. Such is the case when adjacent words collocate so strongly with each other that they form a LEXICAL ROUTINE.[46] Some lexical routines are pure

[46] Lexical routines are similar to Pawley and Syder's (1983) concept of lexicalized sentence stems. In the language acquisition literature, lexical routines are often referred to as "lexical phrases".

WT, others are pure Malay, while still others are mixed. There are three criteria by which such lexical routines may be identified.

First, it may be impossible to switch languages between words, because neither of the words have a (near-)equivalent in the other language. Such is the case with the examples in (28). Although several of these are Malay genitive structures, they are apparently borrowed as phrases, and therefore, the occurrence of Malay syntax does not constitute evidence for Malay mode.

(28) *ayat-ayat•Alkitáb* 'Bible verses'
jam•kantor 'office hours'
kotak•suara 'ballot box'
sekolah•SMA 'high school'
balai•desa 'town hall'
hukum•adat 'traditional law'

Second, even if one or more (even all) the words in a string have equivalents in the other language, the corpus itself may prove the strength of a collocation, if the concept encoded by the string of words occurs frequently enough to demonstrate that the string represents a frozen sequence of words (i.e., a lexical routine).[47] Some such collocations consist of words from a single language as in (29) and (30), while others are mixed as in (31). (Malay words are set in bold italic.) In some instances, there is an equivalent for only one of the words in the other language; the word string as a whole corresponds to a single word in that language as in (32). Finally, a string of words in one language may correspond to a string of words in the other, or to a mixed string of words, as in table 3.1.

(29) WT lexical routines
rat•lat 'three hundred[48]'
ko•tên 'then immediately' (59 instances)
lakaria•mo 'after a while then' (12 instances)

(30) Malay lexical routines
dua•puluh•anam 'twenty-six'
memang•batúl 'indeed correct' (10 instances)
yang•penting 'what's important (is)' (13 instances)

[47] Lexical routines which occur frequently have the potential to become fused as words (cf. English *nevertheless*, *moreover*, *nobody*; WT *bôsian* 'area under house', from *bôt* 'house' + *sian* 'under').

[48] Number phrases in the corpus were always either completely WT or completely Malay, so I concluded that they are phrases which do not allow internal switching and connected them as lexical units.

(31) Mixed lexical routines

*kanei•**suka**•aka*	‘he/she wants’ [literally ‘his/her desire (is) for’] (7 instances)
***ampung**•aka*	‘affection to’ (35 instances)
***oras**•on*	‘now’ [literally ‘this time’] (76 instances)
*sakali•**pusing***	‘don’t care’ [literally ‘not dizzy’] (10 instances)

(32) WT lexical routines equivalent to Malay words

arei•ba (WT) = ***berapa*** (Malay)	‘how much, how many’
dam•ba (WT) = ***bagaimana*** (Malay)	‘how’
dam•on (WT) = ***begini*** (Malay)	‘like this’
leilei•ia (WT) = ***sedikit*** (Malay)	‘a little bit’

Table 3.1. WT lexical routines equivalent to Malay or mixed lexical routines

WT	Malay or mixed	
sakali•lebá•ia (19 instances)	***seng•apa-apa*** (4 instances)	‘it doesn’t matter’
amai•mirmirna (2 instances)	***bapa•bongso*** (10 instances)	‘youngest “father”’
nung•kane (18 instances)	***rupa**•kane* (23 instances)	‘like, similar to’
maera•ne (214 instances)	***waktu**•ne* (69 instances)	‘at that time’
maera•êr (82 instances)		

Since relative frequency is a crucial concept in the lexical analysis of chapter 4, it was essential to compare equivalent lexical units as wholes, rather than comparing the words that comprise the lexical units with their own equivalents.

Third, where frequency of occurrence of a concept is too low to demonstrate the strength of a collocation, the reaction of interviewees can serve as an indicator, although less reliable. For example, WT *kada* ‘pants’ is equivalent to Malay *celana*, but when I suggested the mixed phrase *kada rok* as a possible equivalent of the Malay phrasal entry *celana rok* ‘split skirt’ [literally ‘pants skirt’], all three interviewees rejected it. Thus, a word string was identifiable as a lexical unit if the interviewees considered a language switch between words to be unacceptable, or at least very odd.

In conclusion, studies of CS (including studies of syntactic constraints on CS) must begin with a lexical analysis of the corpus, and that lexical analysis must begin with the identification of lexical units. Syntactic constraints should not be proposed to account for the lack of a switch when collocation is the real explanation; conversely, syntactic constraints should not be

rejected when collocation is the real reason for a switch between two words. If there exists a strong collocation between two or more words, then switching, or the lack of switching, between them is not significant. Therefore, I removed from the list of possible switch-points those word boundaries at which CS could not occur for lexical reasons. Also, I removed from the list of actual switch-points those word boundaries where a switch does occur for lexical reasons—namely, mixed lexical units. In chapter 4, I attempt the kind of lexical analysis I am calling for here.

3.2 The corpus

The corpus consists of 16 conversations, varying in length from 20 minutes to 3 hours, making a total of 15 hours. In this corpus there are 28 native speakers of WT, which I have categorized as primary, secondary, or tertiary, depending on the volume of their contributions.

The most prevalent type of LCP in this corpus is single-word Malay insertions. Brief Malay phrases are also frequent. Complete Malay sentences also occur, usually in direct quotes reporting the speech of non-WT speakers. This corpus, then, is in striking contrast to some others studied in the CS literature, in which, for example, entire paragraphs of one language are followed by entire paragraphs in another, with very few single-word insertions. Chapters 4 and 5 will present a detailed analysis of the various kinds of Malay insertions in the corpus.

3.2.1 Participants

I have categorized speakers as shown in table 3.2, according to the number of lexical unit tokens they produce in the corpus. The primary speakers, in order of amount of speech in the corpus, are listed in table 3.3. For the remainder of this section, I present some information about these speakers' personal linguistic histories which is relevant to their language mixing behavior.

Table 3.2. Speakers categorized by volume of speech

	Total	Lexical unit tokens per speaker
Primary Speakers	6	3000–24000
Secondary Speakers	16	220–1700
Tertiary Speakers	6	less than 120

Table 3.3. The six primary speakers

Speaker	Sex	Birthyear	Origin	Lexical unit tokens
AD	f	1949	Feruni	23478
LL	f	1972	Feruni	13650
OK	f	1961	Doka Timur	10915
YL	f	1962	Feruni	10089
TN	f	1972	Ngaiguli	5739
SD	m	1945	Feruni	3079

AD was born in the WT-speaking village of Feruni, and when she was eight months old her mother moved to Dobo, leaving AD in the care of the protestant pastor in Feruni, who was a Malay-speaker. Before entering elementary school, she moved in with her aunt, who mixed WT and Malay in her speech. During her elementary years she went back and forth between Feruni and Dobo; in the village she and her friends spoke WT with each other, while on her trips to Dobo she spoke mixed WT and Malay. After elementary school she attended the SKPK (Sekolah Kepandaian Putri Kristen = Christian girls' home economics school) in Dobo. From 1972 to 1975 she attended various courses to become a midwife. She lived in Tual, the capital of the Southeast Maluku regency (where Malay is the lingua franca and Kei is the indigenous language), from 1978 to 1980, and in Surabaya, east Java, from 1983 to 1988. In the GPM (Protestant Church of Maluku) church in Dobo she has served as *kolektaan* (offering collector) from 1980 to 1983, and as *pengasuh* (children's teacher) from 1980 to 1983 and 1988 to the present. She has also worked with my wife and me as a language assistant from 1990 to 1991, and as domestic help for an Ambonese family in Dobo from 1991 to the present. She made a brief trip to Ambon in August 1995.

LL and AD are distantly related, and have known each other since LL was a girl. LL attended elementary school in her home village of Feruni, then attended SMP (Sekolah Menengah Pertama = middle school) in Dobo beginning in 1985, and SMEA (Sekolah Menengah Ekonomi Atas = business high school) first in Dobo and then in Tual, graduating in 1994, after which she returned to live with her parents in Dobo. Her language environment as a child, both with her parents and with her friends, was mixed WT and Malay. She made a brief trip to Ambon in August 1995, after which she worked as a language helper for my wife and me and also as domestic help for a Chinese family in Dobo.

OK is the only WTB speaker in the corpus. She was born in her mother's village of Doka Timur, but before she was one year old her father died,

and she was taken to live in her father's home village of Doka Barat, while her older brother remained in Doka Timur. Both are WTB villages. The language of her parents and childhood friends was strictly WT—she reports that while the people of Doka Timur freely mix WT and Malay, the people of Doka Barat prefer not to. She did not learn Malay until she attended elementary school. She finished elementary school in Doka Barat, then attended middle school in Doka Timur for two years but did not finish. She lived in Meror, an East Tarangan village, for one year, then worked for a Japanese pearl company near Tabarfane village (whose dialect is transitional between WTB and Manombai) for three years. She then moved to Dobo, and had an extended relationship with a man from the Kei Islands. In 1982 she worked for a Chinese merchant for a year, and in 1990 to 1991 and 1995 to 1997 worked as domestic help for my wife and me in Dobo. In 1994 she worked at the Djajanti Group shrimp operation near Benjina village (in the Manombai language area), where hundreds of speakers of various languages work and Malay is the lingua franca. She said she mixes a lot of Malay in her WT now because she has been employed by various non-WT speakers, and also because there are not many WT speakers in her neighborhood in Dobo.

YL was born in Feruni village, born into the same clan as LL. She finished elementary school there but did not enter middle school. The language of both her parents and her childhood friends was a mixture of WT and Malay. After elementary school, she lived in Dobo for one year, then at the Djajanti Group shrimp company near Benjina for three years. She returned to Dobo and worked at a government office for seven years, then moved to Ambon where she worked as a cook from 1984 to 1986, married a man from Tepa (a Malay-speaking town in southeast Maluku) in 1986, and continued living in Ambon until 1987, when she returned to Dobo. She lived in Dobo from 1987 to 1992, in Tepa from 1993 to 1995, and again in Dobo from 1995 to 1996.

TN and LL were best friends at the time of the study; they are related and were friends since TN first came to Dobo in 1988. TN was born in the WTA village of Ngaiguli, and attended elementary school first in the neighboring village of Fatural and then finished in her home village. She attended middle school in Jeroil, a WTB village, then entered the SPK (Sekolah Pendidikan Kesehatan = nursing school) in Tual, but after only two weeks decided to return to her home village instead. She then attended SMEA (business high school) in Dobo from 1988 to 1990. In 1993 she worked for a Chinese family in Ujung Pandang, Sulawesi for four months. Both her parents and her childhood friends spoke mixed WT and Malay.

SD was AD's older brother. After elementary school in Feruni village, he attended SMEP Kristen (Sekolah Menengah Ekonomi Pertama Kristen =

Christian junior high business school) in Dobo. He never left Aru, except for one month in Ambon in 1980. He served as Sekretaris (assistant head of village) of Feruni from 1982 to 1987, and Pejabat (acting head of village) from 1987 until his death in 1998. His parents used WT with him at first, then taught him Malay before he began elementary school. He and his childhood friends spoke a mixture of WT and Malay with each other. Although he usually lived in Feruni, he was the main contact person between the people of Feruni and the Indonesian government, and therefore had to use Malay (whether low DM or SIM) frequently.

3.2.2 Situations

Conversations 1–6, 10–12, and 15–16 were all recorded in the homes of one or more of the participants. Conversations 8–9 and 13–14 were recorded in our kitchen. Conversations 6 and 7 were the least natural. In 6, SD tells three stories from his experiences as head of Feruni village; the topics were planned in advance, but there is interaction with other participants, especially his sister AD who insists on telling part of the story from her perspective. Conversation 7 is a complete monologue, in which SD describes how he handles various interpersonal situations as a village leader and as a father.

Conversation 16 was recorded without the participants knowing they were being recorded.[49] In all other conversations, the tape recorder and microphone were visible.

3.3 Conclusion

In §3.1 I have described the methodology used for collecting, transcribing, and analyzing the corpus under study. The analysis involved many hours of interviews with four of the primary speakers in the corpus, and many more hours of computer-assisted lexical analysis. I have also described the primary participants and the situations in which the data were collected. In §3.2, I have given a brief description of the corpus and some relevant information about the linguistic histories of the six primary speakers in the corpus. The corpus itself is significantly different from corpora studied by other LCP researchers (e.g., Myers-Scotton 1993b) in that it contains vast amounts of Malay insertions: although WT tokens far outnumber Malay tokens, Malay types far outnumber WT types.

[49] As noted earlier, they had given prior consent to being recorded without their knowledge.

The kind of study undertaken here could not be performed by a researcher with no previous experience in the languages involved. The present work is grounded on several years of analyzing WT language and culture and building relationships of trust and open communication with WT speakers as well as AM and DM speakers. The great mass of conversation data collected was necessary to ensure that the results of analysis would be valid at a statistically significant level.

In chapter 4 I present the lexical analysis of the corpus, to be followed in chapter 5 by an analysis of the Malay elements not accounted for by the analysis of chapter 4.

4

Prerequisites to LCP Research: Evidence from the WT/Malay Corpus

> *"When* I *use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean—neither more nor less."*
> *"The question is," said Alice, "whether you* can *make words mean so many different things."*
> *"The question is," said Humpty Dumpty, "which is to be master-that's all."*
>
> Lewis Carroll, *Through the Looking Glass*

> *Of course language is not an infallible guide, but it contains, with all its defects, a good deal of stored insight and experience. If you begin by flouting it, it has a way of avenging itself later on. We had better not follow Humpty Dumpty in making words mean whatever we please.*
>
> C. S. Lewis, *The Four Loves*

In this chapter I discuss prerequisites other than those described in §3.1, which are related specifically to LCP research, and which are based on my beliefs and assumptions about the nature of human language as discussed in chapter 1.

After opening with a consideration of the concept of equivalence (§4.1), I argue that LCP research requires a careful examination of particular words (§4.2), particular discourse contexts (§4.3), and particular speakers (§4.4). In §4.5 I discuss the notion of gratuitousness of Malay lexical items, followed by

a discussion in §4.6 of repair sequences involving repairs of language choice. In §4.7 I discuss phonological and morphological integration of EL lexemes, which have been held by some LCP researchers to define the boundary between borrowing and code-switching.

4.1 Equivalence

As stated in chapter 1, the goal of the first and third waves of CS research is to determine why a speaker switches from one language to another, either in terms of a brief foray from the matrix language (ML) into the embedded language (EL) (INSERTION, Muysken 1995) or in terms of a complete switch from language A to language B (ALTERNATION, Muysken 1995). Some corpora are more suited to the latter analysis, some to the former; but it is likely that every sizeable bilingual corpus contains at least a few examples of both. It is necessary, therefore, to attempt to find both in any given bilingual corpus. Otherwise, some insights gained from the study of alternational CS may be imposed upon insertional CS without careful consideration of the validity of such imposition.

When a corpus is characterized by intrasentential rather than intersentential CS, the analyst must begin by attempting to determine the ML.[50] Then, for each inserted EL item, the analyst's first question must be: Does this word have a reasonable equivalent in the ML which could have been used in this context? If there is no EL equivalent, there is no point in trying to determine what social or conversational motivations drove the speaker to express the concept in the EL rather than the ML; the use of the EL word is merely a lexical choice, not a language choice. In the remainder of this section I discuss the different types of equivalence that must be considered.

4.1.1 Static equivalence

In a recent paper, Li (2001) examined a corpus of Hong Kong newspaper articles and determined that many instances of inserted English were due to inequivalence between the lexicon of Chinese and the lexicon of English. For the kinds of equivalence described by Li, I use the term STATIC EQUIVALENCE, to contrast with the concept of DYNAMIC EQUIVALENCE discussed in §4.1.2.

Semantic equivalence. Li demonstrated that much of the English found in his corpus was due to lexical gaps in Chinese—that is, there was no precise

[50] Myers-Scotton (1995:237) outlines her criteria for identifying the ML.

and concise semantic equivalent in Chinese for the concept that the speaker wished to convey, but there was in English. For two terms to be semantically equivalent, they must be equivalent in both denotation and connotation.

The most obvious type of lexical equivalence is equivalence of denotation, i.e., the semantic range of the word. Malay *ikan* does not have the same semantic range as English *fish*, because the former includes fishlike mammals (whales, dolphins) while the latter does not. Although in most contexts *ikan* and *fish* are intertranslatable, owing to the high degree of semantic overlap, a text about *ikan* could not be translated as a text about *fish* if fishlike mammals were an integral part of the text. Similarly, WT *ariran* is a near-equivalent of Malay *burung* and English *bird*, but *ariran* includes bats (excluded by English *bird*) and excludes cassowaries (included by both *bird* and *burung*).

Lexical inequivalence, however, goes well beyond the problem of whether particular items are valid members of sets defined by generic terms like *fish* and *bird*. For example, even when the primary senses of two words are equivalent, they may not share secondary senses or extended senses (including figurative senses). Like Li, I found in my interviews with WT speakers that speakers (in these communities at least) reject attempts to extend the semantic range of an ML word to fit the semantic range of its near-equivalent EL word. It is surprising, then, to find examples such as (33) in the CS literature (Myers-Scotton 1995:244).

(33) *Anakula* ***plate*** *mbili*
He eats two plates.

In discussing this example from a Swahili/English corpus, Myers-Scotton nowhere assures the reader that the Swahili equivalent of *plate* (if in fact there is one) also has the extended sense exemplified by *plate* in (33), namely 'plateful of food'. She merely assumes that *plate* qualifies as an EL word, and uses this clause as an example of a Swahili modifier following an English head noun.

Although Poplack (1980, 1987) and Myers-Scotton (1993b) have disagreed about whether to call single-word EL insertions borrowings or CS, they both acknowledge that cultural imports (i.e., objects or other concepts foreign to traditional ML culture) must be considered borrowings rather than CS. In other words, they qualify as ML words, even though they are etymologically from the EL. In §4.2, I demonstrate that cultural imports is just one category among many in which words from the EL qualify as ML words and, therefore, (a) do not necessarily index a social identity connected with the EL and (b) do not qualify as counter-examples to syntactic constraints on CS.

In addition to denotational equivalence, there is also equivalence of connotation, the feel of the words. Bilingual speakers who are aware of

such connotations may use an EL equivalent because it carries the connotation they want to convey. As Heider (1991) demonstrated, there is a mismatch in connotation between emotion words in English and Indonesian. According to his analysis, English *love* is a happy word, while the Indonesian equivalent *cinta* is a sad word. English has both *surprised* (which often has a good connotation) and *startled* (with a bad connotation), while Indonesian has only *terkejut* (with a bad connotation). Indonesian speakers who are sufficiently proficient in English may, therefore, choose to use *love* or *surprised* when the Indonesian equivalents convey the wrong connotation. Conversely, during my years living in Ambon, I noticed that native English speakers there often inserted Indonesian *bodoh* 'stupid' or *gila* 'crazy' into an English discourse because of the humorous connotations of these words in Ambonese culture.

Farb (1974:89) described how denotationally synonymous words from different languages may take on connotations derived from beliefs about (or attitudes toward) the cultures associated with those languages.

> The habit of creating euphemisms dates back at least to the Norman Conquest of 1066. At that time the community began to make a distinction between a genteel and an obscene vocabulary, between the Latinate words of the upper class and the lusty Anglo-Saxon of the lower. That is why a duchess perspired and expectorated and menstruated—while a kitchen maid sweated and spat and bled. The linguistic gulf between Norman-derived and native Anglo-Saxon words remains as wide as ever after nine hundred years. The farmer today still looks after his Anglo-Saxon cows, calves, swine, and sheep—but once they are served up appetizingly in a restaurant or supermarket, they become French beef, veal, pork, and mutton. And whenever the speech community must discuss anything it deems unpleasant, the discussion is acceptable on the condition that it is carried on in the elegant vocabulary bestowed on English by the Normans.

I argue that such connotations render the words in question inequivalent; in such cases, words which are etymologically from the EL must be treated as ML words in CS research. In a word, they are borrowings.

Equivalence of economy and convenience. In addition to semantic equivalence, Li (2001) cited the principle of ECONOMY: "In the absence of serious semantic discrepancy, a loan word tends to be preferred when it is uttered or written with less linguistic effort." Thus, single lexemes are preferred over phrases and clauses. Li cites examples from his corpus

including *internet* (rather than the Chinese equivalent *gwok³zai³ din⁶lou⁵ mong⁵lok⁶* 'international computer network') and *email* (rather than the Chinese equivalent *din⁶zi² jau⁴gin²* 'electronic mail'). There is support for his hypothesis in the fact that even for monolingual English speakers, *email* and *internet* are preferred over the phrases from which they are abbreviated.

4.1.2 Dynamic equivalence

The types of equivalence just discussed are static in the sense that for the most part they do not vary from one context to another. Here I discuss two types of equivalence that are more variable over time.

Equivalence of frequency. In deciding whether the use of a particular EL word in a particular sentence is an instance of language choice or merely lexical choice, it is essential to consider discourse frequency of the word in question—both the frequency within the current discourse as well as in other discourses in the corpus, and considering speakers both individually and as a whole. EL words which a speaker frequently inserts into ML discourse must be considered loans. Myers-Scotton (1993b:16) acknowledged this principle; her working hypothesis was that any word occurring more than three times in her 20-hour corpus was a borrowing, but she admitted that this figure was arbitrary.

Poplack and Sankoff (1984:102–104) considered previous attempts to define borrowing and found four criteria used by other researchers for identifying loan words; one of these is frequency. However, it is clear from their discussion that their notion of frequency was not what I am proposing here. They were referring primarily to the frequency of usage of a particular item in the speech-community. They were apparently not interested in borrowings on the individual-speaker level but in borrowings on the community-speech level. Poplack, Wheeler, and Westwood (1987/89:136) looked at frequency from the same angle; in fact, after noting that established loan words are "both recurrent in the speech of an individual and widespread in the community" they mentioned that with nonce borrowing, "the social characteristics of recurrence and dispersion need not be satisfied"—thus focusing on social recurrence and ignoring recurrence in the speech of an individual. The present work, in contrast, takes exactly the opposite approach. I am not concerned here with acceptance of a particular lexical item by a speech community, but with the psychological (speaker-internal) reality of the difference between lexical choice and language choice.

It is important in this regard, however, to consider not only absolute frequency but also relative frequency. Myers-Scotton (1993b:194) suggested

that any EL item which occurs at least five percent as often as the equivalent ML term should be considered a borrowing; however, she admits that five percent is an arbitrary figure. She also confesses that relative frequency offers little help when the concept encoded only occurs a few times in a corpus; if, for example, a certain concept is encoded twice by an EL word and not even once by an ML word, we cannot state with confidence that the EL word in question is a borrowing. This points out the need to supplement any computerized lexical analysis of the corpus with native-speaker interviews, to determine what the speakers themselves believe to be the normal way to express a given concept in a given context.

It is possible for the heuristics of absolute and relative frequency to be at odds, of course. An EL word may occur more than three times in twenty hours because the concept it refers to occurs hundreds of times, and the EL word will still be gratuitous because the concept is nearly always expressed by an ML word. I discuss this dilemma in §4.5.

A consideration of frequency should take into account a speaker's context and recent history. For example, on September 2, 1996, my wife Susan was telling another English speaker about our first stays on Tarangan Island. During the conversation she said, "You couldn't buy soap, you couldn't buy matches, you couldn't buy *minyak tanah*." The question is, why did she insert the Malay term *minyak tanah* when there is an exact and common English equivalent, *kerosene*? It is not because *minyak tanah* is any easier to pronounce; rather, it is because although soap and matches are common items in our English-speaking life, kerosene is not—it is part of our Indonesian/Tarangan life. The concept, therefore, was at that time more strongly attached to the Malay word in her lexicon than to the equivalent English word.[51]

While we were resident in Ambon, we observed that expatriate English speakers there had a number of favorite Malay words which recurred frequently in their speech. One of these was *gudang*, which can refer to any kind of storage room from a closet to a large warehouse. Because the semantic range of *gudang* is much broader than high-frequency English terms such as *closet*, *storage shed*, and *warehouse*, *gudang* has a higher frequency than any of these; and because the nearest English equivalent *storeroom* has a relatively low frequency in English, *gudang* was often used by English speakers in Ambon even when referring to a closet.[52]

In translating a written text, frequency considerations may sometimes prove to be more important than precise denotational equivalence—that is, a high-frequency near—equivalent may be preferred over a low-frequency

[51] Susan herself, by the way, independently came up with the same analysis.

[52] This claim about relative frequency is based on my informal impression, not an actual word count.

exact equivalent. In conversation, the same is true for reported conversation. When reporting an original Malay conversation in WT, a speaker sometimes has to decide whether to use an exact denotational equivalent of low frequency in WT or a near-equivalent of higher frequency in WT, or avoid the choice by simply inserting the original Malay word.

Another consideration regarding frequency is the question of performance errors. Muysken (1995:184—5) pointed out:

> A very complicated issue concerns the relation between qualitative structural and quantitative distributional analysis... Since we do not know how the grammar and the lexicon interact with other psychological faculties to produce actual speech, we clearly cannot ignore phenomena such as frequency of occurrence and regularity. This would lead us to take the frequent types of switches as the main body of evidence, and to consider the infrequent ones as possibly fluke phenomena, performance errors and the like.

Thus, while high-frequency Malay items may be excluded from CS analysis as borrowings, low-frequency Malay items may be the result not of speakers indexing social identities but of speakers being tired or whatever else may give rise to performance errors. Since, however, I cannot identify with any certainty such performance errors, in my analysis I do not exclude any low-frequency Malay items on these grounds.

Equivalence in context. The semantic range of a word is restricted by its discourse context. Therefore, LCP researchers examining EL words must consider not the meaning of a word as given, for example, in a bilingual dictionary, but the meaning a word has (for both the speakers and the hearers) in the context where it occurs. Entries in a dictionary are only valid if they accurately reflect the usage of the speakers being studied.

In fact, there may be instances where equivalence of denotation is irrelevant. That is, in a certain discourse context the referential meaning may be more important than strict denotational equivalence. For example, although morning lasts much longer than *pagi*, *good morning* is a fine translation of *selamat pagi* if it is early enough in the morning. Similarly, some Indonesian words referring to emotions and character traits are referentially, but not denotationally, equivalent to WT verbs describing the activities associated with those emotions and traits, as Malay *sabar* 'patient' and WT *-tangara* 'wait'.

Equivalence of implicature should be considered here as well. In some instances, speakers may consider equivalence of implicature more important than equivalence of denotation. Consider YL's use of *tapele* 'blocked off' in (34).

(34) \YD *Ok tôr kem, mo kem jausin.* ***ʏTapele*** *=sin.*
1s call 2p but 2p already blocked.off already
I called to you (PL), but you were already blocked off from my view.

The interviewees considered this use of *tapele* a gratuitous use of Malay; a paraphrase with *surtán-* 'enter' would have conveyed the crucial implicature, namely that the people being called were not visible. Another example of equivalence of implicature is the functional equivalence of the WT greeting *Moimoi ba*? 'Where are you going?' and the English greeting *How's it going*?

Equivalence is dynamic in other ways as well, as indicated by Sperber and Wilson's (1986) Relevance Theory: elements (whether lexemes or longer stretches of speech) from languages A and B may be equivalent in all relevant respects in one context, but not sufficiently equivalent in another context where the relevant respects are different. Sometimes this is entirely dependent on the speaker's whims. In terms of the present study, at one point in a conversation a speaker may use a specific EL term because its meaning is more precise than the equivalent ML term; but in another context where the same object or concept is referred to, the same speaker may decide to use the more generic ML term, feeling that it is not necessary to be so precise.

Unfortunately, LCP researchers have not always taken into account the dynamic and speaker-specific nature of equivalence. It is unrealistic for a researcher to suppose that one can look up words in a bilingual dictionary to determine whether there is an equivalent in the other language. At every point in a discourse, a speaker must decide whether a particular lexical option is semantically specific enough for present purposes, or whether he wants to use a more precise but cumbersome phrase. This is true of both monolingual and bilingual discourse. When bilinguals choose between lexical options, they are not necessarily aware of the language membership of those lexical options. A speaker tends to use the most convenient lexical package to express a given concept without regard for the etymology of that lexical package, unless the speaker feels that the choice of a particular word might convey an undesired meaning simply because of its etymology—for example, a speaker may fear that his choice of an EL word would send an unintended social message, e.g., lack of solidarity with ML society.

In this section I have attempted to demonstrate that equivalence of lexical items is not entirely straightforward. In order to analyze a speaker's linguistic variation at the lexical level, it must be clear that the putative choices are in fact equivalent. It cannot be assumed that lexical choices are in fact language choices simply because the words in question have a different etymology.

In the remainder of this chapter, I propose a variety of motivations to account for many of the Malay items found in the corpus. My goal is to differentiate those lexical items which were chosen for their own sakes (i.e., those which have no convenient WT equivalent) from those which occur precisely because they are Malay. Then in chapter 5, the latter set of Malay items is examined to determine whether their use seems to be related to communicative functions commonly associated with code-switching.

4.2 Lexical prerequisites to LCP research

The analysis presented in this section is based on the notion that different languages have different ways of packaging meaning. For any given concept or proposition which a WT speaker wishes to convey, if there is a convenient lexical package (a word, a phrase, an idiom) in WT for that concept or proposition, then WT will be the default choice; if Malay is used instead, then I label it an instance of insertional code-switching. However, if Malay has a more convenient lexical package than WT for that concept or proposition, then it is less likely that the EL represents a language choice; rather, it is merely a lexical choice, and the chosen lexical item happens to be etymologically Malay. This latter category of insertions should be excluded from analyses of code-switching—for example, they do not qualify as counter-examples to proposed syntactic constraints on CS.

Gardner-Chloros (1995) proposed that there is no definite boundary between CS and other LCP such as borrowing, but that there is a continuum of LCP. I find support for this in the WT/Malay corpus. Just as some Malay lexical units are obvious instances of borrowing (e.g., items imported into WT culture from Malay-speaking culture), so also there are extended stretches of Malay which are clear instances of Malay mode (e.g., direct quotes of non-WT speakers). The latter instances are discussed in chapter 5. The current chapter examines the gray area between the two clear areas in the continuum.

Table 4.1 presents a summary of all lexical units in the corpus which are not in extended stretches of Malay to be defined as instances of Malay mode in chapter 5.[53] In this 15-hour WT/Malay corpus, there is a high number of EL tokens (22 percent being either wholly or partly Malay), and there are more types which are (etymologically) Malay than types which are (etymologically) WT: a full 63 percent of all types are wholly or partly Malay. In other words, while 78 percent of the tokens are pure WT,

[53] As is explained in chapter 5, the term "lone" here does not mean that there is no Malay item adjacent to the Malay item in question, but that if there is such an item, the two do not occur within a stretch of Malay which appears to result from a single language choice—that is, any adjacent Malay item is merely coincidentally adjacent.

only 37 percent of the types are. This disparity is due of course to the relatively small number of very high frequency WT words (e.g., conjunctions, pronouns, high frequency nouns and verbs) which identify the ML as being WT throughout nearly the entire corpus.[54]

Table 4.1. WT and lone Malay lexical units

	Tokens	Types
All lexical units not in stretches of Malay mode	82,973 (100%)	4,730 (100%)
pure WT	64,849 (78%)	1,761 (37%)
wholly or partly Malay	18,124 (22%)	2,969 (63%)

In the remainder of §4.2 I describe the various categories of Malay lexical units for which I argue that the speaker had no reasonable choice but to use Malay, due to the nature of the lexical unit itself. The number of lexical units in each category is shown in table 4.2; the meaning of each category will be explained in the subsequent discussion. My goal in describing these is not to define a set of mutually-exclusive categories, but to describe a variety of possibly overlapping motivations for the use of certain Malay lexical items. The columns in the table categorize Malay items as to whether they only occur as bare Malay forms, or only occur attached to a WT affix or word, or occur both as bare forms and as WT/Malay mixed forms.

Excluded from table 4.2, and from the subsequent discussion, are sixteen types (137 tokens) of Malay origin which the interviewees insisted were WT, not Malay.[55] Some of these are modified from their phonological shape in Malay (e.g., *tungguru* 'teacher' from *tuan guru* 'master teacher', *kotar* 'filth' from *kotor* 'dirty', *kuran* 'insufficient' from *kurang* 'insufficient'); one has undergone semantic shift (*manara* 'stuff' from *manara* 'tool kit'). Others are actually used by monolingual Dobo Malay speakers, but since they are not SIM, these WT speakers apparently believe them to be WT words (e.g., *orang kaya* 'village head' [literally 'rich person'],[56] *malikang* 'only if', cf. BI *melainkan* 'but rather', *pasáng* 'message', cf. BI *pesan* 'message'). The most frequently-occurring lexeme in

[54] It should be noted that for this comparison, an effort was made to keep the total number of Malay types as low as possible by combining all lexical units with the same Malay roots regardless of affixation or variation in senses. In the remainder of this chapter, it will become clear that the total number of Malay types is far higher if I count separately the senses of a Malay item which have different WT equivalents.

[55] These are marked with a superscripted W in examples.

[56] See Ellen (1986) for information on the etymology and range of usage of this term.

this category, however, is *sala* 'wrong' (22 unaffixed occurrences), which has the exact same form and meaning in Dobo Malay, but fits WT phonological word-form patterns perfectly, is often affixed, and was considered to be a WT lexeme by the interviewees. It is only identifiableas a loan by historical reconstruction.[57] Since the current work aims at providing a psychologically realistic account of synchronic language mixing, I ignore etymology, and even ignore usage within the monolingual Malay speech community, and focus instead on WT speakers' own perceptions of an item's language membership.

Table 4.2. Categories of lone Malay insertions

Category	Pure Malay (tokens/types)	Mixed (tokens/types)	Both (tokens/types)
Cultural imports (total)	5,585/1,344	287/160	105/11
Proper NPs	3,452/842	209/119	—
Job titles	277/68	7/6	—
Other	1,856/434	71/35	105/11
Other non-equivalents	3,504/417	437/142	840/52
Near-equivalents which lack some crucial component	256/103	76/32	95/12
Lengthy near-equivalents	72/48	26/14	4/1
Malay idioms and figurative senses	49/33	15/8	21/1
Displacements (total)	3,109/142	332/104	632/12
Kin terms	804/49	103/64	—
Other	2,305/93	229/40	632/12
Exclamative interjections	161/25	—	—
Repetitions	326/243	180/94	—
Metalinguistic usage	28/25	—	—
Non-default	1,116/467	414/185	351/53
Impossible to determine	84/62	19/15	—

[57] Dobo Malay *sala* is a reflex of Proto-Malayo-Polynesian **salaq*. The WT reflex of PMP **s* in initial position is zero, as in **sakay* > *-akai* 'climb' and **suan* > *uan* 'dibble stick'. Thus, sala with an initial /s/ can be identified as a loan word by linguists, if not by WT speakers.

Another word that requires special handling is the conjunction *mo*, which has the same set of functional meanings in both WT and DM. This conjunction has a variety of meanings having to do with contrast, including contrasting two propositions ("but"), contrasting what did not happen with what actually did happen ("instead"), describing two different events occurring simultaneously ("while"), transitioning from a setting clause to an event clause, and others. Since *mo* is neutral with regard to language membership, it was analyzed as a Malay word when the speaker was clearly speaking in Malay mode, otherwise it was considered to be a WT word.

4.2.1 Cultural imports

Cultural imports are the most obvious category of EL words which must be considered borrowings, and LCP researchers universally recognize this fact. However, not all researchers have taken account of this fact in their analyses. Poplack and Meechan (1995), for example, performed a statistical analysis on lone French nouns in Wolof and Fongbe contexts, without first excluding cultural imports from the total set of French nouns in their corpus. They found that most French nouns pattern more like Wolof and Fongbe nouns than like French nouns, and concluded that all such lone nouns are therefore borrowings. However, if a considerable number of these nouns are in fact cultural imports, then they are borrowings by definition and should not be allowed to skew the overall statistical analysis. In terms of both tokens and types, cultural imports are the largest category of lone Malay insertions in the WT/Malay corpus.

Some WT purists (especially, it seems, from the village of Popjetur) are noted for coining WT terms for cultural imports, e.g., *bilabilar* 'rolling thing' (from *da-bilar* '3p-roll') for 'car'. However, most WT speakers consider such word-coining funny and prefer to use the preexisting Malay term. The same negative attitude toward coining terms for cultural imports has been noted in other languages of southeast Maluku, such as Dobel (Jock Hughes, personal communication), Fordata (Craig Marshall, personal communication), Luang (Mark Taber, personal communication), and Selaru (David Coward, personal communication).

Proper noun phrases. Malay names for people, organizations, and so on are EL insertions which are beyond the control of the speaker. WT speakers typically have more than one name, i.e., a Malay name and a WT name (as well as an abbreviated Malay nickname), but usually only one of them is preferred in casual conversation; in Malay-speaking domains, of course, such as church and school activities, the Malay name

is generally preferred. In addition, people who have children are referred to by a teknonym such as *Peki amai* 'Peki's father' or *Losa jinai* 'Losa's mother', using the name of the oldest child; this is considered more respectful than using the person's given name.[58] Sometimes Malay phrases are used as nicknames, as in *Nyong Tua* 'older boy'. Included here are not only Malay names of people,[59] but also place names, names of things (e.g., products, books, songs, prayers, letters, religions, organizations, government programs), and events (including months and days).

Most instances of the Malay word *hari* 'day' are followed immediately by a day name. In fact, in the WT/Malay corpus the day names (which are all Malay) are always preceded by *hari*, as they often are in Indonesian. They are never (in this corpus, or in my experience) preceded by WT *maera* 'day'. I conclude that in these instances, *hari* is actually a part of the day name; in other words, for WT speakers, Malay day names are phrasal lexical entries.

Mixed words and phrases are indicated in the transcription by superscripted '%' and '/', respectively, preceding the lexical category code (in this case, 'P'). Proper names mixed on the phrase level include, for example, a Malay given name with a WT family name (*Anto Bôtmonamona*), a WT teknonym based on a Malay name (*Agus jinai, Eti amai*), a Malay respectful term of address followed by a WT name (*Bu Bôtmir*, *Usi Kabal*), a WT respectful term of address followed by a Malay name (*Taper Noya*), a WT name followed by a Malay descriptor (*Karelau Putih* 'white Karelau'), or a Malay name followed by a WT descriptor (*Naomi toptop* 'short Naomi'). As for mixing within the word, there are Malay roots with WT affixes (*Jepang-ai* 'Japanese-3pa'), WT reduplication (*Tepa-tepa*), or both (*Bu-bugis-na* 'DUP-Bugis-3sa'); there are also two words which exhibit a hybrid pronunciation: *Fatujurin* is a combination of Malay *Fatujuring* (a village in Aru) and WT *Pot Jurin* (that village's WT name), while *Lolor* is a Taranganized pronunciation of the village called *Lorlor* in Malay and *Garjá* in WT. In one instance, there is mixing on both the phrase and word level: *taper Cincinana* 'Chinese old man'.

Job titles. Traditional WT culture is fairly egalitarian; the main divisions of labor are based on gender and age. There are only a few

[58] An abbreviated teknonym may be used if preceded by a Malay kin term, as for example *Om Peki* 'uncle Peki' which LL uses to refer to her uncle, Peki's father.

[59] In the transcription, I linked any preceding title (e.g., *mas* 'young man', *bu* 'older man', *usi* 'older woman', *nyong* 'boy', *nona* 'girl') to the name as part of a single compound Malay name. Malay proper nouns and NPs are marked with a superscript P in examples.

specialists, such as the *orang kaya* 'village head' and the *papa duai* 'local spiritual intermediary' (literally 'master/owner of the land'). Today, however, there are many new titles of jobs and temporary assignments which relate to non-WT culture.[60]

Most of thee job titles, in terms of both tokens and types, are from the domains of government (e.g., *polisi* 'police officer', *tentara* 'soldier', *kapala kampung* 'village head') and church (e.g., *pendeta* 'Protestant pastor', *ketua* 'chairperson', *pengasuh* 'children's religious teacher', *majalís* 'elder'). The next largest set of terms comes from the school system (e.g., *guru* 'teacher', *kapala sekolah* 'principal'). The next most represented area is that of health services (e.g., *mantri* 'health worker', *bidan(g)* 'midwife'). Although these lexical frequency counts merely derive from the topics of conversation which happened to occur in the corpus, they may also be representative of both the extent of impact and the foreignness of these domains of modern WT life.[61]

There is at least one job title borrowed from a Javanese word which is not a job title in Javanese-that is, semantic shift has occurred in the transfer. Around 1990, the number of prostitutes in Dobo increased exponentially, and most of them were young women from Java. The current WT term for 'prostitute' is *ambá*, from Javanese [*mba?*] 'young woman'. In the WT/Malay corpus, *ambá* occurs six times as a common noun meaning prostitute, while *mba* also occurs once as a vocative addressed to a Javanese woman, and once in reference to a Javanese woman who was not a prostitute.

Other cultural imports. There remain over 500 other Malay lexical units (types) in the corpus which represent obvious cultural imports.[62] The vast majority of these are nouns, as might be expected.[63] But it is important to realize that the term "cultural import" here refers to concepts, not only to physical objects. Thus in some cases, while a referred item itself is not a cultural import, a particular perspective on that item is. For example, in (35) AD refers to traditional WT medicine using the Malay term *obat tradisional* 'traditional medicine'.

[60] Malay job titles are marked with a superscript J in examples. A Malay job title followed by a name, such as *Pendeta Layaba* 'Pastor Layaba', was classified as a job title, although it could just as easily have been classified as a proper name. Such overlapping of categories does not affect the present work, since the crucial distinction is between all of these categories together as default Malay, in opposition to non-default Malay.

[61] Interestingly, the WT term *papa duai* 'local spiritual intermediary' never occurs in the corpus, but its Malay equivalent, *tuan tanah*, occurs once.

[62] These miscellaneous Malay cultural imports are marked with a superscript C in examples.

[63] Counting types rather than tokens, there are 344 nouns and 92 verbs.

(35) \AD ***[NF]Memang**, [e] kam' [%N]ma**dapa** [o] on min pei [J]**dokter** [.] aka darokrok [e] [C]**obat-obat•tradisional** êr.*
Indeed, [uh] we got [o] this too from a doctor [.] about looking for [uh] traditional medicine.

Traditional medicine is obviously not a cultural import; but the conceptual distinction between traditional medicine and other kinds of medicine (Western medicine, Chinese medicine) is. By using this term, AD is looking at WT culture from the perspective of an outsider, no doubt using the term used by the doctor in the conversation she is recalling. Although she could have replaced *obat* with the WT equivalent *gakar*, there would be no convenient way to convey the meaning 'traditional' in WT except perhaps by the modifier *jarjár* 'ordinary'.

In (36), TN uses Malay *kapala* 'head' and *polo* 'hug' in secondary senses as names for two types of pillows.

(36) \TN *Kunaka nei. «[Y]**Bantal** din [NF]**kan** lat, [NF]**to**? [C]**Kapala** rua, [C]**polo** ôt. Ken môl [C]**kapala** on, ka nin moi ne,*
I told him. "There are three of these pillows, you know? Two head pillows (literally 'heads'), and one bolster (literally 'hug'). You take this 'head' and sleep over there,

The word *bantal* 'pillow' is itself non-default Malay, since there is a common WT equivalent (*kalangúr*) which occurs five times in contrast to only two instances of *bantal*; *kapala* and *polo* would also be non-default if they were used in their primary senses;[64] but TN informed me with a laugh that the WT equivalents can not be used in any form as names for pillows. In this sense, therefore, *kapala* and *polo* are cultural imports.

Examples (37)–(39) illustrate how puristic tendencies can cause frustration. In turn 1 of (37), SD uses the WT phrase *-tir gar* 'bathe in freshwater' as an uncommon equivalent of the established Malay loan *permandian* 'baptism' (literally 'bathing'). In turn 2, AD challenges his WT phrases, suggesting another (*-uk gar* 'use fresh water'). In turn 3, SD rejects AD's suggestion, but DA joins AD's protest, and in turn 6 SD finally caves in. As turn 6 continues, SD uses another uncommon WT phrase, *gareja kanei gaibúnbúnna* 'church leader' (literally 'church's big-one'), to replace Malay *pendeta* 'pastor'.[65] At the end of turn 6, SD is apparently trying to come up with a WT equivalent of *kapala kampung* 'village head', but after a two-second silence, AD suggests the Malay phrase, and SD uses it.

[64] The words *kapala* and *polo* do not occur at all in the corpus in their primary senses of 'head' and 'hug'; WT *jikala-* 'head' occurs 21 times, and *-tabôr* 'hug' occurs 4 times.

[65] The fact that SD began to say *jinjinaina* 'big-one' rather than *gaibúnbúnna* may constitute evidence that his phrase was an equivalent created on the spur of the moment, rather than a standard WT term for 'pastor'.

(37) 1\SD ᶜ***rapat*** *aka [:] dadem [a]* ᶜ***panitia*** *[.] aka jertir•gar gaibúnbún on,* ᴾ*Taper•****Noya*** ᴰᶠ***waktu•***^*•ne*
meeting for [:] making [uh] a committee [.] for this big baptism, Mr. Noya at that time
2\AD *Jeruk•gar gaibúnbún.*
^ Big baptism.
3\SD *[Je-] Ah* ᴺ***barenti****. Jertir•gar ka* ᴺ***jadi****.*
[je-] Oh, cut it out. "Jertir gar" works too.
4\AD *Daukuk•gar.*
Baptism.
5\DA *Dauk•gar.*
Baptize.
6\SD *Daukuk•gar daidai (permandian ei) Kalakalar on,* ᴾ*Taper•****Noya****,* ᶜ***gareja*** *kanei [a] [jinjin-] gaibúnbúnna ime Perín. Nei ei* ᴶ***ketua****.* ᴰᶠ***Tarús****, nei kanei* ᴶ***sekertaris****,* ᶜ***gareja*** *kanei gaibúnbúnna ina min, ime Ngeíl. Taper•[a]•Kilái. Nampún [a]* ᴶ***wakil•ketua****,* ᴾ***Yohanes•****Lailaiem. Kalakalar kanei [a] [2]*
Baptism at (baptism at) Kalar-Kalar, Mr. Noya, the church's [uh] [big-] leader (literally big-one) who lived at Feruni. He was the chairman. Then, his secretary, another church's leader, who lived at Ngaiguli. Mr. [uh] Kilái. Then [uh] the vice-chair, Yohanes Lailaiem. Kalar-Kalar's [uh] [2]
7\AD ᴶ***kapala•kampung****.*
village head.
8\SD ᴶ***kapala•kampung****.*
village head.

Later in the same conversation, SD successfully remembers AD's preferred WT phrase for baptism in the middle of turn 1, then apparently forgets it at the end of that turn—the two-second pauses before and after jertir gar must reflect either his confusion or some nonverbal communication he is receiving from AD, who corrects him in turn 2.

(38) 1\SD *Galái arei•ba* ᴺᶠ***yang*** *[a]* ᴬ***siap*** *akaka [:] perái ibanabana [:]* ᴾ***Dobo****. Dapalen=ai dasenin, loloar tene [a] [.]* ᴰ***istirahat****, ka [2]* ᴬ*da****cari•tahu*** *[a] [.]* ᶜ⁺***calon-calon*** *[a] [.] daukuk•gar ir. Ja, maera•êr* ᴬ*da****cari•tahu*** *[a]* ᶜ***calon*** *ir,* ᴺ*da****dapa*** *rat•lat•mo•[.]•tarai. Mo [a] pen dabali damin.* ᵞ*Da****janji*** *loloar tene* ᶜ***jam•***ᴺ# ***delapan****, [2] jertir•gar [a] [2]*
The several houses which [uh] were ready for [:] the crowd from [:] Dobo. They finished distributing themselves, the

next day [uh] [.] they rested, and then [2] they went to find out [uh] [.] the prospective [uh] [.] baptizees. So, that day they went to find out about [uh] the candidates, they got three hundred and [.] some. But [uh] some stayed away too. They announced that the next day at eight o'clock, [2] the baptism [uh] [2]

2\AD *Daukuk•gar.*
Baptism.

