Social Lives in Language – Sociolinguistics and multilingual speech communities

Celebrating the work of Gillian Sankoff

EDITED BY

Miriam Meyerhoff
Naomi Nagy

Social Lives in Language –
Sociolinguistics and multilingual speech communities

IMPACT: *Studies in Language and Society*

IMPACT publishes monographs, collective volumes, and text books on topics in sociolinguistics. The scope of the series is broad, with special emphasis on areas such as language planning and language policies; language conflict and language death; language standards and language change; dialectology; diglossia; discourse studies; language and social identity (gender, ethnicity, class, ideology); and history and methods of sociolinguistics.

Volume 24

Social Lives in Language – Sociolinguistics and multilingual speech communities
Celebrating the work of Gillian Sankoff
Edited by Miriam Meyerhoff and Naomi Nagy

Social Lives in Language – Sociolinguistics and multilingual speech communities

Celebrating the work of Gillian Sankoff

Edited by

Miriam Meyerhoff
University of Edinburgh

Naomi Nagy
University of Toronto

For Gillian, of course

Table of contents

**Part I.   Language Ideology: From the speakers, what can we
learn about the language?**

Acknowledgements

Introduction

Social lives in language

Naomi Nagy and Miriam Meyerhoff

1. Why sociolinguistics cares about multilingual speech communities

It is axiomatic in sociolinguistics that language and society are intimately entwined. How people perceive the organisation of society, and their place within it, is realised in often subtle ways through language. The amazingly fine-grained stochastic variability in the surface form of a message may be indicative of socially salient differences among the participants, and how they understand the speech event they are part of. Speakers may present themselves in stances of authority, affection, nurturance and nonchalance, and their interlocutors may or may not agree with them, or may bring different assumptions, or their own presentations of self into the equation.

It is common that the set of linguistic resources speakers bring with them when they interact with other people include a range of languages. That is to say, most of the world is made up of multilingual individuals and most speech communities can be characterised as multilingual. Arguably, therefore, most sociolinguistics happens in contexts of language contact. Yet the sociolinguistics of language contact represents only a small fraction of sociolinguistic research (a point we explore further below).[1]

In the field of language contact, many researchers are unaware of the principles underlying the methods used by most sociolinguists and the theoretical questions of concern to them. Of course, anthropologists, as opposed to linguists, are more likely to be interested in engaging with the social politics of language and ideologies of language, and these topics are immediately to the fore in multilingual communities. So, far from shying away from work in multilingual communities, anthropological linguists are likely to seek them out and embrace the challenges in them (see for example work by Niko Besnier, Peter Garrett, Jane Hill, Don Kulick, Miki Makihara).

---

1. This is not intended to minimise the considerable contributions to sociolinguistics made by researchers who *have* embraced multilingual speech communities as sites of investigation. We pay tribute especially to the work of Robert Bayley, Jack Chambers, Sylvie Dubois, Monica Heller, Lesley Milroy, Carol Myers-Scotton, Shana Poplack, Robin Queen, John Rickford and Ron and Suzie Scollon. But relatively few papers devoted to language contact are presented at major sociolinguistics conferences such as NWAV and the Sociolinguistics Symposia.

And yet, as the papers in this volume show, it is clear that there is an energetic and creative cohort of researchers working in multilingual contexts who nevertheless have been chipping away at fundamentally sociolinguistic questions in the fields of linguistics and anthropology. Not all of them are represented in this volume, by any means, but in gathering these papers together in one volume, we hope that readers will be able to gain a clearer focus on what kinds of questions and issues unite and separate those of us working primarily on data from monolingual speech communities and on data from multilingual speech communities.

The purpose of this introduction is to lay out a few of the major issues that have occurred to us, as editors, in drawing this volume together, and that have been shaping our perceptions about where we might go next. As the title will have already suggested to some readers, its function is also to recognise the importance of a 1980 collection of articles by Gillian Sankoff, which anticipated much of the ground being covered in the different chapters of this volume. In the introduction to her 1980 book, Sankoff challenged the assumption that social factors serve merely to provide a setting or context for language use, instead arguing that social and interactional considerations specific to the verbal exchange of a small number of co-present participants may be absolutely central in shaping the structure of language (1980: xviii). In the next two sections, we review general issues that will contextualise the work on sociolinguistics in multilingual speech communities. In the final section, we review the connections with Sankoff's earlier work, and explain the choice of title for the volume. We will suggest that it remains true today that "the most challenging problems in contemporary sociolinguistics involve putting together the two levels in the society-language relationship" (Sankoff 1980: xxii), but we will also suggest that the papers in the volume represent important and meaningful steps towards the goal Sankoff expressed so concisely more than a quarter of a century ago.

2.   The curious monolingual bias of sociolinguistics

It is extremely unlikely that sociolinguistics students or fledgling researchers have ever been instructed to focus their attention on monolingual speech communities and to ignore multilingual speech communities. Nevertheless, after some forty years of sociolinguistics as a recognised sub-field of linguistic inquiry, the vast majority of sociolinguistic work focuses on monolingual communities. The iconic social dialect studies of the 1960s and 1970s, that is, New York City (Labov 1972), Norwich (Trudgill 1974), and Belfast (Milroy 1980), explicitly or implicitly excluded non-native speakers of English. While these studies have been much copied, it would be too simplistic to assume that it was their model alone which has led to the imbalance in the distribution of sociolinguistic research on multilingual and

monolingual communities. At the same time as the social dialectology of speech communities was being codified into an autonomous and respectable field of linguistic research, linguistic anthropologists continued their research into communities where this monolingual bias was not an issue. As this collection of papers shows, the connections between sociolinguistics and the study of multilingual communities continues to be maintained and nurtured in many cases by researchers whose academic "homes" are in anthropology departments.

There are important practical differences in researching the sociolinguistics of multilingual speech communities and speech communities that are monolingual (but polydialectal or polystylistic). In the field of language contact, it is common to distinguish between the outcomes of mutually intelligible varieties (*dialect levelling* and *koineisation*) and the outcomes of contact between mutually unintelligible varieties (*creolisation* and *externally-motivated change*). Whether there are, in fact, empirical differences between the processes underlying the outcomes of these different contexts of contact and change, and whether there are typologically distinct outcomes from them, remains an open question. However, it is important to recognise that many linguists share strong and honestly-held perceptions that language contact is indeed qualitatively different from dialect contact. This in turn feeds a perception that, until we have answered the questions about the similarities or differences between the underlying processes and eventual outcomes of dialect and language contact, it is prudent for linguists to keep the different contexts of contact apart.

We incline to a somewhat more *im*prudent approach. Like many of the contributors to this volume, we suspect that the study of multilingual speech communities offers insights into social and linguistic processes which are likely to be of broader relevance to sociolinguistics (see contributions by Schieffelin, Jourdan, Daveluy, Mesthrie, Meakins, Thibault and Sankoff for a focus on how the study of multilingual speech communities may shed light on social dynamics; Blondeau, Auger, Meyerhoff, Blondeau and Nagy, Labov and – especially – King for a focus on the broader linguistic insights offered by the study of multilingual speech communities).

We believe that the roots of these perceptions about the differences between the sociolinguistic study of monolingual and multilingual speech communities lie in linguists' conceptions about what the most appropriate linguistic questions are to ask, and what constitutes valid answers to them. In the next section, we will consider some of these empirical and theoretical divergences.

3.   Questions asked and questions answered

Fields of academic inquiry are defined by the kinds of questions they consider meaningful and by what kinds of answers they admit as meaningful to their

questions. The kinds of questions that define sociolinguistics have been shaped by how sociolinguistics conceptualises itself in relation to other sub-fields of linguistics and to sister disciplines such as anthropology and sociology.

3.1   Historical linguistics and sociolinguistics

The study of language variation, for example, is tied closely to the study of language change. Perhaps because of this connection to historical linguistics, sociolinguists have willingly embraced a distinction (common in historical linguistics) between changes that are motivated by language internal factors (those particular to the structure of a language itself) and external factors (factors introduced to a language through contact with other languages). The distinction between internal and external factors here is potentially confusing, since in more recent sociolinguistic work (Labov 1994, 2001), the denotation of *external* factors has shifted slightly so that in sociolinguistics, it now refers specifically to the social factors influencing synchronic variation and change (and these may have nothing to do with language contact).

Nevertheless, both the sociolinguistic and historical linguistics uses of *external* factors do have something important in common. Underlyingly, both hold that:

i.   it is possible to isolate structural factors internal to a given language's system as the object of study, and
ii. a focus on internal factors in variation and change somehow provides a more pristine example of linguistic research (see also discussion in Meyerhoff 2006). In fact, the very labels of the terms suggest that the *internal* factors are more central, while *external* ones are peripheral.

However, among historical linguists who take language contact seriously, both these underlying assumptions are increasingly being questioned. For example, Thomason (2001), Matras (in press) and most of the participants at the 2007 Paris workshop on language contact (Leglise and Chamoreau forthcoming) are inclined to treat the internal/external distinction as otiose. This presents a welcome opportunity: without forfeiting the well-established relationship between sociolinguists' study of synchronic variation and historical linguists' study of diachronic change, the dialogue between socio- and historical linguistics can be extended productively to incorporate multilingual speech communities (Chambers 2003 proposes a typology for dealing with different kinds of multilingual speech communities in variationist sociolinguistics). Some of the quantitative papers in this volume make a direct contribution to establishing both sound methods and sound generalisations that facilitate this continued dialogue.

3.2   The sociolinguistic variable and multilingual/contact linguistics

Within the field of sociolinguistics, too, we find conventions that militate against the study of variation in multilingual speech communities. Specifically, the notion of the sociolinguistic variable itself. As all variationists know, the quantitative analysis of variation requires the researcher to first identify variants that are semantically (or, some would argue, functionally) equivalent, and then explore the (linguistic or social) constraints on the distribution of those variants. The requirement for equivalence may be rather hard to satisfy when you are analysing variation that involves speakers' alternations and selections of variants across different languages. The envelope of variation for a particular variable is often framed in terms of the context: what words or lexical categories immediately precede or follow the element in question. Clearly, it is more difficult to establish a valid description of the envelope of variation in these terms when more than one language is at play.

Various scholars have attempted to address this problem, either from a specifically variationist perspective or from a more general perspective on what happens when speakers have different linguistic systems in play at the same time. The problem is noted in Weinreich (1966, especially chapter 3), and is the central concern of Myers-Scotton's Matrix Language Frame model (1993a, 1993b) of language contact and bilingualism; Mahootian (2006) provides a useful summary of different approaches to formalising the linguistic relationship between multilingual speakers' repertoires. Meechan and Poplack (1995) explore the potential for linguists to work "up" from the specific details of variation itself, instead of "down" from theory-internal presumptions about language structure. In some cases, it may be possible for the variation to describe the points of contact between systems. Cumulatively, it is possible that such work might in the end go beyond descriptive adequacy and offer principled generalisations about language contact and variation (cf. King, this volume).

It is also possible that the problem of defining variables and their variants in multilingual speech communities might prove to be a will o' the wisp. Because the problem is itself an artefact of theory-internal assumptions in sociolinguistics, alternative approaches to studying how multilingual speakers manage their linguistic resources may render it obsolete. Some researchers on multilingualism and language contact consider the "two/three/four languages" model inherent to most if not all of the approaches reviewed above to be fundamentally misguided (e.g., Gafaranga 2007, Matras in press).

In this respect, the work exploring variation in second language (L2) speakers bridges the divide between rather different conceptions of the nature of language, and the methods associated with variationist research have provided opportunities,

not just disadvantages. Sociolinguists tend to adopt a "difference" rather than "deficit" approach when comparing speaker behaviour in first and second language varieties. Many variationists look at L2 speakers as having a system, rather than being in the process of acquiring one.

We also note that even if we find satisfactory ways to address the problem of defining the variable in multilingual speech communities, other challenges present themselves. An attempt to fully explore the sociolinguistics of language contact then needs to engage with other questions: What model(s) of contact should we assume? What are the practical implications of those models in constraining or defining our data collection and our analysis? These questions take us beyond the focus of this particular volume (though Makihara & Schieffelin 2007; Ansaldo et al. 2007, Leglise & Chamoreau forthcoming and Deumert & Durrleman 2006 offer commentary on these questions from anthropological, typological and creolistic viewpoints respectively).

3.3   Practical monolingualism

The last two sections have considered the possibility that the focus on monolingual communities in sociolinguistics may, to some extent, be an artefact of our methods, some of which in turn stem from the research questions we are trying to answer.

However, there is one basic practical limitation to analysis of variation across codes and in situations of contact which must also be considered: it requires researchers themselves to have competency in multiple languages.

We are all aware of the prevalence of monolingualism in the US, where many sociolinguists train. The overall devaluation of the study of foreign languages in the current US educational system has seen even graduate programs in linguistics decrease their requirements for proficiency in multiple languages, thus diminishing the basic ability to conduct research in multilingual communities. It is true that many of us have conducted fieldwork in communities where we did not know some or all of the languages before our inquiry began (a number of the authors in the present volume can be said to be following the successful example of Gillian Sankoff in this respect, too). However, we are not confident that such an approach would be possible for researchers with no prior experience in learning second (third, fourth …) languages.

Furthermore, sociolinguistically interesting situations of multilingualism often involve contact and alternation between internationally dominant languages and socially subordinate ones. Socially subordinate languages aren't usually the ones researchers have the opportunity to study at school or university. So even where academic traditions still include a sufficient component of foreign language

study, the kinds of the languages available for study may be mismatched to the needs of the field.

It is much harder to offer solutions or resolutions to this practical issue than it is to any of the more abstract theoretical issues reviewed above. It is, in all likelihood, a far greater stumbling block to the kind of vital sociolinguistics of multilingual speech communities that we have tried to sketch. Solutions to these problems will have to start long before graduate school, if they are to have a practical impact on the field. Increasingly, linguists are aware of and participating actively in outreach and educational work – they undertake workshops in elementary or high schools (primary or secondary) and craft enticing and rewarding introductory university courses (cf. Adger et al. 2007; Denham & Lobeck 2007; Hazen 2008; Reaser & Adger 2007, Reaser & Wolfram 2005). These are good places to introduce the sociolinguistic dynamics of multilingual speech communities, particularly as the lived world of children in most major cities today is a multilingual speech community. In addition, more practical consultancy work (e.g. advising parents bringing up bi-/multilingual children and exposing common misconceptions about multilingualism) may also encourage new generations of sociolinguistics researchers who take language learning for granted and who have grown up thinking about the sometimes difficult social and linguistic questions that living in a multilingual speech community raises.

In the next section, we examine the extent of this research bias in favour of monolingualism, and provide quantitative evidence in support of our claim that research on multilingual speech communities makes up only a small subset of the work done in the field.

4.   The monolingual bias in quantitative perspective

The dearth of sociolinguistic work in multilingual communities is easily illustrated by examining publication trends. For this purpose, we surveyed two leading sociolinguistic journals, *Language Variation and Change (LVC)* and *Journal of Sociolinguistics (JSL)*, including a sample of articles running through the course of the publication history of each. Articles were sorted into those that examined only one language versus those that examined more than one language.[2] The data for

---

2.   Some articles that examine more than one language are still focusing on monolingual speakers, or at least speakers who are represented monolingually as far as the research collection is concerned. That is, an article may look at how speakers of language A use a certain construction and then at how speakers of language B use a related construction, and will be classified as multilingual in our tally. This method therefore overestimates the rate of occurrence of articles that truly conduct sociolinguistic analysis of multilingual systems.

LVC was gathered by sampling issue two of each volume from 1989 (volume 1) to 2007 (volume 19), for a total of 96 articles in 19 issues. *JSL* was counted exhaustively – all articles published from 1997 (volume 1, issue 1) to 2008 (volume 12, issue 1) were surveyed, for a total of 194 articles.

We see two things. First, the overall rate of publication of multilingual studies is surprisingly low compared to the number of multilingual people and communities on this planet, estimated conservatively at over 50% (Tucker 1999). Overall, 11% of the articles published in *LVC* engage in analysis of more than one language and 28% published in *JSL* do so. These ratios are statistically quite distinct from the 50% (or more) of the world's population which is multilingual.[3] Another way to look at it is that each article in *LVC* examines an average of 1.2 languages, on average, while *JSL* averages 1.5. Given that most individuals live in a community where it is common to use more than one language for one's communicative needs and identity construction, this is oddly at variance with the real world.

Given the regional base of *LVC* in the USA and Canada and the base of *JSL* in the UK and the Pacific, the discrepancy between the journals is perhaps unsurprising. English reigns demographically supreme in the US and Canada. However, we might expect that *JSL,* a journal based in the multilingual zones of the Pacific and Europe, would have more multilingual articles than it in fact does. It seems reasonable to suppose that the hegemony of English in both the Pacific and Europe has had some impact on research trends represented in this journal.

The second, and more worrisome, trend is that there is no sign of improvement in this pattern. Figure 1 illustrates the percentage of multilingual articles published each year in the two journals. The dotted lines are linear trend lines showing the overall decline in annual rate.

Because some parts of the world are more plurilinguistic than others, we also looked at the distribution of where the data for these published articles were collected.

---

3. $\chi^2 = 34$, p < .001. This chi-square calculation compares our observation of the number of articles about monolingual versus multilingual communities (combining our tallies for *LVC* and *JSL*) to an expectation of a fifty-fifty split, representing conservative estimates of how much of the world's population is bilingual:

	Monolingual	Multilingual	Total
observed	213	77	290
expected	145	145	290

(If we instead compared our observation of the monolingual/multilingual ratio within these publications to an estimate of three billion out of six billion for the actual population of the world, the significance level would be far higher.) These statistics were calculated using Preacher's (2001) chi-square calculator.
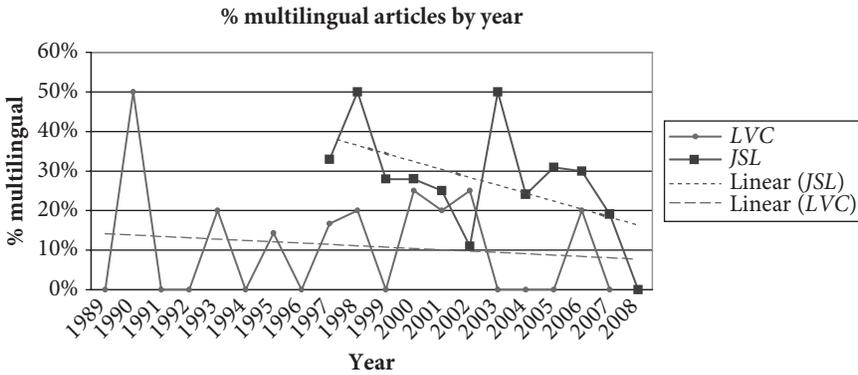
Figure 1. Percentage of multilingual articles by year in two sociolinguistics journals.

Not surprisingly, a large fraction of them report on linguistic situations in the United States, which is known for having a predominantly monolingual population. Both journals publish far fewer articles reporting from regions of greater multilingualism, such as Africa and Oceania. Figures 2 and 3 illustrate these trends for the two journals respectively.
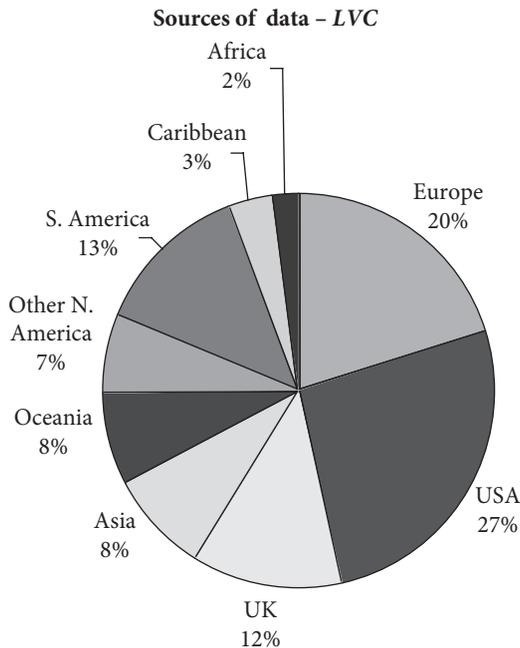
Figure 2. Place of residence of speakers analyzed in 96 *Language Variation and Change* articles.

Sources of data – *JSL*
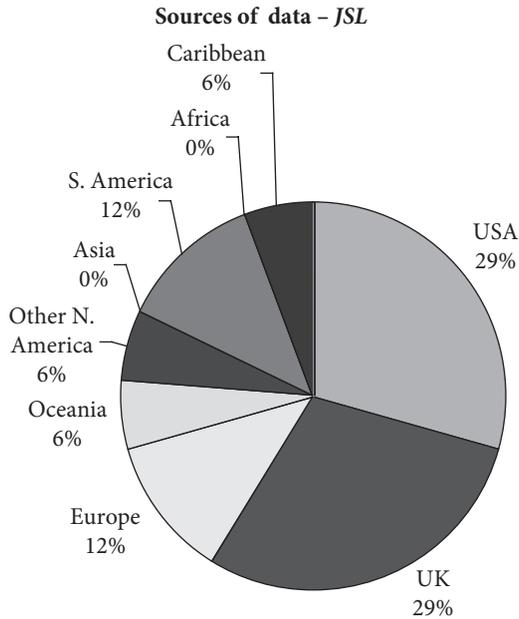
Figure 3. Place of residence of speakers analyzed in 194 *Journal of Sociolinguistics* articles.

These trends provide a further motivation for linguists to increase our efforts to make our research findings accessible to the public. Realistically, a change to the trend documented above is only likely to occur when there has been a change in attitudes and policies regarding the learning of second (and third …) languages in the US and UK, which dominate these publications. Such changes in attitudes are more likely to prepare researchers for work on home-grown contexts of multilingualism, but also in the common, but commonly ignored, language contact situations that involve lesser-known languages.

By extension, these tallies also support the need for an increase in sociolinguistic work on multilingual speech communities – most of the world's people do use a repertoire of more than one language to meet the full range of their communicative needs. To consider single languages in isolation in scholarly work when they don't exist that way in their social lives is to reify idealisations about the discreteness of language systems and the norm of monolingualism (see related discussion in Ansaldo forthcoming). As the figures above indicate, an expansion of research into more multilingual parts of the world could do much to overcome this. The papers collected in this volume hint at the richness of the data and the rewarding generalisations to be gained from such work. They also pay tribute to the innovative path cut out by Gillian Sankoff, the anthropologist and linguist who had the foresight to identify this richness, and outline its potential for linguistics in her 1980 book, *The Social Life of Language*.

5.   A highly social life in language: Gillian Sankoff's contribution
to the sociolinguistics of multilingual speech communities

The papers in this collection exemplify Gillian Sankoff's influence in the field of linguistics. As has already been made clear, the focus in this volume is the sociolinguistic analysis of multilingual communities and situations. The collected papers demonstrate the continued relevance of the issues and priorities Sankoff identified for the study of language in society in 1980. Each of them illustrates her efforts to cross many boundaries that others consider(ed) sacrosanct in her search for the best understanding of how we use language. She looks across divisions between languages, between different levels of linguistics, and between first and second language users. In 1980, she observed that "many linguists still find it difficult to see how … the "social world" is relevant to the internal structure of language" (Sankoff 1980: xix). Today, in large part due to the work of Sankoff and her colleagues, far fewer linguists would be described that way and the roles of individual experience and social context are more likely to be part of the central focus in linguistic research. In other words, people are finding many ways of "putting together the two levels in the society-language relationship" (*ibid*.: xxii); an array of these approaches are collected in this book.

While Sankoff's research has had implications for linguistic study all over the world, her own research has been principally situated in two sites: Papua New Guinea (specifically Morobe Province) and Montreal. Her important contributions to the study of language in society in both locations are reflected in this volume, where the majority of the articles focus on communities in the Pacific Islands and French-speaking Canada. However, geographically, the articles collected here circumnavigate the globe, with ports of call in North America (Blondeau, Blondeau & Nagy, Daveluy, King, Labov and Thibault), Oceania (Jourdan, Meakins, Meyerhoff, D. Sankoff, Schieffelin), Africa (Mesthrie), and Europe (Auger & Villeneuve and Labov). This collection begins to redress the geographic imbalance of the major journal articles described above, as illustrated in Figure 4. We include Figure 4 in order to illustrate Gillian Sankoff's influence. This collection of studies of multilingual communities by her students and colleagues, not surprisingly, shows a much higher rate of representation of Oceanic and "Other North American" (in this case, Canadian) communities.

That Sankoff's work continues to influence colleagues and former students working in this atypical distribution of localities is a tribute to her unusual ability to maintain traditions of work in two very different places: Montreal, giving rise to a large body of work examining the many different ways that French and English interact in Canada, and Oceania, where work explores the interaction between Oceanic languages, regional lingua francas and the colonial languages of English and French.

**Sources of data –
this volume**

Caribbean
0%

USA
0%

Africa
7%

Europe
14%

Asia
0%

UK
0%

Other N.
America
43%

Oceania
36%

S. America
0%
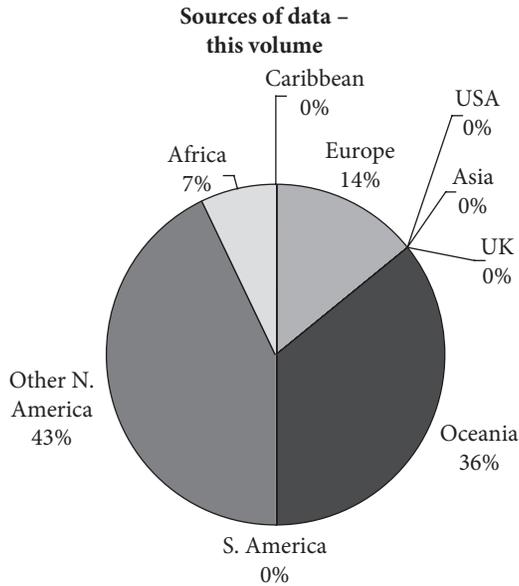
Figure 4. Place of residence of speakers analysed in this volume.

Sankoff's training was originally in anthropology and her Ph.D. at McGill University was supervised by Richard Salisbury, himself a renowned researcher in Papua New Guinea. As a graduate student, she was immersed in the friendly and close-knit atmosphere of McGill's Sociology and Anthropology department, but her own hard work, scientific curiosity, courage and intelligence propelled her to considerable early achievements. A turning point, which Sankoff herself has acknowledged over the years, was attending the Linguistic Society of America's 1964 Summer Institute in Bloomington, Indiana. There, she was able to build on both the descriptive linguistics training she had acquired at McGill and her excellent social sciences background. Sankoff's career ever since has been characterised by a duality of rigour and creativity: she has focused on locally relevant social facts as the basis for linguistic analysis, and she has grounded sociolinguistic generalisations in the complexities of empirical linguistic facts. Friends of Sankoff's at the time she was a student say that with hindsight it is clear that by the mid-sixties, she had had her first important insights into the ways in which linguistics and the social sciences could be studied to their mutual benefit. In the first years of her appointment at the Université de Montréal (c. 1969–1974), there was close collaboration between Sankoff and Henrietta Cedergren, Suzanne Laberge, Diane Vincent, Pierrette Thibault, David Sankoff and others. In terms of the development of the field, as well as of the individuals themselves, this was a remarkably intense period.

It saw the invention of (computer) corpus-based sociolinguistics, and (among other things) a deepening and sophistication in how researchers understood the relationship between language and the social order, and between language (as a system) and real time change, whether in communities or individuals. Sankoff was thinking sociolinguistics long before she (and most other people) had heard the word.

It is probably not a coincidence that this interest in both social and linguistic insights saw light in research on multilingual speech communities. Sankoff's work has demonstrated the wealth of speaker-centred and hearer-centred sociolinguistic data that is offered by the taken-for-granted exchanges of everyday life. Her contributions to research on contact linguistics have strengthened the connections between social anthropologists and linguists working on grammaticalisation and language change.

In this volume, several chapters explore contact between speakers of different language varieties. Some focus on creolised or mixed languages (Jourdan, Meakins, Meyerhoff, D. Sankoff, Schieffelin, Mesthrie), some on variation between and among first and second language users (Blondeau & Nagy, Daveluy), and some focus on variation between local and standard varieties (Auger & Villeneuve, Thibault). Appropriately, given Sankoff's contribution to our understanding of the subtle ways in which substrate, lexifier and cognitive universals interact in the formation and development of contact languages, the role of the substrate in a variety of contact situations is central to a number of the chapters (Labov, King, Meakins, Meyerhoff, Schieffelin). This unusual collection therefore constitutes an innovation in sociolinguistic research publications, as such topics are normally relegated to separate venues.

Similarly, following other paths that Sankoff blazed, chapters in this volume explore and explicitly draw together dimensions of language that are often relegated to the distinct domains of phonology, morphology, syntax, and pragmatics. Sankoff's synthesis of diverse linguistic and social factors in the analysis of variation is developed and thoughtfully explored in several chapters. These accept the need to simultaneously consider several, if not all, of the levels which are traditionally considered distinct. This is most explicitly explored in Meyerhoff's chapter but is also apparent in the consideration of factors from a number of different domains in the papers by Auger & Villeneuve, Blondeau, Blondeau & Nagy, and King.

The title of this volume pays tribute to the importance of Sankoff's 1980 work, *The Social Life of Language*, and several of the themes explored in that monograph are echoed in this collection. We have chosen this as a means to organise the book into sections.

The first section consists of five chapters (Daveluy, Jourdan, Meakins, Mesthrie, Schieffelin) which are more or less concerned with language ideologies or attitudes towards language. These chapters challenge the assumption that social factors

serve merely to provide a setting or context for language usage (cf. Sankoff 1980: xviii). They focus on how people in different kinds of communities understand the place of language in their lives and how people use language in talking about the social world (*ibid.*: xix).

The second section is made up of three chapters which bridge the first and third sections. The work in these chapters (King, D. Sankoff, Thibault) examines sociolinguistic issues both from top-down and bottom-up perspectives, intertwining macro- and micro-linguistic approaches. These chapters illustrate how particular aspects of particular social systems, resulting from particular historical forces, have shaped particular languages (*ibid.*: xx).

The third and final section collects papers that work within a micro-sociolinguistic, quantitative paradigm (Auger & Villeneuve, Blondeau, Blondeau & Nagy, Labov, Meyerhoff). These papers illustrate the variationist approach, in which performance is treated as a sample of the forms that could be generated by grammatical rules. The authors of these chapters share with Sankoff (1980: xviii) the conviction that this approach better matches linguistic reality than a linguist's or speaker's intuitions can, and that by examining language in its natural context of use we may better understand its structure.

We are proud to be able to include works by colleagues at the Université de Montréal, including one of Gillian's first students (Pierrette Thibault), members of her more recent research teams (Hélène Blondeau, Michelle Daveluy, Naomi Nagy), an observer of her anthropological fieldwork in New Guineau (David Sankoff), a fellow contributor to quantitative studies (William Labov), researchers exploring the sociolinguistics of non-western cultures (Christine Jourdan, Felicity Meakins, Raj Mesthrie, Miriam Meyerhoff, and Bambi Schieffelin), all work that presupposes fluency in a typologically and geographically diverse range of languages. This collection marks her enduring influence in the field, bringing together scholars who conduct ethnographic and linguistic work in both western and non-western societies, scholars who combine quantitative and qualitative approaches, and scholars who share her commitment to demonstrating that the behaviour of the world's multilingual speakers should be as much a part of linguistic theory and practice as their more closely-scrutinised monolingual cousins.

Far from being a retrospective, the publication of this volume is an opportunity to consider and be inspired by the new directions in which Sankoff's research agenda continues to take the field. The chapters in this volume also stand as a tribute to Sankoff's on-going leadership in the field of sociolinguistics, especially the ground-breaking work she continues to undertake on the acquisition of variation, variation and stability across the lifespan, and theorising the relationship between variation in the group and the individual (Sankoff 2004, 2005, 2006; Sankoff & Wagner 2006; Sankoff & Blondeau 2007). She is an acknowledged expert on all these topics, currently much in demand for the clarity with which she presents

the complex data that underpins her sociolinguistic insights, as well as for her skills in demonstrating the relevance of individual and group patterns of variation to language acquisition, cognitive linguistics, formal linguistics, and – of course – anthropology and sociolinguistics. It is perhaps appropriate to close by highlighting the fact that her most recent research illustrates that, as speakers, we continue to develop our repertoire throughout our lives, as she has shown that we may do as linguists.

Gillian, tankyu tumas blong yu soemoat rod long ol gudgudfala wok ya long yumi mifala evriwan.

Gillian, merci pour ton exemplaire formidable de comment choisir quoi étudier et comment y achever.

Gillian, thanks for providing such a splendid example of what to study and how to go about it doing it.

References

Adger, C., Wolfram, W. & Christian, D. 2007. *Dialects in Schools and Communities,* 2nd Edn. Mahwah NJ: Lawrence Erlbaum.

Ansaldo, U. Forthcoming. *Contact languages. Ecology and evolution.* Cambridge: CUP.

Ansaldo, U., Matthews, S. & Lim, L. 2007. *Deconstructing Creole* [Typological Studies in Language 73]. Amsterdam: John Benjamins.

Chambers, J.K. 2003. Sociolinguistics of immigration. In *Social Dialectology: In honour of Peter Trudgill,* D. Britain & J. Cheshire (Eds), 97–113. Amsterdam: John Benjamins.

Denham, K. & Lobeck, A. 2007. *TeachLing*. http://teachling.wwu.edu.

Deumert, A. & Durlemann, S. (Eds). 2006. *Structure and Variation in Language Contact.* Amsterdam: John Benjamins.

Gafaranga, J. 2007. *Talk in Two Languages*. Basingstoke: Palgrave Macmillan.

Hazen, K. 2008. Variationist approaches to language and education. In *The Encyclopedia of Language and Education,* 2nd Edn. Vol. 10: *Research methods in language and education*, N. Hornberger & K. King (Eds), 85–98. Heidelberg: Springer.

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia PA: University of Pennsylvania Press.

Labov, W. 1994. *Principles of Linguistic Change,* Vol. 1: *Internal factors*. Oxford: Blackwell.

Labov, W. 2001. *Principles of Linguistic Change*, Vol. 2: *Social factors.* Oxford: Blackwell.

Leglise, I. & Chamoreau, C. (Eds). Forthcoming. *Language Contact and Morphosyntactic Variation and Change.*

Mahootian, S. 2006. Code switching and mixing. In *Encyclopedia of Languages and Linguistics,* 2nd Edn, Vol. 2, K. Brown (Ed.), 511–527. Oxford: Elsevier.

Makihara, M. & Shieffelin, B.B. 2007. *Consequences of Contact*: *Language ideologies and sociocultural transformations in Pacific societies.* Oxford: OUP.

Matras, Y. In press. The borrowability of structural categories. In *Grammatical Borrowing in Cross-linguistic Perspective*, Y. Matras & J. Sakel (Eds). Berlin: Mouton de Gruyter.

Meechan, M. & Poplack, S. 1995. Orphan categories in bilingual discourse: Adjectivization strategies in Wolof-French and Fongbe-French. *Language Variation and Change* 7(2): 169–194.

Meyerhoff, M. 2006. But is it really *linguistics*? Breaking down boundaries. *Journal of Sociolinguistics* 7(1): 65–77.

Myers-Scotton, C. 1993a. *Social Motivations for Codeswitching*. Oxford: OUP.

Myers-Scotton, C. 1993b. *Duelling Languages: Grammatical structure in code-switching.* Oxford: Clarendon Press.

Milroy, L. 1980. *Language and Social Networks*. Oxford: Blackwell.

Preacher, K.J. 2001. Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence [Computer software]. Available from http://www.quantpsy.org

Reaser, J. & Wolfram, W. 2005. *Voices of North Carolina Dialect Awareness Curriculum*. http://www.ncsu.edu/linguistics/research_dialecteducation.php

Reaser, J. & Adger, C. 2007. Developing language awareness materials for non-linguists: Lessons learned from the *Do You Speak American?* Curriculum Development Project. *Language and Linguistic Compass* 1(3):155–167.

Sankoff, G. 1980. *The Social Life of Language*. Philadelphia PA: University of Pennsylvania Press.

Sankoff, G. 2004. Adolescents, young adults and the critical period: Two case studies from Seven Up. In *Sociolinguistic Variation*, C. Fought (Ed.), 121–139. Oxford: OUP.

Sankoff, G. 2005. Cross-sectional and longitudinal studies. In *An International Handbook of the Science of Language and Society*, Vol. 2, U. Ammon, N. Dittmar, K.J. Mattheier & P. Trudgill (Eds), 1003–1013. Berlin: de Gruyter.

Sankoff, G. 2006. Age: Apparent time and real time. In *Encyclopedia of Language and Linguistics*, 2nd edn*,* Vol. 1, K. Brown (Ed.), 110–116. Oxford: Elsevier.

Sankoff, G. & Blondeau, H. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83: 560–588.

Sankoff, G. & Wagner, S. 2006. Age grading in retrograde movement: The inflected future in Montréal French. *Penn Working Papers in Linguistics* 12: 203–216.

Thomason, S. 2001. *Language Contact: An Introduction*. Edinburgh: EUP/Georgetown DC: Georgetown University Press.

Trudgill, P. 1974. *The Social Differentiation of English in Norwich*. Cambridge: CUP.

Tucker, G. Richard. 1999. A global perspective on bilingualism and bilingual education. In *Georgetown University Roundtable on Languages and Linguistics 1999. Language in our time: Bilingual education and Official English, Ebonics and Standard English, immigration and the Unz Initiative,* J. Alatis & A. Tan (Eds), 332–340. Washington DC: Center for Applied Linguistics.

Weinreich, U. 1966. *Languages in Contact: Findings and problems.* The Hague: Mouton.

Photos of Gillian

Photo 1. Gillian Sankoff and Dambi Sanik, visitor from Papua New Guinea, with a decorative stone axe, in 1968.
Photo courtesy of David Sankoff. Source: Staff Photo by Pete Brosseau, *The Montreal Star*, February 21, 1969.

Photo 2. Gillian Sankoff in 2008.
(*Photo by Bill Labov.*)

Biographies of contributors
and email addresses

Julie Auger is Associate Professor of linguistics and French linguistics at Indiana University. After meeting Gillian at a conference in Québec City, she decided to leave the comfort of her home town to pursue a Ph.D. at the University of Pennsylvania. Her dissertation on subject doubling in Québec French, written under Gillian's supervision, allowed her to discover Picard, a regional language of France which bears striking similarities to Québec French but also contains many different and intriguing features. Given the precarious situation of Picard and the lack of attention it received (and is still receiving) among linguists, Julie decided to conduct fieldwork in Vimeu, France and focus a significant part of her research efforts on the phonological and morphosyntactic structure of Picard. Like Gillian, she maintains a strong interest in the structure of Québec French and continues to investigate various features of her native variety.

jauger@indiana.edu

Hélène Blondeau is Assistant Professor of French and Linguistics at the University of Florida. In the mid-nineties, she was involved in a research project led by Gillian Sankoff and Pierrette Thibault on the linguistic repertoire of bilingual Anglophones in Montreal. After completing her own longitudinal research on pronominal variation in Montreal French, she spent the 1999–2000 academic year as a Visiting Scholar in the Department of Linguistics at the University of Pennsylvania where she benefited from Gillian's supervision for her postdoctoral research. In this context, she had the chance to embark with Gillian on the longitudinal study of R in Montreal French trying to understand linguistic behavior over the course of the lifespan. She is currently investigating the language practices of the francophone community of Quebecers in Florida, a community that has been neglected in the large body of research devoted to the situation of French in North America.

blondeau@rll.ufl.edu

Michelle Daveluy is Professor of Anthropology at the University of Alberta. Trained in Montréal, she had the good fortune to be under the influence of Gillian Sankoff early on in her career. As a graduate student, she relocated and re-interviewed many

of the Montrealers originally met in 1971, for the Sankfoff-Cedergren corpus. The year Gillian Sankoff served as assessor for her Ph.D. thesis on French spoken by both Franco- and Anglo-Montrealers, she co-edited a thematic issue of the journal *Culture* (XIV-2) on language variation in Montreal. When she was looking for a textbook to use for her first academic position in Halifax, Nova Scotia, Gillian came up with the suggestion that made teaching challenging enough to be stimulating for her students as well as for herself. Gillian remains an inspiration to her, in particular when tackling aspects of language dynamics that are less often studied, including among multilingual Inuit of the circumpolar world.

michelle.daveluy@ualberta.ca

Christine Jourdan is Professor of Anthropology at Concordia University in Montreal. Her first contact with pidgins and creoles was in Gillian Sankoff's undergraduate class on the Pacific at the Université de Montréal. She went on to write a Ph.D. at the Australian National University in Canberra on the creolization of Solomon Islands Pijin that was centrally inspired by Gillian's work on Tok Pisin in New Guinea. Playing with issues of cultural creolization, multilingualism, the development of urban identities in the Solomon Islands, Jourdan has written on language policy, language and nationalism, language and self, in addition to writing a dictionary of Pijin and a book on the sociolinguistic history of Pijin. Recently she has been working on the linguistic life course and on language ideology among urban middle-class Solomon Islanders.

jourdan@vax2.concordia.ca

Ruth King is Professor of Linguistics and Women's Studies at York University, Toronto, Canada. When she was a student, the first non-textbook linguistics book she bought was Gillian's *The Social Life of Language*. She was inspired by the depth and breadth of Gillian's research contributions and by the fact that another Anglophone had specialized in the study of Canadian French. King has spent a lot of her career working on the grammatical structure of Atlantic Canada Acadian French and related varieties, research which attempts to integrate formal linguistics and sociolinguistics. She also works on general issues around language change, including the process by which language contact leads to linguistic change. In addition, she has a longstanding interest in the relationship between language and social identity, particularly in the field of language and gender studies. Gillian's encouragement and support since their first meeting at NWAV in 1993 has been enormous.

rking@yorku.ca

Bill Labov lays claim to be the other half of Gillian Sankoff. He is the other one of two Professors in the Sociolinguistics Program at the University of Pennsylvania where they have jointly taught many courses since 1977. He has been married to the said Gillian Sankoff since 1993. They jointly inhabit a house in Center City, Philadelphia, where they have raised with great delight and success two children, Alice Goffman and Rebecca Labov. Labov has benefited immensely from Sankoff's carefully constructed study of the Montreal French community and her expertise on far-flung societies, embracing pidgins, creoles, and other contact situations. In his own work, he frequently draws from her current research into language change across the lifespan. He admires but cannot approach her native-like command of many tongues.

wlabov@cis.upenn.edu

Felicity Meakins is a postdoctoral fellow with the University of Manchester, working with Gurindji and Bilinarra people on a DOBES (Documentation of Endangered Languages program funded by the Volkswagen Foundation) project in northern Australia. Her main research interests lie in contact languages, particularly Gurindji Kriol, a north Australian mixed language. In 2007 she finished a Ph.D. dissertation on the change in the form and function of case-marking in this mixed language, basing her work heavily in variationist methodology, with a language contact twist influenced by Gillian Sankoff's work on Tok Pisin.

felicity.meakins@manchester.ac.uk

Miriam Meyerhoff is Professor of Sociolinguistics at the University of Edinburgh. In 1994, she embarked on the study of Bislama in Vanuatu after a throw-away question from Gillian Sankoff ("Have you ever considered doing fieldwork on this?") and since then she has written papers and one book on syntactic variation, reduplication and gendered speech acts in Bislama. More recently she has worked on the sociolinguistic situation of another creole, spoken in Bequia (St Vincent and the Grenadines). Gillian's more recent work on the stability of individual speakers' linguistic systems has played a key role in understanding the sociolinguistic patterns of Bequians who have gone away and then come back to their home community.

mhoff@ling.ed.ac.uk

Rajend Mesthrie has been working on language contact and variation in the South African context since the 1980s. An Educational Opportunities Scholarship from South Africa took him on pilgrimage to Philadelphia for the fall semester of 1989. Here he received guidance from a number of scholars including Sherry Ash,

Bill Labov and – above all – Gillian Sankoff, who read and commented helpfully on his post-doctoral work on South African Indian English and Fanakalo. He has since been informally adopted as an honorary alumnus of the department without actually having graduated there. Rajend heads the Linguistics Section at the University of Cape Town, is serving a second term as President of the Linguistic Society of Southern Africa and holds a National Research Foundation chair in 'Migration, language and social change.' Amongst his publications are the edited volume *Language in South Africa* (CUP 2002).

rajend.mesthrie@uct.ac.za

Naomi Nagy worked with Gillian in two capacities while a graduate student. She collected and analyzed data from bilingual Anglophones in Montreal, a primary motivation for attending the University of Pennsylvania (1990–1996). Gillian also encouraged and supervised her dissertation, assuring her that it would be possible to apply the methods learned from the (Philadelphia) Project on Change and Variation in a new and non-English context. The context was research on Faetar, a contact language spoken in two small villages in southern Italy (and Etobicoke, ON, *inter alia*) that contains elements of Italian and Francoprovençal. Naomi was the head of the Linguistics Program at the University of New Hampshire from 1997 to 2008 and is now Assistant Professor at the University of Toronto. Her work on both of the above topics continues and is regularly influenced by new findings reported by Gillian. She hopes that Toronto will provide a fertile venue for continued work in a plurilinguistic context.

nn.in.nh@gmail.com

David Sankoff accompanied Gillian on her Ph.D. fieldwork in New Guinea, part of which, the lexicostatistics of Morobe District, led to his own Ph.D. research in mathematics. Then, in the early 1970s, he and his collaborator Henrietta Cedergren, working on variation in Panamanian Spanish, joined forces with Gillian to launch the Montreal French project, including the first computerized sociolinguistic corpus. During this time he also wrote the first of the programs for the analysis of variable rules, which eventually became GoldVarb. Aside from his work on phonological and syntactic variation, he has been interested in formal models of bilingual speech modes, quantitative aspects of discourse analysis and the epistemology of sociolinguistics. With Bill Labov and Tony Kroch, he was a founder of the journal *Language Variation and Change*, which he ran from its Université de Montréal office for 20 years. Starting with NWAVE 8, he played a key role in the survival and growth of this conference during a critical period. He is currently

Canada Research Chair in Mathematical Genomics, in the Department of Mathematics and Statistics of the University of Ottawa.

sankoff@uottawa.ca

Bambi B. Schieffelin is Collegiate Professor and Professor of Anthropology at New York University. She first met Gillian at the LSA Institute in Ann Arbor in 1973, and was her colleague at the University of Pennsylvania from 1980–1986, where conversations about language contact and change, and life in general were often mixed. Her initial work on language socialization in Bosavi, Papua New Guinea did not include Tok Pisin, as it was not part of the everyday verbal repertoire of the community. Once Tok Pisin was introduced through Christian missionization, she began to track its impact on the Bosavi language through translation practices over several decades, consulting with Gillian on Tok Pisin issues. Her trilingual Kaluli-Tok Pisin-English dictionary recognized the potential for multilingualism in Bosavi. Always wishing there was more Creole language in her life, she also carried out research on Haitian focusing on language ideology and kréyòl orthography, as well as on code-switching in New York Haitian families. She is currently completing a book on the impact of Christianity and Tok Pisin on the cultural and linguistic lives of Bosavi people.

bs4@nyu.edu

Pierrette Thibault found Gillian Sankoff at the Université de Montréal in 1972, and through her, she discovered Weinreich, Labov and Herzog (1968), which set her on a quest for language change as a dynamic social process. Gillian supervised her Ph.D. thesis, and after Gillian left for Philadelphia, she was hired to replace her. She carried on the sociolinguistic research project on Montreal French, with the help of Diane Vincent and David Sankoff. A new corpus was constituted in 1984, a panel study with some of the same speakers interviewed in 1971 by Gillian and her associates David Sankoff and Henrietta J. Cedergren. Both Thibault and Sankoff have kept working on those two corpora and in the early nineties, they decided to look into L2 Quebec French. Thus, the project on Anglo-Montrealers speaking French started. It has allowed both Gillian and Pierrette to engage new students (Nagy and Blondeau, among others) on the study of languages in contact. They still work together and Pierrette Thibault has research plans for both of them after they retire.

pierrette.thibault@umontreal.ca

Anne-José Villeneuve is a Ph.D. candidate in French Linguistics at Indiana University. A Montréal native, she first became familiar with Gillian's work during

her undergraduate studies at the Université de Montréal. Since then, she has had the chance to meet Gillian at various conferences after being introduced by Julie Auger, Anne-José's thesis advisor and one of Gillian's former graduate students. Anne-José's research focuses on sociolinguistic variation in the spoken French of Vimeu, an area of northwestern France where the regional Oïl language, Picard, still enjoys relative vitality, alongside French.

avillene@indiana.edu

PART I

Language Ideology

From the speakers, what can we learn
about the language?

Language, mobility and (in)security

A Journey through Francophone Canada

Michelle Daveluy
University of Alberta

The proposed journey focuses on localised groups of French speakers in Nova Scotia, New Brunswick, Québec and Alberta but also on transient workers that go back and forth to their workplace and military families who are relocated at regular intervals within the country. Linkages between language, mobility and (in)security are assessed through the analysis of linguistic variables that illustrate the enactment of local norms of interaction among mobile Canadian French speakers. Continuities among groups that may superficially appear, and are often theorised, as disconnected become prominent. I ultimately suggest that Francophone Canada is best grasped as a set of multilingual speech communities rather than as a unidimensionally conceived series of groups sharing the exclusive commonality of speaking French.

Keywords: Francophone Canada; speech communities; language ideology; language rights; language policy; mobile speakers; military personnel; T/V address forms

1. Introduction

The sociolinguistic variationist approach is appealing to assess the diversity of Francophone Canada, in particular in terms of its systematic data collection procedures permitting statistical analysis of linguistic behavior. However, mega-corpora (e.g., Thibault et al. 1990, the Nova-Scotia (Flikeid) and Ottawa (Poplack) corpora, etc.) tend to emphasize similarities and differences in various ways of speaking French across the country while the contexts of the linguistic situations that are accounted for remain underexamined. The journey proposed here starts with clearly localised groups of French speakers in several provinces in Canada, namely (from the east to the west) Nova Scotia, New Brunswick, Québec and Alberta. Well-documented linguistic variables (*tu* and *vous* to address a single speaker (e.g., Brown & Gilman 1960; Thibault 1991; Vincent 2001), *juste* and *seulement* (e.g., Thibault & Daveluy 1989; Nadasdi & Keppie 2004) illustrate the enactment of local norms of interaction in spoken French.

Then the journey broadens its focus to include less often studied categories of French speakers that are nonetheless highly relevant to a thorough representation of contemporary Francophone Canada. Two very mobile groups of French Canadians are specifically discussed: transient workers that go back and forth to their workplace (Daveluy 2007a) and military families who are relocated at regular intervals (Daveluy in press). The overall picture which is obtained in the process is far from exhaustive but it becomes possible to discuss sociolinguistic dimensions of French Canada that are not usually taken into account.

To assess the French language dynamics in Canada, even incompletely and partially, it is useful to shift perspective from a micro-analysis of communities (Daveluy 2005) to processes delimiting relationships among groups of speakers (Irvine & Gal 2000). Such a shift entails linking linguistic variables with language claims made by speakers (Woolard 1998) in order to unravel continuities among groups that may superficially appear, and are often theorised, as disconnected. I ultimately propose that Francophone Canada is best grasped as a set of multilingual speech communities rather than as a unidimensionally conceived series of groups sharing the exclusive commonality of speaking French. Language and (in)security (Daveluy 2007a) emerge as an overarching theme for all groups considered.

2.   Language variation and norms of interaction

Salient linguistic variables are of limited interest for the analysis of the unconscious use of socially meaningful forms of speech, but they are very useful in terms of interaction. They are particularly efficient to assess exchanges between speakers who definitely share a language without necessarily belonging to a single linguistic community, which is the case of Francophones living in various parts of Canada. Highly salient variables (by which I mean variables with relatively high levels of social awareness) provide concrete evidence of the existence of inescapable norms of interaction that tend to remain intangible until challenged. Reactions to the use of such variables point to moments of negotiation in the course of otherwise unmarked conversation between speakers of a given language.

The relatively low frequency of salient variables explains their absence in quantitative analyses of linguistic behavior. Too few occurrences are produced to warrant statistical treatment of tokens of speech that otherwise play a non-negligible role in communication. Indeed, the productivity of salient linguistic variables is ensured by the stylistic effect they convey. Sankoff and Vincent (1980) clearly demonstrated this is how negation works in French spoken in Montréal. The particle *ne* is so rarely uttered that it is tempting to declare its disappearance a done deal, and its absence a typical feature of spoken French in Québec. However,

based on the stylistic impact of *ne … pas*, Sankoff and Vincent instead proposed the particle is productively used by speakers and, as such, constitutes a linguistic resource that cannot be discarded on the basis of its low frequency. Further work supported their analysis (Daveluy 1994, 2005): the use of the particle *ne* maintains itself through time; it is also much more frequently used when speakers read French out loud. It has further been documented that English native speakers who are fluent in French use the particle *ne* according to these same predictable patterns (Daveluy 2005: 51–56, 75–76; see also Blondeau in this volume).[1]

The control of the particle *ne* displayed in the speech of both French and Anglo-Montrealers further signals the strategic use of salient variables. Speakers who can master the use of rare forms draw on their skills for stylistic purposes that carry weight socially. The scarcity of the forms enhances its value in the speech economy, partly explaining their maintenance when obsolescence should have been their fate. But rare forms are also maintained because they are assets for interactive positioning. Another linguistic variant, *vous*, illustrates how this occurs in Francophone Canada.

In French as spoken in Québec, one can choose between *tu* and *vous* when addressing a stranger. However, because *tu* is produced more frequently in a number of contexts, it is often perceived as a typical feature of the language variety. It is in that sense that I refer to *vous* as a rare variant; it remains an option but it is produced sparingly. A number of studies have documented the complexity of this variable (e.g., Thibault 1991; Vincent 2001). As Lyster (1996) shows, it is stylistically as efficient as the complete form of the negation already discussed: *vous* is used more often in formal contexts and in writing. Like *ne*, the fact that *vous* is used in much more clearly identified contexts does not impact its maintenance as a linguistic resource. On the contrary, it ensures its preservation as a highly sociolinguistically marked variant.

In Acadian French (in New-Brunswick and Nova Scotia) *tu* also appears to be predominant. However, to my knowledge no specific study of the variation between *tu* and *vous* to address a single individual in Acadian French has been published to this day. Until such studies become available, it would be unwise to discard evidence from interactions regarding the use of *tu* and *vous* in natural contexts. In such exchanges, it appears addressing a stranger as *vous* while interacting with Acadians clearly discloses one's status as an outsider. For example, when introduced to Acadians in Nova Scotia, I was often surprised, and admittedly disappointed, that conversations starting in French almost invariably switched to English. Since exchanges continued for fairly long periods of time, it was clear to me the desire

---

1.   Gillian Sankoff also presented relevant data on this topic at the 2006 NWAV conference.

to communicate was not at stake and I initially thought the language might be the problem. Then, I realised English native speakers who also speak French were able, in my presence, to converse in French with the same Acadians who were switching to English to talk to me (a French native speaker). So the language was definitively not the issue. At least not on the part of my Acadian interlocutors. Paying attention to what was triggering the code-switch to English, it became clear it was often happening after I had used *vous* rather than *tu* in a sentence.

Becoming self-conscious about my own production of this rather rare form in French, I considered training myself not to use *vous* with strangers. As per the classical explanation provided by Brown and Gilman (1960) *vous* is useful for socially distancing oneself while *tu* shows solidarity. In that sense, dropping *vous* from my repertoire appeared as a desirable strategy. However, it proved much more difficult to achieve than it seems and I never became very good at it. Perhaps eliminating highly efficient and socially valuable linguistic resources from one's repertoire is not so simple a matter. Indeed, in Québec, the proper use of *vous* is a distinguishing feature among English native speakers. Bilinguals do manage to use *vous* more or less appropriately. Those who don't are not considered fully bilingual. Lyster (1996) describes the challenges second language speakers face in this regard. So, I realised I was more attached to *vous* than I would have thought. *Vous* was keeping me away from Acadians, but it was keeping me in touch with, for example, some English-Montrealers (those who speak French).

Of course, English native speakers who are also comfortable in French do not have such strong feelings about *vous*, even if they learned to use it appropriately. To them, it remains a challenging form. Not using it with Acadians while speaking in French can be a relief. As a consequence they can adapt to the Acadian norms of interaction fluidly and continue a conversation in French more easily than I could, at least in this respect. To me, such a use of *tu* by native English speakers and Acadians certainly points to solidarity but it also indicates shared power in front of a French native speaker.

This example of pronoun selection, while anecdotal, points to ironies of interaction: French native speakers (here an Acadian and a Québécoise) switching to another, less challenging, language of interaction (in this case English) while English does not seem necessary for English native speakers and Acadians to hold a conversation (since they apparently sometimes prefer to stick to French). More importantly, this situation underlines how misleading it can be to assume that French is the exclusive common denominator of Francophones in Canada. In fact, Francophones in Canada rather have their relationship to English native speakers in common. In New Brunswick in particular, regional differences are striking in this regard. Keppie (2008) documents how the Brayon, those living in the Acadian peninsula, and those from the Moncton area (where Chiac is used) relate to each

other, how their positioning towards Québec differ, and how the role of English changes from one region to the next.

In the instance described here, bilingual English native speakers keep the two French communities in contact. Through the mediation of a third party, the Québec and Acadian communities are connected. Left to themselves, they might not even maintain a relationship. For example, at this gas station in Hubbards, Nova Scotia, a lady once explained to me that my French was preventing her from speaking French with me. She said, "You're Québécoise, you speak good French. We don't. It's a pleasure to chat with you but don't expect me to answer in French. I would not dare." Clearly, French is not always the factor connecting groups of speakers in Francophone Canada. On the contrary, it even sometimes appears as a threat.

Figure 1. Hubbards, Nova Scotia (photo: Thaddeus Holownia).

In a discussion on language and resilience, I have addressed unexpected but efficient strategies speakers sometimes adopt that are not well modeled by linguists who are particularly concerned by language revitalization (Daveluy 2007b). The behaviors discussed here could easily be interpreted under the paradigm of loss. At the micro-level, I pointed to the loss or restricted use of linguistic resources and variants, like *vous*. More generally, I also noted the partial loss or limited use of the language in favor of another that suits the circumstances better (alternatively, silence may be favored at the individual level). These strategies deserve better assessment. They are reminiscent of the necessary physical spatial positioning in a household, described by Ochs (1986), to successfully elicit speech since conversations are constrained in the part of the house reserved for visitors. Paying attention to such social positioning seems very important in order to move away from the widespread and oversimplistic view that Francophones in Canada form separate groups based on differences in their respective ways of speaking French.

Alternatively, the type of behaviors I described can be understood as long term investments in language sustainability rather than evidence of individuals

failing to do what is required for their language to maintain itself. The reaction of my Acadian interlocutor at the gas station is often (mis-)interpreted as linguistic insecurity leading to code switching. However, sacrificing a linguistic variant in order to adjust to the norms of interaction of other speakers may be a worthwhile and necessary step (that I was myself unable to accomplish in the case of *vous*). Such a strategy could as easily be interpreted as an indication of adaptation to the widening of one's linguistic community. I will argue later that mobility is another feature of contemporary Francophone Canada that requires better analysis. Even if it has so far been under-examined, mobility within Francophone Canada influences one's sense of belonging and participating to a linguistic community. Negotiations across localised communities of the type we have seen in this section are frequent in that context.

3.   Language insecurity and ideologies

Recognising linkages between various groups of French speakers in Canada does not imply they should be thought of as a single and unified entity. Differentiation remains an effective mechanism, both at the level of the language itself and in terms of the historical development of locally oriented ideologies. For example, Acadians and Western Francophones are distinct as far as the conditions under which they maintained themselves as French speaking groups are concerned. Given the forced exodus the Acadians faced in 1755, overcoming adversity is grounded in historical facts for them. For Franco-Albertans persistence through time is not so overtly associated with specific events. A sense of achievement, even if relative, which can be perceived among Acadians, triggers pride in belonging to the group. Cautious self-confidence is not so apparent in the West and in the case of Franco-Albertans specifically. I argue below that a language ideology fostering insecurity is actively nurtured in Alberta (Daveluy 2007a, 2007c).

The emphasis on the exogamic nature of marital arrangements within the French community is one of the mechanisms contributing to the ideology of insecurity in Alberta. Risks of assimilation are (considered) higher when the rate of exogamous marriages increases. Exogamic marriages occur across cultural or linguistic boundaries rather than within one's own group (endogamy). Linguistic exogamy as practiced in the Vaupés area in the Amazon region in South America is a classical example within the field of linguistic anthropology. The case has been extensively described by Gomez-Imbert (1996), Jackson (1983) and Hugh-Jones (1979). In this part of the world, marriage is acceptable exclusively between persons of different language groups. In Canada, we can certainly not talk of marriage rules or obligations, but individuals with different mother-tongues, if these are official languages, are considered exogamic.

Simply put, English and French native speakers can form an exogamous couple. Still, French and English individuals choosing to marry do it on their own volition, without particular social pressure. However, a systematic comparison of linguistic exogamy in the Vaupés and in Canada shows the concept remains relevant in the Canadian context (Daveluy 2007c).

Exogamic marriages between speakers of the two official languages in Canada are of particular interest since 1982, when the Charter of Rights was adopted. Article 23 in this Charter warrants the right to education in French to those who learn French as a mother-tongue. In that context, the language children of exogamous parents speak is not a trivial affair. Prior to that date, exogamic couples were simply labeled as mixed marriages. Nowadays, individuals of all origins can still contract a mixed marriage while the union of French and English Canadians is interpreted as exogamic. Such linguistic exogamy occurs all over Canada but it is linked to the demographic balance between the two ethnoliguistic groups. For example, its prevalence is greater in Alberta than in New Brunswick, where the French population represents a higher proportion of the population.

As one of the strategies adopted to counter the threat of assimilation, children schooled in French in Alberta are explicitly socialised as right-holders (Daveluy 2007c). They are actually referred to as *ayant-droit*s "having rights" if they qualify to register at French schools. In the school system, they appear less as students than citizens familiarising themselves with the role their linguistic origin calls for. Right-holders are socialised as members of a minority group, though a protected one, considering the official status of their mother-tongue. Still, they are constantly reminded of their privileges and responsibilities towards the local community, the French language, and Canada.

In fact, the Franco-Albertan elite fully endorses its status of official minority. Upon arrival in the West, Québécois are repeatedly reminded that they are ill-equipped to face the minority situation they will deal with on a daily basis in Alberta. Québécois are indeed socialised to think of themselves as members of a majority in a circumscribed territory. However, Acadians, Franco-Albertans and Québécois certainly do share the status of French-Canadian minority communities everywhere in the country, Québec excepted. In that context, Franco-Albertans have developed an expertise in French minority affairs that Québécois completely lack. Acadians and Franco-Albertans have similar experiences in this regard but the precariousness of the gains obtained in the West is greater than in the East (in New-Brunswick particularly). The absence of foundational traumatic events in the West also contributes to the development of an ideology of survival anchored in non-confrontational persistence. As a consequence, uncertainty about current conditions is nurtured in Alberta and the sustainability of the French community is not taken for granted by Franco-Albertans.

It is worth noting that speaking French more easily, as Québécois do compared to some Franco-Albertans, does not warrant long term maintenance of the

language. On the contrary, speaking French that does not blend in the political context may trigger repercussions and jeopardize the achieved state of equilibrium. Even if the enemy of the French language has not been clearly identified in the West, like it had been done among Acadians centuries ago, where English plays that role, such an enemy exists nonetheless. Behaviors that would attract attention could trigger intolerance on the part of the unknown enemy. This is not to say that Franco-Albertans are unable to use French as they wish. They use it strategically, with proper motivation clearly laid out in social positioning activities like lobbying efforts.

The fact that some of the Franco-Albertans have arrived in the West as recently as fifty years ago is also interesting from a sociolinguistic perspective. In some instances, groups (rather than individuals) migrated *en masse* from Québec to establish themselves in localities that are nowadays accepted as part of the Franco-Albertan landscape. While conducting research, it became evident that some of the typical linguistic features of Québécois remain in the speech of early settlers (Nahachewsky et al. 2005; Daveluy et al. 2004). For example, it is the case of *seulement que*, a restrictive form that Thibault & Daveluy (1989) associated with older segments of the population in data from Québec. *Seulement que* alternates with *juste*, which tends to replace the former in the speech of younger speakers. Tendencies identified in Québec French (Thibault & Daveluy 1989; Daveluy 2005) are also found among Franco-Albertans (Nadasdi & Keppie 2004). Tracing typical Québécois characteristics in the ways contemporary Franco-Albertans speak certainly contributes to an appropriate documentation of the linkages between them. However, considering the differences Québécois and Franco-Albertans demonstrate with regards to their relationship with the French and English languages, the analysis of such variation must be properly contextualised to address the sociolinguistic dynamic in western Canada.

Indeed, the discrepancies in the ideologies the Franco-Albertans, the Acadians and the Québécois respectively foster are numerous. The ideology of insecurity the Franco-Albertans cultivate contrasts with the Québécois identity which revolves around a sense of self-confidence. Relocated in a minority context in the West, the linguistic security of the Québécois appears arrogant, ignorant and risk-laden. Compared to Franco-Albertans, Québécois rarely question the survival of their mother-tongue. In their secure environment, Québécois have recently learned to indulge in action rather than constantly justify their existence.

This contrast in ideologies is sharply displayed in matters related to language rights. Franco-Albertans are focusing on raising the awareness of future generations by encouraging them to claim their right to education in French. In so doing, they are attempting to maintain the number of students required by governments

to fund programs in French. On the other hand, Québécois take for granted that these services are guaranteed by law all over the country. As predicted by Franco-Albertans, they are bound to experience a rough ride in the West because of their limited knowledge of local socio-political contingencies. National laws are not necessarily diligently implemented everywhere in Canada because jurisdictions overlap. Education, for example, is a provincial responsibility while official languages are a national matter.

Finally, it is important to acknowledge that individual commitment to French is a key factor in the ideology of insecurity in Alberta while language issues are perceived and dealt with as collective affairs in the Québécois perspective. Documenting these very different strategies among Francophones in Canada can certainly contribute to a better understanding of the conditions enhancing language revitalization, maintenance and sustainability respectively.

4.   Language and mobility within Canada

Local norms of interactions and ideologies are productive even when unnoticed but they become apparent if changes happen in the social environment where they apply. Between 2001 and 2006, a population increase of over 10% occurred in Alberta (Statistics Canada, 2007). Meanwhile the population in other parts of Canada remained stable or decreased. The booming oil industry is in great part responsible for this increase of population in Alberta but there is of course a domino effect making the shortage of workers an issue in all sectors of employment. To aid in hiring, substantial bonuses are offered even for low income jobs in convenience stores and gas stations in various towns in Alberta.

The oil and gas industry has attracted a sizable quantity of its workforce from within the country. The movement of this internal migration is from the East to the West.[2] In the process, previously segregated segments of the French-Canadian population are coming into more sustained contact. Upon arrival in the West, Acadians and Québécois, who have well established ways of speaking and clearly stated identities in the East, face a well-organised Franco-Albertan community they might not have been particularly aware of prior to moving. These French-Canadian workers experience various levels of resocialisation in Alberta.

---

2.   King (1989: 141–142) notes similar work related movements among young French Newfoundlanders without studying the phenomenon specifically.

First, when exposed to the Franco-Albertan community, these secure French speakers must realise how fragile gains made in the West are and develop an attitude fitting these conditions. To adapt they must also understand that demonstrating their language proficiency is not particularly relevant in a context where lasting is the main issue at stake.[3] Lasting often entails downplaying one's language proficiency, to claim it back later on, when more favourable circumstances permit doing so (for a discussion of how this occurs in Alberta see Daveluy 2007c, and for other cases Daveluy 2007b).

The fact that French workers coming to Alberta rarely stay very long also plays a role in how they are perceived by Franco-Albertans. They have a reputation of coming, rocking the boat, and going back home upon deciding things change too slowly for their taste in Western Canada. In so doing, they fulfill expectations towards them as transient workers rather than committed members of the French Canadian community.

The few who stay long enough to get acquainted with the local community face a dilemma which is not always fully grasped: being a native French speaker does not automatically grant membership into the group. For individuals who rarely doubt their ethnolinguistic identity, this situation is often puzzling. Many react by distancing themselves from this mysterious Franco-Albertan community that apparently does not seem to value speaking French or consider it an asset. Their journey to the West often becomes the chance of a lifetime to learn English and use it in the workplace. Their reaction feeds into the ideology of insecurity since they, perhaps inadvertently, contribute to the risks the local French speaking community already deals with. Their choice to take advantage of Alberta as an English environment undermines the community. Not to mention that the noticeable presence of these workers has made safety in the workplace a more prominent matter in Alberta. Some have been taxed for jeopardizing the safety of work sites and fired for not understanding work procedures explained in English. (For how safety issues play out in the mining industry, see Lindsay Bell's ongoing Ph.D research.)

Quebecer calls Alberta oilpatch firing discrimination, March 7, 2007

A Quebec steelworker who lost his job in Alberta because he can't speak English says he was discriminated against for being French. Carol Rioux was fired

---

3. Considering the official status of French in Canada, the Franco-Albertan community does not appear concerned about surviving per se nowadays. Surviving has physical connotations which were relevant during the colonization of the West in the nineteenth and twentieth centuries. In the contemporary context, community maintenance is rather framed in the long term, around issues pertaining to endurance and commitment, best summarised under the idea of lasting.

after spending two days at the Suncor Energy Inc. plant in Fort McMurray. During a safety orientation session at the energy company, an instructor showed Rioux an English video and handed him a questionnaire. His bilingual friend translated for him, but the instructor took Rioux aside and asked him what the questions meant. Rioux couldn't answer and was fired. Rioux admitted in an interview with Radio-Canada Wednesday that his English was poor, but said it wouldn't create a safety problem on the job. "When the siren rings, it doesn't ring in English and it doesn't ring in French," he said in French. "Everybody understands." His union says other companies hire translators for immigrants or place people who speak the same language together in the same work camp – and Suncor should do the same. They do have workers from overseas who don't necessarily speak perfect English, but they do seem to be accommodated, union representative Cleo Basque told CBC News Wednesday. "We're at home and we don't want to receive any lesser treatment than anybody else gets."

But Suncor spokeswoman Patti Lewis said everyone at Suncor has to meet the same standard and denied it was a case of discrimination. "The first step before you are able to work on our site is that you must pass this test," Lewis told CBC News. "We welcome everyone with open arms. We just want to make sure that they're going to be able to work very safe and it's with a lot of regret that we had to turn this individual away." But Basque said Rioux could understand the colour-coded industrial procedures and is highly qualified for the job. "Mr. Rioux has 25 year's experience, so we feel he's highly skilled and would know exactly what to do in these situations." Rioux has already been working in Alberta for the past five months and says he'll look for a job with another company.

(CBC News http://www.cbc.ca/canada/story/2007/03/07/quebecois-fired. html#skip300×250)

Fluent speakers of a language have long been considered, and still are, a necessary component of language maintenance (Hinton 2007) and a stable source of evidence of language variation (Sankoff 1980). As a consequence, the case of French-Canadian transient workers is not accounted for in the sociolinguistic description of Francophones in the country even if it is clear they actively participate in the language dynamic.

Based on results obtained in recent research with Canadian military personnel and their families (Daveluy in press, 2007d; Asselin 2007), it would be unwise however to simply discard the status of native speaker as a key factor. Fieldwork was conducted onboard a Canadian patrol frigate, the NCSM VILLE DE QUÉBEC.[4] Data was collected in 2004 in Halifax, Nova Scotia, where the ship is stationed; and

---

at sea, on a training mission, in July 2005. This type of ship is manned by a crew of up to 225 shipmates born all over Canada. Women serve in the Canadian Forces, and some are posted on the NCSM VILLE DE QUÉBEC; they represent about 5% of the crew. According to the language model implemented by the Canadian Forces (Asselin 2006), the ship is designated as a French unit. As such, 80% of its staff could be French-speaking; in a unit like the NCMS VILLE DE QUÉBEC, this represents up to 45 shipmates who do not speak French. Still, the language most likely to be used on board is French. From both a linguistic and a social perspective, the ship represents a Canadian mircrocosm.

At sea, young sailors spontaneously volunteered personal stories relating how important it is for them to have the opportunity to work in French. They told me when, where and why, when they were young, their families deliberately decided to stop using French. To this day, they have very clear memories of the time and day of the week, and even the weather, when this decision was made during a formal family gathering held for this purpose. Many of them are children of military personnel. In Canada, these families are regularly relocated from one base to another. Claire Corriveau's film, *Nomad's Life*, clearly shows how the military mode of life is based on family mobility. Her film addresses specifically how spouses are affected by this constant mobility. In his study in the Esquimalt region in British Columbia, where one of the two Canadian Naval bases are, Gabriel Asselin (2007) documents how being posted in an English environment adds to the loneliness and marginalization of Québécoises married to military personnel. According to Anne-Marie McDonald, a military brat herself, and famous novelist:

> If you move around all your life, you can't find where you come from on a map. All those places where you lived are just that: places. You don't come from any of them; you come from a series of events. And those are mapped in memory. Contingent, precarious events, without the counterpane of place to muffle the knowledge of how unlikely we are. Almost not born at every turn. Without a place, events slow-tumbling through time become your roots. Stories shading into one another. You come from a plane crash. From a war that brought your parents together. (McDonald 2003: 36)

Schieffelin (2007) also addresses how children's memories and relationship to place are relevant to language.

The stories I was told at sea underline the consequences mobility has on children. In many instances, speaking English provided a relative sense of stability and family integrity. These young sailors claim working in French, if only a little, was the occasion to reappropriate their language. Teary-eyed yet smiling, they explain how, since being posted on the NCSM VILLE DE QUÉBEC, they have started speaking French again with their mother. They are very proud to now be able to speak French to their own children. The only French unit of the Canadian Navy apparently had an instrumental role in their life. If two generations were sacrificed, it is fair to say there is hope for the third one.

It is important to note the Canadian Forces manage bilingualism to a great extent according to the civilian model put in place by government agencies (Pariseau & Bernier 1987, 1988; Letellier 1987; Bernier & Pariseau 1994, 1991). For example, individuals have to pick a language when enrolling. A young sailor once related to me he had to fill out the form twice at the enrollment office because he had selected both French and English as first languages. The clerk explained to him language is like gender: everyone has to choose one. It is not clear to me if recruits are aware of the consequences of the choice they make when they fill out that form. Considering bilingualism remains a criteria for promotion in the Canadian Forces, there is some irony in not being able to claim one's bilingualism in a country that has been officially bilingual for decades. Still, the Canadian Forces approach is typical of most Canadian institutions in this regard. All Canadians are considered monolingual in demolinguistic statistics about mother tongues.

I have argued (Daveluy 2007c) that denying one's bilingualism is essential for linguistic exogamy to preserve itself. The logic of the system requires individuals to deny their bilingualism in order to play their role as a speaker of one language. Castonguay (1981: 17) labeled such behaviors as cases of retro-exogamy. Indeed, it is very difficult to understand how children raised in a household where two languages are spoken (by their parents) can later present themselves as monolinguals unless they at some point abandoned one or the other of the languages they were exposed to in their family.

This kind of behavior encouraged either by institutions (like the Canadian Forces and the school system) or community norms fits in well with the notion of age-grading phenomena (Labov 1972). Parenting is often accompanied by the adoption of more conservative features of the language. Comparing the situation of various types of Francophones in Canada documents how social pressures trigger such language phenomena that are necessary for the maintenance of norms sustaining the long-term existence of communities.

5.   Conclusion

Comparing Francophones in Canada, the role French plays in the delimitation of groups of speakers does not fit the usual approach. Traditionally, language is associated with birth or early exposure for individuals to qualify as adequate representatives of its use. We have seen that proficiency in French and the ability to use the language are contested in various locales and by Francophones of different kinds as sufficient criteria to justify inclusion in one's group. Comparative studies of specific linguistic variables across Canada would enhance our understanding of sociolinguistic parameters at stake in the country's language dynamics. Focusing on language ideologies has exposed the minorisation process mobile French speakers

go through across the country. Language security in terms of linguistic capacity as much as attitudinal posture is desirable in certain contexts but not in others. Finally, the comparison points to pitfalls in the perspective approaching Francophone communities as monolingual entities. Many French speakers in Canada are bilingual or want to learn English. The strategic value of French for bilingual English native speakers was also highlighted. For all these reasons, it seems appropriate to consider Francophone Canada as a set of multilingual speech communities rather than as a series of groups sharing the exclusive commonality of speaking French.

References

Asselin, G. 2006. Linguistically-defined environments: Continuity and isolation in linguistic communities of Canadian navy bases. Paper presented at the *Annual Meeting of the Society for Applied Anthropology*, Vancouver, BC.

Asselin, G. 2007. Ni civiles ni militaires. Marginalité et identité parmi les familles militaires francophones de la Base de Forces Canadiennes Esquimalt en Colombie-Britannique. MA dissertation, University of Alberta.

Bernier, S. & Pariseau, J. 1991. *Les canadiens français et le bilinguisme dans les forces armées canadiennes. Tome II 1969–1987. Les langues officielles: La volonté gouvernementale et la réponse de la Défense nationale.* Service historique de la Défense nationale.

Bernier, S. & Pariseau, J. 1994. *French Canadians and bilingualism in the Canadian armed forces, Volume II, 1969–1987: National Defense's response to the federal policy.* Socio-Military Series No. 2, Ottawa, Directorate of History of the Department of National Defense.

Brown, R. & Gilman, A. 1960. The pronouns of power and solidarity. In *Style in Language,* T. Sebeok (Ed.), 253–276. Cambridge MA: Massachusetts Institute of Technology.

Castonguay, C. 1981. *Exogamie et anglicisation dans les régions de Montréal, Hull, Ottawa et Sudbury.* Québec: Université Laval, Centre international de recherche sur le bilinguisme.

Corriveau, C. 2007. *Nomad's life. Les épouses de l'armée.* National Film Board of Canada.

Daveluy, M. 1994. Continuité et changement dans la communauté linguistique: Le français parlé par des francophones et des anglophones de Montréal. *Culture* 14(2): 85–93.

Daveluy, M. 2005. *Les langues étendards. Allégeances langagières en français parlé à Montréal.* Collection Langues et pratiques discursives. Québec: Nota bene.

Daveluy, M. 2007a. New perspectives on native speakers? *American Anthropology Association Annual Conference*, Washington DC, November 28–December 2.

Daveluy, M. 2007b. Resilience and language. In *Resilience in Arctic Societies*, L. Kaplan & M. Daveluy (Eds), 89–96. Fairbanks: Alaska Native Language Centre.

Daveluy, M. 2007c. L'exogamie langagière en Amazonie et au Canada. *Anthropologie et Sociétés* 31(1): 55–73.

Daveluy, M. 2007d. Langues, mobilité et sécurité dans les forces armées canadienne et américaine. *Colloque du trentième anniversaire de la revue Anthropologie et Sociétés*, Université Laval, Québec.

Daveluy, M. In press. Communicating among linguistic communities onboard a Canadian Navy ship. In *Proceedings of the 2006 Atlantic Provinces Linguistic Association Conference,* M. Harry & C. Beaudoin-Lietz (Eds). Halifax: Saint Mary's University.

Daveluy, M., Dalley, P., McMahon, F. & LaChance, D. 2004. Marginal individuals in the defini-tion of communities in the Canadian Prairies. / Marges communautaires dans les Prairies canadiennes. Paper presented at the *31ˢᵗ Annual Meeting of the Canadian Anthropology Society/Colloque annuel de la Société canadienne d'anthropologie*, The University of Western Ontario, London ON.

Gomez Imbert, E. 1996. When animals become 'rounded' and 'feminine': Conceptual categories and linguistic classification in a multilingual setting. In *Rethinking linguistic relativity*, J.J. Gumperz & S.C. Levinson (Eds), 438–469. Cambridge: CUP.

Hinton, L. 2007. Revitalizing endangered languages, empowering speakers. Presidential Address. *American Anthropology Association Annual Conference*, Washington DC, November 28–December 2.

Hugh-Jones, C. 1979. *From the Milk River: Spatial and temporal protests in Northwest Amazonia*. Cambridge: CUP.

Jackson, J. 1983. *The Fish People: Linguistic exogamy and Tukanoan identity in Northwest Amazonia*. Cambridge: CUP.

Irvine, J.T. & Gal, S. 2000. Language ideology and linguistic differentiation. In *Regimes of Language: Ideologies, polities, and identities*, P.V. Kroskrity (Ed.), 35–83. Santa Fe, NM: School of American Research Press/Oxford: James Currey.

Keppie, C. 2008. Ideological Discourses among New Brunswick Francophones. Ph.D. dissertation. University of Alberta.

King, R. 1989. On the social meaning of linguistic variability in language death situations: Variation in Newfoundland French. In *Investigating Obsolescence: Studies in language contraction and death* [Studies in the Social and Cultural Foundations of Language 7], N. Dorian, (Ed.), 139–148. Cambridge: CUP.

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

Letellier, A. 1987. *Réforme linguistique à la Défense nationale: La mise en marche des programmes de bilinguisme 1967–1977*. DND Language Reform: Staffing the Bilingualism Programs, 1967–1977, Socio-Military Series No. 3, Ottawa, Directorate of History of the Department of National Defense.

Lyster, R. 1996. Question forms, conditionals, and second-person pronouns used by adolescent native speakers across two levels of formality in written and spoken French. *The Modern Language Journal* 80: 165–182.

McDonald, A.-M. 2003. *The Way the Crow Flies*. New York: Alfred A. Knopf.

Nadasdi, T. & Keppie, C. 2004. Le mot *juste* en français albertain. *Cahiers franco-canadiens de l'ouest* 16(1–2): 67–78.

Nahachewsky, A. (Ed.). 2005. *Local culture and diversity on the Prairies. A Project Report*. Friends of the Ukrainian Folklore Centre. (With the collaboration of F. McMahon & M. Daveluy).

Ochs, E. 1986. From feeling to grammar: A Samoan case study. In *Language Socialization across Cultures*, E. Ochs & B. B. Schieffelin (Eds), 251–272. Cambridge: CUP.

Pariseau, J. & Bernier, S. 1987. *Les Canadiens français et le bilinguisme dans les Forces armées canadiennes. Tome I 1763–1969. Le spectre d'une armée bicéphale*. Service historique de la Défense nationale.

Pariseau, J. & Bernier, S. 1988. *French Canadians and Bilingualism in the Canadian Armed Forces,* Vol. I*, 1763–1969: The Fear of a Parallel Army*, Socio-Military Series No. 2, Ottawa, Direc-torate of History of the Department of National Defense.

Sankoff, G. 1980. *The Social Life of Language*. Philadelphia, PA: University of Pennsylvania Press.

Sankoff, G. & Vincent, D. 1980. The productive use of *ne* in spoken Montréal French. In *The Social Life of Language: Collected essays by Gillian Sankoff*, 295–310. Philadelphia, PA: University of Pennsylvania Press.

Schieffelin, B.B. 2007. Langue et lieu dans l'univers de l'enfance. *Anthropologie et Sociétés* 31(1): 15–37.

Thibault, P. 1991. La langue en mouvement: Simplification, régularisation, restructuration. *LINX-Linguistique Paris X-Nanterre* 25: 89–92.

Thibault, P. & Daveluy, M. 1989. Quelques traces du passage du temps dans le parler des Montréalais, 1971–1984. *Language Variation and Change* 1(1): 19–45.

Thibault, P. & Vincent, D. avec la participation de Audet, G., Daveluy, M., Laforest, M., Lessard, P. & Rains, C. 1990. *Un corpus de français parlé. Montréal 84: Historique, méthodes et perspectives de recherche*. Québec: Recherches sociolinguistiques.

Vincent, D. 2001. Remarques sur le tutoiement et le vouvoiement en français parlé au Québec. Actes du colloque *La journée du Québec*, Institut d'études romanes, Université de Copenhague, 11–22.

Woolard, K.A. 1998. Language ideology as a field of inquiry. In *Language Ideologies: Practice and theory*, B.B. Schieffelin, K.A. Woolard & P.V. Kroskrity (Eds), 3–42. Oxford: OUP.

Language repertoires and the middle-class in urban Solomon Islands[1]

Christine Jourdan
Department of Anthropology & Sociology,
Concordia University

In Honiara, Solomon Islands, 64 vernacular languages coexist with Pijin, the lingua franca and linguistic cement of the town, and with English, the former colonial language. The chapter shows how the modalities of urban linguistic repertoires vary with different phases of Honiara's transformation and with the life course of individuals. There is a reconfiguration of the linguistic repertoires of most urbanites and language shift for some, particularly the younger urban middle-class. The first section presents a background on multilingualism in the Solomon Islands and the *typical* linguistic repertoires prevalent before urbanization. An analysis of the development of the middle class in Honiara follows. Finally, it addresses the social forces shaping the language practices of the middle-class. The paper shows that, if societal language shift is the trend in Honiara, young members of the urban middle-class are at the forefront of this change.

Keywords: Solomon Islands Pijin; *langgus* (vernacular languages); social class; social change; families; generational change; real time change

In her work, be it on the sociolinguistics of French in Montreal or on Tok Pisin in Papua New Guinea, Gillian Sankoff has been a trailblazer. Ever since, many scholars have been trying to catch up, either by replicating what she has done in other locales or by setting up research agendas in relation to the results she has reached earlier. Hopefully this paper will do more than try to catch up with her, a task in itself quite daunting, but will contribute to further our understanding of the social processes that drive the life of pidgins in Melanesia, and more specifically in the Solomon Islands.

In one of her famous papers on the sociolinguistics of Tok Pisin in Papua New Guinea, Sankoff wrote:

> What appears to have happened is that the original tok ples/tok Pisin dichotomy has been largely replaced by tok Pisin/English as a symbolic marker of power and status in the urban society. (Sankoff 1980a: 26)

If we were to replace Tok Pisin with Pijin and *tok ples* with *langgus* (i.e., the vernacular, indigenous languages in Papua New Guinea and the Solomon Islands respectively), we would be describing the language shift currently taking place in Honiara, the capital city of the Solomon Islands: albeit with slight differences that are linked to the intricacies of changing language ideologies in Solomon Islands.

In this paper honoring the work and intellectual legacy of Gillian Sankoff, and keeping in mind the conclusion she reached above on the linguistic situation in Papua New Guinea thirty years ago, I will show how the modalities of urban linguistic repertoires vary with the different phases of Honiara's transformation. Much of the codeswitching and language shift taking place in Honiara today can be associated with three essential variables: age, gender and social class. For the purpose of this article, I will concentrate on connecting language shift to the development of the urban middle-class. The first section of this article will present a background on multilingualism in the Solomon Islands and the *typical* linguistic repertoires prevalent before urbanization. It will be followed by an analysis of the development of the middle class in Honiara. Finally, I will address the social forces shaping the language practices of the middle-class. I will show here that, if societal language shift is the trend in Honiara, then young members of the urban middle-class are at the forefront of this change.

Two claims are central to this paper. First, the situation encountered today in Honiara of rapid social change has not led to the disappearance of webs of meaning that are sustained by various languages in various contexts. Rather, the diversity of context of use and ideologies at hand has effloresced, requiring individuals to remain multilingual in order to belong and be in the world. Like elsewhere, languages in the Solomon Islands have always been constitutive of the person: in Honiara, they are constitutive of multiple social personhoods (see Mauss 1938; Rumsey 2000), in an increasingly complex web of social relationships. Transitional social moments, such as they are currently in Honiara, are also transitional linguistic moments. When social diversity and complex life ideologies and practices effloresce, so do language choices. Second, linguistic repertoires (including registers) index the social subject. Therefore, changes in linguistic repertoires index the social trajectories acted upon and shaped by individuals and groups in the course of social and ideological changes. The linguistic repertoires of individuals or groups thus effloresce and shrink, or are reconfigured with different languages or different registers at various moments in the history of social groups (Gal 1979).

We are at a point in time when Honiara is in continuous social transformation, and where linguistic repertoires are yet again being reconfigured. However, it is essential to distinguish between individual reconfiguration and societal reconfiguration. In the case of Honiara, both will eventually lead to language shift across generations as the number of long term urbanites increases.

1.   Multilingualism and linguistic repertoires in the Solomon Islands

The sociolinguistic history of the Solomon Islands reveals that the transformation of the linguistic repertoires of Solomon Islanders, as can be expected, developed alongside the social transformations that took place in the country from the moment of contact with Europeans. Three phases are particularly important: the early period of colonization (1893–1920); the development and expansion of the local plantation economy (1920–1960); the beginning of urbanization and the development of the middle-class. I will summarize them.[2]

In the Solomon Islands, as with many Melanesian societies where multilingualism is common (see Sankoff 1980b) rather than the exception, people have been multilingual for pragmatic reasons as well as symbolic ones. Thus they have made use of their linguistic repertoire not only to find a place in exchange networks, but also to express their identity. Active bilingualism or multilingualism is not only a response to linguistic diversity; it is also an act of identity creation.

In pre-colonial times (the islands were a British protectorate from 1893 until 1978), multilingualism seems to have been reciprocal between the ethno-linguistic groups that lived in proximity of each other.[3] Not all members of the groups were multilingual to the same extent. While men, who traditionally were engaged in trade, were actively multilingual, women, who traditionally stayed home, were not. In Jourdan (2007a), I proposed that societal multilingualism indexes a concern for reciprocity at the core of Melanesian sociality.

---

2.   Some parts of this section have appeared in Jourdan (2007a) where the issues have been explored in greater depth.

3.   Note that people living at the geographical boundaries of their ethno-linguistic groups are more likely to be bilingual or multilingual than people residing further away from these boundaries. In the same manner, people involved in trading are likely to be more multilingual than people who are not, etc. The analysis made by Sankoff (1980a), in which she discusses variables such as geographical locations, trading activities, size of ethno-linguistic groups, dialects, mutual intelligibility, etc. are relevant to the situation I am summarizing here.

> It could be argued that this linguistic practice has at its heart a concern for reciprocity that extends beyond the linguistic sphere, a concern that we find often associated with exchange networks or clanic obligations … Reciprocal or balanced multilingualism as I like to call it …., is a linguistic practice that … may stem from a more encompassing ideology of egalitarianism between ethno-linguistic groups (2007a: 32).

The centrality of multilingualism in reciprocal exchange between groups permeates Malinowski's work on the Kula (1961), and that of Jackson (1983) and Aikhenvald (2002) in Amazonia. Bambi Schieffelin's work (1990) on the socialization of young Kaluli children in Papua New Guinea takes this theme one step further and shows that while they are learning Kaluli, the children are also socialized into negotiating reciprocity and the sharing of food. She demonstrates that speech is the avenue of reciprocal exchange. The "give and take of everyday life" (Schieffelin 1990) that takes place within social groups is mediated by appropriate speech events and registers. She says:

> Sharing must be negotiated through talk. In interactions involving food, participants often differ in their goals; reciprocated sharing is not always the desired end for everyone. This creates a tension between autonomy and interdependence, expressed through daily demands and appeals to share food, to cooperate in tasks, to participate in conversation, and even to pay attention. In this egalitarian society, individuals cannot easily compel one another to act. They depend on sociable relationships in which they can expect or negotiate daily cooperation and reciprocity, and through which they can define their personal autonomy. Caregivers use language to constitute these sociable relationships between their children and themselves, mediating both the sharing and talking about it. (Schieffelin 1990: 137)

Linguistic exchange is more than a metaphor for social exchange; it embodies exchange.

Despite the ethos of egalitarianism that was part of sociality in most places, inequality within groups and between groups existed. Clearly, someone speaking a few languages rather than one only, derived some degree of prestige during exchanges with neighboring groups. But within groups it was the knowledge of registers that conferred advantages to some speakers: the knowledge of sacred registers, of word taboos, and of incantations, for instance. There are distinct symbolic and tangible rewards attached to speaking a language *well* even in egalitarian Melanesian societies. Mastering a language and being a good orator carried much weight in societies where much of the local political power rested upon an individual's ability to convince others and to speak to them in eloquent ways (Sankoff 1980a; Telban 1996; Robbins 2007 among others) and to stage rituals or divination (Malinowski 1965). But language use more than language repertoire, it seems to me, was the essence of power: language use as *action* and language use as *performance*. Early on

Malinowski noted the importance of language as action. In *Coral Gardens*, he says: "language is primarily an instrument of action and not a means of telling a tale, of entertaining or instructing from a purely intellectual point of view" (1965: 52). Language use or speech and performance are inextricably linked, in that language use reveals knowledge, performance "informed" (see Fabian 1990: 11) and both are central to power. Beyond their aesthetic dimensions, speech events and performances such as the singing of songs, the telling of myths, and the recounting of genealogies "formalize stored knowledge throughout Melanesia" (Lindstrom 1990: 107) and give prestige to those who know and control this knowledge. Yet, languages themselves were never at stake. Language use was.

Colonization, post-colonial politics and urbanization have significantly altered this context: new languages appeared on the language scene along with different ways of thinking about language. In the early days of the colonial period, various church denominations freshly arrived in the archipelago were looking for local souls, and introduced missionary lingua francas to facilitate their task.[4] Britain, the colonial power, introduced English and imposed it as the language of the administration. Finally, Pijin arrived on the footsteps of the returnees from the sugar cane fields of Queensland.[5] These *new* languages found their way into the local social scenes together with new ways of thinking about languages, where expatriates were ranking the languages at hand along a hierarchy that saw English at the top, Pijin at the bottom, and the local languages somewhere in between.

Solomon Islanders recognized the advantages derived from speaking the language of the *white man* and Pijin. Commenting on the role and importance of linguistic knowledge in colonial Papua New Guinea, Sankoff writes:

> But knowledge was another matter, and especially valuable was the knowledge of Tok Pisin, referred to as *bubum ayez* 'the white men's language', the language which permitted access to the colonial society. (1980a: 21)

The same was true in the Solomon Islands where some speakers were quick to realize the usefulness of being able to speak with the *white men*. Those who already knew Pijin, were able to secure employment in the colonial society: they became

---

4.   For instance, the Anglican Church in Solomon Islands used Mota, a language from the Banks archipelago in Vanuatu. The Methodist Church made use of Roviana, a language of the New Georgia group, as their missionary lingua franca in that part of the Solomons. The South Sea Evangelical Church, which originated in Queensland, and established mission stations on Malaita at the beginning of the 20th century, used the Queensland pidgin English spoken by the returnees from Queensland.

