

**NATIONAL SCIENCE FOUNDATION**



**NATIONAL ENDOWMENT FOR THE HUMANITIES**



---

## Documenting Endangered Languages

### SAMPLE APPLICATION NARRATIVE

This sample narrative conforms to a past set of guidelines. Please consult the guidelines for the current competition on NSF's website at:

<http://www.nsf.gov/pubs/2006/nsf06577/nsf06577.htm>

## Proposal Summary

### A Reference Grammar of Cemaun Arapesh

The overarching goal of this project is to document the endangered Arapesh languages, which are traditionally spoken by Melanesian people living along the New Guinea north coast in the East Sepik and Sandaun provinces of Papua New Guinea (PNG). Drawing on my past fieldwork on Arapesh, as well as on the fruits of collaborative work currently underway with support from a DEL-NEH institutional grant, the Arapesh Digital Language Archive, I will devote the fellowship year to writing the core chapters of an Arapesh reference grammar. Arapesh is of special significance to linguistic theory for its typologically unusual system of noun classification that elaborates phonological, as opposed to semantic, principles of morphological class assignment and syntactic agreement. Like other Arapesh varieties, the Cemaun dialect of Mountain Arapesh this project documents most extensively is seriously endangered. In the Cemaun villages, language shift to Tok Pisin is well advanced, to the point that there are fewer than 100 fluent speakers, none of whom are under 40, and none of whom are Arapesh monolingual.

The current documentation of Arapesh is fragmentary, inaccessible, and incoherent. Existing sources suffer from serious deficiencies (e.g., problematic transcription, word-breaking, and glossing) that make them inadequate for linguistic research, language preservation, and use by Arapesh people. The confusing patchwork of language and locality names and analytical inconsistencies that they present limit linguists' ability to make confident assertions about Arapesh and to develop a clear picture of the relationships that hold within the Arapesh family.

The reference grammar I propose to write will be theoretically informed but ecumenical, comprehensively describing the Cemaun dialect while synthesizing the data on variation across the Arapesh family as a whole. This will provide a sound basis for comparative and typological work on the Torricelli grouping to which Arapesh belongs, in turn contributing to the interdisciplinary study of Sepik prehistory. The grammar will be based upon elicited linguistic data and more than 50 texts representing different genres of discourse, including conversation, that I audiorecorded, transcribed, and annotated in the field. Claims about the language will be richly supported with examples from these texts.

In writing the grammar I will benefit from access to the multimedia Arapesh Digital Language Archive, which I am currently building in collaboration with experts in digital humanities, database construction, and text encoding at the University of Virginia. The Archive consists of a lexical and grammatical database linked to a collection of digitized, marked-up texts with time-aligned audio, enabling searches across both parts of the Archive. It is being developed in accordance with current standards for linguistic documentation and archiving, providing a robust, enduring record of this endangered language as well as a flexible research tool. A website displaying selected Arapesh text, audio, and grammatical features is also being constructed as an accessible educational resource on the language. Together with the Archive, the grammar will lay the foundation for pedagogical materials to be created later for use in Arapesh village schools.

## TABLE OF CONTENTS

---

For font size and page formatting specifications, see GPG section II.C.

	<b>Total No. of Pages</b>	<b>Page No.* (Optional)*</b>
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	_____
Table of Contents	1	_____
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) <b>(Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)</b>	13	_____
References Cited	6	_____
Biographical Sketches (Not to exceed 2 pages each)	2	_____
Budget (Plus up to 3 pages of budget justification)	3	_____
Current and Pending Support	1	_____
Facilities, Equipment and Other Resources	1	_____
Special Information/Supplementary Documentation	2	_____
Appendix (List below. ) <b>(Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)</b>	_____	_____
Appendix Items:		

\*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

---

A Reference Grammar of Cemaun Arapesh  
[aon, ape, aoj]

Lise M. Dobrin  
Department of Anthropology  
University of Virginia

## 0. Overview

In the broadest terms, the goal of this project is to document the endangered Arapesh languages, which are traditionally spoken by Melanesian people living along the New Guinea north coast in the East Sepik and Sandaun provinces of Papua New Guinea (PNG). Drawing on my past fieldwork on Arapesh, as well as on the fruits of collaborative work currently underway with support from a DEL-NEH institutional grant (the Arapesh Digital Language Archive, described in section 4 below), I will devote the fellowship year to writing the core chapters of an Arapesh reference grammar that will provide a synthetic overview of the Arapesh language family in addition to comprehensively describing contemporary Cemaun Arapesh synchronic grammar. Arapesh is of special significance to linguistic theory for its typologically unusual system of noun classification that elaborates phonological, as opposed to semantic, principles of morphological class assignment and syntactic agreement. Like other Arapesh varieties, the Cemaun dialect of Mountain Arapesh which this project documents most extensively is seriously endangered. In the Cemaun villages, language shift to Tok Pisin is well advanced, to the point that there are fewer than 100 fluent speakers, none of whom are under 40, and none of whom are Arapesh monolingual.

This project description begins with some background information on Arapesh in section 1. 1.1 briefly describes the Arapesh people; 1.2 discusses the relationships among languages within the Arapesh family; 1.3 characterizes the language's endangered status along with the sociohistorical and cultural factors that shape the current linguistic situation; and 1.4 surveys the existing documentation for Arapesh, showing how it forms an inadequate basis for language preservation and for drawing sound linguistic conclusions. Section 2 describes the original source materials analyzed, digitized, and represented in the Arapesh Digital Language Archive, the central research tool drawn upon in writing the grammar. Section 3.1 describes the main features of the reference grammar, section 3.2 the contribution it will make to our understanding of grammatical architecture and the comparative study of New Guinea languages. In section 4, the scope and functionality of the Arapesh Digital Language Archive are summarized. Planned future projects that will bring some of the results of this work to the Arapesh people are mentioned in section 5. Finally, in section 6, the work plan for the fellowship year is outlined.

## 1. The Language

### 1.1 Ethnographic Setting

The Arapesh languages are traditionally spoken across a region that extends inland from the north coast of PNG in East Sepik and Sandaun Provinces. The region was colonized a century ago by Germany, followed by Australia, until 1975 when Papua New Guinea received independence. Due to the long history of Catholic mission activity in the region,

most Arapesh villagers are Catholic, though the traditional Catholicism is now giving way to Pentecostal and charismatic forms of Christianity. Arapesh villagers subsist mostly on sago and the varied produce of traditional swidden gardens, supplemented by store-bought commodity foods like rice and tinfish. Money is obtained by cash-cropping (chiefly coffee and copra), selling produce in markets, and by remittance from urban wage-earning kin. In the 1930s, the “Mountain Arapesh” people were the subject of a famous study by the anthropologist Margaret Mead (1935), who portrayed men and women alike as sharing a nurturing and maternal “feminine temperament,” an oversimplification challenged in a little-known essay by the anthropologist and linguist Reo Fortune (1939), who had been her husband and partner during their troubled 8-month-long Arapesh fieldwork (Dobrin and Bashkow in prep, Bashkow and Dobrin in prep; see also Dobrin and Bashkow in press on the two scholars differing interpretations of Arapesh inter-village relations). It is Fortune’s 1942 grammar of Rohwim Arapesh that remains the most widely cited linguistic resource on the languages of this family. Recent ethnographic studies of other Arapesh-speaking people include Tuzin on the Ilahita Arapesh (1976, 1980, 1997) and Leavitt on the Bumbita (1989, 1995, 1997, 2000).