3\SD *Daukuk•gar êr [a] [.]* ***[N]muláí.***
The baptism [uh] [.] would begin.

Still later in the same conversation, example (39), SD again uses AD's suggested phrase at the beginning of turn 1, then at the end of that turn is apparently confused again about which one to use; YD (his son) suggests the phrase SD had wanted to use previously, and surprisingly AD accommodatingly echoes YD's suggestion, at the same time that SD remembers and uses AD's suggested equivalent.

(39) 1\SD *dapiam =na aka [.] itorai daukuk•gar. Nei [a] nena* ***[N]pertahankan*** *êr, baitan [a]* ***[J]Ketua•Klasis*** *itora nei ayei daruk=ai dasí, ka [.] dam•on; «Kenjou ken [a] [kanám•suka-] kanám narpet aka lebá•ia, ja morsirpei.» Nei ersirpei nekanei narpet [n],* ***[J]Ketua•Klasis*** *itora nei ayei* ***[NF]musti*** *damarer aka nei. (Dam•on,) «Oh, ken (mo) morpet one, ja (ma) morsirpei (môl a)* ***[N]muláí*** *êr, to [.] kama mamarer aka ken. Mo one [%N]su****susah****na tareidi =e?» (na) daramasal dal nei, ka [.]* ***[C]lonceng*** *aka [:] jertirgar êr, mo [eiaa]* ***[C]jam•[N#]sepuluh*** *=si. Ja maera•ne, êra* ***[J]pandeta*** *[:] urpapa•lat•mo•tarai (aa) dam êra rat•lat•mo•tarai êr [da-]*
They persuaded her to [.] go along and be baptized. She [.] was holding out, later [uh] the head of the Klasis together with his wife went in to see her, and then [.] said, "If you [uh] [want-] want something, say so." She told her desire, the head of the Klasis together with his wife must stand for her. (They said,) "Oh, (if) you want that, then you (should have) said so (from) the beginning, so that [.] we would stand for you. Is that so difficult?" They strove and succeeded to get her, and then [.] the bell rang for [:] the baptism, and [eiaa] already ten o'clock. So that day, those 30-some pastors (aa) made those 300-some people [3p-]

2\YD *tirtir•gar*
be baptized.

3\SD ^ *daukuk•gar.*
be baptized.

4\AD ^ *tirtir•gar*.
be baptized.
5\SD *Êra ina na [i a] emarer, ka= [:] inam êra urpapai•ia [i] dauk•gar, [.] ja elapei, mo êra ina [:] isí*.
Each one of them [i a] stood, and then [:] he made ten of them [i] be baptized, [.] then came out, and another of them [:] went.

4.2.2 Other Malay items with no WT equivalent

There is a large category of lexical units which apparently have no WT equivalent, but are not necessarily cultural imports.[66] The vast majority of types in this category fail Myers-Scotton's three-occurrence rule—that is, they only occur once or twice in the corpus—and therefore Myers-Scotton might consider them instances of CS rather than borrowing. Yet they are clearly instances of necessary Malay—a speaker would be hard pressed to convey the same concept in WT.

There are over 4,000 tokens and over 600 types in this category. In stark contrast to the category of cultural imports, there are about twice as many verb types as noun types here. The lack of one-to-one correspondence between the semantic ranges of lexical items of two languages is more pronounced in verbs than in nouns, since semantic groupings of activities and states do not have naturally discrete boundaries like many groups of objects do (or at least seem to). As a result, a WT speaker bilingual in Malay may decide that the semantic range of a particular Malay verb which does not correspond to the semantic range of any WT verb is convenient enough to merit use of the Malay verb, particularly if it is used in a domain which is more salient in modern (Indonesianized) WT culture than it was in traditional WT culture.

Only eighty-eight of the pure Malay items in this category occur five times or more. Ten of these items occur over fifty times; their high frequency is due to the fact that they are members of closed syntactic classes. These ten are the conjunctions *padahal* 'however' (110 instances), *biar* 'although' (71 instances), *berarti* 'it means that' (64 instances), and *baru* 'what's more' (58 instances); the adverbs *memang* 'indeed' (121 instances), *lebái* ~ *lebe•bai* 'preferably' (66 instances), and *macang* ~ *macam* 'kind of' (59 instances); the relativizer *yang* 'which'

[66] These Malay items with no WT equivalent are marked with a superscript N in examples. Some of these items may well encode concepts which are cultural imports, but since I do not know them to be such I have categorized them here. Again, it is not important to the main thrust of the present work to establish exact boundaries between the various types of default Malay.

(381 instances); the auxiliary *bisa* 'can' (202 instances); and the copula *ada* 'exist, be at, have' (193 instances).

4.2.3 Malay items with difficult WT near-equivalents

In this corpus, the genre of discourse is casual conversation. The whole point of conversation is to communicate—not merely referential meaning, of course, but also communication of identity, and of perceived or desired social relationships, statuses, and roles. For some purists, it is important to communicate to others their devotion to language purity, and I have met a few WT speakers like that. But the vast majority of WT speakers are not interested in impressing their close friends with their ability to avoid Malay. For the most part, the speakers in this corpus seem to choose the most convenient way of expressing referential information—SD is in effect the exception that proves the rule, since his obviously careful style includes a browbeaten purism alien to the rest of the corpus. In this section, I group together four categories of Malay items which would not be impossible to express in WT, but seem to be more conveniently expressed in Malay.

Lengthy near-equivalents. Some Malay lexical units have WT near-equivalents which are significantly longer than the Malay item.[67] Nearly all of these have only one or two occurrences, and there are only about 65 types in this category. I agree with Li's (2001) principle of economy which states that speakers will often choose the most concise way of expressing a concept no matter what its etymology, unless other factors are considered more important. Therefore, despite their low frequency of occurrence in the corpus, I claim that these represent lexical choice rather than language choice. In table 4.3, I display a few of the lexical units in this category, along with what the interviewees considered to be the best equivalent in WT for the context where the Malay items occurred.

[67] Malay items with lengthy WT near-equivalents are marked with a superscript L in examples.

Table 4.3. Some Malay items with lengthy WT equivalents

Malay	WT
(ibu) janda 'widow'	*konar yang idá lêt dapan dapeipeiai* 'women whose husbands have fallen from them'
kontáng 'paid in full'	*dapayar dasena* 'they paid completely'
ada timbangan 'there is balance'	*takom itil tantan pênai ia* 'so it doesn't lean against one side'
balas 'take revenge'	*nar êra min* 'hit them too'
siku 'jab with the elbow'	*uk ik matai aka* 'use the elbow on'
tobat 'stop being bad'	*takoman dadem minmin* 'they do not do it again'
takaná 'affected by an evil spirit'	*ertom enen samasamai* 'encountered an evil spirit'

Overlapping with the category of figurative senses (see below) are words like *simpang*, a Malay verb meaning 'to store' which also has the figurative sense 'to nurse a grudge against (someone)'. OK, who used the word in the corpus, offered in the interview the WT equivalent *kukatút koi ok abalnga* 'I store it in my soul'. The WT verb *-katút* 'store' does not have the figurative sense that *simpang* does, so a longer phrase must be used to avoid the literal meaning.

AD demonstrates the difficulty of finding brief WT equivalents in (40). She pauses, possibly searching for the WT equivalent of *ibu-ibu janda* 'widows', but then uses the Malay term. Immediately, apparently as an afterthought, she replaces this choice of words with the less specific WT term *gasirasira* 'old women'.

(40) \AD *maera ia, kam'* %N*maban***bantu** *[e] [2]* L+***ibu-ibu•janda*** *on. Gasirasira natapen êr.*
one day, we were helping [uh] [2] these widows. The old ladies of the village.

In the interviews, AD offered *konar yang idá lêt dapan dapeipeiai* 'women whose husbands have fallen from them' as an equivalent for the Malay *ibu janda*. But in the course of actual conversation, in an apparent case of replacing her Malay choice of words with a WT equivalent, she chose a word (*gasirasira*) which did not convey the full meaning of *ibu janda*. If she had used only the WT term, the intended referents would probably have been wrongly identified by the hearers.

As another example, Farb (1974:226) stated that the Chinese kin term *piaomei* is equivalent to a long phrase in English, 'a female cousin on my mother's side and younger than myself'. While this is a suitable dictionary definition, we would not expect a Chinese/English bilingual to use it when speaking English and referring to such a relative. Rather, when referring to such a relative in English discourse with another Chinese/English (or English/Chinese) bilingual, he would have to decide how many semantic components of *piaomei* were relevant to the message he wanted to convey. If all or even most were relevant, he would certainly use the Chinese word rather than a cumbersome English phrase.

Near-equivalents which lack some crucial component. In contrast to the category of lexical units described in §4.2.2, where neither I nor the interviewees could come up with a WT equivalent, there are Malay items for which a WT near-equivalent can be proposed but is lacking some denotational or connotational component which seems to be desirable to the speaker. In other words, although a WT term might be usable, the best term for the given context seems to be the Malay term.[68]

This category contrasts with the category of lengthy WT equivalents, for which a longer and natural-sounding WT phrase would be just as appropriate as the Malay term except for its length. In this category, I was dealing with a speaker's subjective sense that one word is better than another.[69] If a speaker merely likes the feel of one word over another, they are not really equivalent even if they are denotationally or referentially equivalent.

Although the decision here is not as clear-cut as in other categories (either for the speakers or for the analyst), I propose that these Malay items represent the most convenient lexical package for the concept that the speaker wished to convey. As such, they still represent lexical choice rather than language choice. The problem with using bilingual dictionaries for identifying borrowings is especially evident in this category: the dictionary maker considers it his duty to fill in the blank with the best near-equivalent available,[70] but this does not mean that a bilingual speaker considers the pair of words to be truly equivalent.

When a speaker claims that two words "feel different", there may be a variety of factors involved. Perhaps the words have different connotations,

[68] Malay items in this category are marked with a superscript A in examples.

[69] The analysis at this point is necessarily subjective as well, and to that extent less reliable than it is in other sections.

[70] As noted in chapter 3, my interviewees at first had the same desire to merely fill in the blanks with WT equivalents, until I showed them the discourse contexts from which the words in question were extracted.

or belong to different registers; one may be inherently humorous or sarcastic; the two may have different implicatures; the use of a particular word may call to mind a previous experience shared by speaker and hearer; there may even be a sense of phonological aesthetics which renders one word better than another for a given context.[71] Or, the two words may have slightly different semantic ranges—the relation may be one of generic to specific, or the two may be generally equivalent but fail to completely overlap semantically. In a context where a Malay word fits well and its best WT equivalent is not quite as appropriate for some reason, one should expect the typical, nonpuristic speaker to simply use the Malay word, rather than attempting to stretch the WT word to fit and hoping that the hearers understand.

Again in this category, the vast majority of items occur only once or twice; only eleven occur five times or more. The most frequently-occurring item in this category is the conjunction *jadi* 'therefore' (95 instances), which has a much narrower semantic range than the nearest WT equivalent *ja* 'so, then' (1,853 instances). Although *ja* is a conjunction of high discourse frequency, and would probably fit in every instance where *jadi* occurs, *ja* has a broader semantic range; it occurs in contexts where *jadi* would not fit, as in (41).

(41) \TB *Dabebar, ja daltúk! Jonjou [e]* ***N barani****, ja* ***NF bisa*** *%D da****sampi kôla****.*
If they're scared, then they come back up [to the surface]! If [uh] brave, then they can reach the sand.

The second most frequent item in this category is *manusia* 'human being' (25 occurrences). The word *manusia* has a slightly narrower semantic range than WT *tamata* 'person', which usually refers to human beings but also includes ethereal beings. Thus, when a speaker wishes to contrast these two types of persons as in (42), *manusia* is the best choice.

(42) \AD *«Ok gulkalang on,* ***PD Tuhan*** *ijir. Mo sakali* ***A manusia*** *ina ijirjir.»*
"This hair of mine, God created it. It wasn't a human being who created it."

In addition, *manusia* often carries a humorous tone, often implying a proud attitude on the part of the person referred to (or, perhaps, a sarcastically demeaning attitude on the part of the speaker) as in (43).

(43) \TB ***A Manusia*** *[dan-] danonongau %DF rupa****rupa*** *ine, danar =ai baitan* ***S stengah-mati NF boleh****.*
Human beings stealing like those ones do, it would be best to beat them half dead.

[71] This is certainly the case with poetry, but I have no evidence for aesthetically-based choices in these conversations and will not further explore the possibility.

Another Malay item with a WT equivalent that does not quite feel right is bajual 'to sell (habitually)', which occurs six times, as in (44).

(44) \YL *Erteya, ka maera•êr* ^{A}*bajual má [:] on, ja ok* ^{Y}***tagor*** *= na = sin.*
She got married, and then one day she came here selling, so I called a greeting to her.

WT has a transitive verb *-gong* 'sell', occurring twenty-nine times in the corpus, which can be made roughly equivalent to *bajual* by the addition of the actor-focus prefix *r-* (*da-r-gong* '3p-sell things' intransitive). In fact, *-r-gong* does occur four times in the corpus, but always immediately followed by a noun (*sima* 'fish', *letai* 'sailboat') specifying what was sold; the function of the actor-focus prefix is to indicate that the postverbal noun has no specific referent, it merely indicates the type of thing sold. In contrast, *bajual* is never (in this corpus) followed by a noun indicating what was sold, even in a generic sense. The WT verb *-r-gong* is therefore not quite equivalent to *bajual*. WT speakers apparently feel no need to stretch (even slightly) the usage of a WT verb when there already exists a high-frequency Malay verb. Coining a new usage for *-r-gong* might cause confusion or make the speaker seem odd to his peers.

Malay idioms and figurative senses. There are a few Malay idioms found in the corpus; it would not be reasonable to expect speakers to translate or calque[72] them into WT rather than use them as is. In addition, sometimes the stylistic choice of a figurative sense involves an unavoidable EL insertion, since the equivalent WT lexeme cannot carry the same figurative sense. These, therefore, are further examples of unavoidable Malay forms. Again, nearly all have only one or two occurrences in the corpus, but must nevertheless be considered borrowings.[73]

Once AD referred to YL being *basah* 'wet', meaning *tangan basah* 'caught red-handed' (literally 'wet hands'). The equivalent WT *jemil* 'wet' cannot have that meaning. Similarly, OK used Malay *makan* 'eat' in a figurative sense when referring to the walk-in freezer at the local shrimp company, saying *es makan batúl-batúl* 'the ice really eats (into your skin).' Although this entire sentence is Malay, all three words are default Malay, and the preceding and following sentences are mixtures of WT and Malay, with only one non-default Malay word. I conclude that OK was not in Malay mode here. Interviewees AD and LL told me it might be possible to use WT *-ka* 'eat (transitive)' in that figurative sense, but TN disagreed with them.

The most frequent Malay lexeme with a figurative sense (21 instances) is *stengah-mati* 'undergoing extreme difficulty' (literally, 'half-dead'), which has a

[72] The term "calque" as used here refers to ML words in an EL structure.

[73] Malay idioms and items with figurative senses are marked with a superscript S in examples.

fairly broad semantic range not matched by any WT lexeme. This lexeme also has a high frequency of occurrence in monolingual Malay speech. The next most frequent is the mixed phrase *-gong obat* 'spread lies' (literally, 'sell medicine'), which occurs seven times and is apparently a calque of the Malay idiom *jual jamu* 'spread lies' (literally, 'sell herbal tonic'), which also occurs once.

Other idioms resist calquing, however. For example, *cari jalang* 'look for a way to do something' (literally, 'look for a path') consists of two Malay words which are firmly dispreferred, though not quite gratuitous.[74] Attempting to replace them with their common WT equivalents, however, results in a phrase (*-r-ok jala*) which can only be understood literally.

The third most frequent lexeme of this category, with four tokens, is Abu Nawas, the name of a trickster character in southeast Asian folklore. In (45), the name Abu Nawas is actually used as a name, not figuratively; in the figurative usage, as represented in (46), Abu Nawas is used as the name of a character trait.[75]

(45) \YL *One* ***[N]riwayat [Y]itu,*** *ono* *[a] [.]* *moiraka*
that.INAN life.story that that.INAN uh [.] 2s.know
[P]Abu•Nawas.
Abu_Nawas
That, that life story, that, uh, you know about Abu Nawas.

(46) \AD *Jou* *nen* ***[P]Au*** *kanei* ***[SP]Abu•Nawas*** *on.»*
see this.AN Au POSS.3s Abu_Nawas this. INAN
Look at this Au's deceitfulness.

Displacements. It is axiomatic in historical linguistics that basic vocabulary is the most stable part of the lexicon. Basic vocabulary items have a high discourse frequency, so these lexical items are less likely to be replaced by innovative forms. In the same way, a concept which occurs frequently in monolingual Malay discourse becomes strongly associated with the Malay form that expresses the concept. If the same concept is not very frequently referred to in monolingual WT discourse, the stage is set for displacement of the WT term by the equivalent Malay term. The tenuous status of the WT words is weakened further by the fact that WT does not have the same official and institutional support as Malay, nor is the speech community as large or prestigious.

Some Malay words classified as displacements here represent concepts known to have been lexified in WT in the past, as evidenced by data from older speakers. Some of these are shown in table 4.4. Other displacements

[74] Malay *jalan(g)* 'path' occurs twice, compared to 27 instances of WT *jala*; Malay *cari* 'look for' occurs ten times, compared with 72 instances of WT *-r-ok*.

[75] The underlining in (45) indicates an instance of Malay mode, to be discussed in chapter 5.

have WT equivalents which are still current in the language community,[76] but the speakers in this corpus have chosen to use the Malay form instead; some of these are shown in table 4.5. If the interviewees claimed that a given Malay lexical unit was gratuitous, but lexical analysis of the corpus showed that the Malay was used at least as often as its most frequent WT equivalent, I overruled the interviewees' judgments and coded the lexical unit as an instance of displacement. Finally, there are a few words, such as Malay *sayang* 'love', which I consider displacements because I assume that the referred concept was lexified in WT in the past, although I have no evidence of such.[77]

Table 4.4. Displacements with obsolete equivalents

	Malay (instances)	WT (instances)
flag	*bendera* (5)	*lopilopi* (0)
behavior	*kelakuan* ~ *kalakuan(g)* (4)	*jêrai* (0)
ship	*kapal* (20)	*kabal* (0)
gold	*mas* (7)	*pulán* (0)
namesake	*sinama* (5)	*temun* (0)
lean against	*sandar* (4)	*-til nata* (0)
traveling food	*bakál* (9)	*sean* (0)
room	*kamar* (38)	*godong* (0)

Table 4.5. Displacements with current equivalents

	Malay (instances)	WT (instances)
hot drink	*air panas* (6)	*gar rararai* (2)
rear	*belakang* ~ *balakang* (33)	*mir* (10)
relative	*sudara* ~ *basudara* (32)	*jerakata* (6)
pay	*bayar* (18)	*-payar* (11)
stupid	*bodoh* (7)	*duku-* (2)
cost	*harga* (6)	*pêl* (4)
bump into	*kanál* (13)	*-r-tom* (9)
plan	*rencana* (12)	*-ser* (3)
arrive at	*sampi* ~ *sampe* (103)	*-pan* ~ *-pan nal* (24)
midday	*tengah-hari* (5)	*gasira erua* (3)
post	*tiang* (8)	*anga* (3)
feel	*rasa* (26)	*-perat* (3)

[76] I believe that anyone considered by other WT speakers to be a fluent WT speaker would know these words.

[77] A more complete historical reconstruction of an immediate ancestor of the languages of Aru might turn up such evidence.

One conspicuous lexical domain in which Malay terms seem to be displacing the WT terms is that of kinship terminology. Although nearly all of the WT kinship terms are well-known to all of the speakers recorded in the corpus, they are in the habit of using Malay equivalents instead. In contrast to some language contact settings, most of the kinship terms in the local variant, Dobo Malay, are semantically equivalent to the WT terms.[78] Some of these kin terms are shown in table 4.6.

Table 4.6. Some WT kin terms undergoing displacement

	Malay (instances)	WT (instances)
younger sibling	*adi* (65)	*jel-* (16)
father	*bapa* (179)	*bôy* (18)
		ama- (3)
sibling-in-law	*ipar* (34)	*joir* (0)
uncle	*om* (43)	*duai* (0)
		jai (0)
aunt	*tanta* (37)	*tití* (0)

I speculate that the reason these WT speakers use Malay kin terms so much is that they have contracted a number of close relationships in the Dobo community, both true kin and pseudo-kin, with Malay speakers. Moreover, in Malay these terms are honorifics, so they are high-frequency terms used even with strangers. I predict that rural WT speakers will not have such a high frequency of Malay kin terms; further research will be required to determine whether or not this is so.[79]

In addition to kin terms, there are a number of other Malay terms which represent more or less recent displacements of WT lexical units which are equivalent in every way, including convenience.[80] For example, the WT term for year is *narak*, but many speakers under age 30 do not know it. In the corpus we find Malay *tahun* (with both standard and nonstandard pronunciations) throughout, except for four instances in SD's hypercorrect speech.[81] There are two evidences of SD's hypercorrectness in (47): first,

[78] The English glosses are of course inaccurate, since the WT kinship system is very different from the kinship system encoded by English. See S. Nivens (1990) for details.

[79] Phrases consisting of kin terms combined with names were categorized as proper nouns in this analysis.

[80] Malay kin terms are marked with a superscript K in examples. Other Malay displacements are marked with a superscript D in examples.

[81] The situation here is therefore different from that reported for the Asilulu language of Ambon Island by James Collins (personal communication), where people use Asilulu *nale* 'year' to say how many years old a person is, but use Malay *tahun* to name the year a person was born.

he says *tahun* then repairs it to become *narak*; second, he even gives the number of the year in WT, which is very odd.

(47) */DF**Waktu**•ne [D**tahun** a] [.] narak•ur•dubám•mo•[:]•rua.*
At_that_time [year uh] [.] year_seventy_and_[:]_two
That was in [the year uh] [.] the year 1972.

Similarly, SD is the only speaker who never refers to God as *Tuhan* 'Lord' or *Tuhan Allah* 'Lord God'; in four occurrences of the concept, he uses only the WT name *Jirjir Duai* 'master/owner of creation', which is as far as I have been able to determine an exact equivalent[82] of the Malay *Tuhan (Allah)*. In contrast, AD uses *Jirjir Duai* five times and *Tuhan* or *Tuhan Allah* fifteen times, while OK uses the Malay term six times and never uses the WT term. None of the speakers ever uses the WT term as a vocative.[83] The displacement of *Jirjir Duai* by *Tuhan* is obviously related to the fact that nearly all WT speakers participate in Malay-speaking religious organizations.

In contrast to the other categories of default Malay described thus far, only about one-third of the lexemes in the displacement category (not counting kin terms) have only one or two occurrences. Over half have five or more occurrences. This is due to the nature of displacements, or perhaps I should say due to my method of identifying displacements: there are a number of lexical items identified as displacements based on the fact that they occur more frequently in this corpus than their WT equivalents do.[84] Some of these are items which were identified by the interviewees as gratuitous; but since it is common in linguistic research for an informant's self-reporting to be different from his actual speech output, I am taking frequency to be a stronger indication than the interviewees' claims as to whether the normal expression of a given concept is in WT or Malay.

In examining the corpus, I found it essential to pursue a fine-grained lexical analysis rather than focus on word forms alone. For example, the concept 'near (in distance)' occurs nineteen times as WT *den*, and only eleven times as Malay *dekat*. But the concept 'near (in time)' occurs six times as *dekat*, and never as *den*. I therefore conclude that *dekat* referring to distance is non-default Malay, but *dekat* referring to time is default

[82] That is, for any given speaker the two terms are equivalent. It is unknown whether the term *Jirjir Duai* predated the arrival of the GPM among the Taranganese; the Dobel language, at least, has no cognate term (Jock Hughes, personal communication), nor does the Barakai language (Patricia Spyer, personal communication).

[83] The vocative use of God's name is functionally different from the expletive use, though formally similar; the expletive use is excluded from the figures given here.

[84] Recall that I am only considering here Malay lexical units which do not occur within longer stretches of Malay mode, to be defined in chapter 5. If Malay lexemes within Malay mode were counted as well, the Malay frequencies of many of these items would be even higher.

Malay. I do not know if *den* has ever been used in actual speech to refer to nearness in time, but my transcribers twice replaced a temporal usage of *dekat* with *denden* in the transcription.[85] In any event, I assume there was some way in a previous variety of WT to refer to nearness in time; I therefore conclude that this usage of *dekat* is most likely a displacement of that hypothetical previous lexical item.

Members of closed syntactic classes typically have high discourse frequency, and in this corpus several such items are ubiquitous, including the conjunctions *tarús* ~ *terus* 'then' (729 occurrences), *tapi* ~ *tetapi* 'but' (392 occurrences), *barang* 'because' (177 occurrences), and *jadi* 'therefore' (95 occurrences), as well as the preposition *sampi* ~ *sampai* 'until' (161 occurrences) and the adverb *langsung* 'immediately' (101 occurrences). All of these have current WT equivalents, although *tapi* and *jadi* have narrower semantic ranges than their respective WT equivalents *mo* and *ja*. I consider them all to be displacements; although *tapi* and *jadi* have not completely displaced *mo* and *ja*, they have clearly displaced a portion of their semantic ranges.[86]

4.2.4 Exclamative interjections

Poplack, Wheeler, and Westwood (1987/89:139) defined TAGS as "a freely moveable category with no syntactic relation to the rest of the sentence," giving such English examples as *no way!* and *thank God!* I assume that non-exclamative interjections such as *you know* would fall in this category as well; in fact, the difference between this category, which has become entrenched in the CS literature, and the traditional category of interjections is unclear to me. Poplack et al. report that in their Finnish/English corpus, about 19 percent of the English intrasentential insertions are tags. They contrast this relatively small number with the ubiquitous use of Spanish tags by nonfluent bilinguals in the New York Puerto Rican community, apparently as ethnic identity markers.

In the WT/Malay corpus, I find only about fifty types of Malay exclamatives totalling about 160 tokens.[87] Even if I add to this number the sentence-medial particle *kan* 'you know' (129 occurrences) and sentence-final particle *to* 'you know' (392 occurrences), the relative

[85] These were in the transcriptions of the first two recorded conversations, before I realized that the transcribers were editing out some Malay words and instructed them not to do so.

[86] If the conjunction *mo* in DM is a loan from WT, then it is ironic that in DM *mo* has displaced part of the semantic range of *tapi*, the reverse of what is happening in WT.

[87] It may be appropriate to consider these Malay interjections to be displacements, since I assume the emotions they give voice to were already present in WT culture and that equivalent WT exclamations were once available (and perhaps still are).

proportion of tags is far less than even that reported for the Finnish-English corpus.[88]

Most of the Malay exclamatives in the corpus are words or phrases which have meaning in other contexts. Variations on *Tuhan (Allah)* 'God' are the most prevalent (51 tokens), followed by *kasi(h)an(g)* 'sympathy' (27 tokens). But there are also exclamatives which have no meaning other than the exclamative sense, the most frequent being *co(u)*, indicating astonishment, which occurs thirty times. Native WT exclamatives, in contrast, are nearly all of the latter type; the only exception known is *sopar* 'blood', which occurs nine times.

4.2.5 Numbers

The use of numbers in this corpus clearly demonstrates the need not only to examine lexical units and domains separately, but also to examine the contexts in which lexical units occur. In this corpus, numbers referring to dates, times, sizes of boat engines, and school grades are almost always Malay, while numbers referring to money may be either Malay or WT, and other numbers are typically WT.[89]

Dates, times, boat engines, school grades, and money are clearly cultural imports, so it should not surprise us that the numbers associated with them tend to be Malay as well. In fact, there is a clear reason for the difference in number patterns between dates, times, and school grades on the one hand and money on the other: the numbers used with the former are labels, not numbers actually used for counting amounts of things. With money, however, numbers are used for counting amounts—note that *doit* 'money' is a count noun for WT speakers, not a mass noun as 'money' is in English. Note also that a difference between the syntax of Malay and WT is relevant here: as mentioned in chapter 2, all noun modifiers in WT (including numbers) follow the noun, but in Dobo Malay as evidenced in this corpus (i.e., as spoken by these speakers), all noun modifiers except counting numbers follow the noun.[90] Thus, the Malay phrases j*am dua* 'two o'clock' and *dua jam* 'two hours' would both be rendered *jam rua* if a speaker insisted on using WT *rua* 'two'; apparently, to avoid such ambiguity, WT speakers use WT numbers for counting hours (*jam rua* 'two hours') and Malay numbers for telling time (*jam dua* 'two o'clock'). Similarly, *tanggal dua* 'the second day of the month' (literally, 'date two') is used rather than *tanggal rua* 'two dates' (e.g., two dates to choose from). But no such problem exists with money: both *dua ribu rupiah*

[88] Malay exclamative interjections are marked with a superscript X in examples.

[89] Malay numbers are marked with a superscript # in examples, following one of the other codes mentioned above.

[90] In Dobo Malay, sometimes even counting numbers follow the noun.

'two thousand rupiah' and *rupiah ripun rua* 'two thousand rupiah' are acceptable and unambiguous (though the latter pattern never occurs in this corpus).

Myers-Scotton (1993b:195–201) mentioned a similar differential usage of numbers in her discussion on numbers in her Zimbabwe (Shona/English) corpus.[91] In that corpus, 86 percent (1,079) of all numerical expressions (N = 1,257) were in English, leaving only 14 percent (178) in Shona. She considers a variety of semantic categories, reproduced here in table 4.7 along with the relative frequency of numbers in these categories (and others) found in the WT/Malay corpus.[92]

Table 4.7. Numbers in two corpora

Concept	English in Shona	Malay in WT
Duration of time	123/176 (70%)	20/51 (39%)
Point in time	476/500 (95%)	144/153 (94%)
> year number		5/9
> time of day		89/92
> calendar date		48/50
> month		2/2
Count	91/164 (55%)	7/328 (2%)
Money	44/46 (96%)	59/72 (82%)
Labels	5/5 (100%)	28/28 (100%)
Age	154/179 (86%)	—
Educational level	186/187 (99%)	—
# by #	—	1/7 (14%)
'one' = 'same'	—	8/16 (50%)
Measurements	—	12/17 (71%)
# of times	—	7/33 (21%)
Ordinal numbers	—	21/21 (100%)

In both corpora, nearly all point-in-time references use EL numbers. In fact, I propose that there are no WT numbers which may appropriately represent a point in time. All four WT year numbers and one of the WT

[91] In stark contrast to both of these corpora, Nortier (1990:145) found only four Dutch numbers in her Moroccan Arabic/Dutch corpus.

[92] I have excluded from the table seventeen Malay numbers occurring within direct quotes which are entirely Malay, and four Malay numbers which are part of local place names. In the Shona/English corpus, 'labels' refer to room numbers, while in the WT/Malay corpus, 'labels' refer to school grades (nineteen instances, following Malay *kelas* 'grade, class'), names of numbers (seven instances, following Malay *nomor* 'number'), and labels of groups of people (two instances, following *tarop* 'group').

date numbers are instances of SD's hypercorrectness: not only is he the only speaker in the corpus who uses the obsolete WT word *narak* 'year', but he only uses it in these four instances, followed by a completely WT number naming the year, e.g., *narak ripun ia mo rat sêra mo ur dubám mo rua* 'the year one thousand and nine hundred and seventy and two', as well as the abbreviated *narak ur dubám mo rua* 'the year seventy and two'. Similarly, although SD once expresses a date by using WT *maera* 'day' followed by a WT number, throughout the remainder of that conversation he uses *tanggal* 'date' followed by a Malay number seven times, possibly because after trying it once, he decided it was just too odd to use WT to express a calendar date. Of the four remaining WT point-in-time numbers, three are the expression *arei ba* 'how much, how many' (after Malay *tanggal* 'date' and *jam* 'hour'), while the other, *jam lêma* '5 o'clock' is so close to its Malay equivalent *jam lima* that it may be best to consider it a performance error.

In contrast, numbers used for counting are almost entirely WT numbers. Of the seven Malay numbers used for counting in this corpus, six occur as dependents of (and adjacent to) Malay head nouns and are, therefore, to be considered instances of Malay mode in the analysis to be presented in chapter 5. The remaining non-default Malay number is seen in (48).

(48) 1\LD *Japún [a] so loloar ne,* ***ᴾSemol*** *so=* *itora* ***ᴾAgus*** *dagong kataler min.* ***ᴰSampi*** ***ᴾKampung•Cina****, dartom•dauk êra in, êra* ***ʸ#tujuh.***
Then [uh] yesterday, Semol and Agus sold vegetables again. They reached Kampung China, they encountered these ones, they were seven.

2\YN *Dubám.*
Seven.

3\LD ***ᴿTujuh*** *ne (...) tei.*
That seven (...) isn't it.

4\FD *Dubám.*
Seven.

Here we see that when LD uses Malay *tujuh* 'seven' in a counting usage, with no Malay head noun, it is so gratuitous that both YN and FD correct his language choice, insisting that he say WT *dubám* rather than Malay *tujuh*.

The occurrence or nonoccurrence of a Malay head is also involved in the relative frequency of time duration numbers. The twenty instances of Malay numbers in this category are all dependents of (and adjacent to) Malay head nouns and are, therefore, excludable as instances of COLLOCATION SEQUENCES

(discussed in chapter 5); in contrast, six of the thirty-one WT numbers in this category occur as dependents of Malay head nouns, indicating that the preference for the use of WT numbers in this category is at times stronger than the strength of Malay collocation. Similarly, all twelve of the Malay numbers used for measurement occur due to collocation with a Malay unit of measurement, but all five WT numbers used for measurements also occur adjacent to a Malay head noun.

Collocation does not explain the use of Malay in the money category, however, except that whenever a Malay head noun (*rupiah, doit*) does occur, the number is always Malay. But that only covers eight instances. In the sixty-four instances where no head noun occurs, the number is Malay in fifty-one instances and WT in thirteen instances. Although such a distribution of frequencies is similar to Malay lexical units in the displacement category, clearly no historical displacement has taken place here, since the Indonesian rupiah is a fairly recent introduction into WT culture. For this reason, money is conceptually both a cultural import and a countable object. I conclude, therefore, that when Malay numbers are used in reference to money, they represent a usage of numbers which is a cultural import; these Malay numbers, then, have been included in the count of cultural imports above rather than being considered non-default.

To sum up, I consider Malay cardinal numbers to be instances of default Malay if they are used to indicate point in time or are used as labels. Enumeration of money typically, though not necessarily, makes use of Malay numbers. Malay cardinal numbers used to indicate duration of time, measurements, counting, and the other categories in table 4.7 are instances of non-default Malay, often resulting from collocation. Malay ordinal numbers are displacements.

4.3 Discourse-induced Malay lexical units

So far, I have discussed a number of lexical units which should be considered borrowings either because they fill a lexical gap, or because their frequency of usage shows them to be clearly established in the speaker's repertoire. In this section I consider lexical units which might be considered gratuitous in another discourse context, but in the discourse contexts examined are not gratuitous at all.

4.3.1 Repetitions

There are many instances of repetition—either a speaker repeating the lexical choice of another speaker (often in backchanneling) or his

own lexical choice. In either case, once the choice is made, immediately subsequent uses of the lexical unit do not represent independent lexical choices; they are at least partially determined by the preceding discourse context. Indeed, sometimes in the corpus a speaker does choose to replace the Malay choice of another speaker (or even his own choice) with an equivalent WT lexical unit; but in doing so a speaker runs the risk of damaging social relationships. Therefore, the default choice will be to accept the choice already made, rather than replace the Malay form with an equivalent WT form. For this reason, it is inappropriate to consider such repetitions to be code-switches themselves; they cannot qualify as counter-examples to syntactic constraints, and should not be included in statistical counts of code-switching.[93]

In some instances, a repeated Malay word adjacent to one or more other Malay words gives the appearance of an extended stretch of Malay mode, as in (49) and (50). In (49), LL's turn is an example of a TRIGGERED SEQUENCE (discussed in chapter 5) containing a triggered Malay locative preposition *di*. (Underlining indicates an instance of Malay mode, to be discussed in chapter 5.) In TN's subsequent turn, however, it is likely that *di* is a repetition of LL's lexical choice rather than another instance of triggering. Note that TN ends her sentence with a WT tag rather than a Malay tag.

(49) 1\LL ***P**STM **Y**di **P**Langgur* ^ ***Y**kah*
Technical School at Langgur or
2\TN ^***P**STM **R**di **P**Tual* *te.*
Technical School at Tual of course.

In (50), OK slips into Malay mode briefly for the NP *orang yang tidak tahu* 'people who do not know', then uses the contrastive NP *orang yang tahu* 'people who know'. Since all three words of the second NP are repetitions of words in the first NP, they do not represent independent lexical choices and therefore do not necessarily constitute a second brief instance of Malay mode.

(50) \OK *kursir kunga•ne aka [:]* ***Y**orang **NF**yang **Y**tidak **Y**tahu. Mo **R**orang **NF**yang **R**tahu, ja ina danga [:] ok ona kokaleka ken kanám jipjupin ja (jijangal).»*
I talk like that for [:] people who do not know. But people who know, they are saying [:] I am hiding your rottenness."

Tail-head linkage involves the repetition of the final clause of one sentence as the initial clause of the next sentence. When the tail includes a non-default Malay item, the repetition of that item in the head does not

[93] Discourse-induced repetitions of Malay items are marked with a superscript R in examples.

qualify as a second non-default instance, since its usage is required, e.g., (51).

(51) \YL *DF**Tarús**, nei na ernai, %Y**Ø-antar** manám on sí. Ja nei %R**Ø-antar** manám on sí, ka [:] nei papangan nen [a] P**Mesak**.*
Then, she cooked, she brought this food. So she brought this food, then [:] she awoke this [uh] Mesak.

But fifty-seven turns later, when YL refers to the incident again, she uses the equivalent WT verb in (52).

(52) \YL *Ok ma korsirpeipei êr, nei ertar manám on sí,*
What I said previously, she brought this food,

Some instances of non-default Malay may be due to a lexical access problem. In (53) this seems to be the case, as TN refers to the concept 'choose (a leader)' three times, the first using Malay (*pilih*), the second again Malay but preceded by a pause (possibly indicating a brief mental struggle over lexical choice), and finally using WT (*-nal*).

(53) \TN *«YF**Kenapa** ok me P**Ambon** êr mo kem [a:] DF**langsung** %Y**mipilih** J**ketua** Y**baru**. Nungatin metangarangara ok ka= kmá ka [.] %Y**dapilipilih** apúkpúkna, mo kem DF**langsung** minal J**ketua** apúkpúkna.»*
"Why is it that when I was in Ambon, you (PL) [uh] immediately chose a new leader. You didn't wait for me to come to [.] choose a different one, rather you immediately chose a different leader."

When lexical repetition is used to maintain rapport among speakers, it does not always involve choosing between Malay and WT. In (54), AD uses Malay *mungkin* 'maybe', after which LL replies with the Malay synonym *mangkali*. Both of these are non-default, since there exists a common WT equivalent *kôtan*. In turn 4, AD echoes LL's lexical choice; since it is a repetition, this instance of *mangkali* is by default.

(54) 1\AD *Nei %N**()-lapor**, nei YF**mungkin** %Y**Ø-ingin** ken.*
He reported [you], maybe he wanted you.
2\LL *Y**Mangkali**!*
Maybe!
3\OK *[Satu•kali,]*
[One time,]
4\AD R**Mangkali** *[e]* S**perasaan•belok**.*
Maybe [uh] feelings of infidelity (literally, 'feelings turned').

So strong is the desire to maintain rapport, WTA speakers sometimes echo a WTB word even though they are not proficient in WTB. In (55), AD

repeats OK's WTB form *jiljula* 'drunkenness'; both the root and the form of reduplication are alien to WTA. But the use of *seina* 'alone' after this nominalized verb is a WTA pattern (indicating emphasis) which WTB speakers do not use.

(55) 1\AD *sakali dademdem* ***[N]karjá*** *lebá•ia.*
They don't do any work at all.
2\OK *Jiljula.*
Drunkenness.
3\AD *Jiljula se::ina.*
Nothing at all but drunkenness.

Similarly, in (56) LL repeats OK's WTB verb root *-dom* 'go', properly inflecting it; *-dom* is a verb which is known by WTA speakers because of its high frequency in WTB, but it is not used by WTA speakers.

(56) 1\OK ***[NF]Serta*** *madom ko makel [:] [.]* ***[C]kuningan****. Madom lia.*
And we went to dig up [:] [.] brass. We went to a cave.
2\LL *[mi-] Midom lebá?*
[2p-] You went to a what?

4.3.2 Metalinguistic usage of Malay

Sometimes Malay is used metalinguistically, as, for example, when a speaker gives a Malay translation of a WT term. As translations into Malay, they are obviously instances of default Malay.[94]

(57) 1\LL *Dua no lebá,* ***[K]Bongso****?*
What is dua, Youngest-sibling?
2\AD *Mo [-]* ***[M]pela****, te.*
Why [-] pela, of course.

(58) \AD *Gum ia, tei! Gum bôrabôrar dir! Nung damdam* ***[M]rotang-jawa*** *dir.*
A piece of rattan, you know? Those small rattans! What they usually call rotang-jawa.

A large number of items in this category, however, represent uses of Malay sayings and proverbs, as in (59).

(59) \AD *E:h! [.] On [bodoh-]* ***[M]bodoh•pimpin•bodoh****.*
Hey[.] This [fool-] fool_lead_fool
Hey! This is a fool leading a fool.

All CS researchers have noted that direct quotes are a favored site for code-switching, and in chapter 5, I discuss the occurrence of Malay within

[94] Malay items used metalinguistically are marked with a superscript M in examples.

direct quotes. However, while the language choice within direct quotes sometimes reflects the language used in the original speech being reported, direct quotes do not necessarily reflect the exact wording of the original speech. I consider exact quoting to be another metalinguistic usage in which the speaker has no freedom of lexical choice, let alone language choice. Such instances, as in (60), should therefore be excluded from a database of code-switches.[95]

(60) \LL *One ja ok sêrang•kôl* ***[K]ipar*** *nêr, «****[MH]Inilah•hari•****{laughing}*****•perhentian-ku****.»*
Thus I remember my brother-in-law (saying), "This is the day of my end."

Similarly, in (61) YL reports a bilingual joke which plays on the phonological similarity between Malay *pulauan* 'island(s)' and WT *pêluan* 'vampire/witch' (Malay *suanggi*).[96]

(61) \YL ***[M]Satu•pulauan*** *(nen, itorai to) ijirjir [set-]* ***[R]satu*** *pêluan êr*
(This) "one island", (she went along and) drew [...] that "one vampire/witch"

Reporting the joke requires using the original words of the joke; therefore, *pulauan* here is default, even though it is equivalent to the very common WT word *garia* 'island, island group'.

4.4 Idiolectal usage of Malay

Gardner-Chloros (1995:84) claimed that CS is a highly individualistic phenomenon and stated that "the description of both inter- and intra-individual variation in the same communities has hardly even begun." One aim of the present work is precisely to begin investigating such variation. As stated in chapter 3, I interviewed three of the primary WTA speakers as a group to arrive at decisions as to whether any given Malay lexical units had convenient WT equivalents. This had the advantage of the speakers helping each other with the mental exercise of translating, which is different from the activity of actually using words in live conversation. But it had the definite disadvantage of losing information on the individual speakers' differential competences. It is possible, however, to gain knowledge on the latter by examining speaker-specific word lists for contrasts in lexical units used and their frequencies.

[95] The superscript H indicates that the Malay item is High Malay.

[96] The form *pulauan* is not actually a word in either standard Indonesian or Ambonese Malay (Tjia, personal communication); *pulau* means island, and *kepulauan* means archipelago. It is unknown whether YL considers *pulauan* to be a valid Dobo Malay form.

First, there is an overall difference among speakers with regard to the amount of Malay used. Considering tokens rather than types, table 4.8 shows that OK and TN use a greater proportion of Malay than the other primary speakers, while SD, AD, and YL are more puristic, using more WT.[97] Similarly, with regard to non-default Malay alone, OK uses the most, while SD, YL, LL, and AD use the least.[98]

Table 4.8. Primary speakers' usage of WT and Malay

	WT tokens		Default Malay tokens		Non-default Malay tokens		Total tokens
		Percent		Percent		Percent	
AD	18,755	80.5	4,033	17.3	521	2.2	23,309
LL	10,266	76.0	2,955	21.9	284	2.1	13,505
OK	7,702	**71.9**	2,642	24.6	375	**3.5**	10,719
YL	7,995	80.0	1,797	18.0	207	2.1	9,999
TN	4,077	**72.3**	1,396	24.8	164	**2.9**	5,637
SD	2,524	83.0	459	15.1	57	1.9	3,040

Furthermore, everyone has certain favorite words and expressions which he uses more than other speakers tend to; in a bilingual speech community, some such favorites may come from the EL. Thus, a word may be a default Malay item for one speaker and a non-default item for another. Since a complete analysis of speaker-specific use of Malay items would be a massive undertaking, I discuss only a few illustrative lexical items here.

First, consider Malay *muka* 'face' and its WT equivalents, *lunga-/nunga-* and *lunga- ja mata-*. In table 4.9 statistics of usage are listed for all six primary speakers, as well as the aggregate of all WTA speakers. The table contains only those instances of *muka* which are equivalent to these WT terms. Although the aggregate figures for WTA as a whole seem to indicate that DM *muka* is in a stronger position than WT *lunga-* (etc.), i.e., that the WT lexical units are being displaced, this is certainly not the case for each individual speaker. Although the frequencies are too low to be certain, they seem to indicate that TN has all but replaced *lunga-* with *muka*,

[97] The differences between SD, AD, and YL, and likewise between TN and OK, are shown not to be significant at the $p = .05$ level by the chi-square test. But the difference between SD/AD/YL and LL is significant at $p < .001$, and the difference between LL and TN/OK is significant at $p < .01$.

[98] The differences between SD, YL, LL, and AD are not significant at $p = .05$, but these four are significantly different from TN at $p < .01$, and TN is significantly different from OK at $p < .05$.

while *lunga-* is still stronger than *muka* in the lexicons of AD, SD, and OK.[99]

Table 4.9. Terms meaning 'face'

	DM *muka*	WT *lunga-*, etc.
All WTA	**28**	20
AD	5	**11**
LL	3	2
YL	3	3
TN	**5**	1
SD	—	**3**
OK (WTB)	—	**3**

Despite his careful speech, however, SD shows a surprisingly high degree of Malay in the phrase meaning 'at that time', as seen in table 4.10. All the WTA speakers except TN use the mixed phrase *waktu ne* 'that time';[100] here, then, TN is using less Malay than the rest, rather than more. OK also uses the mixed phrase *oras ne* 'that time'.[101] The pure WT equivalents are *maera ne* or *maera êr* 'that day' (except WTB has *mera* for *maera*); the pure Malay equivalent *waktu itu* does not occur at all in the corpus. For this semantic item, AD, LL, YL, and SD all roughly follow the WTA aggregate figures, though for AD and SD the Malay is stronger than for LL and YL. The absence of Malay for this item in TN's speech could be due to her smaller overall contribution. But it is clear that OK prefers Malay for this item; unfortunately I have no other WTB speakers in this corpus to compare with her.