5.   I documented in Jourdan (1985 and 2007b), the various social events that have shaped the history of Pijin in the Solomon Islands.

house boys or house girls; they joined the constabulary; enrolled as plantation labor; or became catechists. A few others tried to learn English (Fifi'i 1989).[6] Knowing particular languages, rather than speaking one's own well, opened the door to social advancement and to other types of social differentiation, and even better, to paid employment. This new concept resulted in the expansion of the linguistic repertoire of the Solomon Islands. For some of them, it now also included Pijin (the lingua franca) and Mota or Roviana (the languages used by some missions), and yet for others it might have included English as well.

In terms of ideology, the difference was dramatic. The ethos of linguistic reciprocity that fostered reciprocal bilingualism and multilingualism was now challenged by a new linguistic ideology that recognized *value* to languages, and concomitantly, to their speakers. This new way of thinking about languages transformed them into quasi commodities: goods to be had. In contrast with the precolonial period, and with regards to colonial relationships, the language one knew rather than how one spoke it was now the avenue to status and power in colonial Solomon Islands. The languages of the Solomon Islands were not subjugated to English the way their speakers were subjugated by the colonial power. Yet it became clear that a new linguistic order developed and toppled that which existed before. The immediate outcome was the development of some form of diglossic relationships between the languages. As Gal said about colonial linguistic situations in general:

> … relatively egalitarian linguistic diversity, based on small-scale languages whose speakers believe their own language to be superior, [has been changed] into stratified diversity: local languages are abandoned or subordinated to "world languages" in diglossic relations … (Gal 1989).

Urbanization, and particularly the development of the capital city of Honiara, brought further changes to the situation. As the country became independent (1978), Solomon Islanders flocked to the town. The population censuses taken over the years show that many of the country's sixty-four languages are represented to some degree in town. Not to the same extent, of course: in some cases, only a few speakers of any given language live in town. Rapid social change also took place: education levels increased; nuclear families became the favored model of residence (a model difficult to emulate given the shortage and cost of housing in Honiara); social differentiation kept increasing; a middle class developed and is now reproducing itself (Gooberman-Hill 1999).

---

6.   Interestingly, many Pijin speakers thought, and some still do, that they were speaking English every time they spoke Pijin.

Yet, with all this rapid social change, ancestral languages remained strong and central to the life of the citizens. In fact, for a long time the following hypothesis held true: in situations such as that encountered in Honiara, rapid social change "has not led to the disappearance of webs of meaning that are sustained by various languages in various contexts. Rather, the diversity of context of use and ideologies at hand has effloresced, requiring individuals to remain multilingual in order to belong and be in the world" (Jourdan 2005: 4).

Very active circular migration between the islands of the Solomon archipelago and the capital city of Honiara ensured, and continues to ensure, a continual flow of cultural and linguistic contact exists between rural areas and Honiara. When people started to settle in town, their links with their home places, languages and cultural worlds of origin never really severed. They were kept alive by a stream of visitors, migrants, and would-be adventure seekers, even though they, themselves, might not have returned home. The result is a Babel-like town where multilingualism is rampant and where vibrant ethno-linguistic communities exist. They ensured that the home languages would have a role to play in the daily social lives of the urbanites. Learning Pijin was not an option any longer for anyone who wanted to have a life in town and find a job. English was part of the school curriculum: all children at school were exposed to it, successfully more or less.

For a long period in the history of Honiara (1950–1980), the home languages and Pijin coexisted on an even footing, each in their special sociolinguistic niches: home languages were spoken with members of one's ethnic group; Pijin was used with everyone else. I argued elsewhere (Jourdan 2007b) that the expansion of Pijin as the lingua franca of the country was curtailed, during most of its history: (1) by the fact that most adults were bilingual or multilingual; and (2) that Pijin was not associated with a cultural world that had legitimacy. This is no longer the case. At some point in the town's history, an ideological shift took place in association with the development of an urban identity. It was more typical of the young urbanites than the older ones but the trend was there. After independence in 1978, it became accentuated and progressively tilted the linguistic practice of urbanites in favor of Pijin and away from vernaculars (Jourdan 1985).

2.   Of repertoires and shift in Honiara

More recently, the balance of power between all the languages at play in Honiara has changed once again. The linguistic ideology that was dominant during pre-colonial times (reciprocal multilingualism) and the one associated with the colonial order (linguistic hegemony and hierarchy) seem to have been replaced by multiple and competing linguistic ideologies where contextuality rules. The linguistic

hierarchy created by the colonial order seems to have been challenged by the fact that Pijin is now the de facto national language of the country: it is used as a second language in the archipelago and has become the main language of the town for all concerned, adults and children.[7] In the course of schooling, children have picked up English with more or less fluency, depending on social class and the school they went to. A great number of Honiarans have been raised with Pijin as their sole mother tongue or concurrently with a vernacular. These were the children I worked with during my first fieldwork in Honiara in 1982. They are now raising their own children exclusively in Pijin. Most of them do not know the mother tongue of their parents. This is even more the case when their parents belong to different ethno-linguistic groups (see Sankoff 1980c; Smith 2003; for Papua New Guinea, and Barnèche 2004 for New Caledonia). For all of them, gone is the pretence that they might teach *langgus* to their children, a pretense that their own parents had entertained: they simply are not able to do so. In addition, many of these young adults do not see the necessity for it. It is clear that some measure of language shift is taking place across generation and that the dyad *langgus*/Pijin is being replaced by the dyad Pijin/English.

To verify this hypothesis, a random sample of eighty households of Vura, an inner suburb of Honiara, was surveyed in 1997.[8] For this study, only individuals above the age of five and who had lived in Honiara for at least five years were selected to participate. Given the pattern of social and geographical mobility, this selection eliminated individuals who were visiting from the rural areas for short times and whose language practices were different; it also eliminated very young children. A total of 412 individuals were included. Table 1 shows how the various repertoires are distributed across gender lines and age groups.

A cursory look at Table 1 reveals three phenomena:

1.   Attrition of the number of languages that are part of one's language repertoire in the course of generational change. A comparison of line 3 with line 12 in each of columns C (number of individuals speaking two ancestral languages plus Pijin) and E (number of individuals speaking two ancestral languages plus Pijin and English) shows the trend clearly.
2.   Obvious generational language shift away from the vernaculars and in the direction of Pijin and English. In columns A and F, we observe that no speakers

---

Table 1. Numbers of speakers by age group and type of repertoire (*Vern.* = vernacular language; *Pij.* = Pijin; *Eng.* = English)

	A Pijin		B 1 Vern. + Pijin		C 2 Vern. + Pijin		D 1 Vern. + Pij. + Eng.		E 2 Vern. + Pij. + Eng.		F Pijin + English		Totals
	n =	%	n =	%	n =	%	n =	%	n =	%	n =	%	
45–60													
1 Men			7	3.38	4	10.52	5	6.02	9	36.00			N = 25
2 Women			6	2.89	4	10.52	4	4.81	4	16.00			N = 18
3 Total			13	6.28	8	21.05	9	10.84	13	52.00			N = 43
30–44													
4 Men			14	6.76	4	10.52	20	24.09	4	16.00			N = 42
5 Women			18	8.69	5	13.15	13	15.66	2	8.00			N = 38
6 Total			32	15.45	9	23.68	33	39.75	6	24.00			N = 80
15–29													
7 Men	4	10.00	39	18.84	9	23.68	17	20.48	4	16.00	8	42.11	N = 81
8 Women	9	22.50	57	27.53	11	28.94	18	21.68	2	8.00	7	36.84	N = 104
9 Total	13	32.50	96	46.37	20	52.63	35	42.16	6	24.00	15	78.95	N = 185
5–14													
10 Boys	14	35.00	32	15.45	1	2.63	2	2.41			2	10.52	N = 51
11 Girls	13	32.50	34	16.42			4	4.82			2	10.52	N = 53
12 Total	27	67.50	66	31.88	1	2.63	6	7.32			4	21.05	N = 104
Grand total	40	100	207	100	38	100	83	100	25	100	19	100	N = 412

Different types of repertoires

above the age of 30 speaks either Pijin only or Pijin and English only, while 40 people below the age of 30 have Pijin as their sole language (column A) and 19 people in the same age group speak Pijin and English but no ancestral language. Similarly, a comparison of lines 3 and 6 of columns C and D shows a steady progression of English, while a comparison of the same lines of columns D and E shows a loss of ancestral languages.

3.   The young people drive this change, as witnessed by line 9 in all the columns. The differences between line 9 and line 12 in columns D, E and F is explained by the fact that children learn English at school (age 7) and become really fluent in English once they reach the end of secondary school (age 18).

While at the societal level all these trends are clear, at the individual level they are not that straightforward. While it is true that the younger generation born in town heavily influences the trend toward language shift, at every generation the life trajectory of individuals may lead to the acquisition of at least one new language, usually Pijin. When two languages are acquired during the life course, the first one is Pijin and the second is English. These complexities are illustrated in the case study below.

Case Study 1

Diagram 1 (a,b,c) illustrates very clearly the processes of efflorescence and contraction of the linguistic repertoire and the language shift that has taken place in John's family in Honiara since 1983. After a period of expansion of the repertoire through addition of Pijin or/and English (generations 1 and 2) the contraction begins. The vernaculars tend to disappear from the linguistic repertoire of the children born into this family as of the third generation, while Pijin becomes the mother tongue of almost all the children who subsequently learned English. Two processes develop in parallel: the transformation of individual repertoires and the more general reconfiguration of repertoires over four generations. Let us look at each generation in detail.

Diagram 1a: 1983. In generation 1, all adults married someone from their own ethnic group.[9] Mr. 2 and Mr. 3 were important men of their time, busy in government and the public service, and acquired English early on, in addition to Pijin. Note that their wives, who typically stayed behind in the village most of the time, did not even know Pijin. Also note that by 1990, both women made

---

9.   Against my own preference, but for the sake of clarity, I have opted to remove names, even pseudonyms as is traditional, and to use numbers to refer to individuals in this study. It allows the readers to easily follow the demonstration while protecting the anonymity of the participants in this study.
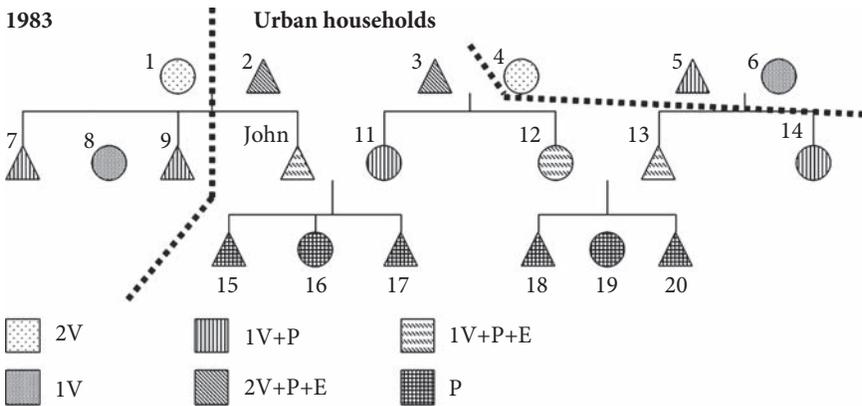
1983   **Urban households**

2V

1V+P

1V+P+E

1V

2V+P+E

P

Diagram 1a. Language repertoires of John's family in 1983.

several lengthy sojourns to Honiara and were able to converse in Pijin. Mr. 5 and Mrs. 6 lived on a small Polynesian atoll on which every one belonged to the same language group. Mrs. 6 never left the island and remained unilingual in Anutan language until her death; Mr. 5 acquired Pijin while working on a coconut plantation located close to Honiara.

In generation 2, all couples married someone from another ethnic group; they met their spouses in town. It is well established that urbanization increases the choice of marriage partners, and these couples are good examples of this trend. John (Mr. 10), well educated and upwardly mobile, acquired English at school and only uses it at work in formal settings. In most other cases, like most other Solomon Islanders, he uses Pijin. His repertoire is similar to that of Mr. 2 and Mr. 3 and typical of young urban professionals who grew up in rural areas but migrated to town early in their life. Symbolically and practically, he straddles the divide between village and town. His brothers (Mr. 7 and Mr. 9) have remained in the village: their language repertoire is typical of young rural males who have traveled in the archipelago in search of work on plantations or in logging companies where they learned Pijin. They have completed primary schooling. Mrs. 12 and Mr. 13 are well educated and speak English fluently in addition to Pijin and their own mother tongue. They have well paid jobs for which good command of English is required, even though most of their day-to-day professional interactions take place in Pijin. Ms. 14 arrived in town a year earlier to serve as a house girl in her elder brother's household; she learned Pijin very quickly.

In generation 3, Pijin is the mother tongue of all these urban children. They live in households where Pijin is the only language of communication amongst

children, between parents, and between parents and children. They do not know any vernacular nor are they being taught any, except for the terms they use to address their grandparents. They speak to the latter in Pijin. Mrs. 11 and Mrs. 12, two sisters, repeatedly told me that it was important for the children to learn Pijin as it was the language of the town. If opportunities arose, the children would learn a vernacular. There are plans to send them to private schools where they will be able to learn English well. To the contrary, both young fathers would like for their children to learn their home language. They saw it as their responsibility to pass on *kastom* (custom practices, tradition) to their children, they say, and ancestral language was part of it. Yet, they were not able to find the time or to establish language routines that would allow the children to learn it. They insisted that children be sent home to the village during school vacations so that they could connect first hand with local customary practices. But the children found it difficult to learn the home language and people accommodated them in Pijin as best as they could. The children quickly gave up learning *langgus*.

The effect of gender on language shift or language maintenance is now well established: The situation I present here resonates very well with those described by Gal for Austria (1979) and more recently with Garrett for Santa Lucia (2005), and Cavanaugh (2006) for Northern Italy. Cavanaugh makes it explicit that language gendering plays an important role in language shift. Women make choices from sociological positions that affect language choice. In her study of the social interplay of Bergamesco, a language of Northern Italy, and standard Italian, she shows that women play a key role in language shift to standard Italian.

The two sisters of generation 2 (Mrs. 11 and Mrs. 12) made linguistic choices that indexed their social positioning and their understanding of the dominant linguistic practice of the town. They wanted an urban life for their children and told me, time and again, that life in the village was simply not an option for them or their children; they saw Pijin as the language of urban rooting. Outside of their immediate family, the social networks of these two women hardly included members of their own ethnic groups. Their husbands, on the other hand, were very ambivalent. Like their wives, they derived their identity from different meaning systems; but the most important was that which was attached to the *wantok* networks (networks defined by shared vernacular, where vernaculars ruled) and to the world of tradition.[10] However, the situation is very different from that described by Kulick (1990) for Gapun. He explains that Tok Pisin is the language used by men for

---

10.   *Wantok* is a Pijin word, from English "one" and "talk": literally, someone belongs to the same language groups. Today the semantic field of *wantok* is larger than it was when the word first appeared in Pijin (early plantation period) and is used also to refer to fictive kin and to address close friends.

contact with the outside world and that the women are the *guardians* of the vernacular. In Gapun, men are driving language shift. This is probably also true of rural Solomon Islands, as was explained earlier. However, in Honiara, the situation of women is radically different from that of women in the villages. This new position, which sometimes transforms them into housewives living in multiethnic residential neighborhoods, or into a wage earners, demands knowledge and use of Pijin in contexts similar to that usually associated with men's activities: work, exchange, inter-ethnic communication. Their social networks extend outside of the ethnic networks into networks where Pijin is central to sociality. Pijin thus, is a form of empowerment for urban women.
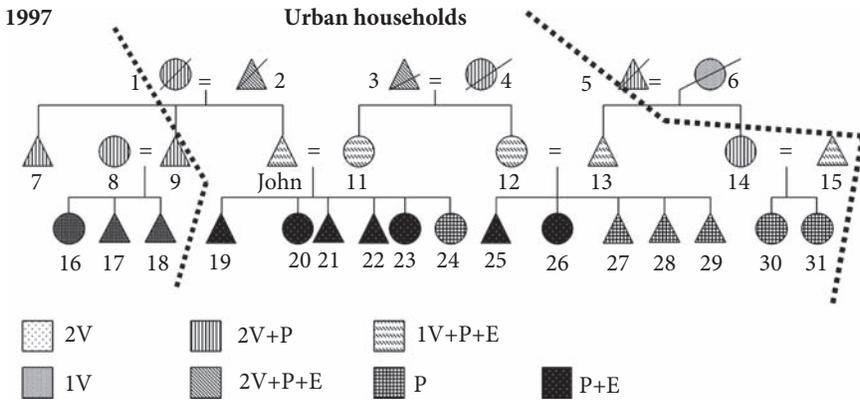

Diagram 1b. Language repertoires of John's family in 1997.

Diagram 1b, 1997. By 1997, many changes took place in the linguistic repertoire of this family. After repeated and lengthy sojourns in town, Mrs. 1 and Mrs. 4 added Pijin to their repertoire and by the time they died, both were able to converse very fluently in Pijin. For professional reasons, Mr. 10 (John) and Mrs. 11 moved their family to New Zealand for a couple of years. Mrs. 11 learned English informally and all their children were sent to school. They formally learned English at school and perfected it through playing with friends and watching television. By the time they returned to Honiara, all the children of that family, except the last born, were fluent in English. They ranged between 18 and 4 years of age. Their cousins, Mr. 25 and Ms. 26 graduated from private secondary schools. Their English was also very good: it gave them access to well paid employment in the public service. Their younger siblings were still at school where they were exposed to English. By that time, Mrs. 12 and Mr. 13 abandoned the idea of sending their children to private schools (too expensive) and sent their younger children to the local neighborhood

schools. Ms. 14 married a young educated man from her own ethnic group, who had found employment in another Pacific Island country. They moved there with their two young girls who, despite living in a monoethnic family, were never taught Anutan, the language of their parents. They were raised in Pijin only. Back in the village, the other children of Mrs. 8 and Mr. 9 were raised in *langgus* only. When they came to visit their uncle in Honiara, they were not able speak to their cousins until they picked up Pijin. In general, it does not take a long time to learn Pijin, given the close proximity of the morpho-syntax and the phonology of this particular *langgus* and Pijin.
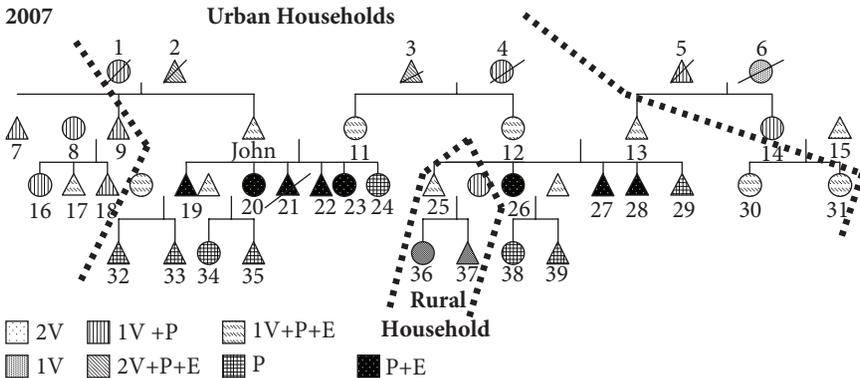

Diagram 1c.   Language repertoires of John's family in 2007.

Diagram 1c, 2007. Ten years later, the repertoire of the adults of the second generation has remained unchanged except for Mrs. 8. In order to bring in some needed cash, and for the last five years, she has made a habit of coming to town once a month to sell her garden produce and her betel nuts. During her so-journs in town, she has picked up some Pijin that comes in handy for her business transactions.

On the other hand, much change is taking place in generation three. Mr. 17 has now acquired Pijin and English. He lives in Honiara where he is studying at the college level. Compare his situation to that of his elder sister, Ms. 16. As is often the case, the girl was pulled out of school at the end of the primary level, and is living in the village where she is helping her mother with the gardening. Her repertoire remains unchanged.

Participant Mr. 25 has an interesting history. When visiting the home place of his father on Anuta, he fell in love and decided to stay on the island. He has been living there for almost eight years and works as a schoolteacher. He is now fluent in Anutan. His wife, who has left the island a few times, has acquired some Pijin while visiting family in Honiara. Also interesting is the case of Ms. 30 and Ms. 31. The two sisters moved to Tuvalu with their parents when they were very young.

At the time, their only language was Pijin. In Tuvalu, where they lived for ten years, they learned Tuvaluan along with English at school. They have just returned to Honiara for further studies: knowing Pijin and English has served them in good stead after such a long sojourn overseas.

Three members of this generation are now married with children. In all cases, their spouses were born in the rural areas and each of them grew up with a different vernacular as a mother tongue. They acquired Pijin and English when they moved to town for greater educational opportunities. All three have advanced education training and have paid employment in the public service. Mr. 19 is married to someone from his father's home village. Yet, he himself does not know the *langgus*, they use Pijin in all contexts of family life.

It is clear that the changes taking place in generation three are either linked to the migration to town (the subject often acquires Pijin in the process) or to rural areas (the subject acquires a vernacular in the process). In all cases, increased education levels are linked to the addition of English in the repertoire of the younger speakers. However, English is always kept for formal settings such as work, encounters with expatriates, formal addresses, etc. These observations corroborate the process of repertoire expansion in members of the preceding generations.

As to generation four, all its members are exclusively raised in Pijin. Ms. 34 and Ms. 38 go to an exclusive primary school. The father of Ms. 38 explained that it was essential for his daughter (and later on for his son) to go to the best school possible and learn English early. This, he said, demanded financial sacrifices on their part as the school fees are very expensive. They saw it as an investment that would ensure their daughter's academic success in the future. By insisting that English, the socially dominant language, was essential to their daughter, Ms. 38's father expressed his class position in a sociolinguistic situation where language shift is taking place. As with John twenty-four years earlier, the father of Ms. 38 lamented the fact that his children did not know his home language. He himself spoke it fluently, but he never addressed his children other than in Pijin (or once in a while English). During the time I lived with this family, both parents enlisted my help in tutoring their daughter in English every day after school. By contrast, in the same urban household, two women (aged fifty-four and eighteen) were totally illiterate: they had never been to school. They were fluent in vernacular and Pijin.

Despite the fact that this family belongs to the middle-class, it is not exceptional at all. The data I have of the ten Honiara families with whom I have closely worked during the last twenty-five years exhibit the same patterns: expansion, contraction and reconfiguration of repertoires are common throughout. This is to be expected in societies in rapid flux where social units are being reconfigured and new ideologies and lifestyles appear. In all the cases in my data, changes to

language repertoires index life trajectory and language ideologies and are associated with geographical and social mobility. As we shall see below, they reveal societal trends that seem to be more accentuated among middle-class families.

3.   Language and the middle-class in Honiara

During the last twenty years, scholars of urban Melanesia have paid attention to the development of urban cultures and social differentiation. Of particular interest has been the development of social stratification and of an urban-based middle-class: it is a growing phenomenon in the contemporary Pacific. Well educated, and sometimes, quite cosmopolitan, this middle-class often acts as a cultural broker between the outside world and the country they live in. As to the South Pacific elites, and because of a shared regional economy, it seems that the elites and the middle-class of different Pacific Island countries resemble each other culturally more and more (Hau'ofa 1987). Observations by Philibert in Vanuatu and Jourdan in Solomon Islands (Jourdan & Philibert 1994; Gooberman-Hill 1999) and in Papua New Guinea (Gewertz & Errington 1999) show that urban local elites play major cultural roles: they serve as a cultural buffer between the local cultural worlds and the various incarnations of globalisation; they shape the development of local urban popular cultures by setting up youth development and activity programs; they foster the development of the nation through school curricula and the ideology of *kastom*.

In the Solomon Islands, the middle-class is a new social feature that has developed over the last fifteen years in association with the localization of government and business jobs that followed Independence. Many of its members are themselves children of individuals who had acquired some standing during the colonial period. Some family names have indeed become household names. For the project I am currently researching in Honiara, the Honiaran middle-class is identified according to the following criteria: income; occupation; status in church; leisure activities; general consumption patterns; social relationships; home ownership; knowledge of English (see Gooberman-Hill 1999). As Angeli remarks (2008) this Weberian approach to the middle-class, also used by Gooberman-Hill (ibid.) for the Solomons, Gewertz & Errington (1999) for Papua New Guinea and James (2003) for Tonga, puts the emphasis on similar patterns of consumption, discourse and desirable lifestyle. Accordingly, members of the middle-class can be found in various occupational niches: owners of small businesses; low ranking civil servants; secondary or tertiary education teachers; well-educated middle range professionals. At the upper spectrum of the middle-class, to which my friend Diake belongs

(see case study 2 below), political connections and economic networking associate members of the upper-middle class to the elite from which they are otherwise clearly distinguished by a lack of financial means that limit their aspirations.

What is particularly interesting here is that class and new patterns of language ideology and use are emerging at the same time. Sociolinguists will not be surprised by this state of affairs. Language (here English) is used as a tool for social differentiation and for distinction; it is part of the process of emergence of class. Thirty years after Independence, it is clear that Honiarans, in general and the members of the middle-class, use language to signal affinity with and departure from various social worlds that seem to confront each other. Playing with different languages and different registers, the adult members of the urban middle-class express social difference through language use. The case of my friend Diake is particularly telling.

Case study 2

A senior member of the public service and the elder son of a Bigman from Kwara'ae, on the island of Malaita, Diake is forty-five years old. Married to a woman from Lau, also on the Island of Malaita, Diake speaks four languages: Kwara'ae, his mother tongue; Lau, the language of his wife; Pijin, which he learned in school through contact with other children; and English which he learned in high school and perfected in New Zealand.[11] By Solomon Islands standards, Diake has done well in life: he owns a large house in a nice suburb of Honiara, drives an imported second hand car, and sends his children to a private school. He is also a member of the Honiara baseball club and of the local yacht club; he regularly travels overseas on official trips; rents videos almost every night; and under pressure from his wife, is increasingly trying to keep at bay the demands from his kin and members of his lineage.[12] Diake is the only wage earner of his family and the rest of his siblings live on Malaita through subsistence agriculture and fishing. Like so many wage earners in the Pacific, Diake sends remittances to his village-based kin that allow him to keep his place in the system of obligation and exchange. When called upon, he participates in bridewealth exchanges, pays compensation for mortuary feasts, and offers a pig in sacrifice in order to propitiate an ancestor, the latter being particularly noteworthy in that Diake is a practicing member of the Seventh Day

---

11.   In urban school and in some regional provincial schools, Pijin is the language spoken in the school playground.

12.   This is seen as an indication that he is wealthy enough to own a VCR and TV, and can afford the daily rentals.

Adventist Church, a Christian fundamentalist church that bans the consumption of pork.

Diake has progressively developed middle-class values that increased the distance between him and his kin in the village. For instance, he is often berated by his kin and extended family members who accuse him of selfishness; they reproach him for not sharing with them everything they see as his wealth. This, they claim, is a sign that he no longer respects the core social values that the villagers hold dear. But he differs from his kin in many other ways, none as spectacular as how he makes use of the languages he knows. It is clear that in many occasions the languages at play are not simply marking ethnicity as was the case in pre-colonial times. For people like Diake, they now index social class, age group, gender and urban identity. Predictably, Diake will use the Kwara'ae language with his Kwara'ae friends or kin, but he may also use Pijin or English with the same interlocutors according to context. If he switches to English, a language he speaks fluently, in the middle of a sentence in Pijin, he signals that he belongs to a social world most of his interlocutors do not know. Or that he has access to new forms of knowledge by knowing English. English increases his status and social standing. Conversely, using Pijin when people expect Kwara'ae puts Diake's intervention in another frame: it signals distancing from the *wantok* networks.

Alternately, all these languages allow Diake to establish distance between *we* and *they*; between *informal* and *formal*; between *private* and *public*, etc. In most cases Kwara'ae indexes the *we*, and the *private*; and sometimes, according to context, it indexes the *formal*. Pijin most often indexes the *we* and the *private* but never the *formal*. All this code-switching allows Diake to play the linguistic market and come out ahead.

What is particularly interesting about Diake and fellow members of the middle-class who belong to the same age group and have a similar life trajectory, is the two essential roles they play in language shift. On the one hand, they are cultural buffers between the world of their parents, village-based, and that of their children, born and raised in town. They live in a continuum of cultural experiences that require them to enrich their linguistic repertoire. As they move into forms of sociality that are organized around different sets of values (associated usually with an ideology of the *modern*) they are not jettisoning old systems of meaning and belonging that are shaping their primary identity. Throughout their life though, this primary ethnic identity may become subsumed to class interests, as I have explained in Jourdan (1996, 2007a). For the children of the middle class who have been raised in town, and particularly if they have experienced the cultural world still predominant in the villages their parents are from,

the situation is drastically different. At the level of discourse, some measure of identity may be proclaimed in ethnic terms (*Mi blong Sa'a* 'I am from Sa'a'; or *Mi blong Soesol* 'I am from Choiseul'). At the level of practice, much of how they conceive of their identity is revealed in ways of acting, consuming, speaking, that cut across ethnic lines and embrace social affiliations. While the repertoire of their own parents consisted of one or two vernaculars, plus Pijin, and sometime English for the most socially prominent of them, that of their own children is much simpler and is often limited to Pijin and English, and for the young children, to Pijin only.

It is true that some ethnic groups such as the Lau of Malaita, or the people from Ontong Java, for instance, are known in town to encourage the learning of *langgus* across generations. They constitute tight cultural enclaves in Honiara and until recently, tended to in-marry. This is changing though, and the language practice of the younger urban segments of these ethnic groups tend to resemble more and more that of the other young urban people. Diagram 2 illustrates this shift.
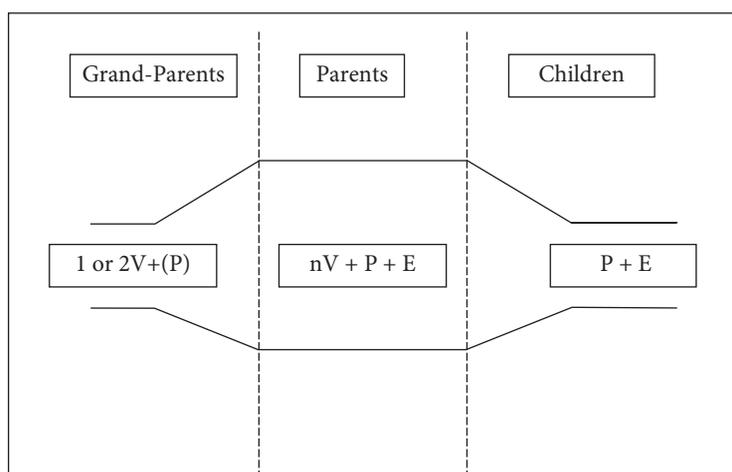

Diagram 2. Shift in urban language repertoires across generations of the middle-class.

The diagram shows that the process of expansion or attrition of the collective repertoire takes place over a few generations. Of course, there is a high degree of individual variation in each of the generations mentioned here and at each transitional moment. For instance John's father and father-in-law (case study 1) knew English but it was not common for people in that generation to do so, and while their repertoire expanded, that of their wives and other men in their generation,

did not. A distinction between societal and individual expansion and contraction needs to be introduced at this juncture as:

1.   Both do not necessarily take place at the same time. In other words, and as is clear from case study 1, individual life trajectories may anticipate or lag behind collective efflorescence.
2.   They do not necessarily take place for the same reasons. While it is fair to say that the collective efflorescence found in Honiara in generation two is linked to an increase in social opportunities and demands associated with urbanization, it is not true that all speakers will respond to them in the same manner, or that urbanization per se is a deciding factor.
3.   They do not take place at the same rhythm. While in some families, and for individuals within families, efflorescence and contraction may take place in two or three generations, in others they may take place over four or five. The factors that mitigate the impact that new linguistic ideologies and practice associated with urbanization and with *modernity* may have on repertoires are linked to ethnic ideologies of belonging and membership.
4.   Finally, individual shift may not take place at all. Factors such as strength of linguistic communities, gender, definition of self (here, rural versus urban; class membership) and opportunities (education; employment) explain much of the variations found in individual repertoires at each generation. Yet the societal trend towards a different repertoire is clear from the data (Table 1). This type of shift is also typical, with the caveats expressed above, of the second and third generation immigrants in North America or with rural migrants to the urban centers of multilingual and multiethnic states (Gal 1979) and in rural areas of Papua New Guinea for instance (Kulick 1990).

As can be expected, the middle class of Honiara is not homogenous: social groups range from lower middle-class to upper middle-class along a continuum of values, practices and ideologies that are separating its members from the working class segment of the urban population. Defining factors involve: levels of income; occupation; political connections; economic networking; status in church; leisure activities; general consumption patterns; social relationships; home ownership; and fluency in English (see Gooberman-Hill 1999).

Ethnographic research carried out in Honiara since 1982 reveals that the general reconfiguration of urban repertoires is linked to various sociographic factors: amount of time spent in Honiara; level of education; mono or bi-ethnic marriages; contact with *hom* (place of family origin). These variables play out differently according to individuals and according to households. They influence the choice of language, or combination of languages that will be used as the main

language of households. Table 2 shows these differences when sorting the data of Table 1 according to social classes and mono-ethnic (language endogamous) or bi-ethnic (language exogamous) households.

Table 2. Number of middle class and working class households by type of repertoire, Honiara 1997

Main language at home	Middle class					Working class				
	Mono-ethnic households		Bi-ethnic households		Total	Mono-ethnic households		Bi-ethnic households		Total
	N	%	N	%	N	N	%	N	%	N
Vernacular only	4	50	2	6.25	6	10	62.50	2	8.33	12
Vernacular and Pijin	3	37.5	5	15.63	8	5	31.25	8	33.33	13
Pijin only	1	1.25	25	78.12	26	1	6.25	14	58.34	15
Total	8	100	32	100%	40	16	100%	24	100%	40

Given the size of the sample, a statistical analysis is not significant, but I have opted to leave the percentages in place in this Table only for the sake of facilitating the comparison between households. However the raw numbers speak for themselves and show the trends and allow us to reach three conclusions.

First, there are more linguistic exogamous couples than linguistic endogamous ones. This confirms that urbanization increases the choice of marriage partners.

Second, and despite the first point above, it is among the middle-class that we find the highest number of linguistic exogamous couples. Two essential factors explain this situation: couples of the middle class are younger than those of the working class: 57% of middle-class couples are below the age of 35 compared to 32% in the working class. Few among these middle-class couples are migrants in Honiara; most of them are born in town of parents who migrated there. These younger people are also better educated. Social mobility is favored by better education and by parental family lifestyle and values.

Third, it is among the middle-class that we find the highest number of households using Pijin exclusively as the main language. Again this correlates with the number of exogamous couples and their age.

To the sociographic variables enunciated above, one must add more ideological variables such as the school selected for one's children; the linguistic ideology and practice that governs the construction of one's language repertoire; the social networks one belongs to; the life trajectory one builds for oneself. I call these variables 'sites of distinction' (see Bourdieu 1982). They are central to the construction of one's social urban identity and are revealed by language repertoire and language use.

4. Conclusion

Two arguments informed this article: The first is that language is not simply a means of communication or a means of action; language selection also reflects the speaker's agency on the road to a constantly redefined sociality. What is better than the inescapability of language to index one's places in a social world that one contributes to create? The second research argument is that the relative status of the languages at play in the Honiara social scene (64 vernacular languages, one pidgin language known locally as Pijin and English, the language inherited from the colonial period) articulates with the social worlds that frame speech events. Each of these languages has acquired legitimacy and is associated with different spheres of meaning: the vernaculars are the language of *kastom*; Pijin is that of urban life; and English that of social advancement. Vernacular languages have not disappeared from the linguistic landscape of the town because the cultural contexts where they are primordial (ethnic cultural meaning and belonging) are still strong (Jourdan 2007a), and because the urban pull is so strong. Generation after generation, each new rural migrant to Honiara contributes, in the short term, to the strengthening of the vernaculars and enriches the pool of *langgus* speakers. But the same migrant contributes in the long term to the societal weakening of the same *langgus* if s/he wishes to stay in town and raise their children there. The data presented here have shown that the expansion of the repertoire of the newly arrived individual is transformed into a reconfiguration and contraction of the repertoire in the next generation. This contradiction bears heavily on the general linguistic situation in the Solomon Islands.

An additional quirk is the development of urban registers of Pijin: some are more or less basilectal, others are more or less acrolectal. Each is associated with particular age groups and social segments of the society: the acrolectal with the young urban crowd born and raised in town; the basilectal with the older folks who have maintained strong ties with the *village* and whose Pijin is heavily influenced by their vernaculars. But while some segments of the young urban crowd find the basilectal varieties to be *old fashion* and often make fun of those who speak them, others associate these varieties with *true* Pijin. Yet, the linguistic distance between the generations is increasing: not only are the linguistic repertoires reconfigured, but value judgments are attached to registers. Each generation seems to have developed acceptance levels of the varieties of Pijin and to recognize implicit norms about good Pijin and bad Pijin. Among the young members of the well-educated urban middle class, much code-switching is taking place between Pijin and English in a linguistic continuum such that it is difficult to know where Pijin stops and where English begins. This situation is reminiscent of that described by Smith for Tok Pisin. He says: "Between the two poles lies a grey area where it may

be difficult to say exactly what language certain forms belong to" (Smith 2003: 202). It is my contention that in many linguistic situations in Honiara, code-switching is the code.[13] Once social groups stabilize and new cultural formations appear, here an urban cultural world, how one uses the languages in play is essential to further social distinctions.

In Honiara, the trend in the reconfiguration of repertoires is clear: the dyad Pijin/English is progressively replacing the dyad *langgus*/Pijin for long term residents of Honiara and particularly for those who were born and brought up in town. The change takes place along life cycle, age grading, and along social classes. For the moment, powers still reside in the hands of those whose linguistic repertoire is the richest: the urban adults of the middle class who speak *langgus* along with Pijin and English. This complex multilingualism reveals memberships into different sociocultural worlds that often collide with each other, but which must be navigated in order to remain ahead.

In addition, the data show that the shift to the dyad *langgus*/Pijin seems to be driven by the young urban middle class, the segment of urban population who will take over power from the current multilingual urban generation. And not surprisingly, as Gillian Sankoff (1980a) told us, urban power will reside in the following generations, in the hands of those who control the two languages.

References

Aikhenvald, A. 2002. *Language Contact in Amazonia*. Oxford: OUP.

Angeli, J. 2008. 'Mi no waetman, I mean': Attitudes toward English and Pijin among Solomon students. MA dissertation: Concordia University.

Barnèche, S. 2004. *Gens de Nouméa, gens des îles, gens d'ailleurs: Langues et identités en Nouvelle-Calédonie*. Paris: L'Harmattan.

Bourdieu, P. 1982. *Ce que parler veut dire: l'Économie des échanges linguistiques.* Paris: Fayard.

Cavanaugh, J. 2006. Little women and vital champions: Gendered language shift in a northern Italian town. *Journal of Linguistic Anthropology* 16(2): 194–210.

Fabian, J. 1990. *Power and Performance: Ethnographic explorations through proverbial wisdom and theater in Saba, Zaire*. Madison WI: The University of Wisconsin Press.

Fifi'i, J. 1989. *From Pig Thefts to Parliament: My life between worlds*, Roger Keesing (Ed.). Honiara: SICHE and the University of the South Pacific.

Gal, S. 1979. *Language Shift: Social determinants of linguistic change in bilingual Austria*. New York NY: Academic Press.

Gal, S. 1989. Language and political economy. *Annual Review of Anthropology* 18: 345–367.

---

13.   I have deliberately kept away from engaging the literature on code-switching in this article as it deserves fuller treatment than I have space for here.

Garrett, P. 2005. What language is good for: Language socialization, language loss and the persistence of code-specific genres in St. Lucia. *Language in Society* 34(3): 327–361.

Gewertz, D. & Errington, F. 1999. *Emerging class in Papua New Guinea*. Cambridge: CUP.

Gooberman-Hill, R. 1999. The constraints of "feeling free": Becoming middle-class in Honiara. Ph.D. dissertation, University of Edinburgh.

Hau'ofa, E. 1987. The new South Pacific society: Integration and independence. In *Class and Culture in the South Pacific*. A. Hooper et al. (Eds), 1–12. Auckland: University of Auckland.

Jackson, J. 1983. *The Fish People: Linguistic exogamy and Tukanoan identity in Northwest Amazonia*. Cambridge: CUP.

James, K. 2003. Is there a Tongan middle class? Hierarchy and protest in contemporary Tonga. *The Contemporary Pacific* 15(2): 309–336.

Jourdan, C. 1985. Sapos iumi mitim iumi: Urbanization and creolization of Solomon Islands Pijin. Ph.D. dissertation, Australian National University.

Jourdan, C. 1996. Where have all the cultures gone? Sociocultural creolization in the Solomon Islands. In *Melanesian Modernities*. J. Friedman & J. Carrier (Ed.), 34–52. Lund: University of Lund Press.

Jourdan, C. 2005. Linguistic paths to urban self. Research Proposal Submitted to Social Sciences Research Council of Canada.

Jourdan, C. 2007a. Linguistic paths to urban self in postcolonial Solomon Islands. In *Consequences of Contact*. B. Schieffelin & M. Makihara (Eds), 30–48. Oxford: OUP.

Jourdan, C. 2007b. *Parlons Pijin: Histoire sociale et description du pidgin des Îles Salomon*. Paris: L'Harmattan.

Jourdan, C. & Philibert, J.M. 1994. Urbi et orbi: construction identitaire et cultures urbaines en Mélanésie. In *Journal de la Société des Océanistes* 99(2): 159–166.

Kulick, D. 1990. Having head and showing knowledge: Language shift, Christianity and notions of self in a Papua New Guinea village. Ph.D. dissertation, University of Stockholm.

Lindstrom, L. 1990. *Knowledge and Power in a South Pacific Society*. Washington DC: The Smithsonian Institute.

Malinowski, B. 1961[1922]. *Argonauts of the Western Pacific*. New York NY: Dutton.

Malinowski, B. 1965[1935]. The language of magic and gardening. In *Coral gardens and their Magic,* Vol. 2. Bloomington IN: Indiana University Press.

Mauss, M. 1938. Une catégorie de l'esprit humain: La notion de personne, celle de moi. *The Journal of the Royal Anthropological Institute* 68: 263–281.

Rumsey, A. 2000. Agency, personhood and the 'I' discourse in the Pacific and beyond. *The Journal of the Royal Anthropological Institute* 6(1): 101–115.

Sankoff, G. 1980a. Political power and linguistic inequality in Papua New Guinea. In *The Social Life of Language: Collected essays by Gillian Sankoff*, 5–27. Philadelphia PA: University of Pennsylvania Press.

Sankoff, G. 1980b. Language use in multilingual societies. In *The Social Life of Language: Collected essays by Gillian Sankoff*, 29–46. Philadelphia PA: University of Pennsylvania Press.

Sankoff, G. 1980c. On the acquisition of native speakers by a language. In *The Social Life of Language: Collected essays by Gillian Sankoff*, 195–210. Philadelphia PA: University of Pennsylvania Press.

Robbins, J. 2007. You can't talk behind the Holy Spirit's back: Christianity and changing language ideologies in a Papua New Guinea society. In *Consequences of Contact*. B. Schieffelin & M. Makihara (Eds), 125–139. Oxford: OUP.

Schieffelin, B. 1990. *The Give and Take of Every Day Life*. Cambridge: CUP.

Smith, G. 2003. *Growing up with Tok Pisin: Contact, creolization, and change in Papua New Guinea's national language*. London: Battlebridge.

Telban, B. 1996. The mystery of identity. *Anthropological Notebooks* 2(1): 5–26.

Land, language and identity

The socio-political origins of Gurindji Kriol[1,2]

Felicity Meakins
University of Manchester

Empirical evidence for the sociolinguistic origins of mixed languages has often proven elusive due to the paucity of historical material on the linguistic and political situation at the point of their genesis. Gurindji Kriol is a mixed language spoken by Gurindji people in northern Australia. The socio-historical circumstances of its emergence are well-documented due to the role that Gurindji people played in Australian politics at the time. Between 1966 and 1975, they led a landmark political struggle to regain control of their traditional lands. Gurindji Kriol found its genesis during this period and its emergence is significant given that many other Aboriginal groups in the area were shifting to Kriol, an English-based creole language. In this chapter I argue that the retention of Gurindji features in the mixed language occurred in parallel with the land rights movement and both can be considered expressions of the persistence of a Gurindji identity.

Keywords: Australian languages; Kriol; Gurindji; Aboriginal peoples; Aboriginal English; language attitudes; language change; mixed languages; cultural change; lingua franca; Wavehill Walk-Off

---

1.   Introduction

Gurindji Kriol is a mixed language which is spoken by the Gurindji people who live at Kalkaringi[3] in the Victoria River District of the Northern Territory, Australia (see map). It combines the structure and lexicon of Gurindji, a Pama-Nyungan language, with that of Kriol, an English-lexifier creole language. Gurindji Kriol is a young language, which only emerged in the 1970s. Its genesis is interesting given that, elsewhere in northern Australia at this time, Kriol was steadily replacing traditional Aboriginal languages (Munro 2000). In this chapter, I examine the socio-political origins of Gurindji Kriol in terms of the historical events, changing social and geographical relations, and the language ecology of the Victoria River District during the period leading up to its genesis. Specifically, I suggest that the maintenance of Gurindji in the mixed language is an act of resistance against the massive cultural incursion which accompanied colonisation. This motivation for the formation of Gurindji Kriol is also reflected in a land dispute between the Gurindji people and non-indigenous occupiers which occurred in 1960–70s. I claim that both of these acts were different arms in the expression of a continuing Gurindji identity, as marked through land or language. I follow the lead of sociolinguists such as Gillian Sankoff (e.g., 1980) in situating language within a social and historical space, looking beyond the description of language as a communicative resource but also as a symbol of identity.

Reconstructing the emergence of contact languages is, by nature, a somewhat speculative exercise, with descriptions of the language ecology and the socio-historical circumstances at the time of genesis frequently based on fragmented historical records and oral accounts, or inferred from the resultant structures of the languages themselves. In the case of mixed languages, the situation is more difficult given the often transient nature of these contact languages. In many instances, they represent a prolonged stage of language change which precedes a shift to the stronger of the languages in the mix, with a notable exception being Michif where the Métis shifted to English as opposed to French or Cree, the source languages (Bakker 1997). Thus the existence of mixed languages often goes by unobserved, with accounts of inter-group relations at the time of emergence and the direction of language shift an interpretative exercise. Hypotheses about their origins are often gleaned, then, from the contemporary speech of the remaining older speakers who are members of speech communities where the mixed language is no longer the main language. The result has been conflicting explanations about the mechanisms

---

3.   I use the umbrella term "Kalkaringi" to refer to the two neighbouring Gurindji communities of Kalkaringi and Daguragu. Kalkaringi is the better known of these, hence the use of this name.
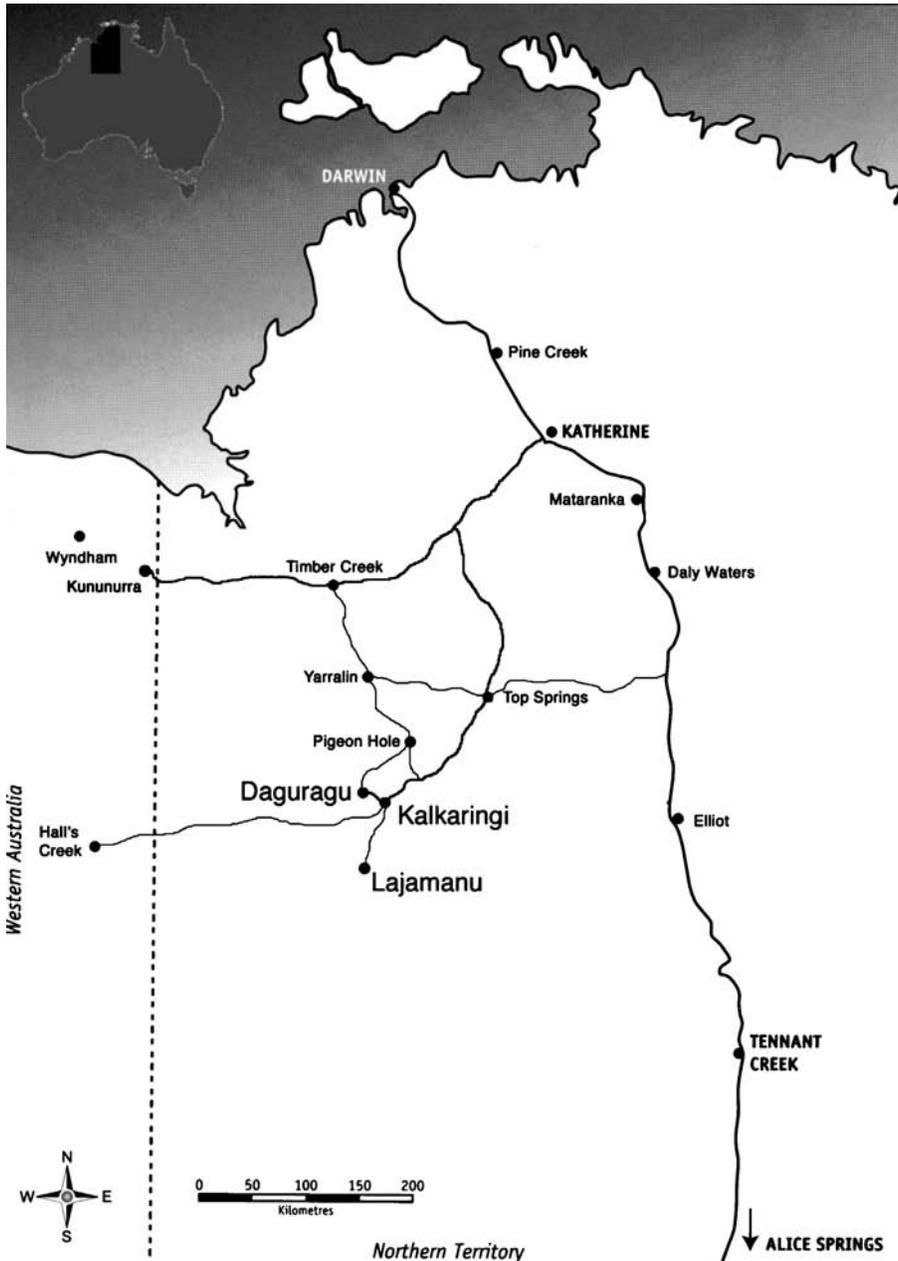
which underlie the formation of these languages. For example Angloromani, a mixed language which combines English grammar with some Romani lexicon, has been variously described as the result of a U-turn where Romani people who had become full English speakers reclaimed their cultural heritage through the selective replacement of English words with Romani equivalents (Boretzky & Igla

1994), or as the massive borrowing of English structure by Romani speakers under intense cultural pressure (Thomason 2001). Even where mixed languages have more currency among younger members of a language community and therefore more and richer data is available, attempts at reconstructing social relations and language shift at the time of genesis has produced conflicting stories, such as Mous (2003) and Thomason's (1997) differing accounts of Ma'á and its formation.

In the case of Gurindji Kriol, detailed information about the sociolinguistic and political circumstances of its emergence is available in the form of oral and written histories, and anthropological accounts of the Gurindji people, as well as descriptions of the linguistic repertoires and practices of the Gurindji people during the 1970s. Much of this interest in the Gurindji people is derived from their influence on Australian politics during the 1960s–70s. Like many Aboriginal groups in north Australia, cattle stations were established on their land and they were put to work in slave-like conditions in the early 1900s. In 1966, the Gurindji initiated a workers' strike to protest against the poor conditions of their employment and ultimately recover control of their traditional lands. Their campaign went on for nine years and resulted in the first successful land claim by an Aboriginal group, which in turn generated further energy for the broader land rights movement. At this time, McConvell (1988) observed that code-switching between Kriol and Gurindji was the dominant language practice of Gurindji people, which was similar to the linguistic practices of many Aboriginal groups across northern Australia. However, where Kriol replaced the traditional language of many other groups and code-switching was indicative of a decline in traditional language use, a mixed language originated from similar circumstances at Kalkaringi (McConvell & Meakins 2005).

This socio-historical and linguistic context is important to understanding where Gurindji Kriol finds its origins. In this chapter, I review the events immediately preceding the emergence of Gurindji Kriol, and the linguistic practices of the Gurindji at the time (§3), I relate these events to an intrinsic link between land, language and identity for the Gurindji. I suggest that the land rights campaign and the continuation of Gurindji in the mixed language were both borne out of the severe social disruption which accompanied the arrival of white colonists, and are both expressions of the persistence of the Gurindji identity. In this respect, the political climate of the 1960s–70s in the Victoria River District contributed to the maintenance of Gurindji which was under functional pressure from Kriol. The effect of these dual pressures of Gurindji maintenance and Kriol influence helped stimulate the emergence of a mixed language at Kalkaringi (§4). The story that I present of Gurindji Kriol is based on my own fieldwork which included informal discussions and more formal (question and answer) interviews with Gurindji

people between 2004–2007, as well as primary and secondary historical materials. All contemporary Gurindji Kriol examples are drawn from my own corpus of 80 hours, and are supplemented by McConvell's Gurindji-Kriol code-switching data from the 1970s. I begin with an overview of the structure of Gurindji Kriol (§2) before suggesting how the formation of this mixed language occurred.

2.   The structure of Gurindji Kriol

The structure and lexicon of Gurindji Kriol is based on the traditional language of the region, Gurindji, which is a highly endangered language, and Kriol, which is the *lingua franca* amongst Aboriginal people in northern Australia. Gurindji is a typical Pama-Nyungan language in that it is a dependant-marking and suffixing language (McConvell 1996). Kriol is an analytical language with an SVO word order, prepositions, transitive marking and independent tense and mood markers (Hudson 1983; Munro 2005; Sandefur 1979). In this sense, it bears a strong resemblance to other English-based creole languages in the Pacific region, including Tok Pisin (Sankoff 1980; Wurm & Mühlhäusler 1985) and Bislama (Crowley 1990; Meyerhoff 2000).

What is remarkable about Gurindji Kriol is the manner in which the source languages have fused. Gurindji Kriol exhibits a structural split between the noun phrase system and the verb phrase system, but is lexically quite mixed. In terms of structure, Kriol contributes much of the verbal grammar including tense and mood auxiliaries, and transitive, aspect and derivational morphemes. Gurindji supplies most of the NP structure including case and derivational morphology. Both languages also contribute small amounts of grammar to the systems they do not dominate. For example, the Gurindji continuative suffix is found in the VP, and Kriol determiners are common in the NP. Kriol also provides Gurindji Kriol with an SVO word order, though the word order is more flexible than Kriol with information structure determining word order to some extent. Complex clauses are constructed using both Gurindji and Kriol strategies, for example coordinating clauses use Kriol conjunctions, and subordinate clauses are formed using Gurindji-derived case and inchoative marking (Meakins 2007; Meakins & O'Shannessy forthcoming).

Example (1) below demonstrates Gurindji Kriol's structural split and lexical mixing. In this example, the core VP structure *i bin baitim im* ('it bit him') including the tense auxiliary *bin* and transitive marker *-im* is drawn from Kriol, while the NP frame, including ergative and locative case marking, are from Gurindji. Note that the lexicon is mixed, and in particular Kriol stems may combine with

Gurindji suffixes. For example, both a Gurindji noun, *warlaku* 'dog', and a Kriol noun, leg, are present, with leg inflected for locative case with a Gurindji form. The Gurindji elements are given in italics, and plain font is used for Kriol elements. I will use this style throughout to differentiate Gurindji and Kriol.

(1) dat *walaku-ngku* i   bin bait-im im   *karu* leg-*ta*
     DET dog-ERG   3SG.S NF bite-TRN 3SG.O child leg-LOC
     'The dog bit the child on the leg.'

The following example typifies the mixed character of Gurindji Kriol. This excerpt is from the Frog story (Mayer 1994 [1969]) and begins as the dog's head becomes stuck in a glass bottle.

(2) dat   *warlaku-ngku* i=m   *warlakap*   *nyila-ngka* botul-*ta*.
     DET dog-ERG   3SG.S=NF look.around that-LOC   bottle-LOC
     'The dog searched (for the frog) in that bottle.'

     abta-*ma* i=m   gon autsaid windou-*nginyi* jing-in-at
     after-DIS 3SG.S=NF   go outside window-ABL   call-CONT-out

     bo   dat *ngakparn*, *warlaku* jeya botul-*jawung* *ngarlaka-ngka*.
     PREP DET frog   dog   there bottle-PROP   head-LOC

     'After that, he (the boy) went outside of the window calling for the frog.
     The dog is there with a bottle on its head.'

     binij i   bin baldan *warlaku* baldan *kanyjurra-k*
     finish 3SG.S NF fall   dog   fall   down-ALL

     windou-*nginyi*, dat *karu* i=m   *karrap* im   baldan.
     window-ABL   DET child 3SG.S=NF look.at 3SG.O fall

     'That's it, the dog fell out of the window, and the child watched it as it fell down.'

In this example, the verbal frame is Kriol with basic meaning verbs such as *gon* 'go' and *baldan* 'fall', and tense marking, *bin* and =*m* (non-future) derived from this language. The NP matrix is predominantly Gurindji. Present is Gurindji inflectional morphology including case marking, for example, ergative -*ngku*, locative -*ngka*/-*ta*, ablative -*nginyi*, and allative marking -*k*. Also present from Gurindji is discourse marking -*ma* and derivational morphology such as -*yawung* (proprietive, 'having'). Lexically there is a mix between Kriol and Gurindji with some verbs derived from Kriol, for example, *gon* 'go' and *jinginat* 'call', and others from Gurindji, for example, *warlakap* 'look for' and *karrap* 'look at'. Similarly nouns from both languages are present – *windou* 'window' and *botul* 'bottle' from Kriol; and *karu* 'child' and *warlaku* 'dog' from Gurindji. The use of inflection does not depend on the language of the stem. For example, case-inflected nouns are of

Gurindji and Kriol origin, *windou-nginyi* 'window_{Kriol}-ABL_{Gurindji}' and *warlaku-ngku* 'dog_{Gurindji}-ERG_{Gurindji}'.

Structural splits between the nominal and verbal systems appear to be quite rare, with grammar-lexicon splits found more often. For example Media Lengua combines Spanish vocabulary with a Quechua grammar (Muysken 1997). The most commonly cited example of a V-N split is Michif, a Canadian mixed language, which combines Cree (VP) and French (NP) (Bakker 1997). Though Gurindji Kriol bears some resemblance to Michif, they differ in their lexical mixes. In Michif, Cree also provides most of the verbs, and French, the nouns. On the other hand, Gurindji Kriol does not follow this language-structure divide. Though Gurindji provides the grammatical frame for the nominal system, nominals themselves are derived from both Gurindji and Kriol. The same is true of the verbal system. In this respect, Gurindji Kriol patterns most closely with Mednyj Aleut, a Russian-Aleutian mixed language (Golovko 1994; Thomason 1997) and Light Warlpiri, a neighbouring Australian mixed language. This language is spoken 100 km from Kalkaringi at Lajamanu (see map) and mixes the structures of Kriol (VP) and Warlpiri (NP). Lexically nominals are also derived from both languages; however verbs are almost solely of Kriol origin (O'Shannessy 2005; Meakins and O'Shannessy 2005). The origins of Light Warlpiri in relation to Gurindji Kriol are discussed in §4.

The degree of structural intertwining described in this section is due, in the broadest sense, to the rapid change in Gurindji society as a result of European colonisation. A number of important historical events have contributed to the level of language mixing observed today. The following section will discuss the history of the Gurindji people and their linguistic practices at key points in time.

3.   The socio-political and linguistic history of the Gurindji people

The post-contact history of the Gurindji people is perhaps one of the better-documented periods of Aboriginal history. A number of accounts of this time come from historians (Long 1996; Doolan 1977; Mulligan 1999; Riddett 1997; Hokari 2000, 2002), anthropologists (Berndt 1950; McConvell 1985, 1998, 2002; Lauridsen 1990; Berndt & Berndt 1987), activists (Dodson 2000; Hardy 1968) and the Gurindji themselves (Wavehill 2000; Rangiari 1997, 1998; Daguragu-Community-Council 2000; Kijngayarri 1986 [1974]; Donald 1998; Frith 1998). Much of the interest in the Gurindji people is derived from their nine year workers strike protesting against the poor conditions of employment on cattle stations (1966–75), and their subsequent pastoral lease (1975) and land claim (1986). This

claim was the first of its kind in Australia.[4] It provided impetus for the Aboriginal Land Rights (Northern Territory) Act (1976), and heralded a fresh wave of Aboriginal activism and non-indigenous interest in the plight of Aboriginal people. However not all of Gurindji history tells such as positive story. Accounts of earlier and darker periods of contact such as massacre stories and virtual slave labour also exist largely as a result of information which emerged in land claim hearings, Berndt and Berndt's (1987) *End of an era*, and oral history projects run through the regional Aboriginal language centre. The language environment and development of the mixed language has been influenced by this history. In this section, I give an account of the history of the Gurindji people, and the language practices of the Gurindji. I begin with Gurindji pre-contact history, setting the scene for the importance of later post-colonial events, particularly the significance of land to the language and identity of the Gurindji people (§3.1). In §3.2, I focus on the strike and land claim period. I argue that the socio-political climate of the time in the Victoria River District contributed to the retention of Gurindji through the mixed language where other areas were shifting to Kriol.

3.1   Pre-contact history and the language situation

Gurindji history begins with the formation of the landscape during a period called the Dreaming. Dreaming creatures traversed the land, shaping its features in a series of journeys referred to as Dreaming tracks or lines. These creatures took many forms. They were animals, humans or natural phenomena such as rain or lightning, and were responsible for the creation of hills, rocks, waterholes and clusters of trees. A number of Dreaming tracks criss-cross Gurindji country including *Ngawa* 'rain', *Martilyi* 'plains kangaroo', *Wampana* 'hare-wallaby' and *Kajirrikujarra warlakukujarra* 'two old women and two dogs'. The maintenance of these lines and their associated sites is essential for the physical and spiritual well-being of the Gurindji people. Some sites are imbued with procreative powers themselves such as Karungkarni, a hill near Kalkaringi which provides the Gurindji with their children. Other sites do not contribute directly to the health of the Gurindji, however the destruction of these places can cause mass sickness. Land and language

---

4.   In fact Aboriginal land rights first hit the headlines in 1963 when the Yirrkala people of Arnhem Land in the Northern Territory presented a bark petition to the Federal Government to stop their traditional lands being handed to mining interests. Workers' strikes similar to that of the Gurindji also occurred in the Pilbara region in Western Australia. Despite these protests, the Gurindji people's petitioning of the government had the earliest legislative impact.

are tightly interwoven. The Dreaming creatures sang the land into being, and the stories of the Dreaming are recounted in songs which also act to help maintain the land. These songs are passed down through family lines which are determined by the Gurindji social structure. Gurindji society is divided into two moieties: *Jalmawuny* – Heron moiety and *Warlawurruwuny* – Eaglehawk moiety. These two moieties provide the basis for land ownership and management. Any one area is owned by people from one moiety and cared for by people from the other. These moieties are further divided into four subsections (skins) which form the basis of kinship relations, dictate behaviour between family members and designate appropriate marriage partners. This general description of the Gurindji belief system and social structure applies to many Aboriginal groups across the north-central area of Australia.

Before European contact, the Gurindji were semi-nomadic, travelling mostly within their traditional land and subsisting on seasonally available animal and plant food. Contact with their neighbours was common. The Gurindji's closest neighbours were the Bilinarra and Ngarinyman to the north-west, the Jaru people to the west, the Karrangpurru to the northeast, the Mudburra people to the east and the Warlpiri to the south. Warfare between the Gurindji and nearby desert people occurred, however the neighbouring groups also shared many cultural practices and would come together once a year for ceremony time. For a fuller account of traditional Gurindji society see McConvell (1976, 1985) and for Ngarinyman and Bilinarra people see Rose (1991, 2000).

The Gurindji characterise the time before European invasion as an unchanging but cyclical period of social and natural order, and predictability. Indeed much of the cosmology of the Gurindji people is quite old, partly demonstrated in the archaic form of Gurindji still used in the Dreaming songs. Other practices have been introduced more recently. For example, the songs and ceremony of the Mungamunga women who come from the Roper River region of Arnhem Land dates back only to the early 20th century. These two women brought potent love songs and secret ceremonies to Gurindji women via Bilinarra women (Berndt 1950; Lauridsen 1990).[5] Up until recently women still received songs and ceremony from the Mungamunga through dreams. Though some of the songs discuss traditional law and sexual conduct, much of the content is about more contemporary issues.

---

5.   The songs of the Mungamunga are also found from the Roper River region to Borroloola, the Barkley Tablelands and to Tennant Creek (see map).

Little is known about the language practices of the Gurindji before European settlement, however McConvell (1988) suggests that the Gurindji and other Aboriginal groups have probably always been highly multilingual, with language mixing an unmarked form of communication. It is possible that this level of mixing is associated with the severe language shift to Kriol and English seen in many parts of Australia. However code-switching between traditional languages suggests that mixing was a common practice before European contact, and these contact languages were merely added to the repertoire.

3.2   The European invasion and its effect on the language ecology

First contact with *kartiya*[6] was a brutal period. The black soil plains of the Victoria River District were attractive to white settlers who were looking for good pastoral land to set up cattle stations. The first party of European explorers was led by the Gregory brothers, Francis and Henry. In late 1855 they arrived from the north. They followed the Victoria River and its tributaries and came upon the Victoria River District which they decided was suitable grazing land (Makin 1999: 43 onwards). Bilinarra, Ngarinyman and Karrangpurru country were the first to be stocked with cattle in 1883. In the process, the *kartiya* brought with them diseases that Aboriginal immune systems and traditional bush medicines could not cope with. These diseases actually slightly predated the arrival of the *kartiya* in the Victoria River District as a wave of illness which came from already-settled areas in the north. Rose (1991: 75 onwards) suggests that small pox almost completely devastated the Karrangpurru before the settlers virtually wiped out this group in a series of massacres. Now only a handful of people from one family claim some Karrangpurru heritage. The Bilinarra and Ngarinyman fared little better, but Rose suggests that, perhaps due to the rocky nature of their country, they were able to hide, and put up a greater resistance to the settlers. The killing sprees ended with a round-up of the survivors who were then put to work as stockmen and kitchen hands on the cattle stations, where they also lived in fringe camps. However by this stage the numbers of Aboriginal people in the Victoria River District had diminished significantly. For instance, when Berndt and Berndt (1948) first encountered the Bilinarra, they were working for the Australian Investment Agency surveying Aboriginal populations on cattle stations. They observed that the population was top-heavy with few children making it into adulthood. However with peace and

---

6.   *Kartiya* is the Gurindji word for "white people", perhaps derived from "guardian". It also may be a Gurindji word for 'ghost' which broadened to include "white people" due to their skin colour and aggressive behaviour towards Aboriginal people.

better health care the Bilinarra now live in greater numbers mostly at Nijburru (Pigeon Hole) and also at Yarralin (see map).

Though the Gurindji people lived further south, they did not escape the onslaught of the white pastoralists either. In an oral history account of this period, Ronnie Wavehill Jangala (2000) recalls similarly bloody periods where the settlers went on killing sprees. These massacres were a disproportionate response to the Gurindji stealing their cattle. However the battles were not always one sided and the Gurindji sought their own revenge for these massacres. Wavehill tells of one massacre that occurred at Warlakurla (Seale Gorge) which is west of Daguragu along the Seale River. This was a place where Mudburra, Gurindji and Ngarinyman people met up at on their travels. A group of pastoralists went to Warlakurla and shot dead everyone camped there, women and children included. Two men later stayed behind to burn their bodies, which was contrary to the traditional mortuary practices of the Gurindji who put dead bodies on high platforms to allow the deceased's spirit to pass on. The two pastoralists were ambushed by two Gurindji men, who killed and burned them in retaliation. This story is typical of the attacks and counter attacks which were common during this time. However the colonists soon wiped out the Gurindji, probably because they had better firearms, and the remaining people were brought under the control of pastoralists.

Most of the Gurindji lived and worked at Jinparrak (Old Wavehill Station), along with members of the Mudburra and Warlpiri tribes. This station was owned by the English Lord Vestey, who was the largest land holder in Australia at the time, owning a number of cattle stations across the north of Australia. The conditions of the Aboriginal people working and living on the stations were appalling, particularly given the profitability of these stations. Oral accounts from Gurindji people (Daguragu-Community-Council 2000; Donald 1998; Kijngayarri 1986 [1974]) and a report by Berndt and Berndt (1948) which was commissioned by the Vesteys to investigate the welfare of Aboriginal employees concur, describing the conditions as substandard. Two hundred and fifty people including 92 men lived in a small area. Gurindji people received no wages for their work. They worked as station hands and stockmen in exchange for goods such as tobacco, salted meat, flour, sugar and tea, and occasionally clothes and blankets. Gurindji women were often forced into sexual liaisons with *kartiya* stockmen. The Gurindji lived in humpies (shelters) which were constructed from discarded material from the station. Fresh water had to be drawn and carried some distance from a well. As a result the general health of people was low and the infant mortality rate very high. The Gurindji commonly likened these conditions to being treated like dogs, and despite Berndt and Berndt's candid report, little was changed.

Discontent ran high amongst the Aboriginal workers. Though many seemed resigned to their predicament, one Gurindji stockman, Sandy Moray Tipujurn,

started agitating amongst the Gurindji. He had spent time travelling to other cattle stations in Queensland and Western Australia and had seen better examples of race relations and employment conditions. Tipujurn had big ideas which went beyond an industrial dispute. He wanted the Gurindji to retrieve their land and run their own cattle station. The opportunity to begin this process arose when another Gurindji stockman, Vincent Lingiari, was thrown from his horse and sent to Darwin to be treated. There he met Aboriginal unionists, Dexter Daniels and Bobby Tudawali, who said that the NAWU (North Australian Workers Union) would support the Gurindji if they decided strike. When Lingiari returned to Wavehill station, he informed the station manager, Tom Fisher, of their intention to strike. Then on the 23 August 1966, Lingiari gathered his people and they walked 16 kilometres to Jurnani (Gordy Creek) and later to Daguragu which is a Ngamanpurru (Conkerberry) Dreaming place (8 kilometres from Kalkaringi and now an established Gurindji settlement). This event is now known as the Gurindji Walk-Off. Various attempts over the years to convince the Gurindji to return to the station failed. Eventually they were offered wages equal to those of white stockmen. However the Gurindji stood their ground. Although their protest had taken the form of a workers' strike, they had not stopped talking about reclaiming their traditional lands which had been taken over by the Vesteys. The NAWU, and in particular a union activist from Sydney called Frank Hardy, continued to support the Gurindji. He helped them petition the federal government, and raised money to fly Vincent Lingiari and another Gurindji stockman, Mick Rangiari, to Sydney on a couple of occasions to talk to union and university crowds about station conditions and land issues. In 1975, after nine years of persistent campaigning and a change to a more liberal federal government, the then Australian Prime Minister Gough Whitlam flew to Daguragu to grant the Gurindji a lease for 3236 km[2] of land around Daguragu. This event has been immortalised by a photo of Gough Whitlam pouring soil into Vincent Lingiari's hands who was, by this stage, a much older man, and blind. Twenty years later, in 1986, they were granted the security of inalienable freehold title under the Northern Territory Land Rights Act.[7] For a more detailed personal account of this sequence of events, see Frank Hardy's *The Unlucky Australians* (1968). Other oral accounts from Gurindji people and

---

7.   Despite the apparent security of this piece of legislation, in August 2007, the Australian government over-turned it, exploiting a loophole in Australian law which allows the Federal government to intervene in Territory politics. The two Australian *territories* – the Northern Territory and the Australian Capital Territory (where the Federal parliament is found) have a lower status than Australian *states*, and are still under the control of the Federal government, in this respect. The Australian government has taken possession of all Aboriginal land in the Northern Territory for a period of 5 years under the guise of child protection. Many critics see this as an attempt to regain control of hard-fought-for traditional lands.

interesting interpretations of this period can be found in two articles by Minoru Hokari (2000, 2002).

Little is known about the language situation at Wavehill station during the cattle station days, however reports from Berndt and Berndt (1987) paint a picture of multilingualism, with Gurindji and Mudburra as the dominant languages, and an Aboriginal variety of English emerging from contact with white station labour.

> Wavehill was a centre of gradual but continuous intermingling of what have sometimes been called tribes with differing language, territorial and cultural affiliations … for general purposes the *lingua franca* was either Gurindji or Mudbara (Mudburra) or usually a mixture of both. Few of the non-Walbiri (Warlpiri) people could either speak or understand more than a few words of the language spoken by the Walbri … On account of their contact with Europeans, by whom so many of them were employed, most of the station people found it necessary to learn a certain amount of English (Berndt & Berndt 1987: 59).

The language environment in the mid-1970s, the final stage of the strike, is better documented. At this time, McConvell (1985: 96) observed that the most pervasive discourse practice among Gurindji people was code-switching between different dialects of Gurindji and Kriol. In a recording of a conversation between Gurindji stockmen who were butchering a cow in a bush paddock near Kalkaringi, McConvell (1988: 97) found that approximately a third of the utterances were monolingual Gurindji, one third Kriol, and the remaining third involved intra-sentential code-switching. Resonances of the mixed language, Gurindji Kriol, can be found in the patterns of code-switching from this time. In a comparison of this 1970s code-switching data with modern mixed language data, McConvell and Meakins (2005) calculate that 73% of the code-switched utterances used a Kriol verbal structure, including tense and aspect morphemes. It appears that during this period the Kriol verbal structure was already becoming dominant. Despite the predominance of Kriol in the VP of the code-switching, Gurindji morphology, including case and derivational morphemes, was also present in the structure of the noun phrases in code-switched utterances. For example, a Kriol VP is found in conjunction with a Gurindji ergative marker in (3), a Gurindji paucal suffix in (4), and Gurindji dative markers in (5).

(3) *kaa-rni-mpal*   said orait   yutubala kat-im   *ngaji-rlang-**kulu**.
     east-up-across   side alright 2DU   cut-TRN   father-DYAD-**ERG**

   'You two, father & son, cut it across the east (side of the cow).'
   (1970s Gurindji-Kriol CS, McConvell corpus)

(4) wi   neba bin bring-im *kartak-walija*.
     1PL.S NEG   PST bring-TRN container-PAUC
     'We didn't bring any buckets.' (1970s Gurindji-Kriol CS, McConvell corpus)

(5) gib-it   langa im *murlu-wu Malingu-wu*.
     give-TRN   PREP   3SG this-DAT   NAME-DAT
     'Give it to this Malingu.' (1970s Gurindji-Kriol CS, McConvell corpus)

These code-switched utterances from the 1970s bear a strong resemblance to the mixed language spoken today. Indeed now Gurindji Kriol maintains both a Kriol VP structure and Gurindji nominal morphology, along with more general lexical mixing, as was shown in §2. [8]

The dominance of Kriol in the code-switching of the 1970s is probably a result of the socio-historical factors discussed above. At this time, Gurindji people were broadening their associations with other Aboriginal people and non-indigenous people through the cattle stations, with Kriol acting as the *lingua franca* for these groups. Already Kriol was becoming the main language of many Aboriginal people, and the fact that Kriol was the dominant language of the VP in this code-switching sample may be the result of Gurindji people following the socio-linguistic trends of the 1970s. Indeed the situation at Kalkaringi probably differed little from other places in Australia in the 1970s. Observations of mixing between traditional languages and English have been widespread. For example in the 1970s, code-switching between Dyirbal and English was a common discourse practice of Dyirbal people in north Queensland, and people in Maningrida (Arnhem Land) also mixed traditional languages with each other and with English in conversation. In the Torres Strait a discourse style called Ap-ne-ap (<"half-and-half") was characterised by frequent switches between Torres-Strait Pidgin and Kalaw Lagaw Ya, and finally code-switching between English and Guugu Yimidhirr was observed as the unmarked register of the people of Hopevale in Queensland (reported in McConvell 1988).

3.3   The Gurindji people today

The invasion and subsequent effects of European language and culture has had a lasting effect on the Gurindji. Nowadays most people live a more sedentary life in Kalkaringi and Daguragu, though travel between neighbouring communities and to the nearby town of Katherine is common. Knowledge of the Dreaming is still important to people's lives, and it continues to be passed down through the

---

8.   It is worth pointing to two ways in which Gurindji Kriol, the mixed language, differs from Gurindji-Kriol code-switching. Firstly there is a great deal of inter- and intra- speaker consistency, and secondly, though Gurindji Kriol draws its structural components from Gurindji and Kriol, these components function in different ways in the mixed language. These arguments are presented in more detail in Meakins (2007: 51–60).

generations albeit in a somewhat reduced form. Gurindji cosmology has also been augmented with Christian belief systems. Christian missionaries, who arrived in Kalkaringi in the 1960s, have been less harsh on Gurindji spirituality than missionaries elsewhere in Australia. Bible passages and hymns have been translated into Gurindji, and Gurindji beliefs are rarely directly challenged. In many ways, the Gurindji people have integrated elements of Christian symbolism into the Dreaming. For example the herd of wild donkeys which inhabit Daguragu are considered to be sacred, related to the donkey that Jesus rode on as he entered Jerusalem. The Gurindji say that these donkeys walk down from Marlukalarni, a nearby hill, where God places them. Older people get very upset when *kartiya* carry out annual culls to reduce their numbers. The communities of Kalkaringi and Daguragu are administered by a council office, and other facilities include an employment and welfare centre, health clinic, bakery, abattoir, store and school. There is little paid employment in the communities, and most people rely on welfare payments. The government-run Community Development and Employment Program (CDEP) program is a work-for-the-dole scheme which provides some part-time employment. Other people are employed as health workers or teaching assistants at the local clinic and school. Younger women look after their children and are primarily responsible for the well-being of older people. People's diets are based on the limited range of food found at the Kalkaringi store and are only supplemented in a minor way by bush foods. As a result there is a high incidence of diabetes and associated kidney failure among the Gurindji people. Though Kalkaringi has a government-run health clinic, traditional bush medicines are still used to treat many common ailments, and traditional doctors are called on for more mysterious illnesses. Tensions between *kartiya* and the Gurindji often run high with both sides frustrated with the lack of understanding of each others' ways of operating. Many Gurindji people are clearly weighed down by the incompatibility of systems, finding the constant grind of negotiating the *kartiya* world very difficult. Other people exhibit an extraordinary resilience despite the imposition of *kartiya* practices and clear clashes between these and their own Gurindji systems.

The post-contact language situation developed in a number of ways. Although Gurindji is the traditional language of Kalkaringi and the surrounding area, Gurindji Kriol is now the main language spoken in Kalkaringi, and the first language of all Gurindji people under 35 years of age. Moreover Gurindji Kriol has gained momentum among younger people of the Victoria River District and has spread north into Bilinarra and Ngarinyman country. Despite the mixed nature of Gurindji Kriol, it is largely viewed in terms of the maintenance of Gurindji. For example, the mixed language is not distinguished from Gurindji with regard to naming practices, with the term "Gurindji Kriol" created during a language workshop about how to refer to language varieties at Kalkaringi (Charola 2002). The term "Gurindji Kriol" has

little currency in the community, and it is rarely used to denote the mixed language.[9] In fact this mixed language is usually called "Gurindji". If a distinction between Gurindji and Gurindji Kriol is required, Gurindji is usually referred to as "hard Gurindji", "rough Gurindji" or "proper Gurindji", and Gurindji Kriol as "Gurindji". The term "Gurindji", it seems, is a relative term used to signify the main language used by the community rather than a particular language form. The use of "Gurindji" to refer to Gurindji Kriol consciously emphasises the Gurindji, rather than Kriol, content, and marks a desire to continue the tradition of the Gurindji people. Despite the dominant use of Gurindji Kriol at Kalkaringi, the picture is far from monolingual. Gurindji Kriol is situated within a complex picture of multilingualism, contact and code-switching (Meakins 2008). Kriol, Aboriginal English and Standard Australian English are also found, along with a neighbouring traditional Australian language, Warlpiri, though their use is more marginal. In this respect, Gurindji Kriol continues to be spoken alongside Gurindji and Kriol, and is a "symbiotic" mixed language (cf. Smith 2000).

Additionally, code-switching continues to be an everyday practice at Kalkaringi, and it is common to find code-switching between Gurindji and Kriol, and Gurindji Kriol and its source languages. Code-switching occurs between speakers where one person speaks one language and the other person replies in another language. It can also occur within one speaker's sentence. Code-switching by one speaker occurs as *insertional* and *alternational* code-switching (cf. Muysken 2000). For example, it is quite common for speakers to maintain their own speech styles in the course of a conversation, as in the following exchange in (6) where a 43 year old woman (FO) is asking her 21 year old adopted daughter (CA) where another group have gone fishing. In this case, one speaker is associated with one language, and code-switching occurs between speakers. FO begins in Gurindji, and CA replies in Gurindji Kriol, and FO again in Gurindji.

(6)   FO: *wanyjika-warla   nyila   ngu-lu   ya-ni*?
          where-FOC   that   CAT-3PL.S   go-PST
          'Where did they go?'

       CA: dei   neba   tok   *ngayiny*
          3PL.S   never   talk   1SG-DAT
          'They didn't tell me.'

       Fo: wal   *yangki   pa-rra   nganayirla*?
          well   ask   hit-IMP   what's.it.name
          'Well ask what's-his-name.'

---

9. Note that in the 2006 Australian Bureau of Statistics Census, four women stated they spoke Gurindji Kriol at home, rather than Gurindji, but these women were language workers.

Insertional code-switching, where fragments from one language, usually lexical items, are embedded within another's grammar, is also common. Speakers thirty five years old and over often use Gurindji as their base language, and insert Kriol items. For example, in (7) a 42 year old speaker is rebuffing a child who is straying close to a fire. She uses the Kriol/English noun *faya* (fire) within a Gurindji sentence. The Gurindji inflecting verb and bound pronouns are used here where they are never found in Gurindji Kriol.

(7)   *ngaja-ngku* **faya**-*ngku jiya-rnana.*
      LEST-2SG.O **fire**-ERG   burn-PRS.IM
      'The **fire** might burn you.'

Alternational code-switching is also commonly used in Kalkaringi to mix Gurindji, Kriol and Gurindji Kriol. In alternational code-switching a language is not structurally nested within another language to the same extent as insertional code-switching. In most cases, an utterance begins with a clause in one language and finishes in another. For example in (8) a 46 year old speaker is telling her sister that their grandchildren are digging for frogs (to use as fish bait) in the wrong place. She begins in Gurindji Kriol and switches to Gurindji in the main clause. In (9), a group of 20 year old women are walking around Jinparrak, the old cattle station where their grandparents were employed (see §3.2). One of the women suggests to another woman where they should go, beginning with a Gurindji Kriol clause, then adding an English prepositional phrase, and finally switching back to Gurindji Kriol.

(8)   i=m   tumaj   *partawarn* dat *janyja/*
      3SG.S=NF because hard   DET ground/

      ngu-lu-rla   *kurrij-karra* *pung-ana*   *yipurrk.*
      CAT-3PL.S-3DAT scratch-CONT poke-PRS.IM in.vain

      'Because the ground is too hard, they're digging for frogs in vain.'

(9)   wi   gon   *kanyjurra nyawa-ma* riba-*ngka* *ngawa-ngka*
      1PL.S   go   down   this-DIS   river-LOC   water-LOC

      kol-wan-*ta/*   down the creek/ *kanyjurra-k.*
      cold-NOM-LOC/ down the creek/ down-ALL

      'We're going down the creek through the cold water, down the creek, downwards.'

4.   Land, language and identity – motivations for the formation of a mixed language

It is significant that a mixed language arose in Kalkaringi, because elsewhere in north Australia, Kriol has steadily replaced the traditional languages. For example,

to the north of Kalkaringi in Timber Creek (see map), the traditional languages of the Aboriginal people in the town and its satellite communities are Jaminjung, Ngaliwurru, Ngarinyman and Nungali. However the main language now spoken is Kriol. Jaminjung, Ngaliwurru, Ngarinyman are only used by older speakers, and are usually mixed with Kriol through borrowing and code-switching. Younger speakers are not proficient speakers of these languages, however they do incorporate single words, usually nouns into the Kriol. In this way the traditional languages function as markers of Jaminjung, Ngaliwurru, or Ngarinyman identity. Similar situations can be found all across the north of Australia, except in many parts of Arnhem Land and the Daly River region in the north eastern part of the Northern Territory where traditional languages remain strong. Despite the dominance of Kriol, Gurindji has remained remarkably resilient. The question is then: why did a mixed language form in Kalkaringi where the rest of the north shifted to a variety of Kriol, with some exceptions including Lajamanu and the Daly River area which I discuss below? In this section I suggest that the formation of Gurindji Kriol is very much a product of the linguistic environment and socio-political history of Kalkaringi. The link between land, language and identity is an important ingredient in the formation of Gurindji Kriol. As was shown in §3.1, a strong connection between land and language exists for the Gurindji. Thus the fight for the right to custody of their traditional lands occurred in parallel to changing linguistic practices. In particular, the code-switching practices and political events of the 1970s provided the seeds for the emergence of this mixed language.

During the period between 1966–75 following the Wavehill Walk-Off, Gurindji people gained notoriety for their persistence in fighting the Vesteys and the Australian government for the return of their traditional lands. Other Aboriginal people and sympathetic *kartiya* regarded them as a strong and courageous group for resisting what many saw as the inevitable dominance of the *kartiya* over their land, language and law. At this time, the Gurindji people set about establishing their own cattle station at Daguragu and steadfastly refused government assistance from Kalkaringi which was a welfare settlement at the time. They only accepted help from people who supported their cause, such as Union and Communist party members. During this period, code-mixing between traditional languages and Kriol was the unmarked discourse practice, and in the case of Kalkaringi, Kriol was the language which provided the main grammatical frame for code-switching, as was shown in §3.2. A similar situation was most likely found in communities to the north. However mixing practices in these more northern areas represented a transitional stage in the shift to Kriol, while at Kalkaringi the code-switching gradually stabilised into the mixed language spoken today. Thus Gurindji Kriol represents the maintenance of Gurindji, in part, as is evidenced in the naming practices discussed in §3.3. Where Kriol gradually replaced almost all of the lexicon and

structure of the traditional languages in the north, significant amounts of Gurindji vocabulary and grammatical features remain in the mixed language, making it unintelligible to Kriol speakers. For example, when Gurindji people travel to predominantly Kriol-speaking communities, they use Gurindji Kriol in the presence of Kriol-only speakers to conduct private exchanges among themselves.

The increasing use of Kriol and the homogenisation of the language of many Aboriginal groups was one of the results of drawing together different groups of Aboriginal people on cattle stations, welfare settlements and missions. This has resulted in the creation of a pan-north Australian Aboriginal identity that overarches individual group allegiances, and becomes particularly prominent in politics between Aboriginal and non-Aboriginal people. However, in the case of Gurindji Kriol, a separate Gurindji identity is both recognised and enacted through the continuing use of Gurindji in the mixed language. The Gurindji separate themselves from other Aboriginal people by staking claim to strength and respect that is associated with their name. This identity remains salient, for example, in the practice of referring to Gurindji Kriol as "Gurindji" despite the significant amount of Kriol content. Speakers view this language as the maintenance of Gurindji, and the language of their people.

The fact that "Gurindji-ness" found its expression in both the mixed language and the land claim speaks to the interwoven nature of land, language and identity. Note, though, that it is unlikely to be the case that the fight for the return of traditional lands led directly to the formation of the mixed language.[10] Indeed Gurindji people themselves do not make this connection. Instead it is more likely that the struggle over land ownership and maintenance of Gurindji in the form of the mixed language both stemmed from a resistance to the cultural incursion which accompanied the arrival of the *kartiya*. A separate Gurindji identity is marked out through both of these acts. Thus the land claim and mixed language can both be viewed in terms of the persistence of the Gurindji identity under pressure from some of the follow-on effects of colonisation such as the subsumption of Gurindji land for cattle stations and the spread of Kriol.

However, although socio-political factors clearly played an important role in the formation/stabilisation of Gurindji Kriol, it cannot be claimed that socio-political factors are the principal and necessary conditions under which these kinds of mixed languages always emerge. For instance, a typologically similar mixed language, Light Warlpiri, is spoken in the nearby community of Lajamanu (O'Shannessy 2006), a community with a very different recent history.

---

10.   My thanks to Christine Dureau for demonstrating the difficulties with such a cause-effect analysis in creating an invalid syllogism.

The Warlpiri people from Lajamanu are not associated with any landmark political event which may have strengthened their sense of identity. In fact the Warlpiri at Lajamanu occupy Gurindji land, which is the cause of much friction between Gurindji and Warlpiri people and necessarily means that continued use of Warlpiri, even in a mixed code, cannot be seen simplistically as the continued maintenance of an indigenous right to the land.

The presence of Warlpiri people on Gurindji land is the result of a *kartiya* decision. A number of Warlpiri families were brought from Yuendumu from 1949 onwards to prevent overcrowding, and the community in Lajamanu grew from there (O'Shannessy pers. comm.). Meakins and O'Shannessy (forthcoming) suggest that other factors, far from grounding their identity in their land, may have contributed to the formation of Light Warlpiri. First Warlpiri is spoken in other communities such as Yuendumu, Nyirrpi and Willowra to the south of Lajamanu. Lajamanu people constantly travel south to visit family and take part in ceremonies. Knowledge of Warlpiri is therefore essential for maintaining familial and ceremonial links with these communities. This situation is therefore quite different to Kalkaringi, which is the only Gurindji-identifying community, though Gurindji and the traditional languages spoken in Pigeon Hole and Yarralin (see map), Bilinarra and Ngarinyman respectively, are mutually intelligible.

In addition, Lajamanu also has a bilingual school (Warlpiri and English) which has operated since the 1980s. Children are taught in Warlpiri in the earliest years before transitioning into English, and Warlpiri continues to be a medium of instruction to a varying extent for the rest of their time at school. This bilingual program has probably also contributed to the continuing use of Warlpiri in the community, both as Warlpiri and within Light Warlpiri. In contrast, Kalkaringi has an English-medium school, as was noted in §3.3. Small Gurindji language programs have existed periodically, however the bulk of Gurindji children's schooling is delivered in English.

The differences between the socio-linguistic contexts of Kalkaringi and Lajamanu suggest that the period of intense social upheaval may not be the sole factor which provided the necessary social conditions for the emergence of a mixed language like Gurindji Kriol. Nevertheless, there is one factor common to these two communities and which contrasts with many other places in northern Australia: only one traditional language is associated with these communities. For example, in Timber Creek and its surrounding communities, people from at least four language groups were brought together. The communities were artificial social constructions, with different cultural groups living in much closer proximity than was traditionally found. Hence, this greater mixture of cultures and languages differs from both the situations in Kalkaringi and Lajamanu where the communities grew from family groups into geographically denser versions of the traditional social

structure. In these communities only one main traditional language was spoken. McConvell (2008) suggests that the number of languages spoken in a community provides an essential clue as to why Kriol gained currency in some communities and not others. He proposes that a *lingua franca* was needed amongst community members, and Kriol suited this purpose. However in places where a common language was already spoken, Kriol was not required as the sole medium of communication.

While McConvell's account squares well with common explanations for the spread of creole languages in Australia and Melanesia, I would suggest that the number of languages represented in a community only provides part of the explanation for the almost complete shift to Kriol in many parts of northern Australia. In other areas where people from a number of language groups were brought together in equally disruptive circumstances, Kriol did not become the dominant language. For example, large numbers of Aboriginal people living in the Daly River area to the north of Timber Creek were shifted to the Catholic mission community of Wadeye. Although many languages were spoken by these people, Murrinh-patha, which is the language of the surrounding country and the church in Wadeye, became the *lingua franca* of this community. The loss of other languages of this area is ongoing, as Murrinh-patha becomes the dominant language (Nordlinger, pers. comm.). In other words, the need for a *lingua franca* need not lead to the selection of Kriol.

Another problem with proposing that the shift to Kriol was the result of the need for a *lingua franca* is that it does not explain why Kriol seeped into communities, such as Kalkaringi and Lajamanu, which already had common languages, Gurindji and Warlpiri, respectively. Although versions of Aboriginal English and cattle station pidgin were being used with the *kartiya* on the cattle stations, and although Kriol had some function as a *lingua franca* with imported Aboriginal labour (see §3.2), the appeal of Kriol within the Gurindji group itself is not clear. It may have been the case that Kriol gained some currency because it helped the Gurindji and Warlpiri communicate with other groups and communities who were losing their traditional languages. Identity reasons may have also played a role. For example, the use of Kriol may have helped the Gurindji and Warlpiri link with a cross-cultural Aboriginal identity which only became salient with the arrival of *kartiya*.

In general it is likely that the combination of all of these social factors provided optimal conditions for the emergence of Gurindji Kriol. This mixed language is probably the result of simultaneous pressure from Kriol and the desire to maintain Gurindji for reasons of identity marking. Preserving Gurindji elements in Gurindji Kriol was made somewhat easier by the fact that Gurindji was the dominant language associated with Kalkaringi. Moreover the already existing

language practice of code-switching provided the basis for the expression of a mixed identity, which could then be gradually systematised.

5.   Conclusion: The future of Gurindji Kriol

In many ways, Gurindji Kriol provides a unique opportunity to study the birth and life of a mixed language. The socio-historical period prior to the genesis of Gurindji Kriol is well-documented due to the political struggle of the Gurindji people in reclaiming their traditional lands. Additionally, language data, from the time immediately preceding the emergence of this mixed language, is available as a result of work done by Patrick McConvell in the 1970s. With all of these resources, a picture of the emergence of Gurindji Kriol is possible.

It is likely that continuing contact with its source languages will precipitate further change in Gurindji Kriol. Indeed aspects of Kriol grammar such as prepositions are beginning to creep into the largely Gurindji-dominated NP structure in the speech of younger Gurindji Kriol speakers. For example, though the Gurindji locative case marker is the main form used for marking topological relations for all speakers (10), children are more likely to use the Kriol-derived *langa* preposition (11) or a double-marked construction using both the preposition and case marker (12), than their 20 year old counter-parts (Meakins 2007).

(10)   najan   *warlaku makin* tebul-***ta***   *kanyjurra*.
       another   dog   sleep   table-LOC   down
       'Another dog is sleeping **under** the table.'

(11)   an dat   *warlaku* i=m   top *nyantu-warinyj* **la**   dat fens.
       and DET dog   3SG.S=NF be 3SG-alone   PREP DET fence
       'And the dog is sitting by itself next to the fence.'

(12)   dat   *warlaku makin* **langa** dat   tebul-***ta***   *kanyjurra*.
       DET dog   sleep PREP DET table-LOC down
       'The dog is sleeping **under** the table.'

