

# *Writing a linguistic symphony: Analyzing variation while doing language documentation*

MIRIAM MEYERHOFF  
Victoria University of Wellington  
[miriam.meyerhoff@vuw.ac.nz](mailto:miriam.meyerhoff@vuw.ac.nz)

---

## *Abstract*

Typically, a study of variation starts from the known and works its way into the unknown. But what happens when you are analyzing variation at the same time as you are grappling with the fundamental structure of the language? Whereas variationist methods often involve doing strategic violence to the data, isolating single variables, documentation tends to encourage a broader perspective. This article shows how documentation of Nkep (Central Eastern Oceanic, Vanuatu) has progressed when guided by a focus on internal and social variation. Three variables are discussed (the near merger of two front vowels, lexical borrowing, and the expression of subject agreement) to highlight the rewards and challenges associated with drawing together two subdisciplines (variation and documentation) that have not traditionally had much to say to each other. Analyzing variation alongside documentation encourages us to write ‘symphonies of variation’, as opposed to ‘sonatas’ of individual variables.

**Keywords:** documentation, variation, Vanuatu, vowel merger, borrowing, subject-verb prefixes

## *Résumé*

Traditionnellement, une étude de variation démarre dans le domaine du connu pour se poursuivre dans l’inconnu. Mais que se passe-t-il lorsqu’on analyse la variation d’une langue tout en s’attaquant à la structure fondamentale de cette langue ? Les méthodes variationnistes commettent une violence stratégique à l’égard des données en isolant des variables individuelles, alors que la documentation favorise une perspective plus large. Cet article relate comment la documentation du Nkep (langue Océanienne du Centre Est, Vanuatu) a pu progresser, une fois axée sur la variation linguistique et sociale. Trois variables y sont examinées (la quasi-convergence de deux voyelles antérieures, l’emprunt lexical, et l’expression de l’accord

---

My thanks for constructive criticism and advice to academic colleagues at VUW, ANU and NWAV 44, plus two *CJL* reviewers. Marie-France Duhamel, Erin Hall, Tamsin Porter and Sarah Truesdale assisted with data handling and data analysis. Shirley Warput and Sapo Warput actively supported my work in Hog Harbour. I am grateful to them for their insights about Nkep and about life in general.

All my work on Nkep is dedicated to the memory of Sapo Warput (15 January 1950–14 August 2015) – *a prince among men*.

du sujet) afin de mettre en évidence les avantages et les inconvénients liés à la réunion de deux disciplines que l'on a rarement fait jouer de concert. L'analyse simultanée de la variation et de la documentation nous incite à écrire des « symphonies de variation », plutôt que des « sonates » de variables individuelles.

**Mots-clés:** documentation, variation, Vanuatu, convergence de voyelles, emprunts, préfixe sujet-verbe

## 1. “TWO HOUSEHOLDS, BOTH ALIKE IN DIGNITY”: LANGUAGE DOCUMENTATION AND VARIATIONIST SOCIOLINGUISTICS

A productive alliance has recently emerged across sub-fields of linguistics.<sup>1</sup> Recognising that variation (in one manner of speaking or another) is both unavoidable in language documentation and central to the systematic study of sociolinguistics (known as variationist sociolinguistics), an increasing number of linguists are exploring the intersections between the two areas. Understanding the productive points of contact between the subdisciplines also requires us to consider some of the fundamental differences, in particular the differences in how they handle variation within a linguistic system.

Let us consider the differences by adopting an even broader perspective on variation than linguists in either field have access to. By the age of three, the average English-speaking child has heard approximately 10 million words embedded in the context of 2.5 million sentences. On the basis of this she will have learned about 1000 words, and for the next decade, based on her exposure to spoken and (eventually) written language, she will continue to add 1000 words per year (Biemiller and Slonim 2001, Biemiller 2005, Coxhead et al. 2015).

Research by Smith et al. (2007, 2009) suggests that in these early years, our child is also avidly analyzing the variation in the speech around her. She learns what social activities and social stances have value in her community, she starts to work out what forms alternate as variants and she learns what significance those variants have as markers of the activities and stances she has simultaneously analyzed. She starts to map the variation she hears onto these abstract forms of human behaviour in ways that indicate she is paying attention to which alternations between variants are above the level of conscious awareness in the speech community and which ones are below (Smith et al. 2007). We infer this because she begins with the variables that are above conscious awareness and only later adds in the variation associated with ones that are below conscious awareness. By the time a child is five years old, she is likely to have heard more than 16 million words and to have used this

---

<sup>1</sup>Abbreviations used in this article, in addition to the abbreviations in the Leipzig Glossing Rules (available at <<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>>): df: degrees of freedom; F1, F\*1: first formant (measured in Hz); F\*2: second formant (measured in Hz); Hz: hertz; N: noun; NP: noun phrase; p: probability that distribution occurred due to chance; p.c.: personal communication; prep.: prepositional; SIL: Summer Institute of Linguistics; Z1-Z3; first-third formant (measured on Bark scale)

tsunami of exposure to inductively derive some of the more salient facts about how variation fits into the larger linguistic and social systems that is also a focus of her attention. That is, the child has begun her successful engagement with what D. Sankoff (1988) characterized as the descriptive *and* interpretive enterprise that is variationist sociolinguistics.

The quantity and quality of a child's exposure to her target language is hard to replicate at any other point in our lives. Usually linguists applying themselves to the descriptive and interpretive work of variationist sociolinguistics are assisted in this task by having quite detailed knowledge of the target language in which they want to describe the variation, supplemented by detailed ethnographic information about the communities and speakers using it. But linguists who decide in adulthood to learn and document the structure of an un(der)described language generally do not have this luxury. Even if they are lucky, they may measure their time in the field, listening to and recording their target language, in months. Assuming they immerse themselves in the language (and many or most of us do not; we generally operate at least partly through a lingua franca of some sort), they are likely to be exposed to only a couple of million words. The corpus they record and use as the basis for detailed analysis may only be measured in the tens of thousands. Clearly, this differs vastly from the input available to a child. Field linguists can, of course, compensate for their disadvantage; their experience with and knowledge about other languages gives them a head start in analyzing and extrapolating meaningful generalizations from new data. Linguists undertaking language documentation exercises are not blind to variation; within the field, there is a tradition of trying to document variability across varieties (local and social) (Evans 2003, Himmelmann 2006), and most substantial descriptive grammars that have arisen from documentation projects record multiple form-function pairs. However, it is probably fair to say that generally the focus when documenting variation has been on accounting for variation that is constrained by the linguistic system (allophony, allomorphy), occasionally to the point of recording the linguist's perception of social constraints (e.g., this form is more common in younger speakers), rather than in providing the kind of systematic, quantitative analysis of variation that Sankoff was talking about and that has been the stock in trade of variationist sociolinguistics since the 1960s.

It is worth noting that there is some difference of opinion about the scope of the term *language documentation*. Himmelmann (1998) made a strong case for distinguishing between the documentation of primary materials and the analysis of the documentary evidence (which might include a linguistic description), but in practice this division is not easily maintained. Evans (2008: 348, citing Colette Grinevald [Craig 2001]) argues for documentation as part of "an eternal spiral [...] through the elements of the classic Boasian trilogy – grammar, texts... and dictionary," and Dobrin et al. (2009) also argue against rigid differentiation of documentation and linguistic analysis. Woodbury (2011: 170) talks of modern documentary linguistics as "an ambitious rewelding of the splintered pieces of the Boasian framework." I use the term *language documentation* in a manner more sympathetic with these last researchers.

My sense of *sociolinguistics* is also particular. There are long and deep ties between language documentation and anthropology (Woodbury 2011, Hill 2006),

and documentation was central to the work of Gumperz and Hymes, two key figures in early sociolinguistics. So the intersection between documentation and ethnography is a historical fact. Where there has not been such a clear intersection is with variationist sociolinguistics, and it is the more quantitative, Labovian model of sociolinguistics that is the focus in this article.

The problems with attempting to incorporate a variationist analysis alongside documentation are clear: generally, language documentation is undertaken by someone who lacks the detailed linguistic and ethnographic knowledge that Sankoff noted is needed to accurately describe and interpret variability that is sociolinguistically constrained.