Table 4.10. Terms meaning 'at that time'

	Mixed *waktu ne*	Mixed *oras ne*	WT *maera êr/ne*
All WTA	54	—	**291**
AD	22	—	**51**
LL	8	—	**57**
YL	4	—	**22**
TN	—	—	**14**
SD	16	—	**26**
OK (WTB)	**15**	**17**	15

[99] Remember, though, that SD's contributions are careful speech, and thus have a greater proportion of WT lexical units.

[100] *waktu* is Malay, borrowed from Arabic.

[101] *oras* is Dobo Malay, borrowed from Portuguese and Dutch.

Third, consider Malay *luar* and WT *loloir* 'outside' in table 4.11. Although the sample is too small to draw many firm conclusions, it seems that LL is the only one of the primary speakers with a definite preference for Malay for this semantic item. Other speakers either have Malay and WT on an equal footing for this item, or perhaps show a slight preference for WT.

Table 4.11. Termsmeaning 'outside'

	DM *luar*	WT *loloir*
All WTA	**19**	12
AD	6	6
LL	**6**	1
YL	2	3
TN	—	—
SD	4	**3**
OK (WTB)	—	—

A fourth example is the Dobo Malay[102]conjunction *par* ~ *por* 'for'. YO is the only speaker who uses it extensively (ten occurrences), despite the fact that YO's total contribution to the corpus is much less than several other speakers. Aside from a single use by YN, all other speakers use *par* only in extended Malay stretches. Similarly, of the forty instances of the Malay adverb *artinya* 'what I mean is', AD is responsible for thirty-six.

I conclude that what is a default choice for one speaker is not necessarily so for another. In practical terms, this means that my coding of lexical units as default and non-default is only a rough approximation; such a distinction is impossible to pinpoint conclusively, given both the dynamic nature of a speaker's competence as well as inconsistencies of performance. Still, despite this caveat, I am confident that I now have a better set of data to analyze for language choice patterns than if the preceding lexical analysis had not been performed at all.

Finally, when there are synonyms in Malay, different speakers may have idiosyncratic preferences. As seen in table 4.12, WT[103] *bela* 'friend' is being displaced by not one but two Malay equivalents, *teman* ~ *tamán* ~ *tamáng* and *kawan* ~ *kawang*. Although the data are sparse, there does seem to be a distinction between those who prefer *teman* ~ *tamán(g)* and those who prefer *kawan(g)*—as in fact there is among monolingual Malay speakers.

[102] This conjunction was borrowed into Malay from Portuguese and Dutch.

[103] Etymologically, *bela* apparently came into WT from Malay; according to de Casparis (1997:12), Malay had previously borrowed it from Sanskrit. But it is not currently used in Dobo Malay, and WT speakers consider it to be WT, not Malay.

Table 4.12. Terms meaning 'friend'

	bela	*teman ~ tamán(g)*	*kawan(g)*
YL	2	2	—
TN	2	2	—
LL	2	**6**	—
OK	—	**7**	—
AD	3	**6**	1
HT	—	1	3
FG	—	—	1
WG	—	—	1

Such differential preferences also occur with completely gratuitous Malay items. All speakers use the WT negator *sakali ~ sikali ~ kali* far more than the Malay[104] equivalents *tida(k)*, *seng*, and *sondor*,[105] but the choice of which Malay negator to use is idiosyncratic, as seen in table 4.13. TN is the only speaker who seems to prefer *seng*, MG the only speaker who seems to prefer *tida(k)*.

Table 4.13. Negators

	seng	*sondor*	*tida(k)*
TN	**4**	1	—
AD	2	**12**	1
YL	—	**3**	1
SR	—	1	—
LL	—	**3**	—
OK	2	3	3
MG	—	—	**3**

Unfortunately, the data are too sparse to draw any conclusions about idiolectal patterns of language preferences. I assume that the preferences in each case derive from each speaker's unique linguistic history, but demonstrating this is far beyond the scope of the present work.

[104] The form *seng* is a loan derived from Portuguese *sem* 'without', while *sondor* is a loan derived from Dutch *zonder* 'without'.

[105] I am excluding here the occurrences of these negators in phrasal lemmas like *tida boleh* 'forbidden' (which OK uses eleven times, other speakers not at all).

4.5 Lone non-default Malay lexical units

One of the main goals of the present work is to account for every Malay lexical unit which occurs in the corpus, or at least as many as possible, and to do so from a perspective of psycholinguistic processes which produce and interpret LCP rather than from a large-scale societal perspective. Therefore, in §§4.1 and 4.2 I presented a number of categories of Malay lexical units which represent default use of Malay in the corpus, either because of various features of the lexical units themselves or because of the interaction between participants in the discourse. I am arguing here that these items represent not language choice but lexical choice, and therefore should not be used by LCP researchers either as examples of speakers indexing dual social identities or as counter-examples to any proposed syntactic constraints on code-switching.

There remains, however, a rather large number of lone[106] Malay lexical units which cannot be excluded on such grounds; they seem to have been used in contexts where equivalent WT lexical units could have been used instead. In addition, there is no obvious discourse-based reason for using them. They are not repetitions of a lexical choice previously made, nor are they metalinguistic uses, nor are they adjacent to another Malay word. In this section I discuss the status of these lone non-default Malay lexical units.[107]

Myers-Scotton (1993b:16, 207) considered all words which occurred at least three times in her 20-hour corpus to be borrowings. She admitted that the figure is arbitrary, and gave no detailed explanation of whether she included repetitions in the count, or whether she considered different senses of a word to be the same word or not. However, since an EL item which occurs dozens of times may still represent a rather low frequency in comparison to its ML equivalent, she also argued that for very high-frequency EL items, relative frequency might tell more than absolute frequency—in fact, she proposed that we should consider any EL item with a relative frequency of 5 percent or higher to be a borrowing.[108] But she also realized that statistical significance requires that the relative frequency test can only be applied to concepts which have a high absolute frequency; as a result, by her three-occurrence rule, all of these high-frequency items must be considered borrowings no matter how low their frequency relative to their ML equivalents.

[106] Here I am only considering non-default Malay items which do not occur adjacent to another Malay item. Non-default Malay items adjacent to other Malay items are considered in chapter 5.

[107] Lone non-default Malay items are marked with a superscript Y in examples. In addition, the combination YC is used to mark imported concepts for which a WT term has been coined, while YN indicates terms (mainly causative constructions) for which a WT wording is readily available, but would involve a syntactic restructuring of the sentence.

[108] A relative frequency of 5 percent would mean that for every five occurrences of an EL item, its equivalent ML item would occur ninety-five times.

How should the competing claims of relative frequency and absolute frequency be balanced? I propose applying Myers-Scotton's two rules in the opposite order, considering relative frequency more important than absolute frequency. That is, I consider any Malay lexical unit with a relative frequency of less than 5 percent to be a gratuitous usage of Malay, no matter whether its absolute frequency is one or fifty.[109] Any item with a relative frequency of 5–49 percent and which occurs at least three times will be labeled DISPREFERRED rather than gratuitous.[110] These two categories together will be referred to as DEFAULT MALAY.

Most of the pure Malay non-default types (366/556 = 66 percent) occur only once in the corpus; another 100 types (18 percent) occur only twice. For some of these, even though the interviewees were able to provide WT equivalents, those WT equivalents also occur only once or twice in the corpus. For such items where the referred concept has a very low frequency of occurrence, it is very difficult to say whether the use of Malay was by default or not, since statistical significance is impossible to obtain with such small numbers. Therefore, any Malay item with a relative frequency of 5–100 percent but which occurs only once or twice will simply be assumed to be dispreferred, as long as a WT equivalent exists; the low frequency of occurrence (or even nonoccurrence) of its WT equivalent in the corpus may well be due merely to chance.

There is in fact no binary distinction between gratuitous and dispreferred, or between default and non-default; rather, like so many other things in language (and other human behavior), there is a continuum. To be specific, I am defining a continuum from GRATUITOUS (relative frequency less than 5 percent) through DISPREFERRED EL items (relative frequency 5–49 percent) through PREFERRED EL items (relative frequency 50–99 percent) to NECESSARY EL items (for which no reasonably convenient WT equivalent exists).[111] I believe that this scheme is more psycholinguistically realistic than either Myers-Scotton's three-occurrence rule (which ignores relative frequency) or Poplack's category of nonce borrowing which applies to all EL items occurring only once in a corpus, whether they have ML equivalents or not, and no matter what the relative frequency is. I also believe that framing analyses of LCP in these terms is more useful than debating whether lone EL words should be considered borrowings or as single-word CS.

[109] In the latter case, of course, the Malay item would be overshadowed by over 1,000 equivalent WT items.

[110] Although I am allowing in principle that an EL item with a relative frequency of 49 percent be considered dispreferred, in fact the highest percentage in this category in my corpus is 43 percent.

[111] Clark (1982) briefly discussed various types of necessary borrowing and unnecessary borrowing he had observed in Vanuatu.

Why do non-default (and especially gratuitous) lexical units occur? Some may indeed represent an attempt by a speaker to index a Malay social identity or affect a Malayish style by sprinkling Malay words into his speech; others may merely be the result of psycholinguistic performance factors such as inattention, weariness, and so on, leading to lexical access problems, rather than lexical choice which represents the speaker's bilingual competence. I assume there are performance errors in bilingual data, just as there are in monolingual data. In particular, when a bilingual has a tip-of-tongue problem, i.e., difficulty in accessing an intended ML word, he may choose to use a near-equivalent EL word rather than wait until he remembers the ML word. Since this can occur even with high-frequency ML words, I do not assume that all instances of non-default EL words actually represent language choice.

As evidence for the label "gratuitous" I cite not only relative frequency but speaker interaction. For example, in (62) HT twice reacts to the use of Malay *sapa* 'who', pointing out that it is not WT. Other EL insertions, however, both default and non-default, do not provoke such a reaction. Note also that in his reaction against the use of Malay, HT himself uses several non-default, even gratuitous Malay words (*paki* 'use', *bahasa* 'vernacular', *ada* (progressive aspect), *ini* 'this, here'). This may be an example of accommodating language choice (discussed in chapter 5).

(62) 1\YN *Ma marokrok* ^{K}*Mama aka Orun (ma= imámá) êr,*
We were looking for Mama because of Orun (who came),
2\CH *Kai Orun.*
That Orun!
3\ED ^{Y}***Sapa****?*
Who?
4\HT *(bela) =ei* ^{Y}***paki*** ^{Y}***bahasa****.*
Hey (friend), use the vernacular.
5\YD *[Ragoi-] Ragoi* ^{P}***Usi•Tin*** *ma= isí erjaman ei bôt min.*
[So that's why-] So that's why Ms. Tin went and asked at the house too.
6\CH ^{Y}***Sapa****?*
Who?
7\HT ^{Y}***Paki*** ^{Y}***bahasa****, ka* ^{Y}***ada*** ^{C}***rekam*** ^{Y}***ini****.*
Use the vernacular, because this is being recorded.

In 4.6 I consider more thoroughly the phenomenon of speakers correcting both their own language choices as well as those of others.

4.6 Modification and negotiation of lexical choices

Example (48) illustrated that Malay *tujuh* 'seven' was so gratuitous that two addressees insisted on replacing it with WT *dubám*. Again in (62), a speaker who used Malay *sapa* 'who' was chided for not speaking WT. These are in contrast to the absence of comment for the vast majority of Malay items. In the present section, I present more examples which clearly indicate that some Malay lexemes are more noticeably Malay (more gratuitous) than others.

There are nearly ninety instances in the corpus of a speaker either changing his mind about which language to use for a particular lexical item, or attempting to repair another speaker's lexical choice by replacing it with an equivalent in the other language. Such LANGUAGE REPAIRS involve a choice between semantically equivalent lexical items of WT and Malay, not items of different meanings, such as in the following self-repair.

(63) \TN *Kali. [i- i- i- i-] Idem =na tan [a: nobar*
no [3s-3s-3s-3s-] 3s.do =3sa onto [uh nobar
*titdi.] **ᴺpapang** titdi.*
top-3pi] board top.3pn
No. [3s- 3s- 3s- 3s-] He did it on [uh top of the nobar-wood.] top of the boards.

As seen in table 4.14, most instances involve a speaker changing his own lexical choice, and most involve a move toward WT rather than toward Malay. By "toward WT" I mean one or two steps along the scale Malay > mixed > WT; the opposite for "toward Malay". ("Mixed" may be a mixed word or a mixed phrase.) Flip-flops involve a change from one language to the other and back again. Other terms in the table are explained below.

Table 4.14. Language repairs

	Toward WT	Toward Malay	Flip-flop	Total	(%)
Self (subtle)	27	4	1	32	(38)
Self (abrupt)	19	2	1	22	(26)
Other (subtle)	10	4	—	14	(16)
Other (abrupt)	6	1	—	7	(8)
Other (uncertain)	9	—	1	10	(12)
Total	71	11	3	85	(100)
(Percent)	(84)	(13)	(4)	(100)	

4.6.1 Self-repairs

Subtle. Subtle self-modifications are the largest category of language repairs; with these, the flow of speech is smooth, and there may be no indication that the speaker is even aware of making a new lexical choice. Sometimes, however, the speaker clearly realizes that he has changed his mind about a term, but still makes the change without abruptly interrupting his flow of speech. In (64) and (65), the speakers paraphrase themselves, switching from Malay to WT.

(64) \MG *kem ina* ***[Y]barangkat****? Kem minamnam =e?»*
are you (PL) going? Are you (PL) going?

(65) \OK ***[Y]Cakalang****.* ***[C#]Dua•stengah****. Ripun•rua•mo•rat•lêma.*
Tuna. Two and a half. Two thousand and five hundred.

In (66), AD refers to colorful mollusks as *yang babunga ir* 'those which are colorful' at the end of turn 1, using non-default Malay *babunga* 'colorful' (and necessary Malay *yang* 'which'). At the beginning of turn 3, however, she decides to replace that with a slightly longer WT phrase, *darpopo koukou dir* 'those which have colors'.

(66) 1\AD *{laughing} (Kupo tan) jiljíl ino, ko [katur e] [.] kelat [a]* ***[NF]yang [Y]babunga*** *ir.*
{laughing} (I took) those little mollusks, then [I lined up uh] [.] I chose [uh] the colored ones.
2\LL *Eya,*
Yes,
3\AD *[Da] Darpopo koukou dir, ko* ***[%Y]katur*** *kortêr bel ne.*
[3p] The colored ones, I lined them up along that beach.

In (67), LL begins her sentence with the Malay topic marker *kalau* 'if, as for', but begins the second clause of the sentence with the equivalent WT phrase *kenjou danaka* 'if talking about'.

(67) \LL ***[D]Kalau [P]Bu•Yopi*** *nono, nei* ***[/A]kali•bagitu****, kenjou danaka [:]* ***[P]Bu•Peki, [NF]boleh****.*
As for that Mr. Yopi, he isn't so much, as for [:] Mr. Peki, he can.

While the above examples involved pure Malay items being replaced by pure WT items, (68)–(71) show a mixed word or phrase repaired to become WT:[112]

(68) 1\AD *abil epir [p s] peda, ja* ***[NF]bisa*** *[%Y]da****tunju*** *dai* ***[D]luar****.*
heart good [...] first, then they can point things out to others.

[112] In (70), the final /a/ of Malay *gila* 'crazy' is deleted by a regular WT morphophonemic rule.

2\YL *Eya,*
Yes,
3\AD *Dasanou dai* [D]***luar****, oh,* [D]***batúl****.*
Point things out to others, oh, correct.

(69) \AD *Lêt* [%Y]*bara****barat****na ina itorai min. Aparaparana nono ina,*
A western man was with them too. A western one,

(70) \YL [/DF]***waktu****•ne, ja* [NF]***rekeng*** *[a] [ida:] [2]* [NF]***macang*** [%Y]*gili****gil****inai, te? Jerijerinai.*
At that time, it's like [uh] [3p :] [2] like they're crazy or something? They're crazy.

While (68)–(70) involved replacing non-default Malay with WT, in (71) TN decides to replace a default Malay verb root—a root which is identified as default in this corpus merely by relative frequency, despite the fact that its WT equivalent also has a fairly high frequency.

(71) 1\TN [NF]***Sidangkan*** *nei kakai gatan,* [%D]***ibayar****. Ipayar,* [NF]***to****?*
Even his own older sibling paid. Paid, you know?
2\LL *M-m. Eya, mom—Eya. Ipayar.*
Uh-huh. Yes, say - Yes. Paid.

This contrasts with (72), which is another instance of SD's hypercorrectness. Here, he tries to avoid the (necessary Malay) cultural import *pandeta* 'protestant pastor' by inventing a more cumbersome WT phrase (which took him four seconds to come up with). Even that phrase, however, contains the cultural import *gareja* 'church', demonstrating the difficulty inherent in attempting to maintain language purity.

(72) \SD *Ja* [/DF]***waktu****•ne,* [C]***panitia*** *jan, kam' mabana Perín, [4] matora [:]* [/J]***Pandeta****•Ngeíl. [4]* [C]***Gareja*** *kanei [e] gaibúnbúnna ina* [NF]***yang*** *ime Ngeíl.*
So at that time, the committees, we from Feruni, [4] together with [:] the pastor of Ngaiguli. [4] The church's [uh] leader (literally, 'big one') who lived in Ngaiguli.

In (73), AD replaces her own Malay *nakal terhormat* 'respectable misbehavior' with the mixed phrase *nakal yang epepir* 'good misbehavior'. WT does not lexicalize the fine semantic distinctions of Malay (or English) relating to specific kinds of goodness or badness; thus AD was forced to replace the more specific Malay *terhormat* 'respectable' with the generic WT *epir* 'good'. In order to replace Malay *nakal* 'misbehavior' with a near-equivalent WT term, she would have to use the same strategy, using the generic WT *samai* 'bad' (which covers many kinds of badness, including *nakal*), resulting in the absurd phrase *samasamai yang epepir* 'good badness'. Instead, she produces a mixed phrase.

(73) \AD *«Oh, eya,* ***[D]batúl****. Ok kanáng bôy nen, ime* ***[Y]bujang****, [.]* ***[%N]nakal****na* ***[DF]tapi*** *[e]* ***[DF]bukan*** ***[N]nakal*** *[a]* ***[YF]untuk*** *[e]* ***[%YN]ikasi•jatuh*** ***[Y]orang*** *te* ***[AH]merusakkan*** ***[R]orang****. Sakali.* ***[N]Nakal*** ***[Y]terhormat****. [Nak-]* ***[N]Nakal*** ***[NF]yang*** *epepir. [2]* ***[NF]Yang*** ***[NF]bisa*** ***[%Y]ibina*** ***[A]manusia*** *aka epepir.*
"Oh, yes, correct. My father, when he was still single, [.] misbehaved but [uh] not misbehavior [uh] for [uh] making people fall or destroying people. No. Respectable misbehavior. [...] Good misbehavior. [2] Which can help people for goodness."

The above examples represent the majority of language repairs, those away from Malay and toward WT. In (74), however, YL follows Malay *farei* 'free of charge' with the synonymous mixed phrase *sakali bayar* 'not pay'. The difference in etymology between these two expressions may not constitute a repair of language choice, however; it may merely be a paraphrase for emphasis, as is common in WT with or without a language switch.

(74) \YL *Daetar* ***[C]pesawat****. [pa-]* ***[Y]Farei****. Sakali* ***[D]bayar****.»*
"They rode an airplane. Free. Didn't pay."

While the example just considered involves a mixed string of words, (75) and (76) involve a change from pure WT to pure Malay.

(75) \OK ***[NF]memang*** *kama [m] martom=kam,* ***[Y]ketemu*** *marsir marká=kama,*
indeed we [...] met each other, met and spoke to each other,

(76) \AD *Eparai dauk.* ***[NF]Yang*** *ino, [.]* ***[YF]paling*** *eparai.*
Very good. Those ones, are very good.

Such examples stand out as peculiar contrasts to the more typical examples of a non-default Malay lexical choice being abandoned in favor of a WT equivalent.

Abrupt. Other instances of self-repair are more abrupt, with the speaker often cutting himself off in the middle of a word and then repeating the previous phrase with some modification. In these instances of abrupt self-repair, square brackets in the transcription mark the repaired sequence, and such bracketed stretches are not included in the lexical analysis presented in this chapter, since the speaker himself rejected his first choice of words. Some are clear examples of hypercorrection, replacing a default Malay item with a low-frequency WT equivalent. For example, in (77) AD initially chose the high-frequency displacement *Tuhan* 'God', then decided on the somewhat lower frequency WT equivalent *Jirjir Duai* instead.

(77) \AD *sakali //[C]damdam•**sombayang**, [.] dananaka [**Tuhan** ka inal a] Jirjir•Duai ka= inal [a] epepir akaka êra,*
they don't pray, [.] tell [God to give uh] God to give [uh] goodness to them,

Such an unnecessary self-repair, however, involving mere differences in relative frequency, does not stand out nearly as much as SD's self-repairs involving obsolete WT *narpet* 'desire' in (78), or his use of obsolete WT *narak* 'year' along with a WT number naming the year in (79).

(78) \SD *dam•on; «Kenjou ken [a] [kanám•**suka-**] kanám narpet aka lebá ia, ja morsirpei.»*
they said, "If you [uh] [want-] want something, then say so."

(79) \SD */[DF]**Waktu**•ne [**tahun** a] [.] narak ur•dubám•mo•[:]•rua.*
At that time it was [the year uh] [.] the year seventy and two.

In (80), SD hypercorrectly replaces *bulan* 'moon, month' with WT *pôlan* 'moon, month', even though he is about to name the month—a context in which Malay is the default choice. Note that his stammering may be an indication that he is not certain of the elegance of such a lexical choice.

(80) \SD *Nene ipel [**bulan** aa pôlan] [.] /[P]pôlan•**Agustus**.*
That was in [the month uh month] [.] the month of August.

In (81), AD replaces the cultural import *lampu* 'lamp' with (innovative?) WT *ngel* 'light'. Again, this may be an instance of hypercorrection, since WT speakers generally do not invent WT names for cultural imports.

(81) \AD *ok jou êra [dam [C]**lampu** êr ime a] dam one kanei ngel ime apúk [ng],*
I saw them [put the lamp at uh] put its light at another place,

Most abrupt self-repairs are not instances of hypercorrection, although they may have been motivated by the desire to produce a more pure WT language sample for the tape recorder. Regardless of the motivation for self-repair, the point to keep in mind here is that some Malay lexical items are noticed by speakers as being Malay and are replaced. Except for obvious cases of hypercorrectness, these all tend to be items classified as non-default (often even gratuitous) by the lexical analysis presented in this chapter. In (82), for example, TB replaces non-default *makan* 'food' with WT *manám* (though he seems to have a bit of trouble doing it).

(82) \TB *to= imá, [.] to= ipo [[Y]**makan** tarei te, [Y]**makan** e] manám tarei te,*
I wish he would come, so he would bring [some food maybe, food uh] some food maybe,

In (83), LL replaces non-default *sore* 'late afternoon' (which has five other lone occurrences in the corpus) with the WT equivalent *dededam* (which has thirty-six occurrences in the corpus). Since the relative frequency is 12 percent, sore is not gratuitous but it is non-default.

(83) \LL *kôlkôl•tuk [^{Y}**sore** er, de-] dededam er.*
I returned [that afternoon, aft-] that afternoon.

In (84), AD replaces two initial Malay attempts. The first is an error (*tuan tanah* 'master/owner of the land'), while the second is a gratuitous Malay lexical item (*kabong* 'garden').

(84) \AD ***^{NF}Kecuali [tuan•t- ^{Y}kabong** nono a] balár duai nono ^{NF}**ada**.*
Unless [the master/owner of the l- the garden uh] the garden's master/owner is there.

In (85), LL replaces *balakang* 'back, behind' with the WT equivalent *mir*, which occurs in the corpus but with a lower frequency than *balakang*.[113]

(85) \LL *Apún, $^{J+}$**supír-supír**, ^{J}**kanek**, ka dame [^{D}**balakang** no,] mir no,*
Then, the drivers, the conductor, were in [the back,] the back,

In (86) and (87), speakers replace mixed verbs with non-default Malay roots with pure WT verbs.

(86) \AD *Kenjou [.] erkôr =si, ja [:] [$^{\%Y}$**ringan**na, te:] marangái pei [:]*
If [.] it's dry already, then [:] [it's lightweight, te:] lighter than [:]

(87) \YL *Ok kom•on, «^{NF}**Lebái** [$^{\%Y}$tas**ingga** tai a] tatongan tai in [a] bôy nono ká,*
I said, "Preferably [we stop by at uh] we rest at these [uh] those men's place,"

4.6.2 Other-repairs

When one speaker uses a lexical item shortly after another speaker has used a synonymous term in another language, it is not always apparent whether he is trying to overtly correct the other speaker, or whether he is subtly suggesting the term, or whether he has no idea of considering his term as better than the other term. Where it seems to be clear, I have divided these cases into subtle and abrupt; the remainder I have labeled as uncertain.

Subtle. Subtle repairs are the most frequent type of other-repairs. Most are in the direction of WT, and most involve non-default Malay items, as in (88), where SR replaces YL's gratuitous *lihat* 'see' with the WT equivalent *i-jou* '3s-see'.

[113] The superscript + indicates that Malay reduplication, marking plurality, occurs in the word thus marked.

(88) 1\YL ***[Y]Sondor*** *[%R]**epariksa!*** ***[NF]Cuma*** ***[Y]lihat.***
He didn't inspect (them)! (He) only looked (at them).
2\SR ***[NF]Cuma*** *ijou ka [.]*
He only looked at it and then [.]

Similarly, in (89) AD replaces YL's mixed verb *da-tunggu* '3p-wait for' with the equivalent WT *da-tangara*, without calling attention to the fact that she is countering YL's non-default lexical choice.

(89) 1\YL ***[%Y]Datunggu*** *nei gatan!*
Just wait for her!
2\AD *Datangara nei, [.]* ***[NF]bisa.***
Wait for her, [.] we can do that.

In (90), AD again replaces YL's non-default Malay lexical choice; this time, YL overtly accepts AD's suggested replacement.

(90) 1\YL *Lêt nen [a]* ***[P]Edison*** ***[%Y]rakus****na.*
This boy [uh] Edison is greedy.
2\AD *{to Edison} Ken kadárka ^ dengal.*
{to Edison} You are very greedy.
3\YL ^ *Kadárna.*
He's greedy.

But language repairs by one speaker are not necessarily accepted by another, as seen in (91). Here, in turn 2, YL suggests non-default Malay *dalam* 'inside' which LL obligingly accepts. In turn 4, however, YL herself replaces *dalam* with the equivalent WT *abil*. But when the concept recurs in LL's turn 7, she again says *dalam*.

(91) 1\LL *Eya,* ***[DF]barang*** *etanatan [a:]*
Yes, because it's still [uh]
2\YL ***[Y]dalam*** *= e?*
inside?
3\LL ***[R]dalam.***
inside.
4\YL *Eranatan abil, ja one ^* ***[NF]rekeng*** *samai!*
If it's still inside, then that's ^ really bad!
5\LL ^ *Ja one, nei ma= ipo ei* ***[C]rumah-sakit,*** ***[NF]tetap*** *ma* ***[C]operasi.***
Ma dasiar.
^ so like that, she took it to the clinic, certainly they operated.
They cut it open.
6\YD *Ja sima•rir êr, kali ma darabarabak dal peipei = e?*
So that fish spike, they didn't pull it out?
7\LL *Sakali* ***[K]adi!*** *Etanatan* ***[Y]dalam.***
No, little brother! It's still inside.

Abrupt. There are only seven instances of someone confrontationally correcting another speaker's lexical language choice, and all but one are in the direction of WT. All six of these, in fact, are AD correcting SD (her older brother), and in each case SD agrees by reframing the sentence using AD's suggestion. Apparently, AD wanted SD to produce pure WT for the tape recorder. Examples (92) and (93) are typical of these abrupt other-corrections; in (93) the correction involves the Malay name of a village versus the WT name of the same village.

(92) 1\SD *Êra daser aka dalálá má* ***ᴾDobo*** *on.* ***ᴰᶠTapi ᵞᶠkarna*** *[:]* ***ᵞketemu*** *ei* ***ᵞperjalanan****, ja êrdá [:] [ren-] [.]*
They planned to run away to Dobo here. But because [:] they met them on the way, their [:] [plan-] [.]

2\AD *Darotan êra.*
They met them.

3\SD *Eya, darotan êra, ja [.] êrdá* ***ᴸniat*** *êr [a]* ***ᵞgagal****. …*
Yes, they met them, so [.] their scheme [uh] failed.

(93) 1\SD *ma darluan = ai bana* ***ᴾJerol****, ka ina* ***%ᴾLolor****.*
they crossed over from Jerol, and are at Lorlor.

2\AD *Garjá.*
Garjá.

3\SD *Ina Garjá.* ***ᴰᶠTarús****, [a]*
They are at Garjá. Then, [uh]

In (94), however, there is another hypercorrection. SD uses a Malay kin term, which is a category of default Malay lexemes. However, since there is also a high-frequency WT equivalent, AD corrects SD's term, and SD agrees.

(94) *1\SD Mo dededam êr, kama [:] maela, mo [.] êra (ma-) ma dal [a]* ***ᴾFrans*** *kanei* ***ᴷbapa•bongso****. [.]* */ᴾTaper•****Simun****•Parjer.*
But that afternoon, we [:] left, and [.] they (ma) took [uh] Frans's youngest father. [.] Mr. Simun Parjer.

2\AD Nei amai•mirmirna.
His youngest father.

3\SD [Eya, mo e nei-] Eya, nei amai•mirmirna.
[yes, but uh he-] Yes, his youngest father.

That this is a hypercorrection is demonstrated by the fact that these are the only two occurrences of *amai mirmirna* in the corpus, in contrast to ten instances of *bapa bon(g)so*. Furthermore, *mirmirna* 'lastborn' alone occurs only twice, while *bon(g)so* 'lastborn' alone occurs forty-four times. Yet *mirmirna* is still considered the correct WT term even by younger speakers, such as my research assistants who once replaced *bongso* with *mirmirna* in the transcription.

Uncertain. Since I am not Taranganese myself, I am not always certain whether the speakers themselves consider particular instances of other-repair to be subtle or abrupt. I have labeled abrupt only those instances where (a) the correcting speaker spoke only the word or phrase in question, and (b) the corrected speaker overtly acknowledges in the next turn that he has been corrected. There remain ten instances where the correcting speaker may have been intending to point out another speaker's lexical language choice error, but no error was acknowledged. Nine of these ten paraphrases are in the direction of WT, as in (95) and (96). In (97), YL repeats LL's Malay phrasal verb *pegang tangan* 'shake hands' (literally, 'hold hands'), adding a WT prefix; then she replaces it with the equivalent WT phrasal verb.

(95) 1\AD ***N*sama *Y*juga**.
just the same.
2\OK ***N*Sama** *gatan*.
just the same.

(96) 1\LL ***NF*Padahál,** ***/PK*Kris•***kanei*•**bapa** *ká ma dalalú aka* ***YC*pegang-tangan**.
However, Kris's dad and others went down to shake hands,
2\YL *%R*Da**pegang-tangan** *dai ba? Dasakata lêmin dai ba min?*
Shake hands where? Shake hands where?

A final example, possibly involving a performance error, is where AD paraphrases SD's appropriate WT lexical choice *nian* 'heavy' by herself using the Malay word *barát* 'heavy'; then SD repeats *barát* in accommodating fashion, only to have AD correct herself by using the WT word (*nian*) used by SD in the first place.

(97) 1\SD *Êra %N*da**pikir** *ok kotabei [a] dua tantan êra on [a] nian,*
They thought that my throwing [uh] alliance onto them was [uh] heavy,
2\AD *Êra* ***R*pikir** *one* ***Y*barát**.
They thought that was heavy.
3\SD *Eya.* ^ *Êra %R*da**pikir** ***R*barát**.
Yes. ^ They thought it was heavy.
4\AD ^ *(Ka dam) on nian dengal*
^ (And then they said) this was very heavy

4.6.3 Summary

To sum up, the overwhelming number of self-repairs from Malay to WT involve the replacement of non-default Malay items. This provides further

evidence for the distinction between default and non-default Malay, and especially between default and gratuitous Malay. Further study may indicate that the more gratuitous an item, the more likely it will be repaired. Other self-repairs involve Malay items which are not necessary Malay, but are classified as default due to a higher text frequency than their WT equivalents. These instances point out a speaker's indecision regarding where a given lexical item lies on the default-non-default continuum. Finally, we may safely assume that language repairs involving necessary Malay are instances of hypercorrectness.

The seventy-one instances of non-default Malay lexical items being corrected or paraphrased toward WT are far outnumbered by over 1,800 unrepaired non-default (either pure Malay or mixed) lexical items (tokens) in the corpus (see table 4.2). In the next chapter I consider the theoretical status within a larger theory of language contact phenomena of both lone non-default Malay lexical units and the longer stretches of Malay occurring in the corpus.

4.7 Phonological and morphological incorporation of Malay items into WT

Some LCP researchers attempting to find a clear boundary between CS and borrowing have divided lone EL words into two categories, those which act like ML words (phonetically, phonologically, morphologically, syntactically) and those which do not. The assumption held by these researchers has been that lone EL words which act like ML words are borrowings, while those which do not are instances of single-word CS.

Poplack and Sankoff (1984:103–104) made a first step toward testing this assumption by attempting to find a correlation between morphosyntactic integration and phonological assimilation. However, their study merely showed a correlation between the degree to which Spanish items were being displaced by English items and the degree to which English items were phonologically assimilated into Spanish patterns. Their methodology was quite different from my own, involving only word-list interviews under ideal recording conditions, rather than natural discourse in natural settings.

Unlike many such studies of language contact phenomena, the present work does not focus on whether Malay words are phonologically assimilated or morphologically integrated into WT. But in this section I do consider the possibility of using the WT/Malay corpus to test the assumption that phonological and morphological incorporation of Malay items constitute evidence of borrowing, and that lack of such incorporation constitutes evidence against a given item being borrowed.

4.7.1 Phonological assimilation

It is not clear-cut whether default Malay items tend to be phonologically assimilated and non-default Malay items tend not to be. The phonetic and phonological differences between SIM and DM complicate the study of phonological assimilation into WT, since WT phonological patterns are in some aspects similar to DM as opposed to SIM, and in other aspects similar to SIM as opposed to DM. A complete analysis is beyond the scope of the present work; in any event, the transcriptions in their current state are not accurate enough to perform such an analysis, as Malay items were often written in accordance with SIM orthography rather than phonetically. Often, in fact, the recording is not of sufficient quality to determine whether, for example, a speaker said (DM) *saki* or (SIM) *sakit*, (DM) *minya* or (SIM) *minyak*, (DM) *kasiang* or (SIM) *kasihan*, (DM) *ibur* or (SIM) *hibur*, (DM) *cape* or (SIM) *capai*, (DM) *balakang* or (SIM) *belakang*. I will, however, make a few comments here regarding phonological assimilation of Malay items which may be relevant to the present study.

There are three levels at which phonological assimilation can take place. The first of these is at the level of phonemes and underlying prosodic structures. A comparison of phonemes in WTA, WTB, DM, and SIM was given in chapter 2; there it was shown that DM has no vowels not shared by WT, though SIM has the additional vowel /ə/. As for consonants, /ñ/, /c/, and /h/ are completely lacking from WT,[114] while Malay /p/, /g/, and /j/ correspond only to allophones of WT /ɸ/, /w/,[115] and /y/, respectively. The low-frequency Malay consonant /f/ is considered below.

WT speakers using Malay words that contain these non-native sounds usually adhere to the Malay pronunciation rather than converting the Malay phonemes to WT phonemes. However, I have observed the following exceptions. First, some older speakers say *persaya* for *percaya* 'believe', *suma* for *cuma* 'only', *so(u)* for *co(u)* (an exclamative expressing astonishment), and *duna* for *duña* 'world'. This may be an indication that previous generations of WT speakers have converted /c/ to /s/ and /ñ/ to /n/ in loans; alternatively, these pronunciations may represent a previous stage of DM itself.[116] At any rate, these pronunciations have been (or are being) displaced by words containing the Malay sounds. In the WT/Malay

[114] In addition, the rare SM consonant /ʃ/ is lacking from both DM and WT.

[115] WTB has /g/ in place of /w/, and does not have [w] as an allophone of any other phoneme

[116] Wallace (1869:354) reported that the people of Wanumbai village referred to the "dancing parties" of the birds of paradise as *sácaleli*, an apparent adaptation of AM *cakalele* 'war dance'. It is not known whether this represented the DM pronunciation of the time, or signified the adaptation of the word into the language of Wanumbai village.

corpus, *percaya* occurs twelve times, and *persaya* not at all; *cuma* occurs ninety-six times, while *suma* occurs twice; *co(u)* occurs thirty times, while *so* with that meaning occurs only three times; *duna* occurs only once, while *duña* occurs eight times. Aside from these, 175 lone Malay lexical items (types) containing /c/, and forty-two containing /ñ/, occur in the corpus, plus those that occur in extended stretches of Malay.

The case of Malay /p/ and /f/ is slightly more complex. WT /ɸ/ has the allophone [p] obligatory syllable-finally, and optional elsewhere. The vast majority of the fifty-seven Malay types containing /f/ are pronounced with [f]; only four convert the /f/ to [ɸ] in prevocalic position, and a few more convert the /f/ to [p] in syllable-final position. Similarly, nearly all Malay types with /p/ are always pronounced with /p/ as [p], rather than being converted to [ɸ]. The sole exception is *pikir* 'think', a verb which nearly always occurs in the corpus with a WT agreement prefix attached and is nearly always pronounced [ɸikir].[117] For this one default Malay item, then, both phonological and morphological evidence indicates that the word is an established loan. However, I have found no general correlation between default Malay items and phonological assimilation. Rather, EL items are pronounced as they are in monolingual EL discourse, rather than being assimilated into ML patterns.

In addition to the occurrence or nonoccurrence of EL phonemes, consider their distributions. For example, in WTA (but not WTB), no polysyllabic morpheme can end in a mid vowel. DM and SIM have no such restriction, but Malay words which violate this WT constraint are not modified to fit WT sound patterns, whether the Malay item is default (e.g., *lenso* 'handkerchief', *pele* 'separate') or non-default (*goso* 'rub', *cape* 'tired').[118]

Nivens (1992a:129–37) described the stress patterns of WT. Except for about 10 percent of attested roots which have exceptional final stress, WT stress is root-penultimate unless the final syllable contains a mid vowel, in which case stress must be final. Malay items in the corpus matching this latter criterion, however, are not altered to fit the WT stress pattern, whether they are default Malay (e.g., *coret* 'cross out', *kantor* 'office') or non-default (e.g., *loteng* 'attic', *sondor* 'no').

The second level at which phonological assimilation can take place is at the level of morphophonemic rules of affixation and reduplication. WT

[117] For some speakers of standard Indonesian, *pikir* also has the alternate pronunciation [fikir]. However, I am not aware that this pronunciation is used in Aru.

[118] However, at least two Dobo Malay words, *pake* 'use, wear' (SIM *pakai*) and *sampe* 'reach, until' (SIM *sampai*) are usually pronounced *paki* and *sampi* in the corpus. Further research is needed to determine whether WT speakers use these variant pronunciations when speaking Malay; if they do, then these pronunciations do not constitute evidence of phonological assimilation to WT patterns.

morpho-phonemic rules of affixation do apply to Malay items in the corpus whenever WT morphology occurs, as for example the deletion of final /a/ when an /i/-initial suffix is attached:

(98) */DUP + gila + inay/* *[gili'gilinay]*
they are crazy

Similarly, the form of WT agreement prefixes is sensitive to the position of stress on the Malay root, just as it is with WT roots: *ku-* (1s), *mu-* (2s), *i-* (3s), and *mi-* (2p) become [ko], [mo], [e], and [me], respectively, either when the following syllable bears neither primary nor secondary stress or when the prefix *r-* follows the agreement prefix. Thus, such forms as *e-tarima* '3s-receive' occur but never forms like **i-tarima*; but since *percaya* 'believe' has secondary stress on the first syllable,[119] the forms *mi-pèrcacáya* '2p-DUP-believe' and *ku-pèrcacáya* '1s-DUP-believe' are found.

Moreover, when a Malay item is reduplicated to fulfill the requirements of WT morphosyntax, it follows WT reduplication patterns rather than Malay reduplication patterns.

(99) */DUP + mi + paksa/* *[mipak'paksa]*
you (PL) force

(100) */DUP + ta + nonton/* *[tano'nonton]*
we (inclusive) watch

Note that in (99) and (100), CVC reduplication occurs as prescribed by WT morphophonemic rules, even though the bases *paksa* and *nonton* are not permissible WT morpheme shapes.[120] In contrast, Malay reduplication usually involves replicating either the whole root or the whole word (Macdonald 1976:32–36). In (100), moreover, WT stress rules would require the root to bear final stress, since the vowel in the final syllable is a mid vowel; although the stress is not altered to fit WT stress patterns, the reduplication is a WT pattern.

The third level at which phonological assimilation can take place is at the level of phonological processes. Examples (101)–(102) list the phonological processes described by Nivens (1992a, 1992c).[121]

(101) Fortitions
a. Glide-to-obstruent fortition (p. 139):
Glides /y/, /w/ become obstruents [dʒ], [g] morpheme-initially and stressed-syllable-initially.

[119] In DM as used by WT speakers at least, *percaya* is pronounced [pèrcáya], in contrast to SIM [pərcáya]. It is unknown whether the assignment of secondary stress to the first syllable of *percaya* is a regular feature of DM or a symptom of assimilation into WT.

[120] WT consonant clusters can only begin with /r/, /l/, or /y/ (Nivens 1992a:147–48).

[121] Page numbers refer to Nivens (1992a).

b. Fricative-to-obstruent fortition (p. 144):
Fricatives /ɸ/, /s/ become obstruents [p], [t] syllable-finally.

c. /ɸ/-fortition (p. 145):
/ɸ/ may become [p] anywhere.

d. Nasal release (p. 209):
Phrase-final obstruents may be released with a homorganic nasal.

(102) Lenitions

a. /k/-lenition (p. 146):
/k/ may become glottal stop between two non-high vowels.

b. High-vowel laxing (p. 154):
High vowels /i/, /u/ may be slightly laxed in closed unstressed syllables.

c. /a/-lenition (pp. 154–55):
Unstressed /a/ usually becomes [ə].

d. /i/-devoicing (p. 156):
Unstressed /i/ may devoice following /s/.

e. /a/-deletion before stressed vowel (p. 205)
Unstressed /a/ is usually deleted before a word boundary if the following segment is a stressed vowel within the same phonological phrase.

Most of these processes are optional, and therefore the lack of their application for a given item cannot be used as evidence against phonological assimilation of that item. However, the first two fortitions listed in (101) are obligatory in WT, and neither of them apply to any of the Malay items in the corpus. The third fortition is irrelevant, since Malay does not have the phoneme /ɸ/. The last fortition listed, nasal release of final non-sonorant consonants, occurs not only with WT words as in (103) and (104), but also with Malay morphemes, both default as in (105) and non-default as in (106).

(103) \LL *ᵖ**Naomi** [datora nei-] datora nei dai bôt [n], mo ᵖ**Terosina** ertêr ela,*
Naomi [went with her-] went with her to the house, while Terosina walked away,

(104) \AD *jir in i= dakarakara ok [ŋ].»*
these cockatoos were biting me."

(105) \AD *Damagôr, eya, aka ᶜ**es** [n].*
Trembling, yes, because of the ice.

(106) \LL *maet [:] tan* ***[N]jalan•raya,*** ***[Y]isterahat*** *[n], mo [a:] nen [a]* ***[P]Petrus,*** *nei tortora nen* ***[P]Mesak*** *dartêr daidai bôt.*
We broke through [:] onto the highway, rested, but [uh] this [uh] Petrus, he along with this Mesak walked to the house.

It should be noted that AD uses this nasal release when speaking in monolingual Malay mode as well.

All of the lenitions are optional, although /a/-deletion before stressed vowel and /a/-lenition are typical in casual speech. In the WT/Malay corpus, /a/-deletion does occur when a Malay item ending in unstressed /a/ precedes the WT vocative particle *ei*, as in (107). But as far as I am able to determine, none of the other lenitions apply to any Malay items in the corpus.

(107)

[K]Bapa *= ei*	[ba'pey]	Hey Dad,
[K]Tanta *= ei,*	[tan'tey]	Hey Aunt,
[K]Mama *= ei*	[ma'mey]	Hey Mom,
[P]Ita *= ei,*	[i'tey]	Hey Ita,

Having considered phonological assimilation on three levels, I conclude that in this corpus phonological assimilation is not a reliable indicator of a difference between borrowing and code-switching. In fact, the bulk of phonological variation in this corpus relates to stylistic differences between high Malay and low Malay, a phenomenon which is beyond the scope of the present work.

4.7.2 Morphological integration

WT verbs generally carry a verb agreement affix, and verbs which modify nouns are nearly always reduplicated.[122] Many, but not all, Malay verbs in the corpus take verb agreement affixes—as with native WT verbs, those which are semantically active take an agreement prefix, while those which are semantically stative take an agreement suffix. Therefore, one could propose that lone Malay verbs which are morphologically marked in these ways are borrowings, while lone Malay verbs which display no such morphological integration are instances of single-word CS. One might also want to see if there is a correlation between morphological integration and relative frequency, in order to test the continuum proposed in §4.5.

Before jumping to any conclusions, however, it is necessary to have a thorough understanding of WT morphology, which in many details is very complex. For example, the second person singular agreement prefix is zero **if** the first syllable of the verb root is stressed **and** the verb root is

[122] See Nivens 1992a, 1992b, and 1993a.

consonant-initial **and** the prefix *r-* does not intervene, **unless** (in WTA, not in WTB) the verb is preceded by one of a small set of vowel-final words (auxiliaries and conjunctions) to which the agreement prefix may cliticize—**except** in TN's village, where there **is** an overt prefix even with consonant-initial, stress-initial verb roots. Similarly, in all villages, the third person singular agreement prefix is zero if the verb root is vowel-initial and the prefix *r-* does not intervene. Furthermore, the stative verb agreement suffix marking an inanimate third person singular subject is zero. Therefore, a Malay active verb with a second or third person subject, or a Malay stative verb with a third person singular inanimate subject, should not be counted as nonintegrated if a WT verb in the same context would have no prefix; any statistical assessment of morphological integration should exclude all items where integration is impossible.

Furthermore, there are certain contexts in which some verb agreement affixes are not obligatory. In general, the singular agreement prefixes may be dropped if morphophonological rules would allow them to surface with no change to their underlying form.[123] In fact, the first person singular agreement prefix *ku-* is typically dropped if it is immediately preceded by the first person singular pronoun *ok*, as long as no morphophonological rule would change the prefix to *k-* or *ko-*. Similarly, the first person plural exclusive and first person plural inclusive agreement prefixes are typically dropped if immediately following their respective pronouns. In addition, the third person singular prefix *i-* is often dropped as long as no morphophonological rule would change it to *e-*; if preceded by a word ending in /i/, the prefix, if there, is typically inaudible if the rate of speech is normal. Furthermore, a stative verb modifying a noun does not usually take the third person plural animate agreement suffix *-na* if it is immediately followed by a determiner.[124]

Moreover, there are no WT verbs which begin with an unstressed vowel. But there are Malay verbs beginning with an unstressed vowel, such as *usaha(kan)* 'strive' and *istirahat* 'rest'. These occur several times in the corpus, and never receive a WT prefix. I argue that since there are no WT verbs with a similar phonological shape, WT speakers would not know how to attach a prefix to such verbs. These verbs, then, should not be included in a study of morphological integration.