These age-related differences may represent a youth style of Gurindji Kriol; however I suggest a diachronic interpretation which points to further development and change in Gurindji Kriol. If these changes continue along the trajectory of Kriol structural features replacing Gurindji equivalents, the predicted endpoint will be an entirely Kriol structure with some lexical contributions from Gurindji. Over time this hypothesis will become testable. With successive generations, it remains to be seen whether the trend towards the increasing use of Kriol structural elements continues, or whether Gurindji features are maintained or even revitalised. Given that Gurindji Kriol remains in contact with its source languages,

all of these options are possible. The direction Gurindji Kriol takes, then, will be largely dependent on what the new generation of Gurindji people wishes to mark with this mixed language.

References

Bakker, P. 1997. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis* [Oxford studies in Anthropological Linguistics 10]. Oxford: OUP.

Berndt, C. 1950. *Women's Changing Ceremonies in Northern Australia*. France: Librairie scientifique Herrman.

Berndt, C. & Berndt, R. 1948. Pastoral stations in the Northern Territory and native welfare. *Aborigines Protector* 2(4): 13–16.

Berndt, R. & Berndt, C. 1987. *End of an Era: Aboriginal labour in the Northern Territory*. Canberra: AIAS.

Berndt, R. & Berndt, C. 1948. A Northern Territory problem: Aboriginal labour in a pastoral area. Sydney.

Boretzky, N. & Igla, B. 1994. Romani mixed dialects. In *Mixed Languages: 15 case studies in language intertwining*, P. Bakker & M. Mous (Eds). Amsterdam: IFOTT.

Charola, E. 2002. The verb phrase structure of Gurindji Kriol. Honours dissertation, Melbourne University.

Crowley, T. 1990. *Beach-la-Mar to Bislama: The emergence of a national language in Vanuatu*. Oxford: OUP.

Daguragu-Community-Council. 2000. Mumkurla-nginyi-ma parrngalinyparla (From the darkness into the light, Gurindji freedom banners: A celebration of the determination and vision of the people of Daguragu and Kalkaringi. Kalkaringi.

Dodson, P. 2000. Lingiari: Until the chains are broken. In *Reconciliation*, M. Grattan (Ed.). Melbourne: Black.

Donald, V. 1998. We're not coming back. In *Take Power like this old man here: An anthology celebrating 20 yrs of land rights in Central Australia*, A. Wright (Ed.), Alice Springs: IAD Press.

Doolan, J.K. 1977. Walk-off (and later return) of various Aboriginal groups from cattle stations: Victoria River District, Northern Territory. In *Aborigines and Change: Australia in the 70s*, R.M. Berndt (Ed.). Canberra: AIATSIS.

Frith, N. 1998. We went on strike. In *Take Power like this old man here: An anthology celebrating 20 yrs of land rights in Central Australia*, A. Wright (Ed.). Alice Springs: IAD Press.

Golovko, E.V. 1994. Mednyj Aleut or Copper Island Aleut: An Aleut-Russian mixed language. In *Mixed Languages: 15 case studies in language intertwining*, P. Bakker & M. Mous (Eds). Amsterdam: IFOTT.

Hardy, F. 1968. *The Unlucky Australians*. Melbourne: Nelson.

Hokari, M. 2000. From Wattie Creek to Wattie Creek: An oral historical approach to the Gurindji walk-off. *Aboriginal History* 24: 98–116.

Hokari, M. 2002. Reading oral histories from the pastoral frontier: A critical revision. *Journal of Australian Studies: New Talents, Jumping the Queue* 72: 21–28.

Hudson, J. 1983. *Grammatical and semantic aspects of Fitzroy Valley Kriol* [Work Paper of SIL-AAB A/8]. Darwin: SIL.

Kijngayarri, L.J. 1986 (1974). The Wave Hill strike. In *This is What Happened: Historical narratives by Aborigines*, L. Hercus & P. Sutton (Eds). Canberra: AIAS Press.

Lauridsen, J. 1990. Women's jarata of north central Australia. Unpublished ms.

Long, J. 1996. Frank Hardy and the 1966 Wave Hill walk-off. *Northern Perspective* 19(2): 1–9.

Makin, J. 1999. *The Big Run: The story of Victoria River Downs station*. South Australia: J.B. Books.

Mayer, M. 1994 [1969]. *Frog, Where Are You?* China: Puffin.

McConvell, P. 1976. *The Aborigines of the northern Tanami area*. Darwin.

McConvell, P. 1985. Domains and codeswitching among bilingual Aborigines. In *Australia, Meeting Place of Languages*, M. Clyne (Ed.). Canberra: Pacific Linguistics.

McConvell, P. 1988. Mix-im-up: Aboriginal codeswitching old and new. In *Codeswitching: Anthropological and sociolinguistic perspectives*, M. Heller (Ed.). Berlin: Mouton de Gruyter.

McConvell, P. 1996. Gurindji grammar. Canberra: AIATSIS.

McConvell, P. 1998. 'Born is nothing': Roots, family trees and other attachments to land in the Victoria River District and the Kimberleys. *Aboriginal History* 22: 180–202.

McConvell, P. 2002. Changing places: European and Aboriginal styles. In *The Land is a Map: Placenames of indigenous origin in Australia*, L. Hercus, F. Hodges & J. Simpson (Eds). Canberra: Pandanus Books.

McConvell, P. 2008. Language mixing and language shift in indigenous Australia. In *Children's Language and Multilingualism: Indigenous language use at home and school*, J. Simpson & G. Wigglesworth (Eds). London: Continuum.

McConvell, P. & Meakins, F. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics* 25(1): 9–30.

Meakins, F. 2007. Case Marking in Contact: The development and function of case morphology in Gurindji Kriol, an Australian mixed language. Ph.D. dissertation, University of Melbourne.

Meakins, F. 2008. Unravelling languages: Multilingualism and language contact in Kalkaringi. In *Children's Language and Multilingualism: Indigenous language use at home and school*, J. Simpson & G. Wigglesworth (Eds). London: Continuum.

Meakins, F. & O'Shannessy, C. 2005. Possessing variation: Age and inalienability related variables in the possessive constructions of two Australian mixed languages. *Monash University Linguistics Papers* 4(2): 43–63.

Meakins, F. & O'Shannessy, C. Forthcoming. Ordering arguments about: Word order and discourse motivations in the development and use of the ergative marker in two Australian mixed languages. In *Lingua (Special Issue on Optional Ergativity)*, W. McGregor & J.-C.Verstraete (Eds).

Meyerhoff, M. 2000. *Constraints on Null Subjects in Bislama (Vanuatu): Social and linguistic factors*. Canberra: Pacific Linguistics.

Mous, M. 2003. *The Making of a Mixed Language: The case of Ma'á/Mbugu*. Amsterdam: John Benjamins.

Mulligan, M. 1999. Reading storied landscapes. *Arena Magazine* 39: 39–42.

Munro, J. 2000. Kriol on the move: A case of language spread and shift in Northern Australia. In *Processes of Language Contact: Studies from Australia and the South Pacific*, J. Siegel (Ed.). Saint-Laurent: Fides.

Munro, J. 2005. Substrate Language Influence in Kriol: The application of transfer constraints to language contact in northern Australia. Ph.D. dissertation, University of New England.

Muysken, P. 1997. Media Lengua. In *Contact Languages: A wider perspective*, S.G. Thomason (Ed.). Amsterdam: John Benjamins.

Muysken, P. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: CUP.

O'Shannessy, C. 2005. Light Warlpiri: A new language. *Australian Journal of Linguistics* 25(1): 31–57.

O'Shannessy, C. 2006. Language Contact and Children's Bilingual Language Acquisition: Learning a mixed language and Warlpiri in northern Australia. Ph.D. dissertation, University of Sydney & MPI, Nijmegen.

Rangiari, M. 1997. Talking history. In *Our Land is Our Life*, G. Yunupingu (Ed.), Brisbane: UQ Press.

Rangiari, M. 1998. They been get rich by the Gurindji people. In *Take Power like this old man here: An anthology celebrating 20 yrs of land rights in Central Australia*, A. Wright (Ed.). Alice Springs: IAD Press.

Riddett, L.A. 1997. The strike that became a land rights movement: A southern 'do-gooder' reflects in Wattie Creek 1966–74. *Labour History* 72: 50–64.

Rose, D. 1991. *Hidden Histories: Black stories from Victoria River Downs, Humbert River and Wavehill stations*. Canberra: Aboriginal Studies Press.

Rose, D. 2000. *Dingo makes us human: Life and land in an Australian Aboriginal culture*. Cambridge: Cambridge University Press.

Sandefur, J. 1979. *An Australian Creole in the Northern Territory: A description of Ngukurr-Bamyili dialects* (Part 1) [Work Papers of SIL-AAB series B, Vol. 3]. Darwin: SIL.

Sankoff, G. 1980. *The Social Life of Language*. Pennsylvania PA: University of Pennsylvania Press.

Smith, N. 2000. Symbiotic mixed languages: A question of terminology. *Bilingualism: Language and Cognition* 3(2): 122–123.

Thomason, S. 1997. Ma'a (Mbugu). In *Contact Languages: A wider perspective*, S.G. Thomason (Ed.). Amsterdam: John Benjamins.

Thomason, S. 1997. Mednyj Aleut. In *Contact Languages: A wider perspective*, S.G. Thomason (Ed.). Amsterdam: John Benjamins.

Thomason, S. 2001. *Language Contact: An introduction*. Edinburgh & Washington DC: Edinburgh University Press & Georgetown University Press.

Wavehill, R. 2000. *Nyawayinangkulu larrpa kujilirli yuwanani ngumpit-ma kartiyarlu*. Translated by E. Charola. Katherine: Diwurruwurru-jaru Aboriginal Corporation.

Wurm, S. & Mühlhäusler, P. 1985. *Handbook of Tok Pisin (New Guinea Pidgin)*. Canberra: Pacific Linguistics.

List of abbreviations

ABL	ablative		OBL	oblique
ALL	allative		PAUC	paucal
CAT	catalyst (auxiliary)		PL	plural
CONT	continuative		PREP	preposition
DAT	dative		PRS	present
DET	determiner		PROP	proprietive (having)
DIS	discourse		PST	past

DOUBT	doubt	REDUP	reduplication
DU	dual	REFLX	reflexive
DYAD	dyad (of kinship)	S	subject
ERG	ergative	SG	singular
FOC	focus	TRN	transitive
IM	imperfect	1	first person
LEST	for fear that it might	2	second person
LOC	locative	3	third person
NEG	negative	-	morpheme break
NF	non-future	=	clitic break
O	object		

"I've been speaking Tsotsitaal all my life without knowing it"

Towards a unified account of Tsotsitaals in South Africa*

Rajend Mesthrie
University of Cape Town

> "Long after the year 2000 AD professors and pimps will still enjoy and be fascinated by slanguages and ganguages"
>
> – Buntu Mfenyana (1981: 302)

This chapter focuses on varieties that flourish in South African townships with names like *Tsotsitaal, Flaaitaal, Iscamtho, Gamtaal* etc. Some speakers and scholars argue that these are new languages arising out of urbanisation and underworld culture in multilingual settings. This chapter examines the extent of overlap in names and defining characteristics of these varieties. It concludes that we may possibly be dealing with just one phenomenon: essentially a set of lexical items associated with gang and prison culture at one end and that of youth culture at the other, which is attached to the syntax of previously existing languages. Evidence for this claim comes from unearthing a similar variety that uses English as its base language.

Keywords: *Tsotsitaal; Flaaitaal; Iscamtho*; new languages; code-switching; code-mixing; relexification; slang; secret language; anti-language; South African English; Zulu

1. Introduction

This paper examines a loose set of varieties that flourish in South Africa's townships and which go by various names, the chief of which are *Tsotsitaal, Flaaitaal*

---

and *Iscamtho*. Other names abound for these entities and their subvarieties, e.g., *mense taal, withi, lingo* – Makhudu (2002: 399) lists 16 such terms from the northern provinces. These varieties have been subject to rich sociological and linguistic analyses ranging from the historical work of Glaser (2000) and Ntshangase (1993), the sociological accounts of Stone (1991) and Makhudu (2002), the contact perspectives of Calteaux (1996) and Slabbert & Myers-Scotton (1997), overviews by Childs (1996) and Kiessling & Mous (2004) and the dictionary by Molamu (2003). However, the status of these varieties amongst both speakers and linguists is varied. Amongst linguists they have been described in various terms ranging from slang (Slabbert 1994; Mfusi 1992; Schuring 1983), new languages (Ntshangase 2002; Makhudu 2002), antilanguages (Stone 2002; Makhudu 2002), as manifestations of code-switching, some of which have fossilised (Slabbert & Myers-Scotton 1997: 325) and as possible outgrowths of an original pidgin-like variety (Makhudu 2002). Functionally, some of their subvarieties have been hailed as secret languages (e.g., the prison variety Shalombombo – Ntshangase 1993, Schuring 1983); others as new lingua francas of the townships (Janson 1984). Some scholars see them as the new informal varieties of established languages – e.g., Rudwick (2005) argues that Tsotsitaal shares a diglossic relation with standard Zulu in Durban's townships. Clearly the entities being described are elusive and flexible, yet visible enough to elicit such varied and contradictory interpretations. This paper attempts to clarify some of the nomenclature and the nature of the varieties involved. It does so by reflecting on a variety that itself has no name and no previous claim to be any of the above, but which in fact belongs to the same group of epiphenomena.

To render things less mysterious, let me initially propose *tsotsitaal* (in lower case) as a generic term for ALL the varieties concerned. (At the beginning of a sentence I use the typographical convention ⟨[T]⟩ to denote what is really lower case generic *tsotsitaal*.) This proposal implies that all the varieties can indeed be related to each other under one umbrella grouping. Thus Flaaitaal and Iscamtho are a type of tsotsitaal as are all the other "taals" (or languages/ varieties) in the above nomenclature. Unfortunately, some of the specific varieties within the umbrella may also be termed *Tsotsitaal* (some with an Afrikaans base, others with an African language base). For convenience this variety will be referred to with an upper case *T* in this paper or by its equally well-known name, *Flaaitaal* (following Makhudu 2002). Most analysts would not accept this formulation, and it is therefore necessary to begin by motivating a thesis of similarity, rather than dissimilarity. In section 2 I explain some of the nomenclature and their etymologies. Section 3 summarises the characterisations of previous scholars, which are often analytically and empirically rich. However, I suggest some reasons for doubting some accounts of the structure and boundedness of

the varieties. I then propose a unified account of tsotsitaal in section 4, testing it in section 5 against a nameless variety spoken extensively by young Indian males in KwaZulu-Natal province. I conclude in section 6 that tsotsitaal is not a language after all, but a lexical code ("slang", or perhaps something more) that has the permeable, areal quality of penetrating just about any prior-existing variety in certain gender-specific sub-cultures, domains and semantic fields. It has analogues elsewhere, notably in Sheng, an East African youth language based on Swahili, English and other languages (see Kiessling & Mous 2004).

2.   Nomenclature and exemplification

The main terms to be explained and exemplified are *Tsotsitaal*, *Flaaitaal* and *Iscamtho*. The root *taal* which is the most common descriptor for the varieties is Afrikaans for 'language', pointing to the importance of this language in the genesis of tsotsitaal. *Tsotsi* is the word for 'gangster, criminal' that may have originated towards the end of the 19th C in the townships that were established around the mines: Sophiatown to the west and Alexandra to the north of Johannesburg. The origins of the word are not entirely clear; but the most commonly cited etymology is from *Zoot Suit*, referring to the narrow-bottomed trousers worn by gangsters influenced by North American culture via the movies of the 1950s (Glaser 2000). *Flaai* most probably comes from English slang *fly* 'artful, informed' (Stone 2002), suggesting the city-slick associations that the variety has always carried. *Iscamtho* as a name is harder to pin down: *isi-* is a prefix from Zulu denoting a noun class (7) that includes language names, so that *isi-* is really an analogue of Afrikaans *taal*. *Camtho* now popularly denotes a young, urban way of speaking (parallel to *fly*). Ntshangase (2002: 409) gives its etymology as possibly from *ukuqamunda* 'to talk volubly'. Of the less widely-used terms: *mense taal* derives from Afrikaans *mense* 'people'; *withi* from Zulu and Xhosa *ukuthi* 'to speak', while *lingo* is transparent in international English. These varieties are today associated mainly with young, Black, urban males. Yet as Makhudu (2002) points out, the connection with Afrikaans suggests origins amongst Coloured people.[1] Coloured males on the Witwaters and are among the most proficient users of Flaaitaal. In Cape Town similar varieties are spoken by Coloured males with gangster or street culture associations, going by names like

---

1.   Traditional South African terms distinguish "Black" from "Coloured", the former are prototypically indigenous people speaking a Bantu language, the latter are prototypically of multiple ancestries speaking English or Afrikaans or in small numbers a Khoe-San language or other African language.

Gamtaal, *Skollietaal*, or *die Sleng* (Stone 1991). Stone links *Gam* with 'Ham',[2] denoting the outcast status of the Biblical children of Ham, while *skollie* is the Cape Afrikaans equivalent to *tsotsi*. Stone was the first to associate the variety with Halliday's notion of antilanguage: a variety symbolically and literally in opposition to an established language in the same way that a sub-culture (or antisociety) stood in relation to mainstream society. This characterisation fits all the other varieties very well: Stone and Makhudu give ample examples of relexicalisations (new terms for old) and overlexicalisations (proliferation of synonyms in key areas) in the respective varieties they studied. They also give examples of the vividness of metaphors, metonymy and the like. Most accounts discuss Tsotsitaal and its analogues in terms of a continuum. At one end is the most secret lexis associated with prison and criminals (and sometimes carrying more specific names), at the other a plethora of established and newly coined terms associated with male, urban slang familiar in most societies.

3.   Characterisations of tsotsitaal

In initial work on Flaaitaal, Makhudu (1980) tended to draw on pidgin and creole linguistics. The idea was that if speakers of diverse African languages (both Bantu and Khoe-San) from all over Southern Africa were being propelled into a highly multilingual industrial sphere, in which English and Afrikaans were the superstrate languages, then models from creolistics were to be considered. Makhudu (2002: 398–9) writes: "There is, […] the possibility that Flaaitaal originated as a type of proto-pidgin, fashioned by expediency to lay the foundations for new communication systems." However, he concedes (2002: 399) that the variety is a "spontaneous in-group result of social and linguistic interaction amongst equals". One could rephrase this to say that Flaaitaal developed as a mode of "horizontal communication" par excellence amongst fellow workers and township dwellers in the mining areas. However, for "vertical" communication between employers and employees, it was not Flaaitaal but a "crystallised" pidgin Fanakalo that was used. Fanakalo, which originated elsewhere in the early 19th century was taken to the mines as the most convenient language to control a large and diverse workforce. It was mentioned in the mines as early as 1890 in a poem by Wilson-Moore

---

2.   The editors of this volume inform me that *gam* was also the word used in 19th C whaling English to refer to the social occasions when whaling boats met up in the middle of the ocean for sailors to chat and exchange news during their three year voyages. It would be interesting to establish whether English *gam* [gæm] is linked to Dutch/Afrikaans *gam* [xam].

and Wilson-Moore, published in their collection *Diggers Doggerel* (see Adendorff 2002: 194). It has all the hallmarks of a pidgin (Cole 1953; Mesthrie 1989), with Zulu (and to a lesser extent the closely related Xhosa) as superstrates and English and Afrikaans as substrates. Its lexis comes primarily from the superstrate, with a small amount of admixture from Afrikaans and English; its syntax is a simple kind of SVO, which could be a near-universal of contact or (less plausibly) based on simple English grammar. From the perspective being developed here, the differences between Flaaitaal and Fanakalo are important. Fanakalo is relatively stable in structure across speakers of different backgrounds that include Germanic, Indic, Dravidian, Bantu and Khoe-San. Its success in the mining context is its compactness: a small core vocabulary with minimal variation, and even less variation in syntax. There is little room for secrecy in Fanakalo: it is all work and no play. This is in stark contrast to the antilanguage character of the in-group Flaaitaal with its syllable inversions, its reduplications, its predilection for incorporating fossilised suffixes from diverse sources, its lexical metaphors and its occasional conversions of non-nasals consonants to nasals – all succinctly illustrated in Makhudu (2002). For its younger members with no criminal associations Flaaitaal is all play and no work. One point worth raising is that though Makhudu documents a rich array of variant suffixes (e.g., locatives like *-eni* from Nguni languages, diminutives like *-kie* from Afrikaans, occasional plural *-s* from English, verbal *-a* from Bantu languages) their effect is to create a vibrant lexis, rather than a new morphology or syntax. e.g., *lonjani* is diachronically a combination of English *lounge* plus the Nguni locative *-ani ~ -eni* (Makhudu 2002: 404), but its synchronic effect is that of an unmarked noun. I assume that 'in the lounge' in Flaaitaal would be expressed using Afrikaans syntax as *na die lonjani*, with no equivalent noun *lonj* or *lonja*. And even the English plural *-s* would not have plural effect in Flaaitaal.

One important aspect in which Flaaitaal is different from the other tsotsitaals is that the base language is not generally spoken as a mother tongue by its Black speakers, who have always maintained one or more of their ancestral languages, albeit in a changed urban form.[3] Makhudu's evocation of creolistics is therefore not out of place, since speakers of many languages that were not all mutually intelligible adopted a new mode of communication based on a superstrate language. Crucial differences from the Atlantic Creole model, however, are the retention of the mother tongues, and the restriction of the new lingua franca to males, often in "expressive" contexts, rather than serious communication.

---

3.   There still is some language shift within families from one Bantu language to another depending on home and neighbourhood dynamics.

It is likely that Flaaitaal would have fulfilled a "crossing" function (Rampton 1995; Kiessling & Mous 2004) in its earliest days, as speakers of a Bantu language drew on the syntax of the variety of Afrikaans spoken by Coloured people in the mining areas, who themselves would have been making lexical crossings into Bantu. Flaaitaal is thus based on the L1 of some of its speakers, whilst for others (a majority) it is an L2. Historically, it is the most important of the tsotsitaal varieties: most researchers agree that it was the prototype for its successors (Makhudu 2002; Ntshangase 2002).

Iscamtho is characterised by Ntshangase (2002: 407) as a variety parallel to Flaaitaal, but which has "very strong leanings towards Zulu and Sotho" (two mutually unintelligible Bantu languages of the area) rather than Afrikaans. He sees these two major languages as separate grammatical bases of Iscamtho: "Iscamtho is a language that is used "through" another language – a type of basilect, yet it retains its own defining features, i.e., it has no structure of its own" (2002: 407). These characterisations are apt for all the varieties under discussion. Ntshangase characterizes Iscamtho as a stylistic variety within Zulu (or of Sotho) that is potentially becoming a home language (2002: 413). That two varieties with distinct syntaxes (those of Zulu and Sotho respectively) should carry the same name is not a problem that he addresses, but one that I try to resolve in this paper.

The third important account of relevance to this paper comes from Finlayson, Calteaux and Myers-Scotton (1998) who attempted to get to grips with the urban sociolinguistic fluidities via Myers-Scotton's (1993) matrix-frame model. At that stage the model made several important claims.

1. The complementiser phrase (CP) was the relevant unit to be considered for understanding intrasentential code-switching.
2. In every mixed CP there is a matrix and an embedded language, the former providing the grammatical frame for mixed constituents, into which items from the embedded language may be inserted.
3. System morphemes (roughly: grammatical morphemes) in mixed constituents come from the matrix language, while content morphemes may come from either matrix or embedded language.
4. A CP may contain an embedded language "island": material entirely from the embedded language including system and content morphemes. In other words, within a speaking turn, should a speaker have started with a system morpheme from the "wrong" language (e.g., a pronoun) s/he is obliged to continue for the entire CP in that (embedded) language.

The model makes two further predictions: (a) that the matrix language stays the same in a mixed CP, and (b) that mixed CPs typically have the same matrix language in a speaker's turn.

The accounts by Slabbert & Myers-Scotton (1997) and Finlayson, Calteaux & Myers-Scotton (1998) went a long way to advancing our understanding of township languages and code-switching/mixing. They cautioned against exaggerating the diversity of the syntax of these varieties and, especially, of assuming that a divergent lexis implied a divergent syntax. They showed that especially where code-switching between English and African languages was concerned the claims and predictions of the matrix frame model were upheld.[4] Slabbert & Myers-Scotton (1997) made some observations about the applicability of their model to Tsotsitaal and Iscamtho. They confirm Ntshangase's interpretation of Tsotsitaal having Afrikaans as matrix and Iscamtho having an African language as matrix. "Most often it is Zulu, but Sotho-based Iscamtho is also attested, and there may be other versions with other Bantu languages as their ML [Matrix Language]." This has been borne out by subsequent descriptions of a Venda tsotsitaal (Mulaudzi 1998) and a Xhosa-based tsotsitaal (Hurst 2008). The authors conclude, further, that the code-switching configurations in Tsotsitaal and Iscamtho are not substantially different from those of the matrix-frame model. They concede that the township varieties have a substantial number of neologisms and semantic shifts of content morphemes of the embedded languages. However, they claim that this is a difference of degree, not kind compared to code-switching data from elsewhere in the world (1997: 330). In other words this is regular switching and mixing. Slabbert and Myers-Scotton's analyses seem particularly apt for the highly multilingual Gauteng area that provided their data, where code-switching is a frequent feature. However, it is preferable to keep code-switching data separate from non-switching data in evaluating the status of tsotsitaal in its various settings and manifestations. I argue later that some types of tsotsitaal give no evidence of code-switching.

4.   A unified account of tsotsitaal

I proffer a dynamic rather than individual system account of tsotsitaal that sees it as one areal phenomenon and which accounts for variant nomenclature and claims about structure and status. [T]sotsitaal is, in my view, a slang lexis (or jargon in one sense of this term) of urban South African origin that has penetrated the languages of the cities under various conditions, especially the experience of younger males in urban settings. At one end tsotsitaal has close associations with the hardships and secrecy associated with prison life and a life of

---

4.   Finlayson and Slabbert (1997: 68) subsequently claimed that things appear less certain where switching occurs between African languages.

crime, at another it is a less secretive but more expressive form of young peoples' street speech. [T]sotsitaal is thus prototypically a gendered lexicon, restricted to certain domains (prison and street). Not only does it not have a grammar, but its lexis, whilst rich and expressive in one sense, is functionally restricted, since it is limited to certain semantic fields. Words for women, prison, drinking, smoking, money and chatting abound, but no words have been reported for 'sun', 'moon', 'stars', 'fire', 'grass', 'dog', 'cat', 'book', less still for 'apple', 'potato', 'to sing', 'to dream', 'to pour', 'to go on vacation', 'give birth', 'read' etc. This expressive but ultimately limited vocabulary combines with language X to form a tsotsitaal variety of X. This combination has been called Flaaitaal or Tsotsitaal (in the specific sense) or any other of the numerous names provided by Makhudu (2002). [T]sotsitaal Zulu has been called *Iscamtho* or *Tsotsitaal* or something else by its speakers. [T]sotsitaal Sotho may also be called *Iscamtho* or *Tsotsitaal*. [T]sotsitaal Xhosa is simply called *Tsotsitaal*. There is thus no strong correlation between the name of a specific tsotsitaal and the base language; nor even between its name and the embedded slang lexis. Assuming there are eleven major languages in the country, the use of tsotsitaal does not raise this number at all. What it does is to multiply the potential registers by two. [T]sotsitaal is at least a lexical code that is variable according to time and area, it is likely to be more than that in terms of its subcultural associations (Hurst 2008 gives a detailed account of tsotsitaal as style). All of this suggests that the code-switching model is illuminating, but not literally applicable to the eleven or so major tsotsitaals of the country.[5] When speakers use a specific tsotsitaal they are not switching between antecedent codes, but rather inserting expressive subcultural words into their everyday syntax. These words come preferably from languages that are unknown or little-known to the users, or have no outside source (in being created via neologism, metaphor, inversion etc.) A speaker of Iscamtho may know no Afrikaans, even though he uses many words from this source. The traditional term for this phenomenon is of course *borrowing*, not *switching*. Myers-Scotton's (1993) model unfortunately does not draw on this salient distinction (though one that is sometimes admittedly difficult to tease out in practice). Traditional borrowing in mainstream registers can also throw up words that are unrecognisable in the source language (e.g., South African Bhojpuri *girmit* and *pilāk* are from English *agreement* and *plank*, without any antilinguistic intentions). More significantly for the code switching thesis, my analysis suggests that

---

5.   I do not propose that the languages of the country are entirely discrete, less still their slang varieties. The "eleven or so" is meant to reflect an open-endedness in the matter of language and slang boundaries.

there can be no islands in tsotsitaal: there is no embedded language with grammatical/system morphemes of its own.

5.   An English tsotsitaal?

Given the complexity of language co-existence and urbanisation in the diamond and gold mining areas, it is not surprising that there should be uncertainty over issues surrounding borrowing, switching and the use of slang registers. In order to ascertain what is perhaps essential to tsotsitaal and what is special to unfettered multilingualism I will, instead, draw on a situation that is somewhat less complex. No one has spoken of an English tsotsitaal, but such a variety does exist in KwaZulu-Natal amongst young Indian and Coloured males and has done so from at least the late 1950s. This variety has no name, but is, I argue, an offshoot of tsotsitaal. Only in 2007 did I come across a male teacher describing this variety that he spoke when drinking with his working-class friends as Tsotsitaal. It is to him that I owe the quotation in the title of this paper. I will briefly describe the variety spoken by Indian youths in the 1960s and 1970s which still survives today in robust form. So widespread is this variety in the speech community that some academics and many speakers consider it to be an Indian invention. Lawrie Barnes (1978) wrote a short article "What the chaar ous chune" which had almost cult status amongst some younger speakers because it put their salient slang vocabulary into print for the first time. *Chaar ou* is the slang term for "Indians"; it is not generally derogatory but can become so if used by outsiders with negative intentions. *Chune* is the word for 'to speak' (based on *tune*). Barnes's description did not fully cover the way that Indians spoke, since (a) the dialect involves a much larger body of lexis than that of its male youth, and (b) this way of speaking was shared by working-class youth in the province, amongst Coloureds, and some Whites and Zulus (see Bailey 1985 on male South African English slang). Today we can make the connection between youth language of Coloureds and Indians in KwaZulu-Natal with the argot/ antilanguage/ slang spoken in the Gauteng area, and I shall coin the label "Indian Tsotsitaal" for the previously unnamed variety (abbreviated to *IT*). To my knowledge the history and sociology of the connections between IT and other tsotsitaals have not been studied closely. Here are some examples, adapted from Mesthrie (1992: 148):

1.   You smaak her? ('Do you like her?' *Smaak* shows semantic shift from the Afrikaans original 'to taste')
2.   You scheme he'll vie? ('Do you think he'll go?' *Scheme* is from English with semantic shift; *vie* is ultimately from Portuguese *vai* 'to go')

3.   I chuned him "Let's chuck". ('I told him "Let's go"'; *chun* is from English 'to tune' and *chuck* is from British English slang *to do a chuck*).

The examples are instructive since they show a slang lexis embedded in an informal SAIE syntax. Some of the IT vocabulary is shared with tsotsitaals nationally. From my recollections of schoolboy slang, all the following salient terms date to at least 1960 in IT, showing a direct connection between it and other tsotsitaals:

Terms	Meaning	Ultimate Origins
vie	'to go'	Portuguese *vai* (also Afrikaans *waai*)
laanie	'White man, boss'	Afrikaans *Hollander*
pozi	'house'	English *position*
stekkie	'girl'	Afrikaans *stukkie* 'little piece'
bok	'girl friend'	Afrikaans *bok 'buck'*
blind	'bad, terrible'	English *blind*
laaitie	'boy'	English *light* (adj.)
akser	'young man' (voc.)	Afrikaans *ek sê* 'I say'
maache	'money'	Zulu *amatye* 'stones'

Slightly later, by the 1970s, other tsotsitaal words came into vogue in IT like *cherry* 'girl', *toppie* 'father' and *outie* 'streetwise youth'. In terms of semantic fields IT does have some items related to crime and prisons, but not much: *tronk* 'prison', *brant* 'to inform', *karel* 'policeman'. These are all of Afrikaans origin with deliberate semantic shifts in tsotsitaal in the case of the last two words. Of the prison gang numbers one was salient in the 1960s, but with a shifted (and incorrect sense): *twenty-six* became the young males' term for a male homosexual, rather than the hardened gang member it denoted. (The gang with homosexual associations was in fact the *twenty-eights*, not known in SAIE slang). Other fields with an expansive set of synonyms or near-synonyms are drinking, sex, fighting, speaking and people considered fools (not unlike slang world-wide). A selection from the 1960s follows:

Terms	Meaning	Origins
Drink:		
spark	'liquor'	Possibly British English slang
dronkie	'drunkard'	Afrikaans
dranklop	'drunkard'	Afrikaans *dronklap*
on	'drunk'	British slang (*on the booze*)
shut	'drunk'	English
Talking:		
thrill	'to exaggerate, sweet talk'	English
shoot	'to lie'	English

brom	'to exaggerate'	Afrikaans 'to growl'
bak	'to speak provocatively'	Hindi 'chatter'
fluke	'to pick on, to make a pass'	Afrikaans *vloek* 'to curse'

Females:

bok	'girlfriend'	Afrikaans 'buck'
cherry	'girl'	French *chérie* 'dear'
stekkie	'girl, girlfriend'	Afrikaans *stukkie* 'little piece'
vrou	'wife'	Afrikaans *vrou*
sissi	'sister'	Afrikaans *sussie*

Fighting:

skop	'beat, hit'	Afrikaans 'to kick'
moer	'beat up'	Afrikaans
klap	'slap, smack'	Afrikaans
parre	'a fight'	Afrikaans *pareer* 'ward off, parry'

The high incidence of items of Afrikaans origins is no coincidence: this was a language not known to the Indian community in the early 1960s, and hence provided a source of novel and opaque words. There is a potential contradiction here: if the language was unknown how could speakers take words from it? The answer is, of course, that borrowing is blind to etymology; male Indian youth probably learnt these terms from Coloured youths who had contacts with tsotsitaal speakers and Coloured Afrikaans speakers from the Witwatersrand and Cape Town. Though Afrikaans is viewed favourably by Coloured people in Durban, its use is not widespread. One could exoticise the Indian slang lexicon of the 1960s by pointing to words from Gujarati, Tamil and Hindi and make claims for code-switching. But this is not really the case. Indian youth of the 1960s who were pioneering slang in their community were in fact starting to lose touch with Indian languages in favour of English. The contexts in which the new tsotsitaal was used were precisely ones which excluded ancestral languages of the hearth, home and older generation. Likewise the words from Zulu in IT are not necessarily indicative of proficiency in it. However, deeper IT contained words from Zulu whose meanings were more opaque, secretive and indicative of gang or prison affiliation. One could further exoticise the syntax by showing occasional morphology from would-be embedded languages (like *-a* from Zulu verbs e.g., *zonda* 'to hate', or *-e* as noun plural and *–ie* as diminutive from Afrikaans). But these are not productive; i.e., they do not replicate the morphology of the source languages (*-e* of Afrikaans in fact occurs in double plural form with English *-s*, as in *skoenes* 'shoes', *klares* 'clothes' (the Afrikaans plurals are *skoene* and *kleure* respectively).

I have tried to make a case that Indian (and Coloured) slang of Durban is an offshoot of tsotsitaal, and that it therefore had similar psychosocial functions.

I can now make the following claims from my own knowledge and recollections of the variety as spoken in the 1960s and '70s:

1. It had no name and was not considered by its speakers as a separate variety. It was simply the way that almost all young male Indians spoke, especially those of urban working-class backgrounds and rural, male youth with contacts with the city.
2. It would have been recognised by speakers as a form of English, with an expressive lexicon (though few would have stopped to ponder over the linguistics of their speech ways).
3. The lexicon went best with a basilectal or lower mesolectal syntax (as exemplified above, see further Mesthrie 1992: 148).
4. No code switching was involved. English was certainly the base language but the quintessential vocabulary came from a variety of languages which IT users had no command of. As such there was no embedded language; and there could therefore be no embedded islands. In this regard it is noteworthy that the Afrikaans phrase *ek sê* 'I say' was treated as a single lexical item, a vocative noun, equivalent to 'boy, young man' (and pronounced *akser*).

6.   Conclusion

Most commentators have referred to tsotsitaal as a language, albeit in qualified terms: it is a language without a syntax of its own (Ntshangase 2002); it is a code-switching variety that has fossilised (Slabbert & Myers-Scotton 1997); it is youth language that is moving from "resistance identity" to "urban project identity" (Kiessling & Mous 2004: 335). Slabbert (1994: 38) goes further: "If there is one thing of which this research has convinced us, it is that Tsotsitaal is not a variety of anything; to its speakers it is a language in its own right". Yet, as the author concedes elsewhere, some speakers themselves have spoken of it as "a matter of words" (Slabbert & Myers-Scotton 1997: 328). It is therefore necessary to clarify (and few commentators have done this) whether tsotsitaal is the embedded lexis, or the matrix grammar plus the embedded lexis. While it is true that in some contexts and places the lingua franca function of tsotsitaal for males (especially Black males) is significant, there are many contexts in which this is less relevant, as in tsotsitaal versions of Xhosa in Cape Town, Zulu in Durban and English in Durban. However, given the richness of the subcultural associations of tsotsitaal, it would be selling it short to think of it as mere jargon lexicon (argot/ slang). On the whole I see tsotsitaal not as a language, let alone as a mother tongue, but as a style of speaking a language. [T]sotsitaal is an embedded stylised lexis;

not the whole language system (syntax, phonetics plus vocabulary). Thus, there is tsotsitaal Zulu (or Xhosa or Afrikaans or English, etc). Its speakers would be using an urban or non-standard variety of, say, Zulu, in the tsotsitaal style (with its lexicon, dress, attitudes, gestures, etc.) There is nothing to prevent tsotsitaal Zulu from entering into switching with other codes, say English or even tsotsitaal Afrikaans (=Flaaitaal) if speakers had a command of them. But code-switching is not an essential ingredient for a tsotsitaal to exist. Finally, the special status of Flaaitaal is recognised, as the variety that first blossomed from multilingual urban contacts of the mining era, probably among Coloured speakers interacting with speakers of Bantu languages, who then adopted it as a second language. This second language has the potential of turning into a first language, but it remains to be seen whether it would then lose its antilinguistic character in becoming a language of women and children in the home. But from a synchronic perspective, even Flaaitaal fits the general principles outlined in this paper: it is non-standard Afrikaans with a tsotsitaal lexis. I conclude with a summary of these main principles:

Principles underlying tsotsitaal varieties:

Nomenclature:

a. Although speakers refer to an entire code as "tsotsitaal", and speak of it as a language, from a linguistic viewpoint the "tsotsitaal-ness" inheres mainly in the lexicon (and perhaps other paralanguage features).
b. Specific names for tsotsitaals abound and overlap; there is no strong correlation between the base language and the name, or even the lexicon and the name.

Structural:

c. A previously existing L1, sometimes adopted and adapted as an L2 lingua franca, provides the syntax, into which slang lexis is embedded along with lexical items usually associated with the base language.
d. The syntax will be the most informal/ urban/ non-standard variety of the base language.

Semantic:

e. The slang lexis will preferably come from a little-known language or sources unknown to a majority of speakers, as novelty and opacity are prized.
f. The lexis will show semantic twists as well as neologising.
g. It will involve overlexicalisation in certain areas.
h. It will be restricted to certain semantic fields.

Sociolinguistic:

i. The more marginalised, oppressed or oppositional the speaker, the "deeper" the tsotsitaal. At this end it is spoken as a secret code by people with criminal, gang or prison associations.

j. "Lighter" street varieties are prized for their creativity and sub-cultural associations among male youth, including educated ones.

k. Some of the lexis may be used by females and retained by males as they grow older and hence enter the mainstream L1. (But this does not make the tsotsitaal a home language – all tsotsitaal speakers command a previously existing mainstream language natively).

References

Adendorff, R. 2002. Fanakalo: A pidgin in South Africa. In *Language in South Africa*, R. Mesthrie (Ed.), 179–198. Cambridge: CUP.

Bailey, R. 1985. South African English slang: form, function and origins. *South African Journal of Linguistics* 3(1): 1–42.

Barnes, L. 1978. What the chaar ous chune. *The Eighteen Twenty* 52: 19.

Calteaux, K.V. 1996. *Standard and Non-standard African Language Varieties in the Urban Areas of South Africa: Main report for the STANON Research Programme.* Pretoria: HSRC Publishers.

Childs, G.T. 1996. The status of Isicamtho, an Nguni-based urban variety of Soweto. In *Pidgins and Creoles: Structure and status*, A. Spears & D. Winford (Eds), 341–70. Amsterdam: John Benjamins.

Cole, D.T. 1953. Fanagalo and the Bantu languages in South Africa. *African Studies* 12: 1–9.

Finlayson, R., Calteaux, K. & Myers-Scotton, C. 1998. Orderly mixing: Code-switching and accommodation in South Africa. *Journal of Sociolinguistics* 2(3): 395–420.

Finlayson, R. & Slabbert, S. 1997. "We just mix" – code-switching in a South African township. *International Journal of the Sociology of Language* 125: 65–98.

Glaser, C. 2000. *Bo-tsotsi: The youth gangs of Soweto, 1935–1976.* Cape Town: Heinemann.

Hurst, E. 2008. Structure, style and function in Cape Town Tsotsitaal. Ph D thesis, University of Cape Town.

Janson, T. 1984. A language of Sophiatown, Alexandra and Soweto. *In York Papers in Linguistics 11 – Papers from the York Creole Conference, September 24–27, 1983*, M. Sebba & L. Todd (Eds), 167–180. York: University of York.

Kiessling, R. & Mous, M. 2004. Urban youth languages in Africa. *Anthropological Linguistics* 46(3): 303–341.

Makhudu, K.D.P. 1980. An Etymological and Morpho-phonological Description of Flaaitaal/Tsotsitaal: A sociolinguistic perspective. BA thesis, University of the Witwatersrand.

Makhudu, K.D.P. 2002. An introduction to Flaaitaal. In *Language in South Africa*, R. Mesthrie (Ed.), 398–406. Cambridge: CUP.

Mesthrie, R. 1989. The origins of Fanagalo. *Journal of Pidgin and Creole Studies* 4: 211–240.

Mesthrie, R. 1992. *English in Language Shift: The history, structure and sociolinguistics of South African Indian English*. Cambridge: CUP.

Mfenyana, B. 1981. Isjita-Scamto: The black arts in SasAfrika. In *Reconstruction – 90 Years of Black Historical Literature*, M. Mutloatse (Ed.), 294–304. Johannesburg: Ravan Press.

Mfusi, M. 1992. Soweto Zulu Slang: A sociolinguistic study of an urban vernacular in Soweto. *English Usage in Southern Africa* 23: 39–83.

Molamu, L. 2003. *Tsotsi-taal: A dictionary of the language of Sophiatown*. Pretoria: University of South Africa Press.

Mulaudzi, 1998. The impact of Tsotsitaal on Venda. Paper presented at the Linguistics Society of Southern Africa conference, University of Potchefstroom, July 1998.

Myers-Scotton, C. 1993. *Duelling Languages*. Oxford: Clarendon.

Ntshangase, D.K. 1993. The Social History of Iscamtho. MA thesis, University of the Witwatersrand.

Ntshangase, D.K. 2002. Language practices in Soweto. In *Language in South Africa*, R. Mesthrie (Ed.), 407–418. Cambridge: CUP.

Rampton, B. 1995. *Crossing: Language and ethnicity among adolescents*. London: Longman.

Rudwick, S. 2005. Township language dynamics: isiZulu and isiTsotsi in Umlazi. *Southern African Linguistics and Applied Language Studies* 233: 305–317.

Schuring, G.K. 1983. Flaaitaal. In *Taalverskeidenheid*, G.J. Claassen & M.C.J. Van Rensburg (Eds), Pretoria: Academica.

Slabbert, S. 1994. A re-evaluation of the sociology of Tsotsitaal. *South African Journal of Linguistics* 12(1): 32–41.

Slabbert, S. & Myers-Scotton, C. 1997. The structure of Tsotsitaal and Iscamtho: Code switching and in-group identity in South African townships. *Linguistics* 35: 317–342.

Stone, G.L. 1991. An ethnographic and socio-semantic analysis of lexis among working-class Afrikaans-speaking coloured adolescent and young adult males in the Cape Peninsula, 1963–1990. MA thesis, University of Cape Town.

Stone, G.L. 2002. The lexicon and sociolinguistic codes of the working-class Afrikaans-speaking Cape Peninsula coloured community. In *Language in South Africa*, R. Mesthrie, (Ed.), 381–398. Cambridge: CUP.

Tok Bokis, Tok Piksa

Translating parables in Papua New Guinea

Bambi B. Schieffelin
New York University

This chapter focuses on Bible translation practices, central to Christian missionization in Papua New Guinea, a site of intensive linguistic and cultural contact, and a productive context in which to examine the dynamics of multiple, competing and contradictory conceptualizations about language, language use, and language ideologies. Focusing on the genre "parable," it tracks how translation changes made by New Testament Bible translators working in Tok Pisin, from tok bokis to tok piksa, created ethnopragmatic challenges for Bosavi pastors who struggled in a rapidly shifting metalinguistic terrain to create local meanings across languages and texts. The essay argues that the importance of genre as an interpretive frame cannot be underestimated in terms of understanding changes in linguistic and cultural meanings over time, especially in language contact situations.

Keywords: Tok Pisin; Bosavi; Christianity; missionaries; translation; tok bokis; hidden language; parables; language ideology; language change

Gillian Sankoff speaks of language – as well as languages – as "living in a social world" (1980: xxi) and having social lives. As her research has shown us, those lives are never simple or straightforward. Complicated by competing interests of their speakers, who cannot let them be the same from one setting or genre to the next, and provoked by all kinds of contact, desired and not, speakers transform languages over time and place in subtle but often predictable ways. At times this happens at a leisurely pace, other times shifts are sudden, and even unexpected. Furthermore, the motivations for change may be public, transparent and even conscious, but other times, motivations remain obscure, at least to contemporary speakers and researchers. Understanding these complicated social relationships between languages and speakers requires careful examination of cultural, cognitive and historical influences affecting language use – a hallmark of Sankoff's work.

Inspired by her theoretically original analyses of linguistic transformations in Tok Pisin, and recognizing the ideological complexity and increasing importance of its incorporation throughout Papua New Guinea's verbal repertoires,[1] I examine Tok Pisin's forays into the Bosavi cultural and linguistic landscape. While a few Bosavi men brought some knowledge of Tok Pisin back with them in the late 1960s after contract labor stints on Highlands tea and coffee plantations, its use in the villages was very limited and had little impact on the verbal repertoires of Bosavi speakers. It was not until the early 1970s with the arrival of a Christian mission and the *Nupela Testamen*, the Tok Pisin translation of the New Testament, that Tok Pisin's presence was heard and felt. In Bosavi, as in many other parts of Papua New Guinea, the *Nupela Testamen* became the source text for vernacular Bible translation, but in Bosavi, it also became the major source for hearing Tok Pisin, as it was read aloud and spontaneously translated into the Bosavi language in church services held several times each week. It was in these contexts that the ethnopragmatic and interpretive aspects of cultural and linguistic contact became salient and it is these dimensions of contact, central to translation practices that are foregrounded in the discussion that follows.[2]

To explore these ethnopragmatic and interpretive dimensions I focus on parable, a culturally recognized and linguistically named verbal genre or way of speaking that is relevant in the context of New Testament translation, and has correspondences with traditional ways of speaking that have been documented in different speech communities in Papua New Guinea. I examine the consequences of a set of translation decisions at the metalinguistic level made over time and across linguistic communities that, despite their historical and cultural differences, share an intriguing set of similarities. In translating speech events from Scripture, ascertaining and naming the genre or manner of speaking are based on an assumption of semantic equivalency thought by missionaries to apply across linguistic communities. In this sense,

---

1. Tok Pisin is one of the official languages of Papua New Guinea. Its changing status can be tracked through the ways it has been variously named and evaluated throughout Papua New Guinea's history, for example, Talk-Boy (Mead 1931), Pidgin English (Murphy 1943), Melanesian Pidgin English (Hall 1943), Melanesian Pidgin (Mihalic 1971), though as Sankoff notes (1980: 6), speakers call it Tok Pisin. Mühlhäusler 1979; Sankoff 1980; Smith 2002; and Mühlhäusler, Dutton & Romaine 2003 provide historical and linguistic accounts of Tok Pisin's development.

2. In order to facilitate the comparison between languages and to highlight borrowings and switches between languages, I follow well-established typographical norms. In this chapter, all Bosavi forms are shown in *italics*, all Tok Pisin forms are underlined. English glosses are given in 'single quotes'.

the assumption seems to be that semantic and pragmatic equivalencies can be found, at least as easily as referential meanings (Nida 1964).

In this essay, I suggest that this assumption maybe problematic. Naming genres and speech acts pose their own set of translation issues, extending beyond the domain of semantics into metapragmatics. The decisions involved are particularly critical as such designations not only affect the system of metalinguistic categorization, but also have consequences for how named genres or speech acts are subsequently recognized and translated into vernacular languages. A locally recognizable genre name establishes and enables an interpretive framework for the reader/listener, including speakers of local languages. In missionizing contexts where Bible translation is central to conversion, how a named genre from the source Bible text is translated also positions it semiotically and ethnopragmatically (Duranti 1993) in relation to other vernacular genres, e.g., traditional and introduced, secular and Christian, real or fictive. Once a genre is named, other factors, such as code choice, syntactic structures and other pragmatic dimensions, conventions or modes of interpretation, may also contribute to or even define how such genres are to be understood.

Like other metalinguistic and metapragmatic aspects of language, the assignment and translation of genre names, and an understanding of genres or styles more broadly, do not exist in isolation. In their creation they connect and index other genres and styles, which are embedded in and understood in terms of sociohistorically specific interpretive frameworks. Language ideologies shape these dynamics as well since decisions about what language is most appropriate for bringing the Word depends on choices implicating all domains of language, and such decisions are never neutral. For example, missions may chose to proselytize in a colonial language, a lingua franca or a vernacular (among others), and may use names from any of those languages, draw on traditional, local genres or create neologisms or loan translations to suit their ideological purposes. Such decisions are tied to language ideologies, which are never about language alone. They envision and enact ties of language to identity, epistemology and morality, such that a focus on language ideology provides a bridge between linguistic and social theory (Woolard & Schieffelin 1994). That there are fundamental linkages among such apparently diverse cultural categories as genre, tradition, religion, modernity and power (among others) – an idea shared across many speakers and communities all over the world – allows us to relate the microcultural organization and semiotics of all varieties of talk to broader social domains. In contexts of intensive missionization, where the transformation of persons is the goal, the linkages across genre, tradition and religion themselves – and how they are named – become sites of contestation.

In English, the term parable usually refers to a short narrative with two levels of meaning. Such narratives figure importantly in the Gospel of Mark, and understanding

their meaning and use requires both cultural and linguistic knowledge. In Mark, Jesus speaks in parables as a pedagogical style, and the meaning of this way of teaching (or preaching) has received extensive critical commentary from theologians, philosophers, and literary analysts, each with a distinct but sometimes overlapping set of textual, interpretive and religious frameworks. There are also several centuries worth of scholarly debates about etymological, philological and subsequent translation histories of the word "parable" in New Testament studies, which include concerns over the pragmatic relationship of parable to two other ways of speaking, riddle and allegory (Green, McKnight & Marshall 1992). Finally, there is a critical literature on whether as a genre, parables obscure or reveal, and to whom, the Good News, especially in Mark 4, which includes the Parable of the Sower.[3]

Shifting our focus to back to Papua New Guinea, the translation of "parable" from English into Tok Pisin, and from Tok Pisin into Bosavi has a much shorter, but nonetheless culturally and linguistically interesting history. Translated in the first edition of the *Nupela Testamen* in 1969 as tok bokis 'secret language', it was changed in the 1978 revision to tok piksa 'parable'. These definitions from Mihalic's *Jacaranda Dictionary* (1971: 195), the most authoritative and widely used dictionary during the time period of translation, are both in the main entry for tok 'talk', where tok bokis and tok piksa are as listed as kinds of tok. The definition for tok bokis varies within the dictionary itself, for example, under the main entry bokis 'box,' tok bokis is also glossed as 'parable' (1971: 74).[4] Furthermore, in the English to Melanesian Pidgin section of the Dictionary, "parable" is multiply glossed as parabel, stori, tok piksa and tok bokis. This same entry lists alternative forms mekim tok parabel and mekim tok bokis 'to speak in parables'. I have not found parabel used in any Tok Pisin texts.

These variable and overlapping translations and alternative glosses in Tok Pisin suggest a culturally intriguing story about tok bokis, and may explain why tok piksa subsequently became the translation term of choice for parable in the Tok Pisin Bible. This shift also had significant interpretive consequences when translated into the Bosavi language, ones that led to cultural and linguistic confusion.

---

3.   See Kermode (1979), especially chapter 2, including footnotes, for a concise and incisive review of this extensive literature.

4.   In contrast, as a kind of tok, Mihalic's earlier *Grammar and Dictionary of Neo-Melanesian* glosses tok bokis as 'a secret language' (1957: 151); tok piksa is not listed. Piksa is glossed as 'a picture, diagram, sketch, movie, cinema' (1957: 107), with no association to forms of talk, similar to Mihalic's later work (1971: 154). The English reversal glosses 'parable' as parabel, stori (1957: 245).

In what follows, I view these outcomes within a framework that argues for the importance of considering metapragmatics and language ideology in situations that involve cultural and linguistic contact, where various types of translation practices are bound to occur, ranging from the mundane and ephemeral, to the sacred and persistent. In addition, shifts in naming and interpretation in Tok Pisin and Bosavi highlight the historically complex, polyvalent nature of the category parable itself and its own contested history in translation across ancient and contemporary languages and religions.

1. Bosavi background

The Bosavi people live north of Mt. Bosavi in the Southern Highlands of Papua New Guinea.[5] In this rainforest environment, 2000 or so Bosavi people inhabit scattered communities ranging from 60–100 people. Traditionally egalitarian, they practice swidden horticulture and hunt and fish. There are four mutually intelligible named dialects of the Bosavi language, one of which is Kaluli. Until the mid-1990s, most people were monolingual in their vernacular, which in 1998 was still the language of village life. As in many other societies in Papua New Guinea, Bosavi people were not literate before government or mission contact. Until 1964, contact with non-indigenous people was very limited. At that time members of an Australian nondenominational Protestant mission[6] organized the construction of a small airstrip/mission station, which they staffed intermittently with native pastors from outside the area. It was not until the early 1970s, however, when two Australian missionaries, members of the Asia Pacific Christian Mission, settled at the mission station and began intensive proselytizing activities that things began to quickly change.

The missionaries considered the Bible the center of all preaching, and as fundamentalist Christians they took a literalist stance toward Scriptural translation and interpretation. They also explicitly rejected the idea that knowledge about local cultural practices might be helpful to their own agenda, and viewed them as not only irrelevant to conversion, but as obstacles to its success. While dismissive of local cultural practices, their mission's language policy dictated working in the vernacular, "the shrine of a people's soul" (Rule 1977: 1341). Fluent Tok Pisin

---

5.   See E.L. Schieffelin (1976); Feld (1982); Schieffelin (1990, 2000, 2002, 2007); Schieffelin & Feld (1998).

6.   The Unevangelized Fields Mission in Papua New Guinea was renamed the Asia Pacific Christian Mission in 1970.

speakers, but lacking linguistic training, the missionaries mediated Christianity through Tok Pisin and the <u>Nupela Testamen</u>, the Tok Pisin Bible. This limited the number of people they could interact with to a small group of younger men who had learned some Tok Pisin outside of the area. Interested in what the missionaries promised both spiritually and materially, they were willing converts. Though they lacked formal schooling, within a short period time they were baptized and sent out to the villages as pastors, and given the authority to preach and convert others. These recently missionized Bosavis became active missionizers and played a major role in producing the language through which people understood Christianity.

Reading aloud from the <u>Nupela Testamen</u> was central to village church services. Pastors haltingly read one verse or segments of verses in Tok Pisin, translating each segment into Bosavi. Pastors struggled to come up with Bosavi equivalencies for Tok Pisin words, idioms, and speech acts, most of which were foreign, and for which glosses were difficult (Schieffelin 2007). As a result, translation practices from Tok Pisin to Bosavi inadvertently added substantial innovations to all levels of the vernacular through loan translations and calques, "literal" translations that stayed very close to the surface features of Tok Pisin expressions. In these ways pastors created a distinct Christian register in the Bosavi language, one that was accompanied by a range of distinct bodily practices. Thus as local pastors preached in the vernacular, it underwent semantic and pragmatic shifts due to contact with Tok Pisin. These spontaneous translation practices are therefore a rich source for understanding metalinguistic struggles and language ideological issues that resulted from contact between languages, cultures and cosmologies. They also illustrate a set of issues specific to genre.

2.   The written texts: *The Good News Bible* and the *Nupela Testamen*

First published in 1969, the *Nupela Testamen* was based on the American Bible Society's *The New Testament in Today's English Version* (1966), which is often referred to as *The Good News Bible*. Its target audience was children and uneducated adults, and "rather than follow traditional vocabulary and style in the historic English Bible versions, it uses standard, everyday, natural forms of English."[7] In addition, *The Good News Bible* also added various "readers' helps" to facilitate comprehension, including dividing the text into sections and adding headings to indicate the contents of each (1992: iii). In Mark 4, for example, seven titled sections not only indicate "content," but some titles include metalinguistic terms to introduce and provide guides for interpreting the text that follows. *The Good*

---

7.   Foreword, *The Good News Bible in Today's English Version* (1992).

News Bible titled verses 1–9 "The Parable of the Sower" and verses 13–20 "Jesus explains the Parable of the Sower."

In Papua New Guinea, the translators of the 1969 *Nupela Testamen* followed this model and used relatively uncomplicated lexical and syntactic expressions in Tok Pisin to make the text accessible to its intended audience. They modified some of the section titles, for example, verses 1–9 ("The Parable of the Sower") did not use metalinguistic terminology (e.g., tok bokis), but instead the title described what a "sower" is, and does, Man i tromwe pikinini kaikai long gaden. While not mentioned in the title, tok bokis first appears in verse 2 referring to what Jesus is doing, Na em i givim planti tok long ol long tok bokis (see Appendix A for English glosses of the Tok Pisin verses). Verses 13–20 were titled and translated close to the English, "The Explanation of the Parable," as As bilong tok bokis bilong pikinini kaikai, explicitly marking the genre explained by Jesus as tok bokis. The term tok bokis is used several times in the relevant 16 verses in this textual selection.

The 1969 *Nupela Testamen* went through two subsequent revisions, 1978 and 1989, to insure that the Tok Pisin was standardized, conventional, and contemporary, though it was based on a relatively rural, more conservative variety, rather than urban varieties which were in more contact with English. One change in the 1978 revision was a shift in the translation of parable from tok bokis to tok piksa. The title for verses 1–9 in 1978 was similar to that of 1969, and the change in genre name initially appears in verse 2, describing what Jesus is doing: Na em i givim planti tok long ol long tok piksa. The title for verses 13–20 also remained consistent, except for the change in genre designation, translated as As bilong tok piksa bilong pikinini kaikai. The 1989 *Buk Baibel*, which added the Old Testament, changed the title of verses 1–9, and with some orthographic and lexical changes, made the genre explicit, Jisas i tok piksa long man i tromoi pikinini wit long gaden. This edition used tok piksa throughout Mark to describe the activity of Jesus speaking in parables. Thus over time, the genre of this text, as a parable or as an explanation, in either case, as a verbal activity closely identified with Jesus' style, became explicitly established as tok piksa in Tok Pisin.

Commentaries about the translation revisions of the Nupela Testamen published by missionary-linguists and linguists have focused on lexical choice and grammatical decisions (c.f. McElhanon 1975a; Mundhenk 1990; Lothmann 2006). There has been surprisingly little published on issues pertaining to translating the pragmatic and metalinguistic dimensions of Scriptural language, such as reflexive language, genre, or speech acts.[8] This reflects a particular set of concerns, those focusing on formal linguistic issues, and indicative of the referentially based

---

8. For an exception see Franklin (1992).

language ideology of missionary-linguists. However, these neglected domains, which are culturally inflected, are precisely the ones that have undergone the most extensive revision in the Tok Pisin Bible. In addition, the ways in which they were translated from Tok Pisin into local vernaculars have had significant and lasting cultural consequences. We examine selections from the Gospel according to Mark as an example of such consequences in Bosavi.

3.   Mark 4, 1–20, with some deletions

In its evangelizing activities, the Asia Pacific Christian Mission always introduced Mark as the first Gospel, and the missionaries working in Bosavi continued this tradition in pastor training and Bible-study classes.[9] Because it was introduced in the early 1970s and used often, it was relatively familiar to church-goers. Several portions of Mark became Bosavi favorites, among them, two named sections of Mark 4, verses 1–9, and verses 13–20. These were repeated in the church services I audio taperecorded and transcribed over numerous ethnographic and linguistic fieldtrips from 1975 through 1998. The verses contain a parable, explain a parable, are about Jesus teaching using parables, and how believers and non-believers heard and understood parables differently, for starters. Thus, these verses of Mark are all about language pedagogy and language socialization, metalanguage and language ideology, and also about interpretation and belief.

The Parable of the Sower (4, 1–9) portrays Jesus addressing a large crowd composed predominantly of outsiders (nonbelievers), teaching by telling the Parable of the Sower. In this parable, He recounts three conditions that inhibit the productivity of most of the seeds that were sown: birds stole them, rocky ground prevented their development of deep roots, and thorns choked the further development of the plants. The fourth condition, however, provides the contrast, productivity, and those seeds that fell on rich soil produced abundant fruit. He opens and closes the parable with exhortations to listen, and pay attention. Verses 10–12, titled "The Purpose of the Parables"[10] begin with a temporal and spatial shift. Jesus is alone with the disciples

---

9.   In 1998, Andy Grosh of the Summer Institute of Linguistics produced a translation of Mark, the first Gospel translated into the Kaluli language. SIL's translation practices differed from the language ideology of APCM in that they did not use the *Tok Pisin Bible* as their source text. Handman 2007 provides an excellent discussion of SIL language ideology and Bible translation in Papua New Guinea.

10.   The meaning and ordering of these verses has received extensive critical commentary in the theological literature, which is beyond the scope of this paper (see Kermode 1979).

and other believers (insiders) who ask Him about the meaning of parables. He responds that they that have been given the Mystery of the Kingdom, but to those outside, everything is in parables, and the question of understanding is raised.

In Bosavi, pastors always omitted these three verses when reading and translating in church services, a point that will be discussed below. They followed verses 1–9 with verses 13–20, titled "Jesus explains the Parable of the Sower." These verses begin with Jesus posing a rhetorical question to the disciples (insiders), the same addressees to whom he was just speaking (in verses 10–12): if they do not understand this parable, how can they understand any of them? (verse 13). Jesus then explains the parable: the Sower sows the Word, resulting in three outcomes which were unsuccessful and one which was successful, each dependent on the context in which the "seed" was sown. While Mark 4, verses 1–9 followed by 13–20 were read in Tok Pisin and translated into the vernacular for over twenty-five years, the meaning, and how to get to it, remained hidden for most Bosavis.

Many factors contributed to this obfuscation. Bosavi people lacked experience with books and printed texts, photographs or pictures, and while many purchased Bibles, they could not read them. More importantly, the performance of Bible reading and translation was unlike any traditional Bosavi genre, including story telling. There was no tradition of one person delivering an uninterrupted monologue. All talk, including stories, was carried out with active audience participation, such that all narratives were co-constructed and through verbal interaction speakers commented, elaborated, clarified, or elicited information, thus making the telling of a story always a dialogic event, an example of a preferred Bosavi aesthetic style of "lift-up-over-sounding" (Feld 1988). This was in many ways the opposite of monologic domination that characterized Christian services, where one speaker, the pastor, considered the most knowledgeable person, showed the Word, and everyone else listened.

The overarching textual or narrative structure of these verses was difficult for pastors to manage, even after many years of practice. The narrative structure in Scripture differed from Bosavi narrative aesthetics, and local pastors had trouble tracking the referents in the text, as well as figuring out what was foregrounded and backgrounded. Members of the congregation struggled to follow the metaphorical shift that takes place in the parable itself and the explanation that followed. In the parable the "seeds" are the foregrounded figure (the Word of God), while the explanation foregrounds the soil, which can be said to represent different types of listeners that are more or less receptive to the Word. Various types of hearers are "sown" in the explanation, v. 15–20.

Another challenge at the broader textual level, for both pastors and listeners, was tracking the shift in Jesus' addressees that occurred across the two sections of Scripture. Verses 1–9 are addressed to the crowds, the outsiders or nonbelievers,

while verses 13–20 include the disciples (insiders) as well. Without verses 10–12, which only address the disciples, their added presence, as well as that of other believers, cannot be retrieved once verse 13 begins. Without verses 10–12, Bosavi listeners would not realize that the explanation of the parable addresses two distinct audiences, outsiders and insiders (nonbelievers and believers), each appearing spatially separated without the textual bridge that joins them.

According to the local pastors, the missionaries had instructed them to omit Verses 10–12, the Purpose of the Parables, because they thought that material would be too difficult for native people to understand. These verses not only add coherence to the entire section, but they contain interpretive material about understanding parables. Without those verses, Bosavi people were excluded as listeners, or hearers. Thus the Scriptural guidelines for how to understand, hear, or most importantly, think about the meaning of those verses, were literally erased. Bosavi people were, by virtue of this omission, designated as outsiders, like the crowds of nonbelievers who also were not the recipients of these verses. And to those outside, everything is in parables, where meaning is left hidden.[11]

While the crowds are not overhearers to the disciples' questions about the parables and Jesus' responses to them, the reader/hearer of the text is included and the existence of the secret, the Mystery of the Kingdom of God, is shared. From a narrative perspective, in terms of Goffman's (1981) participation framework, the reader/hearer of these verses is a type of recipient, a member of the audience of listeners. As such, the reader/hearer is included as a potential member of the inside circle of believers, sharing secret or privileged knowledge. This distinction between inside/outside, which is used in other New Testament passages and in proselytizing, is symbolically and materially potent. As terms of Christian inclusion and exclusion, in Bosavi, the notions of inside and outside (*usa* and *ha:la:ya*) were used to designate Christian converts (*usa* 'inside') from heathens, 'those to the side' or *ha:la:ya*. These categories were not only incorporated into Bosavi religious rhetoric, but the importance of this division was also reflected in the reorganization of Bosavi villages into Christian (inside) and non-Christian (outside) space.

---

11. My own reading of the consequences of this textual omission is that these verses are critical in that they make explicit the distinction between insider/outsider knowledge and status: outsiders (nonbelievers) cannot understand parables, while insiders (believers) do. Without verses 10–12 Bosavi people lacked the textual information to make that connection.

In addition to these broader interpretative factors, the activity of reading in Tok Pisin was itself a challenging task for the few Bosavi pastors who as adults had received literacy training at the mission station. They were the first Bosavis to attain any literacy skills, skills that only included reading, not writing. Many chose to memorize verses, or parts of verses, a choice that became obvious once they were faced with revised Tok Pisin texts. Second, because most of these same pastors could not write, nor were they ever encouraged to do so, in church services they would translate Tok Pisin verses into Bosavi each time they read them. The two languages are typologically dissimilar, and Tok Pisin's verb-medial structure did not map easily into the verb-final structure of Bosavi, a non-Austronesian language, further complicating the translation process. Given that pastors were instructed to stay close to the Tok Pisin text and produce a literal translation, the result was an often disfluent, and sometimes ungrammatical and incomprehensible Bosavi translation. The reading and translation were marked by self-repair, repetition, paraphrase, and dropped as well as added segments as the pastors tried to match phrases and verses of one language that they had just read onto another through loan translations and calques, all the while trying to create some semblance of coherence and authority, as will be illustrated in the examples below drawn from services.

While these performance and textual problems were all obstacles to understanding, the ambiguous interpretive framework or genre of these stories was one of the most significant sources of difficulty. Whether or not they understood the importance of listening and hearing, which pastors repeatedly stressed, without interpretable, interpretive guidelines provided by genre, Bosavi listeners did not know how to take meaning from what they heard. They did not know what the story was really about, or whether there was more than one meaning to be taken from it. All of these different features – cultural, linguistic, and ideological – contributed to the problem of how to take meaning from Scripture. We turn next to examine how "parable" was initially translated in Tok Pisin and then in Bosavi and explore how multilayered, culturally shaped translation practices added to the indeterminate meaning of these verses for Bosavi people.

4.   Tok bokis, tok piksa: Parables in Tok Pisin

The term parable (Greek *parabole*) is used to translate the Hebrew *mashal*, which covers a variety of literary forms, including allegory, similitude, and proverb. In the *New Testament*, what is called parable may also refer to riddles, but its usual application is to short narratives that illustrate comparisons between Christian

truths and everyday life activities. Parables as forms of figurative language require a particular disposition toward the information, that is, they require reflection and thought for understanding.

The 1969 <u>Nupela Testamen</u> translated "parable" as <u>tok bokis</u>, which, like 'parable', covers a broad stylistic range and does not find easy translation equivalents. Composed of <u>tok</u> (English 'talk') and <u>bokis</u> (English 'box'),[12] it is attested in two reliable Pidgin English sources from the early 1940s, though its etymology as a metalinguistic term is not clear.[13] Murphy lists <u>tok bokis</u> as a "metonymous term used to describe an object by applying to it the name of another object or condition to which it has a somewhat metaphoric similarity" (1943: 64). In another early, authoritative Pidgin English dictionary/grammar, Hall defines <u>tok bokis</u> as 'secret talk' (1943: 121). Neither of these two sources lists <u>tok piksa</u>, though both list <u>piksa</u>, Murphy glossing it as "n. photograph, picture; v.i. take a photograph" (1943: 51) and Hall as "picture, movies" (1943: 113), thus one can safely assume that <u>tok bokis</u> predates <u>tok piksa</u> in the lexicon.

Along with the publication of the *Jacaranda Dictionary* and the *Nupela Testamen*, the late 1960s through the 1970s was a time of concentrated attention to and scholarly and public debate about the future of Tok Pisin and its role it in politics, education, and national development more broadly.[14] As evidence of the vitality of the language, Brash (1971), writing about the "imaginative dimensions of Melanesian Pidgin," notes rapid innovations in the area of figurative language and describes several domains of talk. He claims that the first, variously labeled as <u>tok hait</u>, <u>tok baksait</u> or <u>tok bokis</u>, is well recognized in Pidgin and uses ambiguity or a form of double talk to deliberately exclude outsiders, and is used for talking about the boss, when drinking or card-playing, and for sexual themes, among others. The second, <u>tok piksa</u>, are similes, used for "comparing of one person/object to another, common in Pidgin since the language was first used" (1971: 15).

In an extended treatment of the growth of New Guinea Pidgin, Mühlhäusler (1979) describes <u>tok bokis</u> and <u>tok piksa</u> as linguistic registers or special styles, but

---

12.   Murphy (1943) notes that <u>bokis</u> can refer to female genitals, presumably from English slang, and warns non-native speakers to avoid the phrase <u>bokis bilong misis</u>, which does not refer to a suitcase.

13.   Aufinger (1948: 90), in an account of different varieties of secret languages spoken off the coast of Madang, glosses secret language as <u>tok bokis</u>, 'talk that is in a box, enclosed in a disguise or circumlocution.' Unfortunately he does not indicate the source of that explanation, indigenous or self-generated. The term <u>tok bokis</u> is not in the lexicons of Solomon Islands Pijin or Bislama.

14.   See, for example, McElhanon (1975b) for scholarly papers written in Tok Pisin on Tok Pisin language policy and use.

seems to limit tok bokis at least to the lexical level: "What is involved in most cases is a more or less drastic change in the meaning of certain lexical items, changes which may go unnoticed by the outsider who only understands the literal 'innocent' meaning" (1979: 341). Citing Aufinger (1949) Mühlhäusler emphasizes its use in secret or hidden meanings, for taboo topics (death, bodily functions), or in secret varieties known to small groups of initiated people or members of cargo cult movements. He views tok bokis lexical items as being "normal New Guinea Pidgin items with a different meaning."

He distinguishes tok bokis from tok hait, which he claims is used to exclude outsiders, including missionaries from gaining knowledge of cargo cults, and other activities they would disapprove, and like Brash, views tok piksa as the use of similes, but adds the use of metaphors as well.[15] The 1971 *Jacaranda Dictionary* defined tok bokis as 'secret language,' but under the main entry bokis 'box,' tok bokis is glossed as 'parable' (1971: 74). The 2007 online version of the Jacaranda Dictionary glosses tok bokis as 'a means of talking about something without mentioning it, circumlocution, veiled speech.' Tok piksa is glossed as 'an explanatory analogy.'[16] Thus, linguistic sources focusing on Tok Pisin provide little detail in terms of actual uses of tok bokis and tok piksa, and as should be evident, the linguistic functions tend to focus on relatively narrow domains, e.g., lexical choices and phrases.[17]

Anthropologists' reports on language practices, in contrast, tend to use tok bokis to refer to a broad range of indirect styles of speaking that are shared widely throughout Papua New Guinea. They range from secret and ritual languages that are meant to hide meanings from outsiders, to styles of indirect, circumlocutionary or metaphorical speech, which, for example in Mt Hagen are said to "veil, … bend or fold" speech (Strathern 1975: 189) or in Bosavi to use "turned over words" (Feld 1982: 138–139). According to speakers in these and other Papua New Guinea communities, these ways of speaking insure that one cannot always see (or hear) all of the "sides"; what is literally said is not what is really meant. While some scholars glossed vernacular terms with English translations, those working in areas where Tok Pisin was part of the verbal repertoire often provided Tok Pisin glosses. For example, writing about Gapun, Kulick and Stroud discuss tok bokis as "extended metaphorical speech that contains no overt clues to its real meaning"

---

15.   Mühlhäusler, Dutton and Romaine gloss tok bokis as 'hidden languages' (2003: 90) and tok piksa as 'talking in metaphors' (2003: 91).

16.   www.Mihalicdictionary.org (2007)

17.   Franklin (1972) characterizes Kewa "Pandanus language" as a ritual language. He does not categorize it as tok bokis, but his focus is on systematicity in the vernacular language.

(1990: 300), similar to Strathern's (1975) description of veiled speech. They note, however, that it is not as formalized as the metaphorical speech described by Mc-Kellin (1984) for Managalese speakers. Levine describes a vernacular Kafe verbal style (Kafe *fronka ge*), which he glosses as tok bokis, a "highly developed form of circumlocutory speech" (1982: 76) used to avoid directly confronting parties for whom direct verbal attacks would be provocative. Here as elsewhere, these rhetorical styles are linked to broader social strategies of indirection, and preferred when communicating particular types of information that might be politically and/or socially sensitive or restricted. Issues of accountability and responsibility are often linked to these rhetorical styles central to interaction in social and political life in these face-to-face communities (see Schieffelin 2007, fn. 15).

In various ethnographic accounts, when tok bokis is used as a gloss for a local way of speaking, it foregrounds the idea of multiple meanings that are not shared, but rather socially distributed in purposeful ways, meanings that are hidden from some while known to others, meanings that may pertain to the mundane or the sacred, and involve figurative language of various types. One has to know to look for less obvious meanings, or have insider knowledge about something, or have a set of particular beliefs, or an orientation to a specific cosmology, to understand what is really meant, not just what is said, when one hears something spoken in this particular genre or, speech style. This was the interpretive verbal framework in which tok bokis was introduced to Bosavi speakers.

5.   Tok bokis, tok piksa in Bosavi

During the mid-1970s local pastors in Bosavi encountered tok bokis reading Mark in the 1969 *Nupela Testamen* and translated it as *bale to* 'turned over words' or *bale to siyo*: 'said turned over words'. In Bosavi, *bale to* is a locally recognized category of talk that has a concealed meaning, and consists of lexical substitution as well as a range of expressions and speech styles, though it is not used for narrative or story. Bosavi *bale to siyo:* is derived from the combination of two verbs: *balema* 'turn over' + *sama* 'say', and denotes a literal meaning on the surface of what is said (*wa:l*), with an intended meaning underneath (*ha:g*), which can be understood without being directly said. Thus, using *bale to* as a gloss signals that while everyone (outsiders) can understand the literal sense of what is being said, the real meaning is hidden, has to be searched for, and is found "underneath," and shared between those in the know (insiders). One has to know the relevant contextual cues for sense making, and how to think about it. In order to see how Bosavi pastors made sense out of parable as a genre over time, we examine examples of their translation practices from two time periods, 1975 and 1994. During the first

period pastors used the first edition of Mark (1969), and in the second used the 1989 *Buk Baibel*, which incorporated and extended revisions made in the 1978 edition. By the 1990s, the 1989 *Buk Baibel* was in use in the villages. These two time frames highlight continuities and discontinuities in translation choices and practices, and below I focus on the same speaker, the first local pastor in Bona village who was active during this time.

By early 1975 Bosavi pastors were starting to lead services, reading and translating from the 1969 *Nupela Testamen*. Reading phrases, and parts of verses in Tok Pisin, pastors followed each segment by a translation. Mark 4, v. 1 introduces Jesus teaching to the crowds, translated as <u>skulim ol manmeri</u>. The notion of doing school, or even teaching in this manner was alien to Bosavi culture, and Bosavi pastors translated what Jesus was doing as *malolo:wo: wida:i ane* 'going around showing stories.' This named the genre within a local framework as narrative (*malolo:*). Verse 2 introduces the type of talk, <u>tok bokis</u>. These examples, drawn from my audio recordings of church services, represent the reading in Tok Pisin (TP) and translation practices in Bosavi (BSV) of the pastor. Hesitation is marked by a dash (–), and bolding highlights translation of the metalinguistic terms. What is read is not necessarily the printed text, which, along with the English source text is provided in Appendix A. The English is a close gloss of the Bosavi.

(1)   Mark 4, v 2, <u>Nupela Testamen</u> 1969, read by Pastor D in 1975

TP   <u>na em i givim planti tok long ol</u>
BSV *a:la:ta:ga:yo: eyo: mada malolo: towo: – mada modo: wida:i ane sa:la:*
      'so many stories – he went around really showing many stories'

TP   <u>long **tok bokis** tasol</u>
BSV *e malolo: to wida:yo: – bale – **bale sa:lan** – aum bale sa:la:i ane*
      'he was showing stories – speaking turned over words (parables)– like that he went speaking parables'

Another speaker interrupts to correct the pastor:

BSV *bale wida:i ane*
      'went around preaching parables'

Pastor D:

BSV *bale wida:i ane*
      'went around preaching parables'

TP   <u>em i givim – em i givim tok olsem long ol</u>
BSV *eyo: bale widaki – to wida:iyo: – bale widaki o:leo:ngo: widaki ane*
      'he was preaching parables – showing the word – preaching parables – like that he was preaching'

In these passages (1), we see Pastor D's attempts to translate tok bokis, which are corrected by another member of the congregation training to be a pastor. This correction, incorporated in his translation, highlights the difficulty of selecting an appropriate metalinguistic phrase. The pastor shifts from *bale sa:lan* which uses the verb *sama* 'say' to *bale widan*, which draws on the verb *walama* 'show', the verb adopted for the new speech activity 'preach' (*to walama* 'show the word'). This shift from *bale sama* to *bale walama* changes the pragmatic valence from a phrase that is associated with traditional ways of speaking "turned over words" which obscures meaning, to one which is specific only to Christian usage 'showing what is turned over.'

At the end of reading this section, verse 9, Pastor D translates what Jesus has been doing, using *bale siyo: towo:*, 'talk that was turned over', for speaking in parables, again drawing a close connection to a traditional and recognizable genre, still not incorporating the notion of 'showing' (*walama*) the Word in his translations. Announcing that he is moving directly to verse 13, (without mentioning the omission of verses 10–12), he informs the congregation that the verses that follow will explain Jesus' words. He provides multiple local metalinguistic glosses for what the text is "doing" before reading it: the verses will provide the *mo:* 'the reason for or meaning,' the *ha:g* 'underneath' or what is really meant, the *hede* 'truth', and will 'show the other side' of the meaning (*a:na nodo wido:*). These metalinguistic terms make explicit that the story goes beyond its literal meaning, in terms all familiar to Bosavi people, terms associated with *bale siyo: towo:*

Pastor D turns back to reading verse 13, where tok bokis is used in the title of the section, the explanation of the parable, as seen in example (2) below.

(2)   Mark 4, v 13 *Nupela Testamen* 1969, read by Pastor D in 1975

TP   as bilong **tok bokis** bilong pikinini kaikai
BSV *o:leon:go: siyo: ko:lo:lab – o:leongo: siyo: –* **to bale siyo:** *to mo:wo: ma:no: gelano: we a:la: siyo: ko:lo:lab – ma:no: ko:lo: gelano: we*

'this what it is about – it said like this – the explanation for the parable it said it was about planting this food – about planting this food'

Pastor D translates the genre tok bokis as *to bale siyo:* 'turned over words.' Although the subsequent Tok Pisin passages do not mention tok bokis, he nonetheless adds the Bosavi phrase for it, repeatedly reminding the Bosavi listeners of the genre, providing it as a cue for interpretation. This is accompanied by repeated exhortations to listen, to think about what is being said, all given as encouragement to the congregation to look for the meaning of the seeds.

While *bale to* and *to bale siyo:* are pragmatically appropriate and close semantic translations of tok bokis, in terms of traditional ways of speaking, its use in church services is very unconventional. First, *bale to* would never be used to teach or as a

way to impart new knowledge. Deployed as a style closely associated with Jesus, it requires that listeners reinterpret what *bale to* is as a way of speaking that presumably openly discusses a hidden, underlying meaning. This would never occur in Bosavi since *bale to* was the primary expressive resource for avoiding such direct talk. Furthermore, in the exegesis segment of the sermon, Pastor D tries to explain the underneath, but his explanation of the seeds, the ground, the birds, the thorns, and what the surprising bountiful crop means falls apart. He does not have the cultural or linguistic resources to make sense of the parable in any terms, local or otherwise.

The 1978 edition of the *Nupela Testamen* slowly made its way into Bosavi in the late 1980s. Among many other metalinguistic changes in Mark (Schieffelin 2007), tok bokis was replaced with tok piksa. Finding no corresponding metalinguistic term in Bosavi for tok piksa, pastors used *da:fe sama* 'give examples when talking about or explaining something; give instructions', composed of two verbs, *da:fema* 'measure' and *sama* 'say'. This way of speaking characterizes the way in which a more knowledgeable person transmits information to a less knowledgeable one. In traditional usage, this was almost always about an activity, how to do or make something. Like tok piksa, *da:fe sama* is appropriate for explaining ideas that are evident, transparent, that can be easily seen, like a picture. In many ways it is the opposite of *bale sama*, which presumes a nonliteral or hidden meaning. This new Tok Pisin expression and the perspective it promoted was difficult for Bosavi pastors to work with, as evidenced by the ways in which they treated tok piksa when reading these same passages, the first usage of which was in the new title for verses 1–9, Jesus i tok piksa long man i tromoi pikinini wit long gaden. Example (3) is from services audiotaped in 1994. When Pastor D read the Tok Pisin, he did not even translate tok piksa.

(3)   Mark 4, title, *Buk Baibel* (1989), read by Pastor D in 1994

TP   Jisas i **tok piksa** long man i tromoi pikinini – wit long gaden

BSV   *o:leo:ngo a:la: siyo:lo:b – Ya:suwa:yo: – i wiyo: wit – wit a:no: fowo: dia:sa:ga: egelo: isila:ya fifa:i aneka: – a:la: siyo:lo:b*

        'like this it said – Jesus did – the plant name is wheat – having taken the wheat seed he sprinkled it on the cleared garden – it said like that'

As we see in example (4), when reading verse 2, Pastor D translates tok piksa using *da:fe sama* but paraphrases it three different ways marked by hesitation and revision.

(4)   Mark 4, verse 2 *Buk Baibel* (1989), read by Pastor D in 1994

TP   lain tu na Jisas i mekim planti **tok piksa** long ol bilong skulim ol

BSV   *o: Ya:suwa:yo:* **da:fe sa:laki** *– * **da:fe dimida:i** *ha:naki – towo: mada modo: ko:lo: imo:* **da:fe sa:la, sa:la:i** *ane – a:na wida:i ane*

        'and Jesus was explaining – went around explaining – there was a lot of talk, he went around explaining to them – he was preaching there'

At that same service, reading the title of verses 13–20, he slips back into the earlier usage, as shown in example (5).

(5)   Mark 4, title before verse 13 Buk Baibel 1989, read by Pastor D in 1994

TP   <u>as bilong **tok piksa** bilong pikinini wit</u>

BSV   *o: to a:no: wit fo **bale siyo: to** siyo: a:no: hedeleyo: mo:wo: kiyo: we sa:ma:nigo:leka:*

    'so the words that are turned over talk about wheat, this truth, this reason, this bone, is what I am going to say'

Throughout verses 13–20, Pastor D repeatedly alternates between *bale siyo: to*, the gloss associated with <u>tok bokis</u> and *da:fe siyo:*, that associated with <u>tok piksa</u>, even though the Tok Pisin text only uses the latter form. Pastors and others were convinced that there were meanings beyond the literal, and used a variety of metalinguistic terms to refer to their existence, terms that the missionaries did not know or use. While they continued to use Bosavi terms including *mo:* 'the reason for or meaning,' *ha:g* 'underneath' and *hede* 'truth', there were also interesting innovations, one of which is in example 5 above.

    Pastors added a new term for the nonliteral meaning, a loan translation from the Tok Pisin idiom *bun bilong tok bilong God,* literally 'the bone of God's talk.' In sermons in 1994 pastors referred to *kiyo:* 'bone' in the talk, and when I asked about it, they insisted that it was in the Bible though no one could show me where it was. Not part of Scripture, its source was tracked to a Tok Pisin Bible study guide.[18] Pastors read these guides, and translating this idiom back into Bosavi, created a metalinguistic mixture. To Bosavi people, however, this notion of "bone" made little sense, because in their metalinguistic system, meaning was always "underneath," never inside. In translating and sermonic exposition, pastors nonetheless frequently reinserted Bosavi metalinguistic expressions, recalling the original translation from <u>tok bokis</u>, creating competing and contradictory spatial metaphors of where meaning was located, but without making the meaning itself explicit.

6.   Conclusion

In situations of rapid missionization where linguistic and cultural contacts are intensified, as in the Bosavi situation, it was inevitable that competing language ideologies and metalinguistic frameworks became tangled in translation. Given

---

18.   I thank Joel Robbins, who knows his Tok Pisin Bible study guides, among many other things, for suggesting this source. He notes that he always interpreted <u>bun</u> in the title to mean skeleton, since the text was like an outline (pers. comm). Bosavi pastors, however, incorporated this term into their metalinguistic system as an alternative to *ha:g* 'underneath'.

the goals of missionary-linguists who created the Tok Pisin Bible, the Asia Pacific Christian Mission who used it as a source text, and local Bosavi pastors who had to translate the Word in local terms, there was ample opportunity for semiotic slippage – in particular, the creation and transmission of unintended and novel meanings. This is certainly not unique to Bosavi or Papua New Guinea as Biblical scholarship illustrates a long history of competing theories of translation and contested texts.

While I have focused on specific historical, linguistic and ethnographic details of translation practices for Tok Pisin and Bosavi, it is important to remember that these events co-occurred at a particular historical juncture in Bosavi, as well as in Papua New Guinea. The arrival of the mission in Bosavi in the early 1970s co-incided with the first translation, publication, and subsequent revisions of the *Nupela Testamen*. This was also the same time period when Papua New Guinea attained its independence in 1975. Papua New Guinea not only became an independent nation, but also declared itself to be a Christian country. Thus, missionization and Bible translation were essential to nation building, and Tok Pisin, the national lingua franca, was critical to a national identity.

As there seem to be no materials that discuss the decisions behind the translation revision from tok bokis to tok piksa in the *Nupela Testamen* during this time period, I offer my own thoughts on the possible cultural and pragmatic motivations for the shift and the consequences of this in Bosavi. Framed within a broader sociopolitical perspective, I view the shift as iconic, emblematic of the widespread civilizing project of the Christian missions in Papua New Guinea, and of the national government as well.

As a genre, tok bokis characterizes indigenous ways of speaking and interacting recognizable throughout Papua New Guinea. A term of some historical depth, it refers to a genre or style that is not only premised on subtle forms of indirection and ambiguity necessary for social interaction in face-to-face communities, but is also closely associated with "old" ways and traditional social institutions – cargo cults, cannibalism, male initiation rituals, taboo topics that require special lexicons to even be mentioned, and all manner of secrets and secrecy, even secret languages. Tok bokis could be pragmatically loaded with dark connotations. Speakers using tok bokis could potentially exclude not only the uninitiated and other "outsiders" (often women) from understanding the real meaning of what was said and what was going on, but missionaries and government officers as well could be kept in the dark. This was not a desirable state of affairs in a colonial setting with modern and Christian aspirations.[19]

---

19. Kermode (1979: 24, 47) among others scrutinizing the Parable of the Sower in Mark 4, comments on the historical and etymological closeness of riddle and parable as "dark sayings". One wonders about the relationship of coincidence and irony here as parables, with their varying degrees of transparency and opacity, always have the potential for staying dark.

As a gloss for parable, <u>tok piksa</u>, a more recent linguistic innovation, highlights a particular Christian ideology of meaning, one that foregrounds understanding and revelation. It requires a culturally particular mode for accomplishing shared meaning, true meaning. One person as the expert has the knowledge, which others must obtain. There is a right and wrong interpretation, a right and wrong way of doing something, and these ways are already known by the teacher/expert. Furthermore, as a genre, <u>tok piksa</u> is radically different from <u>tok bokis</u> in terms of how knowledge is framed, conceptualized, interpreted, and used. It takes the speaker's intended meaning as authoritative, non-negotiable, something that the listener must apprehend, something that all listeners should apprehend in the same way. The speaker/expert holds all the cards, which in Mark, is Jesus, on the mission station is the missionary, and in the village, was supposed to be the local pastor.

In Bosavi, everyday social interaction based on traditional patterns of interaction drew on meanings co-constructed between speakers and hearers. They deployed their linguistic and social interactional preferences for polyvalent meanings so as to make the responsibility for a particular meaning less than clear to everyone, leaving the control of interpretation, revelation and meaning up to particular participants. Bosavi communication is based on particular notions of intentionality, which differ from those in Western societies, incorporate ambiguity and indirection, and use both as interactional resources (Schieffelin 1990). These strategies functioned well in a society in which everyone could have an opinion and express it, and where what mattered was action and utterance, not intention. *Bale to* made sense, as did <u>tok bokis</u>.

For Bosavi pastors, concerned with self-refashioning, things did not go so smoothly in translation. Anxious to take on what they saw as a Christian subjectivity, they rejected traditional cultural practices and beliefs. Their translation practices show, however, that it was not quite as easy to give up their vernacular metalinguistics, even when such expressions had difficulty matching or mapping on to Tok Pisin Scripture. We see that their acts of Christian identity making also included importing expressions from Tok Pisin, and developing a new style of speaking, as they claimed social and political authority.

These gaps between Tok Pisin texts, missionaries and local language ideologies and practices, mediated by Tok Pisin, point to some of the ways in which the Word was only partially revealed, and substantiated Bosavi doubts about whether or not they were being told everything. They weren't. They did not even hear all of the verses of Mark. The irony here is that naming parable <u>tok piksa</u>, and translating it as *da:fe sama* 'explanation' in Bosavi contributed to keeping its meanings hidden. People did not really know how to look, or what they should look for. Over several decades of intensive Christianizing activities, those who wanted to become Christian were constantly told to have no secrets, to say what was exactly

on their minds, and to confess, and many simply refused (Schieffelin 2008). They knew that getting the message would confirm insider status, but without Christian knowledge, or understanding of how to get it, many Bosavis remained outside.

Languages and their ideologies are instrumental not only to conversion, but also for understanding why it might not happen, or what it means locally to claim convert status. Close analyses of the translating practices also offer insights into agency and responsibility, locating agents and contexts of change and innovation. Such analyses provide insights into how and when multiple, competing and contradictory conceptualizations about language, language use, and language ideologies give rise to hybrid forms. This provides a productive perspective for investigating how new semiotic regimes, developing in the course of intensive missionization through translation activities, transform interpretive practices, ones that have unintended consequences for understanding the Word. Such changes, I argue, arise out of complex, multidirectional interactions between texts and speakers, literacy and orality, languages and language ideologies, and local and transnational agendas. In such missionizing contexts, the importance of genre as an interpretive frame cannot be underestimated in terms of how semiotic and social meanings are introduced and established. I offer this as a complementary perspective to Sankoff's theoretical orientation to understanding language change in progress. Hopefully this perspective sheds light on how translating practices contribute to understanding processes of cultural and linguistic change that are still taking place in complex contact zones across time and place.

Acknowledgments

References

Aufinger, A. 1948. Secret languages of the small islands near Madang. *South Pacific* 3: 90–95.
Aufinger, A. 1949. Secret languages of the small islands near Madang – II. *South Pacific* 3(5): 113–120.

Brash, E. 1971. <u>Tok pilai</u>, <u>tok piksa</u>, <u>na tok bokis</u>: Imaginative dimensions in Melanesian Pidgin. *Kivung* 4(1): 12–20.

Duranti, A. 1993. Intentions, self and responsibility: An essay in Samoan ethnopragmatics. In *Responsibility and Evidence in Oral Discourse*, J. Hill & J. Irvine (Eds), 24–47. Cambridge: CUP.

Feld, S. [1982]1990. *Sound and Sentiment: Birds, weeping, poetics, and song in Kaluli expression*, 2nd Edn. Philadelphia PA: University of Pennsylvania Press.

Feld, S. 1988. Aesthetics as iconicity of style, or 'lift-up-over-sounding': Getting into the Kaluli groove. *Yearbook of Traditional Music* 20: 74–113.

Franklin, K. 1972. A ritual pandanus language of New Guinea. *Oceania* 43(1): 66–76.

Franklin, K. 1992. Speech act verbs and the words of Jesus. In *Language in Context: Essays for Robert Longacre*, S.J.J. Hwang & W.R. Merrifield (Eds), 241-261. Dallas TX: SIL.

Goffman, E. 1981. Footing. In *Forms of Talk*, 124–59. Philadelphia PA: University of Pennsylvania Press.

Green, J., McKnight, S. & Marshall, I.H. (Eds), 1992. *Dictionary of Jesus and the Gospels*. Downers Grove IL: Intervarsity Press.

Hall, R.A. 1943. *Melanesian Pidgin English: Grammar, texts, vocabulary*. Washington DC: Linguistic Society of America.