## **1.2 Genetic Affiliation**

On the basis of lexicostatistical evidence, the Arapesh language family belongs to the Kombio stock of the Torricelli phylum (Laycock 1973, 1975; see also Foley 1986, 2000). Both the Torricelli grouping and the Arapesh languages’ affiliation to it are supported by evidence from reconstructed pronominal forms (Ross 2005a, 2005b). The Arapesh tree branches into three languages: Mountain Arapesh or Bukiyip (Ethnologue code [ape]; 2003 pop. 16,233), Southern Arapesh or Muhiang/Mufian (Ethnologue code [aoj] 1998 pop. 11,000), and Bumbita Arapesh or Weri (Ethnologue code [aon]; 2003 pop. 4,335, Gordon 2005). Nekitel (1985, 1986) carves out from this classification a fourth language, Abu’, though it is open to question whether Abu’ constitutes a distinct language or a northern (Supari) dialect of Muhiang. Given the considerable chaining of phonological, lexical, and grammatical features as one moves from village to village (a distributional pattern not unfamiliar in New Guinea), Arapesh is perhaps best understood as one long dialect chain on which the language classification outlined above is superimposed (Conrad 1978). In Dobrin (1999) I introduced the term “Arapeshan” to disambiguate between the individual languages and the grouping as a whole, but the power of convention is great, and I have not insisted on the term since then. Cemaun is a hitherto undocumented dialect of Mountain Arapesh. It is spoken in Wautogik and Kotai villages, which are located at the northeastern border of the Arapesh dialect continuum in East Sepik Province, where it contacts Boikin, a language of the genealogically distinct Ndu (Sepik-Ramu) family.

## **1.3 Endangered Status**

According to the residential census I took in 1999, the two Cemaun villages Wautogik and Kotai have a combined population of approximately 800 (cf. the official figure of 258 arrived at in the 1990 PNG census; see PNG National Statistical Office 1994). But such population figures present an inaccurate picture of Arapesh vitality, since vernacular fluency is no longer predictable from village of residence. This is increasingly true throughout the entire “West Coast” and “Sepik Highway” areas where Arapesh languages

are spoken. The Sepik provinces have the highest percentage of Tok Pisin users on mainland New Guinea (see Wurm, Mühlhäusler, and Tryon 1996; Aikhenvald 2002), and as with so many other language in this region, Arapesh is giving way to Tok Pisin, and increasingly to English. The advancement of western values and commodities (and with them massive linguistic borrowing) has been unrelenting, and Tok Pisin is now solidly established as the main language used in the Cemaun village communities, as well as by Cemaun people living in town. The youngest good Cemaun speakers are all above forty, putting Arapesh on the verge of “seriously endangered” according to the model of Wurm 1998 (see also Crystal 2000). Many Cemaun speakers show clear deterioration in their grammar and fluency relative to their elders, and their speech in virtually all genres is marked by frequent, unsuppressable code-switching. Below middle age, there is a precipitous drop in vernacular competence in every respect. Young people are unable to produce extended stretches of vernacular speech, and although some young villagers have limited passive competence, most children cannot understand even the simplest everyday commands and greetings. There are fewer than 100 fluent speakers of Cemaun, none of whom are Arapesh monolingual. Underscoring the sense of urgency is the fact that since my fieldwork a number of village elders have died, including the two linguistic consultants I had worked with most closely.

In many ways, the Cemaun Arapesh situation is a textbook case of language shift leading toward extinction. Paradoxically, this process is being accelerated by the Cemaun people’s unusually high rate of success in education and formal urban employment, which makes them reliant on Tok Pisin and English and draws the population, as well as prestige, away from the village and traditional way of life. Indeed, in the last generation, so many people have left Wautogik that more of its people now live in towns and cities than in the village itself. The Wautogik diaspora is unusually well represented in the national elite, and includes several of PNG’s top diplomats, politicians, businessmen, teachers, and professors, as well as several internationally trained PhDs. Fully aware that their language is dying, the people of Wautogik have raised funds for language revitalization activities such as building a vernacular language preschool in their home village. But following an initial optimistic phase, such activities have tended to stall for lack of outside resources and acknowledgment (Dobrin 2006b), and village leaders have been intensely supportive of work on the language, welcoming efforts by western linguists to document the language for their descendants and help reinvigorate its use in the contemporary community.

Language shift among the Arapesh has proceeded in large part in response to culture-external factors associated with colonialism and the postcolonial national framework. These include the use of Tok Pisin as the official medium of Catholic missionization since 1931 and the use of English as the official language of national discourse and schooling since the 1950s (Mihalic 1996, Romaine 1992, Nidue 1990). Perhaps more importantly, however, Arapesh people’s eager embrace of Tok Pisin has been facilitated by the culturally particular ways in which foreign linguistic codes have come to be symbolically associated with the prestige of western modernity, Christianity, and development. While from a linguistic perspective Tok Pisin is undoubtedly a local language, Arapesh people nevertheless strongly associate its use with the West (see also Kulick 1992). I met Arapesh people as young as 50 who resolutely believe the Tok Pisin they speak to be the native language of Germany, an earlier colonial power. (The

glossolalia that occurs in new charismatic forms of Catholic worship is also believed to channel actual European languages such as Polish through local people's voices.) Traditional Arapesh society is culturally predisposed to be open to diffusion and indeed to value most highly those cultural forms, including languages, that can be "pulled in" from outside (hence Mead's 1938 characterization of Arapesh as an "importing culture"). While one hesitates to predict the most dire scenario for Arapesh's linguistic future, the cultural factors that have guided their language's history to this point make the outlook discouragingly dim.

#### **1.4 Previous Documentation**

Cemaun has not been previously documented. Published grammar-length descriptions exist for the neighboring Rohwim dialect of Mountain Arapesh (Fortune 1942), the inland Bukiyip variety (Conrad and Wogiga 1991), and the western coastal But dialect (Gerstner 1963). There are also a few unpublished sources, such as Nekitel's dissertation on Abu' (1985), which provides a linguistic overview of the variety of Arapesh spoken in the author's native village, Womsis; some manuscripts by Conrad (for example Conrad 1996); and a handful of published articles such as Alungum, Conrad, and Lukas 1978.