As a precursor to any statistical study of morphological integration of EL words, therefore, it is necessary to understand not only the

[123] That is, the verb root is consonant-initial, the first syllable of the verb root is stressed, the prefix *r-* does not intervene, and the verb is not preceded by a vowel-final auxiliary to which the prefix may cliticize.

[124] The omission of *-na* in this context is apparently phonologically motivated, since the singular animate determiners all begin with /n/.

(dialect-specific) morphophonological rules determining the surface forms of affixes, but also the phonological patterns of possible verb roots in the two languages occurring in a bilingual corpus. It is also necessary to take into account free variation with regard to the occurrence or nonoccurrence of affixes.

In addition to adding an agreement affix to a Malay verb, WT speakers also create compound verbs by using the semantically bleached verb *-m* 'do' followed by a Malay verb, as in (108).

(108) \OK *[DF]**Tapi*** *kali* *[/Y]da-m•**kanál*** *ken?*
But not 3p-do_know 2s
But they don't know you?

This pattern of integration has been observed by researchers in several other language pairs.[125] Gardner-Chloros (1995:78) pointed out that this strategy can be used even if neither language has such a 'do-construction'. WT, however, does have such a construction—it is used to allow one verb to modify another, to specify the manner in which an action is done, as in (109).

(109) *ko-r-têr* *ko-m* *jomi-jomir*
1s-R-walk 1s-do DUP-fast
I walk quickly

In the WT/Malay corpus, it seems clear that the choice of type of integration is phonologically determined: out of twenty-four verb root types (107 tokens) that occur as complements of WT *-m* 'do', twenty-one (102 tokens) have unstressed initial syllables, as in (108). In contrast, relatively few Malay verb roots which take WT agreement prefixes (fifteen types, sixty-five tokens) have unstressed initial syllables as in (110), in contrast to 348 types (1,183 tokens) with stressed initial syllables which take WT agreement prefixes as in (111).

(110) \WG *êra* *[%N]da**piara*** *êrdá* *[K]**adi*** *babakirna*
3p 3p-take_care_of their younger_sibling DUP-small-3sa
ina.
one (ANIM)
they took care of one of their small younger siblings.

(111) \YN *Éra pen* *[%Y]da**bongkar*** *pel* *pit.*
3p some 3p-tear_apart with night
Some of them tore it apart at night.

[125] Myers-Scotton (1993b:112–113) listed Panjabi/English (Romaine 1989:120–30), Japanese/English (Azuma 1991), Hausa/English (Bickmore 1985 and Madaki 1983), as well as her own Shona/English corpus; Muysken (1995:191–192) mentioned Hindi (and related languages), Tamil, and Navaho as languages which use this integration strategy.

This compounding may also be semantically determined, since nearly all the Malay verbs which are integrated by means of such a do-construction have an active meaning rather than a stative meaning.

Table 4.15 summarizes the distribution of Malay lexical units according to the degree and type of morphological integration into WT.[126] In this table, unlike table 4.1, the total number of Malay types is inflated to the extent that the same Malay lexical unit may occur as two or more different types if its senses occur in different lexical categories. As for the items which are integrated via reduplication only, nearly all are either words which cannot take agreement affixes or verbs in contexts where omission of agreement affixes is typical.

Table 4.15. Morphological integration of lone Malay lexical units

	Tokens	Types
All wholly or partly Malay lexical units	18,124	11,385
1. occurring as pure Malay lexical units only	14,290	10,489
2. occurring as mixed units only	1,786	754
a. mixed phrases	687	272
b. integrated with agreement affixes	1,025	456
c. integrated with *-m* 'do'	28	17
d. integrated via reduplication only	46	9
3. occurring as both pure Malay and mixed	2,048	142
a. mixed phrases	531	11
b. integrated with agreement affixes	1,117	110
c. integrated with *-m* 'do'	346	17
d. integrated via reduplication only	54	4

I am now able to deal with the question of whether morphological integration is a reliable indicator of the borrowedness of a Malay item in the present corpus. While some researchers use morphological integration as a positive identifier of borrowing, Gardner-Chloros claimed, "both loans and code-switches can be morphologically integrated or un-integrated with the surrounding language, depending on a wide variety of personal and linguistic factors" (1995:73). Here I will demonstrate that the occurrence or nonoccurrence of integration is at least partly lexically determined, and to that extent therefore unrelated to the distinction between borrowing and CS.

[126] In passing, it should be noted that although a wide range of Malay affixes occur on Malay roots in the corpus, never do we find a Malay affix on a WT root.

There are a number of Malay verbs (or rather, Malay words used as verbs by WT speakers) which occur frequently enough in the corpus that I can determine whether they tend to be morphologically integrated or not. Since WT morphophonemic rules refer to whether or not the initial syllable of the verb root is stressed, and since semantically active verbs generally take prefixes while semantically stative verbs generally take suffixes, the discussion will refer sequentially to stressed-initial actives, unstressed-initial actives, stressed-initial statives, and unstressed-initial statives.

There are at least three stressed-initial actives which never occur with morphological integration. These are shown in table 4.16. Among unstressed-initial actives, there are likewise nine high-frequency verbs which never receive WT affixation. Among stative verbs, which have a lower overall frequency than active verbs, there are only two high-frequency Malay verbs which are never morphologically integrated. Most of these verbs are default Malay.

Table 4.16. Malay verbs which are never morphologically integrated in this corpus

		active	initial stress	default	tokens
pulang	go home	+	+	+	43
jadi	work out	+	+	+	32
tugas	assigned to	+	+	+	19
barenti	stop	+	–	+	64
berangkat	leave	+	–	–	15
kanál	encounter	+	–	+	15
perlu	need	+	–	+	13
rencana	plan	+	–	+	12
rekam	record	+	–	+	9
ketemu	meet	+	–	–	8
kuliah	attend university	+	–	+	7
tujuan	intend	+	–	+	6
sama	same	–	+	+	5
terlambat	too late	–	–	+	6

In addition to these which consistently occur as bare forms, there are a number of verbs which sometimes occur with integration and sometimes occur as bare forms. These are presented in table 4.17. Some of these tend to be integrated, while others tend not to be, while still others show an even split. Active stressed-initials are integrated by prefixation, while most active unstressed-initials are integrated by forming a do-construction with

the bleached verb *-m* 'do'. There are also stative verbs which are inconsistently integrated; all but one of these are integrated by suffixation. Again, most of these are default Malay.

Table 4.17. Malay verbs which are sometimes morphologically integrated

		active	initial stress	default	bare	prefix	suffix	*-m* 'do'
tunggu	wait (for)	+	+	–	5	12	—	—
cari	look for	+	+	–	1	9	—	—
sampi	arrive, until	+	+	+	13	99	—	—
bayar	pay	+	+	+	3	18	—	—
lewat	pass by	+	+	+	3	17	—	—
sayang	love	+	+	+	1	15	—	—
dapa	get	+	+	+	2	97	—	—
pikir	think	+	+	+	3	41	—	—
kirim	send	+	+	+	1	24	—	—
iku	follow, join in	+	+	+	1	8	—	—
tarima	receive	+	–	+	1	22	—	—
pariksa	inspect	+	–	+	1	10	—	—
sambayang	pray, worship	+	–	+	9	—	—	24
cerita	tell a story	+	–	+	11	—	—	12
bataria	shout	+	–	–	4	—	—	4
karjá	work	+	–	+	23	—	—	21
kanál	know	+	–	–	12	—	—	8
bajaring	fish with nets	+	–	+	6	—	—	3
kuat	strong	–	+	+	2	—	20	—
stengah-mati	in great difficulty	–	+	+	7	—	12	—
susah	difficult	–	+	+	1	—	10	—
mampu	capable	–	+	+	6	—	5	—
pecat	fired	–	–	+	7	—	—	2
barani	bold	–	–	+	2	—	8	—

Finally, there are some Malay verbs which are always morphologically integrated in this corpus, as seen in table 4.18. Again, most of these are default Malay. For whatever reason, the only verbs which occur frequently enough in the corpus to make the claim that they are never integrated are active verbs. All but one of these have initial stress, and even that verb takes a prefix rather than the expected do-construction.

Table 4.18. Malay verbs which are always morphologically integrated

Root		active	initial stress	default	prefix	suffix	*-m* 'do'
antar	escort	+	+	–	19	—	—
singga	stop by	+	+	–	13	—	—
bongkar	take apart	+	+	–	12	—	—
nonton	watch	+	+	–	10	—	—
paki	use	+	+	–	9	—	—
rasa	feel	+	+	+	26	—	—
buru	hurry	+	+	+	12	—	—
dansa	dance	+	+	+	8	—	—
baca	read	+	+	+	8	—	—
suntik	inject	+	+	+	6	—	—
ganti	replace	+	+	+	12	—	—
lapor	report	+	+	+	12	—	—
jaga	guard	+	+	+	12	—	—
bantu	help	+	+	+	11	—	—
goreng	fry	+	+	+	10	—	—
piara	take care of	+	–	+	15	—	—

These data clearly show that even descriptive adequacy cannot be attained by merely comparing the number of verbs which are integrated with the number of verbs which are not. Furthermore, for some verbs I must again rely on a speaker-specific analysis. The Malay verb *kaluar* 'go out' is equivalent to the active WT verb *-lapei*. When OK uses *kaluar*, she integrates it by using a do-construction (five instances). But when TN uses it (three instances), she attaches a WT suffix, making it a stative verb. All other speakers that use *kaluar* (AD, LL, YL) use it as a bare form (five instances total).

In summary, the analysis presented here demonstrates once again that broad statistical analyses of EL insertions must be preceded by a specific analysis which takes into account both individual lexemes and individual speakers. Otherwise, the conclusions drawn will not be valid.

4.8 Conclusions

Some work in the first two waves of LCP research mentioned in chapter 1 has been based on a simplistic and invalid assumption about correspondence between two languages. Although it has been acknowledged for

many years that the lexical entries of different languages do not correspond exactly to each other, this fact has been ignored in practice by some LCP researchers. Sociolinguistic analysis has attempted to explain why an EL word was chosen over an ML word, without considering that no choice was involved because the EL word in itself (i.e., not just as a member of EL) conveyed a meaning that the equivalent ML word could not convey. Hypotheses regarding syntactic constraints on CS may have been prematurely struck down, based on putative counter-examples which do not really qualify as counter-examples because the EL word was chosen for its own sake, not as member of EL. In other words, hypotheses regarding constraints on language choice must be confirmed or refuted based on instances of language choice per se, not based on instances of lexical choice which only secondarily involve language choice.

LCP researchers have disagreed on whether to consider lone EL insertions to be instances of borrowing or of CS. I argue that "lone EL insertions" is merely a wastebasket category, a hodgepodge of items which are rendered as EL for different reasons; the category is not useful for analysis. For most speakers, and in most circumstances, the use of default Malay lexical units requires no explanation at all; rather, what needs to be explained is the puristic avoidance of them. If we seek to understand why bilingual speakers switch codes, we should focus on non-default Malay (especially gratuitous Malay) lexical units. For example, Myers-Scotton (1988) and others have claimed that speakers can index dual social identities by engaging in CS as an unmarked choice, i.e., by massive EL insertions. I propose that it is only in the category of dispreferred EL items that speakers have freedom to use more or less EL for the purpose of indexing dual social identities.

Many LCP researchers have commented that language choice in multilingual settings is functionally similar to stylistic lexical choice in monolingual societies. For example, Gal (1988:286) wrote, "in many multilingual communities the choice of one language over another has the same social function or significance as the selection among lexical variants in monolingual societies." In other words, LCP researchers have focused much on the emotive value of language choice, while ignoring the emotive value of specific words, which are in many cases much stronger. In this chapter I have attempted to demonstrate that in the WT/Malay corpus, the insertion of EL words often represents a greatly expanded lexical choice—that is, what appears to be language choice actually is lexical choice, not bilingual language choice which is analogous to monolingual lexical choice.

Many LCP researchers have noted that nouns are the most common category represented in lone EL insertions. The question is, which nouns?

Are they default or non-default? Nouns which represent cultural imports and other types of default EL material are not relevant to the psycholinguistic questions of bilingual research. The psycholinguistic and sociolinguistic processes involved with the use of gratuitous EL items are different from those involved with the use of necessary EL items.

As stated in chapter 1, the second wave of CS research focused on the search for constraints on how the two languages involved in CS could combine with each other morphosyntactically. A number of constraints were proposed, several of which were found upon cross-linguistic examination to be particular to certain language pairs. Only three have survived as possible universal constraints: the Free Morpheme Constraint, the Equivalence Constraint, and the Government Constraint. In each case, there remain questions as to the proper formulation of the constraint, including the question of whether any such constraint represents an absolute or merely a tendency. Crucial to the debate is a proper understanding of what data qualify as counter-examples; in chapter 4, I have attempted to demonstrate that many of the Malay words occurring in the WT/Malay corpus should not be considered counter-examples to any putative syntactic constraints on CS, since they represent the default way for a WT speaker to express a concept. Whenever putative counter-examples to syntactic constraints to CS are presented, the reader should be informed as to whether EL items in each example are necessary, preferred, dispreferred, or gratuitous.

Poplack and Sankoff (1984) reviewed the literature on borrowing and concluded that researchers have proposed four criteria for identifying loan words: frequency of use, ML synonym displacement, phonological and morphosyntactic integration, and acceptance by native speakers. In addition, there is obviously a need to fill lexical gaps in the (traditional) ML. In this chapter I have pointed out that the vast majority of lone Malay lexical units qualify as borrowings by at least one of these criteria; I have also demonstrated that these criteria are not necessarily correlated, as in the case of certain high frequency verbs which resist morphological integration. I do not believe the lack of such correlation in any way weakens the analysis, since there is no a priori reason why these criteria should correlate.

This chapter began with a discussion of various types and degrees of equivalence which may obtain between lexical items of different languages. Parameters of equivalence may be static (semantic equivalence, equivalence of convenience) or dynamic (equivalence of frequency, equivalence in context). These different kinds of equivalence led me to propose a continuum of Malay items from *necessary* to *preferred* to *dispreferred* to *gratuitous*, the first two being considered *default* and the

latter two *non-default*. Gratuitous and dispreferred items were distinguished based on absolute frequency and relative frequency; in contrast to Myers-Scotton's (1993b) reliance on these same criteria, I considered relative frequency more indicative of the difference than absolute frequency. My approach is also different from that of Poplack (1980, 1987), whose notion of nonce borrowing ignores relative frequency altogether.

In 4.2, I pointed out that the WT/Malay corpus differs substantially from corpora analyzed by other LCP researchers, in that far more EL tokens, and an even higher proportion of EL types, occur. I proposed nine different but partially-overlapping categories of default Malay, attempting to base the analysis on the speakers' perspective, rather than on the etymologies of the words concerned. I found it essential to distinguish different usages of "the same word", as exemplified by the use of Malay numbers.

In addition to Malay lexical units chosen for their own sakes, it was demonstrated (4.3) that some non-default Malay items are induced by the discourse context (including repetitions and metalinguistic usage of Malay). Attention to the discourse context also confirmed the notion of gratuitousness when the use of a certain gratuitous Malay item evoked a negative comment by the hearers. I further determined (4.4) that there is a certain degree of idiolectal variation with regard to lexical choices; what is default Malay for one speaker may be non-default Malay for another. I found significant individual differences in the overall proportion of WT versus Malay used, as well as specific words preferred by individual speakers. I also found that one speaker's contribution was marked by hypercorrectness in the matter of language choice. Further support for the distinction between default and non-default (especially gratuitous) items was found (4.6) in the modification and negotiation of lexical choices (both self-repair and other-repair). A number of repairs (both subtle and abrupt) were shown to involve the replacement of a gratuitous Malay item with an equivalent WT item, though sometimes with loss of semantic detail.

To sum up, I attempted in 4.2–4.6 to distinguish default EL insertions from non-default EL insertions, based on both static (lexical) and dynamic (discourse) factors, while also paying attention to idiolectal differences. This is in contrast to Poplack and Meechan (1995), for example, who lumped all lone EL items together in one category, and Myers-Scotton (1993b), who relied only on frequency to distinguish borrowed items from single-word CS. These sections together constituted an attempt to filter out those EL items which constitute default lexical choice rather than language choice. Identifying a speaker's motives for using Malay as I have here is, I argued, an essential prerequisite to the analysis of longer EL

stretches, including attempts at determining the validity of proposed syntactic constraints on CS. Concepts which are expressed in Malay by default ought not to be taken as strong evidence against a putative syntactic constraint. Unfortunately, the literature on syntactic constraints on CS has given us little indication that the difference between default and non-default EL items has been taken into account.

In 4.7 I considered the question of phonological and morphological incorporation of EL items, a topic treated by Poplack and Sankoff (1984) among others. I proposed that there are three levels of possible phonological assimilation, namely, with regard to phonemes and their distributions, morphophonemics, and phonological processes such as fortitions and lenitions. I decided that for the WT/Malay corpus at least, phonological assimilation is not a reliable indicator of a difference between borrowing and code-switching. Interaction between Malay varieties in the corpus confused the issue; in addition, the desire for natural recording situations reduced the quality of recordings, rendering a precise phonetic account impossible. As for morphological integration of Malay verbs, I determined that some are ineligible for affixation because of their phonological shape; for those which are eligible, the choice between affixation and incorporation into a do-construction is for the most part determined by the phonological shape of the root, but some lexemes simply resist morphological integration. Thus, I found no correlation between morphological integration and the difference between default and non-default Malay; I did find, however, that attention to idiolectal variation was essential here as well, as one verb exhibited speaker-specific variation.

Chapter 4, then, challenged some assumptions which seem to have been implicit in a number of published analyses of CS, the kinds of assumptions noted by Gardner-Chloros (1995). Such assumptions relate to the nature of language itself, in terms of psycholinguistic processing by both speakers and hearers.

In chapter 5, I focus on non-default Malay lexical units which occur in the corpus. I consider separately those which are adjacent to other Malay items and those which are not. I use a three-cycle approach in order not only to provide a complete description of the language contact phenomena found in the corpus, but also to begin to fathom the psycholinguistic motivations and processes which result in the phenomena observed.

5

Code-Switching: Causes, Forms, and Modes

The winds and waves are always on the side of the ablest navigators.

Edward Gibbon, *Decline and Fall of the Roman Empire*

In chapter 4, I argued that in contrast to many published analyses of language contact phenomena, lone EL insertions should not be treated as a homogeneous category. Rather, the analyst must in each instance attempt to determine the psychological motivation for using a particular EL lexical item, and treat default (especially necessary) and non-default (especially gratuitous) items separately as representing the outcomes of different psycholinguistic and neurolinguistic processes. The question now is, what of longer stretches of Malay? Should they all be labeled as code-switching in an attempt at a unified explanation, or divided into psycholinguistically or neurolinguistically homogeneous subcategories first?

In this chapter, I deal with the Malay intrusions which are not considered default use of Malay based on the lexical analysis presented in chapter 4. These range from the non-default single-word insertions mentioned in the previous chapter through phrases to sentences and clusters of sentences. To lay a contrastive background, I begin with an example from a closely-related language of a type of CS which has been cited in the literature but does not occur in the WT/Malay corpus. I then utilize a three-cycle approach to analyzing the WT/Malay corpus. In §5.2, I approach the data from a discourse-functional perspective, while in §5.3 I

attempt to account for the non-default Malay lexical units by dividing all instances into seven form categories. In §5.4, I refine that analysis by considering a number of more subtle psycholinguistic motivations for language mode switching.

5.1 What does not occur in the WT/Malay corpus

To lay a contrastive background to the corpus under study, I present a letter written by a speaker of Dobel, a closely-related language of the Aru Islands. The letter was written by a young Dobel man (living in Ambon) to his father, the head of a Dobel village (living in Dobo) in September 1996 and sent by e-mail via Jock Hughes and me.[127] The letter in (112) begins in SIM (or at least high Dobo Malay), using the standard style of letter writing taught in Indonesian schools. But the second part is in Dobel (including six Malay insertions, in bold face), and the third is in Dobo (or Ambon) Malay.

(112) [Part 1:] *Kepada Yth. Bapa A. P_____*
di Dobo/Puncak
Dengan hormat!
Bersama surat ini saya khabarkan buat Bapa bahwa di saat saya mengirim berita ini saya ada dalam keadaan sehat walaufiat di dala Perlindungan TUHAN kita YESUS KRISTUS, harapan saya semoga Bapa sekeluarga semuanya ada dalam keadaan yang sama pula.
Bapa seperti yang saya sudah katakan di dalam surat saya yang kedua bahwa saya akan tunggu sampai tes dimulai nanti—jadi itu saja yang saya perlu sampaikan.
To the honored: Father/Mr. A. P_____
In Dobo/Puncak
With respect!
With this letter I inform Father that at the moment I send this news I am in healthy condition under the protection of our LORD JESUS CHRIST, my hope is that Father and the whole family are in the same condition.
Father, as I already stated in my second letter, I will wait until the test is begun later—so that is all I need to tell you.

[127] I have deleted the family name of the sender and recipient; the remainder of the letter is as typed by the sender, including any typographical errors.

[Part 2:] ***Bafa*** *si* ***Mama*** *mifantan sa'u minou uballaba Rp. 50.000 ya'a sa'u.* ***Bafa*** *si* ***Mama*** *barang sa'u nda udem noban* ***jadi*** *mito****lon****gngu.*
Dad and **Mom**, help me send Rp. 50,000 to me. **Dad** and **Mom** because I do not do work **so help** me.

[Part 3:] *Jadi uang itu Bapa dong kasih par Pa Rick Nivens di SMA atas, nanti beta terima di Ake sam di Ambon Galala.*
Terima kasih.
Dari anak,
*Y. P*_____.
So that money Father and-others give to Mr. Rick Nivens at SMA Atas, and I will receive it from Ake's dad at Ambon Galala.
Thank you.
From your child,
Y. P_____.

Jock Hughes (personal communication) commented on the correlation between discourse function and language in this letter: in the first paragraph, written in SIM, the writer is softening up his father; in the second paragraph, written in Dobel, the writer makes his request; while in the third paragraph, written in Dobo Malay,[128] the writer spells out the details of how the transaction is to take place.[129] This example of inter-sentential CS is similar to the topic-related CS described by other researchers, where a switch in language choice occurs at (oral) paragraph boundaries; this type of CS does not occur in the WT/Malay corpus at all.

The above letter is of course monologue. Dialogue is more complex, since there are two ways for both participants and analysts to view the switching. First, an utterance may be seen as a switch in contrast to the language of that same speaker's previous contribution; alternatively, it may be seen as a switch in contrast to the preceding speaker's contribution. Thus, a two-dimensional analysis is needed because the contrast is two-dimensional. Auer (1995:124–126) described a typology of sequential patterns of language choice in dialogue as follows.

[128] Except for the Dobel word *sam* 'his/her father', which is part of Jock Hughes' Dobel name. Note that here the strength of the typical name works in the opposite direction from the Malay names seen in chapter 4. In both cases, EL names occur as an unmarked component of ML discourse.

[129] It is possible that the author wrote the last paragraph in Malay in order that I could read it; but it is interesting that the writer chose local Malay rather than standard Indonesian.

1. **Discourse-related CS:** First, both speakers use base language A, then one speaker switches to language B (either at the beginning of his turn or within his turn) and the other speaker follows by using language B.
   Pattern Ia: $1_A\ 2_A\ 1_A\ 2_A$ // $1_B\ 2_B\ 1_B\ 2_B$
   Pattern Ib: $1_A\ 2_A\ 1_A\ 2_A\ 1_{A//B}\ 2_B\ 1_B\ 2_B$
2. **Language negotiation:** Speaker 1 uses language A, speaker 2 uses language B; typically, one speaker eventually changes his language choice and they both end up using either A or B.
   Pattern IIa: $1_A\ 2_B\ 1_A\ 2_B\ 1_A\ 2_B\ 1_A\ 2_B$
   Pattern IIb: $1_A\ 2_B\ 1_A\ 2_B\ 1_A$ // $2_A\ 1_A\ 2_A$
3. [No label given by Auer]: It is impossible to decide which language is the base language; one of the speakers may decide to choose one of the languages as base language, and the other may decide to follow his choice.
   Pattern IIIa: $1_{AB}\ 2_{AB}\ 1_{AB}\ 2_{AB}$
   Pattern IIIb: 1_{AB} // $2_A\ 1_A\ 2_A$
4. **Transfer (not CS):** Within a single speaker's turn, some language B material is inserted into base language A.
   Pattern IV: $1_{A[B]A}$

The WT/Malay corpus has a great deal of pattern IV. Pattern I also occurs, except that instead of one pure language giving way to another pure language, pattern I has a lot of pattern IV mixed into it. Something like pattern II is seen in reported conversations (again with pattern IV mixed in), but it does not represent a negotiation of language choice; the phenomenon is described in §5.2.2. As for pattern III, there are a few possible examples of one speaker (mis)interpreting another speaker's string of default Malay words as a switch to Malay as base language; two of these are mentioned in §5.4.

5.2 First cycle of analysis: Conversational motivations

Auer (1995:117–123) categorized previous theories of the conversational meaning of code-alternation as belonging to one of two opposite extremes. The first extreme position was the notion that certain speech activities are linked to certain languages. In this view it is not the switching which carries meaning, but the objects switched. Auer argued that empirical investigations have not borne this out—the correlations are merely probabilistic.

The second extreme position was the notion that certain speech activiies trigger alternation, without regard to the direction of the alternaion—i.e., that it is the switching itself which carries meaning, not the objects switched. Under this view Auer included all the researchers who have proposed typologies of CS functions, e.g., reported speech, change of participants, parenthesis or side-comments, reiterations (quasi-translations, repetitions), change of activity type, topic shift, puns or wordplay or shift of key, and topicalization (topic-comment structure). He claimed that there are several problems with such lists: (1) analytical categories are often ill-defined; (2) the lists "often confuse conversational structures, linguistic forms and functions of code-alternation"; (3) such listing cannot "tell us anything about **why** code-alternation may have a conversational meaning or function," and since code-alternation is used creatively, the list will be unending; and (4) code-alternation is directional, it is not just the fact of a switch that carries meaning. Auer's view is: "In the typical bilingual speech community, the correlation between language and activity is not strong enough to make code-alternation predictable, but the direction of switching is nevertheless important for reconstructing its conversational meaning" (Auer 1995:123).

With regard to assigning a meaning to any specific instance of CS, Auer stated: "In order to pinpoint the conversational meaning of such a case of code-alternation, we need to know about the 'episode-external' preferences of speakers for one language or the other, or about the community norms for that particular kind of interaction" (1995:121). Based on my own informal participant observation from 1987–91 and 1995–97, I can confidently say that the kind of language mixing found throughout the WT/Malay corpus is an accurate reflection of typical speaker preferences and community norms. Taking that as a given, chapter 4 focused instead on speaker-specific norms for the lexical expression of particular concepts.

In the current section, rather than proposing yet another typology of CS functions, I touch on a few areas where it seems that a switch in language mode was motivated by a particular conversational context. I am leaving for §5.4 a more detailed analysis of psycholinguistic motivations behind language mode switching.

5.2.1 Addressee-related CS

One category of CS often cited by other researchers is that where the switch accompanies a change in addressee, especially to a person who does not understand the current matrix language.[130] For that addressee's sake,

[130] Or base language, depending on one's definitions of those terms.

then, a speaker changes to a different ML. Because of the conversational situations in the WT/Malay corpus, involving only speakers both proficient in WT and preferring WT as their ML, this type of motivation for switching is relatively rare, but does occur thirteen times, in side comments to passers-by and other non-participants. Examples (113) and (114) each demonstrate a single complete turn addressed to a non-participant. In (113), LL speaks to an adult who has entered the room, while in (114) YL scolds her child, who is supposed to be sleeping.

(113) \LL ***Usi•Ani dengar ʸsaja, eh? [.] Jang sambung.***
Ms. Ani, just listen, okay? Don't join the conversation.

(114) \YL *{to child supposed to be sleeping }* ***Biking kaing bagimana?***
What are you doing with the cloth?

Example (115) is slightly more complex, and reveals differing assumptions about default language choices. WT children are often addressed in Malay, so it is not unusual for LL to use Malay in turns 1 and 3 when she scolds YL's child, who had just interrupted her twice. However, in turns 2 and 4 YL addresses her own children using WT—which is somewhat surprising, since YL uses Malay more than LL in this corpus. Thus, there is a contrast here between the default choice for addressing children in general (LL's choice) and a particular parent's default choice for addressing her own children (YL's choice).

(115) 1\LL *{to kids}* ***Pi sana.***
Go over there.
2\YL *Mertêr miela ka mei ne!*
Walk away and go over there!
3\LL ***Pi sana.***
Go over there.
4\YL *Mei loloir ne!*
Go outside there!
5\LL *{to YL} Ok nanaka [:]*
As I was saying,

For the preceding examples of addressee-related CS, I believe that there is a turnover in the ML rather than a brief switch to Malay mode. However, this may not be the case in example (116). As MG and SB are conversing, an intoxicated young man (DR) enters. MG asks him a completely Malay sentence and receives a completely Malay answer, then asks a WT question and again receives a Malay answer. He asks another Malay question (turn 7), but DR is unresponsive, so he paraphrases the question using WT, and receives an answer in turn 10 which is a sentence consisting of two default Malay items. MG continues by speaking to DR using WT.

(116) 1\DR ***Om!!***
Uncle!
2\?? *Eya.*
Yes.
3\MG ***Su= ada kah•apa?***
Are you here already?
4\DR ***Sudah, Om.***
Already, Uncle.
5\MG *Ma= mnun =si =e?*
Did you drink already?
6\DR ***Sudah, Om.***
Already, Uncle.
7\MG ***Lalu Om punya, mana?***
So where is Uncle's?
8\DR *A:h*
A:h
9\MG ***Om*** *kanei?*
Uncle's?
10\DR ***Ada, Om.***
It's here, Uncle.
11\MG *(Ah, ja sudah) Ah! Má, to ^ sêt' %Nta-**bual** ka rat nen [nena-]*
Oh! Come, so that ^ we can converse so this respected man [is-]
12\DR ^ *(...) tei!*
^ (...) really!
13\MG *rat nen nena **Crekam.***
This respected man is recording.
14\SB *(Eya nei nena) **Crekam***
(Yes he's) recording
15\MG *Kama ena **Nbual Ysabarang** on, aka [.] nena **Crekam.** (So!)*
We are conversing casually here, because [.] he's recording. (Hey!)

In this example, it is unclear whether the change to Malay mode represents a turnover of the ML or not. To find out, I would have to know more about MG's perception of DR's preferred matrix language. If MG believes DR's preferred ML is Malay, then this would be another instance of turnover of the ML. Further study of drunken speech would be necessary to determine whether Malay occurs more or less often when a given speaker is intoxicated and whether the patterns of Malay usage are the same or different.[131]

[131] Of course, we might also consider a simpler, purely neurological explanation for why an intoxicated person violates grammatical constraints.

5.2.2 Reported speech

Many if not all analysts have noted the prevalence of CS in direct quotes. In the WT/Malay corpus, Malay in direct quotes is prevalent as well, often reflecting the language used in the original reported interaction.[132] In (117), a direct quote of someone speaking to children is reported in Malay, while in (118) a student speaks to a teacher and the teacher answers. In (119) and (120), the GPM pastor is reported as speaking to WT-speaking children in Malay. In all of these, the use of Malay is an accurate reflection of the typical language choice for interactions between the respective participants, and probably represents the language of the interaction being reported.

(117) \DA *Nei na= iten, ja [n] nei amai inam•on;*
*«[N]**Sudah,** [K]**Mama** ([Y]**ada** [Y]**pi**) [Y]**beli** [a] [C]**gula-gula** [Y]**par** [Y]**kamorang.**»*
Then he cried, so [.] his dad said, "Enough, Mom (is going and) buying [uh] candy for you all."

(118) \SR *%[N]Da-**lapor** kama. [.] «[JK]**Bapa-guru! [P]Yos [Y]deng [e] [P]Nanda [Y]paki [JK]Bapa-guru [Y]pu [C]sandál.**» [.] [X]**Tuhan•Allah**, nei at•nal gangar. [.] Êr nei ne= inal ka= %Yi-**putar** namnam•on, «[Y]**Sini [Y]sini [Y]sini!**»*
They reported us. "Teacher! Yos and [uh] Nanda wore Teacher's sandals." [.] God, he pulled out some rattan. [.] He took it and turned it like this, "Here-here-here!"

(119) \YL *nei tora [JK]**Mama•Pandeta** min ja [A]**cocok**. [3] (Kama a ne) kenjou lo•pit•gair tene mo nei nam•on; [.] «[Y#]**Dua [Y]orang [Y]pi [.][N]tolong [K]Mama [Y]di [N]dapur.**»*
He and the preacher's wife are right for each other. [3] (...) if tomorrow early morning he says, [.] "Two of you go [.] help Mama in the kitchen."

(120) \YL *Tajei dam kupikupin. (To nei) imá, ijou, inam•on, «**Ei! [Y]Yang•itu [Y]cere. [Y]Cere [Y]cere [Y]cere. [Y]La [Y]kamong [Y]makang [Y]itu [NF]sudah.**» Ja [NF]**rekeng** kam' makôr tan pua jan.*
Or else burn it. (...) he came, he saw it, he said, "Hey! That one, divide it up. Divide-divide-divide. Then you all eat it." So figure we beat our chests.

However, the language of direct quotes does not always accurately reflect the language used in the original conversation. There are clear

[132] There are approximately 2,000 direct quotes in the corpus, most of them not being completely Malay.

instances of a non-WT speaker quoted as if he had spoken WT; in (121), for example, LL (who does not speak Kei) reports a conversation between two natives of Kei as if they had spoken WT to each other.

(121) \LL *[DF]**tarús** [J]**Kapolsek** nam•on, «[DF]**Jadi** [N]**skarang** rargá nena ba?» «Oh, [N]**skarang**, ok ma= knaka rargá ka= ipeda [N]**kuburan**, mo ok na= ksika nei ka na masika [PK]**Tete•Kacamata** ká.»*
Then the Police Chief said, "So now where's the girl?" "Oh, now, I told the girl to go ahead to the cemetery, and I would go get her and we would go to Grampa Eyeglasses and those with him."

There are also examples of WT speakers reported as using WT, where in the original conversation they certainly used Malay. In (122), for example, YL quotes herself speaking to two non-WT speakers; her reported speech is mostly WT, but there are a number of non-default Malay words used, which possibly indicates that she remembered bits of her original Malay speech and just fills in the rest with WT. Alternatively, she could just be throwing in a smattering of Malay words to give an overall Malay flavor to her speech, in order to evoke the language of the reported conversation.

(122) \YL *Êra dalapeipei bôt êr, mo ok kom•on; {coughs} «[P]**Hermanus**, [P]**Naomi**, ok sakali na [ka] [%L]ku-**balas** tantan kem, [DF]**tapi** [nanti] [.] [Y]**laing•orang** [Y]**nanti** [%L]da-**balas** aka kem. [YF]**Karna** [.] kêrte ok, ragoi ken memekar nen [P]**Hermanus**. [YF]**Kalau** [k] [LH]**tanpa** ok, pel ok kanáng [D]**orang•tua-tua**, ken sakali [%N]**Ø-dapa** memekar [P]**Hermanus**.»*
They left the house, and I said, {coughs} "Hermanus, Naomi, I will not [.] take revenge on you, but [later] [.] somebody else will take revenge on you. Because [.] thanks to me, you married this Hermanus. If [.] not for me, with my parents, you could not have married Hermanus."

In (123), reporting a conversation between a native of the Kei Islands and some WT speakers, the Kei speaker is reported as speaking Malay, while the WT speakers are reported as using first WT and then Malay. The original was certainly Malay in its entirety.

(123) \ED ***Ei, [X]cuki-mai. [.] [Y]Kamorang [Y]dari [Y]mana?»*** *[DF]**Tarús** kama mam «Kam' mabana Tarangan.»*
"Hey, [expletive]. [.] Where are you (PL) from?" Then we said "We're from Tarangan." [.]

In (124), a WT speaker visiting Java is quoted as if he had spoken WT to a Madurese man; the Madurese man also uses WT in his response.

(124) \AD *(ja dajou) gasua ina,* %P***Ma**dura**dura**-na. (nei nam•on) «Ok* Y***minta•tanya** ken, te?* NF***Kira-kira** [:] on [a] galái* J***pandeta** kanei on [a] kanei [k-] tamata (nata) nen nalbana ba?« (nei nam•on) «*NF***Katanya** nono* Y***orang** P**Ambon**.»*
(so they saw) a young man, a Madurese. (He said) "May I ask you? The people at this preacher's house (literally, 'This preacher's house's people') come from where?" (He said) "They say he's Ambonese."

Some of this may be due to carelessness or memory loss on the part of the quoting speaker. Alternatively, it could be that the convenience of speaking in the language of greatest proficiency (or in the default language of casual conversation) is more important to (some) speakers than accurately conveying the original language choice. However, there are several reported conversations like (123) in which WT speakers are reported as if they had spoken WT, while non-WT speakers are reported as using Malay; this is further illustrated in (125)–(128).

(125) \AD ***«Eh,** P**K**Tanta•Au** Y**ini** N**skarang** Y**lain.** Y**Su**= Y**pi** YH**dengan** N**bule,** DF**jadi** Y**su**= Y**seng** Y**inga** Y**katong** Y**lai.»** (kom•on) «Eh! Ko ok* R***laing** nam•ba?» **«Eh,** Y**seng.** N**Dolo** Y**itu** Y**dekat** Y**deng** Y**katong. (Mo)** N**skarang** Y**ini,** Y**su**= Y**jauh.»** (kom•on) «Ei,* /DF*ja•**barang** kakamá bôt no ona jôu min.»*
"Hey, this Aunt Au is now different. She went with the white person, so she doesn't remember us anymore." (I said) "Hey! How am I different?" "Hey, no. Formerly, you were close to us. (But) now you are distant." (I said) "Hey, because our house is far away,"

(126) \LL *DF**Tarús** P**Dace** má ka= inam•on; **«**Y**Apa** P**Lin?»** [.] «Ei.* P***Dorenci** naka* NF***lebái** tora nei melá* NF***sudah**. (Ka jei, ekanei)* K***mama** ma eiraka =sin. Ja on, ken* NF***kan** moiraka (=si), ken me bôt (deder sangil), (su) moiraka.» [.] Nei nam•on, **«**DF**Jadi** Y**katong** Y**mau** Y**lari** Y**di** Y**mana?»** Kom•on «Êsá! One* D***dari** kem.»*
Then Dace came and said, "What, Lin?" [.] "Hey. Dorenci said why not you and her just elope. (otherwise, her) mom already knows. So here, you, you know, already know, you are in a (...) house (maybe), (already) know." [.] He said, "So where should we elope to?" I said, "Who knows! That's up to you (PL)."

(127) \AD *Kornapal =na.* [DF]***Tarús*** *nei ijou•po ok, mo inam•on;* ***«[P]Au! [.] [Y]Seng [Y]lia [K]Nene [Y]kah?»*** *«Ei,* [K]***Nene***, /[N]***ampung•****aka ken!» Mo jau,* [Y]***loncat*** *ei papa =si.*
I stepped over her. Then she looked up at me, and said, "Au! [.] Didn't you see Grandma?" "Hey, Grandma, affection to you!" But I had already jumped to the ground.

(128) 1\AD *Dame ne, mo [.]* [P]***Minggus*** *inamon* ***«[C]Selamat•malam, [P]Usi•Au!»*** *«Mom «*[R]***Selamat•malam****» ka ^* [N]***beri-beri•bangka****.»*
They were there, and [.] Minggus said "Good evening, Ms. Au!" "If you say good evening, then [may you get] beri-beri."

2\LL ^ *(One kenjou join ja join.)*
(you're too familiar with them)

3\AD *Ok kom•on; «Mom «*[R]***selamat•malam****« ka* [N]***beri-beri•bangka****.» (nei nam•on)* ***«Eh! [P]Usi•Au [Y]ini [N]tinggal [Y]marah-marah.»*** *«Eh!» [.] Jau. [.] (ok kom•on) «Ok kelakela =si. Kupedapeda.»*
I said, "If you say good evening, then [may you get] beri-beri." (He said) "Hey, this Ms. Au just keeps being angry." "Hey!" [.] Enough already. [.] (I said) "I'm leaving. I'm going first."

In these cases, it seems that language choice within direct quotes is used by some speakers as a way of maintaining the ethnic identification of participants in the discourse. In particular, it is a way of maintaining ethnic boundaries and asserting WT and non-WT identities.

5.2.3 Emphasis

There are times when it seems that the choice of Malay adds extra emphasis, as in the following examples.

(129) \TN *A-ah. [.] Êra daparong sena dasí,* [NF]***to****? Nei,* [DF]***tarús*** [NF]***ada*** *[:]* ***[Y]cowo [P]Tepa*** *ina min.* ***[NF]Baru [R]cowo•Tepa [Y]pu•banyak!*** *Nata nen?* [P]***Ulis****.* [P]***Ulis*** *ka= isí min.*
Yeah. [.] They all went, you know? Him, then there was [:] a guy from Tepa too. What's more, there were lots of guys from Tepa! Who was it? Ulis. Ulis went too.

(130) 1\AD *Mo bai* [NF]***yang*** *dakêlkêl =na? ^* [P]***Memed*** *[kanei-]*
But who (PL) bought it? ^ Memed ['s]

2\LL ^ *(najira mai-)* [P]***Larate*** *ká idá* ***[C]mobil [NF]yang [Y]putih*** *[a]* [P]***Lestari*** *nêr,*
^ (Najiron-) Larate and others' white car, the Lestari,

3\AD *Mo [en-] bai ^ dakêlkêl =na?*
But who (PL) ^ bought it?

4\YD ^ ***[P]Lestari.***
^ Lestari.

5\LL *[m-]* ***[P]Memed*** *ká.*
[m-] Memed and his group.

6\AD *(mo- kai-) Takom momom* ***[P]Memed,*** ***[DF]bukan [P]Memed ([NF] yang) [Y]balí.***
(but-) Don't say Memed, it wasn't Memed who bought it.

(131) \AD *Matalênga, mo [.]* ***[P]Johnson*** *karterna. Êr, kama na mapaláu, maparong malá masena. Lêtlêt, konikonir, maparong masena. [.] Nei kana ibana* ***[C]kelas•[N#]enam*** *má, mo [.]* ***[C]kelas•[N#]empat [Y]sondor [NF]ada [Y]jiwa*** *ia. Tamata pui. Ma dalá sena, ona bel. [.]*
We listened, and [.] an outboard sounded. There, we went toward the sea, all of us ran. Boys, girls, all of us. [.] When he came from sixth grade, [.] fourth grade didn't have a single person. Nobody. They all ran, they were on the beach. [.]

Unfortunately, there is no way for me to empirically demonstrate, based on independent evidence, that the function of Malay in these examples is to emphasize, except possibly in (131), where a repetitive WT paraphrase *tamata pui* also indicates emphasis. Perhaps a more careful study of phonological indicators of emphasis would result in a correlation, but here I merely mention the possibility and leave it as an open question for future research.

There are also some reported conversations in the corpus where it seems that the quoted speaker's anger or emphasis is conveyed partly by reporting the speech in Malay. Twice, for example, an angry schoolteacher's speech begins with the Malay command *sini*! 'Come here!' and then continues with a rebuke in WT. Myers-Scotton stated, "One of the most common uses of marked CS is to express authority, along with anger or annoyance" (1993a:132–133). In the WT/Malay corpus, the only instances of angry speech occur in reported speech, due to the casual nature of the recorded conversations.

5.2.4 Style

As noted in chapter 2, Malay (and in particular a high variety of Malay similar to SIM) is the default language of WT church services. Therefore, it is not surprising that when discussing religious matters, or rebuking someone as a preacher would, or praying, WT speakers tend to use more Malay, and often more high Malay. Example (132), long though it is, is only the beginning of an extended section of the conversation where AD's religious style includes a greater frequency of non-default Malay

lexical items sprinkled throughout. (There are also a number of default Malay items, of course, since the Protestant religion is a cultural import.)

(132) 1\AD *Ok kom•on, «[PD]**Tuhan•Allah** [NF]**memang** [e ok ken] [Y]**dari** [Y]**atas** ken jou [ok] ok kanáng [N]**kelemahan** [NF]**ada**. [DF]**Tapi** ken [NF]**yang** tôr ok moi on, ken [NF]**yang** [Y]**kasi** [N]**kekuatan** aka ok. Takom [a] mom ok [YNH]**permalukan** kekanám [Y]**nama** ei [YN]**tengah-tengah** [C]**jemaat•[P]Dobo** on.» [.] [NF]**Memang** [PD]**Tuhan•Allah** nen, [P]**Oya**, one [NF]**yang** [Y]**biking** ok [.] [DF]**biar** [e] [Y]**orang** [Y]**bilang** ok kanáng [C]**goncang•iman** [Y]**bagimana**, [DFH]**tetapi** [.] on [NF]**yang** [N]**jadi•[N]dasar** aka ok.*

I said, "God really [uh I you] from above you see [me] I have weaknesses. But you are the one who called me here, you are the one who gives me strength. Don't [uh] make me shame your name in the midst of this congregation of Dobo." Indeed this God, Oya, that's what makes me [.] even though people say my faith wavers some way, but [.] this is what constitutes a foundation for me.

2\LL */[D]**Batúl**•ken, galian. Kosukata ken lêmam. /[K]Galian•**gandong**.*

You're right, cousin. I shake your hand. My womb cousin.

3\AD *[Bat-] [Y]**Seng, [P]Oya**. Ok goi. On, [Z]**kenyataan** on, on [NF]**yang** [.] [NF]**Artinya**, [NF]**memang** {lowers voice} sêta [A]**manusia** (on) [Y]**samua** [Y]**punya•kelemahan, [NF]to**?*

No, Oya. I'm really serious (literally, 'I die'). This, this reality, this is what [.]-I mean, indeed {lowers voice} we humans all have weaknesses, you know?

4\LL *Eya, te, galian.*

Yes, of course, cousin.

5\AD *[DF]**Tapi** [LH]**tanpa** [:] [ta-] tatôr taká nei, [NF]**memang** [.] [NH]**sulit** min. Sêta [D]**mau** [set-] [%N]ta-**harap** sêtdá [N]**kekuatan** on, [NF]**pasti** [A]**jatuh**. [DF]**Biar** nei ma= itôr sêta, mo kenjou [.] sêta [Y]**bangga** ko tam «Eya. Ok [NF]**bisa**.» [DF]**Tapi** [NF]**bisa** [Y]**dengan** [A]**jalan** nam•ba? Tei? [NF]**Artinya**, nei nal ne min aka [A]**satu** [C]**ujian** aka sêta. «Ken na /[NF]mom•ba•**sampi** [NF]**bisa** [.] [Y]**dekat** pel ok? [.] Te ken na môl [N]**jalan** [:] nam•ba aka [NF]**bisa** [ok rek- kuba-] [%N]ku-**bantu** [d] ken aka [YN]**kasi•[N]kuat** ken?» [P]**Oya** =ei, [N]**pulang**, [NF]**to**? Ok seinga kutalar, seinga kuten. [.] [NF]**Artinya** ok kom•on «Eh, [NF]**memang** [PD]**Tuhan** nen [a] kanei [Y]**kasih** on, tei.*

But unless [:] [1pi-] we call to him, it's really [.] very difficult. If we want [.] to trust our own strength, we'll certainly fall. Even though he has called us, but if [.] we are proud and

say “Yes. I can.” But we can in what way? Huh? I mean, he gives that too as a kind of test for us. “You will do what so that you can be [.] close to me? [.] Or what path will you take so that I can help you to strengthen you?" Hey, Oya, I went home, you know? I sat alone, I cried alone. [.] I mean, I said, “Hey, indeed this God’s love is something, isn’t it.”