Handman, C. 2007. Speaking to the soul: On native language and authenticity in Papua New Guinea Bible translation. In *Consequences of Contact: Language ideologies and sociocultural transformations in Pacific societies*, M. Makihara & B.B. Schieffelin (Eds), 166–188. Oxford: OUP.

Kermode, F. 1979. *The Genesis of Secrecy*. Cambridge MA: Harvard University Press.

Kulick, D. & Stroud, C. 1990. Christianity, cargo and ideas of self: Patterns of literacy in a Papua New Guinean village. *Man (N.S.)* 25: 286–304.

Levine, H. G. 1982. *Tebe kre nentie*: Social learning and behavior therapy among the New Guinea Kafe. *Ethos* 10(1): 66–93.

Lothmann, T. 2006. Considering the notion of adequacy: Notes on the translation of the Bible into Tok Pisin. *Melanesian Journal of Theology* 22(1): 62–86.

McElhanon, K.A. 1975a. Evangelism and the use of Tok Pisin idioms. *Catalyst* 5(3): 49–69.

McElhanon, K.A. (ed). 1975b. *Tok Pisin i go we? Proceedings of a conference held at the University of Papua New Guinea, Port Moresby, PNG 18–21 September 1973* [Kivung special publication 1]. Port Moresby: Linguistic Society of Papua New Guinea.

McKellin, W. 1984. Allegory and inference: Intentional ambiguity in Managalase negotiations. In *Disentangling Conflict: Discourse in Pacific societies*, K. Watson-Gegeo & G. White (Eds), 335–370. Stanford CA: Stanford University Press.

Mead, M. 1931. Talk-Boy. *Asia*, March, 144–151, 191.

Mihalic, F. 1957. *Grammar and Dictionary of Neo-Melanesian*. Techny IL: Mission Press.

Mihalic, F. 1971. *The Jacaranda Dictionary and Grammar of Melanesian Pidgin*. Adelaide: The Jacaranda Press.

Mühlhäusler, P. 1979. *Growth and Structure of the Lexicon of New Guinea Pidgin* [Pacific Linguistics, Series C, NO. 52]. Canberra: Australian National University.

Mühlhäusler, P., Dutton, T.E. & Romaine, S. 2003. *Tok Pisin Texts: From the beginning to the present*. Amsterdam: John Benjamins.

Mundhenk, N. 1990. Linguistic descriptions in the Tok Pisin Bible. In *Melanesian Pidgin and Tok Pisin*, J.W.M. Verhaar (Ed.), 345–373. Amsterdam: John Benjamins.

Murphy, J.J. 1943. *The Book of Pidgin English*, 1st Edn. W.R. Smith and Patterson.

Nida, E. 1964. *Toward a Science of Translating*. Leiden: E.J. Brill.

Nupela Testamen. 1969. Port Moresby: Bible Society of Papua New Guinea.

Nupela Testamen na ol Sam. 1978. Port Moresby: Bible Society of Papua New Guinea.

Nupela Testamen, 1989. In *Buk Baibel*. Port Moresby: Bible Society of Papua New Guinea.

Rule, M. 1977. Institutional framework of language study: The Asia Pacific Christian Mission. In *Language, Culture and Society and the Modern World.* Fascicle 2: *New Guinea area languages and language study,* Vol. 3 [Pacific Linguistics Series C-40], S.A. Wurm, (Ed.), 1341–1344. Canberra: Australian National University.

Sankoff, G. 1980. *The Social Life of Language*. Philadelphia PA: University of Pennsylvania.

Schieffelin, B.B. 1990. *The Give and Take of Everyday Life: Language socialization of kaluli children*. Cambridge: CUP.

Schieffelin, B.B. 2000. Introducing Kaluli literacy: A chronology of influences. In *Regimes of Language*, P. Kroskrity (Ed.), 293–327. Santa Fe NM: School of American Research.

Schieffelin, B.B. 2002. Marking time: The dichotomizing discourse of multiple temporalities. *Current Anthropology* 43: 5–17.

Schieffelin, B.B. 2007. Found in translating. In *Consequences of Contact: Language ideologies and sociocultural transformations in Pacific societies*, M. Makihara & B.B. Schieffelin, (Eds), 140–165. Oxford: OUP.

Schieffelin, B.B. 2008. Speaking only your own mind: Reflections on talk, gossip and intentionality in Bosavi (PNG). *Anthropological Quarterly* 81, 2:431–441

Schieffelin, B.B. & Feld, S. 1998. *Bosavi-English-Tok Pisin Dictionary* [Pacific Linguistics Series C-154]. Canberra: Australian National University.

Schieffelin, E.L. 1976. *The Sorrow of the Lonely and the Burning of the Dancers*. New York NY: St. Martin's Press.

Smith, G.P. 2002. *Growing up with Tok Pisin: Contact, creolization, and change in Papua New Guinea's national language*. London: Battlebridge.

Strathern, A. 1975. Veiled speech in Mt. Hagen. In *Political Language and Oratory in Traditional Society*, M. Bloch (Ed.), 185–204. New York NY: Academic Press.

Woolard, K. & Schieffelin, B.B. 1994. Language ideologies. *Annual Review of Anthropology* 23: 55–82. Palo Alto CA: Annual Reviews.

Appendix A

Mark 4, v. 2
Good News Bible
He used parables to teach many things, saying to them:
   Nupela Testamen 1969
Na em i givim planti tok long ol long **tok bokis**. Em i givim tok olsem long ol:
   Buk Baibel (1989)
Na Jesus i mekim planti **tok piksa** long ol bilong skulim ol. Em i givim tok olsem long ol,
   Mark 4, v. 13
Good News Bible
Then Jesus asked them, "Don't you understand this parable? How, then, will you ever understand any parable?

Nupela Testamen (1969)

<u>Na Jisas i tokim ol, "Yupela i no save long dispela **tok bokis**, a? Na bai yupela i save olsem wanem long olgeta **tok bokis**?</u>

Buk Baibel (1989)

<u>Na Jisas i tokim ol disaipel olsem, "Yupela i no save long as bilong dispela **tok piksa**, a? Olsem bai yupela i save olsem wanem long olgeta arapela **tok piksa**?</u>

Bridging Macro- and Micro-sociolinguistics

Chiac in context

Overview and evaluation of Acadie's *joual*[1]

Ruth King
York University

> Le chiac, mot d'invention assez récente, c'est en réalité du franglais, un
> français corrompu au contact de l'anglais.[2]
>
> <div align="right">Léon Thériault 1975</div>
>
> [Le chiac] est un jeu de codes, une sorte de morse institué au niveau
> d'un peuple.[3]
>
> <div align="right">Dano LeBlanc 2006</div>

I argue that there is little evidence that *chiac*, an often stigmatized variety
of Acadian French spoken in the urban area of Moncton, New Brunswick,
differs dramatically from a number of lesser known Acadian varieties in
terms of the effects of language contact; and that the degree of English
influence claimed is sometimes not supported by the data provided. I begin
with a sociohistorical overview of Acadian French. I then evaluate the
literature on *chiac* and compare it with my own and others' findings
for Acadian varieties spoken in Atlantic Canada. The relationship between
the social context within which *chiac* is spoken and its lexicon and grammar
adds to our knowledge of the linguistic outcomes of language contact, in
addition to providing more detail on variation in North American French.

---

2.   "Chiac, a recently-invented word, is in reality "franglais", a French corrupted through
contact with English." (my translation)

3.   "Chiac is a set of codes, a kind of Morse established at the level of an entire people." (my
translation)

It also illustrates the possibility of a comparative sociolinguistic approach, made possible by variationist methods and theory.

1.   Introduction

In Michel Brault's 1969 documentary *Éloge du chiac*, Moncton, New Brunswick francophone high school students debate the use of a variety labeled *chiac*,[4] which they define as involving traditional Acadian dialect forms and words of English origin, with some aligning themselves passionately with supralocal varieties of French and others, just as passionately, embracing local language use. The film ends with the students leaving the school grounds, with one young woman crying out, "Vive le chiac!" Since first coming under media scrutiny in the 1960s, this variety of Acadian French spoken in close contact with English in the Moncton area has been the subject of intense debate. While Acadian French in general, spoken in Canada's four Atlantic Provinces, has often been stereotyped as *moitié français, moitié anglais* "half French, half English" (Flikeid 1989), *chiac* is the one most often singled out as such by lay people, educators and quite a number of linguists. However, despite a longstanding characterization of *chiac* as "corrupted French", the social situation is more complex in that language attitude studies (e.g., Boudreau 1996, 1998; Keppie 2002) reveal tension between a view of *chiac* as a marker of local identity and *chiac* as an object of derision.[5] And while it is the case that Acadian varieties spoken in close, long-term contact with English typically exhibit extensive codeswitching and lexical borrowing, as I have argued

---

4.   The origin of the label *chiac* is a matter of some debate: the usual story is that it is a deformation of the pronunciation of a neighbouring community, Shédiac. Péronnet (1975), however, has suggested an alternate etymology, citing the existence of the family name Chiaque in southeast New Brunswick, of native origin.

5.   In other work (Comeau & King 2008), I consider the extent to which the scales may be tipping away from negative characterizations of *chiac*, reminiscent of the situation in Quebec in the 1960s and 70s with regards to *joual*. See as well Boudreau (1996: 152) who suggests a possible shift from "*chiac mépris*" ("chiac shame") to "*chiac fierté*" ("chiac pride") along with Gammel and Boudreau (1998) who discuss the role of *chiac* in discourses of resistance in recent Acadian poetry. Finally, Dubois (2003) discusses the (positive) effects of community-based radio on attitudes towards local varieties.

elsewhere (King 2000, 2005) the characterization of the extent of English influence on Acadian varieties tends to be overblown.

While the study of Acadian French in general has seen tremendous growth over the past three decades (cf. Gesner 1986), the study of *chiac* has virtually exploded since the early 1990s.[6] However, much of this literature takes for granted a high degree of English influence, as evidenced by characterizations such as "code mixte" (Gérin 1984), "métissage français/anglais" (Perrot 1995), and "third dialect" (Young 2002). In this chapter I will argue, on the basis of close examination of this literature, that (a) there is little evidence that *chiac* differs dramatically from a number of lesser known Acadian varieties spoken in Nova Scotia and Prince Edward Island in terms of the effects of language contact; and that (b) the degree of English influence claimed is in several cases not supported by the data provided. This is not to suggest that the variety is without interest; rather, the relationship between the particular social context within which it is spoken and *chiac* lexicon and grammar adds to our knowledge of the linguistic outcomes of language contact, in addition to providing more detail on variation in North American French. I begin with a sociohistorical overview of Acadian French in general, and New Brunswick Acadian in particular. I point out the major linguistic characteristics which distinguish Acadian from its better known neighbour, Quebec French. I turn next to the literature on *chiac*, both descriptions of (aspects of) the variety and the empirical support offered for these characterizations. I then compare this literature with my own and others' findings for Acadian varieties spoken in Newfoundland, Prince Edward Island and Nova Scotia and analyze the features regularly identified as *chiac* usage. The chapter ends with a call for a comparative sociolinguistic approach to closely-related language varieties, made possible by variationist methods and theory.

2.   Acadian French: An overview

Acadian French refers to (marginalised) varieties of French spoken principally in the Canadian provinces in New Brunswick, Newfoundland, Nova Scotia, and Prince Edward Island.[7]

---

6.   For instance, of the 26 papers included in the proceedings of a 2004 conference on North American French varieties held at the Université d'Avignon (Brasseur & Falkert 2005), 17 were concerned with Acadian French, and, of these, 12 were concerned wholly or in part with *chiac*.

7.   King (2000) provides a sociohistorical overview and a grammatical sketch of Acadian French.
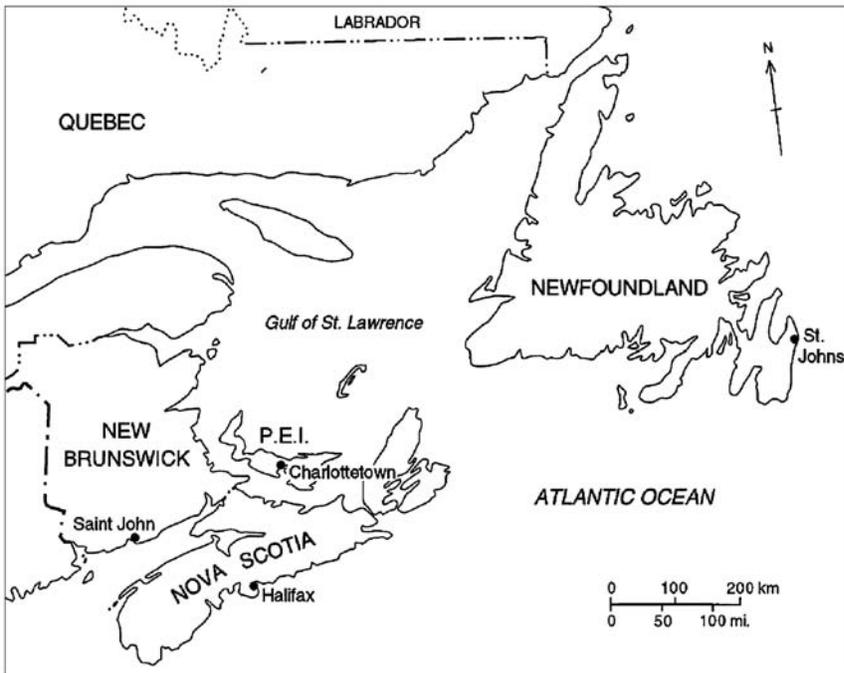
Figure 1.

Acadian differs from its more well-known neighbour, Quebec French, in part due to the different European origins of the colonists, with the majority of Acadian settlers coming from the provinces of the *centre-ouest* of France, whereas Quebec colonists were more diversified, with substantial numbers of settlers from north of the Loire Valley. Another reason for differences between the two varieties, perhaps the most important one, is the relative isolation of Acadians from contact with other francophones and from the normative influences of a French education system, isolation which lasted into the 20th century for some varieties and which continues to the present day for others (Flikeid 1994). For example, Acadian varieties preserve, to varying degrees, first person plural pronominal use (e.g., *je parlons* "we speak") and third person plural verbal morphology (e.g., *ils parlont* "they speak"), the simple past tense (e.g., *il descendit* "he went down"), and the imperfect subjunctive (homophonous with the simple past), all cases of usage which have disappeared from most varieties of spoken French (King 2000; Comeau & King 2007).

In Newfoundland, Nova Scotia and Prince Edward Island, French is spoken by a small proportion of the population, with concentrations of Acadian speakers in isolated regions. According to the 2001 Canadian census just over 2500 residents of Newfoundland and Labrador, just over 6000 residents of Prince Edward Island

and almost 37,000 residents of Nova Scotia spoke French as their mother tongue or as a mother tongue along with English.[8] In contrast, in New Brunswick, French is spoken by a third of the population of 719, 710 (Statistics Canada 2001 census) and has had the status of an official language at the provincial (as well as national) level since 1969.
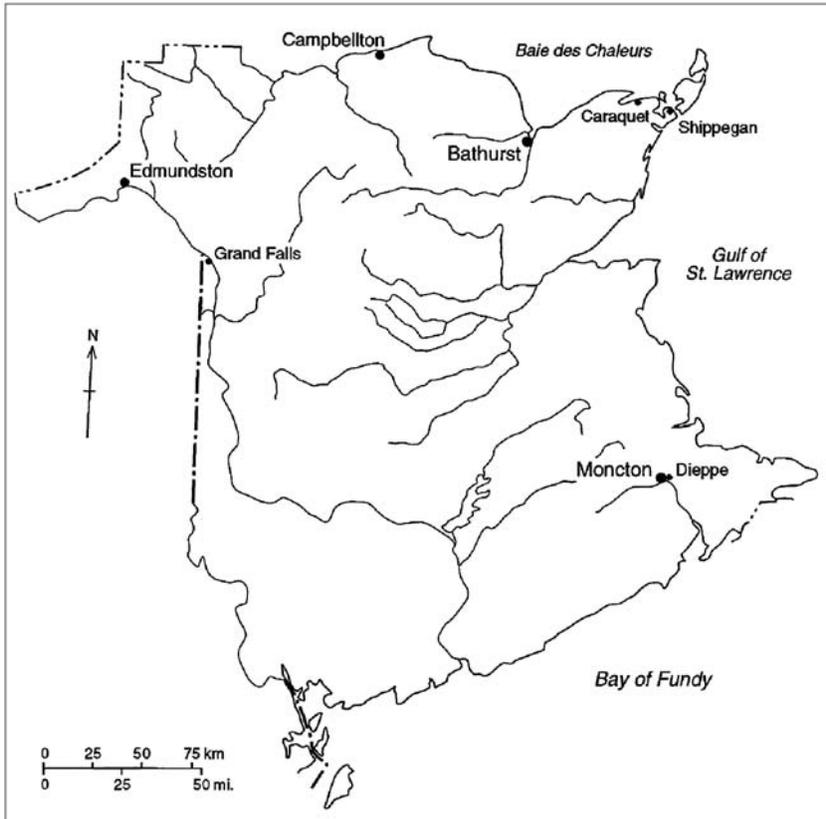
Figure 2.

There are three main dialect areas in New Brunswick: (1) the northwest (known as *le Madawaska*, which comprises Madawaska county and parts of Restigouche and Victoria counties) and includes large towns such as Edmunston; (2) the northeast,

---

8.   Source: www.statcan.ca. In all three provinces Acadian French is spoken in small villages. The Nova Scotia figure is somewhat deceptive as the province has five separate francophone regions, each with a number of small Acadian villages. See Dubois (2005) for detailed analysis of the census data for New Brunswick, Nova Scotia and Prince Edward Island 1981–2001.

which comprises Gloucester county and parts of Northumberland and Restigouche and is usually referred to as *la péninsule acadienne* "the Acadian peninsula", which includes villages popular with tourists such as Caraquet; and the southeast, which comprises Moncton and the surrounding area, along with the towns of Newcastle and Chatham. In Madawaska and particularly on the Acadian peninsula, French is a majority language, indeed in the latter case well over 90% of residents have French as their mother tongue (Beaulieu 1996). The Madawaska variety displays a number of features which may be accounted for in terms of the Quebec origins of a proportion of its settlers or in terms of the present-day dialect contact situation which exists on the Quebec – New Brunswick border (McKillop 1987). As for the Acadian peninsula, it is considered to have been least influenced by English due to its relative isolation and high concentration of francophones. It is in the southeast of the province that French is in a minority position: only a third of Moncton's 115,815 residents were francophone at the time of the 2001 census, a longstanding situation since the same proportion was found in 1976, the time of the first major study of Moncton French (Roy 1979). It is in this minority context that the variety labeled *chiac* emerged.

3.   Acadian French and contact with English: The early research

Relatively little was written until the late 1980s about Acadian French as a contact variety, despite an interest in the description of its linguistic characteristics going back to the turn of the last century.[9] Some Acadian scholars have been noticeably reluctant to work on language contact phenomena or to consider external explanations of language change. This is not surprising, given that minority languages and their speakers are often viewed unfavourably and results of studies of contact-induced linguistic change are easily interpreted negatively. Thus the literature on Acadian French is filled with studies which stress the archaic and conservative nature of the dialect (e.g., Poirier 1993) or which measure the influence of external varieties in recent years (e.g., Flikeid 1984) but until fairly recently treatments of language contact phenomena have been sparse.

   This is not to say that writers have not commented (in passing) on English influence. In Geneviève Massignon's landmark work *Les parlers français d'Acadie*, a (principally) lexical study based on fieldwork[10] conducted in the mid-1940s

---

9.   Early works include Poirier's lexicographical studies and Geddes' treatments of phonetics/phonology. See as well Gesner's (1986) annotated bibliography of Acadian French.

10.   Massignon did fieldwork in the three Maritime provinces (New Brunswick, Nova Scotia and Prince Edward Island) but not in Newfoundland, probably unaware of the Acadian presence there.

but not published until 1962, she justifies her decision to focus on rural speech as follows:

> J'ai laissé de côté, dans mon enquête sur les parlers français de l'Acadie, l'élément francophone «urbain», que j'ai cependant fréquenté pour mes recherches sur les traditions et le folklore acadiens. En effet, les Acadiens des villes parlent un langage beaucoup plus anglicisé que ceux des campagnes, ils forgent à tous moments des calques de l'anglais pour des termes techniques, dont ils ignorent l'équivalent français, et ils n'ont plus le vocabulaire traditionnel inhérent aux choses de la campagne. (p. 88)[11]

Similarly, Vincent Lucci (1972: 15) states that he sought out older, rural informants for his *Phonologie de l'acadien*, rather than urban residents of the Moncton, New Brunswick area, since "on rencontre dans le parler francophone de Moncton tous les degrés *possibles* d'anglicisation. (my emphasis)[12]

   To my knowledge,[13] the first actual study focusing on language contact issues was Marie-Marthe Roy's 1979 M.A. thesis on the use of *but* and *so* in the speech of fourteen working-class Moncton residents, representative of both younger (15–30) and older (60+) age groups and both sexes, based on fifteen hours of taped interviews recorded in 1976. Roy focuses primarily on her consultants' use of *but* and *so* as conjunctions (as in 1) and as discourse markers (as in 2):

(1)   *Y'a pas assez de staff pi ça su le plancher pour le watcher **so** i' l'avont back mis en intensive care*. (p. 116)

   there has not enough of staff and that on the floor for him to-watch so they him have back put in intensive care

   'There weren't enough staff and that on the floor to watch him so they put him back in intensive care.'

---

11.   "I ignored, in my study of the French dialects of Acadia, the urban francophone element, which I did, however, encounter in my research on Acadian folklore and traditions. Indeed, urban Acadians speak a much more anglicized language than those of the country, they often make word-for-word calques of English technical terms, the French equivalent of which they are unaware, and they no longer have the traditional vocabulary relevant to country life." (my translation)

12.   "One finds in Moncton French all possible degrees of anglicization." (my translation) Despite Lucci's stated preference for studying rural speech, there is the occasional comment about (the anglicized nature of) Moncton French scattered throughout the text.

13.   Péronnet (1989) is also of this view. Gesner (1979) is a morphosyntactic study of Baie Sainte-Marie, Nova Scotia Acadian in the Martinet school of functional linguistics which makes occasional reference to English influence, typically in the form of calques, but this is not the central focus of the work.

(2)   *…je veux back aller à l'école parce … toutes mes amis sont là **but** …* (p. 103)
      I want back to-go to the school because all my friends are there but
      '…I want to go back to school because … all my friends are there but …'

Roy found that younger speakers especially tended to use the English-origin vari-
ants (based on their relative frequency she assigns to them the status of borrowings
instead of codeswitches) more than their French-origin equivalents, particularly in
the case of *so* (versus French-origin *ça fait que*). In the case of *but* and its French-
origin equivalents *ben* and *mais*, women tended to be more conservative, some-
what favouring *ben* and *mais*. It should be kept in mind that while Roy extracted
approximately 800 tokens for the *ben – mais –but* variable and approximately 230
tokens for the *ça fait que –so* variable, she reports proportions only, for individual
speakers and for social groupings. No statistical analysis of the data was under-
taken. However, the thesis is rich in detail regarding the variants and syntactic
constraints on their usage. An early chapter also contains a number of important
observations pertaining to other aspects of English influence on the French of her
consultants. For instance, she points to the following examples of English-origin
back: *amener back* 'to bring back' and *je vous dirai pas back* (which she "translates"
as *je vous le dirai pas encore, une fois de plus*) 'I won't tell you again' (p. 65).[14] She
also found a number of other prepositions of English origin in her corpus (she
cites *about*, *on* and *off*) (p. 68) and also gives the following examples of orphan
French prepositions (p. 60).

(3)   *C'est   la   chose   que   je   veux   vous   parler   **de**.*
      it is   the   thing   that   I   want   you   to-speak   of
      'It's the thing I want you to talk about.'

(4)   *Si   que   la   personne   j'ai   adressée … je m'ai   adressé **à** peut*
      if   that   the   person   I have addressed   I REFL have   addressed to can

      pas   me comprendre …
      not   me to-understand

      'If the person I speak to … I speak to cannot understand me …'

While some of the examples are striking indeed, in that other French varieties
investigated to this point did not allow examples like (3) and (4), Roy comes to
the following, careful conclusion regarding the extent of English influence on the
variety:

> Le français de Moncton s'apparente soit au français de d'autres régions, soit
> au français dit standard et que l'influence de l'anglais y est limitée. Le lexique est

---

14.   Consider as well *back's* use preceding the infinitive in (1–2). I will return to the status of
back below.

imprégné de mots anglais mais ces mots anglais seraient utilisés dans des phrases françaises.[15] (p. 77)

Interestingly, while Roy is well aware of the label *chiac*, she chooses to use instead "le français de Moncton" on the grounds that the meaning of the former was a matter of some debate: for some residents, it denoted the speech of the town of Moncton itself, for others, the speech of Moncton and surrounding areas, and for still others, the speech of the entire southeast. She also notes some dispute as to its linguistic characteristics: French mixed with English, French with Acadianisms or French with both English and Acadianisms. She also avoids *chiac* because the label was typically used pejoratively in the community, although she notes the minority opinion of a certain "intellectual elite" embracing the term as signifying the linguistic reality of the city (p. 77).[16]

4.   Chiac as a contact variety: Later research

In a 1989 article, the Acadian linguist Louise Péronnet notes the paucity of research on language contact in the Acadian context, aside from Roy's study. Here she considers borrowings from English in a small corpus of approximately 75,000 words for seven older speakers (aged 65+) from the southeast, long-term residents of majority francophone villages. The study can be seen as providing a baseline for English influence on the French of the southeast. Not surprisingly, Péronnet finds the most frequently occurring words of English origin a particular discourse marker, *well*, several equally predictable nouns such as *job* and *boss* and morphologically-incorporated verbs of English origin such as *watcher* and *driver.* All are longstanding borrowings attested in 19th century Quebec French, according the *Trésor de la langue française au Québec* (http://www.tlfq.ulaval.ca/fichier).[17] *Back* occurs in its canonical English position, as in *A s'en a revenue back* ('She came back') (p. 235) but displays a semantic shift towards "advanced" usage, as in *La vieille, a relisait back* ('The old woman, she read (it) again'), where *back* occurs with the meaning "repeat an action" (p. 234).

---

15.   "Moncton French is related to the French of other regions and to Standard French and the influence of English is limited. The lexicon contains English words but these words are used in French sentences." (my translation)

16.   However, support for the label clearly extended beyond such an elite, given the attitudes of the high school students in Michel Brault's 1969 documentary, noted above.

17.   All are well attested in my 1987 corpus of Prince Edward Island Acadian French: *watcher* and *driver* for instance are in the top three of the most frequently-occurring verbs of English origin.

Such usage was also found by Roy (see the example with *dire* 'to say' above). Péron-net's discussion of syntactic influence from English is limited to calques such as *A se tenait clear de lui.* ('She kept clear of him.' p. 245)

In a 1996 article, Péronnet turns her attention to what she terms new varieties of Acadian French. While keeping in mind the fact that no variety is homoge-neous, she contrasts traditional Acadian French with the French of young consul-tants (aged 12–18) from the southeast recorded in the late 1980s, whose speech she argues to be heavily influenced by English, and the French of other southeast speakers who have adopted supralocal variants (vernacular Quebec French and/or standard French learned at school). The conclusion regarding this latter variety is based in part on local speech heard on the Moncton affiliate of the national radio network, Radio Canada, over a six-month period in 1994 (p. 122).

In this new article, Péronnet contrasts traditional Acadian, which displays the kind of English influence she described in 1989 for older speakers, with the variety spoken by the southeast youth of the late 1980s (she does not mention the *chiac* label, presumably for the same reasons as Roy (1979, who also cites Péronnet 1975 in this regard). Péronnet gives illustrations of what she terms new aspects of lan-guage transfer in the southeast: the use of English *right* as an intensifier (*right out*, *right sharp*), of *kind of* (*kind of peur* 'kind of afraid'), and of *own* (*J'ai ma own car* 'I have my own car').[18] Along with considering overt English influence, Péronnet attributes covert English influence to simplification of the verbal paradigm not found in traditional Acadian varieties, e.g., the conditional of "to do" is formed with the infinitive *faire* rather than with the irregular stem *fer-*, as in, *tu fairais* 'you would do'; similarly with *venir* 'to come', the stem is the regularized *vener-* rather than *viendr-*. However, the link here with English seems to be circumstantial since the argument is that since this kind of simplification is only found in speech which also displays the kind of overt English influence cited above, this simplification must be due to contact with English:

> L'apparition de nouveaux types de simplification, allant au-delà de la variation habituelle et touchant la zone normalement invariable d'une langue … n'apparaît que dans des situations de contact interlinguistique important. Ce genre de transformation linguistique est un indice que la langue concernée est atteinte dans sa structure profonde.[19]

---

18.   In all of the examples but *ma own car* (Péronnet indicated that this type of example actu-ally occurred only once in her corpus), it is not clear to me that these (multi-word) tokens are not better analyzed as codeswitches rather than borrowings. Péronnet does not appear to make a borrowing – codeswitching distinction.

19.   "The appearance of new forms of simplification, going beyond ordinary variation and

Further, Prince Edward Island Acadian also variably exhibits cases of regularized future stems such as in *je voirons* ('we will see'; cf. standard *verrons*) and *ils fini-reront* ('they will finish'; cf. standard *finiront*) (King 2000: 69), a variety which I maintain has not been influenced by English to the extent implied in the above quote. Finally, while Péronnet's article has the merit of attempting to categorize the kinds of French current in New Brunswick in the late 1980s, the discussion of English influence is hampered by lack of evidence of the status of particular usages in the community (a small number of examples are cited), beyond an occasional indication of rarity of particular forms in the corpus.

The data presented as indicative of (extensive) overt English influence by Péronnet are quite similar to those presented earlier by Gérin (1984) in a brief overview in which he makes the case for considering *chiac* a "mixed code," in effect a relexified variety with structural borrowing as well: "les nouvelles généra-tions ont effectué une relexification de leur langue à partir d'emprunts à l'anglais. A ces derniers se sont, en outre, greffés des emprunts morpho-syntaxiques et pho-nologiques." (p. 34)[20] Gérin argues for the following meaning for the term *chiac*:

> Le chiac est la langue socio-maternelle d'une grande part des générations les plus jeunes des francophones du Sud-Est du Nouveau-Brunswick et plus spécialement de la région de Moncton. Sa principale caractéristique est l'abondance d'emprunts faits à l'anglais selon des mécanismes particuliers.[21] (p. 32)

Gérin's outline of its distinguishing features (one or two examples of each are given), based on speech samples for 20-year-olds collected over a twenty-year period, includes the following:

(5) a. Lexical borrowing (note the example given would normally be considered a multi-word codeswitch)

Ex: *Y   sont **all over the place**.*
      they are   all over the place
      'They are all over the place.'

---

touching the normally invariable aspects of a language … appears only in situations of extreme language contact. This type of language change is an indication that the underlying structure of the language in question has been affected." (my translation)

20. "The new generations have brought about a relexification of their language through borrowing from English. To these borrowings are grafted morphosyntactic and phonological borrowings." (my translation)

21. "Chiac is the maternal language of a large proportion of the youngest generations of fran-cophones in southeast New Brunswick, in particular in the Moncton region. Its principal char-acteristic is the abundance of borrowing from English according to particular mechanisms [of transfer]." (my translation)

    b.   Morphologically-incorporated verbs of English origin:

        Ex: *L'hiver, y s'y **restont** pour la grosse été.*
            the winter they REFL there rest-up for the big summer
            'In the winter, they rest up for the big summer [to come].'

    c.   Morphosyntactic borrowing (note the status of "how come" below as a (lexicalized) borrowing or as a codeswitch is unclear; I will return to the status of orphan prepositions, illustrated in the second example, below):

        Ex: ***How come** t'as pas mon p'tit frère dans ton **back seat?***
            how come you have not my little brother in your back seat
            'How come you don't have my little brother in your back seat?'

        Ex: *La première fille que j'ai maké out **avec** c'était une fille du Québec.*
            the first girl that I have make out with it was a girl from Quebec
            'The first girl that I made out with was a girl from Quebec.'

    d.   A system of subordination surviving from earlier stages of the language (in the example below, present-day standard French would have *afin que*)

        Ex: *Flashe moi, **que** j'watch ta fuse.*
            flash me that I watch your fuse
            'Shine (the light) on me so I can see the fuse.'

    e.   Extensive parataxis which "maintient l'anglicisation de la syntaxe à un niveau modéré" ('keeps the anglicization of the syntax to a moderate level'). Note, however, that there is no evidence given in support of this claim.

Gérin concludes by noting that *chiac* was (at the time of writing) a spoken idiom only, due to its short history, concluding its future was threatened by the influence of Quebec-based media and by the prescriptivism of the school system.

    While the 1970s and 1980s saw occasional reference to *chiac* in the literature, the 1990s saw more and more attention paid to this variety. Marie-Eve Perrot's 1995 doctoral dissertation provided the first in-depth study since Roy's late 1970s MA thesis. Although she notes that *chiac* is not reducible to the speech of the young, Perrot concentrates on this age group on the grounds that (she suggests) they exhibit the most "advanced" version of the variety. Thus she studied the speech of high school students aged 16 to 19 years of age in a corpus constructed in 1991. Perrot replaces the traditional interview format with pairs of consultants discussing their responses to a written questionnaire, with no outsider (or adult!) present, an interesting methodological choice. She argues for *chiac* as a system with its own rules, along the lines suggested by Gérin (1984). On examining her data, Perrot argues against the codeswitching-borrowing distinction on the grounds that the degree of interpenetration of English is such that the codeswitching concept is not adequate to express the "réel métissage présentant un remarquable degré de stabilisation" ('actual code-mixing which involves a remarkable stability') (Perrot 1998: 220). Such stability in code-mixing is not argued to be total, as Perrot notes that some

borrowings have no French equivalents in her corpus, other usage seem to her to be instances of (to use Poplack et al.' s 1988 term) nonce borrowings, and still others exhibit alternation with French-language variants. Though recognizing the limitations of her corpus, Perrot maintains that *chiac* is a rapidly changing variety. Since she does not systematically distinguish codeswitches from borrowings, I will discuss multi-word tokens only when they illustrate systematic use of a particular construction type, as in (6f) and in discussion of the borrowing – codeswitching distinction: Examples of code-mixing uncovered by Perrot include the following:

(6)　a.　Advanced usage of English-origin *back*:

　　　Ex: *Je vais　　**back**　watcher ces　funny　　movies.*[22]
　　　　　I　am-going　back　to-watch　these　funny　　movies
　　　　　'I'm going to watch these funny movies again.'

　　b.　Use of English-origin adverbs such as *right* and *probably*:

　　　Ex: *J'　ai　　**right**　aimé ça.*
　　　　　I　have　right　liked that
　　　　　'I really liked that.'

　　　　*Je pourrai　**probably**　aller moi itou.*
　　　　I　can-FUT probably　to-go me too
　　　　'I will probably go, too.'

　　c.　Use of English-origin prepositions:

　　　Ex: *Tu peux parler **about** du stuff qui va on dans ta vie.*
　　　　　you can to-speak about some stuff which goes on in your life
　　　　　'You can talk about stuff that goes on in your life.'

　　d.　Use of English-origin cardinal numbers and non-numeric quantifiers:

　　　Ex: *la　**last**　journée, ta　　**own**　argent, ma　**whole**　été*
　　　　　the　last　day　　your　own　money　my　whole　summer
　　　　　'the last day,' 'your own money', 'my whole summer'

　　　　*Je peux faire　**anything**　que je veux.*
　　　　I　can　to-do　anything　that I　want
　　　　'I can do anything that I want.'

　　e.　Use of English-origin *wh*-words:[23]

　　　Ex: **whoever** *qui travaille à　McDonald's*
　　　　　whoever　that works　at　McDonald's
　　　　　'whoever works at McDonald's'

22.　Perrot notes that the English plural marker is pronounced on *movies*; however, this is unremarkable if the collocation "funny movies" is interpreted as a codeswitch.

23.　Interestingly, Perrot's 1994 article on *chiac* is entitled "Le chiac … ou whatever."

> *whenever* je watch ça
> whenever   I   watch   that
> 'whenever I watch that'

    f.   English matrix clauses with evidential (my classification) verbs appearing with a French complement clause:

    Ex: **I guess** *c'est vraiment beau*.[24]
        I   guess it is really   beautiful
        'I guess it's really beautiful.'

        I hope *que mon père a callé   ma mère.*
        I hope that my father has called my mother
        'I hope that my father called my mother.'

Here we see some similarity with the earlier works (6 a–c, some of the examples in d), albeit more fully discussed in Perrot's dissertation than in book chapters or journal articles. While none of these six features are unknown in other Acadian varieties, several are indeed striking: I will return to their characterization below, where I will argue for a more moderate view of English influence than Perrot's or Gérin's.

    The next major study of *chiac* came in 2002 in the form of a Ph.D. dissertation by Hillary Young, who defines *chiac* as follows: "chiac is a dialect of A[cadian] F[rench] spoken by Moncton teenagers" (p. 9). Young sees *chiac* as one end of a continuum "where there is a relatively high degree of anglicization and codeswitching and a relatively low use of archaic constructions. This correlates with an urban setting, and with speakers who are young and proficient in English. At the other end of the continuum is the speech of older, less educated, more rural speakers (what is typically meant by A[cadian] F[rench])" (p. 10). In a footnote, Young notes that little research has actually been done on the speech of speakers 30–60 but, on the basis of her consultants' discussions with the interviewer, she concludes (contra Perrot) that "chiac is clearly the speech of the young" (p. 10).[25]

    Young's study, set within the framework of Langacker's model of cognitive grammar (where again no distinction is made between borrowings and codeswitches; all are argued to be produced by the same "cognitive system"), attempts to account for the emergence of the variety in the particular social setting and to lay out features of its lexicon, noun phrase and verb phrase. The corpus was constructed over the summer of 2000, with a total of about 53,000 words for 29

---

24.   Although this example displays absence of the *que* complementizer, it is not unexpected since a preceding fricative favours *que* absence in vernacular French (Sankoff et al. 1971, Sankoff 1980; Martineau 1988; Dion 2003; King & Nadasdi 2006).

25.   Among the other writers who share this view are Hamers and Leblanc (1989) and Thompson (1986).

speakers recorded in groups of two or four provided with a set of discussion topics, thus using a methodology that resembled Perrot's. Speakers were largely from the 14–16 age group. With an almost ten-year time gap between the collection of Perrot's data and her own, Young sought to address the question of linguistic change from one decade to the next.

Young's survey centers on phenomena including the following:

(7) a . Use of intensifier **right**:

Ex: *Moi je demande comme des fois   là   comme pas **right** souvent*
me I ask   like   some times there like   not right often
'I ask, like, sometimes, like, not very often'

b. Use of English-origin prepositions:

Ex: *Y'a   rien   que tu   peux dire   **about** it*
there-is nothing that you   can   to-say about it
'There's nothing you can say about it.'

c. *But* and *so* are far more frequent than their French-origin equivalents.

Ex: *Tu étudies **but** comme tu   viens   juste blank des fois.*
you study but like   you become just   blank some times
'You study but, like, you just go blank sometimes.'

d. English plural marking on nouns of English origin:[26]

Ex: *C'est faite   pour les **concerts** [kansərts] te faire mal comme*
it is done   for the   concerts   you REFL   to-do harm like
'It's what concerts are for – like, hurting yourself'

e. Use of English-origin cardinal numbers and non-numeric quantifiers:

Ex: *J' ai   ma **one** bonne spike.*
I   have my one good   spike
'I have my one good spike.'

f. Use of English-origin *back*:

Ex: *I   m'a **back** frappé.*
he me has   back hit
'He hit me back/again.'

g. English-origin verb + particle constructions:

Ex: *Ça m'a   totally **turné** **off** la   dope.*
it   me has totally   turned off the dope
'It totally turned me off dope.'

---

26.   Young notes the noun *concert* occurred 228 times in her corpus with zero suffixation and 229 times with the English –*s* suffix. According to Young's data the full range of English allomorphy occurs with English suffixation.

While Perrot's and Young's studies provide a wealth of data, they disappoint to some extent in that there is little in the way of quantification. To her credit, Young does often provide proportions of English-origin versus French-origin variants (7c) but suggests that the variety exhibits a significant amount of "free variation" (p. 73). In other words, we are left with little indication of the status of English-origin material in the speech of the individual, the community or in the grammar of the host language. The fact that more than one writer elects to discuss the same feature might be argued to alleviate this problem (e.g., if several researchers have the same finding (e.g., "advanced" *back* usage), but only marginally so. What we need are to at least see proportions of use of an English-origin variant against French-origin ones <u>and</u> constraints on their usage, such as Roy provided in her 1979 thesis. In fact, the most concrete evidence we have for any kind of change in *chiac* usage across the decades are Roy's or Péronnet's explicit statements that a particular usage does not exist or occurs rarely and that same usage turning up in later work without any disclaimer; I will point to a number of such cases below. I will also take up the most commonly-occurring features mentioned in the *chiac* literature and compare them to what has been found in other Acadian varieties. First, though, I will complete the actual literature review by summarizing a representative set of recent articles and book chapters which have appeared over the last decade dedicated to particular aspects of *chiac* language use.

The 2005 conference proceedings edited by Patrice Brasseur and Anita Falkert contain a number of articles on aspects of *chiac*, as noted earlier. These articles cover an array of topics, some of which I mentioned above: orphan prepositions (Arrighi), English-origin verb + particle constructions (Chevalier and Long), English-origin intensifiers (Chevalier and Hudson), use of traditional Acadian variants (Pavel), English-origin discourse markers (Petraş) and English-origin swear words (Kasparian). I will summarize those most relevant for the present purposes.[27]

Maria Pavel analyzes language use by the Acadian writer Régis Brun in his novel *La Mariecomo* and cites examples of traditional Acadian usage which has counterparts in the *centre-ouest* of present-day France (she cites e.g., Charpentier 2000 in this regard). The features discussed by Pavel include the traditional morphology of the verb (*je … ons, ils … ont*) mentioned above and the use of the dialectal *–eux,*

---

27.   Thus, for example, I will not consider in detail Kasparian's study of Acadian swear words. Kasparian's study finds, unsurprising for those familiar with cursing in Acadian French, that the swear words used are largely of English origin. It is worth mentioning, however, that, unlike the bulk of research surveyed here, Kasparian's study includes a younger (20–25) and older (45–60) age cohort. The religion-based patterns of swearing so strongly identified with Quebec French (cf. Vincent 1982) are not found.

equivalent to standard French *–eur*, as in *conteux* 'storyteller'. It is certainly worth keeping in mind that *chiac* involves use of traditional dialect features and is not just a mix of some school variety of French with English, but since use of English material figures in all the definitions of *chiac* of which I am aware, it would have been useful had Pavel given examples of usage reflecting that definition. The only evidence provided that one of the varieties Brun presents is in fact *chiac* (the action of the novel takes place in a rural setting) is a Chiac-French lexicon at the end of the book. The topic of "*une grammaire du chiac* " is not taken up by Pavel beyond her saying that this task is best left to linguists working with oral corpora. Thus while Pavel is undoubtedly correct about the use of traditional features by *chiac* speakers (Gérin 1984 also drew attention to this fact), her case would have been made stronger by analyzing material of the English-origin type as well.

By way of comparison, consider another technologically-mediated representation of *chiac* language use, that of the *Acadieman* TV series and comic books.[28] In the animated TV show and comic strips featuring *Acadieman*, the first Acadian superhero (Leblanc 2006, 2007), regional language use and English-origin material are fluidly integrated.[29] This is shown below in an excerpt from the first number (April 2007) of the *Acadieman* comic book. In this extract, one of the characters, Coquille, explains to his new friend, Acadieman, the ease of landing a job in a call center and the job's advantages:

> Moi, ej travaille dans **le call center of the universe**. Zeux, ils changeant tout l'temps d'monde. Ils charchont tout l'temps pour tchequ'un. Tu poudras subventionner ton **superhero stuff** avec c't argent là … pis payer la **rent** itou.[30]

Here the English-origin material is in bold. Traditional Acadian features at the phonological level include the opening of /ɛ/ before /r/, palatalization of /k/ before non-low front vowels, and traditional pronunciations of *je* (*ej*) and *eux* (*zeux*). At the morphological level, both third person plural verbs are conjugated with the *-ont* traditional ending (versus standard French *changent* and *cherchent*, where the

---

28.   Of course, it must be kept in mind that such stylized performances tend to focus on a limited set of features to represent the variety in question (cf. Johnstone 1999; Coupland 2007).

29.   Interestingly, as a superhero Acadieman has no special powers beyond the nerve to speak chiac in public (Inteview with Dano LeBlanc, CBC Newsworld, Dec 6, 2006).

30.   "Me, I work in the call center of the universe. Them, they change personnel all the time. They are always looking for someone. You will be able to support your superhero stuff with that money … and pay the rent, too." (my translation) Note that New Brunswick is often referred to as the call center capital of Canada, in part due to the efforts of former premier Frank McKenna to boost the economy.

plural morpheme is phonetically null). Traditional lexicon is represented in this extract by the archaic *itou*.

Gisèle Chevalier and Michael Long's article in the Brasseur and Falkert volume is based on data from three corpora for the southeast (Anna-Malenfant 1994; Parkton 1994 & *chiac*-Kasparian 1999), all involving adolescent speakers. Chevalier and Long find in these corpora seven English-origin particles:[31] *back*, *out*, *up*, *off*, *on*, *in* and *around*. They leave aside the well-known borrowing *back* (referring the reader to Perrot 1995 for information) and turn their attention to the remaining six, illustrating their use in combination with English-origin and French-origin verbs. They make the claim that these particles are always placed to the immediate right of the verb with which they co-occur in combination (they cite one counterexample from Flikeid (1989) from her Pubnico, Nova Scotia corpus, noting that this is the most heavily anglicized Acadian region of Nova Scotia.)[32]

Chevalier and Long argue that the borrowing of these particles is aided by the fact that they appear in the same position in the clause occupied by the "strong" forms of the French-origin prepositions *sur* 'on' (*dessous*), *sous* 'under' (*en-dessous*) and *dedans* 'in' (*en-dedans*) and by the fact that the intense language contact situation which obtains in southeast New Brunswick is one which is hospitable to the semantic value the English-origin verb + particle combinations offer. I return to such data below, which is also discussed in some detail by Perrot and Young, as noted above.

In another article in the same volume, Chevalier and Chantal Hudson turn their attention to English-origin *right* (also noted by Perrot 1995; Péronnet 1996; and Young 2002). In the Anna-Malenfant corpus for Dieppe, which neighbours Moncton proper, twelve 13–14 year old students were recorded in 1999, with 80 *right* occurrences reported out of a total of 14,800 words. Chevalier and Hudson show modification of an adjective and an adverb are both possible:

(8)   *C'est* **right** *beau.*
      it is   right   pretty
      'It's really pretty.'

      Il   aime **right** *rien.*
      he likes right   nothing
      'He likes absolutely nothing.'

---

31.   I analyze these as prepositions in English (King & Roberge 1990; King 2000), as shown below.

32.   Note Flikeid's example is as follows: *Ils allont out, zeux chuckont leurs … leurs scallops out* 'They go out, they chuck their, their scallops out.' Such usage is also found in Newfoundland and Prince Edward Island Acadian varieties (King 2000).

They suggest that *right* indexes youth identity but argue on the basis of its frequency that it is here to stay (p. 290).

The use of English-origin discourse markers has been a topic of discussion since Roy's 1979 thesis. In Christina Petraş' contribution she considers a range of such markers: *well*, *of course*, *by the way* and *anyway*. The data were taken from an online Acadian discussion forum on the website www.*capacadie.com* (a wide-ranging site offering information on news, arts and culture, economic issues, etc.) over a period of several months in 2003–04. Many of the online contributors were Acadian expatriates, including former and present-day residents of Moncton. The author attempts to determine the function of these various markers, following in the tradition established by Gumperz (1982). Given the diversity of the contributors to these discussion forums, Petraş necessarily appeals to the level of the individual for some of her analysis. However, her analysis of the use of English *well* is in line with that previously proposed by Chevalier (2002) for *chiac*, that of marker of hesitation.

Finally, Laurence Arrighi constructed her own Acadian French corpus in the Maritime Provinces (thus New Brunswick, Nova Scotia and Prince Edward Island) in 2002–03 as well as availing herself of older recordings from more than a decade earlier (1980–1990) housed in the *Centre d'études acadiennes* of the Université de Moncton. While she does not make geographical observations about her data, it is reasonable to assume that some come from *chiac* speakers. Arrighi makes the important point that structures such as exemplified by *beaucoup de personnes que je travaille avec* ('a lot of people I work with') have a long history in the language (see earlier commentary by Bouchard (1982); Vinet (1984) and by Roy (1979) herself). However, Arrighi also suggests that the fact that such orphan preposition usage parallels usage in the contact language, English, may bolster this kind of usage of French orphan prepositions (Auger 2005: 60 makes a similar point regarding Quebec French usage). As Roberge (1998) notes, the set of prepositions which take phonetically null complements is wider in Acadian varieties than in Quebec or Ontario French, since it includes *à* and *de* (see 3–4 above). This difference Roberge takes to supply evidence for lexical diffusion, by which prepositions are added to the set on a one-by-one basis.

5.   Chiac as a contact variety: An evaluation

As we have seen, then, quite a number of scholars have been struck by the "unique" nature of this variety and by the extent of English influence. It is also the case that since the 1960s researchers have suggested that *chiac* is the language of youth (although it should be noted that Perrot's stance is more nuanced). However, the

chiac-speaking teenagers of Brault's 1969 documentary are now well into middle-age: as Sankoff and Blondeau (2007) have demonstrated convincingly, change across the lifespan for particularly linguistic variables is certainly possible, but only in a minority of cases, so we would not want to argue that entire generations have abandoned wholesale the vernacular of their youth. If there is one type of needed research on Acadian French, it would be tapping usage of this middle-aged Moncton cohort.

Before turning to my own analysis of English influence on *chiac*, I should make clear a number of assumptions. I would assume, as I expect most linguists would assume, that multilingual speakers have multiple mental grammars. With respect to mixed language (I intend "mixed" as a neutral term here) discourse, I assume that such discourse is the product of two or more grammars (cf. Poplack 1980, for an early example of work on possible constraints on codeswitching and Muysken 2000 for a recent synthesis). There is a long tradition of assuming that both grammars are drawn upon in multi-word codeswitches, although some scholars, notably Myers-Scotton (1993), have argued against such a borrowing-codeswitching distinction for single-word tokens. How one evaluates such lone tokens is crucial, however, since as Poplack and Meechan 1998: 127 (cited by Sankoff 2002: 650) note: "[i]n virtually all bilingual corpora empirically studied, mixed discourse is overwhelmingly constituted of lone elements, usually major-class content words, of one language embedded in the syntax of another." As Sankoff notes, Poplack and Meechan made an important methodological breakthrough in their 1998 article by comparing the behaviour of single-word tokens according to several diagnostics to that of unproblematic tokens from the same corpus (i.e., to unambiguous multi-word codeswitches, unmixed English language material, unambiguous French language material). The method has been successfully tested for several pairs of genetically-unrelated languages involved in mixing such as Wolof/French and Fongbe/French (Meechan & Poplack 1995).

This method can profitably be applied to languages which are genetically quite similar, as shown by Danielle Turpin's application to Acadian data (Turpin 1998). Turpin's data come from interviews with eight Moncton consultants. She does not identify the variety by the label *chiac* but the following extract (using Turpin's orthographic conventions), shows we are dealing with the same sort of phenomena as discussed by the other researchers:

> Ça a fait un façon de [noise], j'ai hallé la CLUTCH, PULLé OVER, j'ai dit «WHATEVER.» L'engin RUNnait encore, MAJOR PROBLEM. J'ai dit «j'éteins ma BIKE, je PARK.» L'autre, «DRIVE-moi chez nous, je vas quéri mon TRUCK.» Je savais qu'elle DRIVait plus, je savais j'avais BLOWé ma transmission. [Puis il l'avait BLOWé!] J'emmène ça chez nous, dans ma SHED, je 'garde en-dessous, il y avait un trou dans le BLOCK, le WHOLE SECOND GEAR avait tout BLOWé, SHATTERé. Pis là j'avais

mon autre BIKE SO je DRIVais ça en attendant, là je l'ai fait- mon CHUM l'a arrangé là, CHEAP, sept-cents piasses la WHOLE THING. (11/1.281) (p. 221–2)[33]

Turpin extracted from her corpus some 604 lone English-origin nouns, un-surprising since surveys (Flikeid 1989; Poplack, Sankoff & Miller 1988; King 2000) show that tokens for this grammatical category are the most frequently-occurring lone English-origin tokens in French discourse. Further, single nouns occur with a wide range of frequencies across data sets. I will exemplify Turpin's work by choosing a kind of variation mentioned by both Perrot and Young, i.e., use of English plural suffixation. Of the 143 lone English-origin nouns with plural refer-ence in Turpin's corpus, some 42% appear with –*s* suffixation.[34] Turpin found first of all that the appearance of English –*s* suffixation or French zero suffixation[35] is linked to frequency: more frequently-occurring English-origin nouns tended to have zero suffixation (which suggests that they are morphologically-integrated borrowings). Interestingly, the same correlation was found in Poplack, Sankoff and Miller's 1988 study of English loanwords in the Ottawa-Hull sociolinguistic corpus, although it must be noted that the overall proportion of English suffix-ation is much lower than in Turpin's Moncton corpus. Further, some nouns re-ceived categorical English –*s* suffixation (e.g., *friends* categorically occurred with English plural marking). Finally, choice of plural marking varied tremendously from speaker to speaker: for example, one of the eight speakers had 48% English plural marking while the group mean was only 10%. Since the speakers with the most English plural marking also had the greatest proportion of unambiguous

---

33. It made a kind of noise, I hauled in the clutch, pulled over, I said, "Whatever." The engine was still running, major problem. I said, "I'm switching off my bike, I'm parking." The other guy says, "Drive me home, I'm going to get my truck." I knew that it wouldn't go, I knew I had blown the transmission. [And did he ever blow it!] I take it home, put it in my shed, I look underneath, and there was a hole in the block. The whole second gear was completely blown, shattered. And I had my other bike there, so I drove that in the meantime. Then, I had it – my friend fixed it, cheap, seven hundred bucks, the whole thing. (Turpin's translation)

34. Interestingly, Péronnet's 1989 older-speaker study found no evidence of such suffixation and both Flikeid's (1989) and King's (2000) studies report it to be rare in their corpora. Philip Comeau (p.c.) estimates such usage is sporadic among young people's French in Baie Sainte-Marie, Nova Scotia Acadian but notes lack of systematic study.

35. For the sake of brevity I omit the discussion of the intricacies of French plural marking which figured in Turpin's ultimate determination of the status of particular lone nouns. My aim here is to show that there <u>are</u> methods available for deciding the status of such tokens. Since other researchers have taken up this case of variation in *chiac* I exemplify the method with this variable but note that for other cases of variation – patterns of determiner absence and of adjec-tive placement – Turpin obtains even "cleaner" results.

multi-word codeswitches, Turpin suggest that "at least some of the –*s* marked lone English-origin nouns are codeswitches" (p. 231).

However, since I do not have access to any of the corpora mentioned above, I will necessarily rely on patterns of reporting in the literature and with comparison to what obtains in my own Acadian corpora rather than applying the Meechan-Poplack method. I will suggest a possible trajectory for particular borrowings, from codeswitch to nonce borrowing to part of the lexicon of the host language. I will be particularly concerned with what items may be more easily borrowed, with incorporation of borrowed lexicon and embedding in the grammar and with any structural changes such borrowing might be argued to trigger in the host language. I have organized the phenomena presented by more than one author into the following categories:

5.1   Semantic and syntactic reanalysis of words of English origin

Commentators since Geneviève Massignon (whose fieldwork, the reader will recall, was conducted in the 1940s) have been fascinated by the use of English-origin *back* in Acadian French, as the following quote shows:

> On rencontre aussi, chez les éléments les plus humbles de ces régions [acadiennes], l'usure des procédés français de dérivation, tels que la substitution, au préfixe re-, de la préposition anglaise **back**, dans les expressions comme: **il est venu back**, signifiant «il est revenu», et **vous me le donnerez back** «vous me le rendrez».[36] (Massignon 1962: 751)

However, discussion is not limited to the Acadian context. For instance, in his 1955 Ph.D. dissertation on the French of Windsor, Ontario, Alexander Hull gives this example of *back* usage:

(9) *J'ai   jeté   mes roches **back**.*
     I have thrown my rocks back
     'I threw my rocks back.'

Hull's dissertation is based on interviews with four principal informants and some of their family members. The example above comes from the son of a 46-year-old informant who, judging from the context (the clause is contained in a narrative of personal experience), appears to have been a child at the time. Later, Canale et al.

---

36.   One also finds, in the speech of the most humble people of the Acadian areas, the use of French derivational processes, such as the substituation of the English preposition *back* for the *re-* prefix, in expressions such as *il est venu back*, meaning "he came back" and *vous me le donnerez back* "you will give it back to me." (my translation)

(1977) provide the following data from the speech of informants living in the town of Rayside, Ontario:

(10)   *J'ai   l'intention   de revenir   **back** à   Miami.*
        I have the intention of to-come-back back to Miami
        'I plan to come back to Miami.'

(11)   *I'   m'ont   donné mon argent **back**.*
        they me have given my   money back
        'They gave me my money back.'

(12)   *Là, je mettais la   roue   **back** ensemble.*
        there I   put-IMP the wheel back together
        'There, I put the wheel back together.'

In all of these examples, *back* has the same meaning as it does in English and occupies the same syntactic position as it does in the English glosses. As Massignon noted for Acadian, *back* takes on the role of the French prefix *re-* with verbs such as *revenir* 'to come back'. In the Ontario corpora it is also used in calques such as *back ensemble*.

The early *chiac* literature (e.g., Roy 1979) shows that *back* is used with the meaning "return to a former place or state" and is restricted to the position following the main verb or infinitive. However, in the *chiac* data given we also find more "advanced" usage: *back* may also mean "to repeat an action or process," as in Roy's *je vous dirai pas back* ('I won't tell you again'). In terms of syntactic position, "advanced" *back* may precede the past participle, as in (13), taken from Young (2002), or the infinitive, as in (6a), taken from Perrot (1995) and repeated here as (14):

(13)   *Il   m'a   **back** frappé.*
        he   me has back hit
        'He hit me again.'

(14)   *Je vais   **back**   watcher ces   funny movies.*[37]
        I   am-going back to-watch   these funny movies
        'I'm going to watch these funny movies again.'

Raymond Mougeon and his colleagues (1980) rightly link the emergence of *back* in varieties of French in contact with English to the gradual loss in meaning of the French prefix *re-*, a process taking place over the course of

---

37.   Perrot notes that the English plural marker is pronounced on *movies*; however, this is unremarkable if the collocation "funny movies" is interpreted as a codeswitch. See Turpin (1998), discussed above.

several centuries. They note that while in Old French *re-* had several meanings (e.g., the *re-* of *regarder* "to look at" was originally an intensifier), there remain only two productive meanings in modern French, these being "return to a former state or place" and "repeat an action or process". As Mougeon et al. note, evidence that *re-* continues to undergo this process is the use of *rentrer* (Standard French 'to return home') for *entrer* 'to enter' and *rouvrir* (Standard French 'reopen') for *ouvrir* 'to open' in colloquial French . We can view the emergence of *revenir back*, *retourner back*, etc. in contact varieties of French as a consequence of the loss of meaning of the *re-* prefix. Canale, Mougeon et al. (1977) note that it is only with the meaning "return to a former state or place" that *re-* co-occurs with or is replaced by *back* in the high-contact Ontario French communities they studied. In the urban context of Ottawa-Hull located on the Ontario/Quebec border, we find limited use of *back*: an examination of the Ottawa-Hull French corpus collected under the direction of Shana Poplack reveals *back* use in the speech of 21 of 120 informants, with most of the thirty-nine tokens coming from the three Ottawa neighbourhoods sampled, i.e., from neighbourhoods with relatively high proportions of English residents. Only two tokens came from residents of Hull, located in Quebec, and these came from Vieux-Hull, a working class neighbourhood.[38] *Back* is used as it is in the other Ontario localities mentioned above: it means "return to a former place or state" and it occurs immediately following the verb phrase. A clue as to its status in Ottawa-Hull French is given by one consultant who cites *back* usage as an example of anglicized French:

> C'est un patois, par exemple. Il y en a qui disaient «je reviens back». Tu sais? Je riais, je riais à toutes les fois qu'ils disaient ça, A place de dire m - a revenir dans une minute-là, je reviens back.[39] (Inft 034)

What of other varieties of Acadian French? In 1987, I constructed sociolinguistic corpora for two small Prince Edward Island villages, one, Abram-Village spoken in a majority French area of the province and the other, Tignish, located in an area surrounded by English village where the French presence is threatened,

---

38.   I thank Shana Poplack for granting me access to this corpus.

39.   It's a *patois*, for example. There are some who say "je reviens back". You know? I used to laugh, I used to laugh every time they said that. Instead of saying "revenir dans une minute", "je reviens back". (my translation)

for a total of 630,000 words.[40] King (2000) provides an analysis of *back* usage, exemplified for Prince Edward Island French by (15–16):

(15) *Puis je voulais   pas **back** aller.*
      and   I   want-IMP not   back to-go
      'And I didn't want to go back.'

(16) *Je l'avais assez   haï   que je l'ai   jamais **back** fait.*
      I   it had enough hated that I   it have never   back   done
      'I hated it so much that I have never done it again.'

In my analysis I compared quantitatively the distribution of alternatives available to render the two meanings. For Abram-Village, Table 1 records in row 1 frequencies for *back* meaning "return to a former state or place" versus all other possibilities, i.e., RE + VERB + *back* (as in *revenir back*), RE + VERB (as in *revenir*) or the bare verb without either *re-* or *back* (as in *venir* used with the meaning "to come back"). Row 2 compares the frequency of *back* meaning "repeat an action or process" versus all other possibilities, i.e., VERB + *encore* (as in *faire encore*), use of RE + VERB (as in *refaire*), use of RE + VERB + *de nouveau* (as in *refaire de nouveau*) or VERB + *de nouveau* (as in *faire de nouveau*). Table 2 presents the same information for Saint-Louis, where even fewer alternatives to VERB + *back* were found.

Table 1. Frequency of *back* versus other variants in the Abram-Village corpus. In the first row *back* = "return to a former state or place"; in the second row *back* = "repeat an action or process"

Verb + back (e.g., venir back)	Verb (e.g., venir)	Re + Verb (e.g., revenir)	Re + Verb + back (e.g., revenir back)	
121	15	13	2	
Verb + back	**Verb + de nouveau**	**Verb + encore**	**Re + verb**	**Re + verb + de nouveau**
(e.g., faire back)	e.g., (faire de nouveau)	(e.g., faire encore)	(e.g., refaire)	(e.g., refaire de nouveau)
31	11	5	3	1

Table 2. Frequency of *back* versus other variants in the Saint-Louis corpus. In the first row *back* = "return to a former state or place"; in the second row *back* = "repeat an action or process"

Verb + back (e.g., venir back)	Verb (e.g., venir)	Re + Verb (e.g., revenir)	Re + Verb + back (e.g., revenir back)	
123	1			
Verb + back	**Verb + de nouveau**	**Verb + encore**	**Re + verb**	**Re + verb + de nouveau**
(e.g., faire back)	e.g., (faire de nouveau)	(e.g., faire encore)	(e.g., refaire)	(e.g., refaire de nouveau)
32		6		

We see, then, that *back* is the most common way of expressing both "return to a former state or place" *and* "repeat an action or process" in both Prince Edward Island communities.

Other evidence that *back* is in widespread use in Prince Edward Island French is its use in the speech of the interviewers, those in-group residents who were trained in sociolinguistic methodology and instructed to speak as they would at home during the interviews. They contributed an additional thirty-eight tokens with *back*, examples of which are given in (17–18):

(17) *Ça devait être   de la misère pour eux quand   qu'ils avont **back** venu.*
       it   must-IMP to-be of the misery for them when that they have back came
       'That must have been hard on them when they came back.'

(18) *Veux-tu **back** me conter ça?*
       want you back   me to-tell that
       'Do you want to tell me that again?'

The degree of integration of *back* varies according to the degree of intensity of contact with English. In the Newfoundland Acadian French varieties, in which there has been less intense contact with English than in the Moncton or Prince Edward Island cases, and with supralocal varieties of French (King & Butler 2005), the data resemble the Ontario data provided by Mougeon and his colleagues and found in the Ottawa-Hull corpus (King 2000). Karin Flikeid (p.c.) reports that *back* usage in Nova Scotia Acadian French mirrors that of Prince Edward Island, a not unexpected result given the long history of French-English contact in that province. Indeed, in the Baie Sainte-Marie area of Nova Scotia, Comeau (2007: 43–45) reports *back* usage in line with that found in Moncton and in Prince Edward Island:

(19) *Je vais   **back** l'amener.*
       I   am-going back   him to-bring
       'I'll bring him back.'

(20) *Il   a   **back** amené   la   tape.*
he has back brought the tape
'He brought back the tape.'

(21) *Il   en   a   bu   **back** l'autre   weekend.*
he some has drank back the other weekend
'He drank again the other weekend.'

The Butler Sociolinguistic Corpus for Baie Sainte-Marie, Nova Scotia (the source of Comeau's data)[41] also provides examples such as (22), with the morphologically-integrated verb *backer* derived from English "to back up (e.g., one's car)" co-occurring with the *back* usage I have described thus far:

(22) *Puis il   a   **back** backé   à ièlle.*
and he has back backed-up to her
'And he backed up [into her] again.'

As for the syntactic status of *back* in these varieties, consider the fact that they occur in exactly the positions in which one can place French-origin adverbs such as *encore* ("again"):

(23) *Je l'ai   **back** fait./ Je l'ai j   **encore** fait.*
I   it have again done/ I   it have again   done
'I did it again.'

(24) *Veux-tu **back** me conter ça?/Veux-tu **encore** me conter ça?*
want you again me to tell that/want you again me to-tell that
'Do you want to tell me that again?'

Thus an account of the syntax of *back* is readily available if we assume that it has been integrated into the grammar of certain French varieties as an adverb. It is generally held that adverbs may be base generated into a number of syntactic positions, with the set of adverb slots available within a language being subject to crosslinguistic variation. While the distribution of adverbs in French and English overlaps to a large extent, there are certain differences between the two languages. Whereas the adverb may precede or follow the infinitival form of the verb in French i.e., Verb Movement is optional here (Pollock 1989), as shown in (25–26), the first option is ungrammatical in English since the adverb must follow the infinitive:

(25) *Il faut   **back** venir.*
it is-necessary back to-come
*'It is necessary back to come.'

---

(26)   *Il faut   venir   **back**.*
        it is-necessary to-come back
        *'It is necessary to come back.'

To summarize the discussion of *back*, then, *chiac* usage, on the basis of the example sentences found in the literature, closely resembles that of other Acadian communities in close contact with English. Mougeon and his colleagues provide us with a reason for *back*'s integration (to varying degrees) into French grammars in the loss in semantic value of the French prefix *re-*. A close examination of the range of syntactic positions *back* occupies in heavy contact situations is available if we assume its reanalysis as a French adverb.[42]

   Back is not the only word of English origin to undergo such integration into the grammar of the host language. Consider the use of *right* as an intensifier, documented by Perrot (6b), Young (7a) and Chevalier and Hudson (8). Recall that Péronnet cited the use of *right* as an intensifier by young speakers in her (1996) study; in the examples she gives, it modifies English adjectives and prepositions: *right sharp*, *right out*. Young notes similar usages, such as in (27):

(27)   Mes *parents sont **right** cool.*
        my parents are   right cool
        'My parents are really cool.'

As Chevalier and Hudson (2005: 292) note, intensifier *right* is common in the English vernacular of Canada's Maritime Provinces, as it is elsewhere in many other English varieties. They note, however, extension of *right* in *chiac* to contexts where its use is impossible in any English variety of which they are aware, since it can also modify verbs:

(28)   *A   i   donne **right** beaucoup de cadeaux.*
        she to-him gives right a-lot   of presents
        'She really gives him a lot of presents.'

Similarly, Young (2003) notes that *right* occurs frequently in her corpus ("*Right* competes with *vraiment* as a means of evoking emphasis in Chiac", p. 122),

---

42.   Tremblay (2005) makes the interesting point that the Old French particle *arrière* "back" shows strong similarities with Acadian *back*. In Old French *arrière* could precede the infinitive and could also be used with the *re-* prefix (as evidence Tremblay cites examples from Buridant 2000). However, I find it hard to make the case for treating present-day *back* usage as an archaism given that we lack (for several centuries) data from the historical record linking the old particle use of *arrière* with that of present-day *back*. The co-occurrence of *re-* with *arrière*, though, could be part of an argument that *re-* has needed "propping up" for a considerable period in the history of the language.

reminiscent of the case of *back* = "repeat an action or processs", where *back* and *encore* are in competition and occur in the same syntactic environments. Interestingly, Chevalier and Hudson document the same role played by English-origin *full* in Quebec French.

Finally, Comeau (2007) accounts for the emergence of intensifier *tight* in Baie Sainte-Marie, Nova Scotia Acadian French, derived from the English predicate adjective *tight*, a somewhat more complex case than the integration of *right* and *full*. Comeau seeks to account for data such as the following (Comeau 2007: 60):

(29) *Il   est apeuré* **tight**.
      he is   scared tight
      'He is really scared.'

He links such usage to other, more transparent borrowings from English in the Nova Scotia variety, i.e., calques involving English *tight* such as in (30):

(30) *Le   chien était amarré* **tight** *à   un arbre.*
      the dog was   tied   tight to a   tree
      'The dog was tied tight to a tree.'

Comeau argues (on the basis of the distribution of tokens across speakers in the Butler corpus for Baie Sainte-Marie Acadian French) that calques of English idiomatic expressions such as *shut tight* (*fermé tight*), *locked tight* (*barré tight*), etc. serve to (1) extend use of resultative predicates in the variety;[43] and (2) open the door for reanalysis of *tight* as an intensifier. Comeau's study, then, points once again to the fact that borrowings follow particular trajectories in the host language which may be accounted for by grammatical theory.

5.2   Borrowed discourse markers and their effects

The topic of borrowed discourse markers is not a new one in the study of Canadian French. Mougeon and Hébrard (1975) report that English *anyway, well, you know,* etc. are associated with the working-class French of Welland, Ontario, in particular with speakers who speak both French and English on a regular basis. Indeed use of such discourse markers is reported for most studies of North American French: Flikeid (1989), King (2000), King and Butler (2005), Poplack, Sankoff and Miller (1988), to name but a few sociolinguistic studies of Canadian French, document a similar set. As Maschler (1994) notes, discourse markers are often expressed in another language in the conversation of bilinguals. Maschler's own study of the

---

43.   Comeau notes, following Legendre (1997), that while resultative predicates do exist in non-contact varieties of French, they are infrequent and limited in distribution.

speech of two Hebrew-English bilinguals found 46 different switched discourse markers in twenty hours of otherwise English conversation. In such cases, Maschler concludes that the codeswitch itself underscores the verbal activity indicated by the discourse marker.

Based on frequency and distribution in the respective corpora, Roy (1979) and Mougeon and Beniak (1991) assign to the English discourse markers the status of borrowing (versus codeswitch). Mougeon and Beniak (1991: 211) put forth the following hypothesis as to why we might find borrowed discourse markers:

> [T]hat sentence connectors and other kinds of discourse organizers like *so* are so often reported in lists of core lexical borrowings may not be a coincidence, since these items all occur at prime switch points. We would tentatively advance the hypothesis that core lexical borrowings like *so* and other sentence connectors may start out as codeswitches (either as single words or as part of switched sentences) which by dint of repetition become loanwords.

Thus it is not surprising that the earliest *chiac* studies note the presence of English-origin discourse markers and that several articles have been devoted exclusively to the topic (e.g., Chevalier 2002; Petraş 2005). Here I will be concerned with considering how the borrowing of discourse markers may make inroads on the grammar, looking at some of the data presented above in this light. Consider the data from Perrot's dissertation presented in (6f), where we find English matrix clauses taking as complement clauses French-language material. The data are reminiscent of those presented by King and Nadasdi (1999) for Prince Edward Island French:

(31)   ***I guess*** *qu'on est   pas mal tout pareil.*
     I guess that   one   is   not   bad   all   equal
     'I guess that we are just about all equal.'

(32)   **I** ***think*** *j'ai plus peur des   chenilles   qu'une serpent.*
     I think   I   have   more   fear of-the caterpillars   than a   snake
     'I think I'm more afraid of caterpillars than a snake'

However, collocations such as *I guess* and *I think* also appear at the "edges" of utterances, as in:

(33)   *Il   a   marié   la   deuxième femme,* **I guess.**
     he   has married the   second   wife   I   guess
     'He married the second wife, I guess'

In Table 3, I present the distribution of particular evidential verbs of English origin in the two Prince Edward Island communities under investigation. The reader will recall that Saint-Louis has much less contact with supralocal varieties of French and rather more contact with English than the Abram-Village variety.

Table 3. English verbs occuring in codeswitches with first person singular pronouns. Matrix = occurring in matrix clauses; elsewhere = occurring as sentential tags or as single-clause utterances in Prince Edward Island Acadian French*

	Saint-Louis		Abram-Village	
Verb	*Matrix*	*Elsewhere*	*Matrix*	*Elsewhere*
guess	98	85	23	97
imagine	8	8	0	0
know (don't know)	1 (0)	3 (47)	0	2 (4)
suppose	0	6	0	0
think (don't think)	43 (5)	47 (19)	0	5 (3)

*Table adapted from King and Nadasdi (1999).

In my own data and in the data reported in the *chiac* literature, codeswitches such as in (6f) or in (31–32) above occur with a particular class of evidentials, verbs of opinion or belief; note the matrix clause is always in the first person. In the Saint-Louis corpus, matrix clause *I guess* actually outstrips its use as a discourse marker. Even in varieties with less contact with English, such as the Newfoundland varieties, one finds discursive *I guess* and "syntactic" *I guess* (but the set does not extend to *I think*, *I imagine*, *I hope*, etc, as it does in Prince Edward Island Acadian and, from what we can see from the literature, in *chiac*). King and Nadasdi (1999) argue that English matrix clauses involving first person use of evidential verbs serve to underscore a speaker's opinion relative to the veracity of the event. We suggest that the degree of uncertainty which accompanies switching to *I guess* has since become associated with all semantically-related codeswitches to English, particularly in the case of the heavy codeswitchers, who are at the vanguard of such usage in Prince Edward Island.

The title of Perrot's 1994 article "*Le chiac … ou whatever*" points to another frequent loanword in my own corpora, and, I imagine, in most recent corpora for languages spoken in close contact with English. Perrot also points us to the use of English-origin *wh-ever* words in free relative clauses, illustrated in (6e), which have *whoever* and *whenever*. Again such usage is not surprising if discursive *whatever* is in frequent use, as in the Prince Edward Island corpus from which the examples given below are drawn (Flikeid 1989 likewise reports such usage in her Nova Scotia corpus):

(34) *On sautait   de la   corde ou **whatever**.*
      one jump-IMP of the rope   or whatever
      'We used to skip or whatever.'

(35) *J'ai rencontré ma   première girlfriend **whatever**.*
      I   have met   my   first   girlfriend whatever
      'I met my first girlfriend whatever.'

The data in (36–38) show use of *wh-ever* words in the Prince Edward Island varieties:[44]

(36) *Il courait **wherever** que ç'a arrêté.*
  he ran-IMP wherever that it has stopped
  'He ran wherever it stopped.'

(37) *Tu peux peinturer la maison **whichever couleur** que tu veux.*
  you can to-paint the house whichever colour that you want
  'You can paint the house whichever colour you want.'

(38) *Je partirons **whenever** que tu veux.*
  we leave-FUT whenever that you want
  'We will leave whenever you want.'

I want to suggest, then, that discursive *whatever* provides the impetus for change which may result in the addition of other *wh-ever* words of English-origin into French discourse, in situations of intense contact. (In fact, the only English *wh-ever* words which do not occur in the Prince Edward Island corpus are *however* and *whyever*.) But what one does not find, among even my most advanced consultants, is use of lone English *wh*-words *who, what, where*, etc. which would seem, from my reading of the literature, to also obtain in *chiac*. The exception is *why*, also found (albeit rarely) in the Prince Edward Island corpus.

The use of English-origin *why* may be understood from the point of view of current grammatical theory since Rizzi (1990) argues that *why* is a clausal adjunct. Cross-linguistically, *why* seems to be an unreliable diagnostic for the landing site of general "well-behaved" *wh*-phrases. In a number of languages, including Spanish (Suñer 1994; Zubizarreta 1998), Romanian (Gabriela Alboiu, p.c.) and Hungarian (Kiss 1998), there is a requirement that while *wh*-phrases have to be adjacent to the highest verbal head, *why* can instead be followed by topics. So *why* is "different", and, in terms of the discussion here, more amenable to occurring in French-language discourse.

My main point here, then, is that when looks closely, a straightforward account of the use of codeswitches involving a restricted grammatical set and the borrowing of function words is available if one considers the trajectory of integration into the borrowing language. It should be recalled that Roy (1979) made a similar argument regarding discursive and conjunctive uses of *but* and *so*. Roy also gives us a clue that some of the *chiac* usage discussed in this section is relatively new, since in her study of the speech of working class Moncton residents, both young and old, male and female, reports that neither interrogative nor relative pronouns were borrowed.

---

44. Note that Acadian French generally allows so-called Doubly-Filled COMP.

5.3 Loanwords across word classes

As noted above, loanwords from the major lexical categories typically lead in terms of the quantity of both types and tokens in mixed-language corpora: the rank order nouns-verbs-adjectives-adverbs has been found in countless studies of language contact, involving French-English contact and a wide range of other language contact situations (cf. Appel & Muysken 1987). In terms of degree of integration into French, in the *chiac* literature I have looked at, verbs occur with the appropriate French morphological markers, e.g., the infinitival ending -*er* (e.g., *watcher*), the past participle ending -*é* (e.g., *shatteré*), third person plural -*ont* (e.g*., ils charchont*),[45] imperfect marking (e.g., *drivais*). Indeed, in every variety of French of which I am aware, including Quebec French but with the exception of Cajun French,[46] English-origin verbs are morphologically-incorporated into French and *chiac*, too, patterns in this way. Young (2002) provides examples of English-origin participial adjectives which have been morphologically-incorporated into French (e.g., *streaké*) and others which have not been so incorporated (e.g., *baked*). However, since patterns of morphological incorporation are interesting in their categorical absence (for verbs, as in the Cajun case) or categorical presence (for verbs, as is the case for the Prince Edward Island corpus) or in variable presence/absence (as in plural suffixation on nouns in Turpin's data), without access to data for the envelope of variation, there is little on which I can comment here. As Poplack, Sankoff and Miller (1988) argue, and Turpin (1998) exemplifies with Acadian data, morphological incorporation alone is not a sufficient diagnostic for the status of a loanword as a borrowing or a codeswitch. What is more interesting in the *chiac* literature is the attention paid to loanwords from functional categories, which are typically borrowed in situations of intense contact. For instance, Quebec French has not borrowed prepositions from English but most Acadian varieties have done so. Since borrowed prepositions are closely related to the range of orphan prepositions in Acadian varieties (King 2000, 2005), I will consider them separately in section 5.5. Here I will be concerned with use of English numbers and non-numeric quantifiers, the subject of commentary by Péronnet (1996), Perrot (1995) and Young (2002).

---

45. Presence of the *ils* … ont traditional morphology is a variable feature of present-day Acadian varieties in that the traditional variant alternates with a phonetically-null variant (the latter associated with external varieties of French).

46. Rather famously, in Cajun French verbs of English origin are not morphologically incorporated (Brown 1986; Picone 1993; Dubois & Sankoff 1996). Consider Brown's example, "*Il a retire* [rytayr]"('he retired') which would be rendered *Il a retiré* in Acadian French, with the French past participle ending -*é*. I have suggested elsewhere (e.g., Dubois, King & Nadasdi 2004), that lack of morphological incorporation of English loanwords may be related to the general breakdown of the morphology of the verb in Cajun French.

In her data, Young finds the language of the head noun to be a strong predictor of the language of the number (p. 176); in other words, English numbers, both cardinal and ordinal, are most common with English nouns, thus in codeswitches (recall that Young does not make a borrowing-codeswitching distinction). Young does note several occurrences of *every, any*, etc. with French-origin nouns.

Flikeid's (1989) discussion of such data in her five-community Nova Scotia Acadian corpus is useful here. Loanwords follow the usual borrowing hierarchy in this corpus, with all categories represented. Use of English-origin quantifiers was found to be most closely associated with the community in most intense contact with English, Pubnico. However, there is variation among what appear to be members of the same set, in that English loanwords *anyone* and *anything* are fairly widespread while *everything* is attested only once in the Nova Scotia sample.

Given that the number of tokens for members of functional categories will be much smaller than for the lexical categories and that the existing *chiac* corpora are relatively small, with Young's 53,000 word corpus being one of the largest, it is difficult for the authors (or the reader) to assess the status of English-origin quantifiers. For example, even my much larger Prince Edward Island corpora, while it contains over 7000 lone nouns of English origin, has only 170 lone pronouns of English origin, using the category pronoun in the most general sense. The most we can make of Perrot's and Young's findings here is to recall that such data were not mentioned by the early researchers, such as Roy (1979) and Gérin (1984), nor were they discussed with relation to older speakers' language use by Péronnet (1989), suggestive of innovation in the variety. It is also interesting to note that usage of English ordinal numbers is strongly present in the current *Acadieman* TV show and comic books mentioned above: the first season DVD of the TV show is subtitled *La complete first saison* ('The Complete First Season') and *Acadieman* is billed as *Le first superhero acadien* ('The First Acadian Superhero'). Thus use of such quantifiers appears to have become part of the set of "performable" features of the variety in such mediated contexts today.

5.4   Borrowed verb + particle combinations and their effects

Mention of one particular functional category, prepositions, stands out in the *chiac* literature since the 1970s. Roy's (1979) corpus contains a number of instances of prepositions of English origin (*about, on* and *off*). Gérin (1984) draws our attention to the combination *maker out*. Perrot and Young point to "verb + particle" combinations such as *parler about* (Perrot, 6c) and *turné off* (7h) as do Chevalier and Long. Such data, then, are almost as prominent in the *chiac* literature as English-origin *back*.

Likewise, Flikeid's (1989) study of Nova Scotia Acadian found that several prepositions of English origin have been borrowed. She also cites a number of

borrowed "verb + particle" combinations, including *knocker down, picker up*, etc. Flikeid (1989: 225) also draws attention to variation in such usage in Nova Scotia: she notes that in the case of Pubnico, the community in which contact with English is most intense, the corpus contains about a hundred different verb + particle combinations, but in the case of Chéticamp, where contact is least intense, there were only five such examples.

In what follows, I will draw a connection between borrowed prepositions, borrowed verb + particle combinations, and the use of orphan prepositions in interrogatives, illustrated in (39–40) with examples from Chevalier and Long (2005: 202–03):

(39) *Quosses tu   parles **de**?*
       what   you speak of
       'What are you talking about?'

(40) *Qui'ce qu'alle a   donné le   livre **à**?*
       who   that she has given the book to
       'Who did she give the book to?'

First, consider the fact that in the two Prince Edward Island corpora, 17% of the English-origin verbs (n = 2349) occurred in "verb + particle" combinations, or, more technically, in combinations of verbs and intransitive prepositions. While lone English-origin prepositions certainly occurred in the two corpora, combinations such as found in the tables below were particularly frequent (frequently-occurring combinations are in bold).

Several commentators draw attention to such combinations in the *chiac* literature, as we have seen, so it is tempting to infer they are in common use, as in the case of *back* usage. Indeed, while relative frequency is not indicated, in an ap-

Table 4. Verb + Preposition Combinations in the Abram-Village Corpus

bailer out	ganger up	piler up	slipper out
bosser around	getter along	plugger up	**slower down**
builder up	getter over	puller out	smoker up
chickener out	**giver up**	puller through	sporter up
clearer up	**grower up**	putter up	starter off
cusser down	**hanger around**	reflecter off	stepper out
dier down	hanger up	runner around	straightener out
se dresser up	**kicker out**	runner about	tier down
dropper out	layer off	**runner out**	turner down
ender up	looker up	**setter up**	turner on
figurer out	maker up	shipper out	**turner out**
filler out	mixer up	shower off	turner over
finder out	passer out	shower up	walker out
fooler around	picker up	shutter out	walker up
freaker out	picker from	signer in	washer out
fusser pour	picker out	singler out	

Table 5. Verb + Preposition Combinations in the Saint-Louis Corpus

backer up	getter over	puller through	**slower down**
bosser around	**giver up**	putter up	stepper in
clearer up	**grower up**	reflecter off	stepper out
cusser down	**hanger around**	runner around	tier down
se dresser up	layer off	**runner out**	turner down
figurer out	looker up	**setter up**	turner on
filler out	maker up	shower up	**turner out**
finder out	mixer up	**shutter off**	walker out
fooler around	plugger up	signer in	walker up
getter along	puller out	slipper out	

pendix to their 2005 article, Chevalier and Long give eight combinations involving *up* (e.g., *giver up*), thirteen with *out* (e.g., *finder out*), four with *off* (e.g., *dozer off*), two with *on* (e.g., *mettre on*), two with *in* (e.g., *filler in*) and two with *around* (e.g., *hanger around*), found in the three *chiac* corpora they looked at. (Note that the PEI corpora also include combinations of English-origin prepositions with French-origin verbs.) While more attention is paid to the range of combinations themselves in the *chiac* literature than to their use in constructions such as interrogatives or passives (see for example Young 2002) this may be attributable to the fact that these studies are based on small corpora which may not display a range of construction types with much frequency.

In research on Prince Edward Island Acadian French (King & Roberge 1990; King 2000, 2005), I have argued that the borrowing of such combinations has served as a trigger for structural change in this variety, an argument that may be extended to the *chiac* case. Data such as in (39–40) are indicative of movement of the complement of the preposition, a phenomenon known as Preposition Stranding, a rare option among the world's languages, notably found in English and Mainland Scandinavian languages. In French, prepositions have long been argued not to be able to undergo such movement (i.e., the DP is not extractable from the PP) since they have inherently Case-marked complements (see Kayne 1984 for a discussion of the difference between French and English in this regard). Following on Kayne's argument that stranding is a genuine property of the preposition itself, I suggest that the borrowing of English-origin prepositions has brought about structural change: "the direct borrowing of English-origin prepositions has resulted in an extension of a property of English prepositions, the ability to be stranded, to the whole set of English prepositions" (King 2000: 147).[47] On the basis

---

47.   In King (2000, 2005) I present arguments against wholesale importation of English structure but of this one property in particular; I will not rehearse these arguments here as they are not central to the issue at hand.

of Chevalier and Long's (2005) data and discussion (along with data from media representations such as discussed above), it would seem that *chiac* lines up with the Prince Edward Island varieties with respect to this parameter.

6.   Conclusion

I began this chapter with two definitions of *chiac*, both by nonlinguists. Thériault decries the corruption of the language; LeBlanc celebrates a variety which gives a voice to a people. Arguably the situation has changed dramatically since the 1970s though we would not want to argue that negative attitudes have disappeared (cf. Keppie 2002; Young 2002). Still, it is clear that for many *chiac* has become a positive identity marker.

While this chapter does not constitute a full description of *chiac*, I have been able to contextualize earlier studies and (re)analyze some of the data in the light of a particular view of borrowing and codeswitching, one informed by variationist methodology and theory and to some extent by formal (lexicalist) approaches to variation.[48] In general, I suggest a research focus for Acadian French on comparative sociolinguistics, which will dispel (or confirm) any claims of uniqueness of particular usages (Chevalier & Long (2005: 210) likewise argue for more comparison among varieties). By looking at language use in communities which differ in the social context of spoken French, we are able to track the role of social factors in language variation and change (e.g., by comparing the French of a range of New Brunswick communities, including, of course, Moncton, and by comparing with Acadian communities in the other Atlantic Provinces). Further, by investigating substantial differences among varieties with closely-related grammars, we are able to test hypotheses about how particular grammatical properties are linked to one another (e.g., the borrowing of prepositions and the emergence of preposition stranding).

References

Appel, R. & Muysken, P. 1987. *Language Contact and Bilingualism*. London: Edward Arnold.

Arrighi, L. 2005. Des prépositions dans un corpus acadien: Évolution du système linguistique française, archaïsmes et/ou calques à l'anglais et/ou calques de l'anglais? In *Français d'Amérique: Approches morphosyntaxiques. Actes du colloque international Grammaire comparée des variétés de français d'Amérique,* P. Brasseur & A. Falkert (Eds), 239–248. Paris: L'Harmattan.

---

48.   King (2005) provides an elaboration of this perspective.

Auger, J. 2005. Un bastion francophone en Amérique du Nord: Le Québec. In *Le français en Amérique du Nord*, A.Valdman, J. Auger & D. Piston-Hatlen (Eds), 39–80. Québec: Presses de l'Université Laval.

Beaulieu, L. 1996. «Qui se ressemble s'assemble» et à s'assembler on finit par se ressembler: Une analyse sociolinguistique de la variable *si/si que* en français acadien du nord-est du Nouveau-Brunswick. In *Les Acadiens et leur(s) langue(s)*, L. Dubois & A. Boudreau (Eds), 91–111. Moncton: Éditions d'Acadie.

Bouchard, D. 1982. Les constructions relatives en français vernaculaire et en français standard: Étude d'un paramètre. In *La syntaxe comparée du français standard et populaire: Approches formelles et fonctionnelles*, C. Lefebvre (Ed.), 103–34. Québec: Office de la langue française.

Boudreau, A. 1998. Répresentations et attitudes linguistiques des jeunes francophones de l'Acadie du Nouveau-Brunswick. Ph.D. dissertation, Université de Paris.

Boudreau, A. 1996. Les mots des jeunes Acadiens et Acadiennes du Nouveau-Brunswick. In *Les Acadiens et leur(s) langue(s)*, L. Dubois & A. Boudreau (Eds), 137–55. Moncton: Éditions d'Acadie.

Brasseur, P. & Falkert, A. (Eds). 2005. *Français d'Amérique: Approches morphosyntaxiques. Actes du colloque international Grammaire comparée des variétés de français d'Amérique*. Paris: L'Harmattan.

Brault, M. 1969. *Éloge du chiac*. Office national du film du Canada.

Brown, B. 1986. Cajun/English code switching: A test of formal models. In *Diversity and Diachrony* [Current Issues in Linguistic Theory 53], D. Sankoff (Ed.), 399–406. Amsterdam: John Benjamins.

Buridant, C. 2000. *Grammaire nouvelle de l'ancien français*. Paris: Sedes.

Canale, M., Mougeon, R., Bélanger, M. & Main, C. 1977. Recherches en dialectologie franco-ontarienne [Travaux de recherches sur le bilinguisme 14], 1–20. Toronto: Ontario Institute for Studies in Education.

Charpentier, J.-M. 2000. Le parler du Haut Poitou et sa disparition «programmée» en deux temps. In *Actes du XXIe colloque international de linguistique fonctionnelle. Éditions de l'Université Al. I Cuza*, 133–138.

Chevalier, G. 2002. La concurrence entre «ben» et «well» en chiac du sud-est du Nouveau-Brunswick (Canada). *Cahiers de sociolinguistique de Rennes 7: Langues en contact, Canada-Bretagne*, 65–81.

Chevalier, G. & Long, M. 2005. Finder out, pour qu'on les frigge pas up, comment c'qu'i workont out: Les verbes à particules en chiac. In *Français d'Amérique:Aapproches morphosyntaxiques. Actes du colloque international Grammaire comparée des variétés de français d'Amérique*, P. Brasseur & A. Falkert (Eds), 201–212. Paris: L'Harmattan.

Chevalier, G. & Hudson, C. 2005. Deux cousins en français québécois et en chiac de Moncton: *Right* et *full*. In *Français d'Amérique: Approches morphosyntaxiques. Actes du colloque international Grammaire comparée des variétés de français d'Amérique*, P. Brasseur & A. Falkert (Eds), 289–302. Paris: L'Harmattan.

Comeau, P. 2007. The Integration of Words of English Origin in Baie Sainte-Marie Acadian French. MA Major Research Paper, York University.

Comeau, P. & King, R. 2007. *Pas si simple*: Past temporal reference in Acadian French. Paper presented at NWAV 36, University of Pennsylvania.

Comeau, P. & King, R. 2008. Valorizing 'corrupted French': *Chiac* and the *Acadieman* phenomenon. Paper presented at Sociolinguistics Symposium 17, Amsterdam.

Coupland, N. 2007. *Style: Language Variation and Identity*. Cambridge: CUP.

Dion, N. 2003. L'effacement du que en français canadien: Une étude en temps réel. Mémoire, University of Ottawa, University of Ottawa.

Dubois, L. 2003. Radios communautaires acadiennes: Idéologies linguistiques et politiques langagières. In *L'Acadie plurielle. Dynamiques identitaires collectives et développement au sein des réalités acadiennes*, M. Basque, A. Magord & A. Giroux (Eds), 307–323. Moncton: Université de Moncton.

Dubois, L. 2005. Le français en Acadie des Maritimes. In *Le français en Amérique du Nord*, A.Valdman, J. Auger & D. Piston-Hatlen (Eds), 81–98. Québec: Presses de l'Université Laval.

Dubois, S., King, R. & Nadasdi, T. 2004. Past and present agreements: A comparison of third person plural marking in Acadian and Cajun French. Paper presented at Sociolinguistic Symposium 15, Newcastle.

Dubois, S. & Sankoff, D. 1996. L'absence d'inflection sur les emprunts à l'anglais dans le français cadjin. Paper presented at the annual meeting of ACFAS, Montreal.

Flikeid, K. 1984. *La variation phonétique dans le parler acadien du nord-est du Nouveau-Brunswick: Etude sociolinguistique*. Frankfurt: Peter Lang.

Flikeid, K. 1989. 'Moitié anglais, moitié français'? Emprunts et alternance de langues dans les communautés acadiennes de la Nouvelle Ecosse. *Revue québécoise de linguistique théorique et appliquée* 8(2): 177–227.

Flikeid, K. 1994. Origines et évolution du français acadien à la lumière de l'étude de la diversité contemporaine. In *Les origines du français québécois*, R. Mougeon & É. Beniak (Eds), 275–326. Québec: Presses de l'Université Laval.

Gammel, I. & Boudreau, J. 1998. Linguistic schizophrenia: The poetics of Acadian identity construction. *Journal of Canadian Studies* 32(4): 52–68.