Nevertheless, none of these is an adequate comprehensive general reference source on Arapesh. Fortune's grammar succinctly describes the Rohwim dialect's typologically unusual phonologically-based noun classification system, which has attracted recent theoretical interest (Aronoff 1992, 1994, 1998; Fraser and Corbett 1995; Dobrin 1995, 1998, 1999a; see below). But it fails to recognize mood, a ubiquitous and criterial inflectional category on the verb. Fortune's texts, which comprise the bulk (2/3) of the book, are also problematic in that they underrepresent phonological information, and represent it erratically; they also fail to indicate morphemes as such, or distinguish morphemes from words. Using a tagmemic model, Conrad and Wogiga's Bukiyip grammar analyzes syntactic and discourse-level features, especially those relevant to translation. But it gives little attention to lower levels of grammar; the phonology and remarkably complex morphophonemics are together covered in a mere 6 pages. Gerstner's grammar is highly inaccessible (it is a typewritten manuscript with handwritten marginalia, available only on microfilm), and it suffers from severe inadequacies in transcription, e.g., conflating voiced and voiceless obstruent segments in the word-final position crucial for understanding the morphology. It is also couched in terms of traditional grammar, inappropriately applying notions such as "future" and "dative" to a language that has neither tense nor case. In short, the documentary record on Arapesh is uneven and incomplete. It provides an insecure basis for linguistic generalization, and it is inadequate for purposes of preservation, both for the community of linguists and for the Arapesh people. Moreover, as I found when studying the available materials in preparation for fieldwork, the fragmentary literature on Arapesh does not support a coherent picture of the family. It comprises a dizzying patchwork of language and village names, transcription conventions, grammatical categories, and time periods, which limit the ability of typologists and theoreticians to make linguistic assertions about Arapesh with confidence.

During the fellowship period I will work to overcome these limitations by preparing the core chapters of a comprehensive grammar that accurately describes Arapesh's structural particularities and that balances a thorough description of a single variety with

systematic reference to others, thus representing both the remarkable unity and revealing diversity of the Arapesh family as a whole. Clarifying the interrelations among the Arapesh varieties requires a broad range of data that has been framed in comparable terms and that can ideally be interpreted against a background understanding of the settings from which the data was obtained. Thus, the special contribution of this grammar in contrast to the previously published record is to present data on diverse Arapesh varieties that was collected by a single researcher (with a single set of biases) at a single point in time (a single fieldtrip), in addition to interpreting data from earlier sources within a calibrating descriptive framework. Such integration is necessary if we are to develop a coherent picture of the genealogical, areal/typological, and geographic relationships that hold within the Arapesh family.

## **2. Source Data**

The primary source of data for the project are the Arapesh field materials I collected during a 15 month research period in PNG in 1997-1999. Fieldwork was supported by an NSF Dissertation Improvement Grant (Award No. SBR-9707681), a Fulbright-Hays Training Grant for Doctoral Dissertation Research Abroad (Award No. P022A70043), and a Wenner-Gren Foundation for Anthropological Research Predoctoral Grant (Award No. 6156). In their original forms, the materials include the following:

1. 320 handwritten pages of Cemaun Arapesh language texts, which have been transcribed phonemically, annotated and (where necessary) glossed in Tok Pisin, and checked with native speakers while in the field. These texts represent origin/land charter stories, both traditional and modern legends, spontaneous narratives, several linguistic events with multiple participants including a formal meeting and casual conversations, and descriptions of traditional items and events. Another 40-plus pages of texts documenting other Arapesh varieties were collected and have been similarly processed; these include samples of the Arapesh varieties spoken at Woginara (the Rohwim dialect of Mountain Arapesh described by Fortune), Dogur and Balam (two distinct western coastal dialects of Mountain Arapesh), Balif (central Muhiang), Hwamsuk (southern Abu'), and to a lesser extent Bumbita village (Bumbita Arapesh, the lexically and morphologically most divergent Arapesh language which has not previously been documented). There are around 50 transcribed texts in the collection. Their corresponding audio recordings range from two minutes to two hours in length.
2. Cassette tape recordings of nearly all of these texts. To ensure the long-term security of the data, the recordings were copied and the originals archived at the University of Chicago Language Laboratories and Archives.
3. Extensive handwritten notes on grammar and phonology, much of it recopied and reordered to reflect preliminary analyses arrived at in the field. This grammatical data occupies more than a dozen field notebooks and is reasonably well organized. Vocabulary was recorded according to semantic domain (e.g., species of vines, yams, birds; color terms; winds and seasons; kinship terms) or according to formal properties (e.g., plural ending; reduplicative sound structure; verb valency). Many notebook sections are devoted to demonstrating specific grammatical generalizations or justifying particular conclusions (e.g., the syntactic incompatibility of object suffixes and lexical object NPs; lists of minimal pairs; evidence for the borrowed status of phonotactically anomalous forms). Some grammar notebook sections are



indexed to particular transcribed texts in which the examples arise. Dobrin 1999b, 2001, and 2006a are based on some of these notes.

4. A FileMaker Pro database of 1495 nouns created by Dobrin prior to fieldwork that compiles information from all the main published sources on Arapesh and includes not only nouns and glosses, but also notes on their morphological irregularities and attested morphosyntactic behavior along with indexing to the published sources.
5. In addition to the texts described under (1) above, permission has been obtained to use the 60 pages of legends that accompany Reo Fortune's 1942 grammar of Rohwim Arapesh, as well as the Bumbita Arapesh language materials collected in the field by anthropologist Stephen Leavitt in 1984-1986. Leavitt's materials are housed in the Melanesian Studies Resource Center and Melanesian Archive at the University of California, San Diego.

Original field materials were collected with IRB oversight. In selecting examples for the written grammar, attention will be paid to the delicately balanced concerns of acknowledging sources, ensuring the authoritativeness of content, and, where appropriate, maintaining confidentiality.

### **3. Cemaun Reference Grammar to be Written during the Fellowship Period**

#### **3.1 Grammar Features**

The grammar I will work on during the fellowship period will be a theoretically and typologically informed but ecumenical reference grammar that comprehensively describes the Cemaun dialect of Mountain Arapesh while devoting systematic attention to the variation that exists within the Arapesh family. It will begin with a chapter contextualizing the language by describing its sociocultural and historical setting, and it will be illustrated by maps showing the distribution of the languages, the villages of sampling, and important dimensions of variation such as the west-to-east *l* to *r* transition along the coast, the switch in the internal order of possessive NPs which seems to follow a north-south axis in the western Arapesh region, and the dramatic differences in structure and use of kinship vocabularies across the family. Although such dimensions of linguistic variation are well known to the researcher and one other scholar, the SIL linguist who has done translation work on Arapesh, Bob Conrad, nowhere in the literature is this kind of information available. Whenever possible, generalizations presented in the grammar will be exemplified with data from the texts.