Similarly, in (133) OK sermonizes using both default and non-default Malay items (some of them high Malay), as well as some high Malay items which have no WT equivalents but do have low Malay equivalents. AD accommodatingly follows up with a brief stretch of Malay, which she immediately translates into WT.

(133) 1\OK *Sita* [//N]*tam•**kerja** no,* [NH]***walaupun*** [YH]***sedikit****,* [NF]***yang•penting*** *sit’ sidá* [Y]***keringat*** [Y]***sendiri****.*
We work, even if a little, what’s important is that it’s our own sweat.

2\AD *Eya,* [NF]***to****?*
Yes, you know?

3\OK [DF]***Tapi*** [Y]***kalau*** [D]***mau*** [Y]***pakai*** [N+]***cara-cara*** [Y]***seperti*** [Y]***itu****, sakali* [NF]***bisa****.* [PD]***Tuhan*** *sakali* [/D]*kanei•**mau**.* [NH]***Walaupun*** *nonga ná* [Y]***orang*** [C]***kaya****,* [DFH]***tetapi*** *[.] nêr* [NF]***kan*** *(jau n uli,)* [NF]***to****?* [D]***tidak•boleh****. [.] (danonga.) [.] (karna, êr ruai ir in,) ^ (ja one =si,)*
But if you want to use ways like that, you can’t. God doesn’t want it. Even though you steal from a rich person, but [.] he, you know (has spoken,) you know? It’s not allowed. [.] (they stole.) [.] (because, the two of them,) ^ (...)

4\AD *^ (pecat.) ja nam [a] one* [Y]***sudah*** [A]***melanggar•hukum*** *=si. [3] Nekanei* [W]*atorang dir,* [%N]*da-**langgar** =di =si. [.]*
^ (fired.) so like [uh] that, [they] already broke the law. [3] His commandments, they already violated them.

Malay (especially high Malay) seems to be the default choice for voicing authority in those registers or domains (school, church) in which authority is typically exercised by Malay speakers. By sprinkling in a few words of non-default Malay, a WT speaker is playing the role of teacher or preacher, or acting as the voice of God. I predict that authoritative messages in the domain of *adat* would not be expressed in Malay, though I have no evidence in the corpus to confirm this.

In (134), AD speaks as if quoting God and uses a lot of Malay; she first uses a single non-default word, but soon shifts to complete sentences in Malay.[133]

[133] It is unclear where AD intended the quote to end.

(134) 1\AD *Eya, [NF]**to**? [Tuhan na] Sêta na soyarda min, [PD]**Tuhan** na= ijaman ken. «Ken [YF]**selama** me [D]**dunia** on dem lebá? Dem lebá gatan? Ok kôl [C]**tugas** aka ken, [2] ken [Y]**sudah [NH]dibaptis. [YN]Habis•[Y]itu, [Y]sudah [C]sidi.***
Yes, you know? [God will] When we die, God will ask you. "While you lived in this world, what did you do? Just what did you do? I gave you things to do, [2] you have been baptized. After that, you've been confirmed."

2\LL *[C]**Mengaku.***
Confess.

3\AD *[C]**Mengaku [Y]mau [C]iring [P]Kristus.** One ^ inam•ba?*
Confess you want to glorify Christ. How is that?

4\LL *^ (abina:)*
(...)

5\AD *Na //[C]dam•**sembayang [Y]kah** [:]*
will pray or [:]

In (135), OK also uses a lot of Malay words when quoting God; some are default Malay, others are not.

(135) \OK *[Jadi nam- Ataukah-] [Y]**Jangkan [Y]orang•lain.** Mo kem [ona m-on] /[N]**oras**•on, ona mijula minanga•on, [DF]**tapi** kimá [D]**keluarga** /[N]**oras**•on ina [N]**menderita**. (Nono) nata? [.] Ja [C]**berkat** no dapo damá, ja [:] [.] [PD]**Tuhan** inanga «[NH]**Harus [Y]membagi-bagi.** [YF]**Supaya [C]upah** no [YH]**terletak** ná [C]**takta** ne.» [DF]**Tapi** nang•mo mise=kim miká, [NF]**berarti** [.] [Y]**satu•titik** ka [NF]**pasti** [YH]**lenyap**. [PD]**Tuhan** [NF]**pasti** [.] [YN]**kas-kembali** [C]**berkat** no, [YF]**sebab** [C]**berkat** [NF]**yang** nêr inal no, nêr kaní. Ja ken nangmo %[A]m-**atur•[W]salah**, ken [YH]**pun** nêr inal eltúk.«*
[so like- or-] Forget other people. But you (PL) [.] now, you're drunk like this, but your family now are suffering. (that's) who? [.] So that blessing is brought here, so [:] [.] God says, "You must divide it up. So that that reward is located on that throne. But if you alone eat it, it means [.] one day it will certainly vanish. God will certainly [.] take back that blessing, because the blessing which he gives belongs to him. So if you use it wrongly, he will take you back too."

5.2.5 Accommodation and negotiation

In this corpus there are many instances of accommodation, where a speaker alters some feature(s) of his speech with a view to speaking more like the addressee. Dialect accommodation is a frequent characteristic of

OK's speech in this corpus; her WTB is less prestigious than the WTA of AD and LL, with whom she has daily contact. However, there are a few instances of WTA speakers accommodating to OK as well, as in (136), where OK has been the main speaker, and AD asks for clarification using the WTB word *toar* 'house'. In (137) AD uses the WTB phrasal lexical entry *nang mo* 'if', while in (138) she uses the WTB verb *da-uli* '3p-say'. In all three instances, a very common WTB form is used in place of a very common WTA form.

(136) \AD *Kem mime toar bai dá?*
2p 2p.reside house who.(PL) POSS.3p
Whose house did you live in?

(137) \AD *Nangmo ne, mo [dajô-] darjôurau aka lebá*
If that CNTR [...] 3p-speak for what
min?
more
If that's the case, then why do they need to discuss it?

(138) \AD *[Da-] Dauli dam, «*^{P}***SMA***», (one)* ^{N}***jadi.***
[3p-] 3p.say 3p.like high.school (that) works
Say "high school", that'll work.

Aside from dialect accommodation, there are several instances where one speaker switches into Malay mode and the next speaker follows his lead (Auer's discourse-related CS), as in (139).

(139) 1\LL ^{NF}***Katanya****, nei* ^{R}***masu•sel*** *maera rua,* ^{DF}***tarús*** *nei elapei,*
mo ^{Y}***nanti,*** ^{Y}***tunggu*** ^{C}***sidang.***
They say, she went to jail for two days, then she got out, but she'll wait for the trial.
2\YL ^{Y}***Iya,*** ^{Y}***ada•mau*** ^{C}***proses*** ^{Y}***barang*** ^{Y}***itu.***
Yes, they're going to process that matter.

In (140), HT follows the lead of first ED and then CH, each time telling them in Malay not to speak Malay.

(140) 1\YN *Ma marokrok* ^{K}***Mama*** *aka Orun (ma= imámá) êr*
We were looking for Mama because of Orun (who came),
2\CH *Kai Orun.*
That Orun!
3\ED ^{Y}***Sapa?***
Who?
4\HT *(bela) =ei* ^{Y}***paki*** ^{Y}***bahasa.***
Hey (friend), use the vernacular.
5\YD *[Ragoi-] Ragoi* ^{P}***Usi•Tin*** *ma= isí erjaman ei bôt min.*
[So that's why-] So that's why Ms. Tin went and asked at the house too.

6\CH ***YSapa?***
Who?
7\HT ***YPaki Ybahasa****, ka* ***Yada Crekam Yini.***
Use the vernacular, because this is being recorded.

In (141), AD produces a sentence at the end of turn 1 (and continuing into turn 3) which is entirely Malay except for the WT pronoun *êra* (3pa); the final word of the sentence is a triggered[134] non-default Malay word, with Malay reduplication to indicate plurality. LL (in turn 6) follows AD's apparent language choice by using the Malay preposition *di* rather than the WT preposition *ei*. AD, however, apparently did not want to be in Malay mode, so in turn 7 she completes her own sentence by following the pattern LL suggested, but using *ei* instead. Curiously, however, she continues by producing a completely non-default Malay prepositional phrase. LL, who may be confused by now, follows up with a triggered sequence.

(141) 1\AD ***DFTapi*** *ok,* ***DFuntuk•soal Rpakiang****, [ok sakali e] one sakali* ***Njadi•Nmasalah****.* ***NFCuma NFyang Njadi•Nmasalah****, on.* ***NCara Yberbicara****,* ***YFkarna****, êra* ***NFkan Ysudah NHterbiasa*** ^ ***(untuk) NHmenghadapi*** *[.]*
But I, regarding clothes, [I don't uh] that isn't a problem. The only problem is, this. The way to speak, because, they, you know, are already accustomed ^ to face [.]
2\LL ^ *Eya.*
^ Yes,
3\AD ***YC+murid-murid****.*
students.
4\LL ^ ***Rmurid-murid****.*
^ students.
5\AD ^ *[ea]*
^ [uh]
6\LL ***Ydi Csekolah****.*
at school.
7\AD *Ei [.]* ***Csekolah****.* ***YDi Ymana Ysaja***
at school. Wherever
8\LL ***(DFJadi) Rmurid Ybanyak Asekian•banyak Ybagaimana,*** ***DFtapi*** *êra* ***Ysu=*** ***Nbiasa*** *=si.*
(So) many students, however many students, but they are already accustomed to it.'

Another form of accommodation is represented by that category of default Malay lexical items labeled repetition in chapter 4. With these, it is

[134] Triggered sequences will be discussed in 5.3.2.

not the language choice per se which is chosen by one speaker and adopted by another, but a particular lexical item or string of items. When a lexical item is repeated by another speaker, it does not constitute a language choice by that second speaker; thus, repeated Malay only qualifies as a Malay choice for the first speaker. For the second speaker, what matters is the smoothness of social relations and discourse flow which are maintained by the repetition. In (142), WG repeats (with slight modification) AD's entire sentence, while in (143) LL repeats YL's Malay phrase.

(142) 1\AD *(Bu) Bôtmir ká* ***NF yang*** *seiai datalênga daukuk* ^ *(takardi).*
(Mr.) Botmir and his group are the only ones who hear with ^ (their ears).
2\WG^ */P* ***Bu•Bôtmir*** *ká seiai* ***NF yang*** *datalênga daukuk êra takardi.*
^ Mr. Botmir and his group alone are the ones who hear with their ears.

(143) 1\YL *Mo* ***P Dedi*** *ká,* ***P Dedi*** *ekar konar nono, [.]* ***Y# satu Y pulau****.* ***NF Macang*** *[a] Perín ja* ***P Kalar-Kalar*** *êr, [satu kampung, to? e dua-]* ***Y baku-sabláh Y kampung****.*
But Dedi and his group, Dedi married that woman, [.] one island. Like [uh] Feruni and Kalar-Kalar, [one village, you know? uh, two-] neighboring villages.
2\LL ***R Baku-sabláh•kampung****.*
Neighboring villages.

So strong is this desire to affirm the other speaker by repetition that speakers not only repeat lexical items of Malay, a language which they speak, but also sometimes repeat (or attempt to repeat) lexical items of another WT dialect which they cannot speak fluently. In (144), LL repeats OK's WTB verb *i-dom* '3s-go', even though LL cannot speak WTB:

(144) 1\OK *Mo idom* ***P Benjina****,*
But he went to Benjina,
2\LL *Oh, ja= idom* ***P Benjina*** *=sin =e?*
Oh, so he went to Benjina already?

Another indication of the strength of social relations over linguistic competence is evident where speakers confirmingly repeat the performance error of another speaker. In (145), AD says *kegelapan* 'darkness' when she intended to say *gelapkan* 'darken', but before she can repair her error OK repeats *kegelapan*.

(145) 1\AD *Mo kenjou [: e]* ***Y laing Y dari*** *ne, [kegelapan,]*
But if [:] [uh] aside from that, [darkness,]
2\OK ***R Kegelapan****. [.] (pasti.) [.]*
Darkness. [.] (certainly.) [.]

3\AD *[Z]**gelapkan** ne, [NF]**pasti**. Ou na ekapúi.*
darken that, certainly. The fire will destroy it. [2]

In (146), LL ends turn 1 pausing immediately after using a non-default Malay word, and before she can finish her sentence with a pure WT verb phrase, AD jumps in and offers a mixed verb phrase. LL, then, obligingly repeats the mixed word—no doubt she had to, since AD is an older relative.

(146) 1\LL *{laughing} (Ja ne) maera•ne [a] [K]**ipar** nêr nena [L]**marontak** êr, {laughing} [DF]**tarús** (maera•ne) daela datôr [/K]**Bapa**•Garjalái imá, {laughing} isí isursur nei. Inam lalean nêr tan [K]**ipar** [k-] kobir on, nei [a] [Y]bagará [D]satu•kali, {laughing} ka (inam) lalean nêr silna. [.] [D]**Untung** [K]**Bapa** nono [Y]**cepat** [a:] ^ inal =na pei.*
{laughing} (So that) that day [uh] my brother-in-law revolted, {laughing} then (that day) they went and called Mr. Garjalái to come, {laughing} he went and injected him. He put the needle on my brother-in-law's [.] thigh, he suddenly moved, {laughing} and then (he made) the needle bent. [.] It's a good thing he quickly [uh] ^ pulled it out.

2\AD ^ *[%Y]**I-cabu** =na.*
^ He pulled it out.

3\LL *[%R]**I-cabu** =na. Mo (e) kôtan [a] erbata tan [K]**ipar** tapôran.*
He pulled it out. But (uh) maybe [uh] it broke on my brother-in-law's body.

Accommodating repetition works the other direction as well, bringing in WT where there might otherwise be Malay. In (147) YL is using Malay to report a Malay conversation, but in turn 4 LL guesses, using WT, at what YL is about to say. YL then reports the rest of the conversation in WT.

(147) 1\YL *Taper [a] nen, [on a] [/PK]**Edi**•kanei•**bapa** ká. Ruai. Êra [%D]da-**lewat**, [.] «[Y]**Paitua!**»*
This man, [uh] Edi's dad and his group. Two of them. They passed by, [.] "Man!"

2\LL *Eya.*
Yes.

3\YL *[DF]**Tarús** nei nam•on; «[Y]**Iya!**» «[Y]**Dong [Y]pi [Y]mana?**» [DF]**Tarús** taper ruai ir dam on [a]*
Then he said; "Yes! Where are you (PL) going?" Then the two men said [uh]

4\LL *«Maimai [P]**Lokasi**.»*
"We're going to Lokasi."

5\YL *«Kam' maimai* ᴾ***Lokasi*** *on!»* ᴰᶠ***Tarús*** *nei nam•on, «Eh!* ^ *Kem ja joba!*
"We're here going to Lokasi." Then he said, "Hey! You (PL) are fortunate!"
6\LL ^ *Kem joba*
You (PL) are fortunate
7\YL *Aka maera gogongar ka misí, ka [:] mitopatopa* ʸ***taripang*** *ino,*
Because all day long you go, and wash sea cucumbers,
8\LL *Mo ok,* ^ *ja [:]*
But I,
9\YL ^ *Mo ok, ja sakalikang. Êr on, ok jôla [ea] jeresar•papa loloar•pit•gair on, mo ken ja samayarkim. Maera ia, loloar•pit•gair êr ka joujou kem misísí aka mertopatopa [:]* ᴿ***taripang****.»*
"But I, not so. Here it is, I usually sweep the ground early in the morning like this, but you are fortunate. Every day, early in the morning I see you go to wash sea cucumbers."

Following the previous speaker's language choice is by no means automatic or even predictable. In (148), LL pauses and YL offers a Malay end to her sentence, which LL obligingly repeats but then continues her sentence in WT.

(148) 1\LL *Ja [:]* ᴾ***Isak*** *nono,* ᴾ***Usi•Nona*** *ká idá [:]*
So [:] that Isak, is Ms. Nona and her group's [:]
2\YL ***ᴷadi ᴺᶠyang ʸbongso*** ^ ***ʸsekali****.*
younger sibling which is the youngest.
3\LL ^ ᴿ***adi•yang•bongso•[ska-]•sekali*** ᴺᶠ***yang*** *datôr nei damdam* ᴾᴷ***Tete•Dobo*** *nêr, nene =sin =e?*
younger sibling which is the youngest, the one they call Tete Dobo, that one?

Example (149) demonstrates the necessity of examining each language choice in its discourse context, because here there is an overt negotiation of lexical choice which involves language choice. In Malay the collocation *bakar lilin* 'burn candle' is the typical way of saying 'light a candle', even though it could also mean throwing a candle on a fire. In WT, however, the verb *-lan* 'to light' is typically used for lighting candles and lamps, while *-para* 'to burn' is also an option. In this example, TN uses the mixed phrase *iparapara lilin induk* '3s-burn main candle' in turn 3.[135] LL does not correct her, but when she herself wants to refer to the lighting of the candle, she rejects TN's choice, then almost uses the mixed word *ibakar* '3s-burn', hesitates

[135] Note the original context was a Malay-speaking context, so TN's memories may have been in Malay.

twice, then finally pulls *ilan* '3s-light' out of her mental lexicon. TN responds by accepting LL's lexical choice.

(149) 1\TN *Ok kusí kom•on,* **[NF]*to?*** **[DF]*Tarús*** *[e]* **[%Y]*da-minta*** *[:]* **[J]*ketua*** *aka ersirsir. Ja korjou kom•on, sikali kujoujou [a]*
I went like this, you know? Then [uh] they asked [:] the chairman to speak. So I looked like this, I didn't see [uh]

2\LL **[P]*Bu•Pit*** *maerane ime* **[D]*luar***.
Mr. Pit that day was outside.

3\TN **[NF]*Ada!*** *(ime ne ime ne) min,* **[DF]*tapi*** *nei kali [a] Korjou aka =na,-Ei,* **[DF]*bukan*** *aka nei ersirsir, mo iparapara [:]* **[C]*lilin*** **[Y]*induk*** *êr.*
He was there! (He was there) too, but he didn't [uh] I looked at him,-Hey, not for him to speak, but to light [:] the main candle.

4\LL *Eya.*
Yes.

5\TN **[DF]*Tarús*** *ok kom•on; "Ja nen [a bu-]* **[P]*Bu*** *nen ei* **[J]*ketua*** *=sin =e?"*
Then I said, "So this [uh] this guy is the chairman now?"

6\LL *Ja maera•ne* **[P]*Cepi*** *[iba i- i-] ilan =e?*
So that day Cepi [light-] lighted it?

7\TN *Ilan [e]* **[R]*lilin•induk*** *no.*
He lighted [uh] that main candle.

Example (149) demonstrated how speakers sometimes negotiate the best WT equivalent of a Malay word or phrase. Example (150) is another example of this. Here, OK explains that a large arch near her house is called *Pintu Gerbang* by the locals; she describes it using the mixed phrase *pintu jinjinai* 'large door'. But AD is unable to identify the referent. Now in WT, *leipapai* refers to a doorway, while *anam* refers to the door itself; but in Malay, *pintu* refers to either of these.[136] So OK not only gives AD a more detailed explanation of the referent, but in doing so replaces *pintu jinjinai* with *leipapai jinjinai*, to ensure that AD was not thinking of an anam.

(150) 1\OK *Ja [.]* **[NF]*macang*** *danga «*[P]***Puncak****», ***[NF]*yang•penting*** *(dalala tantan kamá no), ja danga [.] «*[P]***Gerbang****». ***[DF]*Barang*** *[a]* **[Y]*pintu jinjinai*** *no, one dam ne itom* **[P]*Pintu•Gerbang***.
So [.] like they say "Puncak", what's important is (they run to our thing), and then they say "Arch." Because [uh] that big door, that they name that "Arch Door".

2\AD ^ **/[R]*Pintu•****jinjinai ba?*
^ What big door?

[136] There is also a Malay phrase, *daun pintu*, which refers only to the door itself.

3\LL ^ ***RPintu•Gerbang.*** *(...)*
^ Arch Door. (...)
4\OK *Mo sita ona %Nta-**masu** tal leipapai jinjinai **PEnce-Ming** kanei (tapaha ada kala no,) te. One ja [.] dam tom ane [:] **PTermenal•Gerbang.** {laughs}*
Well we enter through the big doorway that is Ence-Ming's (...) of course. That one they name [:] Arch Terminal. {laughs}'

5.3 Second cycle of analysis: Form categories

Based on both the interviewees' responses concerning gratuitousness and equivalence, as well as my analysis of frequencies of occurrence of all lexical units in the corpus, and checking all of this against my own (imperfect) intuitions as a speaker of both languages, I have divided the occurrences of non-default Malay into seven form categories which seem to me to be the most fruitful for framing subsequent discussion. The categories are as follows; note that the first three are sequences not containing any adjacent non-default Malay lexical units, while the latter four must contain adjacent non-default Malay lexical units.[137]

- Singly occurring non-default Malay lexical units
- Collocation sequences: a non-default Malay lexical unit which occurs because it collocates with a default Malay word
- Triggered sequences: a non-default Malay lexical unit adjacent to (but probably not collocating with) a default Malay word
- Complete clauses or sentences or clusters of sentences in Malay, containing at least two adjacent non-default Malay lexical units
- Complete sentence fragments in Malay, containing at least two adjacent non-default Malay lexical units
- Near-S: clauses or sentences which contain at least two adjacent non-default Malay lexical units and are completely Malay, except that they either begin with a WT conjunction or end with a WT tag, or both.

[137] Words with pluralizing reduplication are marked with a plus sign (+) and considered non-default for the purpose of identifying CS form categories, even if the Malay root is a default item, since the reduplication itself is gratuitous. In contrast, non-default closed-class items which occurred more than five times in the corpus were considered to be default Malay for the purpose of identifying these stretches of Malay mode; that is, for this stage of the analysis I judged absolute frequency of closed class items to be more important than frequency relative to their WT equivalents.

- Subclausal CS: stretches of Malay which are less than near-S, including some which stress across sentences boundaries, containing at least two adjacent non-default Malay lexical units

It should also be noted here that all of the sequences conform to Malay word order (which is for the most part identical to WT word order). Excluding the thirteen instances of addressee-related code-switching mentioned above (since they involve an additional parameter not in focus in the current discussion), the number of times that each of these seven form categories occur in the corpus is shown in table 5.1 (DQ = direct quotes).

Table 5.1. Form categories of CS in the WT/Malay corpus

	-DQ	(%)	+DQ	(%)	Total	(%)
1. Single non-default items	1205	(69.3)	322	(50.4)	1527	(64.2)
2. Triggered sequences	258	(14.8)	46	(7.2)	304	(12.8)
3. Collocation sequences	150	(8.6)	38	(5.9)	188	(7.9)
4. Complete-S	45	(2.6)	177	(27.8)	222	(9.3)
5. Subclausals	72	(4.1)	25	(3.9)	97	(4.1)
6. Complete S-fragments	8	(0.5)	21	(3.3)	29	(1.2)
7. Near-S	1	(0.06)	8	(1.3)	9	(0.4)
Total	1739	(100.0)	637	(100.0)	2376	(100.0)
(%)	(73.1)		(26.9)		(100.0)	

These seven form categories are now considered in turn. It is crucial to remember throughout the discussion that the assignment of any particular string of words to one category or another is no more valid than the separation of Malay lexical items into default and non-default. Since some non-default items are more gratuitous than others, the categorization is not as certain as I might wish; however, having made the distinction I am much nearer to the truth than if I had merely lumped all the Malay words together or relied purely on frequency of occurrence.

5.3.1 Single non-default lexical units

As in many other studies of LCP, single lexical units dominate table 5.1. There are a number of lexical units (especially conjunctions) which occur

dozens of times; the 1,205 units (tokens) found outside of direct quotes represent only 553 types, while the 322 units found within direct quotes represent only 190 types.[138] Although I expect a high token-type ratio with conjunctions, it is also possible that some of these should be considered established loans (of the displacement category, since they are semantically equivalent to WT conjunctions still in use). However, I have only allowed high-frequency lexical units to bear the label non-default if the WT equivalent has an even higher frequency—in other words, the high frequency in these cases represent not lexical frequency of Malay in contrast to WT, but frequency of the concept. About 79 percent (1205/1527) of the single non-default lexical units occur outside of direct quotes, a fact to which I shall return shortly.

5.3.2 Triggered sequences

The second-largest form category of Malay stretches in the corpus is that of triggered sequences.[139] These are instances of default Malay either preceding or following or in-between non-default Malay, without two non-default Malay items adjacent to each other. In these instances, I propose that the use of a default Malay item created a mental environment which made it easier for the speaker to choose Malay for the preceding or subsequent lexical item. In other words, although in a static lexical analysis (as described in §4.2) a lexical unit may be considered non-default, in a specific discourse context it may actually be more toward the default end of the continuum. These are not instances of collocation but of proximity: the triggering lexical unit and the triggered lexical unit are adjacent. The crucial difference between collocation and triggering is that collocation represents a relatively static relationship (stored in the mental lexicon) between lexical entries which frequently co-occur (and are not necessarily adjacent), whereas triggering involves two or more lexical entries which happen to be adjacent in the sentence currently being processed by the speaker. Collocation, then, is a part of competence, while triggering is closely tied to performance.

In (151) the triggered sequence apparently began with a single lexical choice, *bukan* ‘not’, along with its projected syntactic pattern (*bukan*

[138] As might be expected, this is in sharp contrast to the other six categories, where the number of types and tokens are approximately equal due to the fact that they represent strings of words produced on the spot rather than strings of words which exist as strings in the speaker’s mental lexicon. It should be noted here that there is significant overlap between the lexical units found within direct quotes and those found outside direct quotes, though a number of Malay lexical units are found only within direct quotes.

[139] Not to be confused with the triggering described by Clyne (1967), which is discussed in §5.4.4.

[noun] *yang* [verb] ‘it was not [noun] who [verb]’). Rather than switch back to WT for just the final word of the sentence, AD allowed the preceding words to overwhelm the normal language choice for the concept ‘to buy’ and simply finished the sentence in Malay.

(151) \AD *(mo- kai-) Takom momom* ***[P]Memed,*** ***[DF]bukan [P]Memed ([NF]yang) [Y]balí.***
(but-) Don’t say Memed, it wasn’t Memed (who) bought it.

One caveat in this regard: It is likely that some strings of Malay which I have counted as triggered sequences are merely coincidental juxtapositions of non-default and default Malay items; but since I have no neural data[140] available, I will simply treat all sequences that fit the definition of this category as members of the category. I am not attempting to convince the reader here that my explanation is the correct one, but only that it is a possible one. I know of no way to prove the matter one way or the other. Any lexical items miscategorized here actually belong in the category of single non-default items, making that category even larger than it already is. Like that category, most triggered sequences are found outside of direct quotes (258/304 = 85 percent).

I have claimed in chapter 4 that Malay lexical units which are Malay by default (or the most convenient way to render a concept) do not qualify as indicators of Malay mode and should not be used as counter-examples to proposed syntactic constraints on CS. However, this does not mean that speakers are unaware that these words are Malay words. There is a psycholinguistic continuum, unique for each speaker (and even changeable over time for a single speaker), between established borrowings and what Poplack and others have referrd to as nonce borrowings. The former are truly neutral in their language membership, while the latter have a dual language membership: while speakers consider a Malay word to be the best way to render a concept (in a given lexical and syntactic context), they are still aware (if only subconsciously) that the word is a Malay word. These Malay words not only bring along their own Malay collocations with them in the transfer from Malay to WT, their use also triggers the use of more Malay—not only individual non-default Malay lexical units, as described here, but also longer strings of Malay, as discussed in §5.4.3.

I find justification for this category in the many gratuitous Malay lexical units which never occur except adjacent to other Malay items (whether default or non-default). Included here are the Malay pronouns and determiners, as well as the items in (152)–(154).

[140] It may be that neural scanning equipment will develop in accuracy to the point where neurolinguists can find correlations between neural activity and language choice or lexical choice, but that day has not arrived yet.

(152) nouns

ana(k)	'child'
apa	'what'
beso(k)	'tomorrow'
bini	'wife'
hari	'day'
ikan(g)	'fish'
kampung	'village'
laki	'man'
malam	'night'
mana	'which, where'
orang	'person'
rumah	'house'
sampan(g)	'canoe'
sana	'there'

(153) verbs

berlalu	'passed'
bicara	'speak'
bilang	'say'
jual	'sell'
manangis	'cry'
mari	'come'
pi, pigi	'go'
tanya	'ask'
tatawa, ketawa	'laugh'
tidur	'sleep'

(154) others

begini, bagini	'like this'
di	'at, in, on'
hilang	'lost, gone'
ke, ka	'to'
lai, lagi	'again'
pu(ng), punya	(genitive marker)
sekali	'very'
tadi	'a while ago'

Being gratuitous means that their WT equivalents occur numerous times, indicating that the low Malay frequency is not due to low frequency of the concept. For these concepts, then, avoidance of Malay is the general rule. Therefore, the best explanation for the occurrence of these lexical items in the corpus is that they are triggered by the Malay item next to them. If the adjacent Malay item is Malay by default, the resulting

sequence is counted as a triggered sequence in my analysis; if the adjacent Malay is non-default, however, the sequence is counted as subclausal, complete-S, S-fragment, or near-S.

When a conjunction or preposition is triggered, it is often sandwiched between two default Malay items rather than merely adjacent to one.

(155) 1\YD ***C**Generator **Y**besar* *te* ***C**generator **Y**kecil?*
The big generator or the little generator?
2\YL ^ ***R**Generator•kecil. **P**RKI.*
^ The little generator. RKI.
3\LL ^ ***R**Generator•kecil **Y**di **R**RKI.* *Ena ná [:] kalor sian dir*
^ The little generator at RKI. Under the banana trees.

(156) \LL *«Ja **DF**biar kama marteya min, kama na marpei•sêra kakamá [:] lêt ino, mo kam' masika êra peda. Ka ino (ja) kakamá **D+**teman-teman **Y**dari [e] **P**SMP **DF**sampi [a SMA.] **P**SMEA.*
So even if we get married, we will take leave of our [:] husbands, and go see them. Because those ones are our friends from Junior High through [uh High School.] Business School.

(157) 1\LL *Kêl gakar ka= mnun!*
Buy medicine and take it!
2\YL *(kêl) **Y**obat [a]*
(Buy) the medicine [uh]
3\LL *Kêl **P**Antimo.*
Buy Antimo.
4\YL ***P**Antimo **Y**deng [e]*
Antimo and [uh]
5\LL ***P**Klorokwin.*
Chloroquine.
6\YL ***P**CTM.*
CTM.

(158) \OK ***DF**tapi* *êra sakali **Y**tahu•persis **YH**bahwa [.] «Ok kanáng lêt nen, **C**pegawai•negeri **Y**atau **C**swasta. [.] **Y**ataukah **C**karyawan.*
but they don't know precisely that [.] "My husband is a government employee or private employee. [.] or laborer."

The sandwich trigger is of course not a deterministic phenomenon, since WT prepositions and conjunctions do occur in sandwich trigger environments as shown in (159)–(160).

(159) \SD ***[C]Minggu•[D#]pertama*** *ne,* ***[N]Balai•Desa*** *pel [parigi] [.]* ***[N]parigi*** *dir [a]* ***[DF]hampir*** *puipuidi....ja* ***[DF]waktu*** *[.]* ***[N]perpisahan*** *one, ja [.]* ***[P]Bupati*** *ka elalú,* ***[J]Komandan•Angkatan•Laut*** *ei* ***[P]Ambon*** *ka [.] elalú,*
That first week, the Village Hall and [well] [.] the wells [uh] were almost finished....so when [.] that departure, [.] the Bupati came down, the Navy Commander in Ambon also [.] came down,

(160) \OK *Mo êra dam (tapir) ina* ***[N]sibuk*** *ei* ***[C]kantor,***
But they said (the man) was busy at the office,

In (161) LL permits a sandwich trigger, while AD replaces the triggered word with Malay in her agreeing repetition.

(161) 1\LL *Eya.* ***[D]Tahun•[Y]lama*** ^ ***[Y]deng (tahunbaru)***
Yes. The previous year ^ and the (new year)
2\AD ^ *Eya.* ***[RD]Tahun•[Y]lama*** *pel* ***[RD]tahun•[Y]baru.***
^ Yes. The previous year and the new year.

In (162) AD is attempting to get a two-sentence quote out of her mouth, but it takes her four turns to do it because of WG's conversational participation. Her second sentence begins in turn 5, and WG's suggested continuation of the sentence in turn 6 is wrong on two counts: semantic content and language choice. As AD continues in turn 7, she does end up with a string of six Malay words, but only the last word (*saja* 'just') is non-default (note that when the concept recurs three words later, AD uses the WT equivalent *gatan*). It is likely that *saja* is triggered here by the five preceding default Malay words. The Malay word *orang* 'person' would be gratuitous if not for the fact that the cultural import *petani* 'farmer' requires it.

(162) 1\AD *(kai ka) êra damon; «*[YF]***Kenapa*** *damdam [a on a]*
^ ***[C]Sektor*** *[a]*
(...) they said, "Why did they make [uh uh]
^ Sector [uh]
2\WG ^ ***[C]Sektor*** *no [a]*
^ that sector [uh]
3\AD ***[P]Oret*** *no [a]*
Oret [uh]
4\WG *eyei* ***[C]tuan•rumah****?*
be the host?
5\AD *eyei* ***[C]tuan•rumah****? [.]* ***[C]Sektor*** ***[NF]yang*** *[e]*
be the host? [.] A sektor which is [uh]
6\WG ***[C]Sektor*** ***[NF]yang*** ^ ***[Y]basár*** *[e]*
A sektor which is ^ big [uh]

7\AD ^ ***[NF]cuma*** *[:] [2]* ***[Y]orang•[C]petani*** ^ ***[Y]saja*** *pel* ***[C]nelayan*** *gatagatan ino, mo dam eyei* ***[C]tuan•rumah.»***
^ only [:] [2] farmers ^ only and just fishermen, but they make it the host."
8\?? ^ ***[R]orang•petani•saja. [R]Nelayan.***
^ just farmers. Fishermen.

5.3.3 Collocation sequences

In chapter 3, I explained that a lexical analysis of LCP in a bilingual corpus required dividing the corpus into lexical units—sometimes an individual lexeme, sometimes a phrasal lexical entry (e.g., phrasal names like *Ujung Pandang*), and sometimes a string of words in one language which is equivalent to a single lexical entry in the other language (e.g., WT *nam* on 'like this' which corresponds to Malay *begini*). There has been some debate in the LCP literature regarding the status of "single-word EL insertions"; I argue that the focus should not be on single words or even lexical entries, but instead on these lexical units, which are defined not by how any single language is stored in the speaker's mind, but by how the two languages in a bilingual mind interact with each other.

In contrast to phrasal lexical entries, collocations are relations between individual lexical entries stored separately in the mental lexicon. Like triggered sequences, collocation sequences contain non-default Malay units adjacent to default Malay units, but not adjacent to other non-default Malay units. In contrast to triggered sequences, however, the non-default Malay unit here has a collocational relationship with the default Malay unit. In this category, therefore, it is not that the default Malay lexical unit triggers a Malay language choice; rather, it motivates a lexical choice (which happens to be a Malay lexical item). As mentioned above, triggering is a dynamic phenomenon, situated entirely within the discourse context; collocation is a more static phenomenon, residing in the mental lexicon. In this corpus, most (150/188 = 80 percent) collocation sequences occur outside of direct quotes.

Some collocations are stronger than others; in a strong collocation, the non-default Malay word is more natural than its equivalent WT word and is therefore more default than non-default in that particular context. On the other hand, sometimes a collocation is too weak (or a Malay word is too gratuitous) for the non-default item to appear; different speakers may assign differing strengths to the same collocations, and the same speaker may not be consistent in this regard. Thus, in the corpus I find synonymous phrases, one being mixed and the other being a fully Malay collocation sequence. In (163)

there is an instance of other-repair, which began with AD using the collocation *sama juga* 'just the same'. OK then replaced *juga* with the WT equivalent *gatan*, indicating that the collocation *sama juga* is a weak collocation.

(163) 1\AD *Kenjou nekanei [.]* ***ᴰorang•tua ᴺmampu****. Mo nekanei* ***ʸwatak sikali ᴺmampu,***
If his [.] parents are capable. But if his mind isn't capable,
2\OK *Mm.*
Mm.
3\AD ***ᴺsama ʸjuga****.*
just the same.
4\OK ***ᴺSama*** *gatan.*
just the same. [2]

The words meaning 'day' and 'month' contrast interestingly in this regard. Malay *hari* 'day' never occurs except adjacent to another Malay word; in addition, since the Malay day names never occur (in thirty-nine occurrences) without being preceded by *hari*, I have analyzed them as phrasal lexical entries rather than as triggered sequences. But Malay *bulan* 'moon, month' does occur with no adjacent Malay twice, and the Malay month names do occur without being preceded by *bulan*. Moreover, on two occasions the Malay month name was preceded by WT *pôlan* 'moon, month' as in (164)–(165).

(164) 1\LL *Eh! Nungatin [a]* ***/ᴾpôlan•Maret*** ^ *(emalilia minmin mo a)*
Hey! Not yet [uh] the month/moon of March ^ (appear yet, but uh)
2\YL ^ *[On a:] Aroka na puipuina min =sin?*
^ [uh] Soon it will be gone again?
3\YD *(...)* ÷
(...) ÷
4\YL ^ *Apara on?*
^ This west season?
5\YD ÷ ^ *(utu)* ***ʸbulan ᴾMaret*** *pui.*
÷ ^ (...) the month of March is over.

(165) \SD *Nene ipel [bulan aa pôlan] [.]* ***/ᴾpôlan•Agustus****.*
That was during [the month uh the month] [.] the month of August.

Since the Malay month names occur with *bulan* in ten instances, which is considerably more often than with *pôlan*, I see a strong but not inviolable collocation between the month names and Malay *bulan*.[141] It may be significant that in (164) the WT word *pôlan*, which violates the collocation,

[141] According to Collins (personal communication), the same strong collocation is found in many other languages of Indonesia.

is preceded by a pause filler, and that YD subsequently uses *bulan* rather than repeating *pôlan*. In (165) the speaker actually said *bulan*, then repaired it to become *pôlan*. Since the speaker in this case was SD, this may be an instance of hypercorrection; AD as audience seemed to be more of a purist during the sessions with her brother SD. In this case, then, it appears that social factors were stronger than speaker-internal lexical factors (frequency of collocation).

5.3.4 Complete-S code-switching

The complete-S category consists of clauses or sentences which are completely Malay (including any initial conjunctions), and contain at least two adjacent non-default Malay lexical units. Whenever adjacent clauses or sentences are completely Malay, with no intervening WT elements, they are counted as a single instance in table 5.1.[142] In contrast to the category of single non-default Malay lexical units, the complete-S instances mainly occur within direct quotes (170/214 = 79 percent). Example (166) contains a complete-S instance of Malay outside a direct quote, while (167) contains an instance within a direct quote of a non-WT speaker.

(166) \DA *(...)* ***[Y]rumah•[C]pegawai [Y]apa•sa, [Y]samua [Y]di [C]sektor [Y]dorang [Y]situ.***
(...) employees' houses of any sort, all are in their sector there.

(167) \SR ***«[Y]Kamorang [Y]ini [.][Y]paki [C]sandál [Y]par [Y]apa?»*** ...
"You (PL) [.] wore sandals for what?"

5.3.5 S-fragment code-switching

Sentence fragments stand alone phonologically as if they were complete sentences, but are not complete syntactic sentences. These are often answers to questions or telegraph-style comments. Like instances of complete-S CS, about three-fourths of Malay sentence fragments in the corpus (21/29 = 72 percent) occurred in direct quotes.

(168) 1\OK ***[D]Satu•kali*** *êr aka ok muna* ***[S]barangkat.*** *[.]*
Immediately I was about to leave. [.]
2\LL ***[R]Barangkat*** *ei ba?*
Leave for where?

[142] Whenever an instance of subclausal CS occurred at the end of a clause and was immediately followed by an instance of complete-S CS, the entire stretch was counted as an instance of complete-S CS rather than as subclausal. This occurred only a few times.

3\OK *{laughing} Eh,* ***[Y]ke [Y]mana [Y]saja.*** *[.]* ***[DF]Jadi,*** *[.] barian [a k-] on, ko•tên* ***[%A]ku-paksa*** *ko maera•ne madom* ***[J]dokter.***
{laughing} Hey, to anywhere. [.] So, [.] later [uh] then I forced myself that day to go to the doctor.

(169) \LL ***[P]Tabita*** *naka nei (nam)* ***«Eh, [Y]nyong [Y]pu [K]bapa?»***
Tabita said to him, "Hey, the boy's father?"

In some instances, a sentence fragment was only one or two words long, and there were no consecutive non-default Malay words, but I included it in this category anyway because it was part of a sequence of completely Malay direct quotes, as shown in (170)–(171).

(170) \?? ***«(Kas-makang deng) apa?» «[Y]Pisang [Y]ini.» «Ko kasi berapa buah?» «[Y#]Ampa.»***
"(Fed it with) what?" "These bananas." "How many did you give it?" "Four."

(171) \LL ***[P]Dace*** *naltan.* ***«[Y]Apa [P]Lin?»*** *(Ok one.)* ***«[Y]Ose [Y]deng [P]Dorenci [Y]lari [NF]sudah.»***
Dace arrived. "What, Lin?" (I was there.) "You and Dorenci go ahead and elope."

5.3.6 Near-S code-switching

There are nine clauses or sentences in the corpus which are entirely Malay except that they begin with either a WT conjunction or the emphatic adverb kai, or end with a WT tag; they also contain at least two consecutive non-default Malay lexical units. These I refer to as the near-S category; most of these (8/9 = 89 percent) occur outside of direct quotes. Example (172) begins with *kai* and is the only Near-S example not occurring within a direct quote.

(172) \AL *Kai* ***[Y]dia [Y]pu [K]om [Y]dong [Y]semua [Y]setuju.***
His uncle and his group all agreed!

Five of the nine instances of Near-S involve a clause introduced by the WT conjunction *ja* 'and, so, then'. In one of these, this conjunction occurs at the beginning of the direct quote; in the rest, it occurs between two Malay clauses as in (173).

(173) \SR ***«[DF]Barang [JK]Bapa-guru [Y]tara-paki,*** *ja* ***[Y]katong [Y]ada [Y]paki aju-aju [Y]orang [Y]bajalang.»***
"Because Teacher wasn't wearing them, so we wore them acting like people walking."

In (174), the WT enclitic tag *=si(n)* 'already' occurs at the end of a Malay sentence; similarly, (175) shows a direct quote which is all Malay except for the final WT tag which identifies a yes/no question.

(174) \SR *«(Oh) [Y]seng, [K]Bapa. [Y]Tadi [Y]katong [Y]tinggal [Y]situ mo [Y]pateka ([Y]itu [Y]tadi) [Y#]tiga [Y]itu =si.»*
"(Oh,) no, Father. We were there, and (those) three cucumbers were there."

(175) \WG *Ok kom•on, «Eya. [P]**Usi** ma= ipo =yai. Ir in.» **«([YF]La) [Y]kamong [Y#]dua [Y]beso [a] [Y]jual** =e?»*
I said, "Yes. I brought them. Here they are." "(So) you (PL) two tomorrow [uh] sell?"

5.3.7 Subclausal code-switching

Whenever two or three non-default Malay lexical units occurred in sequence, without qualifying as a complete-S, near-S, or complete S-fragment, I have analyzed them as a stretch of subclausal CS. Any default Malay lexical units contiguous to a sequence of non-default Malay lexical units were also included as part of the subclausal sequence. I also included in this category two non-default Malay lexical units joined by such semantically empty default Malay items as the relativizer *yang*, the copula *ada*, or the tag *kan* 'you know'.

These subclausal stretches are not necessarily complete constituents; in fact, they are not necessarily in the same phrasal constituent at all. Some (though relatively few) constitute the end and beginning of adjacent sentences. While this might raise some eyebrows among those who view lexical insertion as the last step in the construction of sentences, I am allowing the possibility that linear order, rather than (or in addition to) constituent structure, comes into play in switching language modes. In other words, we have no a priori reason to disallow the possibility that a speaker may choose to switch languages without regard to constituent structure.[143] In (176) the Malay head noun of the subject NP is followed by a Malay stative verb, and is analyzed as a subclausal stretch.

(176) \OK *(je kana) ok kanáng [D]**sifat**, ja [NF]**lebái** (tero kol) gun kupei tamata ja kanáng **[Y]hati [Y]senang**.*
(so) my character, preferably (I take) blood from somebody so my heart [will be] happy.

[143] Neither do we have any reason to assume that speakers switch by only one strategy, i.e., either abiding by or ignoring constituent structure.

However, many (but not most) subclausal stretches do in fact correspond to a complete NP, a complete PP, or a complete predicate.[144] These constitute 53 percent (38/72) of subclausals outside of direct quotes, and 28 percent (7/25) of subclausals within direct quotes. As a whole, most subclausal CS stretches occur outside of direct quotes (72/97 = 74 percent).

Because of syntactic and lexical differences between two languages, there are times when a certain lexical choice prevents a switch to the other language. In (177) YL chooses *masih* 'still' rather than the WT equivalent, even though the WT equivalent is more common in the corpus for all speakers, including YL herself.

(177) \YL *Ok kom•on «Ah! Ok mo (ana)* ***ʏmasih [a] ʏbujang*** *nen!*
I said, "Ha! If I were still single!"

One possible explanation is that YL had already chosen *bujang* 'single', which triggered *masih*. Another possibility is that once she chose *masih*, she had no choice but to continue with Malay *bujang*. There is actually no WT lexeme equivalent to *bujang*, but the phrase *nungatin korteteya* 'not_yet 1s-married' is equivalent to either *bujang* or *masih bujang*. Note that YL pauses after *masih*, perhaps realizing she has committed herself to saying *bujang*, since it would not be possible to say *masih nungatin korteteya*. The possibility that *masih bujang* is in fact a phrasal lexical entry rather than a sequence of two lexical entries should be considered; however, *bujang* occurs six other times in the corpus, all without *masih*. I conclude that the relationship between *masih* and *bujang* is most likely one of collocation only.

5.3.8 Refining the form-based analysis: Disjoint and covert collocation

In §5.3.3 above I discussed collocation sequences in which a non-default Malay lexical item occurred as a result of a strong collocation with a default Malay item. In addition to these instances where the two words in collocation are adjacent to each other, there are also instances of disjoint collocation, as seen in the following examples.

In (178) AD uses Malay *nama* 'name' rather than WT *ngaran*, probably because it collocates with the default Malay verb *coret* 'cross out'.

(178) \AD ***ᴘᴋBapa•Noya*** *ner* ***ᴄcoret*** *ok kanáng* ***ʏnama***.
Mr. Noya crossed out my name.

Evidence that *nama* is here the result of collocation comes from the fact that there are only two other instances of *nama* 'name' in the corpus, one immediately subsequent to the mixed verb *da-kas•masu* '3p-insert' and the

[144] That is, a Malay predicate with a WT subject; complete Malay predicates with no subject NP are classified as complete-S.

other being one of the non-default Malay items in a religious-style stretch (and possibly the result of collocation there as well), in contrast to twenty-nine instances of WT *ngaran* or *ngarin-* 'name'. Since sixteen of these are spoken by AD, it is clear that both for AD and for the corpus as a whole, Malay *nama* for 'name' is highly marked.