Geddes, J. 1893–94. Comparison of the two Acadian French dialects spoken in the north-east of North America with the Franco-Canadian dialect spoken at Ste. Anne de Beaupré, province of Quebec. *Modern Language Notes* 8(8): 449–459; 9(1): 1–11; 9(2): 99–115.

Geddes, J. 1908. *Study of an Acadian French Dialect Spoken on the north shore of the Baie-des-Chaleurs*. Halle: Niemeyer.

Gérin, P. 1984. La création d'un troisième code comme mode d'adaptation à une situation où deux langues sont en contact, le chiac. In *Papers from the 7th Annual Meeting of the Atlantic Provinces Linguistic Association*, Helmut Zobl (Ed.), 31–3. Moncton: Université de Moncton.

Gesner, B. 1979. *Etude morphosyntaxique du parler acadien de la Baie Sainte-Marie, Nouvelle-Ecosse (Canada)*. Quebec: Centre for International Research on Bilingualism.

Gesner, B. 1986. *Bibliographie annotée de linguistique acadienne*. Québec: Centre internationale de recherches sur le bilinguisme.

Gumperz, J. 1982. *Discourse Strategies*. Cambridge: CUP.

Hamers, J. & LeBlanc, M. 1989. *Bilinguality and Bilingualism*. Cambridge: CUP.

Hull, A. 1955. The Franco-Canadian Dialect of Windsor, Ontario: A preliminary study. Ph.D. dissertation, University of Washington.

Johnstone, B. 1999. Uses of Southern-sounding speech by contemporary Texas women. *Journal of Sociolinguistics* 3: 505–522.

Kasparian, S. 2005. «Holy fuck ça te coûte cher»: Gros mots et euphémisme linguistique en acadien. In *Français d'Amérique: Approches morphosyntaxiques. Actes du colloque international Grammaire comparée des variétés de français d'Amérique*, P. Brasseur & A. Falkert (Eds), 313–326. Paris: L'Harmattan.

Kayne, R. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris.

Keppie, C. 2002. Les attitudes à l'égard du chiac. MA thesis, Carleton University.

King, R. 2000. *The Lexical Basis of Grammatical Borrowing*. Amsterdam: John Benjamins.

King, R. 2005. Crossing grammatical borders: Tracing the path of contact-induced linguistic change. In *Selected Proceedings of the Eleventh International Conference on Methods in Dialectology*, M. Filppula et al. (Eds), 233–251. Amsterdam: John Benjamins.

King, R. & Butler, G. 2005. Les Franco-Terreneuviens et le franco-terreneuvien. In *Le français en Amérique du Nord*, A.Valdman, J. Auger & D. Piston-Hatlen (Eds), 169–187. Québec: Presses de l'Université Laval.

King, R. & Nadasdi, T. 1999. The expression of evidentiality in French-English bilingual discourse. *Language in Society* 28(3): 355–66.

King, R. & Nadasdi, T. 2006. Another look at Que-deletion. *New Ways of Analyzing Variation 35*, The Ohio State University.

King, R. & Roberge, Y. 1990. Preposition stranding in Prince Edward Island French. *Probus* 2: 351–369.

Kiss, K. 1998. Identificational focus versus information focus. *Language* 74(2): 245–273.

Leblanc, D. 2007. *Acadieman: Ses origines* (bande dessinnée). Moncton: Les Éditions Court-Circuit.

Leblanc, D. 2006. *Acadieman: La complete first season*. DVD Télévision Rogers.

Leblanc, G. 1995. *Éloge du chiac: Poésie*. Moncton: Les Editions Perce-Neige.

Legendre, G. 1997. Secondary predication and functional projections in French. *Natural Language and Linguistic Theory* 15: 43–87.

Lucci, V. 1972. *Phonologie de l'acadien*. Montréal: Didier.

Martineau, F. 1988. Variable deletion of *que* in the spoken French of Ottawa-Hull. In *Advances in Romance Linguistics*, D. Birdsong & J-P Montreuil (Eds), 275–287. Dordrecht: Foris.

Maschler, Y. 1994. Metalanguaging and discourse markers in bilingual conversation. *Language in Society* 23(3): 325–366.

McKillop, A. 1987. Une étude sociolinguistique du parler "brayon" d'Edmunston au Nouveau-Brunswick. MA thesis, University of New Brunswick.

Meechan, M. & Poplack, S. 1995. Orphan categories in bilingual discourse: Adjectivization strategies in Wolof-French and Fongbe-French. *Language Variation and Change* 7: 169–194.

Mougeon, R. & Beniak, É. 1991. *Linguistic Consequences of Language Contact and Restriction: The case of French in Ontario, Canada*. Oxford: OUP.

Mougeon, R., Brent, P., Bélanger, M. & Cichocki, W. 1980. *Le français parlé en situation minoritaire*. Vol. 1: *Emploi et maîtrise du français parlé par les élèves des écoles de langue française dans les communautés franco-ontariennes minoritaires*. Toronto: Ontario Ministry of Education.

Mougeon, R. & Hébrard, P. 1975. Aspects de l'assimilation linguistique dans une communauté francophone de l'Ontario. *Working Papers on Bilingualism* 5: 1–38.

Muysken, P. 2000. *Bilingual Speech*. Cambridge: CUP.

Myers-Scotton, C. 1993. *Duelling Languages: Grammatical structure in code-switching*. Oxford: Clarendon Press.

Pavel, M. 2005. Régionalismes grammaticaux en «chiac». In *Français d'Amérique: Approches morphosyntaxiques. Actes du colloque international Grammaire comparée des variétés de français d'Amérique*, P. Brasseur & A. Falkert (Eds), 49–56. Paris: L'Harmattan.

Péronnet, L. 1975. *Modalités nominales et verbales dans le français franco-acadien de la région du sud-est du Nouveau-Brunswick*. MA thesis, Université de Moncton.

Péronnet, L. 1989. Analyse des emprunts dans un corpus acadien. *Revue québécoise de linguistique théorique et appliquée* 8(2): 229–251.

Péronnet, L. 1996. Nouvelles variétés de français parlé en Acadie du Nouveau-Brunswick. In *Les Acadiens et leur(s) langue(s)*, L. Dubois & A. Boudreau (Eds), 121–133. Moncton: Éditions d'Acadie.

Perrot, M-E. 1994. Le chiac ou … whatever: Le vernaculaire des jeunes d'une école secondaire francophone de Moncton. *Études canadiennes* 20: 237–246.

Perrot, M-E. 1995. Aspects fondamentaux du métissage français/anglais dans le chiac du Moncton (Nouveau-Brunswick, Canada). Ph.D. dissertation, Université de la Sorbonne Nouvelle – Paris III.

Perrot, M.-E. 1998. Les modalités du contact français/anglais dans un corpus chiac: Métissage et alternance codique. *Le français en Afrique francophone*, Paris, Didier-Érudition, Institut national de la langue française, CNRS, nº 12, p. 219–226.

Petraş, C. 2005. Valeurs pragmatiques du contact de langues au niveau des marqueurs discursifs dans un corpus acadien. In *Français d'Amérique: Approches morphosyntaxiques. Actes du colloque international Grammaire comparée des variétés de français d'Amérique*, P. Brasseur & A. Falkert (Eds), 275–288. Paris: L'Harmattan.

Picone, M. 1993. Lexical code-switching in Louisiana French. Paper presented at NWAVE-22, University of Ottawa.

Poirier, P. 1993. *Le glossaire acadien*. Pierre M. Gérin, (Ed). Moncton: Centre d'études acadiennes. (First published in installment form in the newspaper *L'Evangéline* beginning in 1927).

Pollock, J-Y. 1989. Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry* 20: 365–424.

Poplack, S. 1980. Sometimes I'll start a sentence in English *y termino en espagnol*: Toward a typology of code-switching. *Linguistics* 18: 581–618.

Poplack, S. & Meechan, M. 1998. Instant loans, easy conditions: The productivity of bilingual borrowing. *International Journal of Bilingualism* 2(2): 127–234.

Poplack, S., Sankoff, D. & Miller, C. 1988. The social and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26: 47–104.

Rizzi, L. 1990. *Relativized Minimality*. Cambridge MA: The MIT Press.

Roberge, Y. 1998. Les prépositions orphelines dans diverses variétés de français d'Amérique du Nord. In *Français d'Amérique: Variation, créolisation, normalisation*, P. Brasseur (Ed.), 49–60. Avignon: Centre d'études canadiennes.

Roy, M.-M. 1979. Les conjonctions «but» et «so» dans le parler de Moncton. MA thesis, Université du Québec à Montréal.

Sankoff, G. 2002. Linguistic outcomes of language contact. In *The Handbook of Language Variation and Change*, J.K. Chambers & N. Schilling-Estes (Eds), 638–668. Oxford: Blackwell

Sankoff, G. 1980. Above and beyond phonology in variable rules. In *The Social Life of Language*, 81–95. Philadelphia PA: University of Pennsylvania Press.

Sankoff, G. & Blondeau, H. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83(3): 560–88.

Sankoff, G, Sarrasin, R. & Cedergren, H. 1971. Quelques considérations sur la distribution de la variable *que* dans le français de Montréal. Paper presented at the Congrès de l'Association canadienne française pour l'Avancement des Sciences.

Suñer, M. 1994. Verb movement and the licensing of argumental WH-phrases in Spanish. *Natural Language and Linguistic Theory* 12: 335–372.

Thériault, L. 1975. Le parler acadien ou le chiac. *L'Évangéline* October 26 and 27.

Thompson, J. 1986. Chiac: An Example of Dialect Change and Language Transfer in Acadian French. MA thesis, University of Victoria.

Tremblay, M. 2005. *Back* en français acadien: Archaïsme ou innovation? In *Français d'Amérique: Approches morphosyntaxiques. Actes du colloque international Grammaire comparée des variétés de français d'Amérique*, P. Brasseur & A. Falkert (Eds), 263–274. Paris: L'Harmattan.

Turpin, D. 1998. "Le français, c'est le *last* frontier": The status of English-origin nouns in Acadian French. *International Journal of Bilingualism* 2(2): 221–233.

Vincent, D. 1982. *Pressions et impressions sur les sacres au Québec.* Québec: Office de la langue française.

Vinet, M-T. 1984. La syntaxe du québécois et les emprunts de l'anglais. *Revue de l'Association québécoise de linguistique* 3: 221–242.

Young, H. 2002. C'est either que tu parles français, c'est either que tu parles anglais: A Cognitive approach to Chiac as a contact language. Ph.D. dissertation, Rice University.

Zubizarreta, M. 1998. *Prosody, Focus and Word Order*. Cambridge MA: The MIT Press.

How to predict the evolution of a bilingual community

David Sankoff
University of Ottawa

Working for the Bureau of Statistics of Papua and New Guinea in 1967, I carried out a territory-wide population projection, based on a complete collection of village patrol reports and some birth records from a small sample of hospitals. Many years later, I drew on this experience to develop a demolinguistic model for projecting the evolution of a bilingual community, which has been applied to nine language revitalization movements in Spain and in the British Isles. In the light of this, I discuss current literacy rates in Tok Pisin and in English in Papua New Guinea and the social conditioning of the spread of these languages.

Keywords: demolinguistic projection; census data; Papua New Guinea; bilingual community; language shift; revitalization; transmission; acquisition model; Tok Pisin

1.   A short career as a demographer

In January 1967, I left Mambump village, on the headwaters of the Snake River in the Morobe District of New Guinea, and made my way to Port Moresby, the administrative capital of both Papua and New Guinea. I was looking for temporary work to ease a burden on the finances of Gillian's Ph.D. fieldwork project among the Buang; my presence as unofficial assistant anthropologist doubled the burn rate of the subsistence budget in her meagre Canada Council doctoral grant.

In Port Moresby, I found my way to the Bureau of Statistics of the Australian administration of the territories, a small operation, perhaps 15 people all told, and asked for a job. I was hired on the spot by an astonished director, Max Barton, who, along with the rest of the government departmental heads, had been trying with limited success for a number of years to recruit qualified personnel from Australia and overseas, offering attractive travel, vacation and other inducements, and had never previously known an applicant to simply walk in out of the bush and knock on his office door. Especially an applicant with a masters in statistics, which established me as the youngest and, academically, the most qualified employee in

the place, with the title of Senior Research Officer. I remember the faces, but only the first names, of many of my other colleagues in the Bureau and elsewhere in the Public Service, almost all expatriates from Australia and far-flung places such as Ceylon, France, England and Italy.

After a few weeks, I was assigned a project of particular interest to Barton. I was not privy to his motivation for having this done, but was quite happy to take it on. It promised more of a challenge than the dreary interdepartmental task forces in which I had been participating, estimating future personnel requirements within the government. My new task was to compile a statistical picture of the population of Papua and New Guinea based on annual government patrol reports carried out by dozens of police and other officials, visiting villages all across the territories.

1.1   The data

The reports contained handwritten tabulations of total population, broken down by sex and, I believe, age, as well as totals of births and deaths. I don't recall whether there were entries for migration data, but it seems likely there were. I had hundreds of these tabulations, all inscribed on, as I remember, pink legal size forms in sideways format. Needless to say, there was much missing data, many illegible entries and sheets from patrols several years apart in different areas. Nevertheless, the collection was remarkably systematic and I had confidence in my summary statistics for each district, consisting of the number of persons in each district d, subdivided by age group a and sex s, which I will denote by $N(d, a, s)$.

At some point in this work, either at Barton's urging or as a result of my own interest, I started thinking about making population projections. In any case I had Max Barton's total support in this project. Though I had no training in the discipline of demography, I had some familiarity with simple mathematical models of population growth and a reasonably diverse background in science. It didn't take much reflection to realize that only three processes directly influence population size: birth, death and migration. Future migration rates cannot be predicted and can only be assumed to be constant in the short term, but births and deaths are clearly dependent on the age and sex distribution of the population. Even though I had these distributions, I didn't know the nature of the dependence, namely the age-specific death rates and female fertility rates.

I found out about a collaborative project at a number of private hospitals across the territories to institute the collection of vital statistics in a standardized way. For a mother giving birth at the hospital or recorded at the hospital, her age was part of the data recorded. The Bureau arranged for me to visit three of these hospitals, the Baptist Mission Hospital near Mount Hagen in the New Guinea Highlands, the Christian Mission in Many Lands hospital in Anguganak, 100 km

from the famous Sepik River, and the hospital at a Roman Catholic mission near Popondetta in Papua. It was arranged that I should travel to these places, usually in the small aircraft operated by the Mission Aviation Fellowship. All it required in each hospital was a few hours to go through their records, pick out the births, and record the age of the mother; later I would roughly normalize these statistics by taking account of the time span over which they were recorded and by comparing them to the age distribution of females in the district, which I had previously estimated. In addition, Roy Scragg, at the time Director of Public Health in the territories, gave me access to fertility rates from New Ireland, which he had compiled for his 1954 University of Adelaide MD thesis, later reprinted (1957) in Port Moresby and often cited. I was aware of the many methodological difficulties in my procedure, but using these data, differing substantially from one hospital to another, was far superior to not having any age-dependent rates at all for most of the territories, which was the situation at the outset. I assumed that my Popondetta data pertained to the whole territory of Papua, Roy Scragg's data to New Britain, New Ireland, and the nearby islands, the Mount Hagen data to the Highlands, and the Anguganak data to the rest of the New Guinea territory.

I don't remember how I estimated age-specific death rates, but I can imagine several possibilities, either roughly estimating them from the drop-off in the age distributions, using Scragg's data, or some other published source. The details of these rates would have been less critical to mid-term population projections than the age-specific fertility rates.

1.2   The projections

In simple demographic projections, cohorts are defined by administrative district, five-year age group (cohort 0 for 0–4, cohort 1 for 5–9, cohort 2 for 10–14, etc.) and sex (female or male). Each year, every individual in a cohort ages by one year, so that the oldest 20% are promoted to the next older cohort. Every individual has a probability of dying, depending on age and sex, and being removed from the population, and every woman in the 15–49 age range has a probability of giving birth, depending on her age, which shows up as an increment in the 0–4 age cohort. There are inward and outward flows to each cohort due to migration. This may be formalized as:

$$N_y(d, a, s) = N_{\text{year } y}(\text{district } d, \text{ age cohort } a, \text{ sex } s)$$
$$= 0.8 N_{y-1}(d, a, s) + 0.2 N_{y-1}(d, a-1, s) - m(d, a, s)$$
$$+ \sum_{d'} [f_{s,a}(d', d)] - f_{s,a}(d, d')],$$

for $y = 1967, 1968, \ldots$ and $a = 1, 2, \ldots$, where N_y is the number of individuals in the (d, a, s) cohort at the end of year y, the factors 0.8 and 0.2 represent the "promotions" between five-year age cohorts during the year y, m is the annual death rate and $f_{s,a}(d, d')$ and $f_{s,a}(d', d)$ represent the number of individuals in (s, a) cohorts migrating out of district d to d' or into d from d', respectively, and where we include one "dummy" district d' to account for migration out of and into the country. In addition,

$$N_y(d, 0, s) = \sum_{a=4,\ldots,10} \frac{b(a, d)}{2} N (d, a, \text{female}) + 0.8 N_{y-1}(d, 0, s) - m(d, 0, s)$$

$$+ \sum_{d'} [f_{s,0}(d', d) - f_{s,0}(d, d')],$$

where the b/2 represents the division of the births into females and males. Actually, the sex ratio at birth ranges up to 0.55 males versus 0.45 females. Initially,

$$N_{1966}(d, a, s) = N (d, a, s),$$

as compiled from the data I had gathered, described in Section 1.1, the age-specific annual birthrate b comes from the hospital data, and the m and the f are obtained or estimated in various ways.

For each year $y = 1967, 1968, \ldots$ in the projection, the changes in N_y are calculated for each cohort separately, and the total population is then adjusted. Each of the 16 or so districts at the time (there are now 20 provinces) then has some 34 or 36 age/sex cohorts, requiring yearly calculations for almost 600 cohorts. On a computer, this is no problem at all, but in 1967 there was exactly one computer in Port Moresby that I knew about. As the end of my contract approached, I wrote a program to do the projections but it became clear that I would not have enough time and computer access to complete the entry of the program into punch card format, implement and debug the program and carry out and verify the calculations. Instead, Max Barton assigned his secretary to do the computation on a desktop electric calculator. Though I worked out an efficient protocol for her, essentially an old-fashioned manual spreadsheet, she still had to do several days of extremely tedious work.

1.3 The report

I wrote up the whole project by hand, including the district by district projections over the short and mid-terms, and gave it to Barton. I know he read it, but as far as I know it simply ended up on a shelf or in a file drawer. I was proud of it but as an Australian public servant I understood that I had no rights over it. This was in the days where copiers were not available everywhere, or anywhere for that matter, so I did not even have a copy.

I mentioned earlier that I was not aware of why I was asked to do the project. Max did tell me there was talk of undertaking a national census, or rather a sample census, and he wanted to see whether this was really necessary or worthwhile, given that we already had the pertinent information in the patrol reports. And with much better coverage than a sample census could provide. I know now, and indeed I found out only a few years later on another trip to Papua and New Guinea, that a sample census had already been carried out in 1966, before I even got to Port Moresby (National Statistical Office of Papua New Guinea website, 2007). I can only surmise, and I may be completely off base, that Barton was resentful of other experts coming in from Canberra and intruding in what should have been his domain of responsibility and he was using my project to prove they were superfluous. If this is the case, he proved his point to his own satisfaction, because when I visited him on the later trip to Papua and New Guinea, he told me that my projections had been bang on. But it was irrelevant to statistical policy (fortunately) because there was another census in 1971, and in 1980, 1990 and 2000.

I didn't mind too much that my project led nowhere, since the experience was quite an adventure for me, though I may have covertly hoped that I could have written it up for publication.

2.   Demolinguistics

Fast forward 35 years. I was collaborating with Raquel Casesnoves Ferrer, a former PhD student of Pierrette Thibault, on analyzing her data on the factors influencing the attitudes, learning and choice of Catalan versus Castilian in the city of Valencia (Casesnoves Ferrer & Sankoff 2003, 2004a). Though we found many factors affecting these aspects of the revival of Catalan, it became clear that the demographic considerations of language acquisition processes must enter into any prediction of the future composition of a bilingual community. Only then could the language planning process be evaluated.

We would have to be able to predict not only how the different groups in the population would grow or shrink over the years, but also how linguistic revival measures would differentially affect them. It didn't take long for me to realize that I knew just how to do this, even though many years had passed since my first forays into population projection. And so I wrote the demolinguistics program DMLX, where the population dynamics component is no doubt very similar to the stillborn version from the Port Moresby episode. This time I implemented the program and succeeded in running it on real data (Casesnoves Ferrer & Sankoff 2004b).

Over the past few years Casesnoves Ferrer, now at the Universitat Pompeu Fabra in Barcelona, has used DMLX to study the linguistic revival of Catalan in

Catalonia, the Balearic Islands and the Valencian Country, the Basque language in the Basque Country and Navarre, and Galician in Galicia, in collaboration with Maite Turell, and of Welsh, Scottish Gaelic and Irish (Casesnoves Ferrer, Sankoff & Turell 2006; Casesnoves Ferrer 2007).

In the ensuing sections, I will not go into the details of DMLX nor the studies in Spain and the UK. Rather I will give a general exposition of the principles of the method, the types of data needed and the kinds of results it can produce.

2.1 Demolinguistic projection

How can we graft the processes of language shift and language revival onto a demographic projection protocol? The key to answering this is to divide each geographic/age/sex cohort into two, one cohort that has, or self-reports that it has, bilingual ability, and the other not. Then an individual can leave a cohort, not only by dying, migrating, or being promoted to an older cohort, but also by acquiring linguistic ability.

Any projection of the future requires a sufficient characterization of the present. We must then be able to define annual rates of flow into and out of each cohort. Most of the flow is from each five year cohort to the next oldest one of the same sex and linguistic ability in a district, and this needs no additional data, but there is a flow into the 0–4 cohorts due to births and from all cohorts due to deaths, both of which require input in addition to the usual census tables. And finally there is a flow from each cohort without the ability to speak the reviving language into the corresponding cohort where the ability is present. It is the calculation of these latter flows that is the original contribution of this work. We will explain how we use census data to derive these quantities in Subsections 2.2–2.5. For simplicity, we will disregard migration and regional differences in the rest of our discussion.

The key quantity we will discuss is not the number of persons of each age, but the proportion of these persons with an L1 or L2 ability in the reviving language. If there are $N^1(t)$ people of age t with ability in the reviving language and $N^0(t)$ without, we denote this proportion by

$$C(t) = \frac{N^1(t)}{N^1(t) + N^0(t)}$$

As long as we are not seeking overall population statistics, but only the evolution of the age-distribution of abilities, we can disregard N in the modeling, and concentrate on C only. Furthermore, since we are not taking account of migration, and hence the linguistic differences between native-born and immigrants, we can also disregard mortality, which presumably is not affected by linguistic ability.

2.2 The archetypical age distribution of ability during language revival

In the first 20 to 30 years following the initiation of a concerted language revival program, we generally find the emergence of a characteristic age distribution of abilities in the reviving language, as depicted in the plot of $C(t)$ in Figure 1a. There are three recognizable regions, reflecting three distinct processes, though in reality the age intervals affected by the processes are not discrete but overlapping. At the right of the figure, we note a pattern indicative of a long period of language loss. The population has a proportion of mostly L1 speakers of the reviving language, but their numbers are higher in the older age groups, and decline with each younger group of adults. At the left of the figure, the 0–4 age group shows a non-negligible proportion of children with (presumably L1) abilities for the most part, 2–4 years old, or expected abilities, 0–1 year old, though there may be some being raised bilingually. Usually, this transmission-stage proportion will be somewhat less than the average proportion of adults 20–49 with abilities in the reviving language.
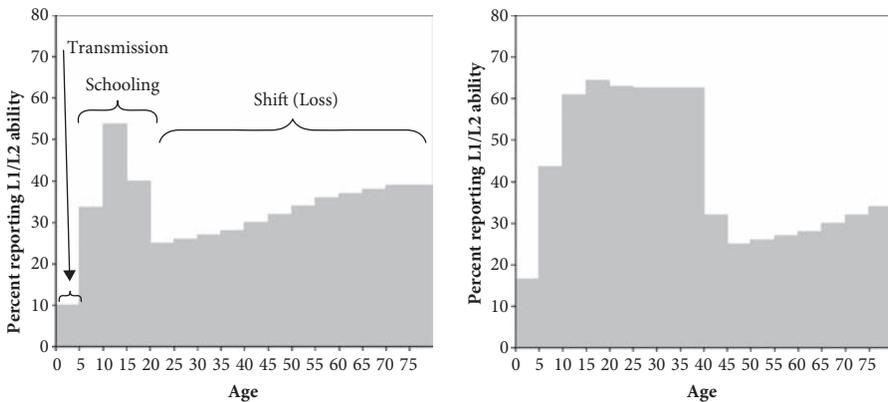
Figure 1. a. (left) Archetypical age distribution of ability during language revival. Gray bars represent (hypothetical) observed data from five-year cohorts. Age spans pertinent to transmission, school acquisition and historical language loss are indicated. b. (right) Projected age distribution after 25 years.

Finally, and most consequentially for the projection, there is the school-age population, 5–19, which shows an initial sharp increase in ability, leveling out or even dropping somewhat toward the end of this period. The increase over the preschoolers can be assumed to consist largely of L2 speakers.

Two quite different exemplars of this pattern, which I drew from data maintained by Casesnoves Ferrer, appear in Figure 2.

Based on the kind of data shown in Figure 1a alone, we can project how the age distribution of ability will evolve over time. For example, we can predict that in 25 years time, the distribution will have changed to be that in Figure 1b. We will describe in Section 2.6 how we do this calculation.
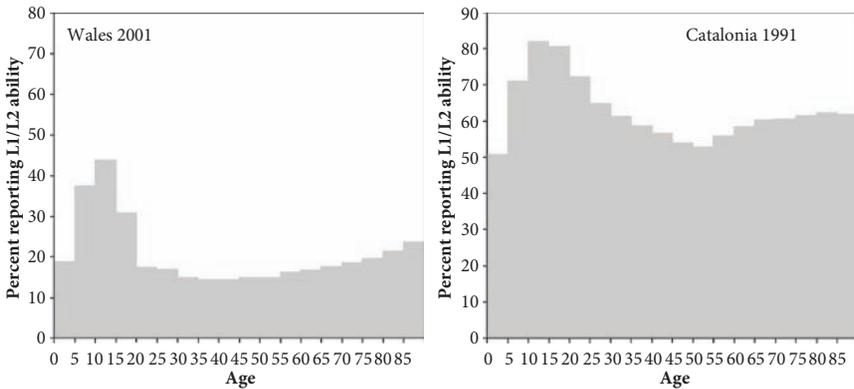
Figure 2. Two typical examples of ability age curve.

2.3   Language shift before revival

Since in our simple model, there is no pattern of adult acquisition nor "forgetting" a language, there is no necessary connection among the $C(t)$ for various values of t, except that in the oldest age groups, generally

$$C(t) \leq C(t+1).$$

2.4   Transmission

Though it is possible to construct elaborate models for transmission rates based on mother's language, father's language and other factors (Casesnoves Ferrer & Sankoff 2004b), for projection purposes it suffices to consider

$$C(0) = \gamma \frac{\sum_{t=t_1,t_2} C(t)}{t_2 \ t_1 + 1},$$

where $[t_1, t_2]$ represents the range of parental ages, reflecting the dominant linguistic input to children 0 to 4. For example, when the data comes in five-year age groups, t_1 might be 15, and t_2 might be 49.

The constant γ is characteristic of the community, and may vary considerably from community to community. It is generally, but not always, less than 1.

2.5   L2 Acquisition in school

The prominent feature of the archetypical age distribution of ability is the rapid rise in ability from transmission levels during the 5–9 school age, with ability falling off in the 15–19 group, sometimes earlier, in the 10–14s and sometimes later, in the 20–24s. How much it rises, usually to a peak in the 10–14s, is a crucial predictor of success or failure of the revival program.

In modeling this pattern of acquisition, we wish first to take account of the relatively rapid initial rise in ability of the 5–9s, and the slower pace in the 10–14s, and the continued slowdown, at best, in the 15–19s. Mathematically, to be able to generate a range of learning curves of this nature requires two parameters, one to account for the initial learning rate and one to determine how this rate slows down.

Figure 3 shows how two very different patterns can be fitted by two curves in the family represented by the model:

$$C(\text{age } t) = C(\text{age } t - 1) + L(\text{age } t),$$

where the C is the ability level at age t and L is the learning increment during year t, measured by the increase in the proportions of the population from age $t - 1$ to age t who report the ability to speak the language, and

$$L(t) = ae^{-\beta(t-5)}.$$
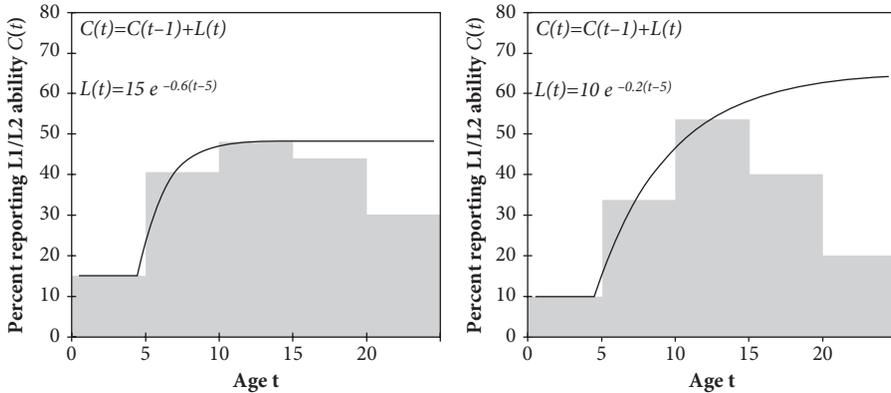
Figure 3. Different learning curves from the same two-parameter family. Gray bars represent (hypothetical) observed data. Parameter values for curves chosen to fit ages under 15.

But we also wish to account for the usual drop-off in abilities, sometimes in the 20–24s, but often in the 15–19s, or even the 10–14s. To model this we make two assumptions. First that the ability level C^* in the age group older than the group where the drop-off occurs represents a sort of background level, unaffected by the improvements in L2 teaching newly instituted during the language revival program. And second, that there is an effective start date for the improvements, so that at least one (the oldest of the five ages) did not benefit from the improvements during the early years. To correct for this, we redefine C starting with the youngest age τ that missed some improved teaching in the early years as follows:

$$C(\tau) = C(\tau - 1) + L(\tau) - L(5),$$
$$C(\tau + 1) = C(\tau) + L(\tau + 1) - L(6),$$

and so on, until the oldest age in the group has been calculated, with the proviso that no $C(t)$ in this group drops below C^*.

Adding the effective start date, whether or not this corresponds to some actual administrative event or is just a parameter representing an initial wearing-in stage, allows us to fit the entire schooling period, as shown in Figure 4.
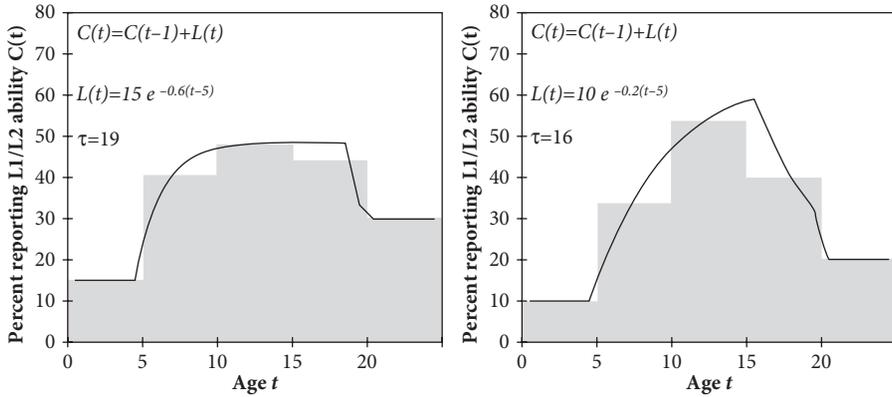
Figure 4. Curves from Figure 3 modified to account for entry into school before the effective start date of the revival program, affecting ages τ and over.

2.6   Inferring and imposing the increments

In the projections, such as the one that produced Figure 1b from Figure 1a, the linguistic abilities in the population are tracked automatically during the aging process. In other words, for each age $t > 19$ in year y, as well as for ages 0–3, we have $C_{y+1}(t + 1) = C_y(t)$.

In addition, when newborn children are assigned age $t = 0$, only a proportion $C_y(0)$ of them, determined by the transmission parameter, described in Section 2.4, are assigned linguistic ability in the reviving language. Of course newborn infants do not know specific languages, but since we are constrained to eventually compare our results with data in 5-year cohorts, with no distinction between age 0 and age 4, say, we simply mean that these newborn are statistically destined to have this ability.

In the treatment of school-age children it would seem reasonable to use the learning increments $L(t)$ to transfer a proportion of each school-age cohort without language ability to the corresponding component with the ability.

Sociolinguists, however, will be familiar with the principle that it may not be appropriate to combine probabilities by adding or subtracting them, partly

because this risks producing probabilities less than zero or greater than 1. Instead, we use the age distribution C, based on the previously estimated parameters, α, β, τ to calculate an increment $l(t)$ by solving

$$\left(\frac{C(t)}{1-C(t)}\right) = \left(\frac{C(t-1)}{1-C(t-1)}\right)\left(\frac{l(t)}{1-l(t)}\right)$$

$$\left(\frac{l(t)}{1-l(t)}\right) = \left(\frac{C(t)[1-C(t-1)]}{C(t-1)[1-C(t)]}\right)$$

$$= X$$

so that

$$l(t) = \frac{X}{1+X}.$$

Then, during the yearly calculations, we combine $C_{y-1}(t-1)$ with $l(t)$ to produce $C_y(t)$, by multiplying

$$\left(\frac{C_{y-1}(t-1)}{1-C_{y-1}(t-1)}\right)\left(\frac{l(t)}{1-l(t)}\right) = \left(\frac{C_y(t)}{1-C_y(t)}\right),$$

for $t = 5, 6 \ldots, 19$.

Unlike the additive increment $L(t)$, the multiplicative use of the odds ratio $l(t)/1-l(t)$ or, equivalently, the additive use of the logit of $l(t)$, assures us that as $C_y(t)$ increases with y and approaches 1, the effect of a fixed value of $l(t)$ will become correspondingly smaller, so that $C_y(t)$ can never exceed 1. This is a principled way of modeling a learning effect $l(t)$ that depends only on age t and learning environment, i.e., the community, and does not change as increased transmission produces more entry-level children with abilities in the reviving language.

Note that after each year τ is increased by 1, so that the curves in Figure 4 eventually give way to the curves in Figure 3. Note as well that $C_1(t)$ as calculated this way is the same as $C(t-1) + L(t)$, but this is only for $y = 1$ and not subsequent years.

In sum, to produce Figure 1b, showing C_{25}, from Figure 1a, showing $C_0 = C$, we simply calculate C_y from C_{y-1} successively for $y = 1, \ldots, 25$, using

$$C_{y+1}(t+1) = C_y(t),$$

for all ages t except for $t = 0$, where

$$C_y(0) = \gamma \frac{\sum_{t=t_1,t_2} C_{y-1}(t)}{(t_2 - t_1) + 1},$$

and for $t = 5, 6 \ldots, 19$, where

$$\left(\frac{C_y(t)}{1-C_y(t)}\right)=\left(\frac{C_{y-1}(t-1)}{1-C_{y-1}(t-1)}\right)\left(\frac{l(t)}{1-l(t)}\right),$$

and the $l(t)$ are adjusted for the first few years to account for the start date.

The very different shape of the curves in Fig. 1a and b suggest the sensitivity of the projections to the details of the acquisition model, particularly for the school-age cohorts. This model is very specific to the context of a language revival process that has been initiated in the last 25 years or so. Where school acquisition rates are less dramatic, variation in transmission rates can also be crucial to long term projections.

3.   Closing the circle

After getting in on the ground floor, if not of the first census of Papua New Guinea, at least of an adjoining and contemporaneous edifice, it would only be appropriate to see if there is any possibility of applying our demolinguistic methodology to current data from that country. The modern census of Papua New Guinea includes a question on literacy in Tok Pisin, English, Motu and "other" languages. None of this involves language revival, but the demolinguistic approach is also pertinent to the very dynamic language situation over the past decades.

I contracted the National Statistics Office of Papua New Guinea (2007), which has responsibility for the census, to prepare cross-tabulations of literacy with age, sex and rural versus urban residence from the 2000 census data. Transforming the raw data into literacy rates produces the results in Figure 5. Literacy data are available only for ages 10 and over. Since literacy is undoubtedly close to zero for age 5, we estimated the literacy for ages 5–9 as half that for the 10–14 cohort, which is likely an overestimation. Unfortunately literacy rates cannot have much meaning for the 0–4 age group, which means that we cannot estimate any analog of the transmission rate.

A glance at these figures reveals a very different pattern from that in Figure 2. While there is of course an increase in literacy in school-age cohorts, this does not necessarily parallel in any predictable way an increase in the knowledge or use of the spoken language. Moreover, an increase in literacy in these age groups also occurs in monolingual communities, and has no necessary connection with bilingualism. There is no indication of a starting point characteristic of a relatively recent revival process, a key feature of our model. And there is no evidence of the gradual loss of a traditional language in the older generations.
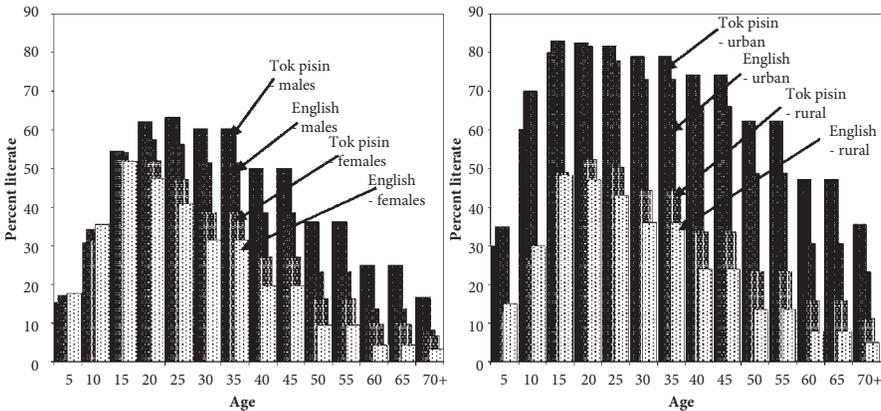
Figure 5. Literacy in Tok Pisin and English, by age group.

The facts that we are essentially dealing with the written language and not the spoken one, and that there are none of the historical indices of language revival, mean that the model developed in Section 2 is not appropriate for the Papua New Guinean data. Nevertheless, the situation is clearly dynamic, based on the changing abilities as a function of age, and we can also discern clear effects of other demographic factors, sex and rural versus urban residence. Thus there is definitely scope for developing demolinguistic models dedicated to this kind of multilingual context.

Here we can only take the first steps towards such a model, by extracting the demographic correlates of the dynamic tendencies.

The relatively small urban population (13%) has dramatically higher literacy rates in all age groups over the majority rural population, while adult males are more literate than females, a tendency which disappears in the school-age population.

Dynamically, we see a rapid buildup of ability in Tok Pisin up to about 40 years ago in males, followed by relative stability, while female rates have continued to rise. English has lagged behind during all this time but is now overtaking Tok Pisin in the school-age population.

Lacking any causal connection between literacy acquisition by children and literacy rates in the adult population at the same time, we cannot model a variable input of literacy levels at early ages, in analogy to the transmission model discussed in Section 2.4.

Any modeling of these data, however, would have to take into account the observations in Figure 5 that acquisition continues later for Tok Pisin than English, suggesting that there is a greater tendency to acquire literacy in Tok Pisin as a young adult, presumably in the work place, whereas English acquisition is more

a function of formal schooling. Acquisition in literacy falls off faster for adult females than males, and faster in rural areas than in the urban context, again likely a function of the use of English and Tok Pisin in the predominantly male urban work environment.

Short of data from the earliest cohorts, and limited to literacy data, our ability to make quantitative predictions is constrained, as these considerations illustrate. Whether the differences in literacy rates correlated with rural versus urban residence and with gender will persist, and for how long, cannot be objectively calculated, though the statistics on the youngest cohorts give us some reason to believe that the gender differences are being attenuated.

Although we have omitted it from Section 2, our DMLX software already handles gender differences, regional variation such as rural-urban differences and migration internally, within a country, and externally. With respect to acquisition, however, it assumes that transmission rates are strictly proportional to the regional prevalence of the language (perhaps not pertinent outside the domain of language shift and revival), and it assumes L2 is acquired either in school or through integration of immigrants, but does not account well for the diversity of patterns of adult acquisition. Most important, aside from the relatively sudden implementation of a language revival program, it takes no account of the long-term evolution of the political, social and educational environment, which is key to understanding the levels of knowledge, use and literacy in Tok Pisin and English in Papua New Guinea. Because this evolution is unique to each country, models should allow for the incorporation of specific histories. This would be a prerequisite for a demolinguistic model of multilingualism for Papua New Guinea that could account for the data in Figure 5, and could project it into the future.

4.   Discussion

The application of demolinguistic methodology depends on the availability of census data, including one or more questions about language. Because of the amount of data in a census, crosstabs of three or four variables are quite feasible. This is essential as a starting point for projections that take into account region, migration, sex and age, all of which can underlie great variation in linguistic abilities.

The vitality of a language depends not so much on how many people know it, which can be derived from the census, than on how much they use it in diverse contexts. Having lived, albeit briefly, in a rural area and in Port Moresby, some forty years ago, helps me little in trying to imagine a pattern of extensive usage of English hinted at by the 2000 literacy statistics. Usage surveys, where they exist,

tend to involve small samples, so it becomes a challenge to try to make some connection between usage studies and census-based demolinguistics.

My interest in quantitative approaches to bilingualism and multilingualism, first stimulated by being born and raised in Quebec, and by collaborating on data and methodologies with Gillian during her studies in anthropology and linguistics at McGill, has taken many turns. On the way to New Guinea in 1966, I stopped in Ireland (traveling to the Gaeltacht), Wales, India and East Pakistan (now Bangladesh), all areas of great bilingual or multilingual interest, meeting with legislators, language planners and scholars, and collecting documentation. More recently I have worked on L2 acquisition (Luján et al. 1984) and code-switching and borrowing, both empirical (Sankoff et al. 1990) and theoretical (Sankoff 1998) aspects. Crossing my own path, as it were, in this present exercise, has been particularly thought-provoking.

Acknowledgements

References

Casesnoves Ferrer, R. & Sankoff, D. 2003. Identity as the primary determinant of language choice in Valencia. *Journal of Sociolinguistics* 7: 50–64.

Casesnoves Ferrer, R. & Sankoff, D. 2004a. The Valencian revival: why usage lags behind competence. *Language in Society* 33: 1–31.

Casesnoves Ferrer, R. & Sankoff, D. 2004b. Transmission, education and integration in projections of language shift in Valencia. *Language Policy* 3: 107–131.

Casesnoves Ferrer, R., Sankoff, D. & Turell, M.T. 2006. Linguistic shift and community language: the effect of demographic factors in the Valencian Region, Balearic Islands and

Catalonia. In *Globalization and Language in the Spanish-speaking World*, C. Mar-Molinero and M. Stewart (Eds), 197–219. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

Casesnoves Ferrer, R. 2007. Prediction of the future efficiency of several language-revitalization programs in Spain and the UK. 11th International Conference on Minority Languages, Pécs, Hungary.

Luján, M., Minaya, L. and Sankoff, D. 1984. The universal consistency hypothesis and the prediction of word order acquisition stages in the speech of bilingual children. *Language* 60: 343–371.

National Statistical Office of Papua New Guinea. 2007. http://www.nso.gov.pg/

Sankoff, D. 1998. A formal production-based explanation of the facts of code-switching. *Bilingualism, Language and Cognition* 1: 39–50.

Sankoff, D., Poplack, S. & Vanniarajan, S. 1990. The case of the nonce loan in Tamil. *Language Variation and Change* 2: 71–101.

Scragg, R.F.R. 1957. Depopulation in New Ireland: A study of demography and fertility. Port Moresby: Printed under the authority of Minister for Territories, Territory of Papua and New Guinea. Also cited as Department of Public Health Monograph No. 3.

How local is local French in Quebec?

Pierrette Thibault
University of Montreal

Looking at the use of French in Stanstead, a small bilingual community located at the border of Quebec and Vermont, several phonological and morphological variables (*ne* presence, verbal ending neutralization, /p/ aspiration, and (R)) are analyzed, using a comparative approach. First language French speakers from Stanstead are compared to Montreal French majority speakers, and to Ontario French minority speakers. Second language speakers of French from Stanstead are compared to English-speaking Montrealers speaking French. The possible influence of English phonology on native French is also explored, with reference to earlier work carried out in Sherbrooke, the largest city of the Eastern Townships, the region Stanstead is part of. Attitudes towards bilingualism and code-switching are investigated as possible characteristics of local identity.

1. Introduction

In Quebec, contact between French and English speakers mainly occurs in the Montreal Metropolitan area. However, along the border between Quebec and Vermont, place names attest to early American settlement: Huntingdon, Bolton, North Hatley, Knowlton, Hemmingford, Stanstead. Over time, French speakers have moved to these parts that were on "*Québécois,*" hence Canadian, soil. The area is called the Eastern Townships and Sherbrooke is its capital city. Starting in 2001, we spent three consecutive summers in Stanstead,[1] a small town located at the border between Vermont and Quebec, as part of a multidisciplinary research team

---

1. As of October 2007, Sherbrooke's population is 147,601; Stanstead's is 3,178, according to the Quebec Towns' Directory: http://www.mamr.gouv.qc.ca.

which set out to characterize the community in terms of its history, environment, and language use.[2]

As anthropologists and sociolinguists, our sub-group focused on a few key questions: Is bilingualism a defining factor of local identity? Is local French tinted by its extended contact with English? To what extent do French speakers behave like the French majority in Quebec, and to what extent do they display minority French behavior?

Three Master's research projects have dealt with some of these questions (Ringuette 2004; Lacasse 2005; Graml 2005). After commenting on their findings, I will introduce additional data pertaining to the same issues, and I will look at community language use from a regional dialect perspective. This discussion will be preceded by a general introduction to the community and its history, from a sociolinguistic perspective.

A little note to end this introduction: In Quebec, as in other parts of Canada, sociolinguistic research has expanded from a wealth of earlier careful empirical work, spearheaded by Gillian Sankoff, David Sankoff and Henrietta Cedergren's sociolinguistic research on Montreal French, conducted in the early seventies. Since then, there have always been paths to follow and to depart from, for Canadian researchers have had golden opportunities to compare new data with well-documented earlier findings.

2.   Language use at the border

Contrary to most Quebec communities where English and French speakers have come to settle, Stanstead stands out in three respects: the English speakers were there first, and they have always dominated numerically until recently (the current population is about half English and half French-speaking), and there is no geographical or economical divide between the two groups.

The people of Stanstead are bilingual in general, but as expected in Quebec, French speakers tend to be more proficient in their neighbors' language than the reverse, although such a generalization does not hold when each age group is

---

considered separately. One fact is clear: bilingualism within the community stems from instrumental and economic benefits which are available to bilingual locals, rather than from contact between American and Canadian citizens.

Traditionally, there has been at least superficial contact across the US/Canada border because of Canadians going to shop in New England. But by 2001, even before September 11, when we started our fieldwork, it was no longer cost-effective for Canadians to cross the border to buy gas or any other goods, because the American dollar was worth a lot more than the Canadian.[3] However, many American license plates could be seen in Stanstead, especially at the supermarket, at the doctor's, and at the drugstore.[4]

The remainder of this section sketches the linguistic context in which our fieldwork was conducted, and describes the domains of French and English, in the past and in the present. In the 19th century, Stanstead's main street displayed at least four huge mansions and a very select private college. At the turn of the 20th century, an opera house opened down the hill from the main street, and by the end of the century, one of the mansions had been turned into the Colby-Curtis Museum, where the Colby family's belongings and archives are a big part of the permanent collection.

Our team looked at one member of the Colby family's personal correspondence,[5] as well as the early issues of the *Stanstead Journal,* the oldest weekly publication in Quebec, for mention of French people or the French language.[6] In the close to a hundred letters we looked at, there are a few mentions of French names, a reference to French courses Emma Colby attends, and her allusion to a local politician's speech in French. She also uses the French *fini* for "finished" or "closed" ("That subject may be classed with the *finis* ones."), and she jokes about the inappropriate use of the French word for "the flu" ("Henry has the *la grippe* as I once heard it called.").

The Stanstead Journal, a local English-language newspaper, was first published in 1845. In an 1853 issue of the *Stanstead Journal*, an entire article in French is published under the heading *Municipalité de Stanstead* ('Stanstead Municipality'). Apart from

---

3.   This situation no longer prevailed in the fall of 2007.

4.   One retired doctor was very popular with American patients, medical care being much cheaper in Canada, and of course the local drugstore benefited from the prescriptions he wrote for these patients.

5.   Emma Frances Cobb Colby (1895–1898), an American-born woman who married into the Colby family. She and her husband would eventually move to Montreal.

6.   Julie Bélisle and Annie Binette thoroughly examined the early *Stanstead Journal* issues as part of their research on the history of Stanstead, under the supervision of my colleague, Laurier Lacroix. I thank them for indicating to us any use of French in the *Journal*.

that, the French presence is only reflected in the advertisement section. More than a century later (in 2001), when the editor evoked the possibility of including French articles in the *Journal*, he received a lukewarm response from the readers.

Overall, there appears to be a clear social divide between the "bourgeois" English community and the adjacent local rural French settlements, at least until industrialization hit Stanstead. After the end of World War II, Stanstead boomed. Several small industries were operating, mostly related to the granite industry, but Butterfield, a large tool-and-dye factory, became the major employer. The Butterfield factory, which has an American section on the other side of the border, attracted a lot of French Canadian workers, who came from as far as the Beauce area, closer to Maine than to Vermont. By then, the wealthy American-born residents had sold their mansions, and both the English[7] and French populations were mainly working-class. Unfortunately for Stanstead, in 1981, Butterfield transferred its Canadian operations to Ontario. This caused a major decline in the population, many young workers having decided to follow the company. The granite industry gradually took over as the major employer, and it remains so.

In the workplace, as well as in shops, English seems to have always been the favored language. During the summer of 2002, we attended an evening softball game. All the instructions yelled to the players on the field by their coach during the game were given in English, while most of the chatting on the bench was done in French.

In Stanstead, as everywhere else in Quebec, all public signs are in French or bilingual, as required by Quebec law. Local residents are addressed in their native tongue in shops and offices, but there is a strong tendency to use English first with visitors from the outside, who are predominantly American.

3.   The way they speak: Sample and method

The first people our research team met in Stanstead were the young people holding summer jobs at the Colby-Curtis Museum. With their cooperation and through their networks of friends and acquaintances, we got to interview twenty-two people, aged from 15 to 26 in 2001.[8] Not only did we discover a thoroughly bilingual

---

7.   The origin of the working-class English-speaking community can be traced to the Loyalists who settled in the Eastern Townships. Irish immigrants also found their way to Stanstead, as a statue of Saint Patrick in the main Catholic church suggests.

8.   The interviews were recorded on digital audio tapes. They have recently been transferred to CDs.

generation of speakers, whose first language we were able to determine only after several minutes of interaction, but we found that the French and the English youth hung out with each other quite freely. Their favorite hangout spot was a rather concealed private beach, by Lake Memphrémagog, which extends for some thirty miles along the border between Quebec and Vermont. This cohort constitutes almost half of our total sample (22 out of 46).

In 2002, we interviewed the generation corresponding more or less to their parents. Fourteen people were interviewed, and their ages range from 37 to 64, which is considerably broader than the younger generation. Two of the interviews were conducted with couples.

We had kept the older generation for the last summer of our project, with the hope they would have heard about our research team by then.[9] We managed to record nine of them, ranging from 66 to 87 years old.[10]

3.1   The Stanstead sample

Some of the speakers we interviewed in Stanstead are clearly from a French-speaking background; others grew up in an English-speaking environment, while a few come from a mixed family and call themselves bilingual. For some of those who claimed English as their mother tongue, we have two interviews, one in their second language and one in their mother tongue.

Three individuals claimed a bilingual identity: two among the younger group, and one in his early sixties. The first two had a totally French schooling background. The third one, however, had married an American, and he had mainly spoken English throughout his adult life. He had a slight English accent, and he made some of the typical gender agreement mistakes characteristic of second-language speech. For the purpose of the studies we have carried out so far, the first two were labeled "French speakers," and the last one an "English speaker".

Thirty-six interviews were conducted with predominantly French-speakers, 15 women and 21 men. However, one of the men, an 87-years old, was born and lived his youth on the American side of the border. Nine interviews involved predominantly English speakers speaking French. Four speakers only spoke English during the interviews: In three cases, the interview was part of a follow-up from

---

9.   In 2002, 2003, and 2004, some members of our team put up exhibitions at the Colby-Curtis Museum. In 2003, the exhibition themes focused on granite and the border. Valérie Maridor and Anne Ringuette, my two research assistants, contributed pictures and excerpts from our interviews (with the agreement of the interviewees).

10.   As the details of the sample will show, three speakers interviewed in 2002 really belong to the older cohort of speakers, being 63 or 64 years old.

the French session. The fourth one participated in a bilingual interview, in which her husband mainly spoke French. Finally, among the couples interviewed, one spouse came from France. All of the interviews conducted in French have been transcribed.

So far, not a single study has drawn on the same speakers' data as input. Therefore, when reporting on each study, I will specify the number and social characteristics of the speakers whose speech was considered.

3.2 A comparative approach

The multidisciplinary team of which we were an active part, along with art historians and visual artists, brought the local identity theme to the center of our collective research quest. The border and the granite industry bore heavily on the general local picture, but from a dynamic perspective, the sociolinguistics and anthropology trio I was part of, with my two graduate students, chose not to consider the possible influence of New England English on local speech for two main reasons: The clear divide between the bourgeois American settlement families and the local population, and the lack of close contact in recent history.[11]

Rather, as stated in the introduction, we set out to compare the use of French and of bilingual speech in Stanstead with Montreal French, both as a first and second language, with Ontario French minority speech, and with Sherbrooke speech, with respect to some regional features. However, in none of the communities we wanted to compare Stanstead with, did researchers find such fluency on the part of most young people from both language groups in each other's mother tongue. Either the English speakers didn't speak much French (as in Ontario), or a minority of the French speakers were bilinguals (as in Montreal). Bilingualism, we thought, might be a strong defining feature of local identity in Stanstead, especially among the young people. This is the hypothesis Anne Ringuette (2004) set out to explore.

4. The emergence of a bilingual community

The American settlers, who are mentioned in Section 2.1, do not seem to have learned French, given their lack of contact with the predominantly rural local population. The industrial boom, however, brought about opportunities for the French speakers of Stanstead to learn and use English. This fact was clearly

---

11. However, Stanstead might be a location for scholars specializing in the diffusion of American dialects to explore around a core question: How Canadian is Stanstead English?

acknowledged by our older interviewees (66 to 87), who had experienced isolation from the English-speaking community in their early lives, stemming from their rural and Catholic background.

Among the 17 (French and English) speakers who were between 37 and 53 at the time of the interview, 9 claimed dual citizenship, being born in Newport, Vermont, as part of a US-Canada agreement. This situation occurred before a highway gave easy access to Sherbrooke and Magog hospitals, and following the home-delivery era, which at least three of our interviewees claimed to be a product of. However, there is no clear indication that they speak English more than their elders. Thus, it appears that generalized bilingualism is a characteristic of the younger generation in Stanstead.

4.1   Bilingualism and code-switching

During the interviews, several questions explored the history of contact with English within the interviewee's family, as well as through his or her own childhood and school years. Although there were no specific questions about code-switching between French and English, 15 out of 22 young speakers, as Ringuette found out, acknowledged its widespread use among their peers.[12]

However, there is hardly any trace of English in the French interviews, apart from quoting something in English. This is reminiscent of the Ottawa-Hull interviews conducted by Poplack in the 80s (Poplack, 1989), and is probably related to normative pressures towards the use of unilingual speech in an interview with an outsider.[13]

As a first step in her quest for a language-related component of local identity in Stanstead, Anne Ringuette (2004) looked at opinions pertaining to the use of code-switching in the young people's interviews. Analyzing all the relevant statements (some of which came from the same speakers, in the course of different speech turns), she found that an overwhelming number of references to code-switching (20) described it as a usual way of speaking. In sum, the young people of Stanstead tend to consider code-switching as normal local behavior.

---

12.   The young people cohort of 22 includes six English-speaking, of whom three acknowledged using code-switching, while the others made no mention of such a practice. Among the 16 French-speaking, one denied switching between languages, and four made no reference to such a behavior. (from Ringuette 2004 Table 3.2, p. 38).

13.   Such an interpretation was confirmed when we asked two of our young interviewees to record spontaneous conversations with their friends. Extensive code-switching was observed during those two conversations, which have not been transcribed yet.

Ringuette took a further step in exploring the status of both bilingualism and code-switching, as defining speech features of local identity. She developed and conducted a reactions test with high school students in Stanstead.

4.2   Recognizing local speech

Ringuette (2004) developed a test inspired by Lambert et al.'s (1960) procedure designed to uncover attitudes towards Quebec French, and modified along the lines of Thibault and Sankoff's more recent experiment (1999). Given the level of bilingualism observed among the young people of Stanstead, she wanted to find out whether high school teenagers would identify more readily with French speech that was totally monolingual, speech that only displayed one-word switches, or speech in which sentence segments alternated from one language to the other.

In order to follow, as closely as possible, Lambert's matched-guise procedure, which involved actors playing out different accents, Ringuette took three excerpts each from two of our Stanstead interviewees, assigning them different names, and mixing them. They were from the only male and female who were recorded in conversations in which code-switching occurred (see note 13).

Thirty-four students, registered in Stanstead's only high school (a relatively low-fee private French school) participated in the experiment. Their ages ranged from 14 to 18. Most were French speaking but the group included some English-speaking bilinguals.

The first question of the test asked the judge to guess the place of origin of the speaker. The choices were: Montreal, The Eastern Townships, Stanstead, and somewhere else in Quebec. For the analysis, these choices were organized according to a proximity to Stanstead scale.[14] As expected, the students tended to associate monolingual speech to others parts of Quebec, code-switching to Stanstead, excerpts containing a single word switch were assigned a middle position on the proximity scale. Such an association was statistically significant.

Other questions explored whether or not different people in the judge's environment (brother, father, teacher) spoke like the person on tape, and in which context. Positive or negative values associated with different speech types were also investigated through questions, such as, "Would you like your children to speak like X?" and "Do you speak like X sometimes?"

---

14.   Wherever the variables could be organized in scales, regression analyses were performed. Otherwise, ANOVA tests were conducted. Analyses were also performed using the mean values for the group, as is the case here.

Ringuette had expected the self-declared bilinguals[15] among the judges (26 out of 34) to associate code-switching, rather than monolingual speech, with their family members, and to view such a practice more positively than the students who labeled themselves either French-speaking (4) or English-speaking (4). Most of the significant results she obtained, however, led to mixed interpretations.

For example, let's consider the question: "Where do you think speaker X mainly speaks the way he does?", to which the judge was asked to answer by choosing between: everywhere, at work, in class, in the school yard, at home, or with friends. Ringuette transformed the answers into a hierarchy of formality associated with each context. She thought that both monolingual excerpts would tend to be linked to the work place and the school, while the excerpts displaying code-switching would be associated with the home, and with friends. However, some excerpts were not judged as expected.

Her results seem to indicate that a normative judgment filtered through the answers. Excerpt No. 1 is an example of monolingual speech on the part of the woman. It reads:

(1)   *Je pense qu'elle avait vingt-et-un ans quand qu'elle est partie en appartement avec son chum.*[16] *Ça faisait sept ans qu'ils étaient ensemble.*

I think she was 21 years old when (that) she moved to an apartment with her boyfriend. They had been together for seven years.

The expression *quand que* is a wide-spread non-standard form. It might be slightly stigmatized, especially in the school context, where the test was performed. This could explain why the excerpt could not be linked to the school context in the judges' mind.

Excerpt No. 2 is an example of code-switching in the man's speech.

(2)   First thing I know five o'clock arrive *puis là, tu sais des fois ils viennent m'aider mais souvent là tout le monde fait ça: Bye!*

La première chose que tu sais cinq heures arrive and then, you know sometimes they come and help me but often everyone says: Bye!

---

15.   The question they were asked was: "Do you consider yourself an English speaker, a French speaker, a bilingual, or other? Three of them chose the label: English-speaking bilingual. They are classified as English-speaking.

16.   *Chum* is the most common word used to designate a boyfriend in Quebec French. It is therefore a totally integrated loanword. The feminine correspondent word is *blonde* ("girlfriend").

In this case, a morphological mistake occurs in the English segment, where the "s" is not heard at the end of the verb "arrive". Moreover, the initial consonant in *thing* is pronounced as a dental stop. Such apparent lack of competence in English might have prevented most judges from associating the excerpts with their family and friends.

Those who chose to call themselves bilinguals, rather than French or English speakers (26 out of 34, a large majority), revealed yet another aspect of local norms, which might be worth looking at in the future. They behaved differently from the others while answering the question: "Do you ever speak like X?", to which the listed answers were: never, sometimes, or often. They acknowledged code-switching less often than their English or French peers, both for the type of speech exemplified by the excerpts containing the one-word switches (*bills* and *mortgage*), and the sentential type of code-switching displayed in excerpt No. 2, where mistakes are observed.

They more readily admitted, however, to speaking the way the woman does in excerpt 3:

(3) *Je me sentais super mal toute la journée après avoir dit ça* so I called her back, I only said I was sorry.

I felt super bad all day long after having said that *ça fait que je l'ai rappelée. J'ai juste dit que j'étais désolée.*

For the excerpts containing English words or segments, the judges were asked through a multiple choice set of possible answers why, in their opinion, the speaker on tape was using English words or speech segments. The most frequent answer, by far (between 19 and 21 of 34), was that it was his or her usual way of speaking. The answer: "She or he is bilingual" was chosen 23 times for excerpt No. 3, and 20 times for excerpt No. 6. Excerpt No. 2 containing the word *mortgage*, and excerpt No. 5 where *bills* appears, were associated with bilingualism only 12 and 7 times, respectively. "Laziness", "lack of knowledge of the French equivalent", and "addressing an English speaking person", were the least often chosen answers (< 10 mentions for all excerpts).

Ringuette's research confirmed that, at least for the bilingual people, who form the large majority among the high school students she reached for her study, code-switching is linked to bilingual competence, an observation well documented in Poplack's research in the New York Puerto Rican community (Poplack 1980). Both her content analysis and the reactions test she conducted tend to show that code-mixing or code-switching are viewed as normal speech behavior in Stanstead. Competence in the two languages is definitely valued by the young people of Stanstead, even when they alternate in short speech segments. The status of single-word switches is not so clear, and it should be looked

at more closely in communities in which normative pressures towards monolingual speech prevail.

In the remaining sections of this paper, we will look at linguistic variables in the speech of Stanstead people of all ages, from a comparative perspective.

5.   Whose language variety is Stanstead French closest to?

Before undertaking our research project, we were quite excited with the prospect of putting to the test some of our intuitions about the weight of (1) the official status of French in Quebec, particularly enforced in the school system, on the language of Stanstead French speakers, and (2) their local status as a linguistic minority in a relatively isolated small community.

For her Master's thesis, Chantal Lacasse (2005) analyzed four morphological variables in the speech of all the individuals recorded in Stanstead. Her choice of variables was determined by previous work done on Montreal majority French (the Sankoff-Cedergren corpus (Sankoff et al. 1976), on English-speaking Montrealers speaking French (Sankoff et al. 1997; Blondeau et al. 2002), and on Ontario French minorities (Mougeon & Beniak 1991, and Mougeon & Rehner 1999, principally).

She studied the distribution of two variants reputed to be associated with stylistic variation, namely the use of *ne* in negative sentences, and the recourse to *nous*, as a first person subject pronoun in the plural. She then looked at two morphosyntactic variations: the use of *avoir* ('to have') versus *être* ('to be'), as auxiliaries in the perfect tenses, and the neutralization of the distinction between the 3rd singular and the 3rd plural endings with certain irregular verbs. In the following sections, I will report on her results from the first and the last of these four variables.One aspect of the language contact situation in Stanstead that surfaced during Ringuette's research is the possible emergence of a bilingual community, through growing interactions between the two language groups. If such is the case, the English-speaking people of Stanstead (especially the younger among them), when speaking French, should be expected to behave even more like their French counterparts than the Montreal English speakers, who speak French in public, but whose personal and social life appears to be conducted mainly in English (Sankoff et al. 1997).

5.1   Negative morphology in Stanstead second-language French

In their study of negation in Montreal French, Sankoff and Vincent (1977) considered a sample of 60 speakers (half of the Sankoff-Cedergren corpus constituted

in the early seventies). They found approximately 10,000 negative sentences, in which *ne* was used 46 times, by 15 different speakers.[17] All the other negative sentences were constructed with *pas* (used for neutral 'not'), *jamais* ('never'), or *plus* ('no longer') alone. Standard French requires *ne* to be placed before the verb and the accompanying negative expression to follow it, as in "*Je ne suis pas fatiguée*" ('I'm not tired'), but even in France, *ne* fails to appear in oral speech, though to a lesser extent than in Quebec (Ashby 1981).

The use of *ne* by one quarter of Sankoff and Vincent's sample of speakers seemed to be related to certain topics, rather than to the social characteristics of the speakers. Therefore, the authors coined *ne* as a stylistic resource available for all community members.

Seventeen English speakers were interviewed in Stanstead, but only eleven spoke French during the interview. Three speakers used *ne*, two of whom produced a single token each. In both cases, the interviewee is talking about someone who doesn't speak French.[18] The third speaker, a 22 year-old woman, is among the least fluent in French.[19] She produced 21 *ne* out of a total of 124 negative sentences, which affords her the lowest percentage of absence of the particle in the whole corpus (83.1% according to Lacasse's Table IV, 2005: 49). In her speech, *ne* seems to be lexicalized since in 20 out of 21 instances, it appears in the expression: *je ne sais pas* ('I don't know').

In their interviews in French, the English speakers of Stanstead used *ne* in 23 of a total of 855 negative sentences, corresponding to a percentage of absence of *ne* reaching 97.3%. As Table 1 shows, this figure is much closer to the Sankoff and Vincent (1977) rate of 99.5% *ne* absence than to the 89% recorded for twenty-one English-speaking Montrealers considered in Sankoff et al. (1997), and Blondeau et al. (2002). Equally of note is the fact that the proportion of Stanstead second-language speakers of French using *ne* (18%) is very close to that of both groups of

---

17. These were what they called productive uses of *ne*, after eliminating occurrences of expletive *ne*, used without *pas*, as in: *ne vous en déplaise* ("should it not displease you"), or cases where *ne* preceded the third person object clitic *en*. In Montreal French, *il en a* is often uttered as [innɑ], and *il n'en a pas* [innɑpɑ], which makes the double 'n' impossible to interpret.

18. Tim says: … "*à cause qu'elle ne savait pas parler le français*" ('because she couldn't speak French'), and Bill mocks his wife, by saying: "*Je ne… je ne peux pas parler français*" ('I can't … I can't speak French')

19. Actually, she speaks French very well, as everyone does in Stanstead, but she tends to hesitate and she has a school-type style of speech.

French native speakers, in clear contrast to the very high proportion of *ne* users among the Montreal English speakers (76%).

Table 1. Comparative use of negative morphemes in Quebec French. Based on Sankoff and Vincent (1977), Blondeau et al. (1995), and Lacasse (2005)[20]

Corpus	# *ne* users / Total # of speakers (%)	# Negative sentences	% Absence of *ne*
Montreal French	15/60 (25%)	10,000*	99.5
L1 Stanstead speakers	6/35 (17%)	3751	99.0
L2 Stanstead speakers	3/17 (18%)	855	97.3
L2 Anglo-Montrealers	16/21 (76%)	2,012	89.0

* Approximate figure

Thus, it appears that with regards to the structure of French negative sentences, Stanstead people behave in a similar way, regardless of their mother tongue. Most of the English Montrealers, for their part, use *ne* in negative sentences, probably owing to school French influence rather than to close contact with their French-speaking neighbors.

5.2 A case of morphological neutralization in Stanstead native French

Among the features studied in Ontario minority French communities, the variable use of 3rd person singular morphology with a plural subject was noted and analyzed by Mougeon and Beniak (1991). Most French verbs sound the same in the 3rd singular and 3rd plural persons, but in the written form, there is a regular ending indicating plurality, as in *ils aiment* ('they like'), as opposed to *il aime* ('he likes'). Referring to the 1980 edition of *Le nouveau Bescherelle* which lists all the verb conjugations, Mougeon and Beniak mention that 600 out of 12,000 French verbs show an irregular ending in the 3rd person plural. Audible differences between the singular and plural forms also characterize the future tense for all French verbs, but its occurrence is very low, given the alternative widespread

---

20.   For this table, I used Chantal Lacasse's raw data found in an appendix to her thesis. In her own tables, she grouped together the data from the very fluent English people with those of the L1 French speakers. As for the data pertaining to the English-speaking Montrealers, I also went back to the raw material to find out how many different speakers used *ne*. In addition, I excluded from the Montreal L2 group one speaker who is a self declared English-speaker but who displays L2 English behavior, making non-native mistakes in English.

future formation with *aller* ('to go')+ infinitive. Here are a few examples of 3rd person oppositions:

3rd **person singular**	3rd **person plural**
Il a ('He has')	*Ils ont* ('They have')
Il aura ('He will have')	*Ils auront* ('They will have')
Il met ('He puts')	*Ils mettent* ('They put')
Il mettra ('He will take')	*Ils mettront* ('They will take')

With *avoir* and *être*, the most frequently used verbs in Ontario French (as in other varieties), such neutralization in favor of the singular form rarely occurs (1% and 3% respectively), but with *mettre* ('to put'), neutralization is found 50% of the time, which is the highest rate recorded in Mougeon and Beniak's data. Thirteen times out of twenty-six, speakers of the four communities studied by Mougeon and Beniak used the singular *met* ('puts') instead of the plural form *mettent* with a plural subject (1991: 97).

Such a lack of subject-verb agreement has been noted for different varieties of French, but only in minority French communities has it shown enough recurrence to be labeled a trend. In Canada for instance, apart from Mougeon and Beniak, Flickeid (1989) and King (1993) have attested to such a tendency in Acadian French from the Maritime Provinces, specifying that it mainly occurs after the relative subject pronoun *qui* ('who').

In the Stanstead corpus, Lacasse (2005) found a very low percentage of neutralization in favor of the singular form (1.5%). Following Mougeon and Beniak's 1995 article, she grouped together the "unrestricted users" of French, among which she included the very proficient second-language speakers of French. In Stanstead, the "restricted speakers" included the six speakers who, though bilingual, were less fluent in French according to Lacasse's classification.

Her results matched those of Mougeon and Beniak in that the "unrestricted speakers" only neutralized the verb endings in 2% of the cases or less, while the "restricted speakers" did it 22% of the time (as compared with 29% for the Ontarians).

Looking at Lacasse's data a little more closely,[21] I noticed that 16 of the 35 French speakers from Stanstead sometimes used a singular form with a plural subject. Since no such widespread behavior has ever been observed in our Montreal majority French corpora, the Sankoff-Cedergren corpus (Sankoff et al. 1976), and the Montreal '84 corpus (Thibault & Daveluy 1989), I decided to submit the data to multivariate analyses, using Goldvarb X (Sankoff et al. 2005). The results appear in Tables 2a and 2b.

---

21.   I thank Chantal Lacasse for giving me access to her raw data.

Table 2a shows the influence of two linguistic factors: the type of verb and the type of subject. Using Lacasse's Table XX (2005: 106), which lists all the irregular verbs found in the Stanstead corpus with their frequency, I grouped under "Other auxiliaries" the following verbs: *aller* ('go'), *faire* ('do'), *venir* ('come'), *vouloir* ('want'), and *pouvoir* ('may, might'). Their occurrence ranges from 261 to 61, while the most frequently found from the "Other verbs" group, *dire* ('say'), appears in 39 relevant contexts.

As expected, the less frequent the verb, the most likely the neutralization will occur (relative weights of .971 and .908 respectively for the future tense and the 45 verbs listed as "Other verbs").

With respect to the subject, the results are different from those observed by Mougeon and Beniak (1995). Given the fact that in Stanstead, as in Montreal, and in Ontario French as well, *ils*, used for both the masculine and the feminine, is pronounced [i] most of the time, the absence of contrast in the singular and plural 3rd person pronouns should favor the occurrence of the neutralization of the verb ending, according to them. Their results for the "unrestricted speakers" show that neutralization only occurs with *qui* ('who'), and *ils*, both unmarked for number or gender.

Table 2a. Analysis of the influence of two linguistic factors on the occurrence of verbal ending neutralization (Input = 0.007)

Factor group	Factor	No. of neutralizations in favor of the singular	Total	% of neutralizations	Relative weight
Type of verb	*Avoir, être* ('To have', 'to be')	10	1919	0.5	.361
	Other auxiliaries	16	547	2.9	.760
	Other verbs	13	194	6.7	.908
	Future tense	1	6	16.7	.971
Subject	*Ils* ('They')	14	1831	0.8	.375
	NP	9	365	2.5	.744
	Qui ('Who')	17	453	3.6	.763
Total		**40**	**2666**	**1.5**	

In Stanstead, we find that *qui* and NPs favor neutralization, while *ils* disfavors its occurrence.[22] Having observed that many NPs corresponded to generic referents, such as "the young people", groups, such as "my family", and to

---

22. In 146 sentences, there was no audible subject. Neutralization occurred in none of these cases. For the analysis, they were grouped together with *ils*.

indeterminate subjects, such as "a few people", I coded half the data for the specificity of the subject referent, a specific referent being "my folks", for instance. This factor group was not selected as significant in the analysis.

Table 2b shows the influence of four different groups of factors on the variable neutralization under study. Men tend to neutralize more than women (.578 versus .392). Older people, on the other hand, are less likely to neutralize than the other age groups. The level of education doesn't influence the variation. Another factor that doesn't seem to enter the equation is the fact that at the interviewee's home, someone is an English speaker.

Table 2b. Analysis of the influence of four social factors on the occurrence of verbal ending neutralization (Input = 0.012)

Factor Group	Factor	No. of neutralizations in favor of the singular	Total	% Neutralizations	Relative weight
Age group	15–26	20	936	2.1	.618
	37–53	15	767	2.0	.606
	63–87	5	963	0.5	.307
Sex	Men	31	1552	2.0	.578
	Women	9	1114	0.8	.392
Schooling	High School	24	1430	1.7	non-sig.
	Junior College	7	958	0.7	non-sig.
	College	9	278	3.2	non-sig.
English at home	Absent	29	1675	1.7	non-sig.
	Present	11	991	1.1	non-sig.
Total		**40**	**2666**	**1.5**	

Thus, the very low rate of neutralization among the French community (1.5%), coupled with the fact that the presence of an English-speaking person at home doesn't influence the variation, indicate that this phenomenon is not related to either the historical numerical minority status of French in Stanstead, nor to the close proximity of English speakers.

Mougeon and Beniak qualified the low level of neutralization found in their unrestricted speakers' speech as *"défaillances épisodiques"* (1995: 53) ("sporadic mistakes"). Lacasse (2005) concurred with their observation, with regards to Stanstead French. However, while she was coding her data for the neutralization in favor of the 3rd person singular form, she noted the converse, that is: neutralization in favor of the plural form, as in the following example:

C'est un ministre de Ayer's Cliff qui vont la marier (Jocelyne)
It's a Minister from Ayer's Cliff who are going to marry her (perform her wedding)

Lacasse recorded 29 cases of neutralization in the other direction, in the French speakers' data. Such vacillation between the two options is reminiscent of a well-documented tendency to alternate between the masculine and the feminine in the assignation of gender to nouns starting with a vowel in most French speaking communities. There is definitely no one-way direction associated with the neutralization of the opposition between 3rd person singular and plural verbal forms.

5.3   Aspirated /p/, a pronunciation foreign to French

In the early seventies, Normand Beauchemin set out to characterize the variety of French spoken in the Eastern Townships, focusing on the Sherbrooke area in particular. His research assistants, mainly using their personal networks, recorded about sixty native French speakers born in the area, 16 years old or older. Given the fact that until the 20th century, the whole area was predominantly English-speaking, Beauchemin (1972) looked at *"quelques traits de prononciation québécoise dans un contexte anglophone qui les influence?"* ("A few features of Quebec pronunciation in an English-speaking environment that influences them?"), the question mark leaving open the interpretation of his results in terms of a possible effect of English on local French.

Aspirated [pʰ], an allophone of the un-aspirated voiceless plosive bilabial /p/, is found in the word-initial position or at the beginning of a stressed syllable in English, but it has not been recorded as a characteristics of any French dialect. Therefore, Beauchemin legitimately thought that its occurrence in the speech of French speakers from the Eastern Townships could be attributed to the influence of English. A list of 25 words, all but one monosyllabic, was given to the speakers to read. The recorded words were transcribed phonetically. A total of 1250 tokens were analyzed (50 speakers x 25 words). Almost half of the initial /p/'s were aspirated (556/1,250, representing 44.5%). The results, displayed in contingency tables, are discussed according to various social groupings. Age did not seem to be a significant factor, but the distribution according to the level of education showed a significant tendency for the least educated group (12 individuals with elementary school background) to aspirate more than the other groups. Another factor that was found to be significant was whether or not one's occupation involves dealing with the public. Those who are not exposed to the public (students, manual workers, farmers, and housewives) tend to favor the use of [pʰ], while their counterparts disfavor its use. Presumably, people who deal with the public at work are expected to use standard French, whereas the others don't feel such pressure.

In Stanstead, some [pʰ]'s can be heard in the spontaneous speech of some French native speakers. My research assistant was instructed to listen carefully to one hundred word-initial /p/'s in a randomly selected segment of the interviews of our sample of 35 native French speakers.[23]

We found 149 [pʰ]'s out of a total of 3,500, which represent 4.3%. Twenty-three different speakers used them, among whom eight only produced one or two tokens. The very large difference between the frequency of occurrence in Beauchemin's data and ours may be partially attributed to the contrasted contexts of production (reading versus spontaneous speech).

Two separate variable rule analyses were performed using Golvarb X. The first one tested the influence of two linguistic factors: the fact that the word was monosyllabic or not, and the fact that the initial /p/ was at the beginning of an utterance or elsewhere in the utterance. Neither factor group proved to have a significant effect.

We then proceeded to analyze the social distribution. The results are displayed in Table 3. Older people, women and the least educated people are the groups who tend to favor the use of [pʰ]. Contrary to Beauchemin's results, age appears to be a significant factor in the Stanstead data. With regards to education, both data sets show a similar trend. As for the occupation or the gender of the speaker, no comparison can be drawn between the Sherbrooke and the Stanstead studies. Gender was not considered in Beauchemin's analyses, and his grouping of occupations is different from ours, which is based on the categories used in all our studies on Montreal French, as explained in Thibault and Daveluy (1989) and Thibault and Vincent (1990).

The results concerning the presence of English in the speaker's home are most interesting. They show that people who have an English-speaking parent are more likely to aspirate their /p/'s than those whose spouse is English, who in turn, are more likely to aspirate than the speakers in whose home no English speakers live. This result, coupled with the low overall rate of occurrence of aspirated /p/, leads me to the following possible interpretation: The use of [pʰ] by the native French speakers of Stanstead is related to intimate relationships with English speakers,

---

23. Claudyne Chevrier, whom I wish to thank for her very careful and thorough work, had no previous formal training in phonetics. A native Quebec French speaker, she used a transcriber with earphones, which allowed her to listen to any occurrence slower or faster, at will. Once she had done a few interviews, we listened to each token together to make sure we heard the same variant. Then, whenever she had problems with one of the tokens, she would ask me to listen to it. Sometimes, we would just discard an occurrence, being unable to decide whether the /p/ was aspirated or not.

such as can be found in a shared home. The fact that the young people, who hang out with the other language group more than their parents, tend to aspirate less than their elders, lends support to such an interpretation. Having English-speaking friends is simply not enough.

Table 3. Analysis of the influence of five social factors on the occurrence of aspirated /p/ (Input = 0.029)

Factor Group	Factor	N. aspirated [ph]	Total	% Aspirations	Relative weight
Age group	63–87	71	1600	6.5	.578
	37–53	25	800	3.1	.464
	15–26	5	1100	3.3	.464
Gender	Feminine	68	1500	4.5	.593
	Masculine	81	2000	4.0	.430
Occupation	Blue collar	69	1200	5.8	non-sig.
	Business owners	15	300	5.0	non-sig.
	White collar	40	1100	3.6	non-sig.
	Professionals	25	900	2.8	non-sig.
Schooling	High school	128	2000	6.4	.664
	Junior college	17	1100	1.5	.255
	College	4	400	1.0	.389
English at home	English parent	39	300	13.0	.807
	English spouse	33	700	4.7	.523
	No English	77	2500	3.1	.451
Total		**149**	**3500**	**4.3**	

5.4 The types of 'r' heard in Stanstead

In our effort to characterize local speech in Stanstead, several issues revolve around the pronunciation of "r". Is American retroflex [ɹ] likely to appear more in "border" French than it does elsewhere in Quebec in contexts other than loanwords? Is Stanstead French displaying the use of traditional regional apical [r], which is associated with Montreal and the Eastern Townships, among other areas, according to Vinay (1950)? Has it taken part in the change from apical to dorsal [R],[24] first acknowledged in Clermont and Cedergren (1979), and Santerre (1979)?

---

24.   Following Sankoff and Blondeau (2007: 563), I use dorsal [R] to represent the velar fricative or uvular trill found in the eastern part of Québec.

Beauchemin (1972) included [ɹ] among the features of English which may have influenced French in the Sherbrooke area. Typically, the retroflex would be heard in the word-ending position, a slot most favorable to the deletion of consonants. The test his team conceived to elicit such [ɹ]'s consisted of fifteen words. In the tables presenting their results, they grouped together what they called "special" /r/'s with inaudible ones.[25] Their conclusion was that "normal" /r/ occurred more than expected (74% of the time).

Carolin Graml[26] did her Master's research on [ɹ], using our Stanstead data. In the 31 interviews she studied, she analyzed more than 22,300 variants of the /r/ phoneme. Graml (2005: 50) found that 87.3% of the total /r/'s were dorsal, and 7.4% were inaudible. Only 2.8% of the /r/'s were apical. She noted almost as many [ɹ]'s as the apical, with 2.4% in words of French origin, 90% of which were found in the word-ending position.

Two major hypotheses had been put forward as to the nature of [ɹ] in Quebec French: that is was a variant of [r] (Vinay 1950), and that the American-sounding sound was in fact a vocalized (actually, a diphthongued) /r/ (Tousignant 1987). Through careful analyses of the linguistic distribution of the various variants, Graml concluded, as reported in her 2006 article, that there was indeed a native Quebec French retroflex, a variant of [r], distinct from the American retroflex, described as "bunched" by Delattre and Freeman (1968). As for diphthongued /r/, which Tousignant presumed was a case of /r/ deletion following a diphthong, Graml (2006) distinguishes between clearly heard /r/'s as in diphthongized *coeur* [kauɹ] ("heart") and *cour* [kouɹ], as opposed to diphthongued *faire* [faɛ] ('to do'), in which no /r/ is heard.

In sum, apart from its occurrence in some English loanwords, there is no strong support for an English influence behind the presence of [ɹ] in the speech of our Stanstead interviewees.

Graml (2005) reported a very low rate of use of the apical [r] variant (2.8%). I wanted to find out whether most speakers used it occasionally or a few used it to a larger extent. In addition, given the wealth of research documenting the change from [r] to [R] in Montreal (Clermont & Cedergren 1979; Santerre 1979;

---

25.   Beauchemin's assistants heard some variants of /r/ which they couldn't clearly label as retroflex or vocalized /r/'s. They chose to call them special /r/'s. As for the inaudible /r/'s, they are presumably deleted.

26.   A Master's student of Roman Philology at Ludwig-Maximilian University in Munich, Carolin wanted to work on Quebec French. Ulrich Detges, who had invited me to Munich around the same time, advised her to ask me for access to data. Carolin has an exceptional ear for Quebec diphthongs, a definite asset for her research, and she got to like Stanstead people as much as we did.

Cedergren 1985; Sankoff et al. 2001; Sankoff & Blondeau 2007), the distribution of [r] in the Stanstead corpus might provide further evidence of the almost completed process.

As she did for [pʰ], my research assistant listened to and analyzed 100 words containing /r/ for each of the 35 native French speakers recorded in Stanstead. Only four speakers used at least one [r]. Table 4 details the results and some personal information on each of the four speakers.

Table 4. Users of apical [r] and some of their personal characteristics

	Albert	Norma	Jean	Léon
N. of [r]	2	19	58	71
N. of [R]	87	61	28	15
N. of [ɹ] or deleted /r/	11	20	14	14
Total	100	100	100	100
Age	64	77	87	78
Place of birth	Stanstead	Beauce	Stanstead	Stanstead
Origin of father	Stanstead area	Beauce	Eastern Townships	Beauce

In the Sankoff-Cedergren corpus of interviews recorded in 1971, Clermont and Cedergren (1979) found a few speakers from the 15 to 25 year-old group for whom the apical variant of /r/ was still dominant. Presumably, the last apical dominant cohort would have been born in the early fifties.

In the Stanstead corpus, the only two apical dominant speakers, Jean and Léon, are among the oldest speakers. A third among the oldest speakers, Florence (87 years old) is a categorical user of [R]. Unlike her age peers, who had worked as manual laborers, she had held jobs involving interactions with the public, in stores and at the post office. Albert[27] and Norma alternate between [r] and [R]. Norma came to Stanstead from the Beauce area, a region where apical /r/ is absent, at the age of twenty-two. She partially adopted the local norm, which she has kept in her repertoire ever since.

In Stanstead, [r] has almost disappeared. Thirty-one of our 35 French speakers use [R] categorically, among whom seven are from the oldest age group. The total occurrence of [r] is 4.3% (150 out of 3,500 tokens).[28] Many observations can be

---

27. The two tokens recorded for Albert in the short segment of 100 hundred /r/'s studied really reflect a variable behavior. In other parts of his interview, the apical variant occurs at a higher frequency.

28. This ratio is slighly higher than that recorded by Graml. Her sample consisted of 31 speakers, and she analyzed all the /r/'s within her corpus.

made in relation to Sankoff and Blondeau's article on the rapid change from apical to dorsal /r/ in Montreal (2007). Following the progression of [R] in the speech of 32 speakers who had been interviewed twice (in 1971 and in 1984), Sankoff and Blondeau (2007) observed an increase from 52% to 64% (a corresponding decrease of [r] from 48% to 36%). An additional 14 speakers were recorded again in 1995 (Vincent et al. 1995). In most of those last speakers, the rate of [R]'s remained stable. The general picture that Sankoff and Blondeau painted was that: "More speakers tended to be categorical than variable" (2007: 580). Such is definitely the case in Stanstead.

Sankoff and Blondeau (idem) go on saying that "those who changed did so very rapidly." Although there is no direct evidence of such a trend in the Stanstead data, the fact that seven speakers out of eleven from our oldest age group are categorical users of [R], all but one born in the area, certainly lends indirect support to their statement. Given the fact that they were around twenty when the change started to spread in the early 1950s, they probably were what Sankoff and Blondeau (2007) called "late adopters" of the new /r/. Ironically, Norma, who was a "late adopter" of [r], having come from a [R] region, has retained it to this day.

6. The borders of Stanstead French

When our research team first got to Stanstead, three things struck us with regards to language use: all business and most public signs were bilingual, despite Bill 101 which requires that public signs be written in French only; whenever you heard short verbal interactions in public places, it was almost impossible to guess the mother tongue of the speaker; once we started interviewing people, we felt they spoke French just like the French-speaking Montrealers among us.

The analyses presented here leave us with an uncomfortable feeling of unfinished business as to the task set forth in the titles of the article, and of section 5. Indeed, how local is Stanstead French? And whose language variety is it closest to?

In Stanstead, both the English-speaking bilinguals and the native French speakers use the *ne* particle in negative sentences with a frequency that emulates French-speaking Montrealers. For the French-speaking Stansteaders, the use of *ne*, being so scarce, doesn't qualify as a good diagnostic feature for minority status. This is the reason why we concentrated on L2 speakers' use of negative *ne*. Their high rate of *ne* deletion (97.3%) confirmed our impression that they sounded more like native speakers of French than the English Montrealers interviewed in the 1990s, as part of the Sankoff and Thibault research project.

With regards to the variable use of 3rd person singular morphology with a plural subject, French-speaking Stansteaders don't display minority French

behavior, as their rate of neutralization in favor of the singular form is similar to the Ontario speakers Mougeon and Beniak (1995) referred to as "unrestricted speakers" (1.5% in Stanstead, 2% in Ontario) living in communities where French is widely used, as opposed to "restricted speakers", who neutralize in favor of singular morphology 22% of the time. However, the observed vacillation between neutralization in favor of the 3rd singular and the 3rd plural in Stanstead is unmatched. Could it be a regional feature or has it just escaped notice in Montreal French?

Aspirated /p/ points to a close-contact situation, namely at home. However, it is doubtful that French-speaking Montrealers with an English parent or spouse would use [pʰ] in Montreal. The Stanstead language contact situation, both demographic and geographic, might be a trigger for such a variant to appear, bilingual interactions being frequent within the community.

The /r/'s present a complex situation in Stanstead. There seems to be a French native retroflex variant studied by Graml (2005) which is likely to be found elsewhere in Québec, as Santerre (1982), and Tousignant (1987) seem to indicate.

On the other end, the almost total disappearance of [r] in Stanstead French is somewhat puzzling. In his recent article on "Transmission and diffusion," Labov (2007) states that changes that proceed through community internal transmission move faster than changes that have reached a community through diffusion from one city to another. Stanstead French has clearly attained a very high level of backward movement of the /r/, given that only four speakers out of thirty-five still use the front [r] variant. Two peculiarities seem to distinguish that change from others: The change has spread from secondary cities and regions to Montréal, the metropolis of Québec, and it went along with a change of the standard norm from [r] to [R], corresponding to a target closer to the French uvular /R/. The question that needs to be asked here is: Since the progress of dorsal /R/ appears to be slower in Montreal than in Stanstead (Sankoff & Blondeau 2007), is Montreal the last bastion of [r] in Québec?

Finally, getting to code-switching and local identity in Stanstead, I wonder if Stanstead, and other small communities near the US and Canadian border, could not become exemplary situations where stable bilingualism and ensuing code-switching would be the main defining traits of local identity, for both the English and the French language groups.

References

Ashby, W. 1981. The loss of the negative particle *ne* in French: A syntactic change in progress. *Language* 57(3): 674–687.

Beauchemin, N. 1972. Quelques traits de prononciation québécoise dans un contexte anglophone qui les influence? *Recherches sociolinguistiques dans la région de Sherbrooke*. Document de travail No. 2.

Blondeau, H., Fonollosa, M.-O., Gagnon, L., Lefebvre, N., Poirier, D. & Thibault, P. 1995. Aspects of L2 competence in a bilingual setting, Paper presented at NWAV 24.

Blondeau, H., Nagy, N., Sankoff, G. & Thibault, P. 2002. La couleur locale du français des Anglo-Montréalais. *In L'acquisition de la variation par les apprenants du français langue seconde* [Acquisition et Interaction en Langue Étrangère 17], J.-M. Dewaele & R. Mougeon (Eds), 73–100.

Cedergren, H. 1985. Une histoire d'R. In *Les tendances dynamiques du français parlé à Montréal*, Vol. 1, M. Lemieux & H.J. Cedergren (Eds), 25–57. Québec: Gouvernement du Québec.

Clermont, J. & Cedergren, H. 1979. Les 'R' de ma mère sont perdus dans l'air. In *Le français parlé. Études sociolinguistiques*, P. Thibault (Ed.), 13–28. Edmonton: Linguistic Research.

Delattre, P. & Freeman, D. 1968. A dialect study of American R's by X-ray motion picture. *Linguistics* 44: 29–68.

Flikeid, K. 1989. Recherches sociolinguistiques sur les parlers acadiens du Nouveau-Brunswick et de la Nouvelle-Écosse. In *Le français parlé hors Québec*, R. Mougeon & É. Beniak (Eds), 183–199. Québec: Les Presses de l'Université Laval.

Graml, C. 2005. Soziale Variation im Französischen Quebecs – Das so genannte retroflex /R/. MA thesis, Ludwig Maximilian University, Munich.

Graml, C. 2006. Ein englisches /R/ im Französischen Quebecs? Eine Gegenthese zur traditionellen Interpretation eines Quebecker Aussprachemerkmals. *Romanistisches Jarbuch* 56: 63–82.

King, R. 1993. Subject-verb agreement in Newfoundland French. Paper presented at NWAV 22, Ottawa.

Labov, W. 2007. Transmission and diffusion. *Language* 83(2): 344–387.

Lacasse, C. 2005. Quelques variations dans le français parlé à Stanstead, une communauté bilingue située à la frontière du Québec et du Vermont. MA thesis, University of Montreal.

Lambert, W., Hodgson, R., Gardner, R. & Fillenbaum, S. 1960. Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60: 44–51.

Mougeon, R. & Beniak, É. 1991. *The Linguistic Consequences of Language Contact and Restriction: the Case of French in Ontario, Canada*. Oxford: OUP.

Mougeon, R. & Beniak, É. 1995. Le non-accord en nombre entre sujet et verbe en français ontarien: un cas de simplification? *Présence francophone* 46: 53–65.

Mougeon, R. & Rehner, K. 1999. Variation in the spoken French of immersion students: To *ne* or not to *ne*, that is the sociolinguistic question. *The Canadian Modern Language Review/ La revue canadienne des langues vivantes* 56(1): 124–154.

Poplack, S. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18: 581–618.

Poplack, S. 1989. The care and handling of a megacorpus: The Ottawa-Hull French project. In *Language Change and Variation*, R. Fasold & D. Schiffren (Eds), 411–451. Amsterdam: John Benjamins.

Quebec Towns' Directory: http://www.mamr.gouv.qc.ca.

Ringuette, A. 2004. Langues sans frontière: le bilinguisme chez les jeunes à Stanstead dans les Cantons de l'Est. MA thesis, University of Montreal.

Sankoff, D., Sankoff, G., Laberge, S. & Topham, M. 1976. Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale. *Cahier de linguistique* 6: 85–125.