The topics addressed by the grammar will be arranged bottom up roughly according to traditional grammatical domains. The section on the phonology will cover the distribution of phones, distinctive features, syllable structure and phonotactics, vowel harmony, and word-level prominence. Because the morphophonemics are rather extensive, alternations involving contrasting sounds and the factors that condition them may warrant a short chapter of their own. Realizational morphology will be conceptually distinguished from morphosyntax, suggesting a division of labor in the grammar outline. The morphology chapter will deal with the types of formal operations used to construct words, derivation and root compounding, inflectional categories and the organization of paradigms, and the two major lexical form classes, nouns and verbs (it is not yet clear whether whether adjectives constitute a proper form class or a subclass of verbs). The chapter on morphosyntax will be a large and important one. It is here that the grammar will describe the principles of noun classification and agreement which constitute the

most central organizing subsystem of the language. By providing numerous examples of how noun class marking works over extended stretches of discourse, the grammar will contextualize the interpenetration of phonology, morphology, and syntax which was elucidated in Dobrin 1999 within an analysis of the overall grammatical and communicative system. Within the rubric of morphosyntax such matters as valency-changing operations and the behavior of clitics will also be treated. The syntax chapter will cover topics including phrase and clause types and structuring, word order, pronominal affixes, interclausal relations, and the function of the ubiquitous discourse and time-marking particles. Given the Arapesh tendency for simple clauses with shallow embedding, it is probable that the discussion of syntax will merge with the discussion of discourse structuring and genres. Some genres, such as traditional moral exhortations, are no longer attested but can be reconstructed and exemplified using archived legacy data collected in the 1930s by Fortune. To the extent allowable by email, the grammar will be checked with a native speaker who resides in New Zealand.

Information about variation within the family will appear throughout the grammar, with the major dimensions of variation described and exemplified as an integral part of the presentation of each construction or phenomenon. Thus, for example, the geographic and lexical distribution of *l* and *r* will be discussed in the section on phonology, making it clear that no Arapesh phonological system uses the two sounds contrastively; it will also be treated in the chapter on morphosyntax insofar as differences in sounds in noun-final position are associated with differences in noun classification and agreement. The detailed descriptions presented in earlier chapters will also be summarized in a final chapter that provides a synthetic overview of the Arapesh grouping as a whole.

### **3.2 Grammar Contribution**

A grammar of Arapesh constructed in this way will provide a solid foundation on which to base hypotheses about the historical relationships that hold among Arapesh varieties; this in turn may lead to progress on the larger question of affiliations among the languages grouped together within the Torricelli phylum. In laying this foundation, the project will contribute not only to Melanesian historical linguistics, but also to a cross-disciplinary conversation regarding prehistoric human migrations in the Sepik area. Anthropological studies integrating evidence from archaeology, oral history, linguistics, biology, and human geography have produced a reconstruction of Sepik prehistory in its very broad outlines (e.g., Swadling 1984). While Torricelli populations in the Sepik appear to be ancient, recent settlement histories indicate that neighboring Ndu-speaking peoples migrated into the region following a northward trajectory, pushing back and exchanging cultural systems with Arapesh-speaking peoples with whom they came in contact (Allen 2006; Roscoe 1989, 1994). The Arapesh area constitutes a miniature “spread zone” embedded in what is otherwise one of the most “residual” regions of New Guinea (Nichols 1992), with the linguistic variation rather more pronounced in inland areas to the west than it is further east and along the coast. Arapesh oral histories corroborate the eastward population movements this implies, echoing other Torricelli-speaking peoples in this regard (Mark Donohue, personal communication). At the same time, there is a high degree of metatypy (Ross 1996, 1999) or morphosyntactic alignment between Arapesh and other Torricelli languages even with very low shared cognate rates, for example, Valman (Brown and Dryer 2006). It is hoped that genealogical and areal

factors such as these can start to be teased apart on the basis of the data analyzed in this study.

Perhaps most immediately, the grammar of Arapesh will contribute to research on the typology of noun classification devices and agreement (Corbett 1991, Aikhenvald 1999, Grinevald 2000, Senft 2000a; see also Corbett 2006:90). Complementing frequently found systems in which noun classes have a semantic basis, the noun class systems of northern Arapesh languages such as Cemaun are unusual in being organized primarily around phonological features. In most cases a noun's class, and hence its plural form, is determined by its final phonological element, usually a consonant phoneme, irrespective of the noun's meaning; thus, *nimbat* 'dog' belongs to the *t*-class, the borrowing *kap* 'cup' to the *p*-class, etc. Moreover, these class-determining sounds are copied surprisingly faithfully in agreement on syntactically related units, yielding a pattern of "literal" alliterative concord across words. This remarkable system of phonological classification and agreement highlights the organizing role phonological form can play in synchronic linguistic categorization, countersupporting the "Saussurean dogma" of modern linguistics, according to which the purpose of form is to encode or realize content, rather than content being limited by form (Nichols 1986, 1989; Aronoff 1992, 1994, 1998; Dobrin 1995, 1998, 1999a).

Arapesh noun classification is also relevant to understanding the path by which noun classes arise diachronically. It is known, for example, that classifier systems may derive from "repeater" constructions in which a copy of the noun itself is used as its classifier (Aikhenvald 2000:361-2; see also Senft 2000b:40). Arapesh noun class agreement bears an obvious resemblance to repeater classification; Arapesh also has a limited number of classifier constructions (Fortune 1942, Dobrin 1999). If it can be established that these are remnants of an earlier classifier system, it would suggest the grammaticalization of repeaters as a possible source for closed grammatical noun classes, a development which has not been attested so far.

The typologically unusual phonologically-based noun class assignment and agreement systems of Arapesh also present the community of linguists with a phenomenon that requires us to rethink fundamental theoretical assumptions about the way the core components of grammar interact. In the lexicalist spirit that has guided much research in the generative paradigm, syntax has been claimed to be "phonology-free" (Zwicky 1987; Zwicky and Pullum 1986; Pullum and Zwicky 1988; Miller, Pullum, and Zwicky 1997; Corbett 2001 provides a clear restatement). If true, this would place significant constraints on grammatical architecture. However, agreement evidence from the northern Arapesh varieties strongly suggests that a phonological system may directly influence syntactic operations, since it is a noun's phonological composition that is often the determining factor not only in selecting among agreement morphemes, but in determining their concrete realization.