In (179) AD uses Malay *makang* 'food' rather than WT *manám*, probably because of a strong collocation with the Malay root of the mixed verb *ma-jamin* '1pe-guarantee'. Note that twenty-two turns later, the adjacent collocation *jaminan makang* 'guarantee of food' occurs (within a direct quote, as in this instance of *makang*).

(179) \AD *ʷTungguru inam•on, [ta-] «Ken me on gatan, mo kama na* *%ᴺma-**jamin** kekanám [a]* *ʸ**makang**.»*
The teacher said, "You just stay here, and we (ex) will guarantee your (PL) food."

In (180)–(181) AD uses the Malay locative preposition *dalam* 'within' rather than the equivalent WT locative noun *abil*, in collocation with the Malay noun *hati* 'heart, soul'.

(180) \AD *ok //ᶜkom•**sombayang** kom [a] on* *ʸ**dalam** ok kanáng* *ʸ**hati** gatan,*
I prayed like [uh] this just within my heart,

(181) \AD *[.]* *ᴰᶠ**Tapi** ok* *ʸ**dalam** ok kanáng* *ʸ**hati** =si.*
But I was pondering. (Literally, 'I was in my heart.')

Elsewhere, the phrase *dalam hati* or *dalang hati* 'within the heart' (referring to speaking to oneself or laughing to oneself) occurs three times. In the transcription I marked *dalam hati* as a phrasal lexical entry; but since it can be split apart into *dalam ok kanáng hati* perhaps it should be considered a collocation sequence.

In (182) AD uses the gratuitous Malay word *ana-ana* 'children' (or possibly 'small child'), triggered by the Malay verb *hardik* 'scold'. Then she uses the Malay stative verb *kecil* 'small', which collocates with *ana-ana*, rather than the equivalent WT *bôrar-ai* 'small-3pa'. This is the only instance in the corpus of Malay *kecil* not being adjacent to another Malay word; in contrast, there are over fifty occurrences of WT *bôrar-*, many of them occurring with WT *gasua* 'child'. I conclude that *kecil* occurs here because of a strong collocation with *ana-ana*.

(182) 1\AD *ᴰᶠ**tapi** kenjou ^ sêta [ta-] [.]*
but if ^ we (inclusive) [1pi-] [.]
2\OK ^ *ᴰᶠ**Tapi [:]** ʸ**dengan** ᴺ**cara** ᴺᶠ**yang*** ^ *ʸ**bagimana?***
^ But with what kind of way?
3\AD ^ *ᴺ**hardik** ʸ+ **ana-ana** no bana* *ʸ**kecil**. Mo= mjou in gatan.*
^scold those children from [when they are] small. Just look at these [children].

In (183) LL uses Malay *terbuka* 'revealed' rather than WT *er-pei=na*, probably due to collocation with Malay *rahasia* 'secret'.

(183) \LL *Dai* [PK]***Tete•Kacamata*** *ká ka datalatalar. On [nungatin:]* [N]***rahasia*** *ungatin* [Y]***terbuka***. [YF]***Karna*** *ok [e] etanatan* [C]***gereja***.
They went to Grampa Eyeglasses and his group's [house] and sat around. At this time [not yet] the secret was not yet revealed. Because I [uh] was still at church.

In (184)–(186) Malay *merah* 'red' and *putih* 'white' are used in collocation with Malay *mobil* 'car' and *oto* 'car' rather than WT *beir-* 'red' and *-lajír* 'white'. These account for three of the seven lone instances of *putih* and *merah* in the corpus.

(184) \LL *Nei* [J]***kanek*** *ei* [C]***mobil*** *[a] [vene- e apa•ini•itu,]* [C]***penginapan*** *kaneina* [Y]***merah*** *nono te.*
He's the conductor in the car of [uh] [Vene- what is it] that inn's red one.

(185) \LL *mo* [C]***oto*** *[a] [.]* [P]***Najiron*** *ká idá* [Y]***putih*** *nêr imá.*
but/while the car [uh] [.] Najiron and his group's white one came.

(186) \LL [P]***Larate*** *ká idá* [C]***mobil*** [NF]***yang*** [Y]***putih*** *[a]* [P]***Lestari*** *nêr*
Larate and his group's car which is white, the Lestari

In (187) OK uses Malay *tangan* 'hands' rather than WT *lima-*, in collocation with Malay *bersih-suci* 'clean and holy'.

(187) \OK %[N]*da-****anggap*** *gatan (ke kpen:) kanám lêt nen [:] kaní* [Y]***tangan*** *dino* [N]***bersih-suci*** *ipe ok. Sakali* [NF]***ada*** *kaní gala•ia aka ok.*
just consider (...) your husband's hands to be completely separated from me. He doesn't have anything for me.

There are also instances of COVERT COLLOCATION, in which a non-default Malay word collocates with a default Malay word which is not actually spoken. In (188) AD describes her state of unconsciousness after a hard fall, using *dalam* 'within' as in (180)–(181), except that in this instance she leaves out *hati* 'heart, soul', using WT *on* 'this' instead.[145]

(188) \AD *Japún* [L]***sadar***. [DF]***Tapi*** *[ok e]* [Y]***dalam*** *[:] ok kanáng on, kunin, mo* [DF]***rupa*** *tamata ina* [N]***layani*** *ok [Ñ].*
Then I woke up. But [I uh] within my "this", I slept, but it was like a person was taking care of me.

[145] If *on* was accompanied by a demonstrative gesture, then 'this' is the correct gloss; however, *on* without any such gesture would be the WT dummy pronoun (cf. English *whatchamacallit* and Malay *ini* 'this'), which AD uses often. Since the data is on audio tape rather than videotape, we cannot know which was intended.

In (189) TN uses Malay numbers in reference to money. As noted in chapter 4, while times and dates never occur with WT numbers in this corpus (except in SD's hypercorrect speech), money, as both a cultural import and as a counted object, occurs with both Malay numbers and WT numbers. In example (189), however, there is no Malay word *rupiah* or *doit* which specifies the object being counted; it is merely implied by the use of Malay numbers.

(189) \TN *«[P]**Ita** =ei môl ok [C#]**lima•ribu**. Eh, [C#]**sepuluh**.» Nei inaka ok. [NF]**Baru** ok kelakela êr [NF]**kan** kupo [C#]**seratus•ribu**, [NF]**to**? Eh, [C#]**seratus•anam•puluh**.*
"Hey Ita, give me 5,000. Eh, 10." He told me. Because when I left, you know, I took 100,000, you know? Eh, 160.

I conclude that just as non-default items within collocation sequences (see §5.3.3) should actually be considered somewhat necessary, so also a number of words counted as single non-default lexical units in table 5.1 are somewhat necessary as well. Ideally, then, each instance of non-default items should be examined to determine whether it is motivated by collocation, either overt or covert. This is but one example of how the CS form categories presented above are merely suggestive of the psychological reality behind language contact phenomena. The next section considers that psychological reality in more detail.

5.3.9 Reconsidering shorter stretches of Malay

With shorter stretches of Malay, there is the possibility that adjacent non-default Malay items are not members of a single stretch of Malay mode, but merely lexical units which are the result of separate non-default language choices and are adjacent only coincidentally. Indeed, given the large number of non-default Malay lexical units in the WT/Malay corpus, it would be strange if none happened to be adjacent to another. Brief collocation sequences, triggered sequences, subclausal CS stretches, and S-fragments need to be checked as to how gratuitous the non-default Malay item is—that is, what is its frequency of occurrence relative to the equivalent WT item. Recall from chapter 4 that items are considered gratuitous if the relative frequency is less than five percent. In fact, many brief collocation sequences and triggered sequences do involve a gratuitous Malay item such as *sana* 'there' in (190), not merely a dispreferred Malay item, thus lending support to the analysis presented here.

(190) \LL *Mo nei* ***[C]tugas*** *ná ne,* ***[P]Telkom*** ***[Y]sana,***
but 3s assigned LOC that (INAN) Telkom there
[NF]to?
you.know
But she was assigned there, Telkom over there, you know?

As for subclausal CS and S-fragment CS, if all the non-default Malay items are in fact gratuitous, as in (191), I would be fairly certain that a switch of language mode has occurred. But if the brief stretch involves only dispreferred items rather than gratuitous items, as in (192), their juxtaposition is more likely to be merely coincidental.[146]

(191) \OK *kursir kunga•ne aka [:]* ***[Y]orang*** ***[NF]yang***
1s.R.speak 1s.like_that for [:] person REL
[Y]tidak ***[Y]tahu.***
not know
I speak like that for people who don't know (about you).

(192) \LD *nei eruk [pakiang]* ***[Y]pakiang*** *(garje) tubôybôy,*
3s 3s.R.use [clothes] clothes (church?) DUP.new
[Y]celana ***[Y]putih.***
pants white
he wore [clothes] new (church?) clothes, white pants.

On the other hand, when a gratuitous item is adjacent to a dispreferred item, the psycholinguistic process involved may be that of triggering as in (193) where *busu-busu* 'thanks to' is dispreferred and *sapa* 'who' is gratuitous; or collocation as in (194), where *dari* 'from' is gratuitous and *atas* 'above' is dispreferred.

(193) \YL ***Mo*** ***[Y]busu-busu*** ***[e]*** ***[Y]sapa?***
but thanks-to [uh] who
But thanks to [uh] who?

(194) \AD *Ok komon, «****[PD]Tuhan•Allah*** ***[NF]memang*** *[e ok ken]* ***[Y]dari*** ***[Y]atas*** *ken jou [ok] ok kanáng* ***[N]kelemahan*** ***[NF]ada.***
I said, "God, indeed [uh I you] from above you see [I] I have weakness."

5.4 Fine-tuning the psycholinguistic approach

In §5.2, I touched on a few conversational contexts of Malay mode observed in this corpus, referring to some of the same phenomena noted by

[146] Additional evidence for Malay mode in this case, however, is seen in the fact that *putih* is not reduplicated as a WT modifier ought to be.

other LCP researchers. In this section, I attempt to provide a more insightful and more comprehensive analysis of LCP, beginning with another look at reported speech.

5.4.1 Form categories and direct quotes

Regarding the occurrence of the seven form categories inside and outside of direct quotes, a clear pattern emerges from table 5.1. Most (1205/1527 = 79 percent) single non-default Malay lexical items, most (258/304 = 85 percent) triggered sequences, most (150/188 = 80 percent) collocation sequences, and most (72/97 = 74 percent) instances of subclausal CS occur outside of direct quotes. In contrast, most (177/222 = 80 percent) instances of complete-S, most (21/29 = 72 percent) complete S-fragments, and most (8/9 = 89 percent) near-S sequences occur **within** direct quotes. For the sake of the present discussion at least, I conflate the former categories as MINOR CS and the latter categories as MAJOR CS. Note that minor CS consists of the three categories with no adjacent non-default Malay lexical units, plus subclausals, and that major CS consists of only clause-like units.

How is the difference between minor CS and major CS to be accounted for? First, keep in mind that most of the corpus does not consist of direct quotes; therefore, the pattern evident in minor CS roughly reflects the pattern of the corpus as a whole. The question then becomes, why is there so much major CS in direct quotes? I propose that when a speaker uses a direct quote, he tends to be more careful about his language choice. In some cases, the choice reflects the language of the original quoted conversation, while in others this is definitely not the case. In either case, though, speakers tend to choose all Malay or all WT for the language of direct quotes. A possible example of this is seen in (195) where YL alternates between WT and Malay in her lexical choice for the concept 'want'. First she quotes herself as saying ***elaka*** *samai, ka samai gatan* 'If she wants bad, then bad is what she'll get'; then a few clauses later says to AD (not in a quote) ***mau*** *samai ka samai gatan*; then almost immediately quotes her aunt as saying, ***Elaka*** *samai ka samai gatan*.

(195) 1\YL ***PK Tanta•[:]•Minggus*** *nam•***on** **[a]** *«Ei,* ***D dekat•P Natal*** *on ja na= m[a][moi a]sika* ***K Mama*** *moi bôt.» Na* ***DF tarús****, ok kom•on; «****K Tanta*** *=ei,* ***N barenti****. On elaka samai, ka samai gatan.* ***YF Sebab*** *[a]* ***NF rekeng*** *[a]* ***A talalu*** *uk =sin.* ***Z Mau*** *[a] nen* ***P Dorenci*** *dem tantan ok te? Êra* ***K adi-adi*** ***NF yang*** ***A sisa*** *in* ***N sama*** *[.] ^ (mo nen Mama)*

Aunt [:] Minggus said [uh] "Hey, when it gets close to Christmas, go see Mama at her house." Then I said, "Oh Aunt, stop.

If she wants badness, then badness is just what she'll get." Because [uh] it's like [uh] much too much already. She wants [uh] Dorenci to do it to me, does she? Those younger siblings remaining are the same [.] ^ (but Mama)

2\AD ^ *(Batúl) kem bôt no abil, tei! Sakali pen [a] abaldi epardi akaka pen.*
^ (True,) your family! Nobody has a good attitude toward the rest.

3\YL ^{DF}***Jadi*** *on,* ^{N}***skarang*** *on, [a]* ^{Y}***mau*** *samai ka samai gatan. [a Mama na er- si-]* ^{K}***Tanta*** *na ersirpei rau on aka* ^{K}***Mama****, «*^{P}***Yos*** *nêr sakali na= imámá aka kem ei bôt minmin. Elaka samai ka samai gatan.*
So now, [uh] if she wants badness, then badness is just what she'll get. [uh Mama will say-] Aunt will say this message to Mama, "Yos will not come to you (PL) at our house anymore. If you want badness, then badness is just what you'll get."

This correlation between form categories and direct quotes validates the distinction I made in chapter 4 between non-default and default Malay lexical items, since the differential patterning of major CS and minor CS in direct quotes would not have been evident without that distinction.

Since there is a clear distinction between the occurrence of these structurally-defined categories inside and outside of direct quotes, there might also be a difference (within each category) between those stretches of Malay which constitute complete syntactic constituents and those which do not. If there is a difference, this might constitute evidence for the psychological reality of phrase structures. However, no such difference is seen. In table 5.2 I repeat the figures from table 5.1 regarding subclausals and complete S-fragments, along with a more detailed analysis of each.

Table 5.2. Subclausal CS and complete S-fragment CS

	–DQ	(%)	+DQ	(%)	Total	(%)
Subclausals	72	(100)	25	(100)	97	(100)
Complete NP, PP	24	(33)	4	(16)	28	(29)
Complete predicate	14	(19)	3	(12)	17	(18)
Non-constituents	34	(47)	18	(72)	52	(54)
Complete S-fragments	8	(100)	21	(100)	29	(100)
Complete NP, PP	3	(38)	9	(43)	12	(41)
Non-constituents	5	(63)	12	(57)	17	(59)

For both subclausals and S-fragments, table 5.2 shows that constituents and non-constituents make up approximately the same relative proportions of Malay stretches inside and outside of direct quotes. That is, the majority of both constituent and nonconstituent subclausals are not in direct quotes, while the majority of both constituent and nonconstituent S-fragments are within direct quotes. Or, comparing the figures horizontally rather than vertically for both subclausals and S-fragments, the relative proportion of constituents inside and outside of direct quotes is approximately the same as the relative proportion of nonconstituents inside and outside of direct quotes. I conclude that the difference between major CS and minor CS is not based on a distinction between constituents and nonconstituents. Therefore, these data do **not** support the notion that the bilingual brain makes crucial reference to phrase structure in order to switch language modes, nor, on a higher level, does it support any particular theory of syntactic structure.

5.4.2 Careful style

It also seems necessary to acknowledge the existence of a difference (that is, a continuum) between careful style and careless style (for lack of a better term) with regard to the language membership of lexical choices. There are two lines of evidence for this in the corpus. [First, as seen in table 5.1, direct quotes are a favored site for the occurrence of complete Malay sentences, complete Malay sentence fragments, or complete Malay clauses with only a conjunction or tag being WT (which I have labeled "Near-S CS").] This is in contrast to the seemingly haphazard occurrence of Malay words outside of direct quotes. Second, there is a clear distinction between SD and the other speakers. As noted in chapters 3 and 4, SD's contribution is more careful than that of the other speakers in the corpus. Strikingly, SD has not a single instance of adjacent non-default Malay lexical units, as seen in table 5.3. Comparing table 5.3 with table 5.1, note that SD has nearly twice as many collocation sequences as triggered sequences, the reverse of what is true of all speakers totaled (304 triggered sequences, 188 collocation sequences). This provides support for my claim that non-default items motivated by collocation are less avoidable (more necessary in that particular context) than non-default items triggered by adjacent default lexical units. It also demonstrates the importance of a speaker-specific (situation-specific) analysis.

Table 5.3. Form categories of CS in SD's speech

	–DQ	(%)	+DQ	(%)	Total	(%)
1. Single non-default items	24	(71)	10	(83)	34	(74)
2. Triggered sequences	3	(9)	1	(8)	4	(9)
3. Collocation sequences	7	(21)	1	(8)	8	(17)
4. Complete-S	–		–		–	
5. Subclausals	–		–		–	
6. Complete S-fragments	–		–		–	
7. Near-S	–		–		–	
Total	34	(100)	12	(100)	46	(100)
(%)	(74)		(26)		(100)	

I have used the words careless and haphazard, and they require some additional comment. There are numerous places in the corpus where I had the impression that a number of single non-default Malay items were sprinkled into WT mode in order to give a more Malay-like flavor to the discourse (see §5.2.2).[147] This would be analogous to the occasional use of high Malay elements (lexical, phonological, morphological, syntactic) in order to convey a more standard style—haphazard in a sense, but also controlled, just as the motion of any particular molecule may be considered random while also fitting into an overall pattern with all other molecules in its vicinity to create measurable temperature and pressure. Thus, it may be necessary to include in a comprehensive model of LCP not only base language **modes**, but also **styles** of language choice, analogous to other styles recognized by sociolinguists. That is, a speaker may be using a Malayish style while still in WT mode. This is obviously a matter which will require further investigation.

5.4.3 WT insertions in Malay mode

Most other-language insertions in this corpus involve the insertion of a non-default Malay lexical unit into WT mode. In this section, however, I describe the reverse: the insertion of WT lexical units into Malay mode.

As noted in chapter 4, certain Malay conjunctions have a high frequency of occurrence in this corpus; I proposed that the high frequency of these conjunctions in monolingual Malay mode makes them feel so natural to WT speakers that they do not hesitate to use them in WT mode, even if there are equivalent high-frequency WT conjunctions. Interestingly, the

[147] Unfortunately, my claim at this point is purely anecdotal; I have not attempted to find any statistical support for my perception.

reverse also occurs: while in Malay mode, a speaker may (carelessly?) use a WT conjunction. There are four instances of the WT conjunction *ja* 'and, so, then' breaking up a stretch of Malay; three of these are shown in (196)–(197). In the analysis as presented up to this point, each of these would be considered an instance of complete-S CS followed closely by an instance of Near-S CS.

(196) 1\SR ***«Kamorang ini [.] paki sandál par apa?»*** *[.]* ***«Seng, katong*** ^ ***(aju-aju paki sandál.)***
"You (PL) [.] wore the sandals for what?" [.]
"No, we ^ (acting like wearing sandals.)"

2\YL ^ ***«DF Barang JK Bapa-guru Y lapás.»***
^ "Because Teacher took them off."

3\SR ***«DF Barang JK Bapa-guru Y tara-paki,*** *ja* ***Y katong Y ada Y paki aju-aju Y orang Y bajalang.»***
«Mo Y kamong Y tatawa Y apa?»
«Y Seng, DF barang [katong ada latih-] JK Bapa-guru Y ada Y kasi Y katong G pekerjaan•rumah, *(to imala), ja* ***Y katong Y ada [e] Y tatawa.»***
"Because Teacher wasn't wearing them, so we wore them acting like people walking."
"But what were you (PL) laughing at?"
"No, because [we practice-] Teacher gave us homework, (or dictation), so we were laughing."

(197) 1\YL ***DF Tapi*** *kenjou* ***NF macang*** *dajaman* ***P Mesak, P Mesak*** *inamon;* ***«N Sudah, Y tanya P Tero.***
But if like they ask Mesak, Mesak says, "Enough. Ask Tero."

2\LL ^ *(Tero.)*
^ (Tero.)

3\YL ^ ***YF Kalau P Tero Y bilang Y betul, Y ya Y betul. W Salah, Y ya W salah.*** *[.] Ja* ***Y dengar Y dari Y si•perempuan.***
^ If Tero says it's true, then it's true. False, then false. [.] So hear it from the female.

For these I propose that what I have counted as two instances of CS in table 5.1 are actually a single instance with a WT conjunction inserted. Similarly, in (198) the final three-word sentence could be analyzed as an instance of Malay mode with a WT conjunction inserted.[148]

[148] Note that since a WT adult is speaking to a WT child here, this is an example of addressee-related CS.

(198) \WG *(Ken)* ***[Y]manangis [Y]par [Y]apa?*** *[.]* ***[R]Manangis*** *ja* ***[Y]bicara!***
(You) are crying for what [.] If you cry, then speak! [3]

As noted in chapter 2, the conjunction *mo* has the same meaning and function in WT and DM. So in data like (199), rather than analyzing *mo* as an instance of an inserted WT conjunction, I consider *mo* to be a member of DM at this point.

(199) \OK *Inanga* ***«[X]Ado:! [Y]Dong [Y]ada [A]asik-asik [N]cerita, mo [Y]liha [:] [P]Lenora [Y]deng [P]Bertha [Y]sana [Y]dong [Y]adadi [:] [P]Tingkat• Jepang [Y]sana.»***
He said "Oh no! They are busy talking, but look at [:] Lenora and Bertha over there, they are on the Japanese floor over there."

However, in (200) the dual-language status of *mo* apparently causes it to function as a neutral site (Clyne 1967), triggering the occurrence of the WT agreement prefix *i-* '3s' on the following verb. Aside from that one WT morpheme, the entire quotation is clearly in Malay mode.

(200) \YL *Eya. [.] «Mo* ***[C]motor•IDT?»*** ***«[P]Bu [Y]ada [Y]bawa [C]motor, [DF]tapi [e] [C]motor [NF]ada [a] [C]papan [Y#]satu [Y]talapás, [DF]jadi [Y]ada [Y]istirahat, mo*** *%[Y]i-*** ***cari [Y]orang [YF]par [Y]pi [N]kerja.»***
Yes. [.] "And the IDT boat?" "He brought the boat, but the boat has [uh] a board loose, so they're resting, and he's looking for a person to go work on it."

If we reanalyze the above instances of Near-S CS as being actually Complete-S CS with a WT conjunction inserted, then the already small category of Near-S CS diminishes from nine to only five instances. Will the WT-insertion analysis be valid for those five as well? Consider the conversation reported by SR in (201). In (b)–(d), the speaker reports a conversation she had as a girl with another WT-speaker. It is entirely WT, except for one necessary Malay kinterm. Then in (f)–(p), the two hold a conversation with a Malay speaker; for the most part, the reported conversation is in Malay, as the actual conversation would have been. But note that in (h), (l), and (m), the WT conjunction *ja* 'and, so, then' occurs, and in (j) the WT enclitic *= si* 'already' occurs. Turns (j), (l), and (m) are three of the five remaining instances of Near-S CS, and it seems that indeed these should be interpreted as Malay mode sentences with high-frequency WT elements carelessly inserted.

(201) a. *(Kama satu kali) mai balár. Ok tora [ba-] Ganái•jinai. Mai [a] in,* ***[P]Oka*** *ká idá balár. Êra temun* ***[D]talalu*** *dengal. Kama ina ka= inoba ina, mo malaba =i ka makáká pênin. Barakala martom* ***[K]Bapa*** *nêr.*
(One time we) went to the garden. I with [.] Ganái's mom. We went to [uh] Oka and his group's garden. They had a lot of

cucumbers. We each carried one, and cut them open and ate half. In a little while we met Father.

b. [DF]***Tarús*** *ok kom•on, «Ah! On sêta puida.»*
Then I said, "Oh! We're dead."

c. *Nei nam•on, «Aka•lebá?»*
She said, "Why?"

d. *«Mo morjou moi jala mona,* [K]***Bapa****.»*
"Just look over there in front, it's Father."

e. *Ganái•jinai etabei (uk nekanei nêr mo) ipo pênai êr, ka [:]* [A]***pura-pura*** *aka erajítjít,*
Ganái's mom threw away (hers and) carried the half and then [:] pretended to urinate,

f. *(tarús Bapa) nam•on* ***«(Mimi =ei,) ko ada biking apa di•situ?»***
(then Father) said "Hey Mimi, what are you doing there?"

g. ***«Tete =ei, beta ada kincing.»*** *Mo nei ne= ikáká.*
"Oh Grampa, I'm urinating." But she was eating it.

h. ***«Eh,*** [Y]***ko*** [Y]***kincing*** *ja {laughing}* [Y]***bagitu, eh?»*** *{laughs}*
"Hey, you urinate {laughing} like that, do you?" {laughs}

i. *(la la aka)* [K]***Bapa*** *ei balár, «Ah!* [X]***Tuhan•Allah****. Temun ir lasai ma puiai.» Nei elalú=na. «Ma minal temun jan (=e)?»*
(then) Father went to the garden, "Oh! God. Three cucumbers are gone." He came down. "Did you take some cucumbers?"

j. ***«(Oh)*** [Y]***seng.*** [K]***Bapa.*** [Y]***Tadi*** [Y]***katong*** [Y]***tinggal*** [Y]***situ mo*** [Y]***pateka (***[Y]***itu*** [Y]***tadi)*** [Y#]***tiga*** [Y]***itu*** *=si.»*
"(Oh,) no, Father. We were there, and those three cucumbers were there."

k. ***«***[Y]***Seng.*** [Y#]***Tiga*** [Y]***hilang.»***
"No. Three are gone."

l. *«Ja* [Y]***tahu!*** [Y]***Mangkali*** [Y]***katong*** [Y]***turung, mo*** [NF]***ada*** [Y]***orang*** [Y]***jalang*** [Y]***di*** [D]***balakang*** [Y]***kah?»***
"Who knows? Maybe we came down here, and somebody was walking behind us?"

m. *«Ja* ***Yitu*** [Y]***ada•baru*** [Y]***petik*** [Y]***akang,*** [Y]***dia*** [Y]***pu [e apa ini ada baru keluar.]*** [Y]***air*** [Y]***baru*** [Y]***keluar.»***
"Somebody just now picked it, its [uh what is it just came out.] water just came out."

n. *«Ja êsá,* [K]***Bapa****.»*
"Who knows, Father."

o. ***«Mo*** [Y]***sapa*** [Y]***ada*** [Y]***angka?»***
"But who picked it up?"

p. *«[Y] **Tadi** [K] **Bapa** [Y] **baku-dapa** [Y] **katong,** [Y] **katong** [Y] **seng** [Y] **bawa** [Y] **apa-apa,** [NF] **to?»** Mo kam' martamarer mam keikei.*
"When Father met us, we weren't carrying anything, you know?" But we were standing as still as trees….

In (202), a reported quote is entirely Malay for the first two clauses, while the final clause is introduced by a Malay conjunction and ends with a Malay word.

(202) \YL …[DF]***tarús*** *nei nam•on;* ***«[P]Ampi [Y]tinggal [NF]sudah. Mo [K]Mama [Y]ada [.] [Y]seng•sanáng [Y]deng [P]Ati,*** [DF]***jadi*** *[e] ken me* [NF]***sudah.»***
…then she said, "Ampi, you stay. Mama is not happy with Ati, so [uh] you just stay.

In the lexical analysis presented in chapter 4, the last two Malay words in (202) were counted as lone Malay lexical items. It seems to me more realistic, however, to say that the entire reported quote is in Malay mode with a brief lapse into WT toward the end. Note the pause-filler [e] just before the two WT words, possibly an indication of a lexical retrieval difficulty. At any rate, there are very few examples like this in the corpus, so this kind of reanalysis would not greatly affect the lexical analysis of chapter 4.

The two remaining examples of Near-S CS are also amenable to reanalysis as Complete-S with an inadvertent slip into WT. They are shown in (203) and (204). In both examples, a single high-frequency WT morpheme occurs: the emphatic adverb *kai* in (203), and the enclitic *=e* in (204), which marks a yes-no question. Neither of these have precise equivalents in Malay. Note also that (204) is part of a reported dialogue in which one speaker uses only WT, while the other uses only Malay except for this one insertion.

(203) *Kai{EMPH}* ***[Y]dia [Y]pu [K]om [Y]dong [Y]semua [Y]setuju.***
His uncle and those with him all agreed!

(204) *([DF]Tarús [P]Pai) nam•on;* ***«[Y]Seng, [P]Usi. [Y]Katong [NF]ada ([Y]datang [Y]cek [Y]par e) [P]Usi [Y]tadi [Y]bawa [Y]jualan?»***
(Then Pai) said, "No, big sister. We have (come to check if uh) big sister brought merchandise?"
Ok kom•on, «Eya. [P]Usi ma= ipo =yai. Ir in.»
I said, "Yes. Big sister brought them. Here they are."
«([Y]La) [Y]kamong [Y#]dua [Y]beso [a] [Y]jual *=e?»*
"(So) the two of you tomorrow [uh] will sell?"
One.
That [is what he said].

Pai inam•on; ***«Oh! [P]Usi = ei, ([NF]lebái) [Y]nanti [Y]beso (da) [Y]bawa [C]kangkung. (mo [Y]beta [Y]nanti) [Y]datang [Y]ambil.»***
Pai said, "Oh! Big sister, (it would be better if) later tomorrow (you) bring watercress. (then I will) come get it."
«(Ja mom). Loloar ja maka.»
"(So do it.) Tomorrow, come for it."

It may be that these inserted conjunctions function sociolinguistically as ethnicity markers.

In contrast to the examples above is (205) where AD reports a conversation in which she inserted a WT word (*sokat* 'interrupt') while rebuking a non-WT speaker. This does not actually count as an instance of WT insertion in the corpus, since the insertion occurred in the original conversation being reported, and the direct quote here is a precise report of the original wording. Apparently, AD inserted the WT word in order to rhetorically overpower the non-WT speaker.

(205)
a. *nei nanaka lebá•ia,* ***[DF]tarús*** *[.] nen,* ***[P]Warkula*** *ma etalêlênga êr, ko= isokat. [.]*
he said something, then this Warkula heard, and he interrupted. [.]

b. *Ka= inam•on [.] [inamba êr?]* ***«Oh, [itu:] [Y]itu [Y]dong [Y]itu [Y]su [Y]dari [N]dulu.»***
He said [.] [what did he say?] "Oh, [that] that, they are from beforehand."

c. ***[DF]Tarús****, [.] Dumgair nam•on, «****[X]Co****, ja ken moraka on kanei* ***[N]tujuan*** *=e?»*
Then, [.] Dumgair said, "Ha! You know the intent of this, do you?"

d. *Nei nam•on,* ***«[YF]La [Y]dong [Y]bilang [Y]apa?»***
He said, "So what did they say?"

e. *Nei nam•on, «E::h! Mo morjaman peda.» [.]*
He said, "Hey! Ask first." [.]

f. ***[DF]Tarús*** *ok kom•on «****[Y]Jangan*** *sokat.» {laughs}*
Then I said, "Don't interrupt." {laughs}

g. *{laughing}****[DF]Tarús*** *(nei nam•on)* ***«(Bilang) [R]jangan•sokat [Y]itu [Y]apa?»*** *{laughs}*
{laughing} Then (he said) "What does (saying) *jangan sokat* mean?" {laughs}

h. *Ok kom•on* ***«Eh, [Y]tanya [Y]dulu. [Y]Kata «[Y]Dong [Y]ada [Y]bicara [Y]apa?»*** *Mo ken o= [%Y]m-ja-****jawab*** *ko (mom•on)* ***[R]dari•dulu*** *aka•lebá?»*
I said, "Hey, ask first." Say "What are they saying?" But you answer (saying) "from beforehand for what?"

i. *Nei nam•on, «Eh, êsá, mo [.] kem o= minanaka lebá?»*
He said, "Oh, who knows, but [.] what are you (PL) talking about?"

This humorous WT insertion is in contrast to the occurrence of the WT word *Tarangan* which occurs within both WT mode and Malay mode in this corpus. *Tarangan*, as the name of an island, a language, and an ethnic group, is an established loan in Dobo Malay. In this corpus it mainly occurs in the place name *Kampung Tarangan*, which is a section of Dobo town.

As mentioned in §5.2.3, sometimes a change of language mode (especially a switch to the language of a politically dominant culture) has been noted to cooccur with a speaker's anger. There is a strange kind of incomplete language mode switching in (206), while the speaker is angry and very animated. In this example, we find the WT nouns *rataún* 'sago paste' and *kataler* 'vegetables' inserted.

(206) \LL *Ela ka ekatút. Ekatút,* ^{DF}***tarús*** *nei inam•on «*^{K}***Mama*** *=ei,* ^{C}***doit*** *ir* ^{NF}***cuma*** $^{C\#}$***tiga•ribu*** *on.» *^{K}***Mama*** *namon «Eya,* $^{C\#}$***tiga•ribu*** *no aka* ^{C}***uang•jajan.»*** *Ok kom•on «*^{Y}***Demi•nama•Tuhan,*** ^{AF}***eh?*** *Aka ok na=* $^{\%N}$*k-****dapa*** ^{C}***doit,*** *melakala na= ukêl manám tan bôt on =e?* ^{Y}***Ada*** ^{Y}***makang*** *rataún* ^{Y}***saja*** ^{Z}***torana*** ^{Y}***deng*** *[e] kataler.* ***Mo*** ^{Y}***pi*** ^{Y}***balí*** ^{Y}***pakiang—***^{D}***mou*** ^{Y}***buru*** ^{D}***mou*** ^{Y}***mati,*** ^{Y}***kah?*** ^{D}***Mou*** ^{Y}***mati,*** ^{Y}***bilang.»***
She went and stored it. She stored it, then she said, "Hey Mama, the change is just 3,000." Mama said, "Yes, that 3,000 is for snack money." I said, "[I swear] by the name of God! When I get some money, do you think I'll buy food for this house? We're eating just sago paste with vegetables. But you go buy clothes—you're in a hurry to die, are you? You wanna die, just say so."

Finally, in (207), YL is telling a story from her childhood about an altercation with the village schoolteacher. In turn 4, SR uses the mixed word *eparikriksa*, derived from the default Malay verb root *pariksa* 'inspect' and integrated into WT by the addition of the agreement prefix *e-* (3s) and by reduplication, which in WTA marks the scope of negation. In turn 5, YL begins by echoing SR's contribution in turn 4, modifying it in the direction of Malay; she then produces a two-word Malay sentence, the second word of which is gratuitous Malay.

(207) 1\YL {laughing} *Ameli nêr!* ***«***^{W}***Salah!*** ^{W}***Salah!»*** ^{Y}***Semua Ckelas*** *abil* ^{W}*salah, sakali ina* ^{NF}***yang*** ^{N}***jawaban*** $^{\%Y}$***ba****túl****túl*** *ia.*
{laughing} That Ameli! "Wrong! Wrong!" The whole class was wrong, not a single person had a correct answer.

2\SR ***NF*Padahál**, *nei kali joujou =di.*
However, he didn't look at them.
3\YL ^ ***NF*Padahál**,
^ However,
4\SR ^ *(Sakali)* *%N*e-**parik*riksa** *=di!*
^ He (didn't) inspect them!
5\YL ***Y*Sondor** *%R*e-**pariksa!** ***NF*Cuma** ***Y*lihat**.
He didn't inspect [them]! [He] only looked [at them].

The only element in turn 5 which is etymologically WT is the verb agreement prefix *e-* '3s', which is merely an echo of its occurrence in SR's mixed word in the previous turn. There are three differences between the repetition and the original worth noting, differences which make the sentence more Malay-like than the original: first, WT *sakali* 'not' is replaced by gratuitous Malay *sondor*. Second, the repetition of *pariksa* is not reduplicated, as would be required by WTA syntax. Third, the WT 3pn enclitic *=di* is missing, even though turns 2 and 4 make it clear that the object is plural. Malay allows a zero inanimate plural object, but WT does not. Therefore, despite the occurrence of the echoed *e-*, the morphosyntactic frame of this sentence seems to be Malay. In contrast to my previous decision to exclude mixed words from the seven form categories, I consider the whole of turn 5 to be a two-sentence stretch of Malay mode[149] with a (modified) WT echo word inserted. Under this analysis, SR integrates the default Malay *pariksa* 'inspect' into WT by means of the agreement affix *e-* (3s) and reduplication; then YL partially re-integrates *eparikriksa* into Malay by the omission of WT-style reduplication.

Thus, many examples of WT lexical items (usually conjunctions) inserted into Malay mode occur in this corpus. I conclude, therefore, that inserting EL conjunctions (both Malay into WT and WT into Malay) is a characteristic of LCP in this community. In §5.4.4 I consider various types of difficulties that arise from the bilingual nature of this interaction.

5.4.4 Interference

Ever since Weinreich's ground-breaking study (1953), numerous LCP researchers have used the term interference, with various definitions. Poplack (1987:72) used the term for "speech errors which involve elements of both languages". This limits the term to a speaker's **production**; in the following discussion, I use the term interference both in Poplack's

[149] The conversational function of the Malay mode could be emphasis, as discussed in §5.2.3.

sense, as well as in the sense of a second language interfering in one speaker's **interpretation** of another.

Misinterpretations. Although speakers usually profit from being able to draw on more than one language in a bilingual interaction, sometimes confusion arises. In (208), for example, YL uses the Malay word *gara-gara* 'just because of', but AD mistakes it as the WT word *garagara* 'born'.

(208) 1\YL ᴷ***Mama*** *na* ᴺ***pulang*** *pei balár, kama bôt olat, mo nei na ersirpei aka* ᴷ***Mama,*** ʸ***saling*** *nam apúk. Nei ersir(pei) aka* ᴷ***Mama,*** *inam [da-] [er]* ᴾ***Monika*** *naka nei* ᴺ***gara-gara*** *êr.* /ᴺᶠ***Padahál•*** *mo nei.*
Mama then came home from the garden, our house was empty, but she then told Mama, reported it otherwise. She told Mama that [.] Monika told her because of that." However, it was her.

2\AD ᴾ***Monika*** *inaka nei ko= igaragara =e?*
Monika told her that she was born?

3\YL *Nei (ma=) inaka nei* ᴺ***gara-gara,*** *aka inam* ᴾ***Monika*** *na= inaka nei ka= inam•on, [a] «*ᴾ***Usi*** *ká, meimei balár, (e) mipo sensena pop no.»*
She told her the cause, she said Monika told her, [uh] "You ladies, you go to the garden, (uh) you take the whole pig with you."

In (209), WG reports a conversation she had with a non-WT speaker. In her turn 1, she reports five complete turns of the conversation and the beginning of a sixth; all but the second reported turn and the beginning of the sixth are in Malay. FG interrupts WG, offering turn 2 in Malay. But as WG continues in Malay in turn 3, FG assumes that WG is correcting her and using Malay outside a direct quote, when in fact WG is merely continuing the direct quote of turn 1 which FG had interrupted. In turn 4, FG mistakenly corrects herself; in turn 5 it is clear what WG meant as she finishes the direct quote.

(209) 1\WG *Eya, (nei nam•on)* ***«***ʸ***Mana*** ᴾᴷ***Adi•Joli?»*** ᴰᶠ***Tarús*** *ok kom•on «Oh, nena natapen.»* ***«***ʸ***Su=*** ʸ***ana*** ʸ#***berapa?»*** *Ok kom•on* ***«***ʸ***Su=*** ʸ***ana*** ʸ#***tiga.»*** *[.]* ***«Eh,*** ᴺᶠ***padahál*** ʸ***itu•***ʸ***hari*** ʸ***pi, mo*** *[ma:-]* ʸ***baru [e]*** ʸ***ini.»*** ᴰᶠ***Tarús*** *ok kom•on,* ^ *«Eya.*
Yes, (she said) "Where's little Joli?" Then I said, "Oh, she's in the village." "How many children now?" I said "Already three children." [.] "Hey, however that day when they left, there was just this one." Then I said, ^ "Yes.

2\FG ^ ***ʏBaru ʏ#satu•ʏbulan ʏkah.***
^ Just a month old or something.
3\WG ***ʏSu= ʏ#tiga.***
Already three.
4\FG ***Eh, ʏ#tiga*** ^ ***ʏbulan.***
Oh, three ^ months.
5\WG ^ ***ʏLaki-laki [e] ʏ#satu, ʏdeng ʏperempuan ʏ#dua, D#Yang• pertama ʏitu ʏperempuan.»***
One boy and two girls. That first one was a girl."

Performance errors. Bilinguals, of course, make performance errors just as monolinguals do; and some of these performance errors are related to the management of the two (or more) languages involved. As already seen in chapter 4, false starts occur throughout the corpus, and many of them involve replacing a Malay lexical item with an equivalent WT lexical item. Sometimes, however, performance errors are not repaired. In (210) AD struggles to find the proper way to morphologically integrate the Malay verb *cerita* 'tell a story'; the form she settles on was rejected as ungrammatical by all the interviewees, including AD herself.

(210) \AD ***ʏBagimana****? [mo mor-] Mo [mo-]* ***%Nmo-r-cerita*** *ton* ***NFbisa*** *[:]*
How? [please 2s-] Please [2s-] tell a story so that it can [:]

In every other instance in the corpus, *cerita* as a verb is either a bare form or else integrated by using the WT verb *-m* 'do'.

Another type of performance error involves language choice. Clyne (1967:17, 84) demonstrated how the presence of a neutral site (i.e., a word or string of words which is the same in both languages) can cause the speaker to lose track of which language he was speaking, and inadvertently switch from one language to the other. Clyne referred to this as TRIGGERING; to avoid confusion with my use of the term TRIGGERED SEQUENCES, I will refer to this phenomenon as TRIGGERED LANGUAGE SWITCHING.[150]

There are no clear instances of triggered language switching from WT to Malay in this corpus; all possible instances involve a very brief stretch

[150] Triggered language switching is similar to monolingual performance errors such as the following, which I heard within a week of each other while writing this chapter: (a) Kris Nivens, Sept.10 1997: *Do you want me to close the door open?* and (b) Collin Seitz, Sept.14 1997: *laying [the foundwork] [.] the foundation.* I propose that in (a) Kris had two functionally equivalent sentences in his mind: *Do you want me to close the door (now)?* and *Do you want me to leave the door open?* These have the same intonation contour, and the NP *the door* was a neutral site, allowing a mid-sentence switch from one sentence to the other. In (b) Collin was thinking simultaneously of the synonyms *foundation* and *groundwork*, both of which collocate with *laying*, and both of which contain [awnd]. This string of phonemes constituted a neutral site, permitting a mid-word switch from one word to the other.

of Malay mode, as in (211). Here, YL answers LL's question about kinship beginning with the Malay kin term *adi* and continuing in Malay to the end of the NP, even though the remainder of the phrase could easily have been said in WT.

(211) 1\LL *Ja [:]* ***[P]Isak*** *nono,* ***[P]Usi•Nona*** *ká idá [:]*
So [:] that Isak is Ms. Nona and her group's [:]
2\YL ***[K]adi [NF]yang [Y]bongso*** ^ ***[Y]sekali.***
younger sibling which is the very youngest.
3\LL ^ ***[R]adi•yang•bongso•[ska-]•sekali [NF]yang*** *datôr nei damdam* ***[PK]Tete•Dobo*** *nêr, nene =sin =e?*
younger sibling which is the very youngest, the one they call Tete Dobo, is that the one?

Note that LL does not follow YL's choice of mode, even though she does repeat the phrase itself.

Another possible example was seen in the first turn of (139), where LL began a turn with two clauses that contained no non-default Malay items, then after the conjunction *mo* produced a clause which was completely Malay, in which two of the words were non-default Malay. As mentioned previously, *mo* is a word with dual language membership and, therefore, constitutes a neutral site which may trigger a language mode switch. Since *mo* was apparently treated by the speaker as a Malay conjunction here, this clause was counted as Complete-S CS. Example (193) may be another instance of non-default Malay triggered by *mo*.

Triggered language switches from Malay to WT, however, do occur in this corpus. In (212), AD reports a conversation between herself and a non-WT speaker, alternating between WT mode and Malay mode. While in Malay mode in turn 3, she apparently experiences some difficulty in retrieving the Malay phrase *abu panas* 'hot ashes'. She begins to say *air panas* 'hot water', but interrupts herself and replaces that with the WT phrase *kôrau rararai* 'hot ashes'. She immediately recovers from this language choice error, however, and continues the reported conversation in Malay.

(212) 1\AD *Masí marsirpei ká* ***[PK]Om•Rumahoru*** *nêr, «****[K]Om*** *=oi! [.] Molá!» [.]* ***«[NF]Ada [Y]apa [A]=lah?»*** *{laughing} Nei nam•on!* ***[DF]Tarús*** *ok komon; «****[K]Om*** *=ei,»*
We went and told Uncle Rumahoru, "Hey Uncle! Run!" [.] "What's up?" {laughing} He said that! Then I said, "Hey Uncle,"
2\LL ***[D]Kanál*** *ekanei ^ jersir êr min.*
Encounter his ^ speaking too.

3\AD ^ *Ah.* ***«[P]Oka [Y]itu [Y]su=*** ***[Y]kanál [air pana-*** *e a on,] kôrau rararai.»* ***[.] «[Y]Sapa•yang [e] [Y]siram [Y]dia [Y]deng [e] [Y]abu [Y]panas [Y]itu?»*** *«Oah! Mortêr. Ko [.] ken na moi ne, ja=* [%YN]*m-**dapa•penjelasan.***
^ Yeah. [.] "Oka got burned by [hot water- uh, I mean,] hot ashes." [.] "Who poured those hot ashes on her?" "Ah! Walk. When you go there, you'll get an explanation."

In (213), YL switches briefly into Malay mode (subclausal CS consisting of three consecutive non-default Malay words) after a string of three default Malay words; this is possibly an instance of triggered language switching from WT to Malay. The brief Malay mode ends with another default Malay word *odol* 'toothpaste', which apparently triggers a language switch back to WT for the final word of the sentence, *on* 'this'.

(213) \YL */[P]**Usi**•Kabal toratora [:] [.] [on a] /[P]**Usi**•Iben. Êra dai bôt, dasí, [DF]**barang [N]dolo [X]kasiang.** **[Y]Sapa [Y]mau [Y]paki** [C]**odol** on.*
Ms. Kabal with [:] [.] [uh] Ms. Iben. They went home, they went there, because back then, too bad. Who would use toothpaste.

In (214) LL is quoting her own conversation with a non-WT speaker from Kei; the original conversation, then, would have been in Malay. I believe there are two performance errors in this example. First, LL uses the WT pronoun *ken* (2s) as the subject of the second sentence in the quote rather than the Malay pronoun *ko*. That same sentence, continuing in turn 3, contains a quote-within-a-quote, in which LL imitates the Kei language. When the quote-within-a-quote ends, LL finishes that sentence in WT rather than Malay, even though it began in Malay.

(214) 1\LL *Ok kom•on, «**[K]Kapolsek ([Y]itu [Y]sa), ([Y]kamong) [Y]pu [Y]orang.*** *Ken* ***[N]bisa***
^***[Y]pi,***
'I said, "(That) police chief is one of (you (PL)). You (SG) can go,"
2\AD ^ *Eh.*
Hey.
3\LL *«**[M]Pelbehe, pelbehe,**» ka jausin!»*
"Pelbehe, pelbehe," and that's it!"