Sankoff, D., Tagliamonte, S. & Smith, E. 2005. Goldvarb X: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.

Sankoff, G. & Blondeau, H. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83(3): 560–588.

Sankoff, G., Blondeau, H. & Charity, A. 2001. Individual roles in a real-time change: Montreal (r-> R) 1947–1995. In *'r-atics. Sociolinguistic, phonetic and phonological characteristics of /r/* [Etudes et Travaux 4], H. van de Velde & R. van Hout (Eds), 141–157. Brussels: Free University of Brussels.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M.-O. & Gagnon, L. 1997. Variation and the use of discourse markers in a language contact situation. *Language Variation and Change* 9: 191–218.

Sankoff, G. & Vincent, D. 1977. L'emploi productif du *ne* dans le français parlé à Montréal. *Langue française* 34: 81–109.

Santerre, L. 1979. Le [r] montréalais en régression rapide. *Prothée,* 117–132, Chicoutimi, Québec.

Santerre, L. 1982. Des [r] montréalais imprévisibles et inouïs. *Revue québécoise de linguistique* 12(1): 77–96.

Thibault, P. & Daveluy, M. 1989. Quelques traces du passage du temps dans le parler des Montréalais, 1971–1984. *Language Variation and Change* 1(1): 19–45.

Thibault, P. & Sankoff, G. 1999. L'évaluation du français des jeunes Anglo-montréalais par leurs pairs francophones. *The Canadian Modern Language Review/ La revue canadienne des langues vivantes* 56(2): 245–281.

Thibault, P. & Vincent, D. 1990 *Un corpus de français parlé*, Québec, CIRAL, coll. Recherches sociolinguistiques.

Tousignant, C. 1987. Les variantes du /R/ montréalais: contextes phonologiques favorisant leur apparition. *Revue québécoise de linguistique théorique et appliquée* 6(3): 73–113.

Vinay, J-P. 1950. Bout de la langue ou fond de la gorge? *French Review* 23(6): 489–498.

Vincent, D., Laforest, M. & Martel, G. 1995. Le corpus de Montréal 1995. Adaptation de la méthodologie sociolinguistique pour l'analyse conversationnelle. *Dialangue* 6: 29–45.

Quantitative Sociolinguistics

From the languages, what can we learn about the speakers?

Ne deletion in Picard and in regional French

Evidence for distinct grammars

Julie Auger & Anne-José Villeneuve
Indiana University

Ne deletion is arguably the best studied variable in French. Despite differences in overall rates of deletion, linguistic factors governing deletion pattern very similarly in most varieties. The comparative study presented here offers a new perspective by using *ne* deletion as a criterion to differentiate two closely-related languages, French and Picard. While patterns of *ne* deletion in the variety of French under study are similar to those found in other French varieties, *ne* in Picard behaves differently from its French counterpart, with respect to both frequency of deletion and linguistic factor effects. We investigate patterns of *ne* deletion in three bilingual speakers for whom we have both written and oral Picard, as well spoken French. We compare their French data with those of monolingual French speakers from the region. This study contributes additional linguistic evidence for the claim that Picard and French are distinct languages.

Keywords: French; Picard; *langues d'oïl*; negation; *ne* deletion; variation; language change; variationist sociolinguistics; quantitative methods; collocations

1. Introduction

One unforeseen consequence of world-wide economic and cultural globalization has been an increased attachment and pride in local values.[1] As a result, instead of observing the eradication of all languages other than English on the internet that had been predicted by many, we are witnessing the appearance of many languages in domains in which they had never been observed. For instance, Wikipedia, the free online encyclopedia, contains articles in numerous different languages, including endangered regional languages such as Corsican, Breton, Occitan, Norman, and Walloon. Similarly, it is now possible to find web sites written in

---

Picard, Poitevin, and Galician, to name just a few "unexpected" languages. Finally, newsgroups such as Achteure, a list that brings together individuals interested in the Picard language, post messages written in languages that, until recently, had been essentially reserved for oral communication.

This trend is undoubtedly encouraging, given that an increased domain of use may help convince community members of the appropriateness of their language to modern life and thus have a positive impact on its preservation or even on its extension. Yet, it must be recognized that many endangered languages continue to face a bleak future. This is especially true of varieties that share many features with neighboring official languages.[2] For instance, no one would deny that the language isolate Basque is not a variety of French or Spanish, or that Breton is not a French dialect. However, with closely related languages, people (particularly non-linguists) may assume that there is one correct standard variety with several sub-standard related dialects. For example, we note that the temptation to consider Catalan as a variety of Spanish was strong during the Spanish Civil War, and that many people continue to regard Gallo-Romance varieties of northern France as examples of poor French (similar attitudes can be observed in many countries with a standardized, national language). In such communities, parents often consider that the best way to guarantee the future success of their children is to ensure that they do not use such "bad" French and that they grow into monolingual speakers of "good" French.[3]

Linguists are well aware of how difficult it is to determine whether a given linguistic variety is an independent language or a dialect of another language. Many introductory linguistics and sociolinguistics courses address this question, discussing criteria such as those proposed by Bell (1976) to decide what is a language and what is a dialect, and pointing out linguistic varieties that pose problems to such criteria (e.g., while Chinese "dialects" are not mutually intelligible, Danish, Norwegian, and Swedish are considered to be distinct languages in spite of the fact that they are largely mutually intelligible). Consequently, the uncertainty experienced by naïve speakers is far from surprising. However, our experience with Picard has taught us that linguists often express similar ambivalent attitudes toward minority varieties. Even worse, some have taken a stand, concluding that Gallo-Romance languages have become extinct in France and

---

2.   However, there are exceptions to this trend. Manzano (2005) discusses the case of Gallo, an Oïl variety that is very closely related to French but that receives a fair amount of recognition as a distinct language thanks to the geographical proximity of Breton.

3.   This tendency corresponds to what some people have termed *language suicide* (Nettle & Romaine 2000: 5–6).

maintaining that the regional variation still observed today is attributable to the regional varieties of French, which have arisen from the contact between French and the Gallo-Romance varieties formerly spoken in a given area. For instance, the first author of this article had to put great efforts into convincing the editor of a series in which she was publishing a book that the brief description of her research interests should refer to *Québec French and Picard* rather than *Québec and Picard French*. Similarly, the lack of interest that most Gallo-Romance languages – and more particularly Oïl varieties spoken in northern France – elicit in France is due in part to the conviction, widespread among linguists and non-linguists alike, that such varieties have already disappeared.

For many linguists, the distinction between language and dialect is perceived to be independent from their linguistic analysis; what is important in linguistic analyses is whether the system corresponds to the linguistic competence of a group of speakers. However, this question is very important to non-linguists for at least two reasons. First, in the minds of its speakers, whether a linguistic variety is perceived as an autonomous language or as a subvariety of another language tends to have a direct impact on their desire to save this variety as a symbol of their identity. Second, to qualify for support under the European Charter for Regional or Minority Languages, which France has yet to ratify, nine years after it signed it, a variety must be recognized as a distinct language rather than as a dialect of the national language. Hence, the contrast between regional languages on the one hand and French and its dialects on the other hand jeopardizes the political recognition of several varieties.

The status of the Gallo-Romance languages of northern France, often referred to as the *Oïl varieties*, has been the object of a heated debate in France. For instance, a report submitted to the prime minister of France by Bernard Poignant in 1998 claims that such languages no longer exist:

> *Ces langues d'oïl, langues utilisées au Moyen-Âge par les seigneurs de ces régions, étaient aussi langues de large communication en milieu rural. Ces langues proprement dites ont disparu et les parlers actuels ont été largement influencés par le français. Leurs locuteurs sont aujourd'hui peu nombreux, mais un réveil s'organise autour de l'université. Les parlers d'oïl, tels que le picard (au nord), le gallo (à l'ouest), le poitevin, le saintongeais, le normand, le morvandiau, le champenois, d'autres encore constituent autant de formes régionales du français.*

> These Oïl languages, which were used in the Middle Ages by lords from these regions, were also extensively used for communication in rural areas. These languages per se have disappeared and the current idioms have been largely influenced by French. Nowadays, there are few speakers of Oïl languages, but a cultural awakening is currently developing around universities. Oïl varieties such as **Picard** (in the north),

Gallo (in the west), **Poitevin**, **Saintongeais**, **Norman**, **Morvandiau**, **Champenois**, and others are various regional forms of French. [emphasis is Poignant's, all translations are ours]

This view of Oïl varieties is contradicted in another report prepared by Bernard Cerquiglini and submitted to the French minister of Education, Research, and Technology and the minister of Culture and Communication in 1999:

Le français "national et standard" d'aujourd'hui possède une individualité forte, qu'a renforcée l'action des écrivains, de l'État, de l'école, des médias. Il en résulte que l'on tiendra pour seuls «dialectes» au sens de la Charte, et donc exclus, les «français régionaux», c'est-à-dire l'infini <sic> variété des façons de parler cette langue (prononciation, vocabulaire, etc.) en chaque point du territoire. Il en découle également que l'écart n'a cessé de se creuser entre le français et les variétés de la langue d'oïl, que l'on ne saurait considérer aujourd'hui comme des «dialectes du français»; franc-comtois, wallon, picard, normand, gallo, poitevin-saintongeais, bourguignon-morvandiau, lorrain doivent être retenus parmi les langues régionales de la France; on les qualifiera dès lors de «langues d'oïl»

Contemporary "national and standard" French has a strong individuality, which was strengthened by actions taken by writers, the State, school, and the media. As a result, we will only consider as dialects, and therefore exclude for the purposes of the Charter, regional French varieties, that is, the infinite number of ways of speaking the French language (pronunciation, vocabulary, etc.) in every location throughout France. Another consequence is that French has become increasingly different from Oïl varieties, which cannot be considered as French dialects today; Franc-Comtois, Walloon, Picard, Norman, Gallo, Poitevin-Saintongeais, Bourguignon-Morvandiau, Lorrain must be included among the regional languages of France; they will be referred to as "Oïl languages"

2.   Picard vs. French: A few differences

From a historical point of view, French and Picard correspond to two different forms that Latin took in Gaul (cf., e.g., Dawson 2002). That is, French and Picard have evolved side by side, and their different fates must be attributed to the fact that French royalty settled in Île-de-France and that their language quickly imposed itself as the only one appropriate for serious purposes. But history is not the only reason why, in our minds, Picard and French must be considered separate languages. Recent research by Auger has provided two kinds of evidence in favor of this conclusion. First, it has established that the structure of Picard differs from that of French. For instance, Vimeu Picard possesses two negative adverbs, *point*

and *mie*, instead of the single *pas* used in contemporary, non-literary French; cf. (1). Another example is provided by the contrast between *a/ch'/Ø* in Picard and *ce/ça* in French. While all these forms can be translated as 'it', different factors constrain their distribution in the two languages. In Picard, the type of predicate determines which of *a/Ø* or *ch'* is used, whereas in French, *ce* is used with the verb *être* 'to be', and *ça* is used with all other verbs.[4,5] Specifically, in Picard, *a/Ø* occurs with verbal and adjectival predicates, as shown in (2), while *ch'* occurs with nominal predicates and in pseudo-cleft constructions, as exemplified in (3).

(1) a. *i   n'   ont   point l' vint din leuz yeux.*   (Joseph L., 6/30/96)[6]
   they NEG have not   the wind in   their eyes
   'they don't have the wind in their eyes'

  b. *Ch 'n'   étoait mie d' chés grands autocars*
   it   NEG   was   not of these large   buses

   tout neus …   (Joseph L., text, p. 18)
   all   new

   'It was not one of these brand new large buses'

(2) a. *a foait   partie d' Moyenneville.*   (Joseph L., 6/30/96)
   it makes part   of Moyenneville
   'It is part of Moyenneville'

  b. *a   n'   est poè difficile à picardiser   des*   (Joseph L., 6/30/96)
   It   NEG is not difficult to make.Picard some

   mots   hein?
   words   right

   'It is not hard to make words into Picard, right?'

---

4. Two of the Picard forms, *a* and *Ø*, are allomorphs: *a* occurs before words that begin with a consonant or a high vowel, while *Ø* occurs before words that start with a mid or low vowel.

5. In many varieties of colloquial French, the use of *ce* is even more restricted: it occurs before vowel-initial forms of *être*, and *ça* occurs with all other forms.

 (i) *C'est beau* 'it's beautiful'
 (ii) *Ça serait beau* 'it would be beautiful'

6. All speakers who have provided oral data are referred to with pseudonyms. These pseudonyms are also used when referring to published written data from the same speakers in order to protect their anonymity. Because the privacy issue does not arise for the older authors for whom we have analyzed only written data, we clearly identify the sources for their data and provide full bibliographical information in the reference section.

(3)   a.   *pasque <u>ch'</u> étoait ch' sujet un peu d' éch live*   (Joseph L., 6/30/96)
           because it was   the topic a bit of the book
           'because it was somewhat the topic of the book'

      b.   <u>ch'</u> *étoait avec eux que o sommes partis*   (Joseph L., 6/30/96)
           it   was   with them that we are   left

           à   l'   évatchuation
           to the evacuation

           'it is with them that we left at the time of the evacuation'

Auger and Villeneuve (submitted) provide evidence that even such a low-level phonological phenomenon as vowel epenthesis operates differently in the two varieties. In addition to involving two different vowels (schwa in French, [e] in Picard), the rules that govern the distribution of the epenthetic vowel result in different insertion sites in many types of environments. For instance, when a verb that begins with a consonant cluster that is not a legitimate onset follows a consonant-final word, French inserts a schwa between the verb's two initial consonants, whereas Picard inserts an [e] before the first two consonants, as shown in (4). Or when the pronominal clitics *j* 'I' and *m* 'me' are combined and used with a consonant-initial verb, French generally inserts its epenthetic vowel after the [m], but Picard inserts it between the two consonants; cf. (5).

(4)   a.   *on s'   relit* [rəli] *jamais assez*   (Joseph L., French)
           one oneself rereads   never   enough
           'one never rereads oneself enough'

      b.   *alors mi j' érbéyoais* [erbejwɛ], *ch' qu' il écrivoait.* (Joseph L., Picard)
           so   me I looked.at   that that he wrote
           'so, I looked at what he wrote'

(5)   a.   *y   a un moment j' me* [ʒmə] *souviens*   (Joseph L., French)
           there is a moment I me   remember

           très bien
           very well'

           'there is a moment I remember very well'

      b.   *J'ém* [ʒem] *souvarai   longtemps d' éch l'air*   (Joseph L., Picard)
           I.me   will.remember long   of the tune
           'I will long remember the tune'

The second type of evidence that French and Picard are distinct languages is provided by a comparison of spoken and written data collected from the same speakers. Recent efforts to revitalize Picard have given rise to a number of written texts. Magazines, novels, collections of short stories, and cartoons are regularly published

(Vigneux ms). This literary movement allows us to examine the choices made by authors and editors in the creation of a literary standard for Picard. Auger (2003 a,b) shows that these choices favor linguistic forms that maximize the distance between Picard and standard French. In other words, Picard authors and editors favor precisely the forms that teachers and grammarians condemn as "bad" French. For instance, in both speech and writing, subject doubling and resumptive pronouns in subject relatives are used virtually categorically; cf. (6). For other features, we observe that their rate of use differs in writing and in speech. For instance, the use of *avoér* 'to have' as auxiliary with verbs of movement and reflexive verbs, cf. (7), is higher in writing than in speech. On the other hand, while *pas* 'not' occurs with a non-negligible frequency in the oral data, Auger did not find a single occurrence of this form in the written texts analyzed. What this analysis shows is that Picard speakers favor different norms for French and Picard: many of the constructions that they avoid in French texts are precisely the ones that they favor when writing in Picard, thus supporting the idea that each language has its own standard.

(6) a. <u>*Parsonne*</u> *<u>i</u>* *n'* *poroait mie vnir* *ll'* *értcheure.* (Joseph L.)
nobody he NEG could not to.come him to.get.back
'Nobody could come and get him'

b. *comme eine brouette* <u>*qu'*</u> <u>*a*</u> *n'* *va* *pus* (Alain Q.)
like a wheelbarrow that she NEG goes anymore
'like a wheelbarrow that's no longer working'

(7) a. *innhui j'<u>ai</u>* *vnu* *aveuc ein live* (Jean-Michel F., oral, Auger in press)
today I.have come with a book
'today I came with a book'

b. *i* *s'<u>a</u>* *cassé* *s'* *gambe* (Jacques V., oral, Auger in press)
he self.has broken his leg
'He broke his leg'

In this paper, we examine *ne* deletion in two language varieties spoken in Vimeu – an area of northwestern France – to provide additional evidence that French and Picard should be considered distinct languages. A cursory look at the two varieties reveals strikingly similar tendencies. In writing and in formal settings, both varieties favor the use of both a preverbal *ne* and a postverbal negative adverb, while in oral performance in informal settings, the absence of the preverbal *ne* is often observed.[7] However, as we will see below, a closer look at our data reveals that *ne*

---

7. Given the tendency to favor structures that maximally distinguish Picard from "good" French, the unexpected *ne* retention in forms such as *ne + point/mie* must be interpreted, in our opinion, as evidence that *ne* is a part of Picard grammar.

deletion occurs with very different frequencies in the two grammars and that it is constrained by different linguistic factors.

3.   *Ne* deletion: An overview of previous research

3.1   Previous studies on French

Ne deletion is arguably the best studied variable in contemporary French (Coveney 1996: 55). This phenomenon has been investigated in many native varieties of Canadian and European French (see Table 1), as well as in the speech of French learners (e.g., Trévise & Noyau 1984; Regan 1996; Rehner & Mougeon 1999; Sax 2003; Dewaele 2004).

Table 1. An overview of previous quantitative studies of *ne* deletion

	Year of survey	Research site	Number of tokens	% of *ne* deletion
European French				
Pohl 1968	early 1950s	Belgium/France	5,308	38.1%
Ashby 1976	1967–68	Paris	1,029	44.2%
Diller 1983	1975	Béarn	641	34.3%
Ashby 1981	1976	Tours	2,818	63.4%
*Coveney 1996**	*1980*	*Somme*	*2,932*	*81.2%*
Moreau 1986	1982–83	Belgium	3,158	49.8%
*Pooley 1996**	*1983*	*Roubaix*	*3,719*	*93%*
Hansen & Malderez 2004	1989–1993	Paris/Oise	1,329	91.8%
Armstrong 2002	1990	Lorraine	2,501	98.2%
*Pooley 1996**	*1995*	*Rouge-Barres (Nord)*	*391*	*99%*
Ashby 2001	1995	Tours	1,593	84.3%
Fonseca-Greber 2007	late 1990s	Switzerland	1,982	97.5%
Canadian French				
Sankoff and Vincent 1977	1971	Montréal	±10,000	99.5%
Poplack and St-Amand 2007	20th century	Ottawa-Hull	61,316	99.8%

* Data from a Picard-speaking area   (table adapted from Armstrong and Smith 2002)

Even though these studies have revealed considerable differences in overall rates of deletion, ranging from 34.3% in Béarn to 99.8% in Canada's Ottawa-Hull region, they have also shown that very similar linguistic and social factors govern *ne* deletion in most varieties. Among determining factors, speech style appears to have the most important impact on rates of negative particle deletion (cf. Pohl 1975; Sankoff & Vincent 1977, among others). Pohl (1975) points to the effect of three primary factors – style of speech, speaker's personality and sentence structure – but

highlights that the use of the more formal 2nd person singular pronoun *vous* instead of informal *tu* interacts with all three factors. He also stresses that geographic region, age, social class, or language contact may affect *ne* deletion.

> *[C]ertaines «qualités» sont favorables […] à l'omission. Parmi ces dernières, notons: l'appartenance à certaines régions, comme Paris, les zones voisines, ou encore le Canada (la Wallonie rurale, au contraire, reste conservatrice); l'habitat urbain; la jeunesse et, plus encore, l'enfance; un niveau socio-culturel modeste […]; certains bilinguismes ou l'influence d'une langue étrangère dont la négation est faite d'un seul mot. (Pohl 1975: 18)*

> Some "qualities" are favorable to deletion. Among these are: being a member of certain regions, such as Paris, neighboring areas, or Canada (rural Wallonia, on the other hand, remains conservative); urban setting; youth and, even more so, childhood; a modest socio-cultural status; some bilingualisms or the influence of a foreign language in which negation is expressed by a single word.

Medium of communication and conversation topic have also been claimed to affect rates of *ne* deletion: face-to-face exchanges favor *ne* deletion more than phone communications (Pohl 1975), while topics that elicit formal speech styles such as language, education, and religion are correlated with higher rates of *ne* retention (Sankoff & Vincent 1977; Blanche-Benveniste 1997; Poplack & St-Amand 2007).

Several linguistic factors have been studied. Among others, frequency of expression, subject and verb type, second negative, and verb length have been found to influence rates of *ne* deletion. High-frequency expressions and lexicalized forms are mentioned by most authors, if only to justify the exclusion of fixed expressions such as *n'est-ce pas* 'right (tag question)' and *n'importe quoi* 'anything' from their corpus. Among others, Pohl (1975) notes that *ne* is never lost in certain ritualized expressions such as *Ne quittez pas* 'hang on (on the telephone)' and *Vous n'avez rien à déclarer?* 'You don't have anything to declare? (at customs)'. Ashby (1976, 1981) comments that in both Paris and Tours, the loss of *ne* is "especially far advanced" in frequent expressions: '[ʃepa] for *je ne sais pas* 'I don't know' and [sepa] for *ce n'est pas* 'it isn't' are popularly recognized stereotypes of *ne* deletion." (Ashby 1981: 678). This holds true not only for native speakers of European French (Moreau 1986; Coveney 1996), but also for learners of French (Sax 2003).

Subject type is also said to affect *ne* retention, with pronouns favoring loss more than full NPs (Ashby 1976, 1981; Diller 1983; Moreau 1986; Hansen & Malderez 2004, among others). Pooley (1996: 173) also notes that the absence of subject doubling favors the retention of *ne*. Further examination of the type of subject pronoun has yielded significant results in Ashby (1976, 1981), with clitic subject

pronouns (*je* 'I', *tu* 'you, SG. familiar', *il* 'he', etc.) favoring deletion more than non-clitic pronouns (e.g., *cela* 'that', *quelqu'un* 'someone'). Armstrong & Smith (2002) further comment on varying effects for different clitic subject pronouns, showing that *ne* is retained more frequently with some clitics (*elle* 'she', *nous* 'we', *vous* 'you PL., SG. formal', *ils* 'they MASC.') than with others (*je* 'I', *ce* 'it').

With respect to the effect of verb type, Ashby (1981: 680) notes that while *ne* tends to be retained with *être* 'to be' and *avoir* 'to have' and with modals such as *devoir* 'must' and *pouvoir* 'to be able to/can', it tends to be deleted with lexical verbs or with the modal *aller* 'to be going to'. This tendency is only partly supported by Moreau (1986): while lower rates of *ne* deletion are indeed observed for *avoir* (37.86%), *ne* is deleted more frequently with *être* (58.32%) and with other modals such as *aller*, *devoir*, *pouvoir*, *savoir* 'to know' and *vouloir* 'to want' (54.01%).

Other negative elements present in the clause also influence *ne* deletion. Ashby (1981) and Pooley (1996) report that *pas* 'not' generally favors *ne* deletion more than *plus* 'no longer'. The presence of a third negative element (e.g., *Tu ne fais jamais <u>rien</u>* 'You never do anything') also tends to favor the retention of *ne* (Pooley 1996), although not significantly so in Paris (Ashby 1976). Lastly, Pohl (1975) found *ne* deletion to be less frequent before longer verbs, claiming that deletion is inversely proportional to the number of syllables that separate the intended location of *ne* from the second negative.

3.2   Negation and *ne* deletion in Picard

When they discuss negation in Picard, grammarians typically focus on the fact that negators other than *pas* characterize these varieties. For instance, in the variety spoken in the Nord department, *nin* and *mie* are frequently used (Dawson 2002: 28–29). In Vimeu, as we saw earlier, *point* and *mie* are the most common forms observed. However, relatively little information is available about *ne* deletion.[8] While, as we have just seen, *ne* deletion is common in colloquial French, it is not typically found in the Vimeu variety of Picard, according to Vasseur (1996). This position is shared by Dawson (2002: 29), who reports that *ne* is deleted less often in Chtimi Picard than in French, and Pooley (1996: 171), who notes that "the omission of *ne* is not characteristic" of Picard, except when "strong structural constraints for its omission are present." One such constraint is the presence of an object pronoun. In fact, Vasseur (1996: 88) notes that the only context in which *ne* deletion occurs in Picard is in front of third person accusative and dative pronouns *l'*, *lé* and *li*.

---

8.   In their description of Picard varieties, neither Debrie (1983) nor Flutre (1955) mention any tendency to drop *ne* in Picard, according to Coveney (1996: 62).

In his study of Chtimi, the urban Picard-French mixed variety spoken around Lille, Pooley (1996: 173) also notes that the presence of non-subject clitics correlates with "a significantly higher rate" of *ne* deletion, as exemplified in (8b–c).[9]

(8)   a.   *I   n   fwé pwé ryē*   (Vasseur 1996: 88)
              he   NEG   do   not   nothing
              'he does not do anything'

        b.   *… k   éj   <u>l</u>   é   pwé vü*
              that   I   him   have   not   see
              'that I did not see him'

        c.   *Ó<u>z-z</u>   óbliræ   pwē*
              you-them   forget.FUT   not
              'you will not forget them'

The occurrence of *ne* deletion in Picard, cf. (8d–e), is also mentioned by Ledieu (1909/2003) and by Dawson (2002, 2003), who describe different varieties of Picard. Yet, unlike Vasseur (1996) and Pooley (1996), none of these sources report any favoring effect of the presence of an object pronoun on the deletion of *ne*.

(8)   d.   *Ch'est point tout …*   (Ledieu 1909/2003: 96)
              it.is   not   all
              'That's not all'

        e.   *J'sai   point*   (Dawson 2003: 30)
              I.know   not
              'I don't know'

4.   Methodology

Given the considerable variation that characterizes Picard and French in northern France and in southern Belgium, our study focuses on the Vimeu region. Located in the western part of the Somme department in France, this region is home to many elderly Picard speakers and a non-negligible number of younger people who speak Picard fluent. Vimeu has been claimed to have greater Picard vitality than other areas which were more severely affected by World War I (Carton 1981). Our corpus of Vimeu French is comprised of approximately 45 minutes of spoken French data for each of three monolingual and three bilingual speakers.

---

9.   Yet, contrary to Vasseur's specific description of Vimeu Picard, Pooley (1996: 174) adds that "relative to the rest of the examples with non-subject clitics, […] *le*, *la*, *l'* and *y* are relatively favorable to the occurrence of the particle" in Chtimi.

The same three bilinguals provided the Vimeu Picard spoken data, which consist of at least 45 minutes of speech for each speaker and of several texts. Since our three bilingual speakers are all actively involved in the Picard revitalization movement, we have also analyzed data from three subjects who can be described as *picardisants du cru*, that is as speakers who use Picard in their daily lives but do not write or read in Picard and do not belong to any Picard association.

In addition to the oral data, we have examined two types of written data. For the bilingual speakers recorded in French and in Picard, we also analyzed their use of *ne* in written texts. Because these three bilingual speakers attend meetings of the *Picardisants du Ponthieu et du Vimeu* and publish in the same magazine, we decided to examine older texts to determine whether the patterns observed in the texts of our bilingual speakers correspond to the wider norm present in the Vimeu community or whether they correspond to a standard created by the *Picardisants* group. While we recognize that written data may not reflect the linguistic competence of their authors as faithfully as spontaneous speech produced in informal settings, these data are important for two reasons. First, as explained above, because they give us access to speakers who belong to different social networks than the bilingual speakers in our corpus. Second, because previous work by the first author has shown that written texts provide a faithful representation of the oral performance of their authors, diverging from speech minimally only in a few features which involve a choice between French and Picard variants (cf. Auger 2002, 2003b). Table 2 summarizes the data used in our analysis.

Table 2. Subjects and demographic information

Group	Speaker	Sex	Occupation
French monolinguals [SF]	Fabienne A.	F	local business owner
	Guy D.	M	farmer
	Annick M.	F	artist
Bilinguals [SF, SP, WP]	Thomas S.	M	teacher
	Jacques V.	M	pharmacist, editor
	Joseph L.	M	retired teacher
Picardisants du cru [SP]	Marcel C.	M	farmer
	Alain Q.	M	retired factory worker
	André L.	M	retired corporate manager
Older authors [WP]	Arthur Lecointe	M	salesperson
	Gaston Vasseur	M	teacher
	Robert Touron	M	priest

SF = spoken French; SP = spoken Picard; WP = written Picard

4.1   Data selection

All instances of a negative construction were targeted in the data; cf. (9)–(10). Negative constructions where *ne* is immediately followed by the second negative, i.e., a sequence of a preposition + *ne* + second negative or of *ne* + second negative + infinitive, are also included in our analysis; cf. (11).

(9)   a.   *Mon ami   (ne)   répond pas à ma question.*
           my   friend (NEG) answers not to my question
           'My friend does not answer my question.'

    b.   *On (ne)   te   demande rien.*
         we (NEG)   you.ACC   ask   nothing
         'We don't ask you anything.'

(10)   *Ce/c' (n')   est pas possible!*
       it   (NEG)   is   not possible
       'It's not possible!'

(11)   a.   *… pour (ne) pas que ça soit visible.*
           for   NEG not that it be   visible
           '… so that it not be visible'

    b.   *… ce qu'il   faut   faire pis (ne) pas faire …*
         what.it   needs do   and NEG not do
         '… what we need to do and not do'

Following most studies of *ne* deletion in French, tokens of negative constructions in which the third person pronoun *on* or object pronoun *en* is followed directly by a vowel-initial verb, as shown in (12), are excluded from the analysis. In such context, the realization of an [n] can be attributed either to liaison or to negative *ne*, making it impossible to determine whether *ne* is realized or not (cf. Armstrong 2002: 159; Armstrong & Smith 2002: 25; and Coveney 1996: 66 for a more detailed account).

(12)   a.   *On était* [ɔ̃netɛ]   *malade.* (affirmative)
           we   were   sick
           'We were sick'

    b.   *On   était* [ɔ̃netɛ]   *pas malade.* (negative without *ne*)
         we   were   not sick
         'We were not sick'

    c.   *On n'   était* [ɔ̃netɛ]   *pas malade.* (negative with *ne*)
         we NEG   were   not sick
         'We were not sick'

Similarly, realizations of the genitive/partitive clitic as [nn] or [nɛ̃] in Picard were excluded from our analysis due to the fact that these forms do not allow for negative *ne* to co-occur.

(13) a. *i* <u>*nn*</u>' *awouot point pour vous tortous.*   (Joseph L., text, p. 23)
It of-it had   not for you.PL all
'There wasn't enough of it for you all'

b. *I* <u>*nn*</u>' *éroait   bré.*   (Joseph L., text, p. 38)
he of-it would-have cried
'He would have cried because of it'

In both French and Picard, *ne* may also be used in *ne … que* restrictive clauses such as *Le bébé ne boit que du lait* 'The baby drinks nothing but milk'. We exclude restrictives from our analysis, focusing strictly on negative constructions in order to ensure functional comparability (Lavandera 1978), and avoid treating two distinct semantic uses of *ne* as one single sociolinguistic variable.

4.2   Coding

French and Picard tokens were coded for several linguistic factors and submitted to GoldVarb 2001 for a multivariate analysis. Given previous findings concerning the effect of subject of the verb, we coded for subject person, subject type (i.e., full NP, relative pronoun, negative quantifier, pronoun), and subject doubling, as illustrated in (14 a–d).

(14) a. <u>*Sa femme, elle*</u> *(ne)   voyage jamais.*   [NP + doubling]
his wife   she (NEG) travel never

b. <u>*Sa femme*</u> *(ne)   voyage jamais.*   [single NP]
his wife   (NEG) travel never
'His wife never travels'

c. <u>*Elle, elle*</u> *(ne)   voyage jamais.*   [pronoun + doubling]
her she (NEG) travel never

d. <u>*Elle*</u> *(ne) voyage jamais.*   [single pronoun]
She (NEG) travel never
'She never travels'

Other linguistic factor groups which were coded for are the occurrence of a following object pronoun, verb type (i.e., lexical verb, *avoir*, *être*, modal verbs such as *vouloir* 'to want', *pouvoir* 'to be able', *devoir/falloir* 'must', and *aller* 'to go', *venir de* 'to have just' when used as modals), and number of realized syllables between the potential *ne* location and the following negative element. With respect to negative elements, we coded for location, (cf. 15), and type of negative element (*pas* 'not',

plus/pus 'anymore', *point* 'not', *mie* 'not', *personne* 'anyone', etc.) in second (*il (n')a vu personne* 'He didn't see anyone') and third position (*il (n')a pas vu personne*), if any.[10]

(15) a. Negative before *ne*: <u>*Rien*</u> *(ne) marche.* 'Nothing is working'
      b. Negative after *ne*: *Claude (ne) regarde* <u>*rien*</u>. 'Claude looks at nothing'

We also distinguished frequent expressions such as *c'est* 'it is', *il y a* 'there is', *il faut* 'there must', *je sais pas* 'I don't know' – or their Picard equivalents – from other constructions. In order to test for interpersonal variation, we coded for individual speakers in both languages. Finally, two social factors that would tap into the influence of formality were added for the French data: interview portion (first 5 minutes, 5–10 minutes, remainder of interview), and interviewer. Even though both interviewers were female native speakers of Québec French, a variety known to have very little *ne* retention (Sankoff & Vincent 1977; Poplack & St-Amand 2007) that may have triggered higher rates of *ne* deletion in their interviewees' speech, their level of familiarity with the speakers differed, as one interviewer knew the speakers prior to the interviews, while the other did not.

5. Results

5.1 *Ne* deletion in Vimeu French

Our multivariate analysis of the Vimeu French data shows a global deletion rate of 79%, similar to Coveney's (1996) rate of 81.2% in data collected in 1980 in the Somme department. The structural constraints selected as significant for *ne* deletion also support previous findings in other French varieties. Table 3, which presents the results of our multivariate analysis, reflects the order in which the significant variables were selected.

In terms of linguistic factors involved, frequency had the strongest effect with frequently occurring expressions showing categorical or quasi categorical rates of *ne* deletion, as seen in Table 4. These rates are not only comparable with Coveney's (1996: 81) results, they also confirm the overall favoring effect of lexicalization on *ne* deletion in French described in much of the literature.

Subject doubling, a well-known feature of colloquial French (cf. Auger 1994; Coveney 2003, among others), also favors *ne* deletion in the variety of French spoken in Vimeu, thus supporting Pooley's (1996) results for Chtimi. This tendency, which

---

10. While we suspect that *point* and *mie* have slightly different meanings, the exact nature of this difference remains to be determined.

Table 3. Factor groups selected for *ne* deletion in French

Factor group	Factor	Weight	% Deletion	N
Frequency***	*il y AVOIR, il FALLOIR, ce/ça ÊTRE*	.916	98%	220/225
	je SAVOIR	.385	85%	44/52
	other constructions	.235	69%	299/434
Subject doubling***	NP + doubling	.733	91%	10/11
	pronoun + doubling	.640	83%	19/23
	single pronoun	.563	83%	422/506
	single NP	.150	35%	31/88
Speaker***	Jacques V.	.745	93%	208/224
	Guy D.	.732	92%	104/113
	Thomas S.	.504	78%	35/45
	Fabienne A.	.287	69%	79/114
	Annick M.	.266	68%	94/138
	Joseph L.	.193	56%	43/77
Negative placement***	after *ne*	.506	80%	562/703
	before *ne*	.098	13%	1/8
Subject person***	2SG *tu*	1.000	100%	5/5
	3PL	.697	59%	38/64
	1SG *je*	.599	80%	158/198
	3SG	.558	55%	54/98
	on	.546	77%	54/70
	2PL *vous*	.409	80%	4/5
	3SG neuter *ce/ça*	.291	90%	169/188
Intervening pronoun*	3rd accusative/dative pronoun	.746	81%	25/31
	other non-subject pronoun	.621	88%	120/136
	no non-subject pronoun	.450	77%	418/544
Interview portion**	after 10 minutes	.533	81%	473/581
	5–10 minutes	.450	73%	49/67
	first 5 minutes	.266	65%	41/63

Input: 0.902; Log likelihood = −236.574; Significance = 0.042; Convergence at Iteration 16

***: $p \leq 0.001$; **: $p = 0.0044$; *: $p = 0.0134$

Table 4. *Ne* deletion in frequent constructions

Construction	% Deletion	*Ne* deleted	*Ne* retained
il y AVOIR 'there BE'	100%	67	0
il FALLOIR 'there NEED'	100%	11	0
ce/ça ÊTRE 'it BE'	97%	142	5
je sais pas 'I don't know'	85%	44	8
all other constructions	69%	299	135

$x^2 = 76.362$; $p \leq 0.001$

is consistent with the fact that both *ne* deletion and subject doubling are typical of informal spoken French, also supports previous findings that pronouns favor deletion more than full subject NPs. Indeed, we can attribute the finding that undoubled NP subjects are the only type of subject to disfavor *ne* deletion (.150) to the fact that they are the only ones that do not contain a subject pronoun. Subject person was also found to play a significant role in *ne* deletion, but low token counts for some categories (e.g., second person *tu* and *vous*) prevent us from making any generalizations at this point.[11]

Other factor groups selected are negative placement and the presence of an intervening non-subject pronoun. A negative element occurring before *ne* strongly disfavors deletion, while the presence of an object pronoun has a favoring effect, the latter supporting Pooley's (1996) evidence for Chtimi. Furthermore, the fact that 3ACC/DAT pronouns more strongly favor *ne* deletion than other intervening pronouns parallels Vasseur's (1996) observation for Vimeu Picard.[12]

Two social factors were selected as significant in our analysis: speaker and interview portion. While we expected bilingual speakers to behave differently from French monolinguals (either because their increased awareness of the distinction between their two linguistic varieties would favor high deletion rates in French or because their low rates of *ne* deletion in Picard would be transferred into French), no

---

11. Although surprising at first glance, the disfavoring effect of the neuter pronoun *ce/ça* (.285, with 90% *ne* deletion) reflects the fact that most tokens of this pronoun are instances of *c'est*. Given that the remaining non-lexicalized tokens of *ce/ça* reveal a deletion rate comparable to what is observed in other non-lexicalized expressions (66%; 27/41), our multivariate analysis attributed the high deletion rate observed with *ce/ça* to lexicalization rather than to the pronoun itself.

12. In our analysis, we have opposed 3SG and 3PL accusative and dative clitics to all other non-subject clitics. Even though Vasseur (1996) only mentions the effect of 3SG clitics on *ne* deletion, he provides an example of *ne* deletion that contains a 3PL accusative clitic. We interpret this example as evidence that both singular and plural clitics favor *ne* deletion.

clear distinction between bilinguals and French monolinguals is apparent. Indeed, bilingual speaker Jacques V. and French monolingual Guy D. have the highest rates of *ne* deletion, while bilingual Joseph L. and monolingual Annick M. show the lowest rates of deletion. Our results for the factor "interview portion" confirm the effect of formality reported in prior studies of French: speakers increasingly dropped the negative particle as they became more comfortable during the interview.

5.2   *Ne* deletion in Vimeu Picard

While patterns of *ne* deletion in the variety of French under study are similar to those of other French varieties, *ne* in Picard behaves very differently from its French counterpart*,* with respect to both frequency of deletion and the linguistic factors involved. Our analysis of the written data confirms Vasseur's (1996: 88) claim that *ne* deletion is very rare in Picard. Table 5 shows that the deletion rate never exceeds 8% in the six authors analyzed and that the overall deletion rate in our sample is 4.2%. However, contrary to what is reported by Vasseur, this analysis reveals that the only context in which *ne* deletion is frequent is with *ch'est* 'it is', as in (16), rather than in clauses containing the clitics *l* '3SG.ACC' and *lé* '3PL.ACC', as in (17): 58.8% of deletion cases in our written corpus are instances of *ch'est point/ mie* 'it's not'.[13] Even in the writings of Gaston Vasseur himself, *ne* is usually present in clitic groups that contain a third person accusative or dative clitic.

(16) *ch'est mie dob- bave   ét crapeud*   (Vasseur, Lettes 008)
      it.is   not some spittle of toad
      'It's not toad's spittle'

(17) *Jé n'   l'érouos   point cru*   (Vasseur, Lettes 023)
      I   NEG it.would.have not   believed
      'I would not have believed it'

Ne deletion in spoken Picard reveals a hybrid system: while its frequency is much higher than in written Picard (39% deletion), we will see that its linguistic conditioning resembles more what we have observed in written Picard than what characterizes French. The results of our GoldVarb analysis are reported in Table 6.

Our GoldVarb analysis reveals both similarities and differences between Picard and French. First, in both varieties, frequency is the most significant factor group: lexicalized expressions such as *(i) feut* 'it is necessary'*, ch'est* 'it is', and *y a* 'there is' favor the absence of *ne*, while *je (ne) sais point* 'I don't know', which we coded separately due to the possibility that many of its instances function as a

---

13.   Given that we have collected only three tokens of dative *li* in negative clauses, we cannot evaluate the impact of this clitic on *ne* deletion.

Table 5. Rates of *ne* deletion in written Picard

Author	% Deletion	N
Touron	4.0%	8/199
Lecointe	4.0%	7/175
Vasseur	7.2%	18/250
Total pre-Ch'Lanchron authors	*5.3%*	*33/624*
Joseph L.	0.3%	2/586
Jacques V.	7.1%	24/363
Thomas S.	6.4%	18/280
Total Ch'Lanchron authors	*3.6%*	*44/1229*
Grand total	**4.2%**	**77/1853**

Table 6. Factor groups selected for *ne* deletion in spoken Picard

Factor group	Factor	Weight	% Deletion	N
Frequency ***	*y AVOÉR, il FOLLOÉR, ch'ÈTE*	.923	86%	127/147
	je SAVOÉR	.377	32%	23/71
	other constructions	.347	26%	137/520
Speaker ***	Thomas S.	.776	70%	83/118
	Joseph L.	.553	34%	52/151
	Marcel C.	.510	40%	44/109
	Jacques V.	.471	34%	75/216
	André L.	.331	26%	4/15
	Alain Q.	.243	22%	29/130
Negation ***	*pas*	.663	52%	72/137
	point	.522	40%	196/486
	pus	.395	22%	16/70
	other	.083	6%	3/44
Third negation **	absent	.511	39%	285/718
	present	.214	16%	4/24

Input: 0.374; Significance = 0.020; Convergence at Iteration 8

***: p ≤ 0.001; **: p = 0.0052

discourse marker, and other non-lexicalized verbs disfavor it. Second, in both varieties, deletion rates vary significantly across speakers. However, in Picard, only Thomas S. has a deletion rate that exceeds 50%: 70%. All other speakers have deletion rates between 22% and 40%. These rates contrast sharply with those obtained for French, which vary between 56% and 93%.

The next two factor groups selected by GoldVarb for Picard were not selected for French. In Picard, the negation adverb is the third most important variable: *pas* 'not' is the only adverb that truly favors deletion. *Point*, which also means 'not' but is clearly associated with Picard, also slightly favors deletion, but all other negative adverbs disfavor it. While *pas* also favors *ne* deletion in spoken French, we attribute this parallel effect to different causes: in French, *ne* deletion is most frequent with *pas* because this is by far the most common negative adverb (78% of all negative clauses). In Picard, *point* is the most frequent element with 65% of all negative clauses and *pas* is found in only 18% of the clauses. In light of the facts described above for written Picard, we interpret the connection between *pas* and *ne* deletion as evidence that *ne* deletion is associated with French rather than with Picard.[14]

Finally, the presence of a third negation, as illustrated in (18–19), strongly disfavors *ne* deletion. Two types of sentences can contain three negative elements in Picard. As in other languages, it is possible to combine negative adverbs that contribute different meanings; cf., e.g., (18) where the adverbs meaning 'anymore' and 'nothing' co-occur in the same clause. However, Picard also allows the combination of *mie* or, more rarely, *point*, two elements that do not convey any obvious meaning beyond negation, with other negative adverbs, as illustrated in (19). We attribute the lower *ne* deletion rate in both types of constructions to a subconscious desire on the part of the speakers to emphasize the negative meaning expressed in the clause by multiplying the number of negative elements overtly realized (cf. Fonseca-Greber 2007 concerning emphatic uses of *ne* in Swiss French).

(18) *i   n'   ont pus   rien   à manger*   (Joseph L. 30/6/96)
     they NEG have anymore nothing to eat
     'They don't have anything to eat anymore'

(19) *O   n'   a   mie janmoais vu   o*   (Joseph L., p. 33)
     one NEG has not never   seen that
     'We have never seen that!'

Ne deletion in Picard also differs from its French counterpart in other ways. For instance, subject doubling, which was the second most significant factor group for French, was not selected at all for Picard. Similarly, subject person and negative placement (i.e., before or after the verb), which were the fourth and fifth most significant factor group for French, were not selected for Picard. One additional fact about subject person further illustrates the difference between the two varieties.

---

14.   The use of *pas* as a French-influenced feature is corroborated by Pooley (1996: 171) in his description of Chtimi.

In Picard, the two third person singular neuter pronouns, *a/Ø* and *ch'*, have a very different impact on *ne* deletion: deletion is most frequent with *ch'* (89%) and rarest with *a/Ø* (24%). While phonology may account for the fact that deletion is more frequent following a consonantal clitic than a vocalic one, a hypothesis supported by the fact that third plural nominative *i* and third singular nominative *i* and *a* have the second lowest deletion rate (27%), the fact that *je* 'I' has a low deletion rate (32%), while *os* 'you.NOM.PL' has a high deletion rate (62%) provides evidence that more than phonology is involved.

6.   Do we really have two grammars?

Even though the analyses presented above provide some evidence that *ne* deletion operates differently in Picard and in Vimeu French, they do not necessarily prove that we are dealing with two distinct grammars. Because the French and the Picard samples analyzed both include heterogeneous groups of speakers, it is possible that the differences observed characterize subsets of speakers rather than all of them. As we saw earlier, three bilingual speakers who participate in activities that promote revitalization of the Picard language are analyzed in both languages. The remaining French data are provided by three monolingual speakers, while additional Picard data come from bilingual speakers who speak the language on a daily basis but are not involved in militant activities for its recognition or survival. Consequently, the possibility exists that the French system described above characterizes only or mostly the monolingual speakers, that the Picard system is due essentially to the non-militant speakers, and that the bilingual speakers' *ne* deletion operates similarly in both varieties. Thus, it is necessary to compare *ne* deletion in French and in Picard for our bilingual speakers.

A comparison of *ne* deletion in oral Picard and French for our three bilingual speakers reveals clear differences between the two varieties. Table 7 contains the individual deletion rates for our three speakers. Table 8 contrasts the factor groups

Table 7. Rates of *ne* deletion in the speech of three bilinguals

Speaker	Picard		French	
	% Deletion	N	% Deletion	N
Thomas S.	70%	83/118	78%	35/45
Joseph L.	34%	52/151	56%	43/77
Jacques V.	34%	75/216	93%	208/224
Total	*43%*	*210/485*	*82%*	*286/346*

selected as significant in either of the two GoldVarb analyses.[15] First, we observe that very different deletion rates characterize the two varieties: 43% in Picard and 82% in French. Individual deletion rates also reveal very interesting differences among the three speakers. While Thomas S. has similar deletion rates in both varieties, Joseph L. and Jacques V. clearly have different rates. The difference is particularly large for Jacques V. Table 8 confirms that, as was discussed in Section 3, many factors have similar effects in Picard and French.

Table 8. Factor groups selected for ne deletion in the spoken Picard of three bilinguals

Factor group	Factor	Picard		French	
		% deletion	N	% deletion	N
Frequency	*(il) y a/(il) faut/c'est*	98%	103/105	100%	112/112
	Je SAVOIR	32%	17/53	80%	21/26
	other constructions	27%	90/237	73%	153/208
Subject doubling	NP doubling	38%	8/21	85%	6/7
	pronoun doubling	12%	1/8	92%	12/13
	single pronoun	42%	173/404	86%	213/246
	single NP	25%	7/28	46%	20/43
Subject person	2SG	50%	3/6	100%	1/1
	3PL	28%	38/74	77%	21/27
	1SG	33%	37/110	84%	88/104
	3SG	33%	38/112	6%	27/48
	On	29%	18/61	67%	19/28
	2PL	66%	4/6	100%	3/3
	3SG. neuter	24%: *a/Ø*	13/53	n/a	n/a
		91%: *ch'*	65/71	93%	92/98
Object pronoun	Non-3ACC/DAT	62%	23/37	93%	14/15
	3ACC/DAT	43%	7/16	88%	53/60
	No pronoun	41%	180/431	80%	219/271
Negative placement	After *ne*	43%	208/475	83%	285/342
	Before *ne*	33%	2/6	25%	1/4
Negation	*Pas*	82%	29/35	86%	240/277
	Point	43%	167/382	n/a	n/a
	P(l)us	25%	11/43	64%	24/37
	Others	12%	3/25	69%	22/32

15. Table 8 reports percentages rather than GoldVarb weights due to the fact that we were unable to obtain convergent models for the French data.

 More importantly, however, it confirms that our bilingual subjects have internalized different rules. For instance, while doubling favors *ne* deletion in Picard and in French, we see that its effect is stronger in French, where doubling is a feature of colloquial speech, than in Picard, where subject doubling is the norm (cf. Auger 2003b). Similarly, we see that while lexicalized phrases favor *ne* deletion categorically or near-categorically in the two varieties, the rates for *je + SAVOIR* and other, non-lexicalized, constructions differ greatly in Picard and in French. Concerning person, the most striking difference between Picard and French is found in the neuter pronouns: while *c'est/ch'est* both favor deletion, Picard *a/Ø* has the lowest deletion rate of all persons. Finally, the results concerning the negator confirm the distinction made by Picard speakers between *pas*, an adverb that is associated with French and that greatly favors *ne* deletion, and *point*, a form associated with Picard and which shows a much lower rate of *ne* deletion.

Even though previous research has provided ample evidence that Picard and French have different structures and constitute different languages, we felt that additional research which would look at a linguistic structure unlikely to be the object of conscious manipulation would strengthen the conclusion reached in those studies. Indeed, given the revitalization process that has characterized Picard in recent decades, the possibility exists that speakers and authors consciously choose words or grammatical constructions that clearly distinguish Picard from French, its close relative. Comparisons of different versions of the same texts (either pre- vs. post-publication or older and more recent versions; cf. Auger 2003b) and interviews with editors and authors provide evidence that this strategy is part of the standardization process that can be observed in the literature. *Ne* deletion appeared to be an ideal variable for this study because of the fact that, contrary to other variables for which the standard Picard form clearly differs from standard French, the standard variant in this case is the same as in standard French. Thus, if the distinct features observed in Auger (2003a,b) result from the efforts of militants and authors to maximize the distance between Picard and French, what we should observe in our Picard data is a tendency toward *ne* deletion. In this sense, the tendency not to delete *ne* that our results have revealed constitutes, in our opinion, strong evidence that, in spite of their obvious similarities, Picard and French do constitute different languages.

References

Armstrong, N. 2002. Variable deletion of French *ne*: A cross-stylistic perspective. *Language Sciences* 24: 153–173.

Armstrong, N. & Smith, A. 2002. The influence of linguistic and social factors on the recent decline of French *ne*. *Journal of French Language Studies* 12: 23–41.

Ashby, W.J. 1976. The loss of the negative morpheme *ne* in Parisian French. *Lingua* 39: 119–137.

Ashby, W.J. 1981. The loss of the negative particle *ne* in French: A syntactic change in progress. *Language* 57(3) : 674–687.

Ashby, W.J. 2001. Un nouveau regard sur la chute du *ne* en français parlé tourangeau: S'agit-il d'un changement en cours? *Journal of French Language Studies* 11: 1–22.

Auger, J. 1994. Pronominal Clitics in Colloquial Québec French: A morphological approach. Ph.D. dissertation, University of Pennsylvania. (Distributed by the Institute for Research in Cognitive Science, University of Pennsylvania, IRCS Report No. 94–29).

Auger, J. 2002. Picard parlé, picard écrit: Dans quelle mesure l'écrit représente-t-il l'oral? In *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache / Romance corpus linguistics: Corpora and spoken language*, C. Pusch & W. Raible (Eds), 267–280. Tübingen: Narr.

Auger, J. 2003a. Picard parlé, picard écrit: Comment s'influencent-ils l'un l'autre? In *Le picard d'hier et d'aujourd'hui* [Bien dire et bien Aprandre 21], J. Landrecies & A. Petit (Eds), 17–32. Villeneuve d'Ascq : Centre d'Études médiévales et Dialectales de Lille 3.

Auger, J. 2003b. The development of a literary standard: The case of Picard in Vimeu-Ponthieu, France. In *When Languages Collide: Perspectives on language conflict, language competition, and language coexistence*, B.D. Joseph et al. (Eds), 141–164. Columbus OH: Ohio State University Press.

Auger, J. In press. Issues of authenticity, purity, and autonomy in minority languages: What is "real" Picard, and who is an "authentic" speaker? In *Minority and Diasporic Languages of Europe* [Berkeley Linguistics Society 29], P. Nowak & C. Yoquelet (Eds), Berkeley CA: Berkeley Linguistics Society.

Auger, J. & Villeneuve, A.-J. Submitted. L'épenthèse vocalique en picard et en français. In *Proceedings of GalRom07 Conference*, G. Brun-Trigaud et al. (Eds), Paris: Presses universitaires de Vincennes.

Bell, R.T. 1976. *Sociolinguistics: Goals, approaches and problems*. London: Batsford.

Blanche-Benveniste, C. 1997. La notion de variation syntaxique dans la langue parlée. *Langue française: La variation en syntaxe* 115: 19–29.

Carton, F. 1981. Les parlers ruraux de la région Nord-Picardie: situation sociolinguistique. *International Journal of the Sociology of Language* 29: 15–28.

Cerquiglini, B. 1999. Les langues de la France. Rapport au Ministre de l'Éducation Nationale, de la Recherche et de la Technologie, et à la Ministre de la Culture et de la Communication. http://www.culture.gouv.fr/culture/dglf/lang-reg/rapport_cerquiglini/langues-france.html.

Coveney, A. 1996. *Variability in Spoken French: A sociolinguistic study of interrogation and negation*. Bristol: Elm Bank Publications.

Coveney, A. 2003. Le redoublement du sujet en français parlé: une approche variationniste. In *Structures linguistiques et interactionnelle*s du *français parlé*, A.B. Hansen & M.-B. Mosegaard Hansen (Eds), 111–143. Copenhagen: Museum Tusculanum Forlag.

Dawson, A. 2002. Histoire du picard. http://www.picard.free.fr/lgpic/Histoire.htm

Dawson, A. 2003. *Le picard de poche*. Chennevières-sur-Marne: Assimil.

Debrie, R. 1983. *Eche pikar bèl é rade*. Paris: Omnivox.

Dewaele, J.-M. 2004. Retention or omission of the *ne* in advanced French interlanguage: The variable effect of extralinguistic factors. *Journal of Sociolinguistics* 8(3): 433–450.

Diller, A.-M. 1983. Subject NP and variable constraints: The case of *ne* deletion. In *Language Use and the Uses of Language*, R.W. Shuy & A. Shnukal (Eds), 68–75. Washington DC: Georgetown University Press.

Flutre, L.-F. 1955. *Le parler picard de Mesnil-Martinsart (Somme)*. Genève/Lille: Droz/Giard.

Fonseca-Greber, B.B. 2007. The emergence of emphatic 'ne' in conversational Swiss French. *Journal of French Language Studies* 17: 249–275.

Hansen, A.B. & Malderez, I. 2004. Le *ne* de négation en région parisienne: Une étude en temps réel. *Langage et société* 107: 5–30.

Lavandera, B. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7: 171–182.

Lecointe, A. 1970. *Farces d'éch' Mawais d'A'ry.* Fressenneville: Imprimerie Carré.

Lecointe, A. 1982. *La mémoire paysanne; Anciens travaux des champs et des fermes,* 1ère partie. Amiens: CNDP.

Ledieu, A. 1909/2003. *Petite grammaire du patois picard*. Bouhet: La découvrance.

Manzano, F. 2005. Les langues régionales de France sont-elles égales dans le recul? *Marges Linguistiques* 10: 133–156.

Moreau, M.-L. 1986. Les séquences préformées: Entre les combinaisons libres et les idiomatismes. Le cas de la négation avec ou sans *ne*. *Le français moderne* 54: 137–160.

Nettle, D. & Romaine, S. 2000. *Vanishing Voices: The extinction of the world's languages.* Oxford: OUP.

Pohl, J. 1968. *Ne* dans le français parlé contemporain: les modalités de son abandon. *Actas XI Congreso internacional de linguistica y filologia romanicas*, 1343–1359. Madrid: Consejo Superior de Investigaciones Científicas.

Pohl, J. 1975. L'omission de *ne* dans le français contemporain. *Le français dans le monde* 111: 17–23.

Poignant, B. 1998. Langues et cultures régionales, rapport de Bernard Poignant, maire de Quimper à Monsieur Lionel Jospin Premier Ministre. http://membres.lycos.fr/insanne/poignant/poignant.PDF

Pooley, T. 1996. *Chtimi: The urban vernaculars of Northern France.* Clevedon: Multilingual Matters.

Poplack, S. & St-Amand, A. 2007. A real-time window on 19th-century vernacular French: The Récits du français québécois d'autrefois. *Language in Society* 36: 707–734.

Regan, V. 1996. Variation in French interlanguage: A longitudinal study of sociolinguistic competence. In *Second Language Acquisition and Linguistic Variation*, R. Bayley & D. Preston (Eds), 177–201. Amsterdam: John Benjamins.

Rehner, K. & Mougeon, R. 1999. Variation in the spoken French of immersion students: To *ne* or not to *ne*, that is the sociolinguistic question. *The Canadian Modern Language Review* 56(1): 124–154.

Sankoff, G. & Vincent, D. 1977. L'emploi productif du *ne* in spoken Montréal French. *Le français moderne* 45: 243–256.

Sax, K. 2003. Acquisition of Stylistic Variation in American Learners of French. Ph.D. dissertation, Indiana University.

Touron, R. [alias Val'ry ch'Bédeu]. 1982. *Eune flopèe d'mintiries; Contes in picard éd Nibos. Vimeu*. Fressenneville: Imprimerie Carré.

Trévise, A. & Noyau, C. 1984. Adult Spanish speakers and the acquisition of French negation forms: Individual variation and linguistic awareness. In *Second Languages: A cross-linguistic perspective*, R. Andersen (Ed.), 165–189. Rowley MA: Newbury House.

Vasseur, G. 1996. *Grammaire des parlers picards du Vimeu (Somme): Avec considération spéciale du dialecte de Nibas*. Abbeville: F. Paillart.

Vasseur, G. 2003. *Lettes à min cousin Polyte*. Abbeville: F. Paillart.

Vigneux, J.-L. 2007. Et le picard du Vimeu a montré sa littérature! Ms., October 2007.

The dynamics of pronouns in the Québec languages in contact dynamics

Hélène Blondeau
University of Florida

This article examines three zones of variation in the French pronoun paradigm in relation to the general issue of language contact in Québec. Three variables, associated with language change in contemporary Montreal French, are analyzed: (1) variation between the pronouns *on* 'one' and *nous* 'we' expressing the first-person plural, (2) variation among *on* 'one', *tu* 'you', and *vous* 'you' to express indefinite reference, and (3) variation between simple and compound forms of plural pronouns with *-autres* 'others'. The article reassesses the situation using data that represent different degrees or dimensions of contact between French and English. A comparison between nineteenth- and twentieth-century Québec French data highlights how certain tendencies were or were not yet implemented at an earlier stage when the contact between French and English was less intense. The article also examines how young bilingual Montrealers behave in relation to these changes in progress.

1.   Introduction

In the early 1970s when the sociolinguistics of French was emerging as a new trend in French linguistics, the received wisdom regarding the language situation in Québec was that the French language was corrupted by contact with the English language. One of the main contributions of Gillian Sankoff and her collaborators, Henrietta Cedergren and David Sankoff, was to document the variety of spoken French in Montreal based on a representative sample of the Francophone population, and to provide linguistic evidence by analyzing several variables at different levels of the linguistic structure, that the observed variation was conditioned by linguistic and extralinguistic factors in a very systematic way. A decade later, in her book *The Social Life of Language*, Gillian Sankoff (1980) reported on the Montreal French variety, among others, including seminal work in collaboration with Suzanne Laberge on

pronominal variation (Laberge & Sankoff 1980). In the introduction to that article, the polysemic nature and the pragmatic versatility of some pronominal forms were highlighted, and the conditioning of the variation at work in the French pronominal subsystem was explicitly analyzed. Since then, many sociolinguists have looked at different aspects of pronominal variation, not only in Montreal French (Thibault 1983, 1991; Vincent et al. 1985; Auger 1994; Blondeau 1999, 2001) or in Québec City French (Deshaies & Ouellet 1982), but also in other varieties of French spoken in Canada (King 1983; Nadasdi 1995) as well as in varieties of spoken French in Europe (Ashby 1992; Coveney 2003; Fonseca-Greber & Waugh 2003). This fruitful body of work confirms that the French pronominal subsystem is highly productive for the study of language variation and change.

This article revisits the variation at work in three zones of the French pronouns paradigm in relation to the general issue of languages in contact in Québec.[1] The three variables investigated here have each been reportedly involved in a change in progress in the Montreal French speech community (Laberge 1977; Thibault 1991; Blondeau 1999, 2001): (1) the variation between the pronominal forms *on* 'one' and *nous* 'we' expressing the first person plural clitic subject with a definite reference, (2) the variation among the pronouns *on* 'one', *tu* 'you', and *vous* 'you' to express an indefinite reference, and (3) the variation between simple and compound forms with *autres* 'others' of plural nonclitic pronouns. These three variables have been associated in different ways with processes of simplification or regularization. As we know from the literature, simplification and regularization are often associated with hypotheses of contact-induced changes (Chaudenson et al. 1993; Mougeon & Beniak 1991). However, those processes could also be part of an internally motivated change (Thibault 1991), and consequently processes *per se* cannot be used as a diagnosis of contact-induced change. Indeed, in order to disentangle if a change is induced by contact or internally motivated, one needs access to points of comparison that represent different situations of language contact. This is precisely the aim of this article: to reexamine the situation for the three previously identified variables based on an extended body of spoken French data that represents different degrees or dimensions of contact between French and English.

The present approach looks at the situation from two angles. On the one hand, the pronominal variation is analyzed in real time, comparing spoken French data that represents nineteenth- and twentieth-century Québec French, expanding the time span and the historical scope of the phenomena under study.[2] These diachronic analyses aim to demonstrate how certain tendencies were or were not yet implemented at an earlier stage of the language in a context where contact between French and English was less intense than it was in Montreal during the last part of the twentieth century. On the other hand, the same three variables are examined from a synchronic point of view in the French of Anglophone Montrealers, a community of bilingual speakers that is now recognized to have incorporated French into its repertoire. These analyses aim to shed light on how bilingual speakers, in daily contact with French, behave regarding the use of variants that are involved in changes in progress in the community as a whole. In research on the acquisition of sociolinguistic competence, "whether the behavior of the L2 speaker is the same in a case of a stable sociolinguistic variable as in the case of a variable that is undergoing change" remains a question to be explored (Bailey & Regan 2004: 325). This synthesis aims to shed light on this question.

2.   Background

Everyone interested in the language dynamics in Québec recognizes that contact situations are far from exceptional, and in fact have been part of the linguistic landscape at least since the arrival of the first French or British colonists in America. Apart from the numerous attested historical reports concerning contact between the newcomers and the native populations at the time of settlement, we also know from historical data that during the New France period, the linguistic landscape in North America partially reflected the contact situation in the French Metropolis (Mougeon & Beniak 1994; Martineau 2005). Furthermore, it is impossible to deny that contact between the French and English after the British conquest of Canada (1760) had a considerable impact; since then the constant contact between migrant populations and English and French has always been important, particularly in urban contexts.

If we want to classify the contact situation between French and English from a sociohistorical point of view, we would consider it a case of populations of distinct languages who have shared a common territory for a long period of time – in other words, a case of long-term coterritoriality (Sankoff 2001). Another important consideration is related to the characterization of the contact with regard to the notion of bilingualism. As far as Québec is concerned, although long-term social bilingualism has certainly played a role, it does not necessarily imply individual bilingualism. Indeed, depending on social factors such as social class, regional origin, or ethnic affiliation, the nature of individual bilingualism has remained highly variable. In addition, the reciprocal positioning of French and English has evolved over time. For example, the changes in the linguistic landscape as illustrated by modification in the uses of language on signs since the so-called "Quiet Revolution" highlight the transformations that occurred in a very short period of time. Although social bilingualism was certainly part of the picture, it would be erroneous to describe it in terms of stability.

3.   Theoretical considerations

Linguistic outcomes of language contact do not appear suddenly. Instead, they are triggered by daily adjustments between speakers in different language situations (Goffman 1981). But how can contact-induced changes be distinguished from internally motivated changes that are also triggered by ordinary interaction in the speech community? As a contribution to this debate empirical evidence is provided regarding the path of changes in the area of pronominal variation in the Québec language contact situation. As a working hypothesis, we state that if language contact is a consequence of multiple day-to-day adjustments, a comparison of a sample of speakers at different points in time and experiencing different sociolinguistic situations will shed light on the effect of language contact.

In contact settings, it is often believed that variation in a language A is due to language contact or linguistic transfers from a language B. For example, over the years many cases of variation attested in varieties of French in North America have been associated with contact with English. However, for many variables outside of the lexicon, the validity of the hypothesis of contact-induced change has not been established beyond any doubt.

King's demonstration (2000) that an apparent structural borrowing from English to French is in fact initially triggered by lexical borrowing is particularly convincing. The emergence of *back* in Prince Edward Island French is viewed as

a lexical innovation and further analyzed in relation to its impact on language-internal structural change. Additionally, Gadet and Jones, (2008), who looked at the contact-induced phenomenon for variables in the area of syntax, have pointed out that many variants of French in North America supposedly associated with the influence of English are in fact features of popular or regional French in the Hexagon, where the influence of English is not suggested as a source. Therefore, before accepting an explanation involving the effect of contact in the area of morphosyntax or syntax, one also has to consider the intrasystemic dynamics of inherent variability. In other words, instead of immediately pointing in the direction of diffusion one has to consider if the case of variation in question might instead be involved in the general transmission processes that trigger general linguistic changes (Labov 2007).

Among the different outcomes associated with language contact in the area of morphosyntax, reduction of morphological oppositions or reduction of grammatical categories are often identified as products of the simplification processes, while paradigmatic symmetry or absence of variation in agreement rules are often associated with the regularization process (Thibault 1991). With regard to the French pronominal paradigm, previous research associated tendencies in the pronominal system as pointing toward simplification or regularization of the system. For example, Chantefort (1976) argued that the tendency to replace *nous* by *on* signified a reduction of the verbal paradigm of French conjugation. The supposed disappearance of the opposition between *tu* and *vous* is also seen by Chaudenson (1998) as evidence of simplification. Finally, the adoption of the nonclitic forms with *-autres* is viewed by Chaudenson (ibid.) as an exemplification of regularization. These arguments regarding simplification or regularization are often followed by an interpretation in terms of convergence toward English without necessarily basing this assumption on empirical evidence. This direct association is questionable.

One of the problems of this association is related to the nature of the comparison. Too often those hypotheses of simplification or regularization have been developed by comparing features present in varieties of French in North America, where English is part of the linguistic landscape, to the idealized standard French social construct that is considered pure and clear of contact-induced influences. If comparison is indeed necessary to assess linguistic change, one needs to have access to a comparable body of data. For example, a presumed contact variety has to be compared with data representing a pre-contact stage of the language. In addition, the comparability of the data sets has not only to be justified from a historical point of view; it has also to be explained with regard to the sociological characteristics of the samples.

4. Methodology and study design

The current research uses the variationist sociolinguistic framework, which adopts a community-based approach, and implies a quantitative paradigm. While one cannot deny that any linguistic changes require innovation from the individual speaker as a point of departure, the focus of this paper is the propagation of a variant within a given community. According to the perspective adopted here, evidence of linguistic change, whether induced by contact or related to internal motivation, needs to be established on a representative sample of the population. This sampling approach is characteristic of classical sociolinguistic corpora. For the present investigation, comparisons will be drawn from three bodies of data that represent different situations of contact between French and English in Québec in order to examine the relations between intrasystemic and contact-induced changes.

Two of the data sets represent Québec French as spoken by native speakers at two points in time, and the third represents the variety of French used today by Anglophone Montrealers. Comparisons are drawn between corpora of contemporary spoken French and a corpus of data representing nineteenth-century French. For both corpora, the informants are native speakers of French, but it is worth noting that in the context of nineteenth-century rural Québec the contact situation with English was far less intense than in Montreal for the last part of the twentieth century.

The twentieth-century situation is documented by the three interrelated sociolinguistic corpora of Montreal French (Sankoff et al. 1976; Thibault & Vincent 1990; Vincent et al. 1995).[3] For this period, references are made to previous analyses of Montreal French corpora, which were collected at the end of the twentieth century (between 1971 and 1995). For the variation between *on* 'one' and *nous* 'we', I refer to Laberge's study based on the 1971 Sankoff-Cedergren corpus (1977). For the variation among *on* 'one', *tu* 'you', and *vous* 'you', I refer to Laberge and Sankoff (1980) and Thibault (1991), respectively based on the Sankoff-Cedergren corpus and on the comparison with the real-time follow-up made possible by the Montreal '84 corpus. Regarding the variation among the nonclitic pronouns, I refer to previous results based on the three corpora of Montreal French collected in 1971, 1984, and 1995 (Blondeau 1999, 2001).

The variation for the three variables is then compared with the nineteenth century situation, as instantiated in the corpus Les Récits du Français Québécois

d'Autrefois (RFQ) (Poplack & St-Amand 2007).[4] This corpus is comprised of forty-four interviews from native speakers of Québec French born in the nineteenth century (between 1846 and 1895), and living in rural areas of the province of Québec. The recordings include folk tales, legends, and personal interviews selected from the original collection of the folklorists Luc Lacoursière and Carmen Roy. These data are considered to be representative of the French spoken in the nineteenth century based on the premise underlying the apparent time construct (Poplack & St-Amand 2007: 709).

The other angle from which we approach the variation involves analyses of the variety of French spoken by young Anglophone Montrealers (AM) recorded at the end of the twentieth century (Sankoff et al. 1997; Blondeau, Nagy, Sankoff & Thibault 2002).[5] This corpus was collected from 1993 to 1994 among speakers aged eighteen to thirty-five years. This generation of speakers is the first one that had access to the immersion program in Montreal developed in the late 1960s. The speakers in this sample have different degrees of exposure to French in terms of school experience and daily interactions in the city (Sankoff et al. 1997). According to the analysis of their representations, we also know that this generation of speakers considers French as an integral part of their linguistic repertoire, although some signs of linguistic insecurity appear in their discourse (Thibault & Sankoff 1993). The fact that today they consider themselves bilingual plays a role in their identity construction. This is an important change compared with older generations of AM, who were not necessarily bilingual and had fewer contacts with the French language despite the majority status of French in the Québec province (Poplack et al. 2006). As bilinguals, the generation under study is currently developing its own variety of French in which substratum influences from English are likely to appear and interact with other tendencies at work in Québec French.

5.   Pronominal variation

As mentioned in the introduction, the three variables under study are hypothesized to have undergone a process of change. In an apparent time study, Laberge (1977) pointed to the replacement of *nous* 'we' by *on* 'one' as a first-person plural

clitic subject, a change near completion according to her interpretation. It was also demonstrated based on apparent- and real-time evidence that the pronominal form *on* 'one' used with an indefinite reference was declining in favor of second-person pronouns, in particular the *tu* 'you' form (Laberge 1977; Laberge & Sankoff 1980; Thibault 1991). In addition, a real-time study of the alternation between simple and compound forms of nonclitic pronouns demonstrated that the commonly used compound forms with *-autres* were declining in Montreal French (Blondeau 1999, 2001). The following subsections discuss each variable from the two angles of comparison, starting first with the comparison between nineteenth- and twentieth-century data, and then looking at the linguistic behavior of bilingual Anglophone Montrealers with regard to the same variables.

5.1 The variation between *nous* and *on* as clitic subject with definite referent

The first variable for our comparison corresponds to the variation between *nous* and *on* as clitic subject with definite referent, as defined in the thorough study by Laberge (1977). For this comparison the direct and indirect object clitics have been excluded from the variable context as well as nonclitic forms.

Laberge identified a correlation between age and the variation, from which she inferred a change in progress based on the apparent time construct (Bailey et al. 1991). In her 1971 data, while most of the speakers opted for the *on* form, *nous* was more frequent in the speech of older speakers. In the data representing the nineteenth century, both variants were also present, sometimes in the speech of the same speaker, as illustrated in (1).

(1) a.   *Astheure <u>on</u> va aller voir mes parents.* (RFQ, 4) [6]
          'Now <u>we</u> are going to see my parents.'

     b.   *<u>Nous</u> avons trouvé une petite bête, elle avait une grande queue.* (RFQ, 4)
          '<u>We</u> found a little animal, it had a long tail.'

Table 1 displays the results for the use of *on* in each body of data. As illustrated in the table, the variant *on* dominated the variable spectrum, being used 99.75 percent of the time in the RFQ Corpus representing the nineteenth century. Out of 3,292 relevant occurrences, the rare variant *nous* was only used eight times, corresponding to 0.25 percent (Blondeau 2003).

---

6. The name of the corpus and the identification number or the pseudonym of the informant are provided in parentheses.

Table 1. Frequency of *on* as a first-person plural clitic with definite reference

Native speaker		Anglo-Montrealer
19th century	20th century	20th century
99.75%	98.4%	97.3 %

This shows that at the time *on* was already widely distributed in the community. Indeed, this rate is very close to the 98.4 percent rate of *on* usage observed in the data representing the twentieth century analyzed by Laberge (1977). According to her interpretation, this change in progress in Montreal French had almost reached completion in 1971, but as one can see by comparing the nineteenth- and twentieth-century data, the difference between the two is negligible. The comparison between the nineteenth and twentieth centuries shows more stability in the general rate of use than initially expected. Indeed, the variation in the corpus representing the nineteenth century shows that *nous* was already a rare form at the time. Based on the rate of usage, it is consequently impossible to consider the nineteenth century as a transition stage in the process of change. Instead, this empirical evidence indicates that the reported change was advanced and in fact already near completion in the nineteenth century. Rather than confirming a hypothesis of a change in progress this diachronic comparison indicates stability over time. In other words, *nous* 'we' was already a relic of a previous stage of the French language.

Considering now the results of the comparison with regard to the contact situation, the hypothesis that the decline of *nous* usage is due to an increasing influence of English in the twentieth century has to be questioned. The informants of the RFQ corpus were not from areas where contact with English was common, and consequently their usage cannot presuppose the effect of contact with English (for a detailed discussion, see Poplack & St-Amand 2007: 718). The fact that the RFQ speakers already usually used *on* instead of *nous,* even slightly more than in the twentieth century, discredits a hypothesis of contact-induced change. In addition, comparison with other varieties of French outside of North America shows that *on* is also a very common variant, for example, in the variety of French spoken in Picardie (Coveney 2000). For these two reasons, the influence of English cannot be considered as an explanatory factor for the presence of *on* in contemporary Québec French.

Looking at the same variable from another angle provides an interesting perspective. As mentioned previously, today's Anglophone Montrealers have integrated French into their repertoire; this is particularly obvious with the generation under study since many of these young Anglo-Montrealers report using French on a day-to-day basis. Both variants were identified in the corpus of Anglo-Montreal French (AMF), as illustrated in (2).

(2) a. *En dixième ou quelque chose de même <u>on</u> a fait un spectacle puis c'était fun.*
        (AMF, Joan)
        'In tenth grade or something like that <u>we</u> did a performance and it was fun.'

    b. *Maintenant <u>nous</u> avez un petit bébé de quatre mois et c'est <u>on</u> trou: <u>on</u> a pas*
        besoin d'une vie externelle maintenant <u>nous</u> sommes occupés avec. (AMF,
        Donald, example from Blondeau, Nagy, Sankoff & Thibault 2002)
        'Now <u>we</u> have a little baby that's four months old and it's: <u>we</u> don't need an
        outside life now <u>we</u> are busy with (the baby).'

As reported in Table 1, Anglophone Montrealers use *on* 97.3 percent of the time.
They highly favor *on*, a tendency corresponding to the usage of native French
speakers in the community. This rate is very close to that of 98.4 percent iden-
tified for the data representing the twentieth century. The general tendency for
both groups to use *on* as a first-person plural clitic with a definite referent offers
a striking parallel. In other words, Anglophone Montrealers have clearly adopted
the community norm for this variable. These results differ from those of Rehner
et al. (2003), who identified a balanced usage of the two variants for advanced
learners of French in an immersion context in Toronto.

    In addition the speakers who use standard *nous* 'we' are those with lower scores
on the environment scale and on the grammatical competence scale (Blondeau,
Nagy, Sankoff & Thibault 2002). The least competent speakers are those who had
the strongest tendency to retain the *nous;* some are still struggling with the verbal
morphology associated with the use of standard *nous,* as illustrated in example (2b).
This pattern of usage suggests that some speakers have not yet identified the com-
munity trend of preferring *on* over *nous*. According to our analysis, Anglophones
with more contact with French tend to behave more like native Francophone
speakers in the community.

    Looking at the use of the variable among Anglophone Montrealers has pro-
vided another perspective on the contact situation. L2 speakers with enough
contact with French clearly have adopted the community norm for this stable zone
of variation in Quebec French. In the project on the young Anglo-Montrealers
(see Sankoff et al. 1997; Blondeau, Nagy, Sankoff & Thibault 2002; Blondeau &
Nagy in this volume), we reported on several other variables in the linguistic rep-
ertoire of the community. For most of the variables, there was a clear correlation
between the individual level of contact with French and the use of variants in the
community.

5.2   The variation of clitic subject pronouns with indefinite reference

The second variable involved in our comparison concerned the variation among
the pronouns expressing a generic indefinite reference. The variants under study for
this variable are: *on* 'one', *tu* 'you', and *vous* 'you' to express an indefinite meaning.

The pronoun *on* used with an indefinite meaning corresponds to the prescription of standard French for this linguistic function. However, in spoken French *tu* and *vous* can also easily convey an indefinite reference. All three variants are common in contemporary spoken French, as illustrated in (3):

(3)   a.   *<u>On</u> recherche plus la paix tu sais quand <u>tu</u> t'approches de la soixantaine.*
            '<u>One</u> looks for peace, you know, when <u>you</u>'re close to the sixties.' (Examples cited in Thibault 1991: 87)

      b.   *Quand <u>on</u> était riche ça allait bien mais quand <u>vous</u> étiez pauvre.*
            'When <u>we</u> were rich it was fine but when <u>you</u> were poor.' (Examples cited in Thibault 1991: 87)

The distribution of the forms in the Montreal data showed robust variation in the 1971 data. (Laberge 1977; Laberge & Sankoff 1980). Far from being considered a case of stable variation, this case was associated with a change in progress within the community. Based on the differential behavior of younger and older speakers, an expansion of the second-person forms, in particular the *tu* 'you' form, was identified. According to the interpretation of the social distribution of the forms, this change in progress was led by young men from working-class backgrounds and therefore was classified as a change from below. On the contrary, older speakers were more likely to retain the *on* 'one' form, showing a more conservative usage. Laberge and Sankoff also identified a significant correlation with lexical and syntactic factors. In addition, they found that discursive and pragmatic constraints played an important role in the variation. These factors will be discussed in more detail in the context of the comparison between the French of native speakers and Anglophone Montrealers.

In a real-time study comparing the 1971 and the 1984 data, Thibault looked at the distribution of the second-person pronouns *tu* and *vous,* investigating the reported disappearance of the opposition between the two pronouns. Her hypothesis was that a stylistic opposition would be maintained at the level of the addressee pronouns but would be neutralized in other functions (1991: 86). She found that the form *vous* was mainly used as an addressee term, but that the form *tu* was much more multifunctional. In particular, the form *tu* with indefinite reference dramatically increased between 1971 and 1984 as did use of the discursive marker *tu sais* 'you know'. Her analysis of the alternation between *tu* and *vous* therefore confirms the hypothesis of an increase of *tu* as an indefinite pronoun, which was developed within the apparent time perspective. What was particularly enlightening in Thibault's real-time study was the fact that she interpreted the results in the dynamic perspective of a restructuring of the clitic subject pronouns paradigm as a whole. She therefore considered that the association of the variation with a complete neutralization of the opposition or with a simplification did not reflect the complexity of the balanced adjustments in the pronominal subsystem.

In a previous study focusing on lifespan change conducted in collaboration with Gillian Sankoff, we were able to trace the path of this change at the individual level from 1971 to 1995 (Blondeau, Sankoff & Charity 2002) and to show how some individuals were highly active in this change in progress, thus confirming previous findings on the same variable. The fact that it was possible to identify adjustments by individual speakers in the course of the lifespan was a clear sign that the change from below was still progressing. The evidence of a community change coupled with the evidence of lifespan change (Sankoff 2005) converge toward an interpretation of a vigorous change in progress. By implementing additional studies adopting this longitudinal perspective, namely the change from apical to dorsal, /r/ we are able to propose that "to the extent that older speakers change in the direction of change in progress in their adult lives, apparent time underestimates the rate of change" (Sankoff & Blondeau 2007: 582). In sum, the data representing the twentieth century all converge toward an interpretation of a recent and vigorous change in progress in the speech community.

Going back in time highlights the historical path of the change. The data representing the nineteenth century, an earlier stage of spoken French, display a different picture. The distribution of *tu* and *vous* in the RFQ shows that both pronouns are used as addressee forms more than 90 percent of the time as well as as discourse markers. However the use of second person pronouns with indefinite reference is almost absent from the corpus. The discourse marker function represents less than 10 percent of the usage. It is worth noting that discourse markers are frequent with the verb *savoir* 'to know' but also with other verb like the verb *voir* 'to see' as illlustrated in the following examples in (4).

(4) a. *Elle s'est mis à dépérir <u>vous savez</u>.* (RFQ, 4)
         'She started to waste away <u>you know</u>.'

     b. *Il passait un missionnaire <u>savez</u>.* (RFQ, 36)
         'A missionnary went by <u>know</u>.'

     c. *J'avais une grosse chevelure <u>savez-vous</u>.* (RFQ, 20)
         'I had big hair <u>you know</u>.'

     d. *Tout d'un coup <u>vois-tu</u> ça crie.* (RFQ 20)
         'All of a sudden <u>you see</u> it yells.'

However the discourse marker *tu-sais*, involved in a dramatic change in the twentieth-century (Thibault 1991), was in competition with other forms constructed with *vous* such as *vous savez* or *savez vous,* two common forms in the RFQ. This has to be interpreted in the general context of the pattern of addressee forms of the nineteenth-century data. Even more important for our purpose, the generic function of second-person pronouns is almost absent from the data. Therefore

it seems that the pronoun *on* 'one' had to fulfill the indefinite pronominal function in the nineteenth-century data.

Since the variable was not common enough in the nineteenth-century data we cannot further pursue the analysis of the linguistic factors involved in the variation and compare them to Laberge and Sankoff's findings. The only thing we can say is that the near absence of second person pronouns with indefinite reference confirms the hypothesis that the change in progress identified for the twentieth century is recent. We also know from the first variable discussed in the previous section of this paper that *nous* 'we' as a first-person pronoun was already a vestigial form in the nineteenth century, but at that time the form *on* was not yet discarded from its general association with indefinite reference. It seems that in the nineteenth century the form *on* had to fulfill two linguistic functions without much competition.

We now consider the role that the contact situation played in the change in progress taking place in the twentieth century. As discussed in the methodology section, the RFQ data representing nineteenth-century Québec French are considered to reflect a precontact stage of Québec French (for a detailed justification see Poplack & St-Amand 2007). If we consider that the twentieth-century data collected in Montreal represents a variety that displays more contact-induced influences, we could propose a hypothesis of change influenced by intersystemic factors. However, I still remain cautious since I cannot provide evidence of a transitional period where the second-person variant would have been emerging. The age factor may play a role here, since we know that in the twentieth-century data older speakers favored *on* in this linguistic function, and the nineteenth-century data comprised only older speakers.

Moreover, I question the interpretation one can develop from the "absence" of a variant in a specific body of data. As we know from the problematic issue of dating linguistic features from a body of written documents, the absence of a form could be attributed to different factors, namely, the role of the writer, the appropriateness of a feature with a specific genre, the normative conventions, and so on (Poplack & St-Amand 2007: 708). Far from saying that the RFQ does not represent vernacular features,[7] I am questioning how we can interpret the absence of a variant for the specific variable I am investigating here. As discussed in Laberge & Sankoff (1980), such a variable involves important pragmatic or discursive constraints. In selecting the data comprised in the RFQ, the researchers focused on discourse genres

---

7.   Poplack and St-Amand (2007) reported the presence of vernacular features in the RFQ and demonstrated that this corpus is well-suited for the exploration of morphosyntactic variables (for an extensive list of variables, see their p. 720–721).

resembling natural conversation: tales, legends, and interviews. While tales and legends represent 77 percent of the data, interviews represent only 23 percent of the data (Poplack & St-Amand 2007: 712). In addition, Poplack and St-Amand noticed that since the field methods and the elicitation techniques employed by the folklorists consisted of participant observation, "most recordings reveal remarkably little, if any, interaction between interviewer and informant" (ibid. 711). For all these reasons, I am particularly cautious about proposing that the absence of second-person pronouns with indefinite meaning constitutes enough evidence to conclude that what happened later in the twentieth century is due mainly to contact-induced influences. In addition, as we know from another real-time study (Jensen 2007), a parallel development of generic use of second-person pronouns occurred in the Danish language, and the spread of this linguistic innovation is better explained by pragmatic changes in the interactional patterns in general rather than by direct contact-induced influences from English.

Now let us look at the situation of the variation from the other angle offered by the examination of the linguistic behavior of 15 young Anglophone Montrealers. As illustrated in the examples in (5), all three variants are present in contemporary AMF.

(5)   a.   *C'était dur au début <u>tu sais</u> parce que quand <u>tu</u> rentres puis <u>tu</u> parles pas un mot.* (AMF, Ted)
'It was hard, <u>you know</u>, because when <u>you</u> arrive and <u>you</u> don't speak a word.'

   b.   *Il y a une section anglaise et <u>vous</u> êtes pas fréquemment dans la même classe.* (AMF, Greg)
'There is an English section and <u>you</u> are not frequently in the same class.'

   c.   *Quand <u>on</u> a treize ans, <u>on</u> est entre l'école primaire et le cours secondaire et oui chaque jour <u>on</u> a les mêmes classes dans le même ordre.* (AMF, Larry)
'When <u>we</u> are thirteen <u>we</u> are between elementary school and the high school and yes, every day <u>we</u> have the same classes in the same order.'

The results show that Anglo-Montrealers do not favor the generic pronoun *on*: it corresponds to less than one-third of the data. They instead clearly prefer to use the second-person pronouns *tu* or *vous*. Since the two forms both appear frequently in AMF, it was possible to undertake a quantitative analysis of the linguistic factors. For the sake of the comparison we tested for the same factors considered in the Laberge & Sankoff study (1980). A multivariate analysis with Goldvarb version 2 (Sankoff & Rand 1990) was performed on the 302 tokens extracted from 15 interviews; the results are displayed in Table 2.

Among the three linguistic factor groups tested, two were selected as significant. The more important factor was the lexical identity of the verb, with a range of 41. As was the case in the Sankoff & Laberge study of L1 French, some verbs clearly favor the use of *on*, such as *appeler* 'to call', for which we notice categorical behavior. The verb *dire* 'to say' generally favors the variant *on* with a weight of .84. Therefore *comme on dit* 'as one says' could easily become *comme tu dis* 'as you say' without implying that the interlocutor is involved in the interaction.

Table 2. Linguistic factors influencing the choice of indefinite *on* 'one' in the French of Anglo-Montrealers

Factor groups	Factors	%	N	Weight
Verb	*appeler* 'to call'	100	2	1.00
	dire 'to say'	82	39	0.84
	vouloir 'to want'	56	9	0.57
	other verbs	25	254	0.43
	range			*41*
Pragmatic	moral	58	122	0.72
	situational insertion	16	180	0.34
	range			*38*
Syntactic	generalization	49	122	NS
	presentative	34	32	NS
	implicative	20	148	NS
Total		30	302	

Input = .304, N = 302

The second significant factor group, with a range of 38, is related to pragmatic considerations. While the formulation of a moral favors the use of *on*, as exemplified in (6), a situational insertion[8] disfavors it, as illustrated in (5a). This reflects the pattern of L1 French. Finally, although the third factor group related to syntactic frames is not selected as significant, the direction of effect goes in the same direction as the one identified in the original study of L1 French. Overall, we can say that Anglophone Montrealers and Francophone native speakers display parallel linguistic conditioning.

---

8. A situational insertion refers here to a generalization having the effect of locating an indefinite person in a potentially repeatable activity or context (Laberge & Sankoff 1980: 280–81).

(6) *Je pense qu'on peut pas deviner le futur avec des rêves.* (AMF, Jack)
     'I think <u>we</u> can't know the future with dreams.'

However for AMF the situation of the English language also has to be taken into consideration. The usage of second person pronouns with indefinite reference is also possible in English, in alternation with other forms like the pronoun *one*. According to our observations of nearly 200 tokens in the English interviews of four speakers (with different degrees of contact with French and with different social backgrounds), they use second-person pronouns even in contexts where the speaker is formulating a moral. Moreover, the pronoun *one* seems to be associated with very formal style or formulaic expression. Further analysis of the pronominal variation in English could shed light on the effect of intersystemic influences.

The analysis of AMF showed a clear parallel with L1 French. The target for Anglophone Montrealers involves the adoption of tendencies identified in the local French variety. Even if second-person pronouns with indefinite reference are not part of the standard French language taught in school and indeed represent a change from below led by young men, it is noteworthy that Anglophone Montrealers are adopting this change in progress corresponding to the emerging norm of their age group. Previous results showed that the more contact Anglophone Montrealers had with native French speakers, the more their French resembled the local French. Here we have another example of this tendency, since they share the same pattern regarding linguistic constraints. This provides a good example of participation in community norms (Daveluy 2006).

However, we cannot discard an intersystemic explanation since their behavior could also be interpreted as a reflection of the patterns of the English language. We saw that the generic use of second person pronouns in English is very common while the use of "one" seems to be associated with formal style. According to Léglise (2007), if a tendency observed in language A meets up with the system of language B, then a snowball effect can occur and an inherent tendency could be reinforced by the contact situation. In other words, the fact that Anglophone Montrealers are clearly favoring the use of generic second-person pronouns could push the change forward, at least in their own repertoire.

5.3 The variation between simple and compound forms of nonclitic plural pronouns

The other variable under study is defined by the alternation between the simple and compound forms of nonclitic plural pronouns. The simple forms (*nous, vous, eux,* and *elles*) correspond to what is taught in school and what is associated with standard French, while the compound forms with *–autres* 'others' (*nous autres, vous autres, eux autres*) are characteristic of ordinary Québec French. The following

example (7) illustrates a case of inherent variation in Montreal French not only in the speech of the same speaker but in the context of the same utterance.

(7) *Nous on filtre les clients, les institutions financiers filtrent les clients puis <u>eux autres</u> les filtrent*. (Montréal-95, #7 or Lysiane B.).

    '<u>We</u> filter the clients, the financial institutions filter the clients and then <u>they</u> filter them.'

As previously demonstrated (Blondeau 1999, 2001), the compound forms with *-autres* are characteristic of the local variety and very frequent in contemporary Montreal spoken French. Table 3 indicates two series of oppositions between subject clitics and nonclitic pronouns for the variants under study. The compound forms with *-autres* can be considered a paradigmatic regularization since they extend the phonological contrast between the clitics and the nonclitics to all plural pronouns, as is the case for the singular forms (for example the opposition between *je* 'I' and *moi* 'me'). In addition, *-autres* could be viewed as a morpheme of plurality showing a regularization for all the plural persons (Blondeau 1999).

Table 3. Contrast between subject clitics and nonclitic plural pronouns

Person	Subject clitic	Nonclitic	
		Simple form	Compound form
1st	*on/nous*	*nous*	*nous autres*
2nd	*vous*	*vous*	*vous autres*
3rd masculine	*ils*	*eux**	*eux autres*
3rd feminine	*elles*	*elles*	*eux autres*

* Within the simple form paradigm *eux* is the only nonclitic pronouns displaying a phonological contrast with the subject clitic form.

Although some researchers considered compound forms with *-autres* as quasi-categorical in Québec French (Auger 1994; Morin 1982), a quantitative analysis of the twentieth-century data shows that compound forms are involved in a change in progress in the community and are less frequent in 1995 in comparison to 1971 (Blondeau 1999, 2001). We can trace the trajectories of the variable over time. Figure 1 combines the data from the RFQ and the Montreal data, for 65 speakers aged 45 and older at time of interview and divided in three groups according to their date of birth.

    We can see that the compound forms with *-autres* were already a well-established feature of local French in the nineteenth century. The data also shows that a recent vigorous change in progress is taking place in which the simple forms progress in the community. There is a clear demarcation according to the birthdate of the speakers. Speakers born later in the twentieth century use the simple forms
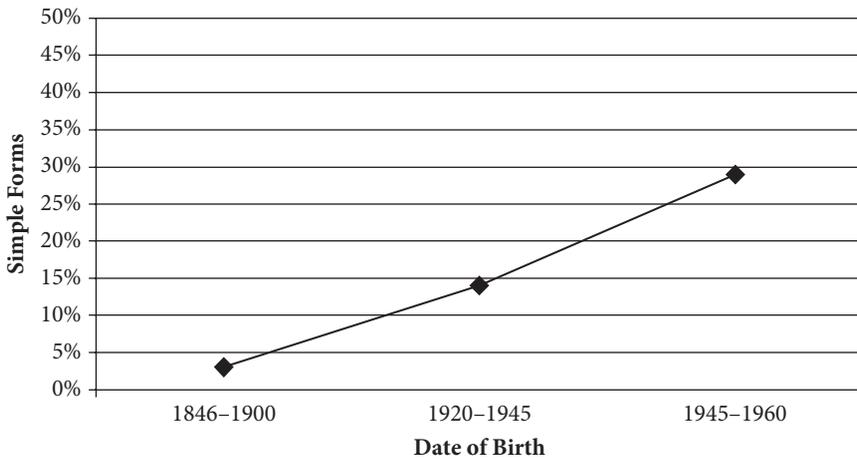
Figure 1. Progression of simple forms for three generations of speakers aged 45 and older at the time of the recording.

more frequently than the other speakers, who seem to keep the compound forms in their repertoire.