One strategy to address this is to partner with native speakers who can fill these gaps. Stanford's work on variation in Sui and Zhuang (Stanford 2008, Stanford and Pan 2013) demonstrates how successful this strategy can be. But in the absence of collaborators like this, linguists usually decide to abstract away from a detailed analysis of the variation they find in the course of their language documentation.

Ulrike Mosel tells the story of her Ph.D. as roughly following this trajectory. Mosel (2014) says that when she set off into the field in the 1970s, she intended to do for the Tolai language and speech community in Papua New Guinea what William Labov (2006) had done for the New York City speech community. However, Mosel says that when her fieldwork was finished, she confessed to her supervisor that she felt it was impossible to combine the work of language documentation and the analysis of variation in one Ph.D. Her documentation of Tolai (Mosel 1980, 1984) is very rich, and one hopes that the notes she must have taken on the sociolinguistic variation that she observed in Tolai will someday become available, to build upon the earlier structural description.

Mosel's autobiographical anecdote highlights the difficulty that people working at the crossroads of sociolinguistics and language documentation face in reconciling the different demands of the two research traditions.

## 2. COMPARING RESEARCH GOALS IN DOCUMENTATION AND VARIATION

Let us consider what exactly the different demands of the traditions are. It might help to summarize them as best as I understand some of the critical differences. [Table 1](#) provides some ideas about the contrastive picture of the major goals and principles in language documentation and variationist sociolinguistics, though this will undoubtedly be refined further by researchers who continue to work at the intersection of the fields. This view of language documentation is admittedly partial. Gippert et al. (2006) offers a broad overview of goals and theory associated with language documentation. I have characterized the goals of language documentation so as to highlight the intimate links between documentation and linguistic typology, partly because it is my impression that the concerns of linguistic typology are seldom integrated into variationist sociolinguistic thinking. This characterization therefore allows us to highlight differences between the enterprises. [Table 1](#) draws on Himmelmann (1998, 2006), and on Bickel (2007), who is the source of the questions in [Table 1](#).

Goal	Language documentation	Variationist sociolinguistics
<b>Describe</b>	... the structure of language <i>L</i> as broadly as possible	... the interaction of linguistic and social structure for a specific variable in language <i>L</i>
<b>Understand</b>	... language diversity and disparity. “Why are these languages here and now?”	... factors structuring variation in language <i>L</i> (i.e., understand that variation is not free)
<b>Derive</b>	... typological principles. “What’s where and why?”	... principles of co-occurrence or directionality (i.e., what governs the distribution of probabilistic patterns in language <i>L</i> and other languages)

**Table 1:** A preliminary distinction between the main goals of language documentation and variationist linguistics. Questions for documentation adapted from Bickel (2007) on linguistic typology.

A notable difference between the goals in the two columns is that of scale. Documentation has long and deep ties to the field of language typology; hence, the questions that are associated with it are rather grander in scale than those associated with sociolinguistics. Variationist sociolinguistics, for instance, has not been concerned with understanding why specific variables are socially salient in a particular language at a given moment in time (that is, variationists seldom ask “why [this variable] here and now?”). Nor has there been much effort to critically examine whether the distribution of variables within and across languages/varieties are random or are themselves subject to some orderliness. Similarly, variationists sometimes compare the constraints on a variable across different varieties (e.g., Rickford and McNair-Knox 1994, Poplack and Tagliamonte 2001, Meyerhoff 2009) but this has been done to establish the relatedness of two or more language varieties or possible points of contact, not to establish a typology of constraints. In other words, variationists are not particularly focused on questions like “what variables/constraints are where, and why?”). However, there has been some recent movement towards this line of questioning in sociolinguistics, a matter to which we return shortly.

At present, we are in the paradoxical situation where variationist sociolinguistics typically starts from a base of very *broad* observations and knowledge about the community using a language. The enquiry process reduces this to very specific characterizations about the relation between the linguistic and social systems. Meanwhile, language documentation starts with very *specific* observations of data points in a single language and has as an end goal the desire to speak broadly about how those specifics shed light on the nature of human knowledge (including knowledge of language).

In recent work in Vanuatu, supported by the Endangered Languages Documentation Project, I have sought to start from the skill set of a variationist sociolinguist and use those skills to help document an underdescribed language. Nagy (2009) outlines the process of writing a “sociogrammar”, perhaps the first systematic

exploration of what sociolinguistics adds to the documentation of a minority language, and this has subsequently been elaborated for a Sub-Saharan African audience in Childs et al. (2014) (the authors note that many of their points were “anticipated” by Nagy 2009).

Language documentation and variationist studies may diverge in their goals, but they share a fundamental commitment to using naturally occurring data as the primary object of study. The “documentation” of a language may be viewed as narrowly as collecting, transcribing and translating primary data (Himmelfmann 1998) but it is generally seen as an enterprise that results in a grammar linked to a dictionary and texts, that is, it is “accountable to a corpus of natural data” (Pensalfini et al. 2014: 1). Increasingly, in language documentation, researchers attempt to construct corpora based on a diverse set of speech acts and communicative events (Woodbury 2011, Himmelfmann 2006, Austin 2006). In variationist studies, methods have been honed over the years to create contexts in which a conversation can be guided to topics that produce very casual speech. This kind of speech is seen as being the closest to the vernacular grammar; however, since the inception of variationist sociolinguistics, many insights on the systematic nature of variation have been garnered by comparing more and less casual speech. Sociolinguists even use direct elicitation techniques (like those used by documentary linguists), though elicitation may take several forms (e.g., the elicitation of minimal pairs and semantic differential tasks, both of which elicit target words but with varying degrees of attention to the word itself. See Meyerhoff et al. 2015 for a review).

In my experience, documentary linguists are pretty quick to see the merits of the methods and sensibilities that variationists bring to the table. For example, the Wellsprings of Linguistic Diversity project (under the direction of Nicholas Evans at the Australian National University 2014–2019) has been designed to explore the relation between language disparity and diversity at the evolutionary level (cf. “Why these languages here and now?”) alongside micro-variation within varieties. The Wellsprings project has incorporated the variationist notion of apparent time into data collection in sites throughout the Pacific and Australia: researchers are gathering information on social networks, and they have adapted the sociolinguist’s ‘danger of death’ narratives (Labov 2006 [1966]) to local norms, introducing ‘coconut stories’ (in some parts of Melanesia, a coconut is planted to mark memorable events). Clearly, the field of language documentation readily adapts variationist methods and principles in order to articulate with its own goals. Nevertheless, detailed quantitative or qualitative analyses of variation, where researchers consider linguistic and social constraints, such as genre/style and speaker age, still tend to be absent from the outputs of language documentation, though social variation is the focus of qualitative attention in applied sociolinguistics, (e.g., Eades 2015), or in anthropology (see Hill 2006 for a review).

A critical difference between variationists and documentarians is that a documentary linguist is concerned with coming to grips with the whole system of the language, while variationists tend to approach a language more atomistically, typically isolating individual variables for a narrow analysis. However, the wide scope of documentary linguistics can complement the focussed depth of variationist sociolinguistics. Moreover, insofar as the breadth of documentation strengthens the context

for identifying new ways of cross-analysing multiple variables, it also offers one way for variationists to address the violence that quantitative analysis necessarily does to our data.

### 3. ON THE VIOLENCE DONE TO DATA<sup>2</sup>

Quantitative (and, arguably, any) analysis requires us to do necessary and strategic violence to the raw data: the practice of isolating one variable at a time for analysis severs a variable from the system that is the language, and the need to transform our data into something that is statistically tractable further reduces and simplifies the richness of the raw data. To give one example, chosen because the authors explicitly acknowledge the reductive violence done to their data, a careful phonetic study of rhoticity in Scottish English (Lawson et al. 2008) deliberately erases considerable phonetic detail in order to finally analyze a binary alternation between “rhotic” (all approximants, trills and taps) and “non-rhotic” realizations.

All quantitative analyses of complex natural systems involve this kind of strategic violence – a recurring criticism of quantitative sociolinguistics from colleagues in linguistic anthropology. Documentary linguistics, too, does its own violence to the data – for example, by extrapolating from variation or making broad generalizations that elide the social and linguistic constraints on the variation. One benefit of documenting a language with a variationist lens is that it renders the documentation more accountable to the primary data being recorded. Conversely, the big-picture, descriptive enterprise of documentary linguistics balances the destructive nature of isolating variables by strengthening the link between individual variables and the larger linguistic system, suggesting new kinds of questions about the nature of language variation. In short, combining the two approaches is one way to simultaneously enhance awareness of the violence we necessarily do to our data, and to mitigate it.

It is possible to find earlier attempts to connect observations about distinct variables with hypotheses about language systems. Observations of variation and change in North American vowels were contextualized in relation to the dynamic nature of the English vowel system as a whole in the *Atlas of North American English* (Labov et al. 2006). Horvath and Sankoff (1987), Rickford and McNair-Knox (1994) and Dubois and Horvath (1999) all consider multiple variables in different kinds of analyses of variation. Recent work (G. Guy 2013, Hinskens and Guy 2016, and articles therein) as well as Walker et al. (2015) and Meyerhoff and Klaere (2017) explore different methods for linking and quantitatively analyzing the variation observed across a number of variables. Obviously, all such work relies on first identifying and analyzing individual variables, so it is clear that we must respect the need for some amount of reductive violence if we wish to eventually arrive at more expressive statements about the languages we are analyzing.

---

<sup>2</sup>‘Doing violence to the data’ is due to my colleague Richard Arnold (Victoria University of Wellington), whose interests include cluster analysis of complex natural systems. I am grateful to Richard for lively discussions about statistics, but he is in no way responsible for how I use his term here.



Ongoing work in a number of different institutions and drawing on transdisciplinary expertise is going some way towards redressing the violence that variationist studies have traditionally done to the raw data that is the language under investigation. The fact that so much of this has emerged from what are fundamentally language documentation projects reflects the logic of linking the two approaches and signals the potential of work at this intersection to redress the violence that all analysis necessarily does to the primary data.

#### 4. DOCUMENTING VARIATION IN VANUATU

Working at the intersection of documentation and variation is one of the benefits of conducting research in Vanuatu, the most linguistically diverse nation on the planet, with over 110 languages in a population of 285,000. The nation recognizes this linguistic diversity as an asset and actively seeks ways to maintain it. Linguistic documentation and collaboration with local communities on language maintenance are important steps towards this.

It was in this context that in 2011, I started working with the community of Hog Harbour, a village on the East Coast of Santo island, to document their language, Nkep. Nkep is closely related to (and mutually intelligible with) the language variety known as Sakao (J. Guy 1972, Touati 2014) used in Port Olry a short distance further north on the East Coast, but for sociopolitical reasons and because of some descriptive linguistic facts, the two villages prefer to use different names. The structure of Nkep/Sakao is extremely unusual among the Oceanic languages of Vanuatu and even in the region of East Santo. Because of many historical changes, Nkep looks and sounds very different from most other North Central Vanuatu languages (Clark 2009) and the variety has a reputation within Vanuatu for being hard to learn as an adult.

The data reported here derives from multiple field trips during which I elicited traditional narratives (these are culturally and pedagogically important) and oral histories. Materials were all transcribed in ELAN (ELAN 2005–2015, Wittenberg et al. 2006) and the data handled in SIL's FieldWorks Language Explorer (FLEX, SIL 2012–2016). The identification and coding of variables was done within ELAN (following Nagy and Meyerhoff 2015).

I present three case studies of variation in Nkep in order to illustrate how combining documentation and variation has been beneficial for both aspects of the project. The first case study shows how the tools of variationist sociolinguistics have helped to resolve a specific descriptive problem with the vowel space in Nkep: the sociolinguistic notion of attention to speech helps clarify a problem in description within the documentary record (cf. Evans's 2008 'spiral', mentioned above).

The second case study considers patterns of lexical borrowing in Nkep. In this case, the apparent-time construct has proven helpful as a way of addressing a problem that was not particularly pressing for me, but which emerged as part of the community's involvement as a joint partner in the documentation project.

The third case study involves the use of multivariate analysis to address variable patterns documented in the distribution of different verbal prefixes. This example highlights how the analysis of variation can be a powerful hermeneutic in cases



where the linguist is unsure about whether and how forms may be (semantically) related.

#### 4.1 Nkep front rounded vowels

The first case involved a descriptive problem with vowels. Nkep is unusual among Oceanic languages in having a large vowel inventory. While most Oceanic languages have either five or seven vowel systems, Nkep has 11. The series of front rounded vowels probably arose through classic umlaut formation, that is, a front vowel in the final syllable was lost and the [+front] feature transferred onto a non-front vowel that remained in the root. So the Proto-North Central Vanuatu form \**kasi* ‘see’ gives modern Nkep / $\gamma\text{æ}\delta$ / through a series of quite straightforward changes discussed in Clark (2009).

Figure 1 shows a plot made using the NORM software (Thomas and Kendall 2007–2015) of all the vowels in the speech of Sapo Warput (b. 1950), taken from a longish narrative he told to me when some of the other members of the community were present as an audience.

Front rounded vowels are noted as <y> (corresponding to /y/), <eu> (corresponding to /ø/) and <oe> (corresponding to /œ/). The vowel space is anchored by /a/ the low central vowel, /i/ the high front vowel, and (to a less clear extent) by /u/ the high back vowel. Figure 1 shows that we have a basically triangular vowel space with a lot of crowding in the front, where the front unrounded and front rounded vowels are very close to each other.

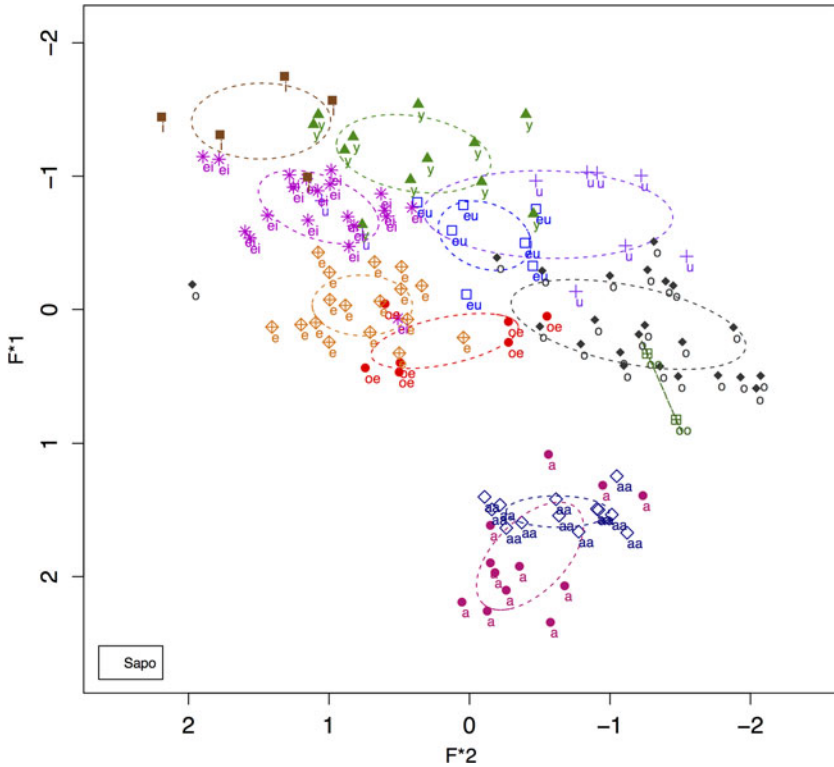
Earlier work based on J. Guy (1972) indicated that Nkep/Sakao have two mid front rounded vowels, /ø/ and /œ/, but minimal pairs for these vowels are very few. There is one culturally salient shibboleth involving two words that have high frequency in Vanuatu village life:

- (1) / $n\text{ø}\delta$ / ‘coconut’
- (2) / $n\text{œ}\delta$ / ‘louse’

But aside from these words, the distinction between the vowels carries a relatively low functional load. Many of the words using the mid-high /ø/ are extremely low frequency or are archaic lexical items (e.g., the names for sweet potatoes and yams, which are no longer widely cultivated).

Moreover, Touati’s (2014) recent description of Sakao observes that the distinction between the two mid front rounded vowels may be neutralized in unstressed syllables. When we look at the realization of these vowels measured in Bark to reflect more accurately how they are perceived (Figure 2), we can see that while they might be distinct in a Lobanov plot, when plotted in Bark, Touati’s observation is backed up. In fact, there is considerable overlap between /ø/ and /œ/ in both stressed and unstressed syllables.