Not all performance errors are equally erroneous. Some, like the erroneous retrieval of *air panas* for *abu panas* in (212), are absolute errors. Others, like the proper way to morphologically integrate *cerita* in (210), or an unusual phonological fortition or lenition, simply violate current norms. Such norms are based on frequency of occurrence; once an error is made, it may be stored in the brains of both speaker and hearers as a possible minority strategy. In particular, younger speakers may not have as strong a sense of the wrongness of an error as those with more experience in the

language. If there is any motivation for making the same error again (e.g., ease of pronunciation), it will be easier the second time, and the more times the same error is made, the less it is seen as an error. One generation's errors can become the next generation's standard.

Hybrids. By hybrids I mean words or strings of words which show characteristics of two languages simultaneously. One example of this is CALQUES, i.e., ML words in an EL structure. There are several possible examples of calques in the corpus. AD twice uses the phrase *galáy* tungguru 'teacher's house', apparently a calque from Malay *rumah guru*, since it follows Malay genitive order rather than WT genitive order (which would be either *tungguru kanei galáy* or, less commonly, *galáy tungguru kanei*).[151] Another calque is seen in (215) where TN imposes the dative-movement construction of Malay onto WT.

(215) \TN «^{P}***Ita*** *=ei môl* *ok* $^{Y\#}$***lima•ribu.***
Ita hey 2s-give 1s five_thousand
"Hey Ita, give me 5,000 (rupiah)."

There are also examples of what one might call HALF-CALQUES, in which Malay words and WT words are combined in a Malay phrase structure. In (216) AD half-calques the Malay phrase *orang mana* to become *orang ba*. Although in other contexts *ba* 'which, where' and *mana* 'which, where' occur in equivalent syntactic contexts, there is no WT phrase structurally equivalent to *orang mana*. The semantic equivalents in WT would be bana ba 'from where?' or the longer *inal bana natapen ba* 'He/she comes from which village?' The semantic equivalent of *orang* in this sense would be *juir* 'native of, person of' which follows the noun, but cannot occur with ba as in **ba juir* 'person from where?' It seems here that *orang* is projecting a syntactic structure into which the WT word *ba* is inserted.

(216) \AD *Nen* ^{P}***Lapono*** $^{/A}$***orang•****ba?*
This Lapono is a native of where?

Similarly, in (217) TN uses a SIM genitive structure with a Malay noun followed by a WT possessor—apparently she resisted the tendency to trigger a Malay pronoun.

(217) \TN ^{C}***surat*** *nêr* ^{NF}***kan*** ^{Y}***paki*** ^{C}***alamat*** *ok.*
that letter, you know, had my address on it.

Note that in both of these the direction of the switch is from Malay to WT, possibly indicating that the speaker did not intend to produce a

[151] The interviewees insisted that *tungguru* is a WT word, even though it is obviously an old loan from Malay *tuan guru* 'master teacher'. For the present work, I consider it a WT word, since the focus is on speakers' perceptions rather than etymologies.

Malay-mode phrase, but once the Malay noun was spoken they had no choice but to use Malay syntax as well. This would typically produce a triggered sequence, but in these two instances the speaker resisted the urge to continue with another Malay word.

Finally, there are other instances of speakers using a mixed phrase in which, although the phrase structure is equivalent in both Malay and WT, the speaker seems to be translating on the fly from a Malay phrase in his mind to the mixed phrase which he actually says. We may call these SEMI-TRANSLATIONS. For example, in (150), repeated here as (218), OK translates the Malay phrase *pintu gerbang* to become *pintu jinjinai*, which confuses AD:

(218) 1\OK *Ja [.] ^{NF}**macang** danga «^{P}**Puncak**», ^{NF}**yang•penting** (dalala tantan kamá no), ja danga [.] «^{P}**Gerbang**». ^{DF}**Barang** [a] ^{Y}**pintu** jinjinai no, one dam ne itom ^{P}**Pintu•Gerbang**.*
So [.] like they say "Puncak", what's important is (they run to our thing), and then they say "Arch." Because [uh] that big door, that they name that "Arch Door".

2\AD ^ /^{R}***Pintu•jinjinai ba?***
What big door?

3\LL ^ ^{R}***Pintu•Gerbang**. (...)*
^ Arch Door. (...)

4\OK *Mo sita ona $^{\%N}$ta-**masu** tal leipapai jinjinai ^{P}**Ence-Ming** kanei (tapaha ada kala no,) te. One ja [.] dam tom ane [:] ^{P}**Termenal• Gerbang**. {laughs}*
Well we enter through the big doorway that is Ence-Ming's (...) of course. That one they name [:] Arch Terminal. {laughs}'

In (219), LL apparently translates the Malay phrase *siap di tempat* 'standing ready' (literally, 'ready in place') to become the mixed phrase *siap ei tempat*. Note that here she uses the High Malay form *tempat*, not the ubiquitous Low Malay form *tampa*, giving the hearer an indication that she is quoting a SIM idiom.

(219) \LL *One tamata ^{NF}**yang** nena [:] inalnal raram nono, nei ^{A}**siap** ei ^{NH}**tempat** gatan aka jernal raram no.*
In that situation, the person who is [:] taking out the bilge water, he is just standing ready to take out the bilge water,

In (220) AD creates the mixed phrase *rasarasa kanonourka* 'you feel hungry', based on the Malay phrase *rasa lapar* 'feel hungry'. In Malay, *lapar* can either stand alone or occur with *rasa*, but in WT *kanoir-* 'hungry' cannot occur with *-perat* 'feel'.

(220) \AD *(Ok on kupipikir!) (nam•on) «Eh! [.] Ken maera on [a] sakali [:]* *%Y**Ø-rasa-**rasa** (kanonourka)?» «Eya. Kali.» Êra damanám.*
(I was thinking!) (he said) "Hey! [.] You today [uh] don't [:] feel (hungry)?" "Yes, I don't." They ate.

In (221), YO manages to say something in one concise mixed phrase that would take more than one phrase to say in either Malay or WT. The phrase is *setiap maera gogongar* 'all day long every day', constructed from (Mixed) *setiap maera* 'each day' and (pure WT) *maera gogongar* 'all day long'. In Malay, both *setiap* 'each' and *sepanjang* 'all...long' would precede the noun *hari* 'day' and could not cooccur, while in WT both *gogongar* 'all...long' and *pelpel* 'each' would follow the noun *maera* 'day', and could not cooccur. By using a mixed phrase, YO is able to put one modifier before the noun and one after, possibly in order to insult his friend more strongly.

(221) \YO *Ken [a] jertir pui, (mo) jeruk* ***C**bedá** *seina. Jeruk* ***C**bedá** *[:],* ***N**setiap** *maera gogongar.*
You [uh] don't bathe at all, (but) do nothing but use powder. Using powder [:], all day long every day.

In chapter 2, I described the Malay causative construction with *kas(i)*. Of sixty-three instances of *kas(i)* with a causative meaning, forty-eight are morphologically integrated by the addition of a WT verb agreement prefix, as in (222) and (223).

(222) \AD *dasí* *%YN**da-**kas•turung** **manara** ei* *ne,*
3p.go 3p.CAUS_descend stuff 3s.at that
they went and put down the stuff there,

(223) \AD *ma =* *%YN**i-**kas•pula**pulang** **C**tabaku* *on,*
REC 3s.CAUS_DUP.go_home tobacco this
she put this tobacco back a while ago,

However, there are four instances in the corpus where YL morphologically integrates a Malay causative phrase with *kas* by inflecting both *kas* and the main verb with a WT subject agreement prefix as in (224)–(226).

(224) \YL ***P**Usi* *= ei,* *%YN**da-**kas•da-**pulang** = na* *gatan.*
lady hey 3p.CAUS_3p-go.home = 3s just
Hey lady, they just sent him home.

(225) \YL *%YN**Da-**kas•da-pula-**pulang** = na* *te.*
3p-CAUS_3p.DUP-go.home = 3s of.course
They sent him home, of course.

(226) \YL *kam' maakai nôr galgal. [.]* %Y*Ma-**kas**•ma-**turung** sena, [.] (dir) mapo =di mai Obanat ne.*
we climbed for young coconuts. [.] We finished bringing them down, [.] (...) we took them over there to Obanat's.

Only two verbs are involved, *pulang* 'go home' and *turung* 'go down'. No other speaker uses such a morphological pattern with *kas(i)*. There is, however, a serial verb construction in WT in which two adjacent verbs are inflected with subject agreement prefixes and share a single object NP as in (227).

(227)

da-nar	*da-siar*	*anam*	*êr*
3p-hit	3p-break	door	DEF.SG.INAN

they broke through the door

Apparently, YL was attempting to morphologically integrate these phrasal entries with *kas(i)* by using this WT serial verb pattern. Again, there is here an odd kind of morphological integration which is apparently not a performance error, since the same speaker persists in using it. But other speakers may consider it a competence flaw, since they never use it themselves.

It is possible that hybrids should be considered a special type of performance error, i.e., a corruption of WT. On the other hand, it may be more realistic (i.e., closer to the speakers' own perspective) to consider some of them to be the output of innovative bilingual speakers who are drawing on the syntactic resources of Malay in order to enrich WT.

Doubling. Myers-Scotton (1993b:132) discussed instances of double morphology[152] and attempted to demonstrate that they do not constitute a problem for her model of CS. She also stated (p. 111): "Double morphology turns up in CS corpora across a number of language pairs. Double-plural affixes are the main exemplars."[153] In all instances of double morphology seen to date, the head is an EL item.

Myers-Scotton (1993b:112) presented data from a Quechua/Spanish corpus which demonstrates that double plural morphology can occur

[152] For example, a Shona prefix which marks plurality (along with noun class) on an English stem containing plural *-s* (Myers Scotton 1993b:61); or a Lingala infinitival prefix on a French stem which contains an infinitival suffix (Bokamba 1988:37). Note that even if the meaning of an affix in one language only partially overlaps the meaning of an affix in the other language, Myers-Scotton still counts the occurrence of both on the same word as double morphology.

[153] She cited Turkish/Dutch (Backus 1990:4), Maori/English (Eliasson 1991:19–20), Lingala and Chiluba/French (Kamwangamalu 1990:5), Mexicano (Nahuatl)/Spanish (Hill and Hill 1986:165), and Quechua/ Spanish (Appel and Muysken 1987:172–173), as well as "all Bantu languages for which CS has been studied".

even if both elements are suffixes.[154] However, I predict that in language pairs where both morphological elements are prefixes or both are suffixes, such double morphology is less likely to occur. The WT/Malay corpus may support this hypothesis. DM *baku-* is a productive transitive verb prefix indicating reciprocal action; it is functionally equivalent to the WT verb structure shown in (228), where a reflexive or reciprocal meaning is indicated by the prefix *r-* and the use of coindexed verb agreement prefixes and suffixes.

(228) *da-r-nar-ai*
3p-R-hit-3pa
they hit each other

Although there are twenty-two instances of *baku-* attached to a Malay[155] transitive verb root in the corpus, and many more instances of the reciprocal WT verb structure (forty-six with a WT verb root and nine with a Malay verb root), they never cooccur.

(229) *da-r-baku-pukul-ai*
3p-r-RECIP-hit-3pa
they hit each other

In contrast, there are many instances of double marking of plurality in the corpus. As noted in chapter 2, there is little productive morphology in DM. The form *baku-* is one productive morphological element; reduplication for marking plurality[156] is another. But plurality in WT is not marked by noun morphology but by determiners as in (230), or by a separate word (jan in WTA, *jakon* in WTB) indicating plurality,[157] as in (231), or both as in (232).

(230) \ES *da-repa* *da-sena* *kei* *ir* *da-pei.*
3p-chop_down 3p-complete wood/tree DEF.PL.AN 3p-from
they chopped down all the trees from there.

(231) \BD *sakali* *ikáká* *nôr* *jan?*
not DUP.3s.eat coconut PL
he doesn't eat coconuts?

[154] The word is *polisiya-s-kuna* 'police-PL-PL'. Myers-Scotton also mentions that double suffixing has been observed in other Central and South American languages. Blust (personal communication) has suggested that a similar process may be behind the Dutch form *schoenen* 'shoe- PL', in which the medial /n/ is a historically fused plural suffix.

[155] There are no instances of *baku-* attached to a WT verb root, which supports the claim that WT is the matrix language wherever the two languages are mixed.

[156] As noted in chapter 2, Malay noun reduplication sometimes indicates not only plurality but also variety. (Malay reduplication has other functions as well.)

[157] In WTA, *jan* occurs after common nouns and is in a paradigmatic relationship with *ka* 'and others' which occurs after proper nouns.

(232) \YL *Itora [ken- e] kekanám marua jan ir*
3s.accompany [2s- uh] POSS.2s in.law PL DEF.PL.AN
da-ela =sin.
3p-go =already
She went away with your sisters-in-law.

Thus, a double marking of plurality in the WT/Malay corpus does not fit Myers-Scotton's concept of morphological doubling; even so, it is still the most frequent type of doubling in the corpus. Perhaps, then, it is not plural morphology per se which is so commonly doubled in bilingual corpora as the plural feature itself, no matter how it is marked. At any rate, when a double marking of plurality occurs in the WT/Malay corpus, a Malay noun (or in one instance, adjective) is reduplicated to indicate plurality and is followed either by the WT plural marker as in (233), or by a WT determiner which also indicates plurality as in (234) and (235).

(233) \OK *Kanáng* $^{D+}$***tamáng-tamáng*** *jakon.*
my DUP-friend PL
My friends.

(234) \YO *Aroka da-rtêrtêr* $^{C+}$***toko-toko*** *din*
soon 3p-R.DUP.walk DUP-store these.INAN
Soon they'll walk from store to store here (literally, 'walk these stores')

(235) \YL *Kama* $^{D+}$***tamáng-tamáng*** *ir da-rteya*
1pe DUP-friend DEF.PL.AN 3p-R.marry
da-sena.
3p-completely
Our friends are all married.

Table 5.4 enumerates the various kinds of plural doubling. In contrast to these are twenty-two Malay nouns reduplicated to indicate plurality which are not included in any of the CS form categories (i.e., lone Malay nouns) and which are not followed by a WT indicator of plurality, as well as the vast majority of plural Malay nouns in the corpus for which plurality is indicated only by the WT determiner and not by reduplication. Finally, there are fourteen lone Malay nouns, and four Malay nouns adjacent to another Malay word, which are reduplicated to indicate plurality and are followed by a singular inanimate WT determiner, as in (236).

(236) \AD $^{J+}$***supír-supír*** ^{C}***oto*** *no ena sena ne.*
DUP-driver car that.INAN LOC.SG completely there
all the car drivers were there.

Table 5.4. Double plural marking

Malay	WT	Instances
Reduplicated noun	plural marker *jan ~ jakon*	3
Reduplicated noun	plural demonstrative determiner	16
Reduplicated noun	plural definite article	9
Reduplicated adjective	plural demonstrative determiner	1

The use of a singular inanimate determiner in reference to a group of referents is a WT device for referring to those referents en masse rather than as separate individuals; I consider a WT noun in such a context to be a mass noun. In instances like (236), then, the noun is curiously marked as plural by Malay morphology and as a mass noun by the WT determiner.

There are several other kinds of doubling in the corpus as well. Most involve nonadjacent words or morphemes, and like plurality, all involve closed syntactic classes. The second most frequent type of doubling is the doubling of degree adverbs, which occurs nine times. There are two WT degree adverbs (*dengal* and *uk*), and these do cooccur twice in the corpus (*dengal uk*, cf. English very very), so it is not surprising to find both a WT and a Malay degree adverb in the same clause; there are nine such instances, six of them spoken by OK. Some involve pure WT verbs, as in (237), some have pure Malay verbs, while others have mixed verbs.

(237) \OK *kaní* ^{D} ***sifat*** *no* ^{YF}***paling*** *epir* *uk.*
POSS.3s character that.INAN very good.3s very
his character is very, very good.

(238) \HT *nen* ^{P}***Dono*** ^{YF}***paling*** $^{\%N}$***nakal****-na* *dengal.*
this.ANIM Dono very misbehave-3sa very
this Dono really misbehaves a lot.

OK is the speaker for six of these nine instances, but this does not necessarily indicate a personal preference for doubling per se. Rather, it fits a more general pattern of her tendency to use Malay degree adverbs. OK is the main user of Malay degree adverbs in the corpus, being responsible for nineteen of the twenty-nine instances of *paling*, and twelve of the sixteen instances of *talalu* (when used as a degree adverb; *talalu* has other usages as well).

Perfective aspect is marked in Malay by the preverbal auxiliary *sudah* (or its proclitic form *su=*), and in WT by the clausal enclitic *=si(n)*. In five instances in the corpus, these two cooccur, as in (239).

(239) \LL *[DF]**tapi** êra [Y]**su**= [N]**biasa** =si.*
but 3p already accustomed already
but they are already accustomed to it.

Other examples of nonadjacent doubling include doubled modals (240), adverbs (241),[158] prepositions (242), and semantically equivalent words which are not in the same category in both languages (243).

(240) \LL *[YF]**mungkin** [P]**Beni** sikali eraraka sangil.*
maybe Beni not 3s.DUP.know maybe
maybe Beni doesn't know about it.

(241) \AD *kama [Y]**semua** [%Y]ma-**masu** sena [.] [C]**panitia**.*
1pe all 1pe-enter completely [.] committee
all of us got on the committee.

(242) \LL *[DF]**Jadi** [N]**justru** [YF]**karna** ne [NF]**yang** kama [:] sikali [%Y]mamasu**masu** [P]**Angkatan•Muda** on aka ne.»*
So just because of that we [:] didn't go to Youth Group because of that.

(243) \AD *[YNF]**Termasuk** [J+]**guru-guru** ko da-me bel min.*
including DUP-teacher also 3p-be.at beach also
even the teachers were on the beach.

In all of the nonadjacent doubling examples above, the first element is Malay and the second WT. This is perhaps an indication of the motivation for such doubling, namely that the speaker realizes midway through the sentence that he has used Malay and still has a chance to make up for that blunder by using the equivalent WT word; he may be trying to cover his tracks without using an abrupt repair sequence. In fact, in (244), AD not only doubles the perfective aspect marker, but also after a 3-second pause retranslates into WT (which includes two default Malay roots).

(244) AD *ja nam [a] one [Y]**sudah** [A]**melanggar•hukum** =si. [3] Nekanei [W]**atorang** dir, [%N]da-**langgar** =di =si.*
so like [uh] that, [they] already broke the law. [3] His commandments, they already violated them.

In contrast to the twenty-six instances of nonadjacent doubling where Malay is followed by WT, there are only two instances of nonadjacent doubling where WT is followed by Malay as shown in (245) where both *-akai* and *nai* mean 'climb', and in (246) where both *joba* and *boleh* mean 'it would be good'.

[158] Double marking could also have occurred involving only WT words: the WT equivalent of *semua*, *ma-parong* '1pe-all', can occur along with *sena*, in the same position as *semua* here.

(245) \TN *surúkai dai* ***[Y]bawah•kolong,*** ***[YF]baru*** *daakai* ***[Y]tangga [Y]nai.***
they went in under the house, then they climbed up the ladder.

(246) \BD *Kai joba* ***[NF]ada*** *nôr* ***[NF]boleh.*** *To daká nôr, mo on nôr ka pui!*
If only there were coconut trees. So that we [could] climb for coconuts, but here there are no coconuts!

In addition to the instances of nonadjacent doubling described above, there are a number of instances of adjacent doubling. First, there are several instances of adjacent synonymous WT and Malay conjunctions, usually in the order WT-Malay: *aka karna* 'because' (once), *ka barang* or *ko barang* 'because' (four times), *ka sebab* 'because' (four times). In each of these instances, the Malay conjunction has a narrower semantic range than the WT conjunction. These are pronounced as if they are compounds, i.e., with no intonation break dividing them; see (247)–(254).

(247) \LL *Tamata jabin dagoi* *[/YF]aka•****karna*** ***[D]tuak*** *on =sin.*
Many people have died because of this liquor.

(248) \YL *Makaina uk.* ***[P]Lin,*** ***[NF]mo,*** *ka nena makaina (aka ka),* *[/DF]ko•****barang*** *kama* *[%N]ma-ak-****akal*** *on =sin.*
She was very angry. Even Lin was angry (about it), because we were lying.

(249) \AD *na minaltúk, mo na mitir sopardi.* *[/DF]Ka•****barang*** *êra ma dal* *[W]pasáng damá =si.«*
when you (PL) return, you will bathe in blood. Because they have sent a message here already.

(250) \OK *Sakali* ***[NF]bisa*** *iká sena,* *[/DF]ko•****barang*** ***[C]baskom*** *jinin ne.*
he couldn't eat it all, because it was a big basin full.

(251) \OK *(Tên kunga•on)* ***«Eh, [Y]sabar [Y]dolo.»*** *[/DF]Ko•****barang*** *jau* ***[YF]talalu*** *makainga,* ***[NF]to?***
(Then I said) "Hey, be patient." Because I was already very angry, you know?

(252) \YL *[/YF]Ka•****sebab*** *one* ***[A]talalu*** *min.*
Because that is just too much.

(253) \YL *[/YF]Ka•****sebab*** *konar nono samaina.*
Because that woman is bad.

(254) \YL *Kai ken na= mjaman ok, ok jesang,* *[/YF]ka•****sebab*** *[/Y]kekanám•****diri*** *on (mo)* *[%W]salah-na.»*
If you were to ask me, I'd refuse, because your character is wrong.

There is also one example of the conjunction compound *mo•tapi* 'but', as seen in (255). The first conjunction in the compound can be considered either WT or DM. To my knowledge, this conjunction compound does not occur in DM, so it is likely that this, too, fits the WT + DM compound pattern seen in (247)–(254).

(255) \AD *Nei nam•on, «Ok* [NF]***memang*** *komalúng aka jesabúr aka [:] [rag-]* [Y+]***lagu-lagu•Aru*** *din,* [DF]***tapi*** *[.] «Kumarera• Ngilangila« on, ok koraka (aka) leilei gatan.* [%Y]*Jer-****sambung*** *kokoyaldi gatan.* [/DF]*Mo•****tapi*** [N]***lagu*** *no, kenjou dasabúr, ok min ko=* [%N]*k-****turut****.*
She said, "I really don't know how to sing these Aruese songs, but [.] *Kumarera Ngilangila* I know just a little. I just join in here and there. But that song, if they sing it, I also will join in."

In one conjunction doublet, however, the order is Malay-WT: *macang•dangala* 'like'. This occurs twice, and OK is the source both times (*dangala* is a WTB word); see examples (256) and (257).

(256) \OK *Ok jausi, kulalú êr, ko•[:]•tên kolalú kutan* [/NF]***macang•****dangala [:] erdem* [D]***kolang****,* [NF]***to****?*
I already went down, then I went down to like [:] it made pits, you know?

(257) \OK *kume tapôran, ja sakali dajola =nga,* [NF]***to****? Ko one, ja* [/NF]***macang•****dangala jau ona* [%Z]***sadar****-nga =si.*
I was in the middle, so they didn't see me, you know? Because there, like I was already aware.

There are also several instances of adjacent doubled words which are not conjunctions, including doubled adverbs (258), auxiliaries (259), nouns (260), and prepositions (261).

(258) \TN [P]***Semi*** *i-jaman nei* [Y]***kembali*** *min.*
Semi 3s-ask 3s again again
Semi asked her again.

(259) \OK *Nêr nenta* [Y]***masih*** [D]***sadiki*** [%N]*pikiran-na* [DF]***jadi****.*
3s still still a.little pensive-3sa therefore
She's still a little bit pensive about it therefore.

(260) \YL *nei sur tan on, ja* [Y]***tambús*** [Y]***sabláh*** *pêna-i on,*
3s stab onto this.INAN so go.through side side-3s this
she stabbed her here, so it went through to this side,

(261) \YO *(da-la)* *%DF**da-sam-**sampi*** *jit-jit* *[a]* *[.]* *C**bensin***
(3p-run) 3p-DUP-until DUP-until [uh] [.] gasoline
one.
that.INAN
(they ran) until they reached [uh] [.] that gasoline depot.

Note that in these examples, there is no intonational cue indicating that the speakers were repairing the first word by replacing it with the second.

Again, some instances of doubling (particularly these latter examples) may best be considered performance errors, while others (e.g., the compounding of semantically broad WT conjunctions with semantically narrow or specific Malay conjunctions) may better be considered a strategy of WT speakers to enrich WT by using Malay without completely replacing the WT lexical item.

5.4.5 Is EL syntax a sufficient indicator of EL mode?

In §2.2.3, I described the DM periphrastic passive construction in which the passivized verb is preceded by the verb *dapa* 'get'. Keenan (1985:258–259) pointed out that in languages which have a periphrastic passive with such a verb of reception, "it is common that the modification of the transitive verb takes the form of a nominalization." No doubt this is due to the fact that verbs meaning 'get' typically take a direct object. In Dobo Malay, the passivized verb is not marked as nominalized; it simply occurs as a bare stem as in *dapa pukul* 'get hit', *dapa tikam* 'get stabbed'. Tjia (1992) has proposed that the lack of any overt marking of nominalization on the passivized verb in Ambonese Malay is simply due to the lack of nominalization morphology in that dialect. In line with the general tendency stated by Keenan, in the WT/Malay corpus nearly every time the passivizing *dapa* occurs with a WT verb complement, that verb is nominalized by means of WT morphology—either by the nominalizing prefix *jer-* as in (262), or by reduplication as in (263), or both as in (264).

(262) \OK *Mo* *ne,* *mo* *ok* *jau* *seta*
but that.INAN but 1s already usually
*%YN**ku-**dapa**•**jer-jar* *YF**tarús.***
1s-get_NZR-hit continually
But at that time, I was already usually getting hit continually.

(263) \AD *Kama* *ru-akam* *[na-]* *na* =
1pe two-1pe [FUT-] FUT
*%YN**ma-dapa**dapa**•**[.]•**nar-nar* *pei-pei* *êra.*
1pe-get_[.]_DUP-hit DUP-from 3p
The two of us [will-] will get hit by them.

(264) \YL *Ka da-ela ka %YN da-**dapa**•je-karírír ei*
then 3p-go then 3p-get_NZR-DUP.chase.away at
on
this.INAN
And then they went and got chased away from here,

Assuming that Tjia is correct in stating that the bare verbs in the Malay construction are actually nominalized, there is more evidence here of a word bringing with it from Malay a projected syntactic structure into which WT words are inserted. The only exceptions are two instances of *-dapa tabái* 'get hit', both produced by AD, in which an uninflected WT verb root is apparently a hybrid form created in imitation of the bare Malay verb roots which occur in the Malay construction. In the WT/Malay corpus, Malay verbs passivized with *dapa* are never incorporated by either of the WT nominalizing strategies; they are probably phrasal lexical entries.

The seven form categories described in §5.3 all have one thing in common by definition: they all contain at least one non-default Malay lexical unit. But in addition to these, there are about fifty sequences of **default** Malay lexical items in Malay order **and** in violation of WT order. For example, Malay *harga* 'price, cost' is a recent displacement of the WT noun *pêl*; as a displacement, I consider it to be default Malay in my analysis. *Harga* occurs nine times in the corpus; in six of these instances it occurs in a genitive relationship with a noun representing the item whose price is being referred to, and five of these are default Malay: *harga gula* 'cost of sugar', *harga oto* 'cost of a car ride', *harga sendál* 'cost of sandals', *harga suntik* 'cost of injection', and *harga tiket* 'cost of a ticket'. In each instance, *harga* precedes the accompanying noun following SIM genitive order, rather than following the noun as with *pêl* in WT, which is an inalienably possessed noun.

The Malay word *dekat* 'near' has three senses in this corpus: near in space, near in time, and near in intimacy. In its spatial meaning it is a non-default equivalent of WT *den*; but *den* never occurs in the temporal sense, so I consider this sense of *dekat* to be by default. In both the temporal and locative senses, *dekat* in this corpus occurs both followed by NP and followed by PP.[159] But in WT, *den* cannot be followed by NP; it must be followed by the preposition *aka*. Therefore, in the four instances where temporal *dekat* is immediately followed by a default Malay noun, a string of default Malay items occurs exhibiting Malay syntax rather than WT syntax.

[159] With either the Malay preposition *deng* 'with' or the WT prepositions *aka* 'to' or *pel* 'with'.

The Malay word *satu* 'one' occurs in several senses in this corpus. One of these senses, 'a kind of', which has no WT equivalent, occurs eight times as seen in (265)–(267).

(265) \AD ***[NF]Artinya**, nei nal ne min aka [A]satu [C]ujian aka sêta.*
I mean, he also gives that as a kind of test for us.

(266) \LL *Ok [A]satu [A]manusia, kenjou kujou tamata (darnun-)*
I'm the kind of person, if I see a person (drink-)

(267) \OK *Ane [:] [Y]merupakan [A]satu [NH]pekerjaan.*
That [:] constitutes a kind of work.

Unlike the default Malay *satu* which occurs as a label for times and dates, this default Malay *satu* precedes the noun it modifies. Since its frequency of occurrence indicates that it is an established loan, I conclude that it is a member of a new syntactic subclass in WT. This syntactic phenomenon is analogous to the addition of a new phoneme when loan words are not assimilated, as in /c/ in WT.[160] It is also analogous to the addition of new phonological patterns, e.g., consonant clusters such as nasal-stop sequences, and stress patterns such as mid vowels in final unstressed syllables, which do not occur in WT.[161]

There are two reasons why I have not included such sequences of only default Malay in the list of CS form categories in §5.3. First, because my focus in the present work is on the necessity of treating non-default EL items differently from default EL items, as a prerequisite to the analysis of longer EL stretches. Second, rather than merely assuming that default Malay lexical units in Malay syntactic structures constitute instances of Malay mode, I first consider the more basic question, What is syntax? The answer to the question, "Is Malay syntax sufficient evidence of Malay mode?" depends upon whether speakers create syntactic structures first and then insert lexemes, or choose lexemes first and then let the lexemes themselves construct the sentences. In other words, do syntactic structures have an independent existence in the mind, or is syntax merely a projection of lexemes themselves?[162] If syntactic structures are created before EL words are inserted, then Malay order is a sufficient condition for identifying Malay mode. However, if lexemes project the syntactic structures, then Malay order is a necessary, but not sufficient, condition for identifying Malay mode.

[160] Recall from chapter 2 that previous generations of WT speakers replaced /c/ with /s/.

[161] This is no doubt one of the mechanisms involved in the spread of Sprachbund-type changes: first, a syntactic subclass enters the language along with loan words, then so many loan words are added to that subclass that it competes with the original syntactic subclass, stealing words from that subclass until it is empty.

[162] Of course, I have no a priori reason for disallowing a synthesis of both of these positions, other than an excessive adherence to Occam's Razor.

Half-calques and semi-translations, described in §5.4.4, are word strings in which a Malay word occurs in a syntactic context which is appropriate for its WT equivalent, but not appropriate for the Malay word (as seen in other occurrences of that word in the corpus). For example, Malay *muka* 'front' is semantically equivalent to WT *mona* 'front'; but syntactically, *muka* normally precedes the noun and *mona* follows the noun. Once, however, LL used the phrase *Yala muka* 'in front of Yala', constructing the phrase as if she were using the WT word *mona*. Clearly, in this particular instance LL constructed a WT syntactic structure first and then inserted a Malay lexeme into it. But since there are many instances of *muka* preceding the noun in this corpus, and only one like this, this particular instance may be a performance error.

A similar example is (268), in which AD uses the preposition *aka* after the Malay verb *akui*, as if she had used the semantically equivalent WT verb *-kôt* (which would take a PP complement with *aka*) rather than Malay *akui* (which should take an NP complement). Again, some analysts might use this example as evidence that WT syntactic structures are created first, after which Malay lexemes replace WT lexemes. But since this is the only such example in the WT/Malay corpus, it may be best to consider this a performance error as well-that is, for such instances it might be assumed that a speaker changed his mind about a lexical choice at the last moment, after his first lexical choice had already created the syntactic structure.

(268) \AD *[DF]**rupa** ok [Y]**akui** aka [:] [/P]**Manu**•Benamen. Nekanei [K]**mama** [DF]**biar** [e] [.] [L]**janda**, [DF]**tapi** [NF]**bisa** [Y]**berusaha**,*
like I speak well of [:] Manu Benamen. His mother even though [uh] [.] widow, but she can work,

In contrast to such examples are those where a WT preposition occurs because a preposition is required by a Malay verb, and the equivalent WT verb would not require a preposition, as in the sequence senang aka in (269).

(269) *OK* *Seta danga (erdem) [D]**kolang**, ja [N]**agar-agar** no (no) [YF]**paling** [A]**senang** aka ne. [DF]**Tapi** ok ona [%Y]k-**anggap** ne, ne limlum.*
Usually like (it makes) a pit, then that *agar-agar* is very happy with that. But I thought that was seaweed.

In Malay, *senang* would require the preposition *deng* ~ *dengan* 'with' (or perhaps *akan* 'with regard to'), but the WT equivalent *-soan* takes a NP complement, not a PP complement.[163] Data like (269) could be used as evidence against the hypothesis that sentences are constructed in the ML first, after which EL words are inserted.

[163] Elsewhere, OK uses the WT preposition *pel* with *senang* rather than the preposition *aka*.

In analyzing such data, it is necessary to grapple with the cyclical relationship between competence and performance. Although I do not wish to draw conclusions about communicative competence based on possible performance errors, I do believe that for both the speaker and the hearers, competence is dynamic and is based on frequency of occurrence. As noted in §5.4.4, an innovation that occurs only once may indeed be rejected as a performance error; but if repeated by other speakers, the innovation may eventually become standard practice.

5.4.6 Syntax 2: Locational genitives

As mentioned in chapter 2, the syntaxes of WT and Malay are similar in many places: both are SVO, both have prepositions and preverbal auxiliaries, both have the basic NP structure Noun-Modifier-Determiner. One area of difference, however, is with genitive structures which function to specify location. In Dobo Malay, there are two strategies. The major strategy is to prepose the locational noun to the object whose location is being predicated as in (270), and the minor strategy is to use the possession word *pu* ~ *pung* ~ *punya* after the object noun and before the locational noun as in (271).[164]

(270) *pinggir* *jalan*
edge road
edge of road

(271) *jalan* *pu* *pinggir*
road POSS edge
edge of road

In WT, on the other hand, locational nouns are postposed to the object whose location is being predicated; some of these locational nouns are members of the inalienably-possessed noun subclass as in (272).

(272) *jala* *jua-i*
road edge-3s
edge of road

In table 5.5, all 162 locative genitive constructions in the corpus are categorized. All of the WT locational nouns follow the WT syntactic pattern, and most of the Malay locational nouns follow the primary strategy described above. The bolded numbers in the table indicate cells where there seems to be a definite preference for expressing the concept in one language or the other.

[164] This structure, which is a minor strategy for genitives of location, is the major strategy for genitives of possession.

Table 5.5. Locational genitive structures

Concept	Malay Loc noun: major strategy	Malay Loc noun: minor strategy	Malay Loc noun follows WT pattern	WT Loc noun follows object
behind	**3**	**4**	—	—
front	**22**	**2**	**1**	7
on top of	—	—	—	**31**
middle	3	—	—	**11**
distant	1	2	—	**5**
beside				
close	3	—	—	**39–40**
beside				
below				**22–23**
Total	34	8	1	119

All but one of the remaining Malay locational genitives follow the minor strategy described above; that one follows the WT pattern as in (273).

(273) ^{P}*Yala* ^{D}*muka*
Yala front
in front of Yala's place

It is probably best to consider this one instance a performance error, a calque of the equivalent WT sequence *Yala mona*. It is also possible that the minor strategy, which like WT has the locational noun after the other noun, occurs because of the similarity to the WT pattern. Note particularly that *belakang* ~ *balakang* 'behind' occurs more frequently in the minor strategy than in the major strategy, which would be very odd coming from a monolingual Malay speaker.

A closer look at table 5.5 reveals, however, that the general tendency to express specific locations by using WT locational nouns is not true for each specific locational concept. Some concepts follow such a general tendency, but 'on top of' (when occurring in a genitive construction) is always expressed with the WT locational noun *tit*, while 'behind' is always expressed by the Malay noun *belakang* ~ *balakang*.[165] In fact, if the number of instances were higher, I would conclude that *belakang* ~ *balakang* was default rather than non-default Malay when occurring in a genitive construction. When standing alone, however, it would still be non-default.

[165] The Malay locational noun *atas* 'on top of' does occur a few times in this corpus, but always as a lone noun, not in a genitive construction. The same is true of the WT locational noun *mir* 'behind'. The Malay locational nouns *belakang* ~ *balakang* 'behind' and *muka* 'front' occur both alone and within genitive constructions.

Even this does not tell the whole story, however. The expression of some locational concepts as Malay or WT depends upon the language used to express the object whose location is being predicated, as seen in table 5.6. Consider the concept 'front'. When the object whose location is being predicated is expressed by Malay (whether default or non-default), the Malay locational noun *muka* occurs in all instances but one. However, when the object is expressed by WT, the WT locational noun *mona* occurs. With the nouns for 'middle' and 'distant beside', there is a similar contrast between non-default Malay and WT, with default Malay split between choosing Malay and WT locational nouns. With the nouns for 'close beside' and 'below', both WT and default Malay objects cooccur with WT locational nouns, with indeterminate results for non-default Malay objects.

In fact, none of the thirty-four instances of Malay locational nouns which follow the major strategy cooccur with a WT noun. I could, therefore, argue that all thirty-four represent subclausal CS stretches, since they seem to be self-contained pockets of Malay. However, some of these may be due to lexical collocations rather than a single language choice covering several adjacent words. For example, there are eleven instances of *muka* 'front' followed by a non-default Malay noun; six of these are *muka rumah* 'in front of the house'. Whenever a particular string of words has a high frequency, I suspect that it is either a phrasal lexical entry or a lexical routine—which, in the terms of my analysis, would make it a single non-default lexical unit rather than a string of lexical units. If it is a single lexical unit, then its internal syntax is irrelevant to the question of whether it is a switch in language mode or a loan, just as the internal structure of a borrowed word is irrelevant to that question. The structure of an item is borrowed as part of the item itself.

The above discussion supports some of the notions I have argued for in chapters 4 and 5, particularly the need to examine lexical items individually rather than assuming homogeneity throughout the lexicon, and the dual-language status of default Malay lexical items. Another point made previously was the need to examine each speaker's contribution separately. The data on locational genitives support this argument as well. Returning to the six instances of *muka rumah* mentioned above, five of them are from LL and the other is from TN. The WT equivalents of this string are either *bôt mona* or *galáy mona*, which occur three times in the corpus—one from LL and two from OK. Although the data are too scanty to draw firm conclusions, it seems that LL and TN prefer to express the concept 'in front of the house' in Malay, while OK prefers to express it in WT. Interestingly, LL and TN are also the source of four of the five remaining instances of *muka* followed by a non-default Malay word. Thus,

there is a preference by particular speakers not just for a certain word, but for a certain word in a certain syntactic context.

Table 5.6. A more detailed look at locational genitives*

Concept	Malay Loc noun precedes object	Malay Loc noun follows object	WT Loc noun follows object
behind	**belakang X_N = 1** **belakang X_D = 2** belakang X_W = 0	**X_D POSS belakang = 4**	—
front	**muka X_N = 11** **muka X_D = 11** muka X_W = 0	X_D POSS muka = 2 X_D muka = 1	X_N mona = 0 X_D mona = 1 **X_W mona = 6**
on top of	—	—	**X_N tit = 2** **X_D tit = 19** **X_W tit = 10**
middle	**tengah X_N = 2** tengah X_D = 1 tengah X_W = 0	—	X_N tapôran- = 0 X_D tapôran- = 1 **X_W tapôran- = 10**
distant beside	**sabláh X_N = 1** sabláh X_D = 0 sabláh X_W = 0	X_N POSS sabláh = 0 X_D POSS sabláh = 1 XW POSS sabláh = 1	X_N pênai = 0 X_D pênai = 2 **X_W pênai = 5**
close beside	**pinggir X_N = 1** pinggir X_D = 2 pinggir X_W = 0	—	X_N juai/juin = 1 (?) **X_D juai/juin = 12** **X_W juai/juin = 27**
below	—	—	X_N sian- = 1 (?) **X_D sian- = 4** **X_W sian- = 18**

*X_N = non-default Malay NP; X_D = default Malay NP; and X_W = WT NP.

The locational genitive data may also support the hypothesis presented earlier that a high frequency of one usage of a word may strengthen another usage of that word in its competition with an equivalent word in the other language of a bilingual speaker's repertoire. It is interesting, for example, that the WT locational nouns *jua-* 'close beside', *tit* 'on top of', and *sian-* 'below', which seem to be the strongest in their competition with Malay, occur elsewhere in the corpus in a number of WT place names (*Got Juai*, *Pou Tit*, *Tabar Tit*, *Lia Tit*, *Kanga Tit*, *Til Siandi*), kin group names (*Nata-sian*, *Aka-sian*), and other phrasal lexical entries (*jarjar tit*, *loloar tit*,

pua siandi, *jek siandi*, *papa sian*, *bel sian*, *lar siandi*)—not to mention phrasal entries which do not happen to occur in this corpus.

My conclusion is that examining syntactic structures without reference to the specific words that make them up and the individual speakers who used them in order to determine whether the syntactic structures in question follow ML or EL patterns, is misguided. Without such attention to detail, striking down proposed syntactic constraints on CS can be merely an extended exercise in knocking down straw men.

5.4.7 Alternation and insertion

Muysken (1995:180–183) discussed two classes of models of CS: alternational models and insertional models. According to the former, two languages take turns on a more or less equal footing; the latter assume that a single matrix language (or base language) controls the overall process. Myers-Scotton's (1993b) Matrix Language-Frame Model is a mixture of both approaches; it allows for alternation on a large scale;[166] on a small scale, however, the ML controls morphosyntactic procedures within ML islands (which contain more than one ML morpheme and no EL morphemes) and in ML + EL constituents, while the EL controls morphosyntactic procedures within EL islands (which contain more than one EL morpheme and no ML morphemes).

I am arguing here for a more fine-grained approach. First, I follow Myers-Scotton in proposing that a matrix language is nearly always identifiable. In saying nearly, I depart from Myers-Scotton; I do allow for the possibility that ML turnover may occur gradually. Identifiability of the ML does not, however, mean that a speaker's language mode is always identifiable; for example, sometimes a speaker produces such a long string of default EL that a hearer misinterprets the language mode, as in (274).

(274) 1\LL *Oh* ***[P]Henrik [NF]kan [N]dolo [NF]kan [C]sekolah [P]STM, [NF]to?***
Oh, Henrik, you know, back then, you know, [was at] Technical School, you know?
2\TN ***[Y]Iyo.***
Yes.

Here, LL produces a topic-comment structure, a sentence pattern available in both WT and Malay, containing nothing but default Malay lexical units. There is, therefore, no firm evidence that she is in Malay mode; however, TN answers her as if she were, by using a gratuitous Malay item.

Second, I predict that many putative instances of Auer's (1995:126) Pattern III code alternation (that is, rapid alternation with no discernible

[166] Her term is "turnover of the matrix language".

matrix language) will usually turn out to have a clear ML once default EL lexical units are excluded.

Third, I propose a subdivision of alternation into CLOSED ALTERNATION and OPEN-ENDED ALTERNATION, depending on whether the speaker is bound to return to the original matrix language or not. Open-ended alternation would include all of the larger-scale situational switching and metaphorical switching in which a complete changeover from one language variety to another coincides with a change of topic, situation, participants, and so on. It would, therefore, include many instances of addressee-related CS, whenever the speaker is leaving open the possibility of continuing to speak to the new addressee. Another example of open-ended alternation is the switching of language modes seen in the letter of the young Dobel man presented in §5.1. Open-ended alternation might also include switching to the power code for the purpose of getting people to do what the speaker wants, as in Scotton and Ury's (1977:16–17) example of the impatient bus passenger who switches to English to get his change back from the conductor (and the conductor answers in English).

Open-ended alternation is equivalent to Myers-Scotton's "synchronic turnover of the ML"; that is, it involves not only a switch in language mode but a switch in matrix language as well. Moreover, with addressee-related CS, if the addressee does not understand the previous ML, the switch is not only a change of language mode and matrix language, but also a switch from bilingual mode to monolingual mode (at least in the present corpus, where all the main participants are usually in bilingual mode).

Closed alternation could be thought of as inserted alternation—that is, it is a turnover of the ML (not merely a change of mode) which is expected to be temporary. The only clear example of this in the WT/Malay corpus is addressee-related CS which is clearly a brief side-comment, not intended to leave open the option of continuing to speak to that new addressee. This is seen in (113), repeated here as (275).

(275) \LL ***^{P}Usi•Ani ^{Y}dengar ^{Y}saja, ^{AF}eh? [.]^{Y}Jang ^{Y}sambung***.
Ms. Ani, just listen, okay? Don't join the conversation.

Of course, if Ani had chosen to disregard LL's request, this could have become an instance of open-ended alternation.

Another consideration is whether direct quotes which are completely Malay introduced by quote formulas which are completely WT might be instances of closed alternation. Clearly, the direct quote itself is an instance of Malay mode, but is completely embedded within WT mode and must by definition have an end. The only question is whether this switch to Malay mode also represents a switch in matrix language. I do not

believe there is sufficient evidence for a turnover of the matrix language in these cases; these are rather instances of Malay mode which in many cases are reflections of Malay as matrix language in the original conversations being reported. If Malay direct quotes represented a turnover of the ML, I would expect speakers to be much more consistent in using only Malay to report original Malay conversations, rather than the tendences noted in §5.2.2.[167]

Another possible example of closed alternation would be a brief self-interrupting side comment, such as the one in (276).

(276) 1\AD *Ok kom•on* ***«Eh, [N]dolo-dolo [e] [P]Weli [Y]ini, [Y]liha [Y]katong [Y]di [Y]jalan [Y]jua, [.] [[Y]katong-] [NF]lebái [a [Y+]orang-orang [NF]yang [Y]punya]***—{aside} *monjá, ok naka ka nei, te*—
I said "Hey, formerly [uh] this Weli, if she saw us in the street, [.] [we-] preferably [people who have]"—{aside} wait a minute, I said it to her, or—

2\OK ^ *(eya)*
^ (Yes)

3\AD ^ ***[Y+]orang-orang [NF]yang [Y]su = [Y]punya [C]titel, [DF]biar [Y]liha [Y]katong [Y]di [Y]jauh, [Y]dia [Y]su = [L]kas-suara.»***
"^ people who have a rank, even though she sees us far off, she greets us."

In quoting herself speaking to a non-WT speaker, AD is completely in Malay mode; but in a brief side comment, she switches completely to WT mode, then back again to Malay mode for the rest of the quote. Again, I do not see sufficient cause for analyzing this instance of Malay mode as an instance of Malay-as-ML.

Fourth, I propose a subdivision of insertion into two types, namely insertion of EL words in EL mode versus insertion of EL words while maintaining ML mode. The latter (EL-word insertion) would include most instances of single non-default EL lexical units, while the former (EL-mode insertion) would include any multi-unit EL stretch in which the morphosyntax is that of the EL; this would include the other six form categories, as well as (some?) calques and other hybrid structures. In terms of the present corpus, EL-mode insertion would involve a string of Malay lexical units (including at least one non-default Malay lexical unit?) displaying Malay morphosyntax, while EL-word insertion would involve a string of Malay lexical units displaying WT morphosyntax.