#### **4. Arapesh Digital Language Archive**

In writing the grammar I will have recourse to a highly powerful tool that is currently being developed by the author in collaboration with humanities computing specialists: the Arapesh Digital Language Archive. The Archive integrates a substantial body of Arapesh linguistic data in both audio and textual format—over 15 hours of audio representing over 50 linguistic events—with a detailed descriptive catalog of the data and a rich

database of lexical and grammatical information. It is being constructed in accordance with current standards and best practice recommendations for linguistic documentation and archiving (Bird and Simons 2003, E-MELD n.d.-a,-b; see 4.5 below), providing a robust and flexible, enduring record of this endangered language.

The Archive consists of two major software components, (1) a Text and Audio Collection and (2), a Lexical and Grammatical Database, linked together in such a way as to enable the isolation and investigation of specific, grammatically relevant features in the corpus of materials from both parts of the Archive. It supports Database lookups when browsing the Text and Audio Collection, and retrieval of all instances of a given Database entry attested in the data. Users can search and sort lexical entries based on grammatical classifications; display tokens in context; and display Database contents in multiple formats and according to multiple criteria. The Archive is constructed so as to facilitate both the maintenance of existing data and the continued incorporation of new data in the form of new texts (audio, video, and/or textual), descriptive data, and markup. Work on the written grammar will thus feed back into the resources available for future research using the Archive.

As with any formalization of grammatical hypotheses, creating the Digital Language Archive requires many aspects of the grammar (e.g., the identification of grammatical categories, word boundaries, base forms, and so on) to be made precise and explicit, making the input and coding of data a direct test of the linguist's analysis. Hence, the process of creating the Archive, with the many analytical decisions this entails, contributes to the grammar. A number of my working hypotheses have been firmed up (e.g., the non-phonemic status of [ɔ]) or brought into question (the extension of alliterative agreement to first and second persons) on the basis of work on the Archive. The Archive's utility as both a form of language documentation and a resource facilitating language description (see Himmelmann 2002 on the distinction) strongly supports the scientific utility of the "documenting endangered languages" agenda being promoted so actively in linguistics today.

#### **4.1. Archive Collaborators**

The Arapesh Digital Language Archive is an interdisciplinary collaboration that brings together several areas of research interest and expertise at the University of Virginia. As the Archive's project director, the present researcher, **Lise Dobrin**, contributes a background in language description and linguistic theory, a professional and personal commitment to the Arapesh people, and the experience and material results of 15 months of field research on Arapesh in Papua New Guinea. **David Golumbia**, a scholar in English and Media Studies at UVA, contributes an interest in the cultural politics of both linguistics and computing (Golumbia 2001, 2003a, 2003b, 2004) and more than a decade's experience developing software in industry and the non-profit sector. The Archive is affiliated with the **Institute for Advanced Technology in the Humanities (IATH)**, an organization whose mission is to facilitate the application of digital technology to humanistic and social scientific research. IATH's facilities and expertise are ideal for the long-term support of work on endangered language preservation; indeed, IATH (with the University of Virginia Library) serves as one of the four hosts for the Text Encoding Initiative (TEI) Consortium, the institution responsible for the development and maintenance of the XML-based standard for the digital representation

of text. Two of IATH's Associate Directors are directly involved in the project, **Daniel Pitti** and **Worthy Martin**. They have both played central roles in the design and development of web-based database projects involving electronic thematic repositories, including the Tibetan-Himalayan Digital Library. Daniel Pitti has extensive experience applying (SGML then XML) markup and database technologies to literary, historical, and cultural studies. He is the principal architect of one established standard (Encoded Archival Description) and one emerging standard (Encoded Archival Context). Worthy Martin is a faculty member in UVA's Computer Science Department, with expertise in database and heterogeneous information structure design. The project has as a resource outside consultant **Bob Conrad**, a semi-retired SIL linguist with an unmatched understanding of Arapesh grammatical structure, dialect variation, and the kind of culturally informed encyclopedic knowledge that can be critical for interpreting vernacular discourse. Finally, the project benefits from the support of an **Advisory Committee** of scholars at UVA with expertise in the areas of nonwestern languages and digital language archiving. Committee members include Ellen Contini-Morava, Professor and Chair of Anthropology at UVA who has worked with IATH in; David Germano, Associate Professor of Religious Studies at UVA and principal architect of the Tibetan-Himalayan Digital Library (THDL n.d.); and David Sapir, a senior anthropologist at UVA who did groundbreaking descriptive and comparative work on West Atlantic languages and who is now at work on a project bringing together linguistic and folklore material from the Kujamaat Jóola language in a digital format. A letter of support from the Archive project's Advisory Committee is appended to the proposal.

#### **4.2 Text and Audio Collection**

The foundation of the digital language archive is a collection of paired digital audio files and XML-encoded transcriptions in the open-source application ELAN, a tool rapidly emerging as a standard in the field of language preservation. Developed by the Max Planck Institute for Psycholinguistics, Nijmegen, ELAN makes use of widely accepted standards for data exchange (XML). The underlying encoding links units of transcription with corresponding units in the audio files via time references. While ELAN is currently the best tool for building, maintaining, and using the Archive, the open standards it uses (WAV, XML) ensure that future access to the data does not depend on ELAN itself.

#### **4.3 Lexical and Grammatical Database**

The Lexical and Grammatical Database allows the researcher to easily find and cross-reference detailed information about a lexeme's lexical category, phonological composition, morphosyntax, morphophonemic behavior, etc., in addition to its gloss in Tok Pisin and English, and, particularly for nouns, limited information on its semantic classification. In addition, Database entries are directly linked to the text and audio files stored in the Archive through unique identifiers that relate token forms to their instantiation in specific source transcriptions, allowing the researcher to see each realization of a lexeme in multiple contexts and to locate the lexeme and associated descriptive and analytic data for any instance in a transcription (i.e., bidirectional concordance functionality). The dictionary is maintained in XML format based on a project specific schema expressed in the Relax NG syntax (ISO/IEC 19757-2:2003). XSLT transformations are used to extract tokens and source identifiers from XML

transcriptions in ELAN and the extracted data is used to interrelate the transcriptions with the dictionary entries. Expansion of the Database is intended to be ongoing, extending beyond the Digital Language Archive's active development period to incorporate new data and enable revision of previously entered data, as well as analysis, in light of new understandings.

#### **4.4 Archive Location and Support**

The Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia has extensive experience with secure long-term migration and archiving of humanities and social science data. IATH is committed to preserving and providing access to its digital collections. System administration for IATH's remote and resident servers (including regular weekly backups, security, and 24x7x365 support) is provided by the University of Virginia's Department of Information Technology and Communication (ITC). The Arapesh Digital Language Archive has a dedicated server housed at IATH, with system administrative support provided by ITC and IATH's resident systems administrator. The original analog audio recordings are archived at the Language Laboratories and Archives of the University of Chicago. The notebooks (see section 2 above) are being digitally captured as 600 dpi TIFF files for secure archiving and backup.