This creates a descriptive problem. Undoubtedly, these forms have different historical derivations, but synchronically, should the data be said to document two distinct vowels? Is Nkep best described as having two mid front rounded vowels, based on the low-frequency but highly salient minimal pair of ‘coconut’ and ‘louse’? Or,



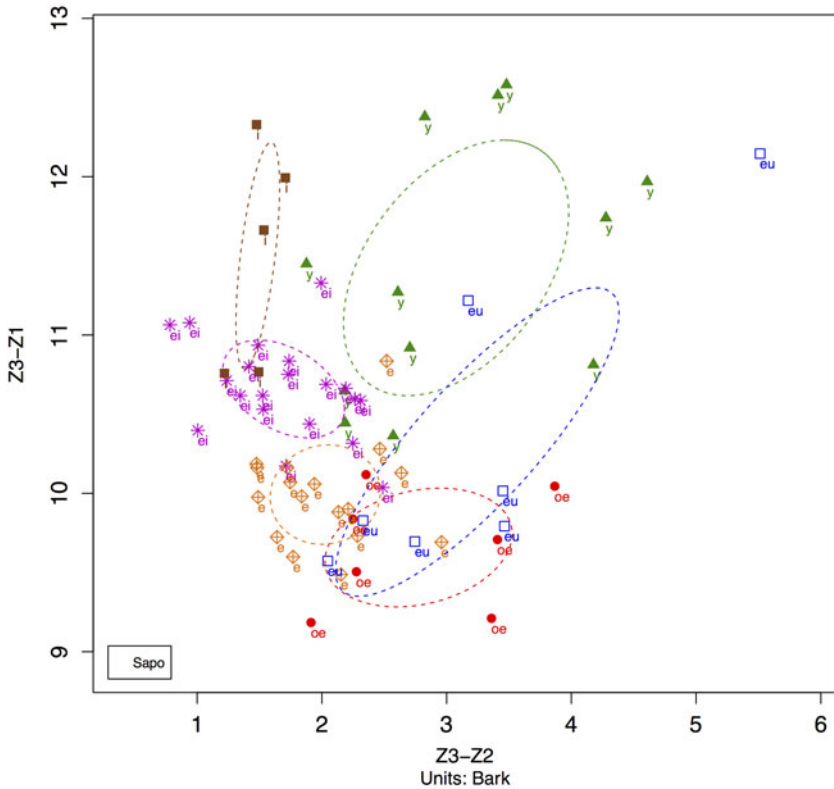
**Figure 1:** NORM plot (Lobanov method) of all vowels in Sapo Warput (b.1950), narrative

regardless of the historical status of these vowels, does the perception and production data suggest that they are synchronically too similar to categorize as two phonemes?

Drawing on variationist methods, we can address these questions in a systematic way. As well as Sapo telling a narrative (and various other spontaneous speech events in the community), he also read some examples of words with the mid front rounded vowels in the carrier sentence shown in (3). When Sapo reached the end of the carrier sentences, he spontaneously produced the two most salient words ('coconut' and 'louse') as a minimal pair twice.

- (3) /jøn namnɔs wartaðɛlp \_\_\_rəvyl/ (phonemic transcription)  
*Yön namnas wartathəlp \_\_\_reviül* (current Nkep orthography)  
 'I want to say XXX again.'

When we compare the F1 values for the vowels in his narrative, his read sentences, and minimal pairs in Table 2, we find that in narratives, the mid-front



**Figure 2:** NORM plot (Bark scale) of front vowels in narrative by Sapo Warput (b.1950)

vowels are barely different, but in tasks where Sapo progressively pays more attention to his speech, the F1 difference between the two vowels is accentuated.

These read sentences involve not only the ‘coconut’/‘louse’ minimal pair, but also some other reasonably familiar lexical items that were matched as closely as possible for preceding and following segment. The overall direction seems clear – there is quite a marked difference between the vowels in minimal pairs, but this difference is largely neutralized in a narrative (as Labov 1994 discusses with near mergers in other languages). Ladefoged and Disner (2012) cite research showing that listeners can accurately discriminate differences as low as 12 Hz, but they note that the just noticeable difference may be different in different parts of the vowel space. Heselwood (2013) suggests that the average just noticeable difference is 60 Hz in F1 and 175 Hz in F2. This may indicate that the difference between Sapo’s mid-front rounded vowels is not perceptually noticeable in narrative style, but is perceptually noticeable when he starts speaking more carefully and self-consciously.

The neutralization Touati (2014) documented for Sakao occurs in Sapo’s Nkep not only in unstressed syllables, but in narrative more generally. Nevertheless, the

	Average Hz F1 /ø/	Average Hz F1 /œ/	F1 difference
Narrative	427	468	41
Reading passage	441	557	116
Minimal pair	403	593	190

**Table 2:** F1 for two front rounded vowels in the speech of Sapo Warput (b.1950) in three different styles

phonemic distinctiveness of the vowels is retained in the system; we can see the difference between them systematically emerge as a speaker pays increasing attention to their speech.

A linguist doing documentation based on the elicitation of controlled sentences might have come more quickly to the conclusion that there are three distinct front rounded vowels in Nkep. But what they would have missed, I suggest, is information that relates to Bickel's big question of "What is found where and why?". If we look at Figure 1 again, it seems that although the front rounded vowels may derive diachronically from the fronting of back vowels, they are synchronically (re)aligned with the front vowels. They are synchronically higher than the back vowels they probably derived from and their current F1 values, especially in most careful speech, create a symmetrical rounded/unrounded series of front vowels.

In other words, using sociolinguistic methods for modelling attention to speech not only generates a clearer picture of the descriptive facts about these vowels, but also allows us to understand the variation in these vowels in the context of a speaker's entire vowel system, revealing how they probably relate to diachronic and synchronic changes across the system as a whole.

#### 4.2 A generational analysis of borrowing

I now turn to my second case study, which focuses on borrowing. In this case, the apparent-time construct has proven to be an effective way of mediating some of the linguistic goals of the project through questions that the community posed to the linguist in the course of the documentation process.

Lexical borrowing occurs in the speech of everyone I recorded in Hog Harbour, as it probably does whenever any bilingual speaker in the world talks to other bilinguals. In Hog Harbour the relevant languages are Nkep and Bislama (the English-lexified creole that is the national language). Speakers tend to have a rather negative view of borrowing; older speakers will strike out words from my word lists if they perceive them to be Bislama, rather than Nkep. This rejection occurs with whole word borrowings and also when a Bislama stem is inflected with Nkep nominal and verbal morphology. In keeping with the (nearly) universal tendency for speakers in older generations to believe that younger speakers are ruining the language,<sup>3</sup> older

<sup>3</sup>A reviewer notes that where change is closely associated with upward educational and social mobility (the superposed variables of Gumperz 1964), this generalization seems not to hold.

speakers of Nkep believe that (i) Bislama borrowings are more common in younger speakers' Nkep than in their own, and (ii) this is a sign of linguistic decline.

However, an analysis of borrowings from Bislama through the lens of apparent time offers no support for (i). All speakers, regardless of age, use Bislama discourse markers and connectives in extended Nkep speech (Meyerhoff 2016). Indeed, connectives like 'but' and markers of narrative/discourse structure like 'so then' are so far below awareness that my language assistants often have to be specifically prompted for 'real Nkep' forms even when they are focussed on the task of retelling a child's story in 'good Nkep'.<sup>4</sup>

Table 3 shows rates of borrowing across three age groups.

There is little difference among them, especially when we consider types rather than tokens.<sup>5</sup> There is a slight increase among the youngest girls, but there is certainly no clear, monotonic pattern of generational change. If we consider the frequency of types/total words, there is a significant increase in borrowings between the middle-aged women and the girls (a test on the raw frequencies, chi-squared = 6.28,  $df = 1$ ,  $p = 0.01$ ) but the differences between older and middle-aged speakers is much less clear.

Table 4 shows the frequency with which borrowings in the three main word classes were subject to any nativization among speakers in the three age groups under consideration.<sup>6</sup>

In this table, we see that there is a decrease in the frequency with which the girls nativize the borrowed words in their Nkep; however, this difference is not significant. (The difference in frequency with which the girls nativize their borrowings is not significantly different from the combined total of all middle and older women; chi-squared with Yates correction,  $p = 0.17$ ). When we look at the data for borrowed verbs, there appears to be a tendency for the girls to nativize borrowed verbs less than the other groups of speakers, but a chi-squared test contrasting girls and the older speakers found that this difference, too, is below the level of significance (girls vs older women, chi-squared with Yates correction = 2.318,  $p = 0.3$ ; aggregating all older speakers versus the girls, chi-squared with Yates correction = 2.734,  $p = 0.098$ ).

In any case, it is not obvious that quantitative analysis of such small numbers is warranted, even if speakers are grouped when possible, given the sample reported here. For all speakers, what seems to most systematically explain switching to Bislama is a qualitative measure of how animated their speech is. So at a moment of high drama in a narrative in Extract 1, we can see that an older woman, Leci Warsal, switches often and seamlessly into Bislama (a phonemic transcription of

---

<sup>4</sup>Other borrowings that don't seem to be noticed are swear words; for example, one way of saying 'he screwed/fucked it up' uses a Bislama lexeme as its stem. The low awareness of these borrowings probably fits with Matras's (2012) observation that bilingual switching often occurs when the speaker is focused on speaker-hearer alignment, rather than content.

<sup>5</sup>The two measures are not significantly different; a t-test returns a value of  $p = 0.064$ .

<sup>6</sup>The numbers of tokens are too small and too unevenly distributed to test, but for the record: girls also nativize 3/12 pragmatic particles; older women nativize 2/21 address/respect terms; middle men nativize 3/12 Proper Ns; older men nativize 1/10 Proper Ns, and 1/2 focus particles.

	Older men	Older women	Middle-aged men	Middle-aged women	Girls
tokens/total words	0.03	0.04	0.06	0.07	0.12
types/total words	0.02	0.02	0.02	0.03	0.05

**Table 3:** Token and type frequency of Bislama borrowings in Nkep across three generations (from Meyerhoff 2016)

	Nouns	Verbs	Prep. phrase
Girls	3/22	3/15	2/2
Middle-aged women	13/17	1/1	NA
Older women	5/14	5/6	1/1
Middle-aged men	23/26	3/4	2/2
Older men	7/27	2/2	1/1

**Table 4:** Frequency of Nkep speakers' nativization of Bislama borrowings in the three most common word classes and across three age groups

her speech is in the first line, followed by Nkep orthography in italics with Bislama lexemes shown by SMALL CAPS).

**Extract 1: Leci's danger of death story**

/we tæmhœ jan ðɔn pɛl ton mɛð ave tmnɛð/  
*wei temhœ yan thaan pel ton, mɛth avei tmnɛth*  
 if we'd run somewhere else, probably we'd be dead

/be tæmhœ janp lðɛ/  
*BE temhœ yanp lthe*  
 BUT we ran and went into the ocean

/ɣam ɣamvɔrye wesi ɣamhœ jan lðɛ/  
*cam cavorcei wesi camhœ yan lthe*  
 there were lots of us, we ran away into the ocean

/yanɣfala ɣamhœ ɣamjan/  
*YANGFALA camhœ camian*  
 and THE YOUNG MEN ran away

/ɣamɣɛr hɔv liviɛyt ðɛ/  
*camɣɛr hov liviɛt the*  
 they swam out to sea

/be ʧam nmama ʧe nwalðay kikri ʧamlro latieð/  
*BE cam- nmama cei nwalthac kikri camlro latieth*

*BUT* us, the mothers and the little children, we hid in the holes in the rock

/ʧaml- ʧamlroke ʧamroke yn ʧaple/  
*caml- camlroke, camroke ün caple*

we were- we were listening, we heard the guns

(NK-20130419-Leci-rebellion1.eaf, 03:25.583-03:40.940)

In this case, a variationist lens applied to the data documented in the project was useful for addressing a descriptive question that has high social salience in the community. The quantitative analysis frames the overall variation space (showing low frequencies of borrowing and no apparently significant differences across age groups) which then invites qualitative analysis to drill down into the uses of variable borrowing even in the speech of older, fluent speakers. The results of this enquiry have been able to allay some of the community concerns about whether lexical borrowing from Bislama in the Nkep of younger speakers is a sign of language decay and have reframed the phenomenon in terms of bilingual competence.

### 4.3 The meaning and aesthetics of pronominal affixes

The third case study involves the use of the classic multivariate analysis of variationist studies as a heuristic for understanding how best to describe the variation in the form of preverbal affixes.

Table 5 shows the prefixes elicited from (especially older) people when giving verb paradigms. They are also the forms that occur most often in the narratives (oral histories and traditional stories) recorded.<sup>7</sup>

The plural prefixes are similar to the free pronouns which themselves derive straightforwardly from the Proto-North Central Vanuatu pronouns (Clark 1985). For example, 1pl excl *\*qam*, 2pl *\*qe* give the modern Nkep forms /ʧam-/ and /ʧem-/. The /-m-/ historically indicates realis mood; it is not clear that speakers perceive it as a discrete morpheme synchronically. The variation between the /ʧ-/ and /t-/ forms are shown in (4)–(7) for 1pl and 3pl, which are the more common plural verb forms in the corpus.

- |        |                     |                               |
|--------|---------------------|-------------------------------|
| (4) a. | <i>ʧam-talpær</i>   | ‘we returned’                 |
|        | b. <i>ʧam-lam</i>   | ‘we went (towards speaker)’   |
|        | c. <i>ʧa-mæmæ</i>   | ‘we prayed’                   |
| (5) a. | <i>təm-ro</i>       | ‘we waited’                   |
|        | b. <i>təm-jan</i>   | ‘we went (away from speaker)’ |
|        | c. <i>təm-l-ŋɔɾ</i> | ‘we were sleeping’            |

<sup>7</sup>These data are discussed in more detail in Meyerhoff 2015. Some of the data here differ from what is shown in that earlier work. This is because, as discussed there, simultaneous documentation and analysis of variation also results in the kind of eternal spiral Evans (2008) sees as desirable.



	singular		plural	
	irrealis	realis	irrealis	realis
1 inclusive			t-	ɣam-, təm-
1 exclusive	naɣ-, n-	nam-	ɣa-	ɣa(m)-, təm-
2	nɛɣ-	m-	ɣ-	ɣɛm-
3	ɣ-, ɣ(V)-	m(V)-	ɣa-, ɣ-	ɣam-, təm-

**Table 5:** Subject indexing prefixes on Nkep verbs. Forms produced most often in direct elicitation and also those observed in narratives

- (6) a. *ɣam-ha*                    ‘they danced’  
       b. *ɣam-hatær*                ‘they sent back’
- (7) a. *təm-rasu*                    ‘they sat down’  
       b. *təm-jan*                    ‘they went’

Touati (2014) gives only t(ə)m-<sup>8</sup> as the prefix for 1pl and 2pl, and ɣa(m)- for 3pl in Sakao, but for Nkep, it is clear that there is variation in the most frequent plural subject forms (1pl and 3pl).

Meyerhoff (2015) uses variationist methods to explore the meaning of the /t-/ initial prefixes based on their variable distribution in the primary data. This was a deliberate deviation from standard variationist practice. Instead of starting with two forms known to be semantically or functionally equivalent, it was hypothesized that because of an observed overlap in the distribution of forms, we may assume that, given the right potential constraints, they may prove to be variants of an underlying abstract variable. That is, the forms may be denotationally equivalent, and the purpose of the analysis of variation is to identify the probabilistic differences in the distribution of the prefixes in order to arrive at a more accurate description of the verb paradigm – a description that can state which forms are the defaults and what social or linguistic features interact productively with the default(s) – essentially adopting a variationist perspective on Bickel’s question “what’s where and why?”.

Meyerhoff (2015) outlines the most significant constraints on the /təm-/ prefixes. They occur mainly with plural subjects, primarily 1pl exclusive, but overwhelmingly with all named discourse participants (i.e., 1pl and 2pl together virtually always occur with /təm-/, supporting Touati’s analysis of them as the 1pl and 2pl prefix). They never occur with a verb designating a future event, but do occur in negative

<sup>8</sup>In the emerging community orthography (associated with government vernacular literacy projects) Hog Harbour speakers prefer to write the prefix with the reduced vowel as <tem-> for ease.

	<i>ɣa(m)-</i> prefix on verb	<i>t(əm)-</i> prefix on verb
1pl inclusive	0	17
1pl exclusive	43	274
2pl	0	32
3pl	275	194
<b>Total</b>	<b>318</b>	<b>517</b>

**Table 6:** Distribution of subject prefixes on Nkep verbs from a corpus of 2539 inflected verb phrases

clauses. This finding provokes further questions for the documentation of Nkep, such as what exactly defines “realis” in this language.<sup>9</sup>

Since the 2015 analysis, documentation of Nkep and the corpus available have expanded and it is now even clearer that the variation between /təm-/ and any other prefix is almost exclusively restricted to 1st person plural exclusive and the 3rd person plural. Table 6 shows the raw number of different prefix forms with plural subjects in a corpus of 2539 inflected verbs.

While 1pl exclusive and 3pl may seem an odd couple in European eyes, there is a language-specific logic to their patterning together. The contrast between 1pl exclusive and inclusive rests on whether some third person (non-discourse participant) is a semantic entailment of the subject (1pl exclusive denotes the speaker and at least one other person who is not the addressee).

If we think back to what I suggested are some of the central goals of language documentation, the analysis of the variation in the preverbal affix system raises some important questions. Answering “why are these forms in Nkep today?” (cf. Bickel’s “why are these languages here and now?”) will need more extensive comparative documentation and analyses of variation than is possible here. It is unclear where the *t(əm)-* forms originally came from – there appears to be no analogue in Proto-North Central Vanuatu.<sup>10</sup> It may be a regional innovation shared by or diffused through several languages in northern Vanuatu and the southern Solomons. Informal reports from linguists documenting other languages in this region (Alex François, Åshild Næss and Brenda Boerger, p.c.) suggest that forms with a similar function – if not a

<sup>9</sup>A reviewer suggests that, given this finding, Table 5 should label the two prefix series as something other than realis/irrealis. Arguments against that are that the distinction between realis/irrealis is highly productive in Vanuatu languages and the /-m-/ can be traced to realis marking in regional proto-languages. As noted here, the fact that negative clauses pattern with realis clauses reinforces the need for language-specific, evidence-based definitions of grammatical categories like ‘realis’.

<sup>10</sup>Though Simon Musgrave (p.c.) notes that Proto-Oceanic \**ta* became Proto-Vanuatu \**ka*. He suggests the emergence of /təm-/ forms alongside /ɣam-/ forms may indicate that there is some phonetic (re)cycling of changes similar to those that have occurred in the past. This seems plausible to me; the reanalysis of the 1pl excl prefix as /t-/ initial differentiates it more clearly from the free pronoun /ɣam(ðøŋ)/ ‘1pl excl’.

similar shape – occur in several languages in the Banks and Torres islands in Vanuatu, or in languages like Äiwoo, spoken in the Solomons. What is interesting for the point about productive intersections between different fields of linguistics is the following: it was only when we treated the distribution of these preverbal affixes in Nkep as a problem in language variation that documentary linguists noticed the commonality among these languages. This is a clear dividend that an analysis at the intersection of typology, history, variation and documentation can pay; a contribution that, I would argue, neither variationist sociolinguistics nor language documentation was likely to make alone.

Probing this variable further with the methods of variationist sociolinguistics also allows us to ask “why these forms here?” (cf. Bickel’s “what’s where and why?”), the answer to which lies at the intersection of a slightly different set of fields: variation and documentation in this case intersect with typology and social aesthetics. Having established some principles about the distribution of the /t-/ initial prefixes, and having determined that they are the default for 1pl and 2pl, we might like to ask under what circumstances we still hear a 1pl/2pl verb with a /ɣ-/ initial prefix. Table 6 showed that the /ɣ-/ prefixes still occur 14% of the time in 1pl exclusive.<sup>11</sup> What factors influence the production of these prefixes in the less monitored speech of narratives?

A further multivariate analysis was conducted of the distribution of /ɣ-/ forms in relation to linguistic factors that included the agency of the subject, whether the subject is a simplex pronoun, a complex one or a NP, and whether there is any intervening material between the subject and the verb.

Since the earlier analysis of /t-/ series prefixes showed a significant effect for subjects that are discourse participants (compared to those which were generic or 3rd person referents), it seemed possible that some subject types might be considered more prototypically subject-like than others and that this might be dependent on verb semantics. The subjects of stative, experiencer, motion and active verbs were coded separately. The phonological shape of the verb stem was also considered, and verbs were coded according to whether or not they contained any back or velar segments that the velar fricative in the /ɣam-/ prefix might be interacting with.

Looking at the data for the 1st plural results in a rather small dataset, but by conducting a multiple regression analysis using the linguistic factors I have outlined as predictors, we achieve a model with  $R^2 = 0.72$  (meaning that over 70% of the variation observed in the data was accounted for by the multiple factors the data was coded for).

Complex pronouns (/ɣamru/ ‘1dual exclusive’, /ɣamðø/ which historically derives from a trial form, but in practice now functions as a general 1pl exclusive) do not co-occur with the /ɣam-/ prefix on the verb. Where the pronoun is the simplex form /ɣam/, it strongly favours the occurrence of a /ɣ-/ series prefix,

---

<sup>11</sup>The frequency of these prefixes is radically different for 3pl (59% of all verbs). For this reason, it seems prudent to initially analyze the variation with the different subject types separately.

however, since there are only 18 tokens of these simplex pronouns, this finding must be interpreted cautiously.

The agency of the subject also emerges as a significant constraint: when the clause refers to a state or a motion event, it strongly favours the use of /*ɣam-*/. Where the main verb is an experiencer verb or any other active verb, this strongly disfavours the use of /*ɣam-*/.<sup>12</sup> Combined with the earlier finding that the /*t-*/ initial prefix is favoured with subjects that are discourse participants, this establishes clear directions for future enquiry into what constitutes a canonical subject in Nkep.

For the purposes of this article, I focus on the additional significant constraint, namely that the /*ɣam-*/ prefix is preferred when the verb stem has a velar segment in it (variants of /*r*/ were coded separately, as there is phonetic variation in place of articulation of /*r*/). All of the forms in (8)–(10) are perfectly grammatical, and attested in the corpus, but not all of them are equally probable, as shown in Table 7.

- (8) a. *təm-ɣæð* 'we saw'  
 b. *ɣam-ɣæð* 'we saw'
- (9) a. *təm-ro* 'we stayed'  
 b. *ɣam-ro* 'we stayed'
- (10) a. *təm-hɔv* 'we followed'  
 b. *ɣam-hɔv* 'we followed'.<sup>13</sup>

This result is a probabilistic association that is so far from categorical that it would be ridiculous to call it consonant harmony or to analyze it in terms of feature spreading. Equally, it is not clear that it warrants an explanation in terms of ease of articulation, since the affinity between velars in adjacent morphemes occurs at such low frequencies. But it appears that speakers of Nkep find it more pleasing – at some level – to have a prefix with a velar segment echoed by a velar in the stem. It may not be harmony, but it does suggest that euphony plays a role in the variation.

	Weighting	%	N
/ <i>ɣ</i> / in stem	0.78	31	49
no dorsal C in stem	0.48	13	100
/ <i>r</i> / in stem	0.24	14	66

**Table 7:** Distribution of the /*ɣam-*/prefix (rather than /*təm-*/) with 1st person plural exclusive according to presence or absence of a velar phoneme in the verb stem

<sup>12</sup>It appears that as far back as Central Eastern Malayo-Polynesian, the subjects of emotion verbs have been treated syntactically like active subjects, cf. Musgrave (2006). Again, I am grateful to Simon Murgave for pointing this out.

<sup>13</sup>The presence of an /*h*/ in the stem had no effect in the first multivariate analysis and was thereafter combined with 'no dorsal'.

At the beginning of this article, I observed that language documentation may feed into formal or structural analyses of the language, but that there is also a tradition of links with ethnographic analysis. In this example, we see that the methods of variationist sociolinguistics, combined with observations in language documentation not only lead us to generalizations of a typological nature, but also remind us that aesthetic or ludic properties of language might also be significant. There is an expressive side to language that language documentation and variationist sociolinguistics can contribute to, even though we usually cede this ground to other disciplines.

## 5. CONCLUDING REMARKS: LANGUAGE AS SYMPHONY, NOT SONATA

In this article, I have outlined productive intersections between the analysis of variation and language documentation in my own work on Nkep. The case studies examined have shown that combining the sensibilities and methods of both fields may be useful for addressing questions of a descriptive nature, a typological nature, a historical nature and a social nature. We have also seen that combining variation and documentation may address questions that animate the community members a linguist works with.

The use of variation as an exploratory tool is not new. D. Sankoff (1988) argued that variationist methods are suitable for the analysis of syntactic, lexical and pragmatic features because any underlying semantic differences between the alternating forms is neutralized in discourse. Retreating from this position somewhat, Torres Cacoullos and Walker (2009: 327) make a convincing case that multivariate analysis can explain how and why a particular syntactic variant is used “to fulfill a particular discourse function”. One way to read their approach, and that of many other people working on syntactic and lexical variation, going back to at least G. Sankoff’s (1980) pioneering work on similar kinds of variation in Tok Pisin, is that if we suspend a hypothesis of semantic distinctiveness (central to the identification of phonological variables), and if we allow for a Sankoff-like neutralization of meaning for syntactic variation in discourse, then a multivariate analysis of variation can reveal traces of historically meaningful distinctions or the persistence of historically meaningful collocation effects.

To suspend what we know about the semantics of the forms being investigated presupposes that we know the semantics. Yet at the early stages of language documentation suspending what is known may be a moot point. We may not yet know whether the forms occurring in similar positions are semantically related, and we may not know in what way they might possibly be related (historically, synchronically, aesthetically). In this situation, I have argued that the methods of language variation are useful hermeneutics that allow us to explore and interpret the variation. Furthermore, I’ve proposed that they lead us to further descriptive insights or enhance our understanding of variation as a component in a linguistic system as a whole.

This perspective on variation as part of a larger linguistic system will, I believe, have a significant impact on the shape of the field of variationist sociolinguistics.

Insofar as sociolinguistic analyses of variation have traditionally focused on one variable at a time, we might say that the results have produced findings that are the linguistic equivalent of a sonata – studies that highlight the rich and beautiful contribution of a single instrument. But the way that language layers modalities and different levels of structure is more like an orchestral piece. We speak in symphonies, not in sonatas. Working at the crossroads of language variation and language documentation layers observations across many levels of linguistic structure and highlights the links between sometimes unexpectedly connected components of the grammar. It compels the researcher to think in terms of symphonies, constantly revising what is known about the language under investigation, and using patterns of variation to revise their understanding of the structure of a language as a whole.

The synthetic perspective on variation that is fostered by language documentation is one route into the exploration of how variables mesh with each other, and this in turn allows us to redress some of the violence that is necessarily involved in transforming primary data for subsequent analysis.

Moreover, in my view, the whole-language perspective of language documentation complements recent trends among variationist sociolinguistics. These have seen researchers begin to explore new quantitative methods for establishing the mutual dependence or independence of several variables in a single linguistic system. However, there are some important differences in how the analysis of multiple variables will play out depending on the kind of language we are dealing with. The selection of variables in better known languages, such as English, Brazilian Portuguese, or Spanish can be guided by theory-driven hypotheses, while the identification of variables that occurs in tandem with documentation work, as described here, may be driven by chance discoveries, throwing up insights that would not be predicted by existing theory (Labov 2015). Moreover, as noted earlier, the corpora in documentation projects may be quite small compared to the corpora available for better known languages. Together, these differences mean that variables identified in documentation projects may not be well-suited to the kinds of quantitative analysis undertaken in Hinskens and Guy (2016) and Meyerhoff and Klaere (2017).<sup>14</sup>

For example, in the Nkep data, some of the methods used in Hinskens and Guy (2016) would be unsuitable given that borrowings from Bislama into Nkep never intersect with the realization of Nkep front rounded vowels. This is because the Bislama vowel system is a proper subset of the Nkep one; phonological adaptations to Bislama loan words happen only with consonants. Similarly, we cannot ask questions about how the type of subject prefix interacts with borrowed verb stems, since Bislama lacks the velar fricative that favours the use of /ɣam-/. A comparison of variables identified during documentation may in some cases require new methods: if we want to examine the co-occurrence of the front rounded vowel variable and the subject prefix variable, we need methods that will allow us to compare variation in a continuous variable (the vowels) and variation in a categorical variable (the subject prefix).

---

<sup>14</sup>Although Meyerhoff and Klaere (2017) deals with an under-described language (Meyerhoff and Walker 2013), the identification of variables benefited from previous analyses of variables in non-standard or creole varieties of English.

Nevertheless, it seems realistic to imagine a future in which some of these questions about co-occurrence become more tractable. Other questions may turn out to be more tractable than expected; as the documentation process continues, additional relevant data may become available for analysis. If documentation and description can exist in a virtual spiral of mutual informativeness, so too can documentation and the analysis of variation.

For some time, sociolinguists have acknowledged the laminated meanings of sociolinguistic variables, recognizing that variants index many social attributes (sometimes at the same time, sometimes in different situations). Some of the recent quantitative approaches to cross-variable analysis have provided one approach for treating the domain of variation as itself being laminated – it may prove to be the case that the variation associated with different variables is itself significantly linked and that these links are a key component to the meaning of variation. Equally, however, the careful building up of observations through the documentation process constitutes another kind of laminated meaning, one that ultimately defines what it is to know a language.

It's clear that the growing interest in new ways of analyzing more than one variable at a time means we are at a crossroads in the field as a whole; one which I certainly hope will lead us to rich and beautiful symphonies in variation in the future.

## REFERENCES

- Austin, Peter K. 2006. Data and language documentation. In *Essentials of language documentation*, ed. Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel, 87–112. New York: Mouton de Gruyter.
- Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1): 239–251.
- Biemiller, Andrew, and Naomi Slonim. 2001. Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology* 93(3): 498–520.
- Biemiller, Andrew. 2005. Size and sequence in vocabulary development. In *Teaching and learning vocabulary: Bringing research into practice*, ed. E. H. Hiebert and M. L. Kamil, 223–242. Mahwah, NJ: Lawrence Erlbaum Associates.
- Childs, Tucker, JeffGood, and Alice Mitchell. 2014. Beyond the ancestral code: Towards a model for sociolinguistic language documentation. *Language Documentation and Conservation* 8: 168–191. <<http://hdl.handle.net/10125/24601>>
- Clark, Ross. 1985. Languages of North and Central Vanuatu: Groups, chains, clusters and waves. In *Austronesian linguistics at the 15th Pacific Science Congress (Pacific Linguistics C-88)*, ed. Andrew Pawley and Lois Carrington, 199–236. Canberra: The Australian National University.
- Clark, Ross. 2009. *\*Leo Tuai: A comparative lexical study of North and Central Vanuatu languages*. Canberra: Pacific Linguistics.
- Coxhead, Averil, Paul Nation, and Dalice Sim. 2015. Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Education Studies*. DOI 10.1007/s40841-015-0002-3.



- Craig, Colette Grinewald. 2001. Encounters at the brink: Linguistic fieldwork among speakers of endangered languages. In *Lectures on endangered languages vol. 2: From Kyoto Conference 2000*, ed. Osamu Sakiyama, 285–314. Kyoto: Endangered Languages of the Pacific Rim.
- Dobrin, Lise M., Peter K. Austin, and David Nathan. 2009. Dying to be counted: the commodification of endangered languages in documentary linguistics. In *Language Documentation and Description*, vol. 6, ed. Peter K. Austin, 37–52. London: SOAS, University of London.
- Dubois, Sylvie, and Barbara Horvath. 1999. When the music changes, you change too: Gender and language change in Cajun English. *Language Variation and Change* 11(3): 287–313.
- Eades, Diana. 2015. Theorising language in sociolinguistics and the law: (How) can sociolinguistics have an impact on inequality in the criminal justice process? In *Sociolinguistics: Theoretical debates*, ed. Nikolas Coupland, 367–388. Cambridge: Cambridge University Press.
- ELAN software package. 2005–2015. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. <http://tla.mpi.nl/tools/tla-tools/elan/>. Accessed 29 November 2016.
- Evans, Nicholas. 2003. *Bininj-Gun-wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: The Australian National University.
- Evans, Nicholas. 2008. Review of Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel, eds., *Essentials of language documentation*. *Language Documentation and Conservation* 2(2): 340–350.
- Evans, Nicholas. 2014–2019. The Wellsprings of Linguistic Diversity. ARC Laureate project, The Australian National University. <<https://chl-old.anu.edu.au/sites/wellsprings/>>.
- FieldWorks Language Explorer (FLEX). 2012–2016. SIL International. <<http://fieldworks.sil.org/flex/>>. Accessed 29 November 2016.
- Gippert, Jost, Nikolaus P. Himmelmann, and Ulrike Mosel, eds. 2006. *Essentials of language documentation*. New York: Mouton de Gruyter.
- Gumperz, John J. 1964. Linguistic and social interaction in two communities. *American Anthropologist* 66(6, part 2): 137–153.
- Guy, Gregory R. 2013. The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics* 52: 63–71.
- Guy, Jacques Bernard Michel. 1972. *A grammar of the northern dialect of Sakao (Pacific Linguistics B-33)*. Canberra: The Australian National University.
- Heselwood, Barry. 2013. *Phonetic transcription in theory and practice*. Edinburgh: Edinburgh University Press.
- Hill, Jane H. 2006. The ethnography of language and language documentation. In *Essentials of language documentation*, ed. Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel, 113–128. New York: Mouton de Gruyter.
- Hinskens, Frans, and Gregory R. Guy, eds. 2016. *Coherence, covariation and bricolage: Various approaches to the systematicity of language variation* (Special issue). *Lingua* 172–173: 1–146.
- Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161–195.
- Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In *Essentials of Language Documentation*, ed. Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel, 1–30. New York: Mouton de Gruyter.
- Horvath Barbara, and David Sankoff. 1987. Delimiting the Sydney speech community. *Language in Society* 16(2): 179–204.

- Labov, William. 1994. *Principles of linguistic change, volume 1: Internal factors*. Oxford: Blackwell.
- Labov, William. 2006. *The social stratification of English in New York City*, 2<sup>nd</sup> ed. Cambridge: Cambridge University Press.
- Labov, William. 2015. The discovery of the unexpected. *Asia-Pacific Language Variation* 1 (1): 7–22.
- Labov, William, Sharon Ash, and Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. New York: Mouton de Gruyter.
- Ladefoged, Peter, and Sandra Ferrari Disner. 2012. *Vowels and consonants*, 3<sup>rd</sup> ed. Oxford: Wiley-Blackwell.
- Lawson, Eleanor, Jane Stuart-Smith, and James Scobbie. 2008. Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticization in Scottish English. *University of Pennsylvania Working Papers in Linguistics* 14(2): Article 13. Available at <<http://repository.upenn.edu/pwpl/vol14/iss2/13>>. Accessed 29 November 2016.
- Matras, Yaron. 2012. An activity-oriented approach to contact-induced language change. In *Dynamics of contact-induced language change*, ed. Claudine Chamoreau and Isabelle Léglise, 17–52. Philadelphia: John Benjamins.
- Meyerhoff, Miriam. 2009. Replication, transfer and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change* 21(3): 297–317.
- Meyerhoff, Miriam. 2015. Turning variation on its head: Analysing subject prefixes in Nkep (Vanuatu) for language documentation. *Asia-Pacific Language Variation* 1: 79–109.
- Meyerhoff, Miriam. 2016. Picking the prettiest pronoun. Paper presented at Sociolinguistics Symposium 21, University of Murcia.
- Meyerhoff, Miriam, and Steffen Klaere. 2017. A case for clustering speakers and linguistic variables: Big issues with smaller samples in language variation. In *Language variation – European perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8)*, ed. Isabelle Buchstaller and Beat Siebenhaar. Amsterdam: John Benjamins.
- Meyerhoff, Miriam, Erik Schleeff, and Laurel Mackenzie. 2015. *Doing Sociolinguistics*. London: Routledge.
- Meyerhoff, Miriam, and James A. Walker. 2013. *Bequia Talk: St. Vincent and the Grenadines*. Westminister: Battlebridge Publishing.
- Mosel, Ulrike. 1980. *Tolai and Tok Pisin: The influence of the substratum on the development of New Guinea Pidgin. (Pacific Linguistics, B-73.)* Canberra: The Australian National University.
- Mosel, Ulrike. 1984. *Tolai syntax and its historical development (Pacific Linguistics, B-92.)* Canberra: The Australian National University.
- Mosel, Ulrike. 2014. *Documenting spoken and written registers in a previously unresearched language*. Workshop at Documentary Linguistics and Variationist Sociolinguistics Summer School, University of Bamberg.
- Musgrave, Simon. 2006. Complex emotion predicates in eastern Indonesia: Evidence for language contact? In *Linguistic Areas: Convergence in Historical and Typological Perspective*, ed. Yaron Matras, April McMahon, and Nigel Vincent, 227–243. New York: Palgrave Macmillan.
- Nagy, Naomi. 2009. The challenges of less commonly studied languages: Writing a socio-grammar of Faetar. In *Variation in indigenous minority languages*, ed. James N. Stanford and Dennis R. Preston, 397–417. Amsterdam: Benjamins.

- Nagy, Naomi, and Miriam Meyerhoff. 2015. Extending ELAN into variationist sociolinguistics. *Linguistics Vanguard* 17 Sept 2015. doi <10.1515/lingvan-2015-0012>
- Pensalfini, Rob, Diana Guillemin, and Myfany Turpin. 2014. In *Language description informed by theory*, ed. Rob Pensalfini, Myfany Turpin and Diana Guillemin, 1–10. Amsterdam: John Benjamins.
- Poplack, Shana, and Sali A. Tagliamonte. 2001. *African American English in the Diaspora*. Oxford: Basil Blackwell.
- Rickford, John R., and Faye McNair-Knox. 1994. Addressee- and topic-influenced style shift. In *Sociolinguistics perspectives on register*, ed. Douglas Biber and Edward Finegan, 235–276. Oxford/New York: Oxford University Press.
- Sankoff, David. 1988. Sociolinguistics and syntactic variation. In *Linguistics: The Cambridge Survey. Vol 4. Language: The socio-cultural context*, ed. Frederick J. Newmeyer, 140–161. Cambridge: Cambridge University Press.
- Sankoff, Gillian. 1980. *The social life of language*. Philadelphia: University of Pennsylvania Press.
- Smith, Jennifer, Mercedes Durham, and Liane Fortune. 2007. “Mam, ma troosers is fa’in doon!” Community, caregiver and child in the acquisition of variation in Scottish dialect. *Language Variation and Change* 19(1): 63–99.
- Smith, Jennifer, Mercedes Durham, and Liane Fortune. 2009. Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Language Variation and Change* 21(1): 69–95.
- Stanford, James N. 2008. A sociotonic analysis of Sui dialect contact. *Language Variation and Change* 20(3): 409–450.
- Stanford, James N., and Yanhong Pan. 2013. The sociolinguistics of exogamy: Dialect acquisition in a Zhuang village. *Journal of Sociolinguistics* 17(5): 573–607.
- Thomas, Erik R., and Tyler Kendall. 2007–2015. NORM: The vowel normalization and plotting suite. <<http://lingtools.uoregon.edu/norm/norm1.php>>. Accessed 29 November 2016.
- Torres Cacoulios, Rena, and James A. Walker. 2009. The present of the English future: Grammatical variation and collocations in discourse. *Language* 85(2): 321–354.
- Touati, Benjamin. 2014. Description du sakao, langue océanienne du nord-est Santo (Vanuatu): Phonologie, morphologie, syntaxe, sémantique et éléments de socio-linguistique. Doctoral dissertation, Université Paris-Sorbonne (Paris IV).
- Walker, James A., Michael Dunn, Aymeric Daval-Markussen, and Miriam Meyerhoff. 2015. Modeling the speech community through multiple variables: Trees, networks and Clades. Poster presented at New Ways of Analyzing Variation 44, York University/University of Toronto.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a professional framework for multimodality research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*. Available at <[http://pubman.mpdl.mpg.de/pubman/item/escidoc:60436:2/component/escidoc:60437/LREC%202006\\_Elan\\_Wi](http://pubman.mpdl.mpg.de/pubman/item/escidoc:60436:2/component/escidoc:60437/LREC%202006_Elan_Wi)>. Accessed 10 January 2017.
- Woodbury, Anthony C. 2011. Language documentation. In *The Cambridge handbook of endangered languages*, ed. Peter K. Austin and Julia Sallabank, 159–186. Cambridge: Cambridge University Press.