Where WT and Malay morphosyntax is identical, it would be impossible to determine which kind of insertion is involved from transcription data

[167] Note that even if the ML is Malay in these direct quotes, the occurrence of an occasional WT conjunction demonstrates that the speaker is still in bilingual mode.

alone; as mentioned previously, some kind of neural scanning data would be required. In other words, adjacent non-default Malay lexical units may be merely coincidentally-adjacent EL-word insertion rather than EL-more insertion. However, where the two languages differ in morphosyntax, a lack of ML morphosyntax is an indication of EL mode. The difference between WT and Malay genitives has already been noted; another difference is that in WT, noun modifiers must be reduplicated. When a Malay noun is modified by an unreduplicated Malay verb, it is EL-mode insertion; note examples (277)–(280).

(277) $_{D+}$***teman-teman*** $_{N}$***yang*** $_{Y}$***kaluar*** *(*not *teman-teman yang kaluluar)*
friends who left

(278) $_{Y}$***celana*** $_{Y}$***putih*** *(*not *celana* ***puti****putih)*
white pants

(279) $_{Y}$***Rumah*** $_{Y}$***lama*** *(*not *rumah* ***lama****lama)*
previous house

(280) $_{C+}$***kota-kota*** $_{Y}$***besar*** *(*not *kota-kota bes****ar****sar)*
big cities

The main reason for distinguishing between EL-word insertion and EL-mode insertion is to account for strings of EL lexical units in ML order (i.e., EL words in ML mode). Although this is rare in the WT/Malay corpus (cf. *Yala muka* 'in front of Yala's place'), and probably in all corpora, it is attested: for example, Myers-Scotton (1995:253, footnote 7) found in her Nairobi corpus four instances of English words inserted in ML order (e.g., *table long*). Note, however, that since such strings are rare they may be instances of performance errors.

How do my two types of alternation and two types of insertion correspond to Myers-Scotton's (1993b) model? And what is the relationship between my modes and her ML versus EL? Although Myers-Scotton claims that her proposed constituents do not involve a turnover of the ML, they definitely involve different modes: ML + EL constituents and ML islands are in ML mode (in her terms, ML morphosyntactic procedures are activated), while EL islands are in EL mode. Applying her conception of ML and EL to the WT/Malay corpus and to my model, I make the following observations:

- WT is the ML, and Malay is the EL throughout the corpus, except for some instances of addressee-related CS where alternation occurs
- My EL-word insertion is a subtype of her ML + EL constituents
- My EL-mode insertion and closed alternation are subtypes of her EL islands

I have offered evidence from the WT/Malay corpus that my conceptual categories are more in line with the psycholinguistic reality which underlies bilingual speech. With regard to the relationship between form categories and psycholinguistic modes, I conclude the following for the Malay found in this corpus.

- All single non-default Malay items are insertions of Malay into WT mode.
- All instances of major CS are instances of Malay mode; more specifically, they are EL-mode insertion (unless some evidence is found for analyzing direct quotes as closed alternation).
- When an instance of subclausal CS displays Malay syntax, and contains only gratuitous Malay lexical items, it is a clear instance of Malay mode (EL-mode insertion). The remaining instances of subclausal CS are either instances of Malay mode, or instances of single non-default Malay items coincidentally adjacent to each other (EL-word insertion).
- Triggered sequences and collocation sequences are brief instances of Malay mode as well; perhaps they should be called asymmetric Malay mode, since the language choice of one lexical item clearly depends on that of another.

5.5 Summary and conclusions

In this chapter I have proposed several possible pathways which have led to the occurrence of non-default Malay lexical items in the WT/Malay corpus.

1. motivated by the needs (or niceties) of the communication situation:
   - communication with an addressee who does not understand WT (or merely prefers Malay)
   - accommodation to another speaker's (perceived) language choice
2. motivated by the speaker's own mental lexicon and quirks of processing:
   - collocations (adjacent, disjoint, and covert)
   - triggering (both triggered sequences and triggered language switching)

- interference (misinterpretations, performance errors, doubling, hybrids)

3. motivated by aesthetics:
   - speaker attitude (anger) and style (emphatic, religious)
   - direct quotations (replicate original, mark ethnicity, exhibit carefulness)

It is not my intent that the above list be taken as a general typology of CS motivations in all languages or even in the Dobo community of WT speakers; rather, it is merely a list of motivations for language mode switching that I have found evidence for in the particular corpus under study here. As Auer (1995:121) pointed out, we will probably never arrive at a complete list of CS functions, because bilinguals use CS creatively.

I have also proposed that some multi-word stretches of Malay are based on syntactic constituents (switching the language mode of an entire constituent, e.g., a clause or phrase), while others are based on linear order (left-to-right switching which pays little heed to constituent structure). This may not be satisfying to syntacticians wishing to find a single principle underlying all language mode switching, but I believe this is what the data support. My goal is not to propose the most efficient and simple model of language possible, but the most efficient and simple model which actually represents the way human beings process language.

It is not always possible for the analyst to determine whether a given stretch of EL words has resulted from separate, word-by-word language choices, or a single collective language choice (language mode). In order to have some principle for categorizing the data, I decided on making a primary distinction between those stretches which contain at least two adjacent non-default Malay lexical units and those which do not. I have noted that this scheme, while useful, is often too crude to reveal the apparent motivations for the use of Malay lexical items, and must be supplemented with a more fine-grained analysis.

To determine the language mode of any given string of words containing EL lexical units, I propose that analysts consider the following four questions.

1. Is the EL lexical unit default or non-default?
2. If non-default, is it gratuitous (based on relative and absolute frequencies)?
3. If there are only default EL lexical units, does the context indicate EL mode (e.g., is it part of a direct quote sequence)?
4. Does the morphosyntax indicate EL mode (when EL morphosyntax differs from ML morphosyntax)?

Only after these questions are answered can one begin to consider the question of syntactic constraints on code-switching (or rather, constraints on switching language modes). It may be that there are separate constraints, or different degrees of constraints, on preferred, dispreferred, and gratuitous EL, and no constraints at all on necessary EL.

In conversation data, there is more than one psychological reality involved. Not only are we interested in knowing the psychological processes of the speaker who produced a stretch of Malay, we also need to consider the interpretation of the hearer(s). A stretch of Malay lexical items which were coincidentally adjacent from the point of view of the speaker's production may be subsequently interpreted as Malay mode by the hearer, possibly resulting in the hearer following the first speaker's (inferred) Malay mode with more Malay mode. Example (281) illustrates the preceding principle, as well as several others in the chapter.

(281) 1\TN *DF**tarús** ok sika P**Henrik**. Nei NF**ada**, NF**to**? Nei [.] YN**habis [a]** C**ujian** êr, ka nei NF**ada**, DF**tarús** kusí. «P**Henrik. (Y**Ini C**surat)** Y**dari** Y**ko** Y**pu** Y**maitua.»* DF**Tarús** *nei nam•on*
Then I went to Henrik. He was there, you know? He [.] after the test, he was there, then I went to him. "Henrik. (Here's a letter) from your wife." Then he said

2\LL *Oh P**Henrik** NF**kan** N**dolo** NF**kan** C**sekolah** P**STM**, NF**to**?*
Oh, Henrik, you know, back then, you know, Technical School, you know?

3\TN *Y**Iyo**.*
Yes.

4\LL *P**STM** Y**di** P**Langgur** ^ Y**kah***
Technical School at Langgur or something

5\TN ^ P**STM** R**di** P**Tual** *te. [.]* DF**Tarús** *nei nam•on, «C**Surat** Y**dari** Y**mana?»* Ok kom•on; «Y**Dari** P**Tina.»*
Technical School at Tual of course. [.] Then he said, "Letter from where?" I said, "From Tina."

6\LL *Nei C**lesten** pel P**Petrus** ká YF**mungkin**.*
He was in the same class as Petrus and his group maybe.

7\TN *Eya, ^ te,* Y**#satu** C**lesten** P**Petrus**,*
Yes, ^ of course, in one class were Petrus,

8\AL ^ *(Eya. ...)*
^ (Yes ...)

9\TN *[eh] ^* P**Robi**,*
[uh] ^ Robi,

10\LL ^ P**Bian** Y**dong**.*
^ Bian and his group,

11\TN[P]***Robi.*** *[.]* [DF]***Tarús*** *[e] [.] ok kom•on; («Eya» nang)* [DF]***tarús*** *nei nam•on; «*[C]***Surat*** *bana ba?»*
Robi. [.] Then [uh] [.] I said; ("Yes" so) then he said; "Letter from where?"

In turn 3, TN apparently takes LL's turn 2 to be an instance of Malay mode, and follows LL's perceived language choice by answering with a gratuitous Malay word. Other observations about this brief interchange are as follows.

- In turns 1, 4, and 10, there occur the triggered sequences *habis ujian* 'after the test', STM *di Langgur kah* 'Technical School in Langgur or something', *Bian dong* 'Bian and his group'. (The Malay locative preposition *di* in turn 5 does not qualify as a non-default lexical unit, since TN is echoing LL's lexical choice in turn 4.)
- In turns 1 and 5, TN is reporting a conversation between herself and Henrik, a non-WT speaker; after an interruption by LL, she resumes the reported conversation in turn 11 by repeating part of turn 5, but in WT.
- In turn 2, LL's topic-comment sentence is completely Malay, but since none of the words are non-default, and since such a topic-comment structure is possible in WT as well as Malay, I have not analyzed it as a stretch of Malay mode.[168]
- In turn 7, the Malay word *satu* 'one' is analyzed as resulting from collocation with the default Malay *lesten* 'class' (in the sense of 'class of 1991'); but it is also possible that this is a triggered sequence rather than a collocation sequence.

As analysts work with recorded and transcribed conversation data, they have the hearer's perspective, and attempt to ascertain the speaker's purposes on the basis of the stream of speech alone. Furthermore, it is a very limited hearer's perspective. Even if there were videotapes to examine, it would still not be the same as actually participating in the conversation. Our understanding of what motivated the speaker's lexical choices and language choices is far less than that of the addressee in most cases, since the speech is tailored for that addressee and (in the case of conversation among old friends) builds upon years of shared experiences. The analyst's lack of understanding is compounded when he is not as fluent a speaker of one of the languages as the subjects are (or in my case, of both languages). In summary, I must admit that there is a certain amount of guesswork involved in inferring psycholinguistic motivations for lexical choices and

[168] There are only a few such sentences in the corpus.

language choices, as I have attempted to do in the present chapter as well as the preceding one.

In this chapter, I have attempted to account for all of the non-default Malay insertions in the corpus from lone words to multi-sentence stretches. In doing so, I subjected the data to three cycles of analysis, first noting (as many other LCP researchers have done) the use of EL to mark various conversational motivations for a change of language mode, such as addressee-related CS, CS in reported speech, CS for emphasis, CS to mark a certain style, and CS for accommodation and negotiation. As Auer (1995:120) pointed out, many studies of CS have focused on developing such typologies of conversational functions for CS. This approach, however, only accounted for a part of the data.

Next, I divided the data into seven form categories: single non-default lexical units, four structurally-defined types of EL sequences which contain adjacent non-default Malay lexical units (Complete-S CS, S-Fragment CS, Near-S CS, and Subclausal CS), and two types of EL sequences not containing any adjacent non-default Malay lexical units. These latter two (triggered sequences and collocation sequences) were posited to be the outcome of neurolinguistic or psycholinguistic processes in which the default Malay items brought about the occurrence of the adjacent non-default Malay items. As in other studies, lone EL lexical units occurred far more frequently than longer EL stretches, even after I excluded lone default Malay items.

In a refinement of Clyne's (1967:84) proposal that neutral sites may trigger a language switch, I proposed that while established loans may indeed be neutral, other default Malay items have dual language membership—that is, while the item has become an integral part of the ways of speaking of WT speakers, they still consider it to be a Malay item. Evidence for this view was seen in the many gratuitous items which only occurred adjacent to other Malay items. I found evidence for different strengths of collocations depending on the lexical items in question and also depending on the speaker; I also presented examples of apparent disjoint and covert collocation. I further proposed that the choice of an EL lexical item sometimes invokes an EL syntactic structure. I suggested that the present analysis could be refined by a closer examination of subclausal stretches, to see whether a sequence of non-default Malay items should be reanalyzed as triggered sequences or collocation sequences, if one of the items was gratuitous and the other was not. To my knowledge, no LCP researcher to date has taken account of such details in published analyses of bilingual corpora.

In the third cycle of analysis, I attempted to identify which EL insertions represented purposeful functions of a switch in language mode (whether

for aesthetic purposes, or for the needs of the communication situation) and which represented unintentional outcomes of neurolinguistic or psycholinguistic processing (some involving a change of language mode and others not). Specific findings included the following:

- In 5.4.1, I noted that a correlation exists in the corpus between major CS (Complete-S, S-Fragment, Near-S) and direct quotes; this correlation was interpreted as evidence for a more careful style within direct quotes. Since finding such a correlation depended on the distinction made between default and non-default Malay items, that distinction is therefore supported by this finding.
- In 5.4.2, I demonstrated that careful style was also evident in the hypercorrect speech of one of the primary speakers. I proposed that a model of LCP should include not only language modes, but styles of language choice; that is, lone Malay items may be sprinkled in to effect a certain style, just as phonological and syntactic elements have been observed by other researchers to mark certain styles (Labov 1966, Trudgill 1974, Chambers and Trudgill 1980:67ff).
- In 5.4.3, I identified some mixed strings not as Malay insertions into WT but as WT insertions into Malay mode. These were mostly WT conjunctions between Malay clauses. It is possible that these inserted WT conjunctions function as ethnicity markers (Giles 1979).
- In 5.4.4, some phenomena were labeled “interference”, including both innovations and errors (misinterpretations, performance errors, hybrids, and doubling), involving both production and interpretation. While Myers-Scotton (1993b) focused on double plural marking as the most common type of double morphology, I found that with this language pair double plural marking was again the most common type of doubling, but not involving affixes from two languages on one noun. I, therefore, suggested that it may be the feature “plural” which is most prone to doubling, whether or not double morphology is involved. I also discussed examples of a wide variety of other types of doubling in the WT/Malay corpus, including a type of compounding of conjunctions in which a semantically broad WT conjunction is followed by a more specific Malay conjunction.
- In 5.4.5 and 5.4.6, I argued that EL syntax is not a sufficient indicator of EL mode. The argument was based on Starosta’s

(1988) notion that "the grammar is in the lexicon", that is, that syntactic structures are projected from lexical features; thus, if a Malay word occurs nested within a Malay syntactic structure foreign to WT, the structure need not be interpreted as evidence of Malay mode. In 5.4.6, I continued this argument, focusing on locational genitives in the corpus.

- Finally, in 5.4.7, I related my findings to Muysken's (1995) distinction between alternational CS and insertional CS, proposing two subtypes of each (closed and open-ended alternation, EL-word insertion, and EL-mode insertion). My model is more fine-grained than Myers-Scotton's Matrix Language Frame model (1993b); and in contrast to that model, I proposed that the ML is not necessarily identifiable at every point in a discourse, though it usually is. I further proposed that a speaker's language mode is not always identifiable; in fact, it is sometimes misinterpreted by the hearers. However, in step with Myers-Scotton's model, I also predicted that any bilingual corpus which seems to have rapid alternation of languages with no discernible ML will turn out to have a clear ML once we take care to distinguish default and non-default EL items.

None of these observations would have been possible without a micro-analysis of the corpus. A broad statistical survey of the corpus (such as that of Poplack and Meechan 1995) would not have helped in my attempt to account for every Malay lexical unit which occurs in the corpus in a way that is psycholinguistically real—that is, from a perspective of psycholinguistic processes which produce and interpret LCP rather than from a large-scale societal perspective.

I have argued that micro-analysis of a bilingual corpus must begin with identifying *lexical units* based on correspondences between the two (or more) languages in the corpus and continue on through frequency counts of those lexical units as used by individual speakers in specific lexical, discourse, and social contexts. A certain amount of macro-analysis must precede the micro-analysis (e.g., to help identify the ML) and follow the micro-analysis (e.g., to form a general theory of conversational functions of language mode switching), but these former and latter phases of research can only be validated by the accompanying micro-analysis.

A number of recent studies (e.g., Bortoni-Ricardo 1985, Milroy 1987, Milroy and Li 1995) have shown that a sociolinguistic analysis based on social networks is more accurate than one based on a statistical analysis of cohorts (age, gender, education, and so on). Similarly, the present study

demonstrates that a lexical network analysis (collocations, senses, and so on) is more accurate than a statistical study of word classes or other categories. I argue that micro-analysis should not be used as a mere patch that we have to resort to in order to make problematic data fit macro-generalizations; instead, micro-analysis must define the categories upon which macro-analysis is performed.

In the present work, I have not attempted to replace any existing models of CS. Rather, I have demonstrated the kind of prerequisite micro-analysis that I believe must be performed *prior* to admitting any data as evidence for or against any particular model of CS or any syntactic constraints on CS.

6

Epilogue: Future Directions

> *In the midst of this sublime and terrible storm, Dame Partington, who lived upon the beach, was seen at the door of her house with mop and pattens, trundling her mop, squeezing out the sea-water, and vigorously pushing away the Atlantic Ocean. The Atlantic was roused; Mrs. Partington's spirit was up. But I need not tell you that the contest was unequal; the Atlantic Ocean beat Mrs. Partington.*
>
> Sydney Smith, *Speech at Taunton*, 1813

In this final chapter, I discuss possibilities for future research to follow up on the research presented in the previous chapters. I close with a discussion of possible futures of the West Tarangan language.

Directions for future research

As with all research, there are still many loose ends and many questions yet unanswered. Many other studies could grow out of the present work, both building upon and refining its conclusions. For example, although I examined the mixing patterns of each individual primary speaker, I did not examine patterns of accommodation—that is, whether or not the same speaker has different patterns dependent on who the other participants in the interaction are. Much more data would have to be collected in order to have enough data to draw valid conclusions in this regard.

Obviously, one could go beyond the individual speaker perspective of the present study to examine social networks, or cohorts based on

parameters such as age, gender, education, and so on. A variety of comparative studies would be possible. The findings of this study (in which most of the speakers are residents of Dobo) could be compared to a similar study conducted in the home villages of the primary speakers of the present study.

As noted in chapter 3, most of the speakers in this study were aware of being recorded, and this undoubtedly affected their speaking style, including their use of Malay. It would be good to do a second study of people who have given prior consent to being recorded without their knowledge and compare the results with those of the present work. In particular, we might examine the phenomena of self-repair and other-repair, and of morphological integration.

Also, since OK was the only WTB speaker in my corpus, data from more WTB speakers could be gathered to determine which characteristics of OK's speech are common to other WTB speakers. One could also look for differences between WTA speakers and WTB speakers, or between speakers of individual villages. According to OK, the people of her village refer to mixed speech as *Rau Garjijá* 'language/speech of Garjá [=Lorlor village]', since (they claim) the people of Lorlor allow more Malay insertions than they do.

Another way to refine the present work would be to include mixed lexical units in the analysis of chapter 5. There are dozens of instances where either a pure non-default Malay item apparently occurs due to collocation with a mixed word as in (282) or else a mixed non-default item occurs due to collocation with a pure Malay word as in (283).

(282)	*da-**kunci** **ʏpintu***	3p-lock door
	*da-**toki** **ʏpintu***	3p-knock_on door
	***ʏmasih** ku-**kuat**-na*	still DUP-strong-3sa
	*da-**kas-masu** **ʏnama***	3p-CAUS-enter name
(283)	*%ʏda-**cari dana***	3p-seek funds
	*%ʏda-**paki semang***	3p-use outrigger
	*%ʏi-**biking masalah***	3s-make problem
	*%ʏmi-**dengar pengumuman***	2p-hear announcement

There may also be examples of triggered sequences involving mixed words as in (284).

(284) \YL *gasua %ᴘKei-**Kei**-na, **ʏatau** %ᴘ**balakang-tanah**-na ina.*
child DUP-Kei-3sa or behind-land-3sa one.
a young man from Kei, or one from the other side of Aru.

Finally, for the purpose of language maintenance research, the current study may be used as the basis for a longitudinal study. According to

Giacalone Ramat (1995:62), "analysis of the development of CS patterns over time" is greatly needed. Every ten years, a few hours of conversation could be recorded, and a lexical analysis similar to that presented in chapter 4 could be undertaken in order to determine whether there has been a shift from WT to Malay or vice versa for particular lexical units. If there are a number of independent lexical units which have shifted in the same direction, this may be taken as an indicator of the direction the language as a whole is going.

At present, I cannot say whether WT will survive many more generations. If language shift to Malay does occur, I hope that researchers interested in examining the precursors to language death will gain insights from the data presented here.

The future of the West Tarangan language

WT is in a similar situation to that described for Alsatian by Gardner-Chloros, "code-switching is...the only 'we code' now available to a large number of younger Alsatian speakers, who are no longer able to speak the dialect in an unmixed form" (1995:80). WT speakers who are attempting to maintain their language are enriching it by adding words which allow more fine-grained semantic distinctions; as a result, the WT which is being maintained includes massive Malay intrusions. As seen here, the presence of default Malay words may trigger the use of other Malay words in collocation sequences or triggered sequences. It is possible that this may lead to language shift; as mentioned in chapter 3, many WT speakers felt they were incapable of conversing in pure WT, which could be an indicator of impending language obsolescence. However, no linguist to date has successfully shown that code-switching necessarily leads to language shift.

As elsewhere in Maluku, Taranganese parents have for years been teaching their children Malay as a first language, in order to prepare them for elementary school. Some wanted to increase their children's chances for gainful employment; others were primarily concerned with protecting their children from the negative consequences of not being proficient in Malay. In any event, while Malay was initially seen purely as a tool, it has increasingly become a symbol of personal identity. In some families Malay actually replaced WT as the language of home and family. In one small village, Ngaiguli, no one under the age of 20 is able to speak WT.

If this were the situation in every village, the outlook for the future of WT would be bleak. However, in most villages this is not the case. In my survey of WT speakers in Dobo, I found that most young people who had

attended elementary school in their home village were fluent in WT, while most who had attended elementary school in Dobo were not. If this is still true for children in WT villages today, then there is still hope that WT will remain a strong language for some time to come.

In addition, the Indonesian government has recently increased its support for local cultures and languages via the *Muatan Lokal* program, in which certain aspects of the local culture are taught in primary schools. Wherever appropriate materials are available, the local language is taught as part of this program. Assuming that such materials are developed and used in the near future, WT parents, even in Ngaiguli, may decide to make sure their children are fluent in WT before entering school.

Time will tell whether the current generation of parents in Ngaiguli will consider WT important enough to begin teaching it to their youngest children as a first language again. Perhaps, with the new support for vernacular languages in the local schools, a reversal of the trend is possible.

If cultural diversity is a highly-valued commodity, then Indonesia is one of the wealthiest nations in the world. Not only so, but speakers of indigenous languages are culturally enriched by joining the mainstream of Indonesian society. In addition, new economic opportunities are available to those who speak Indonesian. However, as seen in other nations, devotion to the national language can inadvertently lead to the death of the indigenous language. Although governments and organizations may strive to prevent language death, decisions about language maintenance are ultimately made on the level of individual speakers. As Indonesia continues on the road of economic development, the speakers of the nation's many indigenous languages will have to decide, one by one, whether such economic development must be at the expense of cultural and linguistic diversity, or whether they can enjoy the best of both linguistic worlds.

References

Ammon, Ulrich, Norbert Dittmar, and Klaus J. Mattheier, eds. 1988. Sociolinguistics: An international handbook of the science of language and society. Berlin and New York: de Gruyter.

Appel, R., and Pieter Muysken. 1987. Language contact and bilingualism. London: Edward Arnold.

Auer, Peter. 1984. Bilingual conversation. Amsterdam: Benjamins.

Auer, Peter. 1995. The pragmatics of code-switching: A sequential approach. In Milroy and Muysken 1995a, 115–135.

Auer, Peter, ed. 1998. Code-switching in conversation: Language, interaction and identity. London: Routledge.

Backus, A. 1990. Turkish-Dutch codeswitching and the frame process model. Paper presented at the International Workshop on Ethnic Minority Languages in Europe, Tilburg, Netherlands, December 1990.

Bentahila, A., and E. D. Davies. 1983. The syntax of Arabic-French code-switching. Lingua 59:301–330.

Berk-Seligson, Susan. 1986. Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. Language in Society 15(3):313–348.

Blom, J. P., and John J. Gumperz. 1972. Social meaning in linguistic structures: Code-switching in Norway. In John J. Gumperz and Dell Hymes, (eds.), Directions in sociolinguistics, 407–434. New York: Holt, Rinehart and Winston.

Blust, Robert. 1993. Central and Central-Eastern Malayo-Polynesian. Oceanic Linguistics 32(2):241–293.

Bokamba, Eyamba G. 1988. Code-mixing, language variation, and linguistic theory: Evidence from Bantu languages. Lingua 76(1):21–62.
Bortoni-Ricardo, Stella Maris. 1985. The urbanization of rural dialect speakers: A sociolinguistic study in Brazil. Cambridge Studies in Linguistics Supplementary Volume. Cambridge: Cambridge University Press.
Breitborde, L. B. 1983. Levels of analysis in sociolinguistic explanation: Bilingual code switching, social relations, and domain theory. International Journal of the Sociology of Linguistics 39:5–43.
Briggs, Charles L. 1986. Learning how to ask: A sociolinguistic appraisal of the role of the interview in social science research. Cambridge: Cambridge University Press.
Cartier, Alice. 1979. De-voiced transitive verb sentences in Formal Indonesian. In Frans Plank (ed.), Ergativity: Towards a theory of grammatical relations, 161–183. New York: Academic Press.
Casparis, J. G. de. 1997. Sanskrit loan-words in Indonesian: An annotated check-list of words from Sanskrit in Indonesian and Traditional Malay. NUSA 41.
Chambers, J. K., and Peter Trudgill. 1980. Dialectology. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
Clark, Ross. 1982. 'Necessary' and 'unnecessary' borrowing. In Amran Halim, Lois Carrington, and S. A. Wurm (eds.), Papers from the Third International Conference on Austronesian Linguistics, Vol. 3: Accent on Variety. Pacific Linguistics C-76:137–143.
Clyne, Michael. 1967. Transference and triggering. The Hague: Nijhoff.
Clyne, Michael. 1988. History of research on language contact. In Ammon et al., 452–459.
Collins, James T. 1980. Ambonese Malay and Creolization theory. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Collins, James T. 1981. Pertembungan linguistik di pulau Ambon: Bahasa Melayu dan Bahasa Asilulu. Jurnal Dewan Bahasa 25(7):30–55.
Collins, James T. 1983. Syntactic change in Ambonese Malay: The possessive construction. In James T. Collins (ed.), Studies in Malay Dialects, Part 2. NUSA 17:28–41.
Collins, James T. 1995. Dialek Melayu di Pulau Kalimantan dan bahasa Bacan: Misanan atau mindoan? In Soenjono Dardjowidjojo (ed.), PELLBA 8, Sewindu. 225–265. Yogyakarta: Penerbit Kanisius.
Cruse, D. A. 1986. Lexical semantics. Cambridge: Cambridge University Press.
Dabène, Louise, and Danièle Moore. 1995. Bilingual speech of migrant people. In Milroy and Muysken 1995a, 17–44.

Dardjowidjojo, Soenjono. 1978. Sentence patterns of Indonesian. Pali language texts: Southeast Asia. Honolulu: University of Hawaii Press.
Denes, I Made, I Nengah Budiasa, Cokorda I Stri Sukrawati, and Ni Wayan Ariani. 1994. Interferensi Bahasa Indonesia Dalam Pemakaian Bahasa Bali di Media Massa. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
DiSciullo, Anna-Maria, Pieter Muysken, and R. Singh. 1986. Government and code-mixing. Journal of Linguistics 22:1–24.
DiSciullo, Anna-Maria, and Edwin Williams. 1989. On the definition of word. Linguistic Inquiry Monograph 14. Cambridge, Mass.: MIT Press.
Dorian, Nancy C, ed. 1989. Investigating obsolescence: Studies in language contraction and death. Cambridge: Cambridge University Press.
Dressler, Wolfgang U. 1988. Language death. In Frederick J. Newmeyer (ed.), Language: The socio-cultural context, Linguistics: The Cambridge Survey, vol. 4, 184–192. Cambridge: Cambridge University Press.
Eckert, Penelope. 1980. Diglossia: Separate and unequal. Linguistics 18(11/12):1053–1064.
Eliasson, S. 1991. Models and constraints in code-switching theory. Papers for the Workshop on Constraints, Conditions, and Models, 17–50. Strasbourg: European Science Foundation.
Ellen, Roy F. 1986. Conundrums about panjandrums: On the use of titles in the relations of politicial subordination in the Moluccas and along the Papuan coast. Indonesia 41:46–62.
Evens, Martha W., Bonnie E. Litowitz, Judith A. Markowitz, Raoul N. Smith, and Oswald Werner. 1983. Lexical-semantic relations: A comparative survey. Edmonton, Alberta: Linguistic Research, Inc.
Farb, Peter. 1974. Word play: What happens when people talk. Toronto, NewYork, London: Bantam.
Ferguson, Charles A. 1959. Diglossia. Word 15:325–340.
Fishman, Joshua A. 1965. Who speaks what language to whom and when? La linguistique 2:67–88.
Fishman, Joshua A. 1971. Sociolinguistics. Rowley: Newbury.
Gal, Susan. 1979. Language Shift: Social determinants of linguistic change in bilingual Austria. New York: Academic Press.
Gal, Susan. 1988. Linguistic repertoire. In Ammon et al., 286–292.
Gardner-Chloros, Penelope. 1991. Language selection and switching in Strasbourg. Oxford: Oxford University Press.
Gardner-Chloros, Penelope. 1995. Code-switching in community, regional and national repertoires: the myth of the discreteness of linguistic systems. In Milroy and Muysken 1995a, 68–89.

Geerts, Guido. 1988. Research on language contact. In Ammon et al., 598–606.
Giles, Howard. 1979. Ethnicity markers in speech. In Klaus R. Scherer and Howard Giles (eds.), Social markers in speech, 251–289. Cambridge: Cambridge University Press.
Grimes, Barbara D. 1991. The development and use of Ambonese Malay. In Hein Steinhauer (ed.), Papers in Austronesian Linguistics No. 1. Pacific Linguistics A-81:83–123.
Grimes, Barbara F., ed. 2000. Ethnologue. Thirteenth edition. Dallas: Summer Institute of Linguistics.
Grosjean, François. 1982. Life with two languages: An introduction to bilingualism. Harvard University Press.
Grosjean, François. 1995. A psycholinguistic approach to code-switching: The recognition of guest words by bilinguals. In Milroy and Muysken 1995a, 259–275.
Gumperz, John J. 1976. The sociolinguistic significance of conversational code-switching. In Jenny Cook-Gumperz and John J. Gumperz (eds.), Papers on Language and Context 46, 1–46. Berkeley: University of California, Language Behavior Research Laboratory.
Gumperz, John J. 1982. Discourse strategies. Studies in Interactional Sociolinguistics 1. Cambridge: Cambridge University Press.
Gutt, Ernst-August. 1991. Translation and relevance: Cognition and context. Oxford and Cambridge, Mass.: Blackwell.
Hagen, Anton M. 1988. Sociolinguistic aspects in dialectology. In Ammon et al., 402–413.
Haugen, Einar. 1950. The analysis of linguistic borrowing. Language 26:210–231.
Haugen, Einar. 1953. The Norwegian language in America: A study in bilingual behavior. Philadelphia: University of Pennsylvania Press.
Haugen, Einar. 1971. The ecology of language. The Linguistic Reporter, Supplement 25, 19–26. [Reprinted in The Ecology of Language: Essays by Einar Haugen, ed. by Anwar S. Dil, 325–39. Stanford: Stanford University Press, 1972.]
Heider, Karl G. 1991. Landscapes of emotion: Mapping three cultures of emotion in Indonesia. Cambridge: Cambridge University Press.
Heller, Monica. 1995. Code-switching and the politics of language. In Milroy and Muysken 1995a, 158–174.
Hill, Jane H., and Kenneth C. Hill. 1986. Speaking Mexicano. Tucson: University of Arizona Press.
Hockett, Charles F. 1958. A course in modern linguistics. New York: MacMillan.

Hoevell, G. W. W. C. Baron van. 1877. Iets over de vijf voornaamste dialecten der Ambonsche landtaal (bahasa tanah). Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch Indie 24:1–136.
Hopper, Paul J. 1983. Ergative, passive and active in Malay narrative. In Flora Klein-Andreu (ed.), Discourse perspectives on syntax, 67–88. New York: Academic Press.
Hudson, Richard. 1980. Sociolinguistics. Cambridge: Cambridge University Press.
Hughes, John Bernard (Jock). 1987. The Languages of Kei, Tanimbar and Aru: A Lexicostatistic classification. In Soenjono Dardjowidjojo (ed.), Linguistic studies of Indonesian and other languages in Indonesia. NUSA 27:71–111.
Hughes, John Bernard (Jock). 1995. Bilingualism on eastern Kobror Island. ms.
Hyltenstam, Kenneth. 1995. The code-switching behaviour of adults with language disorders—with special reference to aphasia and dementia. In Milroy and Muysken 1995a, 302–343.
Hymes, Dell. 1972. On communicative competence. Sociolinguistics: In J. B. Pride and Janet Holmes (eds.), Selected readings, 269–293. Harmondsworth: Penguin.
Jackendoff, Ray. 1983. Semantics and cognition. Cambridge: MIT Press.
Joshi, A. 1985. Processing of sentences with intrasentential code-switching. In D. R. Dowty, L. Karttunen, and A. Zwicky (eds.), Natural language parsing, 190–205. Cambridge: Cambridge University Press.
Kamwangamalu, N. M. 1990. Bilingual code-mixing: Syntax and constraints. Revised version of paper presented at African Linguistics Conference, University of Georgia, Athens. April 1990.
Kaswanti Purwo, Bambang. 1988. The categorial system in contemporary Indonesian: Numerals. In Bambang Kaswanti Purwo (ed.), Towards a description of contemporary Indonesian: Preliminary studies, Part III. NUSA 30:65–90.
Köppe, Regina, and Jürgen M. Meisel. 1995. Code-switching in bilingual first language acquisition. In Milroy and Muysken 1995a, 276–301.
Kruger, Th. Muller. 1959. Sejarah gereja di Indonesia. Jakarta: Badan Penerbit Kristen.
Labov, William. 1966. The social stratification of English in New York City. Washington: Center for Applied Linguistics.
Lambert, Wallace E. 1967. A social psychology of bilingualism. Journal of Social Issues 23(2):91–109.
Lambert, Wallace E., R. C. Hodgson, Robert C. Gardner, and S. Fillenbaum. 1960. Evaluational reactions to spoken languages. Journal of Abnormal and Social Psychology 60:44–51.

LePage, Robert B. 1968. Problems of description in multilingual communities. Transactions of the Philological Society (1968), 189–212. Oxford: Blackwell.
LePage, Robert B. 1978. 'Projection, focussing, diffusion' or, steps towards a sociolinguistic theory of language, illustrated from the sociolinguistic survey of multilingual communities, Stages I: Cayo District, Belize (formerly British Honduras) and II: St. Lucia. School of Education, St. Augustine, Trinidad: Society for Caribbean Linguistics Occasional Paper No. 9.
LePage, Robert B. 1989. What is a language? York Papers in Linguistics 13:9–24.
LePage, Robert B, and Andrée Tabouret-Keller. 1985. Acts of identity: Creole-based approaches to language and ethnicity. Cambridge: Cambridge University Press.
Li, David C. S. 2001. L2 Lexis in L1: Reluctance to translate out of concern for referential meaning. Multilingua 20(1):1–26.
Litamahuputty, Betty. 1994. The use of biking and kasi in Ambonese Malay. Cakalele 5:11–31.
Macdonald, R. Ross. 1976. Indonesian reference grammar. Washington, D.C.: Georgetown University Press.
Mackey, William F. 1965. Bilingual interference: Its analysis and measurement. The Journal of Communication 15:239–249.
Mackey, William F. 1988. Bilingualism and multilingualism. In Ammon et al. 1988, 699–713.
Mæhlum, Brit. 1990. Codeswitching in Hemnesberget-myth or reality? In Ernst Håkon Jahr and Ove Lorentz (eds.), Tromsø linguistics in the eighties. Tromsø Studies in Linguistics 11, 338–355. Oslo: Novus Press.
Martin-Jones, Marilyn. 1995. Code-switching in the classroom: Two decades of research. In Milroy and Muysken 1995a, 90–111.
Milroy, Lesley. 1987. Language and social networks. Second edition. Oxford and New York: Blackwell.
Milroy, Lesley, and Li Wei. 1995. A social network approach to code-switching. In Milroy and Muysken 1995a, 136–157.
Milroy, Lesley, and Pieter Muysken, eds. 1995a. One speaker, two languages: Cross-disciplinary perspectives on code-switching. Cambridge: Cambridge University Press.
Milroy, Lesley, and Pieter Muysken, (eds.). 1995b. Introduction: Code-switching and bilingualism research. In Milroy and Muysken 1995a, 1–14.
Minde, Don van. 1990. The sound system of Ambonese Malay. Cakalele 1:61–98.

Minde, Don van. 1997. Malayu Ambong: Phonology, morphology, syntax. CNWS Publications, vol. 61. Leiden: Research School CNWS.
Mioni, Alberto M. 1988. Domain. In Ammon et al. 1988, 170–179.
Moeliono, Anton M., and Soenjono Dardjowidjojo. 1988. Tata Bahasa Baku Bahasa Indonesia. Jakarta: Balai Pustaka, Departemen Pendidikan dan Kebudayaan.
Moravcsik, Edith A. 1978. Language contact. In Joseph H. Greenberg (ed.), Universals of human language 1:93–122. Stanford: Stanford University Press.
Mustakim. 1994. Interferensi Bahasa Jawa Dalam Surat Kabar Berbahasa Indonesia. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa,
Muysken, Pieter. 1987. Neutrality in code-mixing. Eigen en Vreemd: Handelingen van het 39e Filologencongres Amsterdam, 359–73. Amsterdam: Vrije Universiteit.
Muysken, Pieter. 1995. Code-switching and grammatical theory. In Milroy and Muysken 1995a, 177–198.
Myers-Scotton, Carol. 1993a. Social motivations for codeswitching: Evidence from Africa. Oxford: Clarendon Press.
Myers-Scotton, Carol. 1993b. Duelling languages: Grammatical structure in codeswitching. Oxford: Clarendon Press.
Myers-Scotton, Carol. 1995. A lexically based model of code-switching. In Milroy and Muysken 1995a, 233–256.
Nababan, P. W. J. 1982. Indonesia. In Richard B. Noss (ed.), Language teaching issues in multilingual environments in Southeast Asia. RELC Anthology Series 10, 1–47. Singapore: SEAMEO Regional Language Centre.
Nababan, P. W. J. 1985. Bilingualism in Indonesia: Ethnic language maintenance and the spread of the national language. Southeast Asian Journal of Social Science 13(1):1–18.
Nivens, Richard. 1987. A preliminary sociolinguistic survey of Tarangan island. ms.
Nivens, Richard. 1989. A sketch of West Tarangan grammar. ms.
Nivens, Richard. 1991a. Toward reconstructing an immediate ancestor of the languages of Aru. ms.
Nivens, Richard. 1991b. The prefix *r-* in West Tarangan: Lexicase and relational grammar analyses. ms.
Nivens, Richard. 1992a. A lexical phonology of West Tarangan. In Donald A. Burquest and Wyn D. Laidig (eds.), Phonological studies in four languages of Maluku, 127–227. Arlington: Summer Institute of Linguistics and the University of Texas at Arlington.
Nivens, Richard. 1992b. A sketch of West Tarangan morphology. ms.
Nivens, Richard. 1992c. An overview of natural phonological processes in West Tarangan. ms.

Nivens, Richard. 1993a. Reduplication in four dialects of West Tarangan. Oceanic Linguistics 32(2):353–388.
Nivens, Richard. 1993b. West Tarangan: pronominals, transitivity, ergativity, and auxiliaries. ms.
Nivens, Richard. 1994. Written Ambonese Malay from 1895 to 1992. Cakalele 5:33–75.
Nivens, Richard. 1996. Dialects or dynamic ways of speaking? Evidence from West Tarangan, Aru. ms.
Nivens, Richard. 1998. Review of One speaker, two languages by Lesley Milroy and Pieter Muysken, eds. Notes on Linguistics 80:46–52.
Nivens, Susan. 1990. A sketch of kinship and marriage in West Tarangan, Aru. Workpapers in Indonesian Languages and Cultures 8:81–93.
Nortier, Jacomine. 1990. Dutch-Moroccan Arabic code switching. Dordrecht: Foris.
Pawley, Andrew, and Frances Hodgetts Syder. 1983. Language and communication, ed. by Jack C. Richards and Richard W. Schmidt, 191–226. London: Longman.
Pfaff, Carol W. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. Language 55(2):291–318.
Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. Linguistics 18(7/8):581–618.
Poplack, Shana. 1987. Contrasting patterns of code-switching in two communities. In E. Wande, J. Anward, B. Nordberg, L. Steensland, and M. Thelander (eds.), Aspects of Multilingualism: Proceedings from the Fourth Nordic Symposium on Bilingualism, 1984, 51–77. Sweden: University of Uppsala.
Poplack, Shana, and Marjory Meechan. 1995. Patterns of language mixture: Nominal structures in Wolof-French and Fongbe-French bilingual discourse. In Milroy and Muysken 1995a, 199–232.
Poplack, Shana, and David Sankoff. 1984. Borrowing: The synchrony of integration. Linguistics 22:99–135.
Poplack, Shana, and David Sankoff, and Christopher Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. Linguistics 26:47–104.
Poplack, Shana, Susan Wheeler, and Anneli Westwood. 1989. Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. In K. Hyltenstam and L. K. Obler (eds.), Bilingualism across the lifespan: Aspects of acquisition, maturity and loss, 132–154. Cambridge: Cambridge University Press.

Preston, Dennis R. 1988. Domain-, role- or network specific use of language. In Ammon et al., 690–699.
Pulau Pulau Aru Dalam Angka 1996. 1996. Ambon, Dobo: Mantri Statistik Kecamatan Pulau Pulau Aru.
Rafferty, Ellen. 1982. Discourse structure of the Chinese Indonesian of Malang. Linguistic Studies of Indonesian and Languages in Indonesia. NUSA 12.
Reyes, R. 1976. Language mixing in Chicano bilingual speech. In J. D. Bowen and J. Ornstein (eds), Studies in southwest Spanish, 182–188. Rowley: Newbury House.
Sankoff, David, and Shana Poplack. 1981. A formal grammar for code-switching. Papers in Linguistics: International Journal of Human Communication 14(1):3–45.
Scotton, Carol Myers. 1988. Code switching and types of multilingual communities. In Peter H. Lowenberg (ed.), Georgetown University Round Table on Languages and Linguistics 1987, 61–82. Washington, D.C.: Georgetown University Press.
Scotton, Carol Myers, and W. Ury. 1977. Bilingual strategies: The social function of codeswitching. International Journal of the Sociology of Language 13:5–20.
Showalter, Stuart. 2001. The same but different: Language use and attitudes in four communities of Burkina Faso. SIL International Publications in Sociolinguistics 5. Dallas, Texas.
Sneddon, James N. 1996. Indonesian reference grammar. Sydney: Allen and Unwin.
Sneddon, James N. 1997. Indonesian: A comprehensive grammar. London: Routledge.
Sperber, Dan, and Deirdre Wilson. 1986. Relevance: Communication and cognition. Oxford: Blackwell.
Staden, Miriam van. 1998. Where does Malay end and Tidore begin? In Jelle Miedema, Cecilia Odé, and Rien A.C. Dam (eds), Perspectives on the Bird's Head of Irian Jaya, Indonesia: Proceedings of the Conference, Leiden, 13–17 October 1997, 691–716. Amsterdam and Atlanta: Rodopi.
Starosta, Stanley. 1988. The case for lexicase: An outline of lexicase grammatical theory. London: Pinter Publishers.
Starosta, Stanley. 1991. Grammar, perception, and reality. In Robert Blust (ed.), Currents in Pacific linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace. Pacific Linguistics C-117:477–500.
Steinhauer, Hein. 1991. On Malay in eastern Indonesia in the nineteenth century. In Hein Steinhauer (ed.), Papers in Austronesian Linguistics 1. Pacific Linguistics A-81:197–225.

Stokhof, W. A. L., ed. 1982. Holle lists: vocabularies in languages of Indonesia vol.3/3: Central Moluccas: Seram (III), Haruku, Banda, Ambon (I). Materials in languages of Indonesia 15. Pacific Linguistics D-49.

Syahdan. 1996. Sasak-Indonesian codeswitching (bilingualism). Ph. D. dissertation. Tucson: University of Arizona.

Taber, Mark, ed. 1996. Atlas Bahasa Tanah Maluku. Ambon: Pusat Pengkajian dan Pengembangan Maluku Universitas Pattimura and Summer Institute of Linguistics.

Timm, L. A. 1975. Spanish-English code-switching: el porqué y how not to. Romance Philology 28(4):473–482.

Tjia, Johnny. 1992. Some notes on dapa passive in Ambonese Malay. ms.

Tjia, Johnny. 1995. Sketsa tata bahasa Melayu Ambon. ms.

Tjia, Johnny. 1997. Verb serialization in Ambonese Malay. M. A. thesis. University of Oregon, Eugene.

Trudgill, Peter. 1974. The social differentiation of English in Norwich. Cambridge Studies in Linguistics 13. London: Cambridge University Press.

Umat Beragama di Propinsi Maluku Dalam Angka 1992. 1992. Ambon: Kantor Wilayah Departemen Agama Propinsi Maluku.

Verhaar, John W. M. 1988. Syntactic ergativity in contemporary Indonesian. In Richard McGinn (ed.), Studies in Austronesian Linguistics. Southeast Asian Monograph Series 76, 347–384. Athens: Ohio University Press.

Wallace, Alfred R. 1869. The Malay Archipelago: The land of the orang-utan and the bird of paradise: a narrative of travel, with studies of man and nature. London: Macmillan and Co.

Wattimury, E., Ny. A. Haulussy, and J. Pentry. 1995. Sintaksis Bahasa Tarangan. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Weinreich, Uriel. 1953. Languages in contact: Findings and problems. New York: Linguistic Circle of New York.

Whitney, William D. 1875. The life and growth of language: An outline of linguistic science. New York: D. Appleton and Co.

SIL International
Publications in Sociolinguistics

Recent Publications

8. **Borrowing versus code-switching in West Tarangan (Indonesia)** by Richard J. Nivens, 2002.
7. **The dynamics of Sango language spread** by Mark E. Karan, 2001.
6. **K'iche': A study in the sociology of language** by M. Paul Lewis, 2001.
5. **The same but different: Language use and attitudes in four communities of Burkina Faso** by Stuart Showalter, 2001.
4. **Ashéninka stories of change** by Ronald James Anderson, 2001.
3. **Assessing ethnolinguistic vitality: Theory and practice** Gloria Kindell and M. Paul Lewis, eds., 2000.
2. **The early days of sociolinguistics** Christina Bratt Paulston and G. Richard Tucker, eds., 1997
1. **North Sulawesi language survey** by Scott Merrifield and Martinus Sales, 1996.

For further information or a full listing of SIL publications contact:

International Academic Bookstore
Summer Institute of Linguistics
7500 W. Camp Wisdom Road
Dallas, TX 75236-5699

Voice: 972-708-7404
Fax: 972-708-7363
Email: academic_books@sil.org
Internet: http://www.ethnologue.com