This change in progress has all the characteristics of an internally motivated change involving a combination of social, stylistic and linguistic factors. The presence of a preposition played a role in the variable choice for both centuries (Blondeau 2003), but the factor related to referential contrast was not significant, suggesting that, already in the nineteenth century, compound forms were not characterized by an emphatic function. For the twentieth century data, the change in progress involved social and stylistic factors, and the movement toward an increase of simple forms was triggered by the linguistic factor related to person (Blondeau 1999, 2001). The proposed interpretation involved a hypothesis of so-cio-stylistic specialization of each variant in the context of a grammaticalisation process in which the compound forms with *-autres* no longer carried an emphatic function. The study of *ne* deletion in French offers some parallels. According to an analysis of the situation in Montreal French, the negative particle *ne* is maintained in Montreal French as a stylistic resource (Sankoff & Vincent 1980) and is not on the verge of a complete disappearance although the change is far more advanced. In addition, according to Daveluy (2006), Anglo-Montrealers contribute to the maintenance of the complete form of the negation.

If we now turn to AMF, it is worth noting that the compound variants with –*autres* are less common in this segment of the population, with an overall rate of 41 percent. Most of the young Anglophone Montrealers prefer the simple variant, conforming to the variant associated with the standard language taught in school.

It is worth noting, however, that they also parallel the new trend of native speakers of French who favor the simple forms. In addition, the speakers who use the more compound forms are those who have more contact with French speakers, according to their linguistic environment scale (Blondeau, Nagy, Sankoff & Thibault 2002). They therefore seem to adjust their behavior in relation to the speakers they are in contact with.

6. Conclusion

The context of the language dynamics in Québec always offers an excellent situation for exploring the relation between inherent tendencies and contact-induced phenomenon. As we know from the literature, it is a challenge to distinguish between the two processes, and in fact, in many situations of contact both tendencies work at the same time. One of the keys for elucidating this relation relies on the power of comparability. Light can be shed on the dynamics of language change through access to comparable bodies of data. Moreover, one of the aims of this synthesis was to go further than the detailed examination of one isolated phenomenon that could lead to explanation in terms of simplification or regularization, and instead to look at pronominal variation as a whole to provide insight on how the structure of a linguistic subsystem is in constant movement and implies balancing effects.

In this synthesis I aimed to reexamine the variation for three variables of the pronominal paradigm from two different angles, one involving a diachronic comparison between nineteenth-century and twentieth-century Québec French, the other involving a synchronic comparison between the variety of L2 French incorporated in the repertoire of Anglophone Montrealers and L1 French spoken by native speakers in Montreal.

From the perspective of the data involving native French speakers, the three analyses shed light on different trajectories of change. In the first case, contrary to what was expected from the apparent time perspective, we saw that the replacement of *nous* by *on* was already implemented at an earlier stage of the language. In the nineteenth century, *on* was the most common variant and *nous* was already rare. These results not only expand the historical scope of our understanding of the path of the change but, if we agree that the RFQ represents a precontact stage of Québec French, the diagnosis of a contact-induced change has to be excluded for this variable. In the case of the variation among pronouns with indefinite reference, the results confirmed that the change in favor of second-person pronouns is a recent process peaking at the end of the twentieth century. No evidence of a productive use of indefinite second-person pronouns was found in the nineteenth-

century data. Due to the pragmatic nature of this specific variable, however, we are cautious in our interpretation of the absence of a variant. Therefore the hypothesis of a contact-induced change remains open and to be confirmed by further studies looking at a transitional period. In addition access to data involving more varied contact situations might also be taken into account. For the variable involving the alternation between simple and compound forms of the nonclitic pronouns, the comparison between the nineteenth- and twentieth-century data highlighted a change in progress regarding the rate of usage, and a reorganization of the linguistic conditioning of the forms.

Looking at the Anglophone Montrealers' behavior in comparison with native speakers of French in Montreal, we can see that for all three variables both groups of speakers behave in the same manner. Research has shown that L2 learners in Toronto behave in a different manner than native French speakers regarding their use of informal or vernacular variants (Mougeon et al. 2004). These results, based on school experiences in Ontario, another province of Canada with a large majority of English speakers, indicate a marginal use of those variants by L2 learners. Our analysis of the use of French by Anglo-Montrealers, who are in daily contact with native French speakers, has shown a different pattern. Anglo-Montrealers have clearly adopted the pattern of *on* usage to express the first-person plural with a definite reference. In fact, their rate of usage is similar to that found for native speakers of French. Regarding the use of the generic pronouns, they have not only gone in the same direction of adopting the second-person pronouns as the most common form for indefinite reference, but they also parallel the linguistic conditioning of the form. Finally, their usage of the compound and simple forms also resembles that of younger speakers of L1 French, who seem to adopt the simple forms. Anglophone Montrealers who have integrated French into their repertoire and who are using French on a daily basis clearly participate in the dynamics of the speech community.

References

Ashby, W. 1992. The variable use of *on* versus *tu/vous* for indefinite reference in spoken French. *Journal of French Language Studies* 2: 135–157.

Auger, J. 1994. Pronominal Clitics in Quebec Colloquial French: A morphological analysis. Ph.D. dissertation, University of Pennsylvania.

Bailey, G., Wikle, T., Tillery, J. & Sand, L. 1991. The apparent time construct. *Language, Variation and Change* 3(3): 241–265.

Bayley, R. & Regan, V. 2004. Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics* 8(3): 323–338.

Blondeau, H. 1999. Parcours d'un marqueur sociolinguistique: les pronoms non-clitiques du pluriel en français parlé de Montréal. Ph.D. dissertation, Université de Montréal.

Blondeau, H. 2001. Real-time changes in the paradigm of personal pronouns in Montreal French. *Journal of Sociolinguistics* 5(4): 453–474.

Blondeau, H. 2003. The old *nous* and the new *nous*. A comparison of 19th and 20th century spoken Quebec French. *University of Pennsylvania Working Papers in Linguistics* 9(2): 1–15.

Blondeau, H., Sankoff, G. & Charity, A. 2002. Parcours individuels et changements linguistiques en cours dans la communauté francophone montréalaise. *Revue québécoise de linguistique* 31(1): 13–38.

Blondeau, H., Nagy, N., Sankoff, G. & Thibault, P. 2002. La couleur locale du français des Anglo-Montréalais. *Acquisition et Interaction en Langue Étrangère* 17: 73–100.

Chaudenson, R. 1998. Variation, koïnéisation, créolisation: français d'Amérique et créoles. In *Français d'Amérique: Variation, créolisation, normalisation, Actes du colloque Les français d'Amérique du Nord en situation minoritaire, Université d'Avignon.* P. Brasseur (Ed.), 163–79. Avignon: Centre d'études canadiennes and Université d'Avignon.

Chaudenson, R., Mougeon, R. & Beniak, E. 1993. *Vers une approche panlectale de la variation du français.* Paris: Didier Érudition.

Coveney, A. 2000. Vestiges of *nous* and the first person plural verb in informal spoken French. *Language Sciences* 22: 447–481.

Coveney, A. 2003. 'Anything *you* can do, *tu* can do better': *tu* and *vous* as substitutes for indefinite *on* in French. *Journal of Sociolinguistics* 7: 164–191.

Chantefort, P. 1976. Diglossie au Québec: limites et tendances actuelles. *Cahiers de linguistique de l'Université du Québec* 6: 22–53.

Daveluy, M. 2006. *Les langues étendards: Allégeances langagières en français parlé.* Québec: Éditions Nota Bene.

Deshaies, D. & Ouellet, J. 1982. Contribution à l'analyse des pronoms personnels. *Langues et linguistiques* 8(2): 219–243.

Fonseca-Greber, B. & Waugh, L.R. 2003. On the radical difference between the subject personal pronouns in written and spoken European French. In *Corpus Analysis: Language structure and language use,* P. Leistyna & P.C. Meyer, (Eds), 225–240. Amsterdam: Rodopi.

Gadet, F. & Jones, M.C. 2008. Variation, contact and convergence in French spoken outside France. *Journal of Language Contact* 2.

Goffman, E. 1981. *Forms of Talk.* Philadelphia PA: University of Pennsylvania Press.

Jensen, T.J. 2007. Generic Use of the Second Person Pronoun in Danish – the spreading of a linguistic innovation. Paper presented at ICHL 18. Université du Québec à Montréal.

King, R. 1983. Variation and Change in Newfoundland French: A sociolinguistic study of clitic pronouns. Ph.D. dissertation. Memorial University of Newfoundland.

King, R. 2000. *The Lexical Basis of Grammatical Borrowing: A Prince Edward Island case study.* Amsterdam: John Benjamins.

Labov, W. 2007. Transmission and diffusion. *Language* 83(2): 344–387.

Laberge, S. 1977. Étude de la variation des pronoms sujets définis et indéfinis dans le français parlé à Montréal. Ph.D. dissertation, Université de Montréal.

Laberge, S. & Sankoff, G. 1980. Anything *you* can do. In *The social Life of Language*, G. Sankoff (Ed.), 271–293. Philadelphia PA: University of Pennsylvania Press.

Léglise, I. 2007. Variation in contact settings: The interplay of inherent tendencies and (contact-induced) transfer. Paper presented at the International Workshop Variations et changements morpho-syntaxiques en situation de contact de langues. Paris.

Martineau, F. 2005. Perspectives sur le changement linguistique: Aux sources du français canadien. *Revue canadienne de linguistique* 50(1–4): 1001–1040.

Morin, Y.-C. 1982. De quelques [l] non étymologiques dans le français du Québec: Notes sur les clitiques et la liaison. *Revue québécoise de linguistique* 2(2): 9–47.

Mougeon, R. & Beniak, E. 1991. *Linguistic Consequences of Language Contact and Restriction: The case of French in Ontario.* Oxford: OUP.

Mougeon, R. & Beniak, E. (Eds) 1994. *Les Origines du français québécois.* Québec: Presses de l'Université Laval.

Mougeon, R., Rehner, K. & Nadasdi, T. 2004. The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics* 8(3): 408–432.

Nadasdi, T. 1995. Variation morphosyntaxique et langue minoritaire: Le cas du français ontarien. Ph.D. dissertation, University of Toronto.

Poplack, S. & St-Amand, A. 2007. A real-time window on 19th-century vernacular French: The Récits du français québécois d'autrefois. *Language in Society* 36(5): 707–734.

Poplack, S., Walker, J. & Malcolmson, R. 2006. An English like no other? Language contact and change in Quebec. *Canadian Journal of Linguistics* 51(2): 185–213.

Rehner, K., Mougeon R. & Nadasdi, T. 2003. The learning of sociolinguistic variation by advanced FSL learners: The case of *nous* versus *on* in immersion French. *Studies in Second Language Acquisition* 25: 127–156.

Sankoff, D. & Rand, D. 1990. *Goldvarb Version 2. A variable rule application for the Macintosh.* Montréal: Centre de recherches mathématiques, Université de Montréal.

Sankoff, D., Sankoff, G., Laberge, S. & Topham, M. 1976. Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale. *Cahiers de linguistique de l'Université du Québec* 6: 85–125.

Sankoff, G. (Ed.) 1980. *The Social Life of Language.* Philadelphia PA: University of Pennsylvania Press.

Sankoff, G. 2001. Linguistic outcomes of language contact. In *The Handbook of Language Variation and Change*, J. Chambers, P. Trudgill & N. Schilling-Estes (Eds), 638–68. Oxford: Blackwell.

Sankoff, G. 2005. Cross-sectional and longitudinal studies in sociolinguistics. In *An International Handbook of the Science of Language and Society*, Vol. 2, 2, U. Ammon, N. Dittmar, K.J. Mattheier & P. Trudgill (Eds), 1003–1013. Berlin: de Gruyter.

Sankoff, G. & Blondeau, H. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83(3): 560–588.

Sankoff, G. & Vincent, D. 1980. The productive use of *ne* in spoken Montreal French. In *The Social Life of Language,* G. Sankoff (Ed.), 295–310. Philadelphia PA: University of Pennsylvania Press.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M.-O. & Gagnon, L. 1997. Variation and the use of discourse markers in a language contact situation. *Language Variation and Change* 9: 191–218.

Thibault, P. 1983. Équivalence et grammaticalisation. Ph.D. dissertation, Université de Montréal.

Thibault, P. 1991. La langue en mouvement: Simplification, régularisation, restructuration. *LINX* 25: 79–92.

Thibault, P. & Vincent, D. 1990. *Un corpus de français parlé. Montréal 84: historique, méthodes et perspectives de recherche.* Québec: Département de langues et linguistique, Université Laval.

Thibault, P. & Sankoff, G. 1993. Diverses facettes de l'insécurité linguistique: Vers une analyse comparative des attitudes et du français parlé par des Franco et des Anglo-Montréalais. *Cahiers de l'Institut de linguistique de Louvain* 19: 209–218.

Vincent, D., Rains, C. & Daveluy, M. 1985. When *I* and *we* are in competition. NWAV 14. Georgetown University, Washington DC.

Vincent, D., Laforest, M. & Martel, G. 1995. Le corpus de Montréal 1995: Adaptation de la méthode d'enquête sociolinguistique pour l'analyse conversationnelle *Dialangue* 6: 29–46.

Subordinate clause marking in Montreal Anglophone French and English

Hélène Blondeau & Naomi Nagy
University of Florida, University of Toronto[1]

We investigate the variable presence of overt complementizers in the bilingual repertoire of young Anglophone Montrealers, examining approximately 1,600 sentences in spoken French and English. The effects of linguistic constraints are compared between their two languages, and also to recent research on Quebec City Anglophones' and to L1 Quebec Francophones' speech. We also examine instances of usage of the verbs of quotation *be like* and *être comme*. Patterns of several linguistic constraints affecting the variation help us understand the intersection of subordinate clause marking and the grammaticization of *be like/être comme* as verbs of quotation as well as better understand advanced stages of second language acquisition.

Keywords: Second language acquisition (SLA); vernacular; subordinate clause; complementizer (COMP); verb of quotation (VOQ); *be like*; syntactic variation; Montreal; French; English; frequency effects; collocations

1. Introduction

In 1993, Gillian Sankoff and Pierrette Thibault began an innovative investigation of variation in the speech of young Anglophone Montrealers (AM), with an aim toward understanding how the degree and types of exposure to French influenced the way these speakers use French. This ambitious project benefitted from the years

of experience with sociolinguistic investigation in multilingual communities, and in Montreal in particular, that Gillian brought to the project. We are very pleased to have had the opportunity to work with her on this project and to now have this opportunity to thank her for both her earlier groundbreaking research and her excellent skills as a mentor. We are grateful to her for conceiving of this important project and for allowing us access to the data, but even more for having set the stage in so many ways for the study of the interaction of linguistic and social variation and for having trained so many of us who work in this area, including Pierrette Thibault, the co-PI for the original Montreal Anglophone project, and both authors of this paper. Since this study began, many aspects of both the French and English of these speakers have been investigated. An overview may be found in Blondeau et al. (2002). In this chapter, we aim to follow Gillian's example of examining linguistic variation in its social context, as we investigate one more pattern of variation in AM speech.

French and English both permit the variable presence of overt complementizers as the head of subordinate clauses. English allows the presence or absence of the complementizer *that* as illustrated in (1) and French allows the presence or absence of the complementizer *que* as illustrated in (2). Previous studies show that this variable is constrained by a set of linguistic and extra-linguistic constraints in each language (cf. Biber 1999 for English; and for a recent account of a variety of Quebec English see Torres Cacoullos & Walker *forthcoming*; for a variety of Quebec French, see Dion 2003). Since French is now an integral part of the linguistic repertoire of Anglophone Montrealers, we compare the patterns of usage of our bilingual speakers to both of these other studies, looking at similarities and differences in the effects of linguistic constraints. Specifically, we analyze and compare the linguistic factors that contribute to the variable presence of overt complementizers (COMP) in the English L1 and the French L2 included in the linguistic repertoire of young Anglophone Montrealers, who produce tokens such as those shown in (3) and (4).

(1)   L1 English examples (from the *Quebec English Corpus*, Poplack, Walker & Malcolmson 2006, cited in Torres Cacoullos & Walker forthcoming: 1)

   a.   And I let it slip <u>that</u> Darth Vader was Luke's father.
   b.   I can't even believe Ø I just said that.

(2)   L1 French examples (from the Ottawa-Hull Corpus (Vieux-Hull neighborhood), Poplack 1989, cited in Dion 2003: 12).

   a.   *Moi je leur disais Ø c'est pareil comme nous autres si on irait dans leur pays.*
        I told them <u>Ø</u> it's the same, like we–if we went in their country.

      b.   *Bien parce qu'elle disait <u>que</u> les filles étaient pas assez distinguées.*
           Well, because she said <u>that</u> the girls weren't distinguished enough.

(3)   Anglo-Montrealer example in French
      Je pense Ø c'est plus anglophone je pense <u>que</u> les compagnies sont: donnent
      un sens de anglophone. (Greg)

      I think <u>Ø</u> it's more Anglophone. I think <u>that</u> the companies give a sense
      of being Anglophone.

(4)   Anglo-Montrealer example in English
      I think <u>Ø</u> he thought <u>Ø</u> it was really cool <u>that</u> I spoke French and <u>that</u>
      I was bilingual. (Liz)

In current spoken English it is common to have subordinate clauses introduced without an overt COMP, as in (1 and 4). The absence of COMP, although possible, is less common in French (as in 2 and 3). These AM speakers respect this difference: they exhibit only 27% COMP presence in English but 77% in French. Their rate of usage in English is higher than the ~15% reported by Biber (1999: 145) for the conversational register, but given methodological differences, the difference is not noteworthy. In our analysis of the varation, the factors examined include the form and reference of the matrix and subordinate clause subjects, identity of the subordinating verb, frequency of collocation of these two as well as lexical frequency of the matrix verb, phonological context, and several measures of semantic and syntactic complexity. Throughout, we present the application value as COMP presence. That is, we compare how often there is a surface *that* or *que*. This makes our study compatible with other studies of the variable in English, although French studies tend to report the rate of COMP absence.

    After a literature review summarizing the constraints on variation in the two language varieties, we explicitly circumscribe the variable context for our study and define the linguistic variables to be examined. The next section of our paper provides an analysis of the similarities and differences both within the speech of these Anglophone Montrealers and with other recent studies of the same variable. Then we briefly explore how *like* could be reanalysed as a COMP since it appears in the same surface position as the (deletable) COMP, and make predictions for the evolution of *comme* in French. This provides an analysis for sentences like:

(5)   Because I felt like, Uh: OK, I understand that the French people have to
      protect themselves. (Louisa)

We then position our analysis within the social context of the variation.

2. Previous studies of the variable presence of COMP

2.1 French

The presence or absence of the COMP in subordinate clauses has been examined from a number of variationist perspectives. In her seminal article, "A quantitative paradigm for the study of communicative competence," Sankoff refers to the first analysis of the deletion of COMP *que* in Montreal French (Sankoff et al. 1971, cited in Sankoff 1980: 66). She uses this example to emphasize the structure of variable phonological deletion constraints. In another article devoted to the role of phonology in variable rules, Sankoff and Brown (1976) refer to another construction containing the form *que* in co-occurrence with wh-forms.[2] *Que*-presence is variable not only for Quebec French but also for other varieties of French, as commented upon in Gadet (2007), making *que* an interesting candidate to examine from a variationist perspective. In this article we restrict ourselves to the variable presence of the COMP *que*.

Most researchers agree that the variable presence of COMP *que* in French is constrained by social factors and that the absence of *que* is stigmatized. However for many years there has been debate over the role of linguistic factors in the explanation of the variability. Three types of conditioning are identified: phonological, syntactic and lexical, but the relative importance of each is disputed.

Sankoff et al. (1971) and Sankoff (1980) first looked at this variable, identifying the phonological environment as the major constraint on the variation. Their analysis of the distribution of the variable among sixteen Montrealers led them to identify the preceding and following phonological environment as relevant factors. The analysis identified sibilants as a favorable context for *que* deletion, and the interpretation provided was one of consonant cluster simplification (Sankoff 1980: 66). Working with the same data, Connors (1975) proposed an alternative explanation involving syntactic factors. According to her, the deletion was conditioned by the nature of the subject in the subordinate clause. If a pronoun appeared in the subordinate clause rather than a noun, a null COMP was favored. According to Connors, the phonological environment was an epiphenomenon of the syntactic effect, since several frequent subject pronouns in French start with a sibilant (i.e., *je* 'I' and *ça* and *ce* 'that'). The debate remained open until new studies were undertaken.

Ten years later, a study of the same variable in the French of Ottawa-Hull was undertaken by Martineau (1985), followed by a real time study of Montreal

---

2. This variable context was also investigated for Acadian French by Beaulieu and Cichocki (2002) and by King and Nadasdi (2006).

French (Warren 1994). Both found that the less sonorant the following segment was the more the deletion of *que* was probable, at least partially supporting Sankoff's interpretation rather than Connors'. However they also both discarded the effect of the phonological environment preceding the COMP and indicated a possible lexical effect. Warren also pointed out the importance of a lexical effect and discussed extensively the specificity of COMP deletion following the collocation *je pense*.

In a recent real-time study on the same variable in Canadian French, Dion (2003) tried to reconcile differences in reported signficant factors by carefully examining the interaction between factor groups. She was able to convincingly disentangle the effect of each factor and identified only two factors contributing independent effects. In the preceding context, the lexical identity of the matrix verb plays an important role in determining the likelihood of deletion. In the following context, Dion discarded the syntactic explanation in favor of a phonological one, since all pronouns did not behave in the same way. Pronouns starting with a vowel clearly disfavor deletion, while pronouns starting with an initial consonant favor it. The NP's were at an intermediate point but this was also related to their phono-logical status, most of them starting with nasals and liquids. Her demonstration therefore convincingly supports Sankoff's, Martineau's and Warren's accounts, and her findings were further supported by additional data presented in King & Nadasdi's (2006) report on Acadian French *que*.

2.2   English

The variable presence of complementizers in English has been well documented in a number of publications. The most recent that we are aware of is Torres Cacoullos & Walker (*forthcoming*) and we refer readers to Section 2 of that paper for a thorough review of earlier work as well as a succinct history of the development of the COMP *that*. Torres Cacoullos & Walker present their study of this variable in the Quebec English Corpus (Poplack, Walker & Malcolmson 2006) in a multivariate framework similar to ours (cf. Nagy & Blondeau 2005) but are innovative in their investigation of the possibility that certain collocations that frequently occur between the subject and verb of the matrix clause may in fact have grammaticized into discourse markers, at least in some of their occurences. They make the distinction that a sentence can truly be conceived of as containing a matrix and subordinate clause only when there is semantic content on both sides of the COMP. When there is no content in the putative matrix clause, the collocation may be better analyzed as a discourse marker, with typical examples being *you know* and *I think*. Their quantitative analyses of ~3,000 tokens, considering common and rarer collocations separately, convincingly support their arguments. This is relevant

to our analysis as they found that the COMP is much less likely to surface when the "matrix clause" consisted of these common collocations, although the same linguistic factors constrain the variation in COMP presence in both common and rarer structures.

3.   Methods

In this section, we briefly summarize the methods used to collect our data. They are presented more completely in Sankoff et al. (1997) and Blondeau et al. (2002). We then turn to the specific manner of coding and analysis to show how COMP behaves in our corpus and how this compares to its behavior in the studies mentioned in the preceding section.

3.1   The speakers[3]

The speakers are 18 Anglophones born in metropolitan Montreal. Their ages were 20–34 when they were first interviewed (1993–1995). Many of the speakers are a self-selected group of volunteers. The remainder are a sample of the 1990 graduating class of an English language high school on the border between an Anglophone and a bilingual neighborhood. The school offers two levels of French immersion as well as French taught as a subject (for those less advanced in French). It had recently received an influx of Francophone speakers. All speakers lived in the Greater Montreal area at the time of the interview and spoke English with their parents. They differ, however, in their mode of acquisition of French, the type of exposure they had to French as children, their current degree of contact with Francophones socially and in the workplace, and in the degree to which they use French in their daily lives.

3.2   Corpus development

The self-selected participants were recruited by newspaper ads seeking bilingual speakers placed in two free Montreal newspapers. The high school graduates were located by cross-referencing their high school yearbook and a telephone directory. Potential participants were telephoned and asked (in English) if they would participate in an interview about bilingualism in Montreal.

---

3.   These paragraphs on the speakers and corpus development are adapted, more or less wholesale, from Nagy et al. (2003).

Each participant was interviewed first in French by a Francophone and a few weeks later in English by a non-local Anglophone. Topics for both interviews included school and family background, use of French in various contexts, attitudes toward French politics, people, and culture, and incidents where language differences played a significant role. Each interview lasted about one hour. All interviews were tape-recorded and fully transcribed. All our data is, therefore, from the conversational register, as defined in Biber (1999).

3.3   Methods of analysis

We first examined the frequency of subordinate clauses in the corpus. We then conducted a variety of multivariate logistic regression analyses using Goldvarb X (Sankoff et al. 2005) to determine the relative effects of the linguistic factors on the rate of COMP presence in various subsets of our data. Once a set of significant linguistic factors was determined, social factors were added in. Separate analyses were conducted for the French and English data. Goldvarb X produces factor weights indicating the strength of each factor, or, in essence, how likely a particular sentence is to contain a surface COMP if it has a particular linguistic or social attribute, independent of its other attributes. Factor weights were calculated using the "one-level" analysis, and the relative significance of each group was determined using the "step-up, step-down" analysis and consideration of the range of weights of the factors.

3.4   Linguistic data coding

All sentences illustrating the variable presence of a COMP as the head of a subordinate clause were extracted from the corpus, providing about 500 tokens in French and about 1,000 in English. We also extracted all examples of English *be like* (N ≈ 90) and French *être comme* (N ≈ 20) used to introduce reported speech, for reasons explained in Section 6. Sentences such as those shown in (6) were excluded, as variation in COMP presence is not possible.

(6)   a.   I never know like [*that/Ø which one is which] (Liz)

b.   It's [that/*Ø the person speaks both and I speak both] (Louisa)

c.   I love [that/*Ø I can]. Um: and [that I can read, just at all, y'know, just that I'm literate in two languages.] (Louisa)

d.   *Ça dépend [qu'/*Ø est ce qui joue]* (Joan)
That depends who's playing.

Each subordinate clause was coded for the presence or absence of a COMP and for the independent internal and external factors described below.

3.5 Dependent variable

Each token sentence was first coded for the dependent variable: presence or absence of a COMP at the head of the subordinate clause. In English, *that* and *like* were considered as surface COMP forms, but the *like* tokens are relegated to a separate analysis, discussed in Section 6. In French, the surface COMP forms are *que* and *qu'*, and *comme* (again, see Section 6).

3.6 Independent linguistic variables

The list of independent variables is provided in the appendices, and each possible variant is exemplified by a sample sentence from our corpus. The distribution of forms for each variable is shown in Tables 1 and 2, discussed below. The variables are grouped into four categories: those related to lexical identity and frequency, syntax, phonology, and semantics. Several earlier studies have tried to tease apart the effects of lexical identity of the matrix clause verb and its frequency. To contribute further data to this complex question with important ramifications, we coded for both. The syntactic factors relate to the structural complexity of the sentence and are included to bolster previous reports that COMP is more likely to surface in more complex sentence structures (Torres Cacoullos & Walker *forthcoming*: 11, Biber 1999). Semantic factors are similarly important, in that previous reports suggest that COMP will surface more often when there is less semantic cohesion between the matrix and subordinate clause, that is, less co-referentiality of the two parts (Torres Cacoullos & Walker *forthcoming*). Phonological factors are included because of the possibility that COMP presence/absence is (partially) determined by marked vs. unmarked syllable structure, as has been shown for French.

3.7 Independent social variables

We have selected speakers who differ in terms of their means of acquisition of French, degree of contact with French, and ability with prescriptive French rules. For ease of comparison, five types of information are shown for each speaker in Table 3. The "Contact" column indicates the degree to which they have been exposed to French. "High" indicates the most contact, whether it be through school, other activities, or friends and family; and "low" indicates least. "Education" focuses on the type of schooling they received. 1 point was granted for attending an English school with French as a subject. 2 points were granted for partial attendance in a French medium or immersion school. 3 points were granted for full attendance in a French medium school or (post-) immersion program. Students who changed program type receive intermediary scores. Scores of 5 and above are shown as "high;" scores below 3 are "low;" others are "medium." The "Grammar score" indicates the percentage of nouns (out of 20) that they correctly and unambiguously marked for gender in a stretch

Table 1. Distribution of tokens for AME subordinate clauses

| Factor group | Factor | N | | | % |
		that	Ø	Total	*that*
Verb identity and frequency factors					
Lexical frequency	> 1000 (in BNC)	125	539	664	18.8
	> 100 but < 1000	87	134	221	39.4
	< 100	40	42	82	48.8
Verb identity	*realize*	15	6	21	71.4
	tell	15	8	23	65.2
	other *find*	5	3	8	62.5
	know	32	69	101	31.7
	I find	4	9	13	30.8
	say	16	39	55	29.1
	other *think*	13	75	88	14.8
	I think	38	266	304	12.5
	I'm sure	2	14	16	12.5
	I guess	4	34	38	10.5
	I wish	1	9	10	10
	I remember	3	30	33	9.1
	I don't think	5	69	74	6.8
	Other	99	84	183	54.1
Syntactic factors					
Subject type: matrix clause	1st SG.	148	618	766	19.3
	2nd SG.	8	18	26	30.8
	3rd SG. pronoun, human	15	19	34	44.1
	Pleonastic/neuter	34	15	49	69.4
	Singular noun	5	9	14	35.7
	1st PL.	3	3	6	50
	3rd PL. pronoun, human	22	20	42	52.4
	Plural noun	17	13	30	56.7
Subject type: subordinate clause	1st SG.	153	620	773	19.8
	2nd SG.	9	19	28	32.1
	3rd SG. pronoun, human	15	19	34	44.1
	3rd SG. pronoun neuter	24	11	35	68.6
	Singular noun	5	9	14	35.7
	2nd plural	3	4	7	42.9
	3rd PL. pronoun, human	24	20	44	54.5
	Plural noun	19	13	32	59.4
Intervening material	none	202	645	847	23.8
	argument (IO, PP)	12	5	17	70.6
	other	38	65	103	36.9
Other *that* in COMP-position	no	245	671	916	26.7
	yes	7	44	51	13.7
Transitivity of subord. clause	intransitive	31	84	115	27
	transitive	221	631	852	25.9
Semantic factors					
Animacy: matrix subject	human	219	701	920	23.8
	pleonastic	14	5	19	73.7
	verbal	8	3	11	72.7
	abstract	11	6	17	64.7

(*Continued*)

Table 1. Continued

Factor group	Factor	N that	N Ø	N Total	% that
Animacy: subord. subject	human	167	418	585	28.5
	abstract	79	291	370	21.4
	inanimate, concrete	6	6	12	50
Polarity	matrix neg., subord pos.	48	116	164	29.3
	both positive	202	597	799	25.3
	both negative	2	2	4	50
Co-referentiality	coreferential	46	140	186	24.7
	not coreferential	206	575	781	26.4
Semantic class of matrix verb	comment	21	10	31	67.7
	attitude	100	524	624	16
	extraposition	13	9	22	59.1
	knowledge	70	118	188	37.2
	utterance	38	50	88	43.2
	suasive	10	4	14	71.4
Finiteness: matrix-clause	finite	239	709	948	25.2
	non-finite	13	6	19	68.4
Phonological factor					
Segment following COMP	vowel	127	404	531	23.9
	liquid	3	6	9	33.3
	nasal	7	11	18	38.9
	fricative	105	277	382	27.5
	stop or affricate	10	17	27	37.0
Total		**252**	**715**	**967**	**26.1**

of their French interview, included to show their general level of proficiency. Apart from the Grammar score, information is based on self-reports.

4. Data

In this section, we look first at the distribution of subordinate clauses across the two languages and across speakers. We then define the envelope of variation with respect to the different linguistic variables considered. Finally, we turn to an examination of the frequency of several elements within the token sentences.

4.1 Overall frequency of subordinate clauses

Our first observation is that the overall frequency of subordinate clauses does not differ greatly between the English- and French-language interviews in our corpus. Thus, the cross-linguistic differences discussed below may not be discounted by claiming that the speakers' language is less syntactically complex (in the relevant

manner) in their L2 than their L1. There are ~4.5 subordinate clauses per 1000 words in English, and ~3.5 in French. In the English interviews, this value ranges from 1.2 (Tony) to 9.7 (Donald), and the range does not appear to correspond to any particular social features. The rate of COMP presence in the two languages is not correlated for individual speakers, and is always greater for AMF than AME, as Figure 1 illustrates.
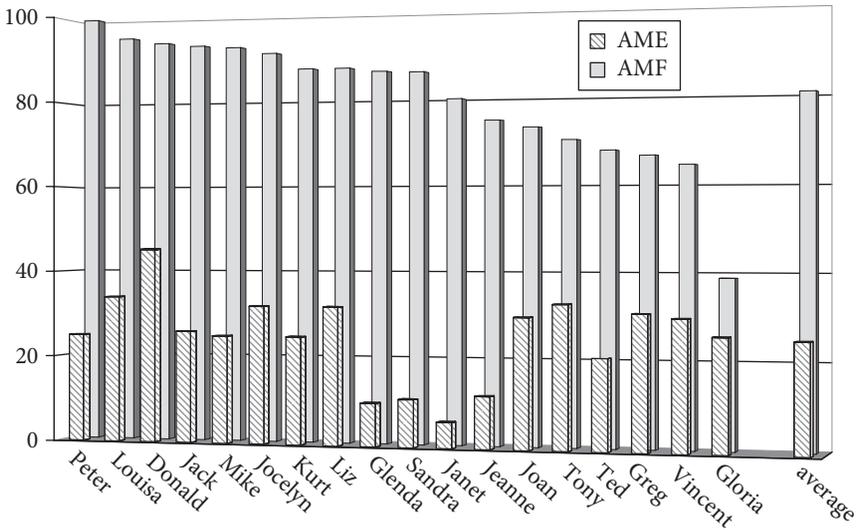
Figure 1. Rate of COMP presence (%) in AME and AMF for each speaker.

The English data is comprised of 1,013 sentences containing a subordinate clause. The distribution of tokens according to all the factors considered is shown in Tables 1 and 2.

Table 2. Distribution of tokens for AMF subordinate clauses

		N			%
Factor group	Factor	*que*	Ø	Total	*que*
Verb identity and frequency factors					
Lexical frequency	>1000 (in Corpaix)	261	82	343	76
	>100 but <1000	97	26	123	79
	<100	29	11	40	73
Verb identity	*sembler*	5	5	10	50
	souvenir	2	2	4	50
	dire	45	19	64	70.3
	penser	87	35	122	71.3
	croire	9	3	12	75
	trouver	55	18	73	75.3

(*Continued*)

Table 2. Continued

Factor group	Factor	N que	Ø	Total	% que
	savoir	30	9	39	76.9
	falloir	45	8	53	84.9
	voir	13	3	16	81.2
	être + X	25	5	30	83.3
	other verbs (1–4)	29	5	34	85
	other verbs (5–15)	41	8	49	83.6
Syntactic factors					
Subject type-matrix clause	1st SG.	211	77	188	73.3
	2nd SG.	5	2	7	71.4
	3rd SG.pronoun*	73	17	90	81.1
	3rd SG + 1st PL *on*	13	7	20	65.0
	2nd plural	1	0	1	100
	3rd plural	22	3	25	78.1
	Indefinite	1	0	1	100
	No subject	11	3	14	78.6
	Impersonal	25	7	32	78.1
	Noun Phrase	25	3	28	89.3
Subject type-subordinate clause	1st SG.	120	31	151	79.5
	2nd SG.	15	7	22	68.2
	3rd SG. m. pronoun	67	9	76	88.2
	3rd SG. f. pronoun	3	1	4	75
	3rd *on*	18	3	21	85.7
	3rd plural	23	7	30	76.7
	Indefinite	2	1	3	66.7
	Demonstrative	1	0	1	100
	No subject	8	0	8	100
	Impersonal	82	45	127	64.6
	Noun Phrase	48	15	63	76.2
Intervening material	none	329	97	426	77.2
	yes	58	21	79	73.1
Semantic factors					
Polarity	positive	360	111	471	76.4
	negative	27	7	34	79.4
Phonological factors					
Segment preceeding COMP	vowel	230	55	285	80.7
	[l]	7	3	10	70
	a	2	0	2	100
	[r, R]	15	5	15	75
	[z]	4	2	6	66.7
	[j]	1	0	1	100
	other fricative	41	16	57	71.9
	sibilant	82	33	115	71.3

Table 2. Continued

Factor group	Factor	N			%
		que	Ø	Total	*que*
	other stops	1	0	1	100
	[t]	3	1	4	75
	[b]	1	3	4	25
	[p]	0	1	1	0
Segment following COMP	vowel	121	19	140	86.4
	liquid	30	7	37	81.1
	nasal consonant	11	5	26	68.8
	fricative	1	1	2	50
	[j]	101	28	129	78.3
	sibilant	78	45	123	63.4
	[k]	7	1	8	87.5
	[p]	3	3	6	50
	[t]	24	6	30	80
	[d]	2	2	4	50
	[b]	9	1	10	90
Total		387	119	506	76.5

* All tokens of this type were for masculine pronouns.

Table 3. Profiles of the speakers

Pseudonym	Sex	Contact	Education	Friends	Work language	Grammar score
Jocelyn	F	high	high	French spouse	both	95–100%
Ted	M	high	high	French friends	both	95–100%
Vincent	M	high	high	French friends	English	95–100%
Liz	F	high	high	no French friends	both	95–100%
Sandra	F	high	high	no French friends	both	95–100%
Joanie	F	high	medium	no French friends	both	85–90%
Louisa	F	medium	high	French spouse	English	85–90%
Kurt	M	medium	high	French friends	both	65–70%
Gloria	F	medium	high	no French friends	both	95–100%
Janie	F	medium	high	no French friends	English	85–90%
Glenda	F	medium	medium	no French friends	both	85–90%
Joan	F	medium	low	French spouse	both	85–90%
Greg	M	low	medium	French friends	both	75–80%
Mike	M	low	medium	French friends	both	75–80%
Jack	M	low	medium	no French friends	English	65–70%
Tony	M	low	low	French spouse	both	75–80%
Peter	M	low	low	no French friends	both	65–70%
Donald	M	low	low	no French friends	English	75–80%

4.2   Frequency of matrix verbs

The lexical frequency of the different matrix verbs may be relevant in accounting for the variable patterns, at least in English (cf. Berkenfield 2001, Torres Cacoullos and Walker *forthcoming*). Tables 4 and 5 provide separate counts for common collocations vs. other uses of frequent verbs. In addition to calculating the relative frequency of each of the matrix verbs in our own corpus, we looked up the frequency of the verbs in the British National Corpus (Leech et al. 2001) and Corpaix (Véronis 2000), corpora of spoken language for which frequency statistics are available online.[4]

There are some noteworthy similarities in matrix verb usage between the two languages. Table 6 compares the most frequent matrix verbs in our corpora of AMF and AME. The same verbs are most common in both languages (*think, say, know* in English and their synonyms *penser, croire, dire, savoir* in French).

Table 4. Frequency counts for matrix verbs in AMF (N = 506 tokens)

Common collocation	N	% of tokens	% of tokens
Je pense "I think"	104	21%	
Je trouve "I find"	58	11%	
Il faut "to be necessary"	32	6%	
C'est X "it is X"	22	4%	
Je dirais "I'd say"	14	3%	56% (N = 283)
Je sais "I know"	12	2%	
Je crois "I believe"	10	2%	
Il me semble "it seems to me"	10	2%	
Étant donné "considering"	8	2%	
On dirait "one would say"	6	1%	
C'est pas X "it is not X"	7	1%	
Other *dire* "to say"	44	9%	
Other *savoir* "to know"	27	5%	44% (N = 223)
Other *falloir* "to have to"	21	4%	
Other *penser* "to think"	18	4%	
other verbs, other forms of the common verbs	113	22%	

---

4.   It is an open question how best to calculate the effect of lexical frequency (cf. Dinkin 2007; Abramowicz 2007: 31–2). Is the frequency of the forms used in the particular variable context what is relevant, or does frequency of use in other syntactic positions also play a role? As a third method, we calculated the frequency of appearance for each matrix verb in the AME corpus overall (not just its uses as a matrix verb). The three types of frequency calculations are highly correlated (AME-overall & BNC: r^2 = 0.40, p < .05; AME-as matrix & AME overall: r^2 = 0.28, p < .05). Further inquiry might reveal which method of frequency calculation best fits the model, but we proceed using the BNC calculations, which contribute a strong effect for lexical frequency in the multivariate analyses presented below.

Table 5. Frequency counts for matrix verbs in AME (N = 1,013 tokens)

Common collocation	N	% of tokens	% of tokens
I think	304	30%	
I know	82	8%	
I don't think	74	7%	
I'd say, I said	34	3%	60% (N = 605)
I guess	38	4%	
I remember	34	3%	
I'm sure	16	2%	
I find	13	1%	
I figure	10	1%	
other *think*	88	9%	
other *know*	19	2%	
other *say*	22	2%	40% (N = 408)
tell	23	2%	
realize	21	2%	
other verbs, other forms of the common verbs	234	23%	

Table 6. Frequency counts for matrix verbs in AMF and AME (N = 967*)

	AME		AMF	
Other Verbs	N	% of tokens	N	% of tokens
think/ penser, croire	88 (+304)	9% (+30%)	134	26%
say/ dire	22 (+34)	2% (+3%)	64	13%
know/savoir	19 (+82)	2% (+8%)	39	8%
falloir			53	10%
tell	23	2%		
realize	21	2%		
voir			16	3%
sembler			10	2%
all other forms & other verbs	234	23%	156	31%
TOTAL	1,061		506	

*N = 907 for the analysis involving lexical frequency because 60 tokens were excluded for which we were not able to establish the lexical frequency of the matrix verb.

This is true whether the frequent collocations of matrix subject *I* + verb are included (the additional values shown in parentheses) or not. Next most frequent, however, we see five verbs which are commonly used in one language, but not the

other (*trouver* 'to find' *falloir* 'to be necessary', *voir* 'to see', and *sembler* 'to seem' in French; *tell* and *realize* in English).

5.   Multivariate analysis

Thus far, we have only presented raw numbers and percentages. We turn next to analyses that allow us to clarify the independent effects of each variable.

5.1   Effects of linguistic factors in English

Among the factors mentioned, there are a number which interact, so a binomial regression analysis cannot be conducted with all factors at once. First, animacy is fairly redundant with subject type (first and second person pronouns always refer to humans, third person masculine and feminine also refer to humans exclusively in the AME corpus and quite frequently in AMF). Inanimates only appear in the pleonastic/neuter or the NP categories. In a recoding of the AME data, in which animacy is only examined within these categories, it was found not to have a significant effect and is therefore not further considered.

Second, there are four factor groups which logically interact with each other. All of them relate to the matrix verb of the sentence (verb identity, semantic class, finiteness, and lexical frequency as determined by the BNC). Four binomial analyses were conducted with just one of these factors each, and their results are compared in Table 7. (Just the weights for the one factor which was switched are presented.) As in all Goldvarb analyses presented in this paper, the application value is presence of the COMP.

Of the four interacting factors that describe the matrix verb, the one which best fits the data is lexical frequency (having the smallest log likelihood). Therefore, we proceed with further analysis including this factor group and excluding the other three.

Table 8 provides factor weights for all significant factor groups in an analysis in which lexical frequency is included, but the other three factor groups relating to matrix verb identity are excluded. Factors are listed in decreasing order of significance.

5.2   Effects of linguistic factors in French

We next present the effects of these factors in French, and then integrate our discussion of the effects in both languages. The French data is comprised of 506 tokens. The rate of COMP presence is 77%. A multivariate analysis was performed on the 506 tokens extracted from the French interviews. The multivariate analysis tested for the following five linguistic factors: preceding phonological environment, following phonological environment, subject of the matrix clause, subject

Table 7. Comparison of four analyses with 1 of 4 interacting factor groups (AME). Application value is COMP presence

Verb identity*	Factor weight	% that	N
realize	*0.886*	*71*	*21*
other *find*	*0.879*	*63*	*8*
know	*0.854*	*53*	*34*
tell	*0.846*	*65*	*23*
other	*0.780*	*54*	*183*
I said	*0.680*	*33*	*6*
say	*0.600*	*35*	*37*
I know	*0.579*	*21*	*67*
I find	*0.533*	*31*	*13*
I'm sure	*0.409*	*13*	*16*
I think	*0.365*	*13*	*302*
I wish	*0.336*	*10*	*10*
think	*0.320*	*14*	*90*
I guess	*0.285*	*11*	*38*
I don't think	*0.236*	*07*	*74*
I'd say	*0.211*	*08*	*12*
I remember	*0.197*	*09*	*33*
Semantic class**			
suasive	*0.880*	*71*	*14*
comment	*0.819*	*70*	*30*
knowledge	*0.664*	*37*	*189*
utterance	*0.604*	*43*	*88*
attitude	*0.411*	*16*	*624*
extraposition	*0.347*	*59*	*22*
Finiteness†			
non-finite	*0.783*	*68*	*19*
finite	*0.494*	*25*	*948*
Lexical frequency (in BNC)‡			
>100	*0.703*	*43*	*84*
<100	*0.614*	*32*	*64*
>1000	*0.436*	*18*	*539*

*Input 0.198, Number of factors 17, Range .689, Log likelihood −411.
**Input 0.218, Number of factors 6, Range .533, Log likelihood −443.
†Input 0.227, Number of factors 2, Range .289, Log likelihood −464.
‡Input 0.201, Number of factors 3, Range .267, Log likelihood −415.

of the subordinate clause and the lexical identity of the matrix verb. The results for the three factors selected as significant are illustrated in Table 9.

The most significant factor is the lexical identity of the matrix verb, with a range of 58 points. The analysis identified five verbs disfavoring the presence of

Table 8. Binomial regression analysis for signficant factors in AME, including lexical frequency; ranked in order of decreasing significance (Application value is COMP presence, Input = .200, N = 869)

Group	Factor		Weight	%
Lexical frequency	Frequent (>100)		0.70	43%
(in BNC)	Rare (<100)		0.60	33%
	Very frequent (>1000)		0.44	19%
		range	.26	
Matrix clause subject	Plural noun		0.82	48%
	3rd person pronoun		0.76	46%
	2nd person pronoun		0.57	31%
	Singular noun		0.49	30%
	1st person pronoun		0.46	19%
		range	.36	
Subordinate clause subject	1st PL.		0.66	31%
	Noun (SG. & PL.)		0.65	41%
	1st & 2nd p. SG.		0.57	26%
	Neuter & plural pronoun (human)		0.51	27%
	Non-human neuter pronoun (e.g. *it/there*)		0.50	19%
	3rd SG. pronoun (human & *that*)		0.22	11%
		range	.44	
Intervening material	argument (indirect object, prep. phrase)		0.86	71%
	other		0.62	36%
	none		0.48	22%
		range	.38	
Following phonological	obstruent		0.61	26%
segment*	vowel or sonorant		0.42	23%
		range	.19	
Polarity	both negative		0.80	50%
	matrix positive, subordinate negative		0.70	44%
	both positive		0.48	23%
	matrix negative, subordinate positive		0.48	18%
		range	.32	

*While this factor is significant and suggests a sonority effect, it is worth pointing out that the effect is non-monotonic when finer divisions are made in the Sonority Hierarchy. See Table 10.

COMP: *sembler* 'to seem', *souvenir* 'to remember', *dire* 'to say', *penser* 'to think' and *croire* 'to believe'. Other verbs, such as *savoir* 'to know', *falloir* 'to be necessary', *voir* 'to see', *trouver* 'to find' and *être* 'to be', seem more favorable to the presence of COMP *que*, as do other infrequent verbs in our data.

There is significantly more COMP presence if there is a vowel or a liquid following the COMP site. Sibilant and stops instead disfavor the presence of COMP

Table 9. Significant linguistic factors in AMF (Application value is COMP presence, Input = 0.735, N = 506)

	Factor	Weight	%
Lexical identity of the matrix verb	other verbs 1–4	.78	85%
	être	.59	83%
	falloir	.58	85%
	voir	.55	81%
	other verbs 5–15	.53	83%
	trouver	.52	75%
	savoir	.50	77%
	croire	.49	75%
	penser	.43	71%
	dire	.40	70%
	souvenir	.25	50%
	sembler	.19	50%
	range	*.58*	
Following phonological environment	vowel	*.65*	86%
	liquid	.55	81%
	fricative	.51	78%
	stop	.46	77%
	sibilant	.34	71%
	range	*.31*	
Subject of the matrix clause	Noun Phrase	.71	89%
	3rd pronoun	.57	79%
	1st & 2nd pronoun	.44	62%
	range	*.27*	

que. However, the preceding phonological environment was rejected from the multivariate analysis. The AMF speakers are not influenced (directly) by this factor in the choice of COMP presence.

As far as the syntactic factors are concerned, the subject of the matrix clause is significant (but with a lower range of 27 points). NP subjects are the most favorable to the presence of COMP *que,* followed by third person pronouns. In contrast, first and second person pronouns disfavor the presence of COMP. We checked if this was related to the collocation effect. For example, the verb *penser* 'to think' appears most of the time in cooccurrence with the pronoun *je* 'I'. This collocation represents 21% of the tokens (see Table 4) and is associated with 71% COMP presence. The second most frequent collocation, *je trouve,* is associated with 74% COMP presence.

For those two collocations, the rate of COMP presence is near the overall rate and therefore does not sufficiently explain of why the syntactic factor is selected by the analysis. Finally, the subject of the subordinate clause was not significant in AMF. We next compare multivariate analyses of the two languages.

5.3   Anglophone Montreal English (AME) vs. French (AMF)

The rate of COMP presence in the two languages differs. In AME, it reaches only 27%, while in AMF it is 77%. This difference of 50 percentage points confirms that the behavior of the informants regarding COMP usage is quite different in French and English. Comparing the general tendency with the most recent study available for L1 Quebec French (speakers living in Vieux-Hull, Quebec) we find a better match. COMP presence there is 63% for the younger speakers and 68% for the older speakers (Dion 2003: 28). Taking the younger speakers of Dion's study as the counterparts of our young Anglo-Montrealers, we observe a difference of 14 percentage points.[5] Therefore, from the strict perspective of rates of usage, one cannot see evidence of convergence or inter-systemic influence. In other words, the lower rate of presence of *that* in English does not seem to interfere with the presence of *que* in French. In addition, we observe that the even higher rate of COMP presence in AMF in comparison with L1 French seems to reflect more conformity with the prescriptive norm of French, where *que* deletion is highly stigmatized. Comparing the input values for the two samples (see Table 10), we see again that there is a much higher rate of COMP presence in AMF than AME – it is not a spurious effect of the differential usage of different types of sentence constructions, etc., but a true difference between the grammars.

The first section of Table 10 compares the factor weights for the various matrix verbs, in decreasing order of frequency of COMP presence in AMF. The nine verbs which appeared most frequently in the AMF corpus are listed individually, then the four next most common are grouped, followed by a grouping of the next 10 most common. Where the same verbs appeared frequently in AME, factor weights are listed. This table illustrates that not all of the same verbs were in the most common set in both languages and that their order in terms of preference for a surface COMP differs as well.

---

5.   We could also compare our results with Warren (1984), which identified a rate of 88% COMP presence. Since her circumscription of the variable context is slightly different, we prefer to compare our results with Dion (2003) who defined the variable context as we did.

Table 10. Comparison of factor weights in AME and AMF (Application value = COMP presence; For AMF: N = 506, input = 0.735; for AME: N = 967, input = 0.198)

Factor group	Factor	Factor weight	
		AMF *que*	**AME** *that*
Lexical identity of the matrix verb	other common verbs (next 4 most common)	0.78	N/A
	être	0.59	N/A
	falloir	0.57	N/A
	voir	0.55	N/A
	other common verbs	0.53	0.78*
	(next 10 most common, after other 4)	0.50	0.85 *know*
	savoir		
			0.58 *I know*
	croire	0.49	N/A
	penser	0.43	0.37 *I think*
			0.24 *I don't think*
			0.32 *think*
	dire	0.40	0.60 *say*
			0.68 *I said*
			0.21 *I'd say*
	souvenir	0.25	0.20
	sembler	0.19	N/A
Following phonological environment	vowels	0.65	0.39
	fricative	0.51	0.66
	liquid	0.50	0.35
	stop	0.46	0.58
	sibilant	0.34	0.48
	NP	0.71	0.39 singular NP
			0.80 plural NP
Subject of the matrix clause	3rd person pronoun	0.57	0.58 feminine
			0.48 masculine
	1st & 2nd person	0.40	0.47 1st sɢ.
			0.23 2nd
			0.81 1st ᴘʟ.

*This value is for all the uncommon English verbs combined (all tokens except the sixteen most frequent collocations and verbs).

Turning to examine the effect of the following phonological environment, we see a difference between the two languages: less sonorant following segments favor the presence of the COMP in AMF, while more sonorant segments favor it in AME, when just a binary [+/–sonorant] distinction is made. As shown in Table 10, however, there is no consistent effect of sonority in the AME data, when finer-grained distinctions are made.

There is a different effect in the two languages for the subject of the matrix clause: NPs most favor COMP presence in AMF, while their weight is between that of first and second person vs. third person pronouns in AME.

In sum, it is not possible to say that the same grammar is used in the two languages – these speakers have acquired distinct grammars for AMF and AME.

5.4 AMF vs. L1 French

We turn next to a comparison of factor effects in L1 French to see whether these L2 speakers have in fact acquired the same patterns as native speakers. For both AMF and L1 French, the lexical identity of the matrix verb is an important factor. In Dion's (2003) findings, *sembler* and *penser* were also among the verbs that disfavor *que* presence. Our results also resemble Martineau's (1985) and Warren's (1994) findings which reported an association with the lexical identity of the verb. In particular, Warren (1994: 45), who devoted a qualitative analysis to frequent collocations with the verbs *dire* and *penser,* indicated that *je pense* was frequently associated with the absence of *que* in Montreal French. However, while *dire* strongly disfavors COMP presence in AMF, this verb was less associated with deletion in Dion's study of L1 French.

Noteworthy is the fact that the verb *falloir* favors *que* in AMF, but not in L1 French (Dion 2003). In our data the collocation *il faut* is the third most frequent (see Table 4). Dion found that *falloir* was very often used without *que* as illustrated in (7a). In our data, *falloir* instead favors *que,* as exemplified in (7b).

(7) a. *Fallait Ø je me place en quelque part* (cited in Dion 2003: 9)
         'It was necessary Ø I put myself somewhere.'

    b. *Il faut <u>que</u> je le fasse tu sais.* (Louisa)
         'It's necessary <u>that</u> I do it, you know.'

This might be related to the distinct syntactic nature of the English equivalent *have to* + infinitive, a construction not involving the presence of a COMP. The constraints associated with this verb might not be acquired in the same way as other verbs with an equivalent structure in both languages. In addition, the strong correlation of *falloir* with the use of the subjunctive (Poplack 2001) may play a role.

In the context of formal acquisition of French L2, the complete collocation *il faut que* is learned in association with the French subjunctive.

Similar results are reported for *falloir* in Acadian French where rates of *que* absence fall below average (King & Nadasdi 2006). This verb is also among the most frequent verbs (Véronis 2000). However this verb is not associated with deletion in our data. At the other end, the verb *trouver* 'to find', which has medium frequency, is more favorable to deletion than the frequent verb *falloir*. Such observations do not support the association between verb frequency and deletion for the French data, as noted by King and Nadasdi (2006).

The subject of the matrix clause plays a role in AMF but with a weaker effect. While it is true that first and second person pronouns disfavor the presence of *que*, we hypothesize that this is related to the frequency of certain types of collocations, e.g., the verb *penser* 'to think' which is the most often used in collocation with the pronoun '*je*' in our data. This was reflected in Warren's (1994) study, the only study that examines collocation effects in French. She devoted a qualitative analysis to three frequent expressions: *je pense* 'I think', *disons* 'let's say' and *parce que* 'because'.[6] In her data, *je pense* also favors the absence of *que*, paralleling our results. In our data the collocation *je trouve* was also associated with the absence of *que*, parralleling Warren's findings. Based on Thompson & Mulac's (1991) proposal, Warren considered these two constructions to be "unitary epistemic phrases." In her data set, these collocations are both associated with a higher degree of autonomy: they may appear in different positions in the utterence (Warren 1994: 47). In contrast, *je trouve* was not specifically associated with COMP absence in Dion's (2003) study.

Regarding the effect of the following phonological environment, AMF speakers exhibit parallel tendencies to L1 French speakers (Sankoff 1980; Martineau 1985; Warren 1994; and Dion 2003). There is no effect of the following syntactic environment in our data. This result may be explained by Dion's (2003) finding that the following syntactic environment factor interacts with the following phonological environment. Because the pronouns are not a homogeneous category, phonologically speaking, she considered the syntactic effect an epiphenomenon of the phonological one.

In sum, many of the linguistic constraints for AMF mirror those found significant in studies of French L1, confirming that these AMF speakers have acquired the grammar of L1 Francophones.

---

6.   The expression *disons* is not frequent enough in our data set (less than 5 occurrences) and is not further discussed. The expression *parce que* is not part of the variable context and is also not further discussed.

5.5 AME vs. Quebec City English (QCE)

Finally, we compare the AME grammar to that of another group of native speakers of English. This group lives in Quebec City and all qualify themselves as bilingual, "having learned French informally" in a city in which Anglophones have minority status (Torres Cacoullos & Walker *forthcoming*: 8). A majority of the speakers also report using French on a daily basis (Poplack, Walker & Malcolmson 2006: 199–201). They do not differ greatly from the AME speakers in terms of reported contact with French.

The first thing to note is the similarity in overall rate of COMP presence: 21% in Quebec City English (QCE) and 26% in AME. In both, the effect of verb identity is strong and significant, and similar verbs favor COMP presence in both languages. A second shared feature is the favoring effect of the presence of other morphemes between the matrix verb and the subordinate clause, a factor to which we return in Section 6. However, there are numerous differences between the Englishes of these two groups.

Our results differ from Torres Cacoullos and Walker's for the effect of both the matrix clause subject and the subordinate clause subject. While they report matrix clause noun subjects to favor *that* more than pronouns (Torres Cacoullos & Walker *forthcoming*: 21), we find a difference in the behavior of singular and plural nouns, and cannot combine them. Plural nouns most strongly favor COMP presence, while singular nouns are the least favoring. Pronouns fall between. We also see salient differences across types of pronouns in AME, which they do not report: third person pronouns have a stronger preference for COMP than first & second person. As far as subordinate clause subjects, there is a significant effect in QC: Nouns favor COMP most, *it* and *there* favor it least, with other pronouns falling in between) in QC (*ibid*: 21). Table 8 illustrates a different pattern for AME.

Two other factors emerge as significant in AME but not in QCE: the sonority of the phonological segment following the COMP (apparently not tested in QC) and polarity differences between the matrix and subordinate clause. (But, see Table 10 regarding the inconsistent sonority effect.) On the other hand, matrix clause finiteness and subordinate clause transitivity are not significant in the AME data but are in QCE (*ibid*.).

5.6 Comparison across four varieties

Table 11 presents a summary of the comparisons we have made. Shading highlights relevant similarities between speech varieties. Looking at AMF, the second language of our speakers, we can see that in several respects (overall frequency of COMP, the effects of the matrix verb, and the two phonological factors) it patterns like that of native speakers' French. There is no evidence that the AME grammar is

Table 11. Comparison of factor effects in L1 and L2 English and French

	QC English	AME	AMF	L1 French
Overall rate of COMP	21%	26%	77%	63%
Lexical Matrix verb	Verb identity	Verb identity/ frequency	Verb identity	Verb identity
Syntactic complexity Material between matrix verb & subord. clause	Yes > No	Yes > No	*not sig.*	Not tested
Morphological Matrix clause subject	NP >>Pronoun	NP-plural >>3rd pronoun >1st & 2nd >> NP-sing *(interacts w/ subord. subject)*	1st & 2nd >>3rd pronoun >> Noun *(interacts w/ verb identity)*	NP >>3rd pronoun >> 1 st & 2nd
Subord. clause subject	NP >> Other pronoun >> *I* >> *it/there*	*not sig.*	*not sig.*	*not sig.*
Phonological Segment after COMP	Not tested	Obs >> Son (?)	Son >> Obs	Son >> Obs
Segment before COMP	Not tested	Not tested	*not sig.*	*not sig.*

interfering – no shared patterns between AME and AMF that are not also shared by L1 French. Interestingly, AME also does not pattern like QCE, except with respect to the syntactic complexity factor, which is shared across the two English varieties but not the two French varieties; the effect of the lexical identity of the verb; and the overall rate of COMP presence.

5.7 The problem with sentences like "*I think …*"

One concern that we had with our data is that certain collocations such as *I think* may not actually form a matrix clause that subcategorizes for a subordinate clause, but rather be a discourse marker or filler (much as *y'know* has come to be). In that case, the next clause is independent and could not be introduced by a surface COMP. It would, therefore, need to be excluded from the envelope of variation. In order to address this, we separated the very frequent forms from our corpus and conducted an analysis using only less common forms, which are not suspected of being on the lexicalization path toward discourse filler status. Following the

method described by Torres Cacoullos and Walker (*forthcoming*), we conducted separate analyses of common [matrix subject + verb] collocations and rarer verbs to see if the same set of linguistic factors are operational in both. This allows us to determine whether, like Torres Cacoullos and Walker's (*forthcoming*: 1), our data show that:

> [g]rammatical conditioning persists in fixed discourse formulas. Despite their high frequency and formulaic status, such formulas are not completely autonomous from the productive constructions from which they emerge … despite slight rates [of COMP presence], the linguistic conditioning of *that* in frequent collocations that behave like discourse formulas parallels its conditioning in the general construction.

This also allows us to contrast the patterning of COMP presence in AME and QCE in one other way. Table 12 shows the similarity of overall effect between AME and QCE. The input values for COMP presence in more vs. less common collocations are similar, and the differences between the two are nearly identical.

Table 12. Input values for subsamples of common vs. rare collocations in AME and QCE. (Application value is COMP presence.)

	Common collocations	Less comon collocations	Difference
AME	0.122	0.415	0.29
QCE	0.049	0.328	0.28

However, when comparing factor effects in these two subsamples, we see different patterns in AME and QCE. The *only* significant independent factor in the AME analysis of common collocations is the presence of material intervening between the matrix verb and the subordinate clause (See Table 13, and see Table 5 for a list of the frequent collocations included). In the AME analysis that excludes the common collocations, the three significant factors are: lexical frequency, subordinate clause subject, and presence of intervening material (See Table 14). In each table, factor groups are listed in decreasing order of significance and only significant groups are listed.

Table 13. Significant factor for common collocations in AME (Analysis including lexical frequency, N = 562, Application value is COMP presence, Input = 0.127)

Intervening material	Weight	%	N
yes	0.66	23	65
no	0.47	12	497

Table 14. Significant factors for less common verbs in AME (Analysis including lexical frequency, N = 377, Application value is COMP presence, Input = 0.435)

	Weight	%	N
Lexical frequency			
< 100	0.75	68	53
> 100	0.58	54	134
> 1000	0.36	31	190
Subordinate subject			
Nouns	0.72	7	63
I	0.47	4	97
other pronouns	0.44	42	119
it/that/there	0.43	34	98
Intervening material			
argument	0.73	64	14
non-argument	0.66	64	36
none	0.47	41	327

In QCE, the same grammatical effects are evident in the common and rarer collocations. Torres Cacoullos and Walker (*forthcoming*: 28) report lexical identity, matrix and subordinate clause subject as significant in their subsample of common collocations. These are also the three most significant factors in their subsample that excludes the common collocations. Our data pattern differently from the Quebec City data. We do *not* see the same factors operating on the common and the rarer types of matrix clause verbs in AME. This is the most striking difference between the Montreal and Quebec City data. It may suggest that the AME variety has moved farther along toward grammaticization of the frequent collocations – and that the frequent collocations have in fact lost all trace of grammatical conditioning.

We are not in a position to definitely say that the grammars of QCE and AME are different, although we must leave that open as a possibility. Several other possible explanations exist for these different patterns. First, the QCE study included 2,820 tokens (*ibid*: 16)– three times as many as our analysis. Our smaller data set also contributes to a methodological problem. Since all of the common collocations are of exactly the form "*I* + matrix verb," several factors must be omitted from the analysis. These include matrix clause subject type (always a first person pronoun), animacy of matrix subject (always [+human]), and matrix clause morphology (always finite). Additionally, among the common collocation tokens, there is not enough variety to be able to examine the effects of polarity (two variants must be omitted) or coreferentiality of matrix and subordinate subjects (too few non-coreferential sentences). Several of these factors are significant in the analysis of less common collocations. Omitting them in one analysis but not the other may skew the results. These considerations notwithstanding, the comparison of

more and less common matrix subject + verb combinations discloses yet another difference between AME and QCE.

6.   Connections with *like* and verbs of quotation

When circumscribing the variable context we came across problematic examples such as (8), where the word *like* appears in the same surface position as a (deletable) COMP.

(8)   a.   And it seems to me [**like** if you're going to put up an ad targeting the English community you should make the effort to make sure it's grammatical.] (Donald)
      b.   Like I feel [**like** we're sort of cut off from the English community] (Donald)

Sometimes, such examples contain verb of quotation (VOQ) usages of *like* (Romaine & Lange 1991; Tagliamonte & D'Arcy 2007), as in (9a). Our labeling of these as VOQ usage of *like* stems from comparison to sentences such as (9b).

(9)   a.   Because I felt **like**, "Uh: OK, I understand that the French people have to protect themselves." (Louisa)
      b.   He was **like**, "No no no no." And they're **like**, "Yes yes yes yes yes." (Liz)

Additionally, in the English data, we find an effect on the rate of COMP presence of material which intervenes between the subject and the matrix verb: there is more COMP deletion when there are intervening words (see Table 13). Specifically, we observe cases where the absence of *that* co-occurs with the presence of *like*, as in (8a and b and 9a).

   Such examples raise several questions. What type of word is *like* in the canonical *feel /seem + like + CP* structure? Do these sentences contain a deleted COMP? Is *like* the COMP? Or is *like* a conjunction? What is the relationship between *like* in the (surface) COMP position and *like* in the VOQ examples (especially *be like*, as in (9b)), given that in both cases *like* has clausal scope? Answering these questions contributes to research on the grammaticization of *like* as a COMP (cf. Meehan 1991; Buchstaller 2001; Meyerhoff 2002; Tagliamonte & D'Arcy 2007). We propose that the discourse marker *like* is being reanalysed as a COMP because of its frequent appearance in the same surface position as the (deletable) COMP *that*. Because *like* in its various usages has always been associated with the concept of inexact comparison, it is an ideal candidate to introduce an inexactly quoted utterance or internal thought. Our final question is: Do AMF or L1 French speakers do anything like this with the comparative French particle *comme*?

6.1   What is *like* like?

From the perspective of grammar, it is possible that *like* is a COMP, a conjunction, or a discourse marker. First, we present grammatical and distributional evidence that *like* in this context is, in fact, a COMP. The behavior of conjunction *like* (CONJ-*like*) is distinct from COMP-*like* in several ways. The first distinction between COMP-*like* and CONJ-*like* is that CONJ-*like* can usually be replaced by *as*, another comparative conjunction (see 10a–b'), but COMP-*like* never can (10c–c').

(10)   a.   Winston tastes good, **like** a cigarette should (taste good).
      a'. Winston tastes good, **as** a cigarette should (taste good).
      b.   I feel **like** I should (feel/go to the park).
      b'. I feel **as** I should (feel/*go to the park).
      c.   She actually feels **like** she's not an Anglophone.
      c'. *She actually feels **as** she's not an Anglophone.
      d.   I think **like** a scientist does (think/work hard).
      d'. I think **as** a scientist does (think/*work hard).

Second, when there is VP-ellipsis, the elided material of CONJ-*like* must find its antecedent in the conjoined clause. In (10a), *like* is unambiguously a conjunction: the elided material must come from the previous clause, but in (10b) and (10d) *like* can be a conjunction or a COMP, giving rise to two possible interpretations. *Feel like*, as in (10b), also patterns both like a conjunction and like a COMP, in allowing either interpretation.

    A third reason to suppose that *like* in the relevant cases may be a COMP is that it not only licenses argument extraction, but it does so with the same asymmetry as unambiguous COMPs *that* and *for*. Objects can be extracted (as in 11a"–c") while subjects cannot (as in 11a'–c').

(11)   a.   She feels **like** her friend deserves the job more.
      b.   She thinks **that** her friend deserves the job more.
      c.   She wants **for** her friend to get the job.

      a'. *Who does she feel **like** deserves the job more?
      b'. *Who does she think **that** deserves the job more?
      c'. *Who does she want **for** to get the job more?

      a". What does she feel **like** her friend deserves?
      b". What does she think **that** her friend deserves?
      c". What does she want **for** her friend to get?

These three grammatical arguments indicate that *like* may be interpreted as a COMP. However, we must also consider the possibility that such sentences actually have a deleted COMP and a discourse marker *like* coincidentally occurring in the same position. The frequency of null-COMP + *like*, however, is actually

quite rare. There are only four examples of unambiguous cases in our corpus, listed in (12):

(12) a. I know **like** the advertisements all focus on this worldwide thing. (Janie)
b. I figure **like** every time: every time the PQ says "Separation" a hundred more businesses leave. (Mike)
c. I think **like** towards the end like 55th, 56th, it would be about a hundred percent. (Peter)
d. And so I remember **like** finally I said: I said "Hi" to them in French. (Donald)

Out of 256 tokens with these verbs, this is 2%. With *feel/seem*, however, 74% of the total (N = 58) are of the form *feel/seem + like*. If the *like* in *feel like* examples were a discourse marker, we would expect it to occur approximately as frequently in the same position with other verbs, but it does not. We conclude, then, that *like* is a COMP in these cases.

6.2 Relationship with VOQ-*like*

There is no one-to-one correlation between using *like* as a VOQ and using it as a COMP. There are speakers who use one, the other, both and neither. These inter-speaker differences notwithstanding, it is worth considering where *like*'s use as a COMP fits into the development of the word more generally, since it has been progressively increasing its range of uses since at least Old English *soðlice* (D'Arcy 2005). Romaine & Lange (1991: 261) propose that *like* became a conjunction before branching off into a discourse marker and a VOQ. They suggest that *like* expanded from a preposition to take sentential complements, at which point it simultaneously became a conjunction and a COMP. (They do not distinguish between *like* as a conjunction and *like* as a COMP.) Vincent & Sankoff (1992) describe its use as a discourse marker.

Quotative *like*, however, is quite new on the linguistic scene: only about 30 years old (Tagliamonte & D'Arcy 2007). Discourse marker *like* in the context of a full clause, however, dates back to the early 20th century (D'Arcy 2005). Taking a clausal complement is, then, a necessary, but not sufficient precursor, and may be part of a conspiracy of multiple causes: *like*'s discourse status, which D'Arcy (2005) notes became progressively freer over the course of the 20th century, along with its subcategorization status (taking a CP complement), put *like* in a position where it could occur before a quotation of any kind.

Recall that we have examples where feel *like* introduces a quotative. We also have examples of VOQ's *you know,* and null verbs co-occurring with *like*:

(13) a. You know **like**: "I don't have a problem with you so you don't have a problem with me." (Greg)
b. Me **like** "Waa:" (Gloria)

The final requirement would be a pragmatic need for a new form. Tagliamonte & D'Arcy (2007: 212) propose that a rising discourse option which just preceded

like's appearance as a VOQ in Toronto, that of reporting inner monologue, was actually the function that VOQ *like* served at its outset of productive use. Thus, it was the combination of *like*'s COMP status, its discourse function and scope-taking ability, and a grammatical niche opening up in the context of quotation/narration strategies, which all conspired to allow *like*'s introduction into the English quotative system.

6.3   VOQ-*comme*

Since French *comme* functions in many similar ways as English *like*, we investigated the uses of *comme* in AMF. There are no published reports of L1 French speakers using *comme* as a VOQ.[7] AMF speakers, however, do use *comme* this way:

(14)   *… tout le monde c'est <u>comme</u> "oh tu parles à toi-même"* (Thomas)
    … everyone is <u>like</u>, "Oh, you talk to yourself"

Figure 2 shows the distribution of speakers' use of *comme* for a variety of functions in L1 and L2 Montreal French.
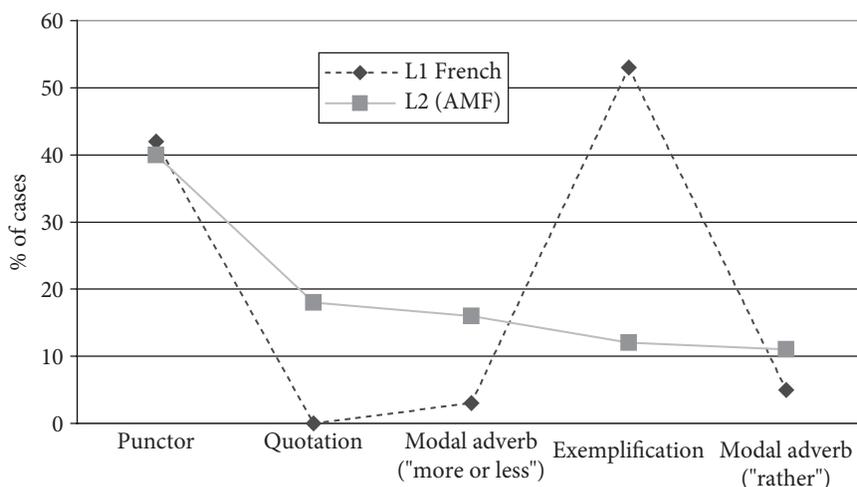
Figure 2. L1 and L2 use of *comme* in Montreal French (adapted from Figure 2 in Sankoff et al. 1997: 208).

While both L1 and L2 speakers use *comme* frequently as a punctor, and share several other usages, its VOQ use is the second most common usage among L2 speakers, but is not attested in these L1 speakers. This probably reflects the L2

---

7.   Dion and Poplack (2007) recently reported the use of VOQ-*comme* to be surprisingly common in the French spoken by students in the Gatineau area of Quebec.

speakers calqueing from English, since they would not have heard it from L1 French speakers.

A natural question is why L1 French speakers would not independently have anything resembling VOQ-*comme*. Returning to the functions of *comme*, a possibility suggests itself: the conspiracy of *comme*-uses was not sufficiently strong. Earlier, we suggested that it was *like*'s multiple functions in discourse as well as its status as a COMP, along with Tagliamonte & D'Arcy's suggestion that there was an increase in use of internal monologue in discourse, that set the stage for VOQ-*like*. *Comme* is a discourse marker with considerable freedom: a conjunction, an exemplifier, an adverb and a preposition. But it does not take a clausal complement. Thus, either one or both of two facts are crucial: (1) *comme* is not a COMP, like *like* is, or (2) French speakers, though they could develop something like VOQ-*comme*, do not have the discourse need. This latter fact has not been empirically verified. If the discourse strategy of French oral narrative does not involve using internal monologue to the extent that it does in English, the necessary discourse requirements would not be present. But L2 speakers would use it, if they retain their L1 discourse strategies.

6.4 Summary for *like* and *comme*

We have suggested that *like* is a COMP with some verbs in some constructions. We provided grammatical and distributional evidence for this claim, and suggest that its COMP status might have been crucial to the development of VOQ-*like*. Finally, we examined the use of VOQ-*comme* in AMF and noted that L1 speakers did not use it contemporaneously. We suggest two possible reasons for its absence (at the time of data collection): either *comme* must take a full clausal complement (be a COMP) before it can be a VOQ, or else the grammar of French offers the possibility, but it is a possibility unrealized due to differing discourse strategies of native French speakers. Dion & Poplack (2007) report that there are now some French speakers who use this construction. This offers the exciting prospect of testing these hypotheses in the future by examining the narrative strategies of the French speakers who use it.

7. Effects of social factors

We close by assessing the roles of the social factors with this variable, trying to better understand the linguistic patterns in their social context. As this paper is part of a bigger initiative to understand the way French and English are used by Montreal Anglophones, we focus on social factors that relate to the use and meaning of COMP variation in the speech of these young bilingual Anglophones.

As one might expect, very little of the variation in the English data may be attributed to the social factors. That is, the amount and type of exposure that native English speakers have to French does not impact the way they mark their subordinate clauses in English. Of the social variables considered, the only one with a significant effect is the language used with friends and family: speakers with a Francophone family or roommate have a higher rate of COMP presence than those who do not (factor weights of 0.72 and 0.44 respectively). While this might intially be interpreted as an effect of the high rate of COMP presence in French, there is no difference between speakers with French friends and those without (both 0.44), the two other options in our scale of French contact, suggesting that the correlation seen for family/roommate may well be coincidental. For completeness, we mention that there is no effect of sex: males and females both have factor weight of 0.50 for COMP presence in English.[8] Because this pattern is not stigmatized in spoken English, nor does it seem to be part of a change in progress, this is not surprising.

In the AMF data, one important finding from our analysis of the effects of social factors is that the more contact with native French speakers a person has, the less frequently they produce the COMP.[9] Speakers with very low contact tend to have a higher rate of COMP *que* presence, following more closely the prescriptive norm taught in school (factor weight of 0.81, vs. 0.31 and 0.47, respectively, for those with medium and high amounts of contact). This behavior resembles what we have found for pronominal variation between *nous* and *on* in French: people with very low contact tend to favor prescriptive *nous* over colloquial *on* (Blondeau et al. 2002).

This complements our analysis of several other linguistic variables (e.g., Nagy et al. 2003; Blondeau et al. 2002) in which Anglo-Montrealers who had more contact with French more closely approximated the quantitiative patterns of Montreal French than did those speakers with less contact.

8. Summary

We have shown that the systems which govern the presence of COMP in AME and AMF are quite separate, do not interact with each other, and are not obviously used as a way of marking affiliation with a particular attitude or identity. Speakers'

---

8. This was determined using the step-up/step-down analysis in Goldvarb, including all the social factors and the linguistic factors which proved significant in the analyses presented above.

9. This was the only social factor selected as significant in the multivariate analysis.

rates of COMP presence respond to different combinations of linguistic and social factors in the two languages. And while the amount of contact that speakers have with French influences COMP rates in both languages, the effect is in the opposite direction in the two languages. The one place where we see interaction between the AME and AMF patterns is in the use of *like* and *comme* as VOQ's. This is a frequent pattern in AME and has extended into AMF in a way that native-speaker French did not exhibit at the time our data was collected. Interestingly, ten years later, its use has risen in native-speaker French to a level which is noted in at least one scholarly presentation, in a region where contact between English and French is relatively high. (Dion & Poplack 2007). This may set the stage for this variable to gain a role in identity-marking, but that question awaits further research.

References

Abramowicz, L. 2007. Sociolinguistics meets exemplar theory: Frequency and recency effects in (ing). *U. Penn Working Papers in Linguistics* 13(2): 15–6.

Beaulieu, L. & Cichocki, W. 2002. Le concept de réseau social dans une communauté acadienne rurale. *Revue canadienne de linguistique* 47(3): 123–150.

Berkenfield, C. 2001. The role of frequency in the realization of English that. In *Frequency and the Emergence of Linguistic Structure*, J. Bybee (Ed.), 281–308. Amsterdam: John Benjamins.

Biber, D. 1999. A register perspective on grammar and discourse: Variability in the form and use of English complement clauses. *Discourse Studies* 1: 131–150.

Blanche-Benveniste, C. 2000. Corpus de français parlé. In *Corpus*: *Méthodologie et applications linguistiques*, M. Bilger (Ed.), 123–8. Paris: Champion et les Presses Universitaires de Perpignan.

Blondeau, H., Nagy, N., Sankoff, G. & Thibault, P. 2002. La couleur locale du français des Anglo-Montréalais. *In L'acquisition de la variation par les apprenants du français langue seconde*. [Acquisition et Interaction en Langue Étrangère 17] J.-M. Dewaele & R. Mougeon (Eds.), 73–100.

Buchstaller, I. 2001. An alternative view of like: Its grammaticalisation in American English and beyond. *Edinburgh Working Papers in Applied Linguistics* 11(1): 21–41.

Connors, K. 1975. L'effacement de *que* - règle syntaxique. *Recherches linguistiques à Montréal* 4: 17–33.

D'Arcy, A. 2005. Like: Syntax and Development. Ph.D. dissertation, University of Toronto.

Dinkin, A. 2007. The real effect of word frequency on phonetic variation. Paper presented at Penn Linguistics Colloquium 31.

Dion, N. 2003. L'effacement du *que* en français canadien: Une étude en temps réel. MA thesis, University of Ottawa.

Dion, N. & Poplack, S. 2007. Linguistic mythbusting: The role of the media in diffusing change. Paper presented at NWAV 36, University of Pennsylvania.

Gadet, F. 2007. *La variation sociale en français*. 2nd edn. Paris: Orphys.

King, R. & Nadasdi, T. 2006. Another look at *que* deletion. Paper presented at NWAV 35, Ohio State University.

Leech, G., Rayson, P. & Wilson, A. 2001. Companion website for word frequencies in written and spoken English: Based on the British National Corpus. Longman: London. http://www.comp.lancs.ac.uk/ucrel/bncfreq/lists/2_2_spokenvwritten.txt.

Martineau, F. 1985. Élision variable de (*que*) dans le parler d'Ottawa-Hull. MA thesis. Université d'Ottawa.

Meehan, T. 1991. It's like, "What's happening in the evolution of 'like'?" A theory of grammaticalization. *Kansas Working Papers in Linguistics* 1: 37–51.

Meyerhoff, M. 2002. All the same? The emergence of complementizers in Bislama. In *Reported Discourse: A meeting ground for different linguistic domains*, T. Güldemann & M. von Roncador (Eds), 341–359. Amsterdam: John Benjamins.

Nagy, N. & Blondeau, H. 2005. I think [$_{CP}$ [$_{COMP}$ Ø] [$_{IP}$ that is deleted in Montreal]]. Poster presented at NWAV 34, New York University.

Nagy, N., Blondeau, H. & Auger, J. 2003. Second language acquisition and "real" French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change* 15(1): 73–103.

Nagy, N., Blondeau, H. & Wood, J. 2007. On était comme, "We think 'like' may be a COMP too." Paper presented at NWAV 36, Philadelphia.

Poplack, S. 1989. The care and handling of a megacorpus: The Ottawa-Hull French project. In *Language Change and Variation*, R. Fasold & D. Schiffren (Eds), 411–451. Amsterdam: John Benjamins.

Poplack, S. 2001. Variability, frequency, and productivity in the irrealis domain of French. In *Frequency and the Emergence of Linguistic Structure*, J. Bybee & P. Hopper (Eds), 405–428. Amsterdam: John Benjamins.

Poplack, S., Walker, J.A. & Malcolmson, R. 2006. An English like no other? Language contact and change in Quebec. *Canadian Journal of Linguistics* 51(2): 185–213.

Romaine, S. & Lange, D. 1991. The use of "like" as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech* 6: 227–279.

Sankoff, D., Tagliamonte, S. & Smith, E. 2005. Goldvarb X: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.

Sankoff, G. 1980. *The Social Life of Language*. Philadelphia PA: University of Pennsylvania Press.

Sankoff, G. & Brown, P. 1976. The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language* 52: 631–66. (Reprinted in *The Social Life of Language*, G. Sankoff (Ed.), 211–255. Philadelphia PA: University of Pennsylvania Press).

Sankoff, G., Sarrasin, R. & Cedergren, H. 1971. Quelques considérations sur la distribution de la variable *que* dans le français de Montréal. Paper presented at the Congrès de l'Association canadienne-française pour l'Avancement des Sciences.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M.-O. & Gagnon, L. 1997. Variation and the use of discourse markers in a language contact situation. *Language Variation and Change* 9: 191–218.

Tagliamonte, S. & D'Arcy, A. 2007. Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change* 19: 199–217.

Thompson, S. & Mulac, A. 1991. The discourse conditions for the use of the complementiser that in conversational English. *Journal of Pragmatics* 15: 237–251.

Torres Cacoullos, R. & Walker, J. A.. Forthcoming. On the persistence of grammar in discourse formulas: A variationist study of that. *Linguistics* 46(6).

Véronis, J. 2000. Fréquences des mots en français parlé. http://www.up.univ-mrs.fr/~veronis/donnees/.

Vincent, D. & Sankoff, D. 1992. Punctors: A Pragmatic variable. *Language Variation and Change* 4: 137–62.

Warren, J. 1994. Plus ça change, plus c'est pareil: The case of 'que' in Montreal French. *Culture* 14(2): 39–49.

Appendix A: Examples for each factor code, in AME

This section provides illustrative examples from our corpus for each factor in all of the linguistic factors considered. Factors are also defined, as necessary.

MATRIX VERB IDENTITY AND FREQUENCY

Lexical frequency of the matrix verb

For the English data, we used Leech et al. (2001), which provides lemmatized frequencies for spoken British English, from the British National Corpus.

very frequent (>1000 in BNC)	That means [we won't have our jobs.] (Vincent)
frequent (>100)	I mentioned [I love uh geography] (Peter)
infrequent (>10)	I would imagine [he didn't.] (Peter)
rare (<10)	It didn't register [that that it was very important] (Peter)
other (not listed in the ANC)	She's just absolutely amazed [that an entire culture speaks two languages.] (Louisa)

Matrix verb identity
Common collocations

I think	I think[it was only two people.] (Liz)
I don't think	I don't think [it's a big problem.] (Sandra)
I guess	I guess [I wasn't persecuted enough.] (Liz)
I remember	I remember [the teacher was always on my back.] (Vincent)
I find	Now I find [they're very judgemental up there] (Ted)
I'm sure	I'm sure [it's helpful] (Liz)
I wish	I wish [that I could do that.] (Donald)

Other forms

realize	That's when I realized [that French was all around me] (Peter)
other forms of *find*	I found [I'd get home] (Mike)
tell	He tells you [well Principal Black might help us.] (Vincent)
know	I know [that it's wrong]. (Louisa)
say	My parents said [that was a dumb move,] (Mike)
remind, other forms of *remember*	reminds them [that I do.] (Liz)

thought, other forms of *think*	I thought [I'd lost it.] (Sandra)
figure	I figure [it should be a business choice.] (Mike)
other	I felt [I was forgetting a bit] (Peter)

SYNTACTIC FACTORS

Subject type (matrix clause)

1st SG.	I know [they speak 2 languages at home.] (Vincent)
2nd SG.	You said [you spoke to him.] (Peter)
3rd SG. f. pronoun	She's just amazed: [that we speak both languages.] (Louisa)
3rd SG. m. pronoun	He finds [that I'm a completely different person.] (Sandra)
3rd SG. Noun Phrase	My mum tells me [I used to come home in tears.] (Sandra)
1st PL.	We never thought [it was important.] (Greg)
3rd PL. pronoun	They could detect [that I I was an Anglophone] (Peter)
3rd PL. Noun Phrase	People are aware [that they'll freak out at you, if you speak in English.] (Kurt)
neuter / pleonastic	It doesn't mean that [the class will not be conducted in English.] (Joan)
other (e.g., gerund, infinitive)	I would hate to think [that it was that]. (Louisa)

Subject type (subordinate clause)

1st SG.	I felt [I was forgetting a bit] (Peter)
2nd (SG./PL.)	I don't imagine [you got along with every one of your teachers.] (Peter)
3rd SG. f. pronoun	I just thought [she was such a pixy.] (Louisa)
3rd SG. m. pronoun	I would imagine [he didn't.] (Peter)
3rd SG. Noun Phrase	I thought [French im: immersion worked really well.] (Louisa)
1st PL.	They think [we talk behind their back and whatever.] (Vincent)
3rd PL. pronoun	I never thought [they were like that.] (Vincent)
3rd PL. Noun Phrase	Cause I know [some people don't.] (Liz)
neuter / pleonastic	Cause I know [there's a lot of politics involved.] (Vincent)

Material between COMP and subordinate clause

intervening clauses, fillers, parentheticals	I suspect maybe [that if you go through the French immersion,] (Ted)
argument	I'd explain it to them [that in Westmount it was still Dorchester.] (Donald)
pause	So I think [that: that: that really was the case.] (Jack)
none	I think [it's just north] (Jack)

Non-COMP *that* in COMP-position

none	I don't think [it's gonna happen again.] (Kurt)
yes	But I don't think [Ø that's gonna support four million people.] (Kurt)

Type of matrix clause verb

non-finite	I would like to believe [that it's possible.] (Liz)
finite	I don't think [he could function well.] (Peter)

Transitivity of subordinate clause

transitive	They think, you know, [we're gonna lose a distinct society and stuff.] (Kurt)
intransitive	I don't think [they really care] (Gloria)

Phonological factor
Phonology of segment following COMP

[+sonorant]
nasal	Cause I knew [that my French wasn't that good.] (Donald)
liquid	But I know [last year she was in Australia.] (Janie)
vowel	I figure [I'll respect that in class] (Ted)

[−sonorant]
affricate	So I figured [just be mov- easier to move elsewhere] (Glenda)
stop	Because I think [both groups called the other group Pepsi.] (Donald)
fricative	They were told [that the classes were full]. (Louisa)
sibilant	Cause I know [some people don't.] (Liz)

Semantic factors
Animacy of subject in matrix clause

human	They never think [I'm from another part of Quebec] (Jocelyn)
abstract	The problem is [they don't have the heart for it.] (Vincent)
pleonastic	It seemed [like there was a lot.] (Glenda)
no subject (e.g., gerund, infinitive)	That person's English too, thinking [I'm French.] (Kurt)

Animacy of subject in subordinate clause

human	I don't think [he could function well.] (Peter)
concrete inanimate	He'd tell me [that this bathroom floor wasn't washed]. (Mike)
abstract	I said [her vocabulary and everything was great] (Liz)

pleonastic   I think [there is a chance.] (Liz)
gerund, infinitive   I don't think [that uh separating's going to give them
                                        any more of an identity.] (Jocelyn)

Polarity

matrix positive, subord. positive   It just seems [they all have this attitude] (Ted)
matrix negative, subord. negative   My mother won't admit [she can't speak French.]
                                        (Donald)
matrix negative, subord: positive   But it doesn't mean [you know what it is.] (Donald)
matrix positive, subord: negative   I felt [that the boss wasn't too pleased] (Peter)

Co-referentiality

coreferential   I'm just glad [I learned both] (Ted)
non-coreferential   I realized [that French was all around me] (Peter)

Semantic class of the matrix verb

comment/factive   I'm just glad [I learned both] (Ted)
extraposition   It just seems [they all have this attitude] (Ted)
suasive   He kept insisting [that he was Trinidadian] (Mike)
knowledge   I knew [it was school.] (Janie)
utterance   You said [you spoke to him.] (Peter)
attitude   I would hate to think [that it was that]. (Louisa)

Appendix B: Examples for each factor code, in AMF

This section provides illustrative examples from our corpus for each factor in all of the linguistic factors considered. Factors are also defined, as necessary.

MATRIX VERB IDENTITY AND FREQUENCY

Matrix verb frequency

For the French data, we used Véronis (2000), which provides lemmatized frequencies for spoken French, based on Corpaix, the 2000 version. For a description of this corpus of spoken French, see Blanche-Benveniste (2000).

>1000 (in Corpaix)   Ça se peut [qu'elle parle anglais.] (Ted)
>100 but <1000   Je rappelle [quand j'étais toute petite tout en anglais.] (Janie)
<100   Je perçois facilement [que la personne est anglophone.] (Tony)

Matrix verb identity

sembler   Il me semble [que toutes ces écoles-là leur programme français
                       c'est pas adéquat.] (Louisa)
souvenir   Je me souviens [que ça avait été bizarre une expérience vraiment
                       étrange.] (Liz)

dire	On dit [qu'il y a beaucoup de drogue.] (Jack)
penser	Je pense pas [que je vais finir le programme.] (Janie)
croire	Je crois [que ça avait affaire avec la loi.] (Joanie)
trouver	Moi je trouve [c'est une question de culture.] (Ted)
savoir	Durant le deuxième terme je savais [c'était pas pour moi.] (Joanie)
falloir	Tu sais faut [que je pense avant de parler.] (Jack)
voir	Alors j'ai vu [que c'est inutilisable en Quebec.] (Jack)
être + X	C'est juste bon [que le monde là-bas chu : c'est les Blocks là les vrais.] (Ted)
Other verbs (1–4)	Ma mère exige [que je travaille.] (Ted)
Other verbs (5–15)	Je remarque [qu'il y en a beaucoup plus.] (Janie)

SYNTACTIC FACTORS

Subject type (matrix clause)

1st SG.	Je ne crois pas [qu'il y a discrimination dans l'emploi.] (Jack)
2nd SG.	Tu veux [que je dise qu'est-ce qu'il a fait?] (Kurt)
3rd SG.	Il faut [que je sois capable de réfléchir comme lui.] (Liz)
3rd sing + 1st PL. *on*	On dirait [que tu es comme ça tout le temps.] (Ted)
2nd PL.	Vous saviez [que j'étais un anglais.] (Ted)
3rd PL.	Ils sauraient [qu'elle aurait de la misère.] (Ted)
Indefinite	C'est pas [qu'ils ont pris une décision.] (Kurt)
No subject	Étant donné [tu peux pas répondre à la demande de tous.] (Vincent)
Impersonal	Personne dans mon famille peut comprendre rien [qu'il a dit.] (Donald)
Noun Phrase	Ma femme a décidé [que il y a beaucoup de monde.] (Donald)

Subject type (subordinate clause)

1st SG.	Je trouve [que je suis plus fort pour faire les courses.] (Donald)
2nd SG.	Je dirais [quand tu fais des recherches sur des costumes.] (Sandra)
3rd SG. m.	Vu [qu'il avait commencé dans l'école anglophone.] (Sandra)
3rd SG. f.	Elle a vu [qu'elle était très responsable.] (Vincent)
3rd sing + 1st PL. on	Mes parents ont voulu [qu'on aille à l'école en français.] (Sandra)
3rd PL.	Ces gens je pense [qu'ils ont trouvé.] (Peter)
Indefinite	Je pense pas [que c'était tant que ça.] (Sandra)
No subject	Je pense [que oui.] (Kurt)
Impersonal	Je pense [que la personne avec qui ils ont fait des enfants.] (Tony)
Noun Phrase	J'ai trouvé [que l'allemand c'était plus facile d'apprendre.] (Peter)

Material between COMP and subordinate clause

Intervening material	Ca se peut très facilement que la première fois [que je parle anglais.] (Louisa)
No material	Je le sens [que je danse pas pareil comme eux.] (Vincent)

Phonology of segment preceding COMP

Vowel	C'est pas [qu'ils vont tuer.] (Vincent)
Liquid	Je rappelle [quand j'étais petite tout en anglais.] (Janie)
Fricative	Moi je trouve [c'est ça c' est les jeunes.] (Ted)
Stop	Je remarque [qu'il y en a beaucoup plus.] (Janie)
Sibilant	Je pense [la plupart sont francophones.] (Jack)

Phonology of segment following COMP

Vowel	J'ai remarqué [on a des plus par exemple.] (Vincent)
Liquid	Je trouve [que la famille noire est plus solidaire.] (Vincent)
Fricative	Je trouve [faut le faire.] (Joan)
Stop	Je pense [que quand j'avais : quand on m'a donné le char il y avait uncouvre-feu.] (Jack)
Sibilant	Je savais [que c'était gagné avec elle.] (Vincent)

Polarity

Positive main clause	Il faut [qu'il travaille dans deux langues.] (Mike)
Negative main clause	Faut pas dire [que les bons professeurs ils font ça.] (Ted)

Mysteries of the substrate

William Labov
University of Pennsylvania

There can be little doubt about the existence of substrate effects in many cases when a whole population abandons their original language and adopts another. But there are situations in which the direction of linguistic influence remains unexplained, the causal connections are obscure, or the expected effect does not occur. We still do not know just how young children in American society manage not to acquire the foreign accent of their parents. If anything, the effect of parents' language may be in the opposite direction from that predicted by contrastive analysis. Several cases of unmistakeable but inexplicable substrate effects are discussed: the initiation of the merger of /o/ and /oh/ by Slavic-speaking coal miners in Eastern Pennsylvania; the use of *later* for *earlier* in the English of Puerto Rican Spanish speakers, and the confusion of *make* and *let* among several generations of Italian-American speakers of English.