#### **4.5 Standards Compliance and Access**

All project collaborators are committed to working within the recommendations and standards proposed by the multi-university E-MELD (Electronic Metastructure for Endangered Languages Data) consortium, which was formed to maximize the interoperability and long-term robustness of electronically archived endangered languages data. The design of the Arapesh Digital Language Archive has been developed in coordination with E-MELD best practice outlines from the outset. The audio tapes have been digitized and stored in WAV format (48 kHz/24-bit). Audio and textual data is presented via the open source ELAN software application. Textual transcriptions maintained and displayed in ELAN are Extensible Markup Language (XML) instances based on an ELAN-specific schema. Descriptive data is being provided to the OLAC (Open Language Archives Community) OAI (Open Archives Initiative) harvesters (see <http://www.language-archives.org>). Descriptive data is based on OLAC and E-MELD recommendations and is maintained in PostgreSQL, a robust open source SQL database. Basic information and sample data will be available directly via the web, and full remote access to the archive will be available to qualified researchers via account setup by the Archive administrators.

As its public face, the Archive has a small, open-access website, designed and hosted by IATH, that will present culturally and sociolinguistically contextualized educational materials about the Arapesh language (a placeholder can be seen at <http://www.arapesh.org>). In addition to including selected photographic images, the public-facing website will present representative linguistic data (with proper rights certification), access to the Lexical and Grammatical Database (unlinked to the Text and Audio Collection as a human subjects protection), and examples of interesting grammatical features, particularly the phonologically-based Arapesh noun classification

and agreement patterns. It will include an audio component, allowing visitors to hear as well as visualize the Arapesh speech samples presented.

### **5. Plans for Future Projects based on the Grammar and Archive**

At present, Cemaun speakers have virtually no access to a written version of their language. At the initiative of village leaders, I helped native speakers of Cemaun compose vernacular versions of The Lord's Prayer and Hail Mary which are regularly recited at Catholic church services in the village. Cemaun speakers can also read Christian pamphlets that are written in the neighboring Rohwim dialect, which is highly comprehensible to Cemaun speakers. But these are a highly limited resource for vernacular literacy. They can only be used by those who are both fluent in the vernacular and literate, and such people are few. Moreover, Cemaun people experience interference from English (the language of schooling) as well as a feeling of estrangement from their own speech when they try to read them (which they typically do haltingly and aloud).

This feeling is intensified by the ubiquitous presence of features (most saliently [ʃ] instead of [tʃ], orthographic *c*) that mark the texts clearly as "Rohwim." This carries infelicitous, counter-motivating social meaning, as Cemaun and Rohwim villages were traditional enemies and remain competitive today (see Dobrin and Bashkow in press).

The documentation produced with the support of a DEL fellowship will serve as a basis for a subsequent, independent project to create printed pedagogical materials oriented toward children, for use in the Cemaun villages' vernacular preschools. These materials will draw on the Arapesh Grammar, and with it the Digital Language Archive, to produce a reverse nominal dictionary of Cemaun that incorporates grammatical information (in addition to glosses in Tok Pisin and English) in a way that graphically demonstrates the language's phonologically based noun classification patterns. The unusual reverse organization of the dictionary is suggested by the linguistic structure of Arapesh itself, and, as my experience eliciting nominal vocabulary in the field suggests, it more closely parallels the formal word-searching mechanisms used spontaneously by Arapesh speakers than does conventional alphabetical order. The reverse nominal dictionary, along with other planned educational materials written in an essentially phonemic orthography, will be particularly beneficial for Arapesh revitalization, because the logic of the pervasive noun classification system is becoming increasingly opaque as key phonological contrasts on which it depends are being lost from young people's speech and linguistic perception. This opacity contributes to further deterioration of the grammar while increasing people's sense that the native language is "too hard" to learn (see also Nekitel 1984, 1992). Laying the groundwork for such pedagogical materials is a broader impact of this language documentation project.

### **6. Work Plan**

During the 9 month fellowship period I will write those chapters of the grammar that deal with the traditional grammatical domains, focusing primarily on Cemaun: the phonology and morphophonemics (September – November 2007); realizational morphology (form classes and their attendant grammatical categories) and morphosyntax, including noun class agreement (December 2007 – February 2008); and syntax and discourse organization (March – May 2008). I have already carried out extensive fieldwork and library research on these topics, and have many preliminary analyses on which to build.

These are moreover topics which can be queried directly using the Archive, laying the foundation for the comparative family overview to be written later on.

Work on the Digital Language Archive will be substantially complete in advance of the start of the fellowship period in September 2007. That project is up to date with respect to its own timeline as the first year's work on it draws to a close: The audio recordings have been digitized and edited; descriptive data has been compiled and is being elaborated on an ongoing basis; text transcription in ELAN is underway with the help of two capable undergraduate assistants now entering their second year associated with the project; and work on the Lexical and Grammatical Database is transitioning out of an intense planning and prototyping phase and moving into an intense data entry phase.

The other chapters of the grammar—the ethnographic and historical introduction and the important concluding chapter on relationships across the family—will be written subsequently to the fellowship period with the help of regular summer faculty research funding from the University of Virginia Vice President for Research (summers 2008 and 2009). Separate funding will also be sought for preparation of the maps. By the end of summer 2009 the manuscript will be completed and ready to submit for publication as a scholarly reference book, the concrete final product of this work.



## References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2002. Language Endangerment in the Sepik Area of Papua New Guinea. *Materials of The Fourth International Conference on Endangered Languages of the Pacific Rim*. Kyoto, 23-25 November 2002, 1-53.
- Alungum, John., Robert J. Conrad, and Joshua Lukas. 1978. Some Muhiang Grammatical Notes. *Workpapers in Papua New Guinea Languages 25: Miscellaneous Papers on Dobu and Arapesh*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Aronoff, Mark. 1992. Noun Classes in Arapesh. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 21-32. Dordrecht: Kluwer.
- Aronoff, Mark. 1994. *Morphology by Itself*. Linguistic Inquiry Monograph 22. Cambridge: MIT Press.
- Aronoff, Mark. 1998. Gender Agreement as Morphology. In Geert Booij, Angela Ralli, and Sergio Scalise (eds.), *Proceedings of the First Mediterranean Conference of Morphology*, 7-18. Patras: University of Patras.
- Bashkow, Ira, and Lise M. Dobrin. In prep. The Great Arc of Human Possibilities and a Small Circle of Friends: The Social Microcosm of Mead's *Sex and Temperament*. For publication in *History of Anthropology*.
- Bird, Steven and Gary Simons. 2003. Seven Dimensions of Portability for Language Documentation and Description. *Language* 79/3:557-582.
- Brown, Lea, and Matthew Dryer. 2006. The Verbs for *And* in Walman, a Torricelli language of Papua New Guinea. Paper presented at the Linguistic Society of America, Albuquerque, NM, January 2006.
- Conrad, Robert J. 1978. A Survey of the Arapesh Language Family of Papua New Guinea. *Workpapers in Papua New Guinea Languages 25: Miscellaneous Papers on Dobu and Arapesh*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Conrad, Robert J. 1987. Kinds of Information in Bukiyip Oral Narrative Discourse. *Language and Linguistics in Melanesia* 16:23-40.
- Conrad, Robert J. 1996. More on Arapesh Noun Classes. Unpublished ms.
- Conrad, Robert J., and Kepas Wogiga. 1991. An Outline of Bukiyip Grammar. *Pacific Linguistics* C 113.
- Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.

- Corbett, Greville. 2001. Why Linguists Need Languages. In Luisa Maffi (ed.), *On Biocultural Diversity: Linking Language, Knowledge, and the Environment*. Washington and London: Smithsonian Institution Press.
- Corbett, Greville. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Crystal, David. 2000. *Language Death*. Cambridge: Cambridge University Press.
- Dobrin, Lise M. 1995. Theoretical Consequences of Literal Alliterative Concord. In . In Dainora et al. (eds.), *The Proceedings of the 31st Regional Meeting of the Chicago Linguistic Society, Volume 1: The Main Session*, 127-142. Chicago: Chicago Linguistic Society
- Dobrin, Lise M. 1998. The Morphosyntactic Reality of Phonological Form. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1997*, 59-81. Dordrecht: Kluwer.
- Dobrin, Lise M. 1999a. *Phonological Form, Morphological Class, and Syntactic Gender: The Noun Class Systems of Papua New Guinea Arapeshan*. Unpublished doctoral dissertation, University of Chicago.
- Dobrin, Lise M. 1999b. Lexical Splitting in the Kinship Vocabulary of the Buki (Arapesh) Languages. *Language and Linguistics in Melanesia* 28:99-119.
- Dobrin, Lise M. 2001. Arapesh. In *Facts About the World's Languages: An Encyclopedia of the World's Languages, Past and Present*. Jane Garry and Carl Rubino (eds.), 33-38. New York: H. W. Wilson.
- Dobrin, Lise M. 2006a. Noun Classification in Weru. Manuscript completed, submitted, and accepted by K. Sumbuk (ed.) for a volume honoring the memory of Papua New Guinean linguist Otto Nekitel. University of Papua New Guinea.
- Dobrin, Lise M. 2006b. From Linguistic Elicitation to Eliciting the Linguist: Lessons in Community Empowerment from Melanesia. Unpublished manuscript.
- Dobrin, Lise M. and Ira Bashkow. In press. 'Pigs for Dance Songs': Reo Fortune's Empathetic Ethnography of the Arapesh Roads. In R. Darnell and F. Gleach (eds.), *Histories of Anthropology Annual*, vol. 2. Lincoln, NE: University of Nebraska Press.
- Dobrin, Lise M. and Ira Bashkow. In prep. Arapesh Warfare: Mead and Fortune's Clash of Ethnographic Temperament. For publication in *History of Anthropology*.
- E-MELD. n.d.-a. Archival Formats for Audio Digitization. Best Practices recommendations. Online at <<http://cf.linguistlist.org/cfdocs/emeld/school/classroom/audio/archive.html>>.
- E-MELD. n.d.-b. What Are Best Practices? Best Practices recommendations. Online at <<http://cf.linguistlist.org/cfdocs/emeld/school/what.html>>.
- Foley, William A. 1986. *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
- Foley, William A. 2000. The Languages of New Guinea. *Annual Review of Anthropology* 29: 357-404.

- Fortune, Reo F. 1939. Arapesh Warfare. *American Anthropologist* 41:22-41.
- Fortune, Reo F. 1942. *Arapesh*. Publications of the American Ethnological Society 19. New York: J. J. Augustin.
- Fraser, Norman M. and Greville G. Corbett. 1995 ms. *Defaults in Arapesh*. University of Surrey.
- Gerstner, P. Andreas. 1963. *Grammatik der Alubänsprache*. Micro-Bibliotheca Anthropos 37. Bonn: St. Augustin.
- Golumbia, David. 2001. Hierarchies, Evidentials, and the History of Constituency. In Sunyoung Oh and Naomi Sawai, eds., *Proceedings of WSCLA 6: The Workshop on Structure and Constituency in Languages of the Americas*, 51-64. Vancouver, BC: University of British Columbia Working Papers in Linguistics.
- Golumbia, David. 2003a. Computation, Gender, and Human Thinking. *differences: A Journal of Feminist Cultural Studies* 14/2:27-48.
- Golumbia, David. 2003b. Metadiversity: On the Unavailability of Alternatives to Information. *Electronic Book Review* 3.0. Online at <[http://www.electronicbookreview.com/v3/servlet/ebr?essay\\_id=golumbiaaltx&command=view\\_essay](http://www.electronicbookreview.com/v3/servlet/ebr?essay_id=golumbiaaltx&command=view_essay)>.
- Golumbia, David. 2004. The Interpretation of Nonconfigurationality. *Language & Communication: An Interdisciplinary Journal* 24:1.
- Gordon, Raymond G., ed. 2005. *Ethnologue: Languages of the World*. 15<sup>th</sup> Edition. Dallas: SIL International. Online at <<http://www.ethnologue.com>>
- Grinevald, Colette. 2000. A Morphosyntactic Typology of Classifiers. In Gunter Senft (ed.), *Nominal Classification*, 50-92. Cambridge: Cambridge University Press.
- Himmelmann, Nikolaus P. 2002. Documentary and Descriptive Linguistics (full version). In Osamu Sakiyama and Fubito Endo (eds.), *Lectures on Endangered Languages* 5, 37-83. Kyoto: Endangered Languages of the Pacific Rim.
- Kulick, Don. 1992. *Language Shift and Cultural Reproduction: Socialization, Self, and Sycretism in a Papua New Guinean Village*. Studies in the Social and Cultural Foundations of Language 14. Cambridge: Cambridge University Press.
- Laycock, Donald. 1973. Sepik Languages: Checklist and Preliminary Classification. *Pacific Linguistics* B25.
- Laycock, Donald. 1975. The Torricelli Phylum. In S. Wurm (ed.), *New Guinea Area Languages and Language Study, 1: Papuan Languages and the New Guinea Linguistic Scene*. Pacific Linguistics C38:767-80.
- Leavitt, Stephen C. 1989. *Cargo, Christ, and Nostalgia for the Dead: Themes of Intimacy and Abandonment in Bumbita Arapesh Social Experience*. Doctoral dissertation, University of California, San Diego.

- Leavitt, Stephen C. 1995. Political Domination and the Absent Oppressor: Images of Europeans in Bumbita Arapesh Narratives. *Ethnology* 34/3:177-189.
- Leavitt, Stephen C. 1997. Cargo Beliefs and Religious Experience. In J. P. Spradley and D. McCurdy (eds.), *Conformity and Conflict*. New York: Longman.
- Leavitt, Stephen C. 2000. The Apotheosis of White Men? A Reexamination of Beliefs about Europeans as Ancestral Spirits. *Oceania* 70:304-323.
- Mead, Margaret. 1935. *Sex and Temperament in Three Primitive Societies*. New York: Morrow.
- Mead, Margaret. 1938. *The Mountain Arapesh. I. An Importing Culture*. New York: American Museum of Natural History.
- Mihalic, Rev. Francis. 1996. The Medium of the Message. *Verbum SVD* 37:167-176.
- Miller, Philip H., Geoffrey K. Pullum, and Arnold M. Zwicky. 1997. The Principle of Phonology-Free Syntax: Four Apparent Counterexamples in French. *Journal of Linguistics* 33:67-90.
- Nekitel, Otto. 1984. What Is Happening to Our Vernaculars? *Bikmaus* 2:89-98.
- Nekitel, Otto. 1985. *Sociolinguistic Aspects of Abu', A Papuan Language of the Sepik Area, Papua New Guinea*. Unpublished doctoral dissertation, Australian National University.
- Nekitel, Otto. 1986. A Sketch of Nominal Concord in Abu' (An Arapesh Language). *Papers in New Guinea Linguistics No. 24, Pacific Linguistics A70*.
- Nekitel, Otto. 1992. Culture Change, Language Change: The Case of Abu' Arapesh, Sandaun Province, Papua New Guinea. In Tom Dutton, ed., *Culture Change, Language Change—Case Studies from Melanesia. Pacific Linguistics C-120*, 49-58.
- Nichols, Johanna. 1986. On Form and Content in Typology. In W. P. Lehmann (ed.), *Language Typology 1985*. (Current Issues in Linguistic Theory 47), 141-162. Amsterdam: John Benjamins.
- Nichols, Johanna. 1989. The Nakh Evidence for the History of Gender in Nakh-Daghestanian. In Howard Aronson (ed.), *The Non-Slavic Languages of the USSR: Linguistic Studies*, 158-175. Chicago: Chicago Linguistic Society.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nidue, Joseph. A. 1990. Language Use in a New Guinea Village: A Triglossic Profile of Makopin I. *Language and Linguistics in Melanesia* 21:47-69.
- Papua New Guinea National Statistical Office. 1994. *1994 Report on the 1990 National Population and Housing Census in East Sepik Province*. Port Moresby: National Statistical Office.
- Pullum, Geofferey. K. and Arnold Zwicky. 1988. The Syntax-Phonology Interface. In Frederick J. Newmeyer, ed., *Linguistics: The Cambridge Survey. I: Linguistics Theory: Foundations*,. Cambridge: Cambridge University Press.

- Romaine, Suzanne. 1992. *Language, Education, and Development: Urban and Rural Tok Pisin in Papua New Guinea*. Oxford: Clarendon Press.
- Roscoe, Paul B. 1989. The Flight from the Fen: The Prehistoric Migrations of the Boiken of the East Sepik Province, Papua New Guinea. *Oceania* 60/2:139-154.
- Roscoe, Paul B. 1994. Who Are the Ndu? Ecology, Migration, and Linguistic and Cultural Change in the Sepik Basin. In Andrew J. Strathern and Gabriele Stürzenhoffer (eds.), *Migration and Transformations: Regional Perspectives on New Guinea*, 49-84. Pittsburgh: University of Pittsburgh Press.
- Ross, Malcolm. 1996. Contact-Induced Change and the Comparative Method: Cases from Papua New Guinea. In Mark Durie and Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, 180-217. New York: Oxford University Press.
- Ross, Malcolm. 1999. Exploring Metatypy: How Does Contact-Induced Typological Change Come About? Keynote address given at the Australian Linguistic Society's annual meeting, Perth, 1999. Online at <[rspas.anu.edu.au/linguistics/mdr/Metatypy.pdf](http://rspas.anu.edu.au/linguistics/mdr/Metatypy.pdf)>
- Ross, Malcolm. 2005a. Pronouns as a Preliminary Diagnostic for Grouping Papuan Languages. In Andrew Pawley, Robert Attenborough, Jack Golson, and Robin Hide (eds.), *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, 15-65. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2005b. Pronouns as preliminary evidence for grouping Papuan languages. To appear in Andrew Pawley, Malcolm Ross and Meredith Osmond (eds.), *Papuan languages and the Trans New Guinea phylum*. Canberra: Pacific Linguistics.
- Senft, Gunter. 2000a. *Systems of Nominal Classification*. Cambridge: Cambridge University of Press.
- Senft, Gunter. 2000b. What Do We Really Know about Nominal Classification Systems? In G. Senft (ed.), *Systems of Nominal Classification*, 11-49. Cambridge: Cambridge University of Press.
- Swadling, Pamela. 1984. Sepik Prehistory. In *Sepik Heritage: Tradition and Change in Papua New Guinea*. Nancy Lutkehaus, Christian Kaufmann, William E. Mitchell, Douglas Newton, Lita Osmundson, and Meinhard Schuster (eds.), 71-86. Durham, N.C.: Carolina Academic Press.
- THDL. n.d. The Tibetan and Himalayan Digital Library and Tibetan Collaborative Dictionaries. Online at <<http://iris.lib.virginia.edu/tibet/reference/dictionary.html>> and <<http://www.thdl.org>>.
- Tuzin, Donald F. 1976. *The Ilahita Arapesh: Dimensions of Unity*. Berkeley: University of California Press.
- Tuzin, Donald F. 1980. *The Voice of the Tambaran: Truth and Illusion in Ilahita Arapesh Religion*. Berkeley: University of California Press.
- Tuzin, Donald F. 1997. *The Cassowary's Revenge*. Chicago: University of Chicago Press.

- Wurm, Stephen A. 1998. Methods of Language Maintenance and Revival, with Selected Cases of Language Endangerment in the World. In Kazuto Matsumora (ed.), *Studies in Endangered Languages*. Tokyo: Hituzi Syobo.
- Wurm, Stephen A., Peter Mühlhäusler, and Darrel T. Tryon (eds.). 1996. *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas*. Trends in Linguistics Documentation 13.
- Zwicky, Arnold M. 1987. *Phonologically Conditioned Agreement and Purely Morphological Features*. Report SRC-87-06 of the Syntax Research Center, University of California, Santa Cruz.
- Zwicky, Arnold M. and Geoffrey K. Pullum. 1986. The Principle of Phonology-Free Syntax: Introductory Remarks. *Ohio State University Working Papers in Linguistics* 32:63-91.