Keywords: substrate effects; foreign accent; /o/ and /oh/ merger; Pennsylvania English; Puerto Rican Spanish, Italian-American English

1. Introduction

The term "substrate" conjures up a range of differing opinions and disputes about the effects of language contact and how the use of one language affects another. In many cases there can be little doubt about the existence of substrate effects when a whole population abandons their original language and adopts another. Hiberno-English is such an example, where even the most naïve and amateur observer can detect substratal effects from Irish Gaelic on current-day English spoken in Ireland. For example, Hiberno-English shows

– *After* as a marker of recent, based on the Irish use of the preposition *tar éis,* *'after',* as in *I was just after getting off a truck.*
– Replies to yes-no questions with reduplicated subject and auxiliary, as *Are you coming home soon? I am.* Irish has no words for "yes" or "no" but expresses agreement and disagreement in this manner.

– The consistent use of dental stops for interdentals, as in the story about the schoolteacher from Cork who said, *Today we're going to study English pronunciation. And we're not going to say [diz] and [dem] and [doz]. We're going to say [diz] and [dem] and [doz]!*

These and many other powerful substrate effects have been incorporated into our general knowledge of languages in contact (Weinreich 1968; Thomason & Kaufman 1988). Sankoff and Brown's (1976, 1980) account of the marker *ia* in Tok Pisin traces its evolution to a relative clause bracket as an internal process of grammaticalization, but then points out that much of this development is parallel to the patterns found in substrate Austronesian languages. They observe that adult second language speakers of Tok Pisin show *ia* bracketing which they could not have learned from their children.

One can speak of a personal substrate, within one speaker, where the effect of L1 on L2 can be charted, but more generally we are speaking of a community effect – that is, a change in the language that is the result of a very large number of bilingual speakers transferring some part of the L2 effect to large numbers of their descendants, speakers of L1.

What then is problematic about substrate influence? In the field of Creole Linguistics, substrate arguments may be opposed to arguments from general historical processes and even innate mechanisms, and this field is known for its sharp disagreements (Bickerton 1981). Substrate arguments in dialectology are equally open to dispute. The explanation of fronting of /u/ and /o/ in French as the effect of a Gaulish substrate may be countered by geographic mismatch of the region occupied by the Gauls and the area of fronting.[1] In this brief treatment, I would like to focus upon unresolved issues of substrate influence that have arisen in the course of sociolinguistic studies of the speech community. They are of two kinds: the absence of substrate effects where we would most expect them to appear, and the explanation of effects that do appear but are not clearly motivated by the structure of the substrate language.

2.   The minimal effect of parental substrate: Children's ability to avoid a foreign accent

Sociolinguistic studies of English speech communities in North America have found extensive linguistic variation conditioned by age, gender, social class and

---

1. See however Weinreich (1963) on how successive stages of migration can lead to such mismatches, and why in general we should not expect substrate influence to coincide geographically with co-territorial areas.

social networks. But ethnicity – Italian, Jewish, Irish, Polish, German family background has not appeared as a major factor.[2] Given the fact that the ethnicity of speakers is the trait most closely related to language, it is quite surprising to find very little differentiation of the English of the population by ethnicity. Of all the social factors examined in sociolinguistic studies of New York (Labov 1966 [2006], Detroit (Shuy, Wolfram & Riley 1967), Philadelphia (Labov 2001), Ottawa (Woods 1979) and the Northern Cities (Labov, Ash & Boberg 2006), the ethnic group of the speaker's family and knowledge of the immigrant language has the least effect.

This minimal effect of ethnicity is consistent with the more general observation that at least in North America, children do not acquire the foreign accent of their parents. No detailed studies have as yet been carried out of the effects of mothers' non-native patterns on children's first stages of language learning. The study of the new dialect of Milton Keynes (Kerswill & Williams 1994; Kerswill 1996) showed that children did show the influence of their parents' dialect at 4 years old. They did not in general acquire the emerging Milton Keynes dialect until 8 or 12 years old. However, no such effect of parents' non-native accent has been as yet recorded, and linguists have reported individually many cases of the reverse (Chambers 2002). In the study of New York City (Labov 1966) a substantial portion of the native speakers were second generation children of parents who spoke Italian, Yiddish and other languages, but no trace of a foreign accent was found for any of those speakers.

We are then left with the task of explaining how children detect and avoid the non-native patterns of their first-generation parents. Chambers (2002) on "The Ethan experience" points out that it is accompanied by an even more surprising phenomenon: that second generation speakers avoid their parents' foreign accent without being aware that they have one. Of all substrate mysteries, this remains the most profound.

3.   Reverse ethnic effects

The preceding discussion should not be taken to indicate that no effects of ethnicity have been observed. The rest of this report will deal with a wide range of linguistic features that are associated with the ethnic background of native speakers of English. In New York City, small but consistent effects were observed. However, what is relevant here is that they were not a replication of the parents'

---

2.   Ethnic identity is here considered distinct from the racial categories African-American, Latino and Asian, which have even greater influence on language behavior than social class, age and gender.

foreign accent, but were in a direction quite different from what would have been predicted from the structure of the immigrant language.

In New York City, all ethnic groups participated in the raising of /æh/ and /oh/ from low to mid to high. Italians showed distinctly higher values of (æh) and Jews favored the raising of (oh) (Labov 1966). This is not what one would have predicted if the second generation (that is, the first native generation) had carried the vowel system of Italian or Yiddish into their English. None of the Yiddish dialects of Eastern Europe have a high back ingliding vowel [oᵊ]; a first generation Yiddish accent in English shows a low back [ɑ] for both /o/ and /oh/ and [ɛ] for /æ/. The Italian dialects of southern Italy do not show a high ingliding vowel [eᵊ] or [iᵊ], and a first generation Italian accent shows [a] for [æ]. This suggested that ethnic influences might generally be the obverse of direct influence, triggered by the general desire to avoid the features of the foreign accent of the speaker's parents.

If such a hypercorrect pattern exists, it is not the immediate result of language contact, but of a more general factor that persists over time as the knowledge of the foreign language diminishes. If it were a result of language contact, one would predict that the advantage would be stronger in the second (first native) generation, and weaker in the third. Table 1 (from Labov 1976) shows such a comparison for two closely matched sets in New York City: upper middle class Jewish younger men, and working class Jewish older women. The scores for the 2nd and 3rd generation do not differ significantly at any point, and for the most part are very close indeed.[3]

Table 1. Phonological variables for subjects with foreign- and native-born parents in New York City

	Upper middle class Jewish Younger men [21–39 yrs]				Working class Jewish Older Women [40–65 yrs]			
	(æh) scores		(oh) scores		(æh) scores		(oh) scores	
	2nd Gen.	3rd Gen.	2nd Gen.	3rd Gen.	2nd Gen.	3rd Gen.	2nd Gen.	3rd Gen.
N	3	6	3	6	6	3	6	3
Mean	31.0	30.3	27.0	24.5	28.6	28.6	19.9	20.6
Std dev	7.8	5.8	1.6	4.8	3.9	10.0	3.5	1.1
t	0.15		0.86		0.47		0.33	

---

3.   A reviewer of this paper reminds me that in some cases the third generation can show more L1 influence than the second, as children return to the patterns of their grandparents (Dubois & Horvath 1998).

This suggestion of ethnic hypercorrection as the basic mechanism is reinforced by the findings of Laferrière (1979) in Boston, a study which focuses primarily on the sociolinguistic patterns of ethnicity. The main linguistic variable in question is the low back vowel [ɔ] of *for, morning, short, fork, or,* which represents the phoneme /ɔhr/ in dialects that distinguish them from the vowel of *four, mourning, port, ore,* etc., pronounced with an upper mid ingliding vowel. Because the low back vowel is stigmatized as stereotypical of the Boston dialect, many speakers show discrete oscillation between the use of /ɔhr/ and /ohr/. Figure 1 shows the percent use of the low back vowel by three ethnic groups. Just as in New York, the Italian speakers are here opposed to the Jewish group. One interpretation is that Jews are here more sensitive to the social impression conveyed by Boston /ɔhr/, or perhaps less inclined to identify themselves as Bostonians. But there is also a linguistically motivated generalization: just as in New York, Jews of the second and third generation favor higher back ingliding vowels.
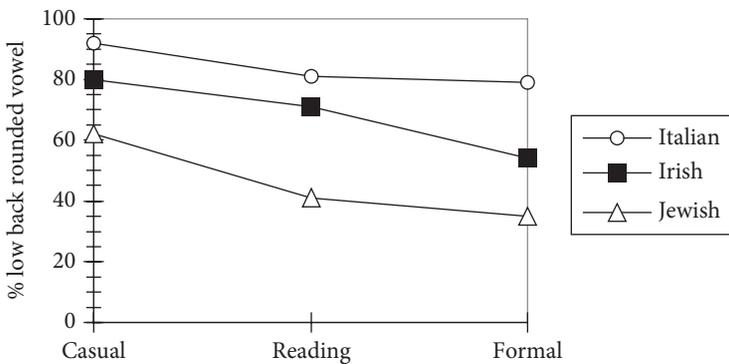
Figure 1. Percent low back rounded [ɔ] in /ɔhr/ class in Boston by ethnic group (from Laferrrière 1979).

A parallel preference for raised /oh/ was found by Knack (1991) in her study of ethnic differences in the Northern Cities Shift of Grand Rapids, Michigan. Here in a very different linguistic and sociolinguistic situation, Jews also showed higher realization of /oh/. Table 2 shows that in a regression analysis of the raising of /oh/, 49% of the variance is accounted for by the effect of Jewish ethnicity.

Table 2. Regression analysis of the effect of ethnicity on the raising of /oh/ in Grand Rapids, Michigan (Knack 1991)

	Coefficient	Probability
Jewish women	0.47	≤ 0.0001
Jewish men	0.27	0.0044
r^2 (adjusted)		49

That is not to say that /oh/ is always higher for Jewish speakers. Where there is no contrast of /o/ and /oh/, and none for /ohr/ and /ɔhr/, no such effect has been found.

4.   Vocalization of /r/ by Italians in an r-pronouncing dialect

Philadelphia is an *r*-pronouncing dialect. Except for certain dissimilating words like *quarter, ordinary, forward* and *corner*, white Philadelphians consistently pronounce postvocalic /r/ as a constricted central consonant. However, on any day in the year, one may hear a very general pattern of vocalized /r/ from the men who run the stalls and shops at the 9th Street Market, an open-air produce market spread across several blocks in the Italian area of South Philadelphia. The first impression is that r-lessness is concentrated in the South Philadelphia Italian community. This was confirmed in our interviews with 60 residents in the study of Linguistic Change and Variation in the 1970s. For those whose parents were both Italian, 75% showed some r-vocalization. Other members of the same community displayed only 39% (p = .009 by Fischer's exact test). In the past few decades, rapid and anonymous studies of the pronunciation of "Market" and "Girard" have consistently shown significant concentrations of r-lessness in the Italian areas.

Is there any linguistic connection between Italian language background and the vocalization of (r)? There is no direct copying here: the southern Italian and Sicilian dialects spoken by the immigrants to South Philadelphia have a strong, apico-alveolar trilled /r/ which does not show the variability in question.[4] Italian /r/ is clearly a consonantal liquid, with a [+consonantal] feature of apical tongue contact that is absent in English /r/. It is possible that the vocalization of /r/ originated in the Italian community when new speakers of English with an Italian phonological system identified English postvocalic /r/ as a glide rather than a consonant, and produced a vocalic glide as their nearest phonological equivalent. If this is indeed the origin of (r) vocalization, there should be some relationship between this variable and knowledge of Italian. In our study of South Philadelphia, speakers' knowledge of a non-native language was rated on six levels:

0   no foreign language background
1   passive understanding of grandparents
2   passive understanding of parents
3   spoke only foreign language up to school age, never much since

---

4.   In the vast majority of cases, Italian /r/ occurs in intervocalic position, but there are a few words like *per* where it occurs finally.

4   occasional use of foreign language with older people
5   regular use of foreign language with older people
6   dominant in foreign language.

Table 3 shows the distribution of (r) vocalization and knowledge of Italian for the 35 speakers whose primary ethnicity was Italian. Despite the small numbers, it is immediately apparent that there is no correlation. It is clear that (r) vocalization is not decreasing with decreasing knowledge of Italian.

Table 3. Vocalization of (r) by knowledge of Italian for South Philadelphians whose parents were both Italian

	Foreign language knowledge							
	0	*1*	*2*	*3*	*4*	*5*	*6*	*Total*
No (r–0)	5	2	1	2	2	0	0	12
Some (r–0)	9	4	2	3	3	1	1	23
Total	14	6	3	5	5	1	1	35

It is has been suggested that r-lessness in Philadelphia is the result of contact with the r-less community of New York City, and closer studies of communication patterns may support such an idea. But I have also noted r-less speech on the docks of San Francisco, in a strongly r-ful environment. An explanation of the mechanism of (r) vocalization in the Italian community must still be linked to some aspect of the Italian language, and that remains to be done.

5.   The Slavic effect on the merger of /o~oh/

Perhaps the most striking example of ethnic influence appears in the sudden expansion of the merger of /oh/ and /o/ in Eastern Pennsylvania, documented by Herold (1990, 1997). Herold demonstrated that this sound change was tightly focused in coal-mining towns, where the population had undergone a sudden reversal of ethnic composition with the immigration of large numbers of Slavic-speaking miners from eastern Europe. However, she was unable to find any mechanism that would connect the languages spoken by the immigrants with this merger. Pittsburgh exhibited the same merger at an earlier period, perhaps associated with the sizeable Polish migration to that city beginning in 1885. Polish has a contrast of a lower mid back rounded and low back unrounded vowel, while the Pittsburgh merger may well be an inheritance of the original Scots-Irish population. No convincing connection has yet been found between a Slavic substratum and the merger of /o/ and /oh/.

6.   The Puerto Rican use of *later*

During the mid 1980's, the research project on the relation of urban minorities to linguistic change examined the contrast of linguistic forms across the Euro-American, Hispanic and African American communities in North Philadelphia. In the course of this work, Wendell Harris as a participant observer noted a number of utterances in the English of Puerto Ricans who had intimate connection with blacks, which reversed the usual use of *later* and *earlier*. He recorded many examples of people saying things like, "I was over your house later, but you weren't there," where the intended meaning was clearly "earlier." Since that time he has continued to note examples of this phenomenon, which is widely current in the English of this section of the Puerto Rican community, but has never been noted from any other speakers. I have consulted with many linguists with an extensive knowledge of this dialect of Spanish and others, but we have been unable to find any contrast between Spanish *tarde* and English *later* that would account for this effect.

7.   The Italian-American confusion of *let* and *make*

In 1978, members of the Penn course on The Study of the Speech Community, working in South Philadelphia, observed a series of utterances that suggested a reversal or neutralization of the distinction between *make* and *let*.

(1)   A woman's husband was observed making over-friendly approaches to a number of other women at a local dance. A friend asked her, "Why do you make your husband do that? You make your husband go … he dances with everybody."

(2)   In a story about a baby brother who swallowed kerosene while the narrator and her sister were supposed to be watching him: "It was my sister's fault and my fault, because we made it happen."

(3)   In the course of a narrative, a man said, "These guys never went to college, and what they told, it would let Jesus shrink."

Sentences (1) and (2) are from Kate G., a 56-year-old housewife and native speaker of English with some understanding of Italian, and (3) is from Jimmy S., a 53-year-old committeeman and native speaker of both English and Italian. More recently, Gillian Sankoff observed another such speech production from a South Philadelphia Italian woman who is a close friend of our family. She was discussing a local restaurant which had added a dance lounge without getting a cabaret license, and was maintaining it in the face of community protests. She said,

(4) "Who does he know at L&I [Licenses & Inspections] who is making
     him keep it open?"

A handful of sentences over many years of observation is not enough to come
to a decisive conclusion about what can be said and who says it. A comparative
inquiry into intuitions was in order. Hoekje et al. (1978) report a field experiment
constructed by members of the class, in which 25 members of the neighborhood
they were studying were asked to give grammaticality judgments for eight sen-
tences containing *make* and *let* in English. The non-Italians agreed in 149 out of
152 judgments, while the subjects of Italian background agreed in only 75 out of
102 judgments on whether to use *make* or *let*.

   This is an extraordinary difference. We do not at present know of any se-
mantic differences in Italian *lasciare* and *fare* and English *make* and *let* that would
account for this ethnic effect.

   In 2006, Marielle Lerner undertook a more extensive investigation of intu-
itions on the choice of *make* and *let*, comparing Italians and non-Italians in South
Philadelphia, Boston's North End, and Court Street in Brooklyn. Building on the
work of Hoekje et al., she constructed a questionnaire with 12 sentences. Lerner
found that the tendency to give non-standard responses was significantly corre-
lated with lower education. In her data, it was also correlated with knowledge of
Italian. Table 4 summarizes responses based on ethnicity.

Table 4. Responses to 12-sentence questionnaire on choice of *make* and *let* (Lerner 2006)

No. non-standard responses	Italian ethnicity	Other ethnicity
0	7	15
1	7	4
2	3	0
3	2	0
4	4	0
5	1	0
6	1	0

   From these results, there can be no doubt that confusion of *make* and *let* is as-
sociated with Italian ethnicity, and by every logic we can summon up, would find
its roots in a structural contrast between Italian and English in the semantics and
syntax of these verbs. Are sentences (1–4) calques of Italian?

   Hoekje et al. (1978) did note the following Italian constructions, which show
differences in the use of these verbs:

(5) *Vado   a   fare   mangiare   mio   marito.*
     I   go   to-make   eat   my   husband.
     'I'm going to fix my husband dinner'

(6)   *Fammi vedere.*
      Make me see
      'Let me see.'

While the Standard English causative "make" always corresponds to causative "fare" in Standard Italian, (5) and (6) are instances when the converse does not apply. Sentences of type (5) were observed in South Philadelphia (*I'm going now to make my husband eat*), but not (6).

Lerner (2006) attempted to confirm the calque hypothesis directly by asking Italian speakers. She asked three native speakers of Italian to translate sentence (1) into Italian. They all used *lasciare,* as in (7).

(7)   Why did you let your husband dance with other women?
      → *Perchè lascia suo marito ballare con altre donne?*

All used *lasciare* to translate the meaning of (2), as in (8):

(8)   We got into trouble because we let it happen.
      → *Siamo finite nei guai perche' abbiamo lasciato che questo accadesse.*

Finally, the three Italian native speakers all used *fare* to translate sentences of type (3). Lerner also found in a Google search of Italian websites that *fare piangere Gesù* was the norm and that *lasciare* was not found in this construction.

Lerner (2006) searches many other avenues of explanation for the *make/let* confusion among Italian-Americans. Since most of the Italian residents in America come from southern Italy, she finds it interesting to note that in the Neapolitan dialect, the construction [*lasciare* + determiner phrase + infinitive] is not possible, though it is in Italian (De Silvio 1915). Lerner continues:

> The alternatives to expressing this sort of construction in Italian, which translates to ["let" + determiner phrase + infinitive verb] in English, would be either to use the Italian causative "fare" or to re-phrase with ["lasciare che" + determinative phrase + present subjunctive verb]. However, in Neapolitan this alternative is not possible in the present tense as there is no present subjunctive verbal form (Ledgeway 1996: 202). The same is true in Sicilian (Leone 1995: 33). These differences do not provide for any more definitive correlation with the non-standard "make" and "let" of Italian-Americans, but rather they further complicate the nature of the correspondence between English and Italian causative constructions.

In short, the *make/let* substitutions are a substrate effect, and remain a mystery for linguistic analysis to solve.

8.   Where should we turn?

Among the many mysteries of substratal effects, a few have been clarified in recent years. We note that speakers of English of Hispanic background differ from others

in the possibility of simplifying clusters with /rd/ and /rt/ (Santa Ana 1991). The percentage is low, but no other group has been observed to do this. *We had a good car' game* is a possible sentence with Latino native speakers of English. My own account of this refers to the asymmetry of production and perception. These second generation speakers produce fully American /r/, but it is possible that their underlying phonological organization shows /r/ as [+consonantal] as opposed to the [–consonantal] glide that is generally characteristic of English. This glide does not trigger consonant cluster simplification for most speakers of English, but it may do so for Latinos.

A similar appeal to differences between production and perception would account for the Latino pronunciation, sometimes heard in New York city, of *Puerto Rican* as [por ərikən], with two American humped /r/ glides. The native New York City r-less pronunciation is [poɾərikən], with a flap and no preceding /r/. Since this flap is identical with the native Spanish /r/, it is possible that the word is re-analyzed as /porərikən/ and produced as such by English rules.

Such a reference to the asymmetry of production and perception might apply to Italian (r) vocalization. Given the consonantal status of Italian /r/, the English /r/ may not be recognized as a consonant at all in syllable-final position but here the situation is much less clear since there are only a few Italian words with /r/ codas.

We cannot doubt that there are powerful substrate effects, but many of them – more than I have touched on here – remain as challenges to our linguistic understanding.

References

Bickerton, D. 1981. *Roots of Language.* Ann Arbor MI: Karoma Press.

Chambers, J. 2002. Dynamics of dialect convergence. In *Investigating Change and Variation through Dialect Contact*, L. Milroy (Ed.), Special issue of *Sociolinguistics* 6: 117–130.

De Silvio, A. 1915. Studies in the Dialect of Basilicata. *PMLA* 30(4): 806.

Dubois, S. & Horvath, B. 1998. Let's tink about dat: Interdental fricatives in Cajun English. *Language Variation and Change* 10: 245–261.

Herold, R. 1990. Mechanisms of Merger: The implementation and distribution of the low back merger in Eastern Pennsylvania. Ph.D. dissertation, University of Pennsylvania.

Herold, R. 1997. Solving the actuation problem: Merger and immigration in eastern Pennsylvania. *Language Variation and Change* 9: 165–189.

Hoekje, B., Nishimura, M. & Nudelman, P. 1978. Bell and Weaver Streets. Final Report for Linguistics 560, The Study of the Speech Community.

Kerswill, P. 1996. Children, adolescents, and language change. *Language Variation and Change* 8: 177–202.

Kerswill, P. & Williams, A. 1994. A new dialect in a new city: Children's and adults' speech in Milton Keynes. Final report to Economic and Social Research Council.

Knack, R. 1991. Ethnic boundaries in linguistic variation. In *New Ways of Analyzing Sound Change,* P. Eckert (Ed.), 251–272. New York NY: Academic Press.

Labov, W. 1966. *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

Labov, W. 1976. The relative influence of family and peers on the learning of language. In *Aspetti Sociolinguistichi. dell' Italia Contemporanea.* R. Simone et al. (Eds), Rome: Bulzoni.

Labov, W. 2001. *Principles of Linguistic Change: Social factors.* Oxford: Blackwell.

Labov, W., Ash, S. & Boberg, C. 2006. *Atlas of North American English: Phonology and sound change*. Berlin: Mouton de Gruyter.

Laferriere, M. 1979. Ethnicity in phonological variation and change. *Language* 55: 603–617.

Ledgeway, A. 1996. The Grammar of Complementation in Neapolitan. Ph.D. dissertation, University of Manchester.

Leone, A. 1995. *Profilo di Sintassi Siciliana.* Palermo: Centro di Studi Filologici e Linguistici Siciliani.

Lerner, M. 2006. Causatives in the English of Italian-Americans. Undergraduates honors thesis, University of Pennsylvania.

Sankoff, G. & Brown, P. 1976 [1980]. The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language* 52(3): 631–666. (Reprinted in G. Sankoff (Ed.), *The Social Life of Language* 211–255. Philadelphia PA: University of Pennsylvania Press).

Santa Ana, O. 1991. Phonetic simplification processes in the English of the Barrio: A Cross-Generational Sociolinguistic Study of the Chicanos of Los Angeles. Ph.D. dissertation, University of Pennsylvania.

Shuy, R., Wolfram, W. & Riley, W. 1967. A study of social dialects in Detroit. *Final Report, Project* 6–1347. Washington DC: Office of Education.

Thomason, S. & Kaufman, T. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.

Weinreich, U. 1963. *Four Riddles in Bilingual Dialectology*. The Hague: Mouton. Weinreich, U. 1968. *Languages in Contact: Findings and problems*. The Hague: Mouton. (Originally published as Publications of the Linguistic Circle of New York, no. 1, 1953).

Woods, H. 1979. A Socio-dialectology Survey of the English Spoken in Ottawa: A study of sociological and stylistic variation in Canadian English. Ph.D. dissertation, University of British Columbia.

Empirical problems with domain-based notions of "simple"

Miriam Meyerhoff
University of Edinburgh

This chapter addresses the on-going debate about the relative "simplicity" of creole languages. It proposes that an evaluation of simplicity/complexity must consider not only categorical features of a language but also probabilistic ones, because (it argues) there is a good deal of linguistic structure encoded stochastically in creoles. To illustrate this, it explores four case studies: the marking of inalienable possession in Bislama (Vanuatu), subject agreement in Bislama, possessive marking in Tayo (New Caledonia), and the recent emergence of a new complementiser in Bislama. Substrate, lexifier and cognitive constraints contribute to the emerging shape of all four features. The data argues for perspectives on creolisation that include non-deterministic features, and for a view of language structure straddling what are sometimes seen as discrete levels of linguistic structure.

Keywords: Vanuatu; Bislama; complementisers; *se* 'say' COMP; *olsem* 'like' COMP; language attitudes; substrate influence; simplicity; direct/indirect possession; alienable/inalienable possession; subject marking; Tayo; possessives

1. Introduction

When creolists talk about the relative *simplicity* and *complexity* of different languages, there is a lot riding on a reliable definition of these terms. It has been argued that, without any reliable way to differentiate creole languages from other natural languages, creolists are essentially out of a job. Or less dramatically, that without a means of identifying creoles as distinct from other languages, there is no cohesion to the debates that characterise the field of creolistics and no clear path of linguistic inquiry underlying what "creolists" (whoever they might be) actually do.

One particularly influential position has, for many years, been that that creole languages are a recognisable class of languages among all natural languages and that they are unified by their structure (Bickerton 1984; McWhorter 2001; Parkvall

2001).[1] That is to say, creoles are a typologically distinct class of languages identifiable by their reduced inflectional morphology and the use of fewer (or less) complex syntactic structures than non-creole languages do. This position has been accepted uncritically by some mainstream linguistics (e.g., the comparison between signed languages and creoles in Aronoff et al. 2005) despite the widespread critique of it within creolistics. DeGraff (2001, 2003) has critiqued this generalisation on empirical grounds; others have suggested that it arises from an undue focus on isolated components of the grammar (see papers in Byrne & Winford 1993; Arends 2001); and still others have argued that this position is only tenable if analysts engage in the (arguably, inappropriate) comparison of creoles with lexifiers, rather than with substrates (Brousseau et al. 1989; Lefebvre 2001; Ansaldo & Matthews 2001). Finally, it has been suggested, even if this claim is true, structural simplicity arises from the social/sociolinguistic isolation of creole-speaking communities rather than linguistic factors (Maher 1984). All these responses to the simplicity argument problematise the status of creoles as a typological class on purely linguistic grounds.

Another view of creoles is to focus squarely on them as the products of socially and sociolinguistically complex speech communities, and to explore them as languages in their own right, in their particular sociolinguistic context. Gillian Sankoff has been pre-eminent among the exponents of this approach and all variationist work on creole languages (and a good deal of other work on language change) has been greatly indebted to the work she has done on the language dynamics of the multilingual speech communities in Papua New Guinea (and in more recent work, other parts of Melanesia).

A practical question arises from this: what does it mean for a linguist to take the sociolinguistic complexity of a creole as the starting point and focus of their inquiry? In Sankoff's case, it has meant approaching creoles with this same kind of empirical rigour that variationist sociolinguists have employed to study monolingual and multilingual speech communities anywhere. I believe this approach affords us a very different view of complexity, and I believe it is a view that is useful for several reasons. First, the data gathered via this approach may help put to rest some of the theory-internal debates about complexity and simplicity in creolistics.

---

1.   Pidgin languages are specifically identified as the only "real primitive languages" in Bauer et al. (2006). They then note that the process of creolisation occurs when child speakers "start adding in a lot of the grammatical complexity that is missing in the pidgin" (2006: 98). Such characterisations of the relationship between pidgins and creoles tend, I think, to perpetuate the idea that creoles do have simpler grammars than other languages, and also to suggest that creoles somehow sit on a developmental cusp between "primitive languages" and whatever non-creole languages are.

Second, this approach provides data that positions the study of creole languages squarely within the general field of language variation and change, moving creole languages out of their position as a side-bar to the discussion of "real" languages, where they are often examined solely in the contexts of theory-internal debates about the genesis and development of natural languages. Whether or not creoles turn out to be identifiable as a natural class of languages on purely structural grounds may prove to be far less interesting or important than how the study of contact languages can inform us about some basic linguistic questions.

2.   Some background issues with "simple" and "complex"[2]

Let us start by back-tracking slightly in order to reconsider the question of creole simplicity/complexity. In all the debates mentioned above, simplicity and complexity have been conceptualised almost entirely in terms of morphological and syntactic complexity (see discussions and summaries in Arends 1995; Mühlhäusler 1997; Sebba 1997; Aboh & Ansaldo 2007). There are some attempts to extend this to phonology (asking, for instance, whether creoles have smaller/simpler phonemic inventories than other languages or their lexifiers, e.g., Klein 2006), but the discussion has seldom, with the recent exception of innovative experimental work by Gil (2007), drifted into the much murkier domain of semantics. The reason for this is obvious: talking about simpler or more complex semantic representations is a very problematic thing to do.

The fundamental problem is that we don't know what semantics (or let us say, the language of the mind) looks like. In the face of this, linguists tend to turn to syntax, on the assumption that syntax is how and where semantics are actualised in natural languages. But any measure of the simplicity or complexity of the syntax-semantics mapping is necessarily theory-dependent (Bach 1964; Chomsky 1965) because there are no *a priori* measures of simplicity (Clark 2001; Brighton 2003). (An analogous point can be made about "similarity" – determining whether a creole is more similar to, say, a lexifier or substrate language is problematic

because similarity measures are theory-dependent too, cf. Kihm 1995; Kouwen-berg 2001; Siegel 2004).

A highly-localised and domain-specific definition of complexity and sim-plicity underlies many linguists' use of the word. In Huddleston & Pullum (2002), *complex* usually describes a property internal to a word or internal to a phrase. And Hauser et al. (2002) identify recursion in syntax as the key form of complexity that is unique to human languages. Hudson's (2000) paper on the un-acceptability of **I amn't* also defines simplicity/complexity within a specific lin-guistic domain. He defines *simple* like this, "a [morphological] paradigm without a gap is presumably simpler than one with a gap, so one might expect learners to take the very easy step of simplifying this paradigm" (2000: 298), where "simplifying" would involve eliminating the gap, or regularising the paradigm in some other way.[3] These three examples, drawn from theoretically very different approaches to language, demonstrate that defining complexity/simplicity in relation to different linguistic domains or modules of the grammar is fundamental and pervasive. It is not restricted to one theoretical model of language structure.

Moreover, note that most linguists operate with fundamentally modular theories of language, that is, the properties and inventory of a language's pho-nology, morphology, syntax (and possibly its discourse[4]) are modelled as being largely independent of each other. If we want to compare or measure correla-tions across modules, we need a theory of language to define the parameters (Muysken 1988 makes this point and exemplifies it in relation to serial verb constructions).

2.1   Determining complexity

Brighton (2003) provides an extensive discussion of different ways for measuring complexity. He reviews the usefulness of Occam's razor and Kolmogorov com-plexity as the basis for comparing the relative complexity of different languages, and he discusses the use of measures such as the Minimal Description Length (MDL) of an utterance and Bayesian probabilities in computational linguistics. I focus here on MDL because it is closest to the intuitive notion of complexity that

---

3.   Naomi Nagy (p.c.) notes that an entirely irregular list (i.e., one with maximal gaps, in Hudson's sense) isn't a paradigm, but this doesn't mean they are necessarily analysed as more complex. She offers the example of English which, lacking a rich array of aspect markers, uses adverbs such as "usually" or "repeatedly" instead of a different verb conjugation. It's not clear whether English is therefore more complex than an inflectionally, aspect-rich language.

4.   Linguists tend to split radically in whether or how far they consider probabilistic features typical of discourse to be "really" linguistics.

most debates about creole languages revolve around. Essentially, the MDL of an utterance describes how many rules or descriptors are required for a computational system to get from nothing to the target sentence/utterance. Exceptionless rules will result in shorter "grammars" for a sentence; allomorphy and syntactic movement will require more complicated rules, increasing the MDL for a given utterance. Arguments for the relative simplicity of creoles that focus on the lack of, for example, person and number agreement on verbs, or case marking on nouns, or noun classes (genders) are essentially arguments that the MDL in that domain of a creole grammar is shorter than the MDL in that domain of a lexifier or substrate language. This is a production-centred notion of simplicity, of course, because for the hearer, this reduction of information creates an interpretative burden (cf. Gil 2007).

There are problems with using MDL as the basis for arguing that creoles are simpler than other languages. First, MDL alone proves to be unsuccessful at grammar induction in various simulations (Clark 2001). Computational systems that are successful at inducing grammar achieve their success by pairing meanings and structural representations (presumably what children do when they learn a language, a point also made by Prince 2001). Pairing an "irregular" form with a meaning is inherently no more complex than pairing a "regular" one with an underlying meaning. Each involves a one-to-one mapping between form and meaning. If grammar is induced from form-meaning pairs, this means that from a computational perspective, grammar induction of morphologically rich languages and morphologically poor languages is no different in complexity.

An additional problem with using MDL or similar measures as the basis for comparing the relative complexity of languages is that they cannot tell us how to evaluate compensatory strategies. Simplification of a paradigm doesn't necessarily entail greater simplicity in a system as a whole. A good illustration of this is the cross-linguistically well-attested relationship between inflectional morphology and word order: the loss of inflectional morphology often proceeds hand in hand with greater word order rigidity. A focus on less-ness in creolistics may be tendentious. It may distract us from compensatory complexity that emerges in other domains of the (socio)linguistic system.[5]

---

5.   A discussion of simplicity and complexity might go well beyond modular theories of language to consider the question of whether or not interpersonal accommodation adds complexity to an interaction or reduces it by eliminating uncertainty about the interlocutor (Meyerhoff 2006). These questions have been raised often with respect to creoles, most cogently by Le Page & Tabouret-Keller 1985 (but see also Kouwenberg 1992; Meyerhoff & Niedzielski 1994; Trudgill 2002).

For example, Faraclas et al. (2007) extend the enquiry of complexity to encompass the political economy of earlier creole-speaking communities, and argue that "the political, economic, and ideological framework within which power relations were manifested" (2007: 258) had significant effects on the emergent structure of different Caribbean creoles. Gillian Sankoff's work on Tok Pisin, and the substantial body of work on other creoles that it has inspired, demonstrates that creoles may have more complexity in their grammars than most linguists and speakers of creoles give them credit for. In particular, when we examine the structured variability in creole languages, we find evidence that speakers are tracking and attending to a good deal more information than we would think if we only look at deterministically marked features.

In the tenor of this earlier work, I will pursue a sociolinguistic investigation of complexity in creoles. I explore the evidence that there are consistent, albeit probabilistic, features in creole grammars. But because evidence for these probabilistic features is generally found by observing the distribution of related variants in different discourse or sentential positions, they also illustrate clearly why domain-based (or modular) definitions of simplicity may be missing large pieces of the overall picture.

3.   Assumptions and methods

Having reviewed some useful and important observations about comparing the simplicity and complexity of natural languages, I will now outline some of the important assumptions that inform the selection of and discussion of my data.

First, I adopt the position that pidgin and creole grammars reflect features that are present in a range of potential sources. In addition to substrate and lexifier language(s), there is a case to be made for perceptual and/or cognitive factors influencing the structure of creole languages. This position is presently uncontentious among creolists, though it's interesting to note that among linguists who do not actively study creoles there is still a strong belief that the grammar of creoles reflects innate, universal properties of the human language faculty, and that they somehow access universal grammar more directly than other languages (cf. comments by Siegel 2006). Among Pacific creolists, like Sankoff, there is wide agreement that multiple sources (linguistic and cognitive) give rise to creole structures, and this agreement unifies researchers working in rather different analytic frameworks (e.g., Baker 1990; Charpentier 1979; Crowley 1989).

I will also adopt the (possibly still contentious) position that non-deterministic properties of a language variety are part of its grammar. Like a number of creolists (Patrick 1999; Rickford 1979; Sankoff 1977, 1980 [1973], 1996; Satyanath 1992;

Sidnell 1999), I will be examining variation using the tools of variationist analysis, considering the potential for both linguistic and social factors to constrain the distribution of variants. In doing this, I hope to demonstrate that (a) aspects of creole grammar should be construed as complex, not simple, and (b) this complexity is seen most clearly when we consider the distribution of linguistic forms across different levels of language structure.

The discussion is organised as follows. First, I briefly review some previous work on two variables in Bislama (an English-lexified creole spoken in Vanuatu). The distribution of variants in these variables is best understood when multiple factors from different levels of linguistic structure and from different source languages are taken into consideration. These include very subtle, probabilistic associations between one variant and semantic or structural features of the clause. I then briefly examine data on possessive marking in Tayo (a French-lexified creole spoken in New Caledonia) and suggest that the three variants used in Tayo likewise favour an account in which constraints on the variation are drawn from several different domains. Finally, I look in greatest detail at data on the distribution of innovative complementisers in Bislama. I will argue that each of these case studies demonstrates considerable complexity underlying the surface-level simplicity of Bislama and Tayo.

4.   Inalienable possession in Bislama

The distinction between inalienable and alienable possession (or direct and indirect possession, as Oceanic linguists tend to identify the distinction now, Lynch 1998), is found in the Austronesian languages that form the substrate to the Melanesian creoles. However, it is widely believed that this semantic distinction is not a candidate for transfer and that it does not show up as a property of the Melanesian creoles. Dutton & Brown (1977: 773–4) state this for Hiri Motu, a Papua New Guinea creole with a non-European lexifier (Motu), and it is taken as axiomatic for the European-lexified creoles of Bislama, Solomon Islands Pijin and Tok Pisin (for example, Charpentier 1979: 340).

However, Sankoff & Mazzie (1991) reported the unexpected finding that NPs in Tok Pisin are more likely to occur with an overt PP of possession if the head N is something that would be considered an inalienable possession. The correlation emerged as a significant factor in a regression analysis looking at the form of NPs in Tok Pisin. Their finding was the first statistical evidence that inalienable (or direct) possessions (such as 'brother' and 'arm') are systematically (though not categorically) differentiated from alienable (indirect) possessions (such as 'tree' or 'story') in a Melanesian creole.

Meyerhoff (2003, 2008) notes a similar statistically significant association between alienable/inalienable possessions and their realisation as phonetically null versus pronominal objects in Bislama. A prototypical inalienable (direct) possessum, *bubu* 'grandmother/father', is shown in (1) and a prototypical alienable (indirect) possessum, *tija* 'teacher', in (2).

(1) *Bubu   blong mi i   no   save draeva.*
    grandparent POSS 1s AGR NEG ABIL drive
    'My grandmother/father doesn't know how to drive.'

(2) *Tija   blong Klas Faev i   no   save draeva.*
    teacher POSS class five AGR NEG ABIL drive
    'The teacher of Class 5 doesn't know how to drive.'

If speakers continue to discuss the *bubu* or *tija*, and if, in the next utterance, they are the direct object of a verb, there is a subtle but statistically significant difference in how they will be overtly realised. In both cases, it is possible to continue with either an overt pronoun, *hem*, or with a phonetically null object, Ø, as shown in (3).

(3) *Bae mi tijim (hem/Ø).*
    IRR 1s teach 3s/Ø
    'I'm going to teach him/her'.

But the overt pronoun option (*hem*) is favoured when the referent is something like *bubu*, that is an inalienable (direct) possessum. An alienable (indirect) possessum, such as *tija* in (2), has no effect one way or another.

The probabilistic effect of the possessum type has apparently been transferred from the substrate languages. This is important because a systematic distinction between inalienable/alienable possession is widely believed not to have been transferred into the Melanesian creoles, and its apparent absence is interpreted as evidence that the creoles are structurally more simple compared to the substrate languages. As this example shows, both the empirical claim and the conclusion drawn from it are questionable. We see this clearly when we consider the interaction between clause-level syntax and discourse-level referent tracking.

5. Subject form and VP agreement in Bislama

Over time, it is clear that some of the pronominal forms in Bislama have been reanalysed so that some of them have now been reallocated (Britain 1997; Trudgill 1986) as agreement markers (or subject indexing markers–what we call them matters much less than the phenomenon), and other forms have taken the place of

free pronouns. So, over time, the pronoun *he* with the phonetic form [i] has been reallocated as the 3s agreement marker with finite verbs. The stressed pronoun *him* has been reanalysed as [hɛm] 's/he; it' and is optional as a subject in most non-contrastive finite clauses. Similarly, English *all* has been reanalysed with the 3s [i] as [oli] and this is now the agreement marker with 3P subjects. English *all-together* has been reanalysed as the free 3P pronoun and has the form [olgɛta] and this too is optional in many cases.

The result of this historical process of reanalysis and change (and some independent parallel developments in the pronoun inventory) is that the Bislama pronominal system and VP now look very similar to the pronominal systems and VP structure of many of the substrate languages (Camden 1979 gives a detailed comparison of Bislama and Tangoa; Crowley discusses parallels with Paamese, 1990: 227). As in many substrate languages, independent subject pronouns in a finite clause are optional and subject reference may be identified through verbal inflections.

But the substrate parallel is not deterministic. As Meyerhoff (2000) has demonstrated, the degree to which subject agreement on the verb is semantically transparent seems to be an important constraint. For example, whereas 3s and 3P agreement are distinct, the agreement associated with 1st and 2nd person is the same (in the singular and also in the plural). This is probably related to the strong preference for 1st and 2nd person subjects to be overtly expressed as pronouns, while 3rd person referents are highly likely to be phonetically null. In addition, information structure plays a statistically significant role. So, for instance, continuity of the same referent in subject position strongly favours a phonetically null subject. Both these factors are illustrated in the sequence of clauses in (4) taken from a narrative told by a young man on Malo island in 1995. Subjects in each clause are underlined.

(4) *Ol man oli kam, Ø oli lukaot hem Ø oli luk*
PL man AGR come Ø AGR look.for 3s Ø agr look

we trak blong leg blong hem i kam go finis nomo long ston.
RELtrack PREP leg PREP 3s AGR come go finish only PREP stone
(M-95-9, Obed)

'Everyone came [and] [they] looked for him. [They] saw (how/that) his footprints went up to the stone and vanished.'

Split pro-drop systems have been attested in the literature (e.g. Finnish, and in some tenses in Hebrew, Borer 1989), but the direction of the split in Bislama is rather unusual. Generally, it is assumed that functional factors, specifically, the immediate presence of 1st and 2nd person referents, will predispose speakers to

omit these pronouns if they omit anything. However, in Bislama the trend is in the other direction, and 3rd person referents are more likely to be phonetically null than 1st and 2nd person subjects are. The counter-functional direction of this shift occurs in several other language contact situations, specifically Cap Verde Creole (Marlyse Baptista, personal communication) and Faetar (Heap & Nagy 1998: 156). Whether this is purely a coincidence, or whether there are factors peculiar to language contact favouring such a split is an open question.

The occurrence of null subjects is ubiquitous in the English-lexified Pacific creoles; they are attested not only in Bislama, but also Tok Pisin and Pidgin (Hawai'i), but at very different rates in each language. In Bislama, they occur approximately 44% of the time in finite clauses where variation is possible. In Tok Pisin, approximately 39% of all 3s subjects are realised as phonetically null. In Pidgin, they occur 8% of the time and tend to be in non-referential positions (where null subjects are prohibited in Standard English). Since the rate for subject pronoun deletion in English is very low even compared to Pidgin (approximately 2% of non-coordinate subjects in English) (Meyerhoff 2008), it is clear that use of phonetically null subject pronouns is not a simple calque from the lexifier.

It is also a feature that hasn't carried over from the substrate in a straightforward manner. I recently compared the distribution of phonetically null and pronominal subjects in a narrative corpus of Tamambo (spoken on Malo island in NW Vanuatu) with the distribution of null and pronominal subjects in a corpus of Bislama I had recorded on Malo (Meyerhoff in press). In a multiple regression analysis using Goldvarb X (Sankoff et al. 2005), I found that different factors were significant in Bislama and Tamambo. Person and number of the subject were not selected as significant factors in the Tamambo corpus (unlike Bislama); subject continuity across clauses was (like Bislama).

The data in Table 1 shows the two factors that were selected as significant constraints on the distribution of pronominal and phonetically null subjects in the Tamambo corpus. It shows that the discourse status of the subject referent had the strongest effect and that the effect is very similar to the effect this factor had for Bislama subjects. It also shows that the animacy of the subject is a significant constraint on the Tamambo subjects. Interestingly, this substrate semantic effect does not appear to have transferred into the Bislama used in the same community, though there is some indication that speakers are a little bit sensitive to animacy when choosing pronouns or phonetically null subjects in Bislama. The principal distinction in Bislama is definitely the person and number of the subject referent, but we see an incipient animacy effect in 3rd person where there is a tendency for 3s and 3p human subjects to be realised by pronouns more than 3s and 3p other animate or inanimate subjects are.

Table 1. Significant constraints governing the use of pronominal subjects in Tamambo (spoken on Malo island, NW Vanuatu) expressed as Goldvarb weightings and percentages

	Goldvarb weight	% Null	Total N
Subject in prior clause	0.686	89	116
Other discourse status	0.228	49	106
Human referent	0.592	82	173
Animate (non-human) referent	0.353	65	74
Inanimate referent	0.298	39	23

These variable patterns reflect an underlying complexity to the creole grammars, one that appears to have multiple sources. The lexifier and the substrate languages all can be argued to have contributed some predisposition to this patterning. But in addition, completely independent factors, such as the semantic transparency of agreement marking and speakers' preferences for overtly tracking topics across stretches of discourse, play a significant role. Like the case study of pronominal versus null objects we looked at first, the distribution of pronominal and null subjects in Bislama reveals a surprising degree of internal complexity. This complexity is expressed through significant probabilistic associations between individual variants and is captured most fully when morphological, syntactic, and information structure factors are all considered.

6.   Tayo possessive marking

Possessive marking in Tayo, a French-lexified creole spoken in New Caledonia (Ehrhart 1993; Ehrhart & Corne 1996), is another case where the convergence of substrate and lexifier systems seems to require more complex generalizations than might at first seem necessary. Tayo emerged following substantial in-migration to the area around the Marist mission at St-Louis, close to Nouméa (New Caledonia). It stabilized around 1855–1880 as the medium of communication among speakers of a number of New Caledonian languages (principally those classified as *Centre North* and *Far South*). In addition, there were a number of French convicts, Indians from Réunion, and later Japanese, Javanese and Vietnamese labourers that also moved into the area (Corne 1990: 5–6; Ehrhart 1993).

Kihm (1995) reviews several aspects of the structure of Tayo VPs and NPs and comments on three ways in which possessive marking may be indexed in NPs (1995: 244–5). All three consist of a NP followed by a PP construction, but the choice of preposition varies depending on the nature of the possessor.

Some examples from Kihm (1995) are shown in (5)–(8). The first line gives the general structural description of the possessive including Kihm's postulated source of the preposition in the lexifier, the second line gives a Tayo example, and the third line gives Kihm's translation.

(5) N + [*pu* + animate N]$_{PP}$ (<French 'pour')
     peti p<u>u</u> *chef de Dubea*
     'a child of the chief of Dumbéa'

(6) N + [*de* + inanimate N]$_{PP}$ (incl. animal by-product) (<French *des* [?])
     chef <u>de</u> *Dumbea*
     'chief of Dumbéa'

(7) *lagres* <u>de</u> *poka*
     'pork fat'

(8) N + [*a* + material N]$_{PP}$ (<French 'en'; not French 'à')
     twa <u>a</u> *paj*
     'thatch roof'

The possessive construction in (6)–(7) N + *de* +N is similar to the default Standard French possessive construction N + [*de* + N]$_{PP}$, though the Tayo preposition has a full vowel (hence Kihm's suggestion that the lexifier source item is *des* rather than *de, d'*). Kihm also argues that, given the way French [a] is realised in other lexical items in Tayo, the possessive construction in (8) must be derived from French *en*, rather than the French preposition *à*. (It is possible that possessives with *pu* < Fr. *pour* were present in the vernacular varieties of French used in New Caledonia in the late 19thC, but this requires further research.)

Kihm (1995) suggests that these three options represent the transfer of substrate patterns into Tayo. He notes that one of Tayo's substrate languages, Nráa Drùbea (an Austronesian language indigenous to the area around Noumea),[6] also has three possessive constructions, however under closer scrutiny, the form and meaning of the Nráa Drùbea possessives seem to be rather different to the forms found in Tayo. This raises questions therefore about the mapping process in general, if indeed there has been transfer of the Nráa Drùbea forms into Tayo. Prince's (2001) discussion of contact-induced features present in Yiddish highlights similar problems; she points out that the borrowing or transfer of form and function may occur independently of one another.

---

6.   *Nráa Drùbea* is the form commonly used by linguists working on New Caledonian languages; the language is identified in the Ethnologue as *Dumbea* [duf] (Gordon 2005). The spelling *Nráa Drùbea* marks the tones that are one of the distinctive characteristics of the language.

The parallels between the Tayo structures and the substrate are very neat for the possessive construction illustrated in (6)–(7). As shown in (9), one of the Nráa Drùbea options for expressing possession involves a sequence of N + *re* + N, where *re* is tentatively identified as a preposition.

(9)   **N+ *re* + N [cf. (6)–(7)]**
    vèdrùu <u>re</u> có
    'spring of water'

In (9), both the form (N + PP) and the function (possessor is an inanimate N) are very similar to the pattern in Tayo (6)–(7); this parallelism seems to argue strongly for a process of fairly straightforward calquing.

However, when we turn our attention to the other means of marking possession in Nráa Drùbea the parallelisms between Nráa Drùbea and Tayo become more attenuated. In (10) we see the second option from Nráa Drùbea, which again involves linking nouns by what is presented as a preposition, here, *a*.

(10)   **N + *a* [prep?] + N**
    trée <u>a</u> vî yô
    'group of CLASS woman'

The possessive construction in (10) superficially resembles the Tayo option illustrated in (8) – both the phrase structure and the phonetic realisation of the preposition are similar in Tayo and Nráa Drùbea. But in Nráa Drùbea the meaning associated with the form *a* is very different from the semantics of *a* in Tayo: Tayo requires the noun that it selects as complement to the preposition to be some form of material. If a process of transfer underlies this parallelism, then that process is more complex than simply carrying over a mapping between form and meaning in Nráa Drùbea. Tayo speakers have reanalysed the material in the process of this transfer. Britain (1997) and Trudgill (1986), among others, have demonstrated that speakers may reallocate variants within a domain (e.g. resulting in the allophonic distribution of variants from different source dialects). When looking at the data from the Pacific creoles (both the case studies of Tayo prepositions and Bislama objects), it becomes clear that reanalysis can involve the reallocation of variants across domains, and this calls into question the extent to which different levels of linguistic structure are in fact autonomous modules for the speakers.

Finally, the third option for expressing possessives in Nráa Drùbea is shown in (11). This option doesn't involve a prepositional phrase, instead possession is marked through concatenation of two nouns, with a change in vowel quality in the head noun.

(11)   **(N + long vowel) + complement**
    mí<u>i</u> dùu
    'branch [of] mangrove'

There is no direct structural parallel between the Nráa Drùbea pattern in (11) (noun compounding and vowel lengthening) and the Tayo patterns in (5)–(8), all of which involve N + PP constructions. If we want to pursue Kihm's line and argue for transfer from Nráa Drùbea to Tayo in the possessive system, we have to expand the notion of "transfer" so it can encompass the reanalysis of a phonological process that operates on the head N in N-N compounds in the substrate as the choice of a particular preposition acting as the complement to the head noun in Tayo.[7]

In short, Kihm is quite right to draw our attention to the parallel between Tayo's three-way possessive system and the three-way possessive marking system in Nráa Drùbea. It's easy to imagine that speakers of Nráa Drùbea might have found it perfectly natural to have three possessive constructions, and to this extent there can be said to be transfer from the substrate. It would also be interesting to see whether other substrate languages, such as Cèmuhî, which had an important founder effect (Mufwene 1996; Sankoff 1980) on Tayo, have the same tripartite possessive system. This would surely strengthen the claim for substrate transfer. But the details of this transfer prove to be quite complex. Phonology, semantics, and syntax are all directly involved in the Tayo patterns, and the solutions that Tayo speakers arrived at demonstrate that the variation is resolved across linguistic domains, not entirely within one domain or another.

In the next section, I return to data from Bislama. As with the previous examples, variability in the Bislama complementiser system suggests that speakers are operating with underlying representations of a system that are more complex than the superficial alternations between forms seem to be.

7.   Emerging complementisers in Bislama

Bislama has a well-established complementiser that can be used to introduce finite subordinate clauses. It has the form *se* [sɛ] and its origins in English "say" and French *c'est* (and semantic parallels with complementisers in some of the Vanuatu substrate languages) are discussed in Crowley (1989). *Se* may also concatenate with other subordinators, especially *from* 'because' and *taem* 'when' and

---

7.   Karin Speedy's work on possible connections between Réunion Creole and Tayo notes that possessive constructions in Réunionnais concatenate Ns with no linking preposition (2007: 205). It is possible that the pattern in (11) owes something to contact between Réunionnais and Tayo.

sapos 'if'.[8] Crowley (1989) also identified *olsem* 'like' as a possible means of introducing a subordinate clause, also with the option of combining with *se*.

(12) *Hem i   wokabaot olsem (se) leg blong hem i   soa.*
      3s   AGR   walk   like   *se* leg of   3s   AGR   sore
      'He is walking as if his leg hurts.' (Crowley 1989: 191, ex. 36)

Meyerhoff (2002) discusses spoken and written uses of *olsem* in a corpus gathered in the 1990s. In this corpus, *olsem* seldom co-occurred with *se* and it seemed that the kinds of clauses *olsem* might introduce were much more restricted than *se*. While *se* has undergone very extensive semantic bleaching and occurs following all kinds of verbs (including propositional attitude verbs such as *biliv* 'believe' and *minim* 'mean'), *olsem* in this corpus was still more likely to introduce subordinate clauses following a main verb of perception.

Many languages systematically distinguish information according to different levels of evidentiality. This may involve a distinction between first hand or reported information, or more rarely according to Aikhenvald (2006: 320) a distinction between visual experience and less direct forms of perception (e.g. hearing and indirect inference). In that light, the division of labour between Bislama complementisers seemed intriguing – perhaps if we were to investigate the synchronic variation in this part of the grammatical system more closely, the results might shed light on the emergence of evidential systems in language in general. As Aikhenvald (2006: 321, 323) notes, different subordination strategies of this kind are a likely source of evidential marking, and evidentiality may be marked in many different parts of the linguistic system. Evidential marking is attested in at least one contact language (Chinese Pidgin Russian, Nichols 1986), so it is reasonable to suppose that the Bislama data might illuminate this general area of linguistic research. While there may be evidential strategies in every natural language, a systematic grammatical system for marking evidentiality is not typically associated with either Bislama's lexifier or substrate languages (a rather different situation, therefore, to the one described by Schieffelin vis-à-vis evidential marking in Bosavi and Tok Pisin, 2007: 151–2). Hence, if there is emergent complexity in this aspect of the grammar, it suggests an internally- or communicatively-driven complexity.

7.1   Defining "evidentiality"

Several recent introductions to evidentiality give an excellent overview of the state of the art (Aikhenvald 2004; Aikhenvald & Dixon 2003). It suffices for this

---

8.   These subordinators may also be followed by *we*, a generalised relative marker. Both *se* and *we* are optional following *from* 'because', *taem* 'time' and *sapos* 'if'.

discussion to define *evidentiality* as the explicit marking of how speakers know what they are reporting, whether this be reported discourse (speaking and thinking) or a report of their sensory experience (hearing, smelling, seeing) or social experience (the speaker's first-hand knowledge, someone else's report, hearsay or inference). In short, it marks the source of the information and the evidence type (Chafe & Nichols 1986; Rooryck 2001). Clearly, all languages can mark source of information and evidence type somehow (optional adverbs, such as *apparently*, convey some of this information in English) but it is useful to distinguish languages that systematically mark evidentiality in, for instance, the tense-mood-aspect system, as is the case for Tibetan (DeLancey 1986) and Bulgarian (Izvorski 1997), or with other clause constituents (Chafe 1986; Floyd 1999: 28; Palmer 1986). Floyd says that the evidential clitics of Wanka Quechua can occur with "almost all lexical classes" (1999: 30).

I will follow some of the experts in this field in saying that languages (like English) where evidentiality is realised by lexical items *mark evidentiality* but do not have an *evidential system*. By contrast, languages where evidentiality is grammaticalised have an *evidential system* (where for the purposes of this discussion, grammaticalised can be taken to mean marked by (more) functional constituents). I also associate the following property with an evidential system: two or more distinctions with respect to the source and type of evidence are realised in the same functional sub-system of the grammar, e.g. unique verb suffixes, auxiliary verbs, complementisers, tense/aspect marking, and so forth. Naturally, speakers may mix evidential *marking* with an evidential *system*, but the distinction is useful because it allows us to approach the new data from Bislama in relation to the structural extremes of lexis and inflectional/clitic morphology.[9] If we adopt this distinction, and if we agree that complementisers are functional not lexical constituents, then the Bislama complementiser variable is a candidate for an emerging evidential system and can be investigated further as such.

7.2   A case for evidentiality in Bislama

Up to this point, I have asserted the emergence of evidential marking in Bislama without providing concrete data to support this. Simply observing the expansion of the complementiser system to include new forms is not sufficient evidence. A contrast between the semantics of the innovative and existing complementisers is

---

9.   Admittedly, the distinction between lexical and functional elements is fuzzy when we are dealing with grammaticalisation–a longitudinal process involving the reanalysis of lexical items as functional items. The distinction between evidential marking and evidential systems may prove to be unnecessary in the long-run, especially where the principal interest is the emergence of an evidential system from lexical marking of evidentiality. But I hope the distinction is a helpful heuristic at this point.

important, as are differences in the typical collocations associated with each. In the case of *olsem* and *se* the fact that one derives from a verb expressing comparison or semblance ('like') and the other derives from a verb which may directly or indirectly report discourse ('say') is one clue. Furthermore, impressionistically, it appears that *se* and *olsem* co-occur with different kinds of main verbs – as noted above, *olsem* favours verbs of perception and *se* is more generally distributed.

But given the semantics of *olsem* 'like' it is hardly surprising to find this general co-occurrence pattern. More compelling evidence that an evidential system is grammaticalising would come from more fine-grained distinctions in how and where the complementisers are used. Specifically, are all verbs of perception equally likely to co-occur with *olsem*? And furthermore, does the subject of the main clause influence the selection of the complementiser? Both the semantics (including tense) of the main verb and the person and number of the subject are widely-recognised to be relevant factors in the systematic grammatical marking of evidentiality (e.g. Floyd 1999: 189–191). Working in a generative framework, Rooryck proposes that the there is an evidential mood phrase which has anaphoric properties. This means that "the matrix subject is responsible for the information status of the sentential complement" (2001: 161), while in directly reported speech he argues that it is coindexed with the speaker. The details of his analysis need not concern us here (for one thing, he analyses a much smaller repertoire of complement clauses than we will consider in Bislama). The important point is that the close associations between verb type and main clause subject and selection of a subordinating complementiser are well-established and implemented in a number of different syntactic approaches.

I will explore these possible interactions in the Bislama data by asking the following questions:

QUESTION 1: Is *se* preferred with matrix verbs of direct perception ('see, hear'); *olsem* with low evidentiality ('feel, smell, appear')?

Given the semantics of *olsem*, we expect it to initially be preferred with verbs selecting complements that are evidentially attenuated and inferential, e.g. 'feel' and 'appear', but *se* to be preferred where the evidence for the proposition in the complement is specified as being more direct, e.g. 'look'. Note that this question is independent of the form of the main verb, rather it depends on context of use. *Luk* can mean either 'see' or 'seem/appear [via visual evidence]'.

QUESTION 2: Is *se* preferred with 3rd person matrix subject; *olsem* preferred with 1s matrix subject?

A speaker necessarily has less direct and solid evidence about how someone else knows or perceives something than the speaker has about how they know or perceive things themself.

7.3 Method

I tested these hypotheses by means of elicited acceptability judgements on 15 paired sentences in which the complementiser was varied. Although this is a small sample, it serves as a useful indication of trends in spoken Bislama in 2001. Examples (13) and (14) illustrate some of the alternations.

(13) *Bang i wantem mi faen from mi ovaspen.*

    a. *Be mi harem **se** bae mi no pem. Folt blong olgeta.*

    b. *Be mi harem **olsem** bae mi no pem. Folt blong olgeta.*

       'The bank wants me to pay a fine because I am in overdraft.
       But I feel (that/like) I won't pay it. It was their fault.'

(14) *Bae mi no go long stoa tedei. Mi go long haos …*

    a. *mo mi luk **se** man blong mi i go long stoa finis from ol samting oli stap long kijin.*

    b. *mo mi luk **olsem** man blong mi i go long stoa finis from ol samting oli stap long kijin.*

       'I'm not going shopping today. I went home and I saw (that/like) my husband had already been shopping because of all the things in the kitchen.'

In terms of the hypotheses, example (14) represents a situation that I expect to favour *se* as the complementiser: both the main verb *luk* 'see' and the subject *mi* 'I' emphasise the direct experience and imply high speaker certainty. On the other hand (13) implies less certainty. Although the subject is still first person, the main verb selecting the complementiser is *harem* with its sense of 'feel'.

The full range of verbs tested were *harem* 'hear', *luk* 'see', *ting* 'think' , *drim* 'dream', *smell* 'smell', *harem* 'feel', *luk* 'appear/seem'. The *olsem* and *se* options were presented in alternating order within the questionnaire, and the questionnaire order was reversed for half the speakers who completed the exercise.

Speakers were asked to rate each option on a four point "smiley" scale, where ☺ was *nambawan* 'perfect' and ☹ was *mi no laekem nating* 'awful'. There were two weaker smiley and sad faces in between glossed as *i oraet* 'it's OK' and *i defren* 'not quite right'. I decided to use smiley faces as the primary stimulus (each smiley had the Bislama gloss in smaller text beneath it) in order to be able to administer the test with pre-literate children and non-literate members of the public.

I report here on the responses from a sub-set of the people who undertook the test in Port Vila in July 2001. Forty-nine people had education to at least Class 6 (age 11–12) and most of them had been educated in English medium. Twenty-one people identified Bislama as their L1; twenty-five identified Bislama as their L2; and three identified Bislama as their L3. These Bislama speakers were able to

complete the questionnaire on paper, hence I was able to gather more responses from them more quickly.

7.4   Results

The investigation yielded 1409 responses. This was out of a possible 1470, but for some questions people didn't give a response at all. This generally was when they had given the alternate complementiser sentence a *nambawan* rating (sometimes an *i oraet*). I suspect, therefore, that many of the "no responses" in practice signify *mi no laekem nating*. The following results report only on the answers people explicitly gave. Some of the speakers that are included in my dataset were relatively young (I was given permission to do the exercise in a classroom of 12 year olds), the class teacher (who has a degree in Applied Linguistics) and I were both available to help the children if they had any queries about the procedure.

Overall, the average acceptability ratings showed speakers find *se* more acceptable than *olsem* across the range of verbs and subjects the complementisers were tested with. The mean acceptability rating for *se* was 1.77 (1=perfect, s.d.=1.09) while for *olsem* the mean was 1.97 (s.d.=1.12). This is what we would predict given that *se* is the more established complementiser and is supposed to be available for use following all main verbs and *olsem* is the innovative variant. Though this difference is small, a comparison of the two averages shows the difference to be significant (p<0.001).[10]

I found no support for the hypothesis that *olsem* would be rated as more acceptable when the matrix subject is the speaker (1s) than when it is a third person subject. The mean acceptability rating for *olsem* with 1s subjects was 1.98 (s.d. = 1.1) and for third person subjects the mean rating was 1.95 (s.d. = 1.14). However, as we will shortly see, this is only true when we consider all matrix verbs types together.

Olsem was rated as more acceptable when the sample sentence had a 'see' or a 'seem/appear' matrix verb than when the matrix was any other verb. The average acceptability rating for *olsem* following each of the different matrix verbs can be seen in Table 2 below.

---

10.   I divided the difference between the mean ratings for *se* and *olsem* by the square root of the standard error of both. I used this method for all the comparisons of means that I report on in this section.

Table 2. Mean acceptability ratings for *olsem* with each matrix verb tested, contrasted with mean for all other verbs (with standard deviation for each)

Matrix verb	Mean acceptability rating (s.d.)	Other verbs mean acceptability rating (s.d.)	Significant?
SEE	1.78 (s.d. = 0.98)	2.01 (s.d. = 1.15)	Yes, p < 0.005
SEEM/APPEAR	1.61 (s.d. = 1.00)	2.02 (s.d. = 1.13)	Yes, p < 0.001
HEAR	2.09 (s.d. = 1.15)	1.95 (s.d. = 1.11)	No
THINK	2.22 (s.d. = 1.26)	1.9 (s.d. = 1.07)	No
DREAM	2.14 (s.d. = 1.15)	1.94 (s.d. = 1.11)	No
FEEL	1.98 (s.d. = 1.03)	1.97 (s.d. = 1.13)	No
SMELL	1.91 (s.d. = 1.03)	1.98 (s.d. = 1.13)	No

Since we are interested in seeing whether the grammaticalisation of *olsem* follows a trajectory that is consistent with cross-linguistically attested patterns of evidentiality, Table 2 considers the rating for *olsem* with each matrix verb and contrasts it with the acceptability of *olsem* elsewhere. As we can see, there is little evidence for anything like an elaborated evidential system. But the evidence suggests that speakers find sentences with *luk olsem* 'see that' more natural than *olsem* with any other verb of perception. *Olsem* is viewed particularly favourably following *luk olsem* 'seem/appear that', e.g. 'the cloud (looks/seems) COMP it is a spider', 'he looks COMP someone hit him'. The use of *olsem* with 'seem/appear' verbs is to be expected because *olsem* as a preposition expresses similarity between two events or things and this meaning is congruent with the similative sense of 'seem'.

This may seem so self-evident that it's worth pointing out that on average 11% of the respondents gave *se* a rating of 1 (*nambawan* 'perfect') with *luk* meaning 'seem/appear'. So this synchronic data is consistent with the historical record about the Bislama complementiser system: *se* is the established complementiser and can be used with any matrix verb. I conclude that where we see a strong preference for *olsem* now (following 'seem/appear') this is a genuine innovation, and that *olsem* has become acceptable with *luk* meaning 'see' partly because it is so well-established with *luk* when it means 'seem/appear'.

However, Table 2 masks further underlying trends. I compared the responses for *ting* 'think' with both a 1s subject and a 3s subject. Reporting one's own thoughts is, obviously, a very different activity to reporting another person's thoughts. We can claim direct experience of our own thoughts and ideas (and as many people have noted, reported thought is often treated no differently from reported speech in discourse, e.g. Güldemann & von Roncador 2002; Buchstaller 2004), however, in order to talk about someone else's thoughts and ideas we have to do a considerable amount of inferencing or rely on hearsay. If *olsem* is being used by speakers to indicate lower levels of confidence in what they are reporting, then we would expect it to be used more with

ting 'think' when the matrix subject is third person than when it is first person. In this data, we find the opposite. There is a significant difference in the acceptability rating for *olsem* when it follows a first and a third person subject: it shows speakers find *olsem* more natural and acceptable with *mi ting* 'I think' than with *hem i ting* 's/he thinks' and *oli ting* '(they) think'. The mean rating for *mi ting* is 1.96 (s.d. = 1.21) and the mean rating for *(hem i/oli) ting* is 2.36 (s.d. = 1.27). There is a non-significant trend in the same direction with *luk* 'see', where *mi luk olsem* gets slightly better acceptability ratings than *hem i luk olsem*. What might this suggest? It is possible that speakers are using *olsem* to hedge when asserting their own thoughts or perceptions. This would make sense semantically, and I do have examples of *olsem* being used as a lexical hedge. Perhaps a preference for *olsem* when following a first person matrix subject indicates that it is being used more systematically in the grammar as a way of indicating speaker stance.

I then explored the responses in more detail, and introduced a strict notion of "preference" to the interpretation of results. This involved looking only at instances where speakers said that a particular complementiser was *nambawan* 'perfect'. The results of this for both *se* and *olsem* are shown in Table 3.

Table 3. Percentage preferred (1= *nambawan* 'perfect') responses for the complementisers *se* and *olsem* following different types of matrix verb

	se % perfect (N)	*olsem* % perfect (N)
SEE	48 % (71)	37 % (55)
HEAR	58 % (57)	20 % (20)
THINK	52 % (76)	17 % (25)
DREAM	59 % (58)	21 % (21)
FEEL[a]	33 % (16)	22 % (11)
SMELL	41 % (40)	38 % (37)
SEEM/APPEAR[b]	11 % (11)	60 % (59)

[a] Only tested with 1s *mi* subject; [b] Only tested with 3s *hem* 'it' subject.

Note that 'seem' matrix verbs were only tested with a 3s matrix subject: while it is possible to construct sentences with a 1s subject and 'seem', they are often pragmatically bizarre. Floyd (1999) notes that in some evidential systems such combinations are simply not possible because the implicature it gives rise to – that the speaker might have to infer information about their state of mind or actions – generates too much cognitive dissonance. 'Feel' was only tested with a 1s subject, and although this was an unintentional gap in the questionnaire, we can make a similar observation about the social and/or cultural felicity of commenting on another's feelings. Since I found several of my sample sentences were being judged on the grounds of cultural acceptability (rather than grammatical acceptability)

anyway, it was probably not a bad thing that both 'seem' and 'feel' occurred only in more natural contexts.[11]

Let us, for the sake of exposition, assume that the matrix verbs in Table 3 reflect a rough evidential hierarchy, with matrix verbs where the speaker has most reliable and direct evidence for the proposition at the top, and those where the speaker has the least direct or reliable evidence at the bottom. When we plot the data in Table 3 as a line chart, we see a general picture of the kind of complementarity that we would expect when we have two variants in competition with each other. Since there is no reason why someone might not find both *se* and *olsem* perfect with all the different verbs they were tested with, it is noteworthy that they do seem to be patterning like complementary variants.

As we have seen above, 'see', 'smell', and 'seem' already accept *olsem* as a complementiser. It would appear that the next matrix verb which may prove receptive to *olsem* is 'feel'. 'Feel' has the fewest responses which pick out *se* and/or *olsem* as 'perfect' complementisers. This may reflect greater uncertainty about whether there is a correct or preferred complementiser for 'feel'.

A further factor which would merit investigation is the extent to which substrate languages provide models for this patterning of *se* and *olsem* complementiser use. We know that numerous substrate languages in Vanuatu have polyfunctional *olsem*-type lexemes which may function as complementisers (Meyerhoff 2002), but there has been no systematic investigation of how deep the parallels between the substrate and Bislama might be. Specifically, future work will need to determine which matrix verbs and/or which matrix subjects favour the *olsem*-type complementisers in the vernacular languages of Vanuatu.

---

11.   For example, some of the respondents in the market were very unhappy with the following sentences and ranked them as *nogud nating* 'awful' for practical reasons (a) and cultural reasons (b):

    a.   *Pakoa i gat wan bot. Hem i se bae i karem mi long Erakor.*
       *Be mi no wantem. Mi luk **se/olsem** bot bae i kafsaed from i gat tumas man.*

       'Pakoa has a boat. He says he'll take me to Erakor. But I don't want him to. (It seems/I see) COMP the boat will capsize because there are too many people in it.'

    b.   *Mama blong Leiwia i ded finis.*
       *Be las naet Leiwia i drim **olsem/se** mama blong hem i stap toktok long hem bakegen.*

       'Leiwia's mother has died. But last night Leiwia dreamed COMP her mother was talking to her again.'

I can't say how much considerations like this may have affected responses in the data I report on here (which individuals completed on their own in a questionnaire format).
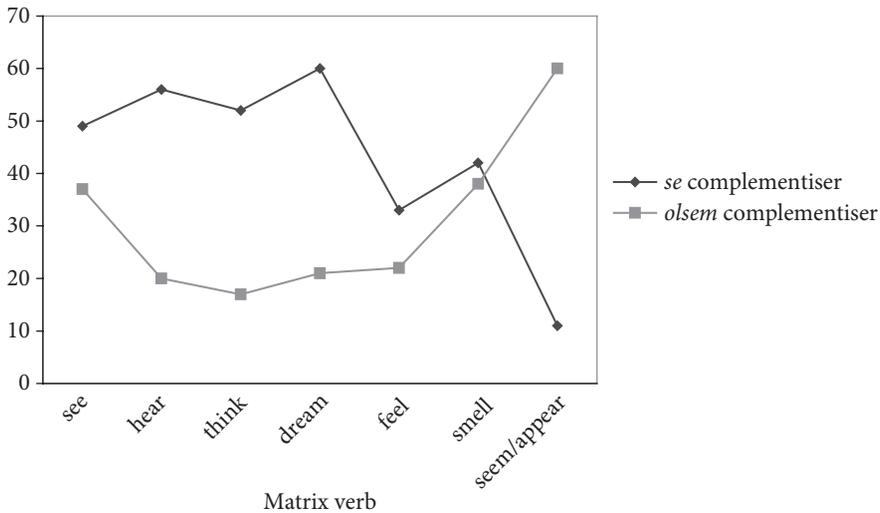
Figure 1. Percentage preferred (=*nambawan* 'perfect') responses for the complementisers *se* and *olsem* following different types of matrix verb. (Matrix verbs ordered left to right in more or less expected decrease in speaker evidentiality).

In Figure 2, I separate out the data for the 22 people who said that their first language is Bislama (six of these people gave Bislama as their only language), and the 27 people who gave other vernacular languages as their first language. The two groups by and large report similar acceptability judgements for the two complementisers, but differ in their complementiser preference following *drim* 'dream' and *harem* with the meaning 'feel'.

The uncertainty with respect to matrix 'feel', which we noted above, seems to be more marked among the people who are L1 speakers of Bislama. A majority of their responses for both *se* and *olsem* were qualified acceptances (mostly 'it's alright' or 'it's different'). The other interesting difference between the groups is in their evaluation of sentences with matrix 'dream', where only seven percent of the ratings by L1 speakers of languages other than Bislama were 'perfect' for *olsem*. Seventy-two percent of the ratings for *drim se* were 'perfect'. This contrasts with 43% 'perfect' ratings given to *drim se* and 39% 'perfect' ratings given to *drim olsem* by L1 Bislama speakers. The ratings given to *drim se* by L1 speakers of Bislama are significantly lower than those given by non-L1 speakers of Bislama (comparison of 'perfect', 'it's alright/it's different' and 'terrible', chi-square= 11.91, p = 0.0025). The greater number of non-L1 Bislama speakers rating *harem se* meaning 'feel' as perfect is not significant ('perfect' versus 'it's alright/it's different' responses in L1 and non-L1 speakers compared, chi-square = 1.79 with Yates correction, p = 0.18).
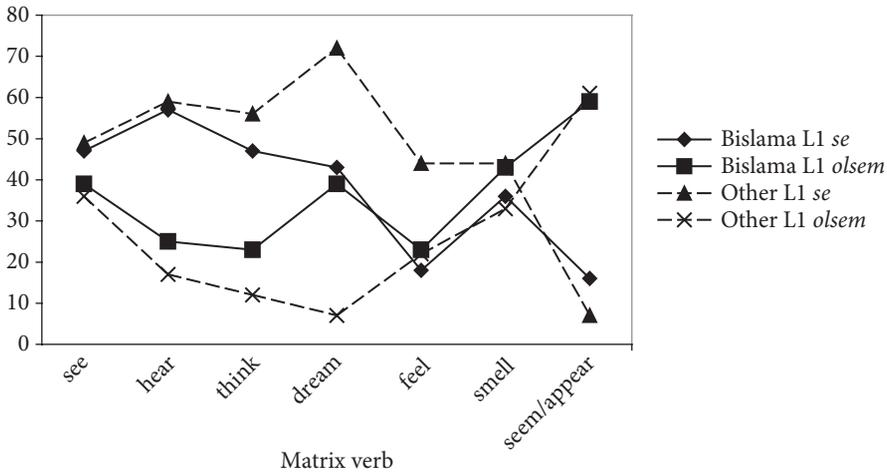
Figure 2. Percentage preferred (=*nambawan* 'perfect') responses for the complementisers *se* and *olsem* by speakers' first language. Bislama L1 contrasted with L1 any other vernacular language of Vanuatu (sample includes some from north, central and southern clusters).

The similarities and differences between the two groups provide some support for the proposal that use of *olsem* is spreading from *luk* (which shows little difference between the groups in both its 'seem/appear' guise and its 'see' guise) to other verbs. I find the pattern with 'hear', 'think' and 'dream' suggestive: in each of these cases, the L1 speakers of Bislama are more likely to give a 'perfect' rating to sentences with *olsem* as a complementiser than people who identified another language as their L1. I take this as possible support for my proposal that *olsem* is grammaticalising and becoming more acceptable in a wider range of linguistic contexts.

In sum, the data on *olsem* use provide very limited support for an emergent complementiser system. It is certainly clear that grammaticalisation is taking place, but so far, use of *olsem* as a complementiser seems to be most preferred with only one verb, *luk* in both its meanings of 'seem/appear' and 'see'. There is some evidence that *olsem* might be preferred with 1st person matrix subjects over 3rd person subjects, though the strength of this generalisation needs to be tested more thoroughly. It is possible that the current data set, based on only 15 sample sentences, is not large enough for systematic effects of subject type to emerge.

It is, however, very clear that there is a good deal of systematicity and structure underlying the surface variation. Again, as we saw with the earlier examples from Bislama and Tayo, understanding the systematicity requires that we consider the interplay between many linguistic factors and also take into consideration likely

sociolinguistic aspects of how and when the innovative variant might be used. Ignoring these factors not only robs us of insights into the synchronic dynamics of long-term language change, but also may feed a misconception that the resolution of complex social and linguistic conditions in creole language communities necessarily leads to simplification. The four examples we have considered here show that the resolution of complexity may require speakers to select from an array of choices and that these choices may be embedded within and across different levels of linguistic structure, possibly drawing on substrate models, lexifier models and unique combinations where both converge on compatible analyses.

8.   Concluding remarks

In this chapter, I have tried to cover ground in several Pacific creoles. My goal was to demonstrate the rich and complex structure that underlies these languages. The fact that this richness of structure underpins aspects of their morphology–when morphological simplicity has frequently been claimed to be a definitional property of these languages–is particularly noteworthy.

I have argued that there is considerable complexity to the variation that is inherent in these linguistics systems. Specifically, I have drawn attention to the need to consider all aspects of linguistic structure, including discourse structure and inter-sentential coherence as part of language variation. In this, my analysis of the data articulates with a growing body of literature that problematises the notion of creole simplicity (Ansaldo et al. 2007). It also articulates with a well-established body of variationist sociolinguistics, in which linguistic, social and cognitive constraints on variation are analysed in relation to each other. By drawing on the empirical insights of both contact linguistics and variationists, I hope this work strengthens our appreciation of the important role played by both fields in understanding the sociolinguistic dynamics of multilingual speech communities.

References

Aboh, E.O. & Ansaldo, U. 2007. The role of typology in language creation: A descriptive take. In *Deconstructing Creole* [Typological Studies in Language 73], U. Ansaldo, S. Matthews & L. Lim (Eds), 39–66. Amsterdam: John Benjamins.

Aikhenvald, A.Y. 2004. *Evidentiality*. Oxford: OUP.

Aikhenvald, A.Y. 2006. Evidentiality in grammar. In *The Encyclopedia of Languages and Linguistics,* Vol. 4, K. Brown (Ed.), 320–325. Oxford: Elsevier.

Aikhenvald, A.Y. & Dixon, R.M.W. (Eds), 2003. *Studies in Evidentiality*. Amsterdam: John Benjamins.

Ansaldo, U. & Matthews, S.J. 2001. Typical creoles and simple languages: The case of Sinitic. *Linguistic Typology* 5: 311–325.

Ansaldo, U., Matthews, S., & Lim, L. (Eds). 2007. *Deconstructing Creole* [Typological Studies in Language 73]. Amsterdam: John Benjamins.

Arends, J. 1995. The socio-historical background of creoles. In *Pidgins and Creoles: An introduction,* J. Arends, P. Muysken & N. Smith (Eds), 15–24. Amsterdam: John Benjamins.

Arends, J. 2001. Simple grammars, complex languages. *Linguistic Typology* 5: 180–182.

Aronoff, M., Meir, I. & Sandler, W. 2005. The paradox of sign language morphology. *Language* 81(2): 301–344.

Bach, E. 1964. *An Introduction to Transformational Grammars*. New York NY: Holt Reinhart & Winston.

Baker, P. 1990. Pacific reorientations. *Journal of Pidgin and Creole Languages* 5: 253–269.

Bauer, L., Holmes, J. & Warren, P. 2006. *Language Matters*. Basingstoke: Palgrave Macmillan.

Bickerton, D. 1984. The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7(2): 173–188.

Borer, H. 1989. Anaphoric AGR. In *The Null Subject Parameter*, O. Jaeggli & K.J. Safir (Eds), 69–109. Dordrecht: Kluwer.

Brighton, H. 2003. Simplicity as a Driving Force in Linguistic Evolution. Ph.D. dissertation, University of Edinburgh.

Britain, D. 1997. Dialect contact and phonological reallocation: 'Canadian Raising' in the English Fens. *Language in Society* 26: 15–46.

Brousseau, A.M., Filipovich, S. & Lefebvre, C. 1989. Morphological processes in Haitian Creole: The question of substratum and simplification. *Journal of Pidgin and Creole Linguistics* 4: 1–36.

Buchstaller, I. 2004. The Sociolinguistic Constraints on the Quotative System: British English and US English compared. Ph.D. dissertation. University of Edinburgh.

Byrne, F. & Winford, D. 1993. *Atlantic meets Pacific: A global view of pidginization and creolization*. Amsterdam: John Benjamins.

Camden, W.G. 1979. Parallels in structure of lexicon and syntax between New Hebrides Bislama and the South Santo language as spoken at Tangoa. In *Papers in Pidgin and Creole Linguistics,* No. 2. [Pacific Linguistics, A-57], 51–117. Canberra: The Australian National University.

Chafe, W. 1986. Evidentiality in English conversation and academic writing. In *Evidentiality: The linguistic coding of epistemology*, W. Chafe & J. Nichols (Eds), 261–272. Norwood NJ: Ablex.

Chafe, W. & Nichols, J. 1986. *Evidentiality: The linguistic coding of epistemology*. Norwood NJ: Ablex.

Charpentier, J.-M. 1979. *Le Pidgin Bislama(n) et le Multilinguisme aux Nouvelles-Hébrides*. Paris: Société d'Etudes Linguistiques et Anthropologiques de France (SELAF).

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge MA: The MIT Press.

Clark, R. 2001. Information theory, complexity, and linguistic descriptions. In *Language Acquisition and Learnability*, S. Bertolo (Ed.), 126–171. Cambridge: CUP.

Corne, C. 1990. Tayo pronouns: A sketch of the pronominal system of a French-lexicon creole language of the South Pacific. *Te Reo* 33: 3–24.

Crowley, T. 1989. Say, *c'est*, and subordinate constructions in Melanesian Pidgin. *Journal of Pidgin and Creole Languages* 4: 185–210.

Crowley, T. 1990. *From Beach-la-Mar to Bislama: The emergence of a national language*. Oxford: OUP.

DeGraff, M. 2001. On the origin of Creoles: A Cartesian critique of Neo-Darwinian linguistics. *Linguistic Typology* 5: 213–310.

DeGraff, M. 2003. Against Creole exceptionalism. *Language* 79: 391–410.

DeLancey, S. 1986. Evidentiality and volitionality in Tibetan. In *Evidentiality: The linguistic coding of epistemology,* W. Chafe & J. Nichols (Eds), 203–13. Norwood NJ: Ablex.

Dutton, T. E. & Brown, H.A. 1977. Hiri Motu: The language itself. In *Language, Culture, Society, and the Modern World, New Guinea Area Languages and Language Study,* Vol. 3, Fasc. 1, S.A. Wurm (Ed.), 759–793. Canberra: Pacific Linguistics.

Ehrhart, S. 1993. *Le créole français de St-Louis (le tayo) en Nouvelle-Calédonie*. Hamburg: Buske.

Ehrhart, S. & Corne, C. 1996. The creole language Tayo and language contact in the 'Far South' region of New Caledonia. In *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas*, Vol. 2, S.A. Wurm, P. Mühlhäusler & D.T. Tryon (Eds), 265–270. Berlin: Mouton de Gruyter.

Faraclas, N., Walicek, D.E., Alleyne, M., Geigel, W. & Ortiz, L. 2007. The complexity that really matters: The role of political economy in creole genesis. In *Deconstructing Creole*, U. Ansaldo, S. Matthews & L. Lim (Eds), 227–264. Amsterdam: John Benjamins.

Floyd, R. 1999. *The Structure of Evidential Categories in Wanka Quechua*. Arlington TX: SIL and University of Texas.

Gil, D. 2007. Creoles, complexity and associational semantics. In *Deconstructing Creole*. U. Ansaldo, S. Matthews & L. Lim (Eds), 67–108. Amsterdam: John Benjamins.

Gordon, R.G. Jr (Ed.) 2005. *Ethnologue: Languages of the world*. 15th ed. Dallas TX: SIL International. Online: http://www.ethnologue.com. Accessed 8 August 2007.

Güldemann, T. & von Roncador, M. (Eds), 2002. *Reported Discourse: A meeting ground for different linguistic domains*. Amsterdam: John Benjamins.

Hauser, M.D., Chomsky, N. & Fitch, W.T. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–1579.

Heap, D. & Nagy, N. 1998. Francoprovençal null subjects and constraint interaction. In *CLS 34: The Panels*, M. Gruber, D. Higgins, K. Olson & T. Wysocki (Eds), 151–66. Chicago IL: Chicago Linguistic Society.

Huddleston, R. & Pullum, G.K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.

Hudson, R. 2000. **I amn't. Language* 76: 297–323.

Izvorski, R. 1997. The present perfect as an epistemic modal. *Proceedings from Semantics and Linguistic Theory (SALT)* 7: 222–239.

Kihm, A. 1995. Tayo, the strange bird from New Caledonia: Determiners and tense-aspect in Tayo and their implications for creolization theories. *Journal of Pidgin and Creole Languages* 10: 225–252.

Klein, T. 2006. Creole phonology typology: Phoneme inventory size, vowel quality distinctions and stop consonant series. In *The Structure of Creole Words: Segmental, syllabic and morphological aspects*, P. Bhatt & I. Plag (Eds), 3–21. Tübingen: Niemeyer.

Kouwenberg, S. 1992. From OV to VO: Linguistic negotiation in the development of Berbice Dutch Creole. *Lingua* 88: 263–299.

Kouwenberg, S. 2001. Convergence and explanations in creole genesis. In *Creolization and Contact*, N. Smith & T. Veenstra (Eds), 219–247. Amsterdam: John Benjamins.

Lefebvre, C. 2001. What you see is not always what you get: Apparent simplicity and hidden complexity in creole languages. *Linguistic Typology* 5: 186–213.

LePage, R.B. & Tabouret-Keller, A. 1985. *Acts of Identity: Creole-based approaches to language and ethnicity*. Cambridge: CUP.

Lynch, J. 1998. *Pacific Languages: An introduction*. Honolulu HI: University of Hawaiʻi Press.

Maher, J. 1984. Sociolinguistic factors in the direction of language change. *Language Problems and Language Planning* 8: 281–293.

McWhorter, J. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2–3): 125–166.

Meyerhoff, M. 2000. The emergence of creole subject-verb agreement and the licensing of null subjects. *Language Variation and Change* 12: 203–230.

Meyerhoff, M. 2002. All the same? The emergence of complementisers in Bislama. In *Reported Discourse: A meeting ground for different linguistic domains* [Typological Studies in Language 52], T. Güldemann & M. von Roncador (Eds), 341–359. Amsterdam: John Benjamins.

Meyerhoff, M. 2003. Formal and cultural constraints on optional objects in Bislama. *Language Variation and Change* 14: 323–346.

Meyerhoff, M. 2006. *Introducing Sociolinguistics*. London: Routledge.

Meyerhoff, M. 2008. Forging Pacific P/C syntax. In *Handbook of Pidgin and Creole Languages*, S. Kouwenberg & J. Singler (Eds). Oxford: Blackwell.

Meyerhoff, M. & Niedzielski, N. 1994. Resistance to creolization: An interpersonal and intergroup account. *Language and Communication* 14: 313–330.

Mufwene, S.S. 1996. The Founder Principle in creole genesis. *Diachronica* 13: 83–134.

Mühlhäusler, P. 1997. *Pidgin and Creole Linguistics*. London: University of Westminster Press.

Muysken, P. 1988. Are creoles a special type of language? In *Linguistics: The Cambridge Survey,* Vol. 2: *Linguistic theory: Extensions and implications*, F.J. Newmeyer (Ed.), 285–301. Cambridge: CUP.

Nichols, J. 1986. The bottom line: Chinese Pidgin Russian. In *Evidentiality: The linguistic encoding of epistemology*, W. Chafe & J. Nichols (Eds), 239–257. Norwood NJ: Ablex.

Palmer, F.R. 1986. *Mood and Modality*. Cambridge: CUP.

Parkvall, M. 2001. Creolistics and the quest for creoleness: A reply to Claire Lefebvre. *Journal of Pidgin and Creole Languages* 16(1): 147–151.

Patrick, P.L. 1999. *Urban Jamaican Creole: Variation in the mesolect*. Amsterdam: John Benjamins.

Prince, E. 2001. Yiddish as a contact language. In *Creolization and Contact*, N. Smith & T. Veenstra (Eds), 263–289. Amsterdam: John Benjamins.

Rickford, J. R. 1979. Variation in a Creole Continuum: Quantitative and implicational approaches. Ph.D. dissertation, University of Pennsylvania.

Rooryck, J. 2001. Evidentiality (Part 1 and 2). *GLOT International* 5: 125–133, 161–168.

Sankoff, D., Tagliamonte, S.A. & Smith, E. 2005. Goldvarb X: A multivariate analysis application. Software available at http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm

Sankoff, G. 1977. Variability and explanation in language and culture: Cliticization in New Guinea Tok Pisin. In *Georgetown University Round Table on Languages and Linguistics*, M. Saville-Troike (Ed.), 59–73. Washington DC: Georgetown University Press. (Ch. 12 in Sankoff 1980).

Sankoff, G. 1980. *The Social Life of Language*. Philadelphia PA: University of Pennsylvania Press.

Sankoff, G. 1996. The Oceanic substrate in Melanesian pidgin/creole revisited: A tribute to Roger Keesing. In *Oceanic Studies: Proceedings of the First International Conference on Oceanic Linguistics*, J. Lynch & F. Pat (Eds), 421–450. Canberra: The Australian National University.

Sankoff, G. & Mazzie, C. 1991. Determining noun phrases in Tok Pisin. *Journal of Pidgin and Creole Languages* 6: 1–24.

Satyanath, S. 1992. Variation and Change: (Daz) in Guyanese. Ph.D. dissertation, University of Pennsylvania.

Schieffelin, B. 2007. Found in translating: Reflexive language across time and texts in Bosavi, Papua New Guinea. In *Consequences of Contact: Language ideologies and sociocultural transformations in Pacific societies*, M. Makihara & B.B. Schieffelin (Eds), 140–165. Oxford: OUP.

Sebba, M. 1997. *Contact Languages: Pidgins and creoles.* New York NY: St Martin's Press.

Siegel, J. 2004. Morphological simplicity in Pidgins and Creoles. *Journal of Pidgin and Creole Languages* 19: 139–162.

Siegel, J. 2006. On the language bioprogram hypothesis. *Language* 82: 2–4.

Sidnell, J. 1999. Gender and pronominal variation in an Indo-Guyanese Creole-speaking community. *Language in Society* 28: 367–399.

Speedy, K. 2007. Reunion Creole in New Caledonia: What influence on Tayo? *Journal of Pidgin and Creole Languages* 22: 193–230.

Trudgill, P. 1986. *Dialects in Contact*. Oxford: Blackwell.

Trudgill, P. 2002. *Sociolinguistic Variation and Change*. Edinburgh: EUP & Washington DC: Georgetown University Press.

Index of names

Index of subjects

In the series *IMPACT: Studies in language and society* the following titles have been published thus far or are scheduled for publication: