

Melanesian Languages on the Edge of Asia: Challenges for the 21st Century

edited by

Nicholas Evans and Marian Klamer



PUBLISHED AS A SPECIAL PUBLICATION OF LANGUAGE DOCUMENTATION & CONSERVATION

LANGUAGE DOCUMENTATION & CONSERVATION
Department of Linguistics, UHM
Moore Hall 569
1890 East-West Road
Honolulu, Hawai'i 96822
USA

<http://nflrc.hawaii.edu/ldc>

UNIVERSITY OF HAWAI'I PRESS
2840 Kolowalu Street
Honolulu, Hawai'i
96822-1888
USA

© All texts and images are copyright to the respective authors, 2012
All chapters are licensed under Creative Commons Licenses

Cover design by Susan Ford incorporating a photograph by Darja Hoenigman

Library of Congress Cataloging in Publication data
ISBN 978-0-9856211-2-4

<http://hdl.handle.net/10125/4557>

Contents

<i>Contributors</i>		iv
1.	Introduction <i>Nicholas Evans and Marian Klamer</i>	1
2.	The languages of Melanesia: Quantifying the level of coverage <i>Harald Hammarström and Sebastian Nordhoff</i>	13
3.	Systematic typological comparison as a tool for investigating language history <i>Ger Reesink and Michael Dunn</i>	34
4.	Papuan-Austronesian language contact: Alorese from an areal perspective <i>Marian Klamer</i>	72
5.	Even more diverse than we had thought: The multiplicity of Trans-Fly languages <i>Nicholas Evans</i>	109
6.	Projecting morphology and agreement in Marori, an isolate of southern New Guinea <i>I Wayan Arka</i>	150
7.	‘Realis’ and ‘irrealis’ in Wogeo: A valid category? <i>Mats Exter</i>	174
8.	From mountain talk to hidden talk: Continuity and change in Awiakay registers <i>Darja Hoenigman</i>	191
9.	Cross-cultural differences in representations and routines for exact number <i>Michael C. Frank</i>	219
10.	Keeping records of language diversity in Melanesia: The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) <i>Nicholas Thieberger and Linda Barwick</i>	239

Contributors

I WAYAN ARKA is affiliated with the Australian National University (as a Fellow in Linguistics at School of Culture, History and Language, College of Asia and the Pacific) and Udayana University Bali (English Department and Graduate Program in Linguistics). His interests are in descriptive, theoretical and typological aspects of Austronesian and Papuan languages of Indonesia. Wayan is currently working on a number of projects: NSF-funded research on voice in the Austronesian languages of eastern Indonesia (2008-2011), ARC-funded projects for the development of computational grammar for Indonesian (2008-2011) and the Languages of Southern New Guinea (2011-2014).

LINDA BARWICK is Associate Professor in the School of Letters, Art and Media at the University of Sydney, and Director of the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). She trained in Italian language and dialectology with Antonio Comin at Adelaide's Flinders University, and in ethnomusicology with the late Catherine Ellis at the University of Adelaide. Since then she has undertaken ethnomusicological fieldwork in Italy, Australia and the Philippines, and has published widely on Australian Indigenous Music as well as continuing her engagement with Italian traditional music.

MICHAEL DUNN is an evolutionary linguist with a background in linguistic typology and language description. His current research takes a quantitative, phylogenetic approach to language change and linguistic diversity. Recent projects have addressed questions of coevolution of typological parameters, as well as the ecological and social factors influencing language change. He has also worked on the classification and prehistory of Papuan languages, and the phylogeography of Indo-European and Aslian language. He leads the Max Planck Research Group "Evolutionary Processes in Language and Culture" at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands.

NICHOLAS EVANS is Distinguished Professor of Linguistics in the College of Asia/Pacific, Australian National University. He has carried out wide-ranging fieldwork on traditional languages of northern Australia and southern Papua New Guinea. The driving interest of his work is the interplay between documenting endangered languages and the many scientific and humanistic questions they can help us answer. In addition to grammars of two Aboriginal languages, Kayardild and Bininj Gun-wok, dictionaries of Dalabon and Kayardild, edited collections on a number of linguistic topics, and over 120 scientific papers, he recently published the widely-acclaimed crossover book *Dying Words: Endangered Languages and What They Have to Tell Us* which sets out a broad program for engaging with the world's dwindling linguistic diversity. He has also worked as a linguist, interpreter and anthropologist in two Native Title claims in northern Australia, and as a promotor of Aboriginal art by the Bentinck Island women's artists.

MATS EXTER studied General Linguistics, Historical-Comparative Linguistics, Phonetics and Finnish Studies at the University of Cologne. He has held post-doctoral research and teaching positions at the Universities of Bonn and Düsseldorf. His work focuses on language description and documentation, experimental phonetics, laboratory phonology, and morphosyntactic typology. He has conducted fieldwork on Wogeo, an Austronesian language of Papua New Guinea, where he has done descriptive research on the phonological and morphosyntactic structure as well as producing a collection of traditional texts. More recently, he has conducted fieldwork on Nluu, a Tuu (formerly ‘Southern Khoisan’) language of South Africa, focusing on the phonetic and phonological structure of the language.

MICHAEL C. FRANK is Assistant Professor of Psychology at Stanford University. He received his PhD from the Department of Brain and Cognitive Sciences at MIT in 2010 and now heads the Language and Cognition Lab, which uses experimental, observational, and computational methods to study language acquisition and language use. He is broadly interested in the reciprocal interactions of language and cognition: both how languages affect the thoughts of their users and how the structure of cognition (especially social cognition) facilitates the acquisition of language in infants and children.

HARALD HAMMARSTRÖM studies linguistics and computer science at Uppsala University (Sweden) and went on to do a PhD in computational linguistics at Chalmers University (Sweden). His interests are linguistic typology (especially numeral systems), Papuan languages, language classification and computational techniques for modeling the diversity of human languages. He is currently working as a PostDoc at the MPI EVA (Germany) and Radboud University (the Netherlands) on the documentation of the Papuan language Mor and on areal linguistics in South America.

DARJA HOENIGMAN is a PhD candidate in Anthropology at The Australian National University, working among the Awiakay, a community of 300 people living in Kanjime village in East Sepik Province of Papua New Guinea. In her current project she is investigating socio-cultural continuity and change in Kanjime and its relation to linguistic registers. In studying these speech varieties and their relation to the overall social scene, she brings together linguistic anthropology and ethnographic filmmaking.

MARIAN KLAMER teaches at Leiden University and has done primary fieldwork on a dozen Austronesian and Papuan languages in Indonesia over the last two decades. Her research centres on language description and documentation, typology, and historical and contact-induced language change. Her publications include *A grammar of Kambera* (1998), *A grammar of Teiwa* (2010), *A short grammar of Alorese* (2011), and over 50 articles on a variety of topics. Klammer has coordinated numerous research projects on languages of Indonesia, including the NOW-VIDI project ‘Linguistic variation in Eastern Indonesia’ (2002–2007) and the EuroBABEL project ‘Alor Pantar languages: Origins and theoretical impact’ (2009–2012), funded by the European Science Foundation

SEBASTIAN NORDHOFF is a postdoctoral researcher at the Max Planck Institute for Evolutionary Anthropology in Leipzig. He specializes in language contact and language change and the interface of language description and documentation on the one hand and electronic publication on the other. He is a member of the working group on Open Data in Linguistics of the Open Knowledge Foundation, where he works on integrating typological data into the Linguistic Linked Open Data Cloud.

GER REESINK studied psychology at the University of Utrecht, after which he spent 15 years in Papua New Guinea under the auspices of SIL. Finishing his affiliation with SIL, he spent more than 15 years at Leiden University, mostly doing research on the languages of the Bird's Head of Papua province, Indonesia. Since 2002 he has been a postdoctoral researcher at the Radboud University and the Max Planck Institute for Psycholinguistics at Nijmegen, involved in typological research of Papuan and Austronesian languages in order to trace the ancient history of genealogical and contact relations.

NICK THIEBERGER is an ARC QEII Fellow at the University of Melbourne. He recorded Paakantyi (NSW) speakers in the early 1980s and then worked with Warnman speakers (Western Australia) when he was setting up the Wangka Maya language centre in Port Hedland. He built the Aboriginal Studies Electronic Data Archive (ASEDA) at AIATSIS in the early 1990s and then was at the Vanuatu Cultural Centre from 1994–1997. He wrote a grammar of South Efate, a language from central Vanuatu, which was the first grammar to cite a digital corpus of recordings in all example sentences and texts. In 2003 he helped establish the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). He taught in the Department of Linguistics at the University of Hawai'i (2008–2010). He is a co-director of the Resource Network for Linguistic Diversity (RNLD) and the editor of the journal *Language Documentation & Conservation*. He is developing methods for the creation of reusable data from fieldwork on previously unrecorded languages and training researchers in those methods.

Introduction: Linguistic challenges of the Papuan region

Nicholas Evans

The Australian National University

Marian Klamer

Universiteit Leiden

The region where Papuan languages are spoken – centred on the Island of New Guinea, with extensions westward into Timor and the islands of eastern Indonesia, and eastward into the Solomon Islands – is at the same time the most linguistically diverse zone of the planet and the part of the logosphere.¹ It packs around 20% of the world’s languages into less than 1% of its surface area and less than 0.1% of its population. The absolute level of linguistic diversity – whether measured in sheer numbers of languages, or in terms of ‘maximal clades’ of unrelatable units – is comparable to the whole of Eurasia.

Getting the right term to describe the region of interest in this collection is a famously difficult problem. Melanesia is a little too broad – extending out to Fiji, Vanuatu and New Caledonia to the east, a little beyond the scope of the present collection, and on the other

¹ We gratefully acknowledge the support of the various bodies which supported the original conference in Manokwari from 8-10 February, 2010, under the title ‘Melanesian Languages on the Edge of Asia’: the Australia-Netherlands Research Collaboration, the Australian National University (Department of Linguistics, College of Asia and the Pacific, and the Stephen and Helen Wurm Endowment), the Max Planck Institute for Evolutionary Anthropology, and the Universitas Negeri Papua (Unipa) and the Centre for Endangered Languages Documentation (CELD) for so generously hosting the conference. We would further like to thank the various colleagues who acted as referees for the present collection (in alphabetical order): Sander Adelaar, Wayan Arka, Johan van der Auwera, Matthew Baerman, Rene van den Berg, Sonia Cristofaro, Mary Dalrymple, Philippe Grangé, Simon Greenhill, Harald Hammarström, Andy Pawley, Ger Reesink, Malcolm Ross, Alan Rumsey, Ruth Singer, Lourens de Vries, as well as to Susan Ford and Aung Si for their editorial assistance. Evans would further like to thank the Alexander von Humboldt Stiftung, through an Anneliese-Maier Forschungspreis, for financial support which assisted in some of the final production of this collection.

hand not generally including the Lesser Sunda islands in the Indonesian archipelago.

Nor are definitions in terms of language families easy to make cleanly. The Austronesian languages have wrapped New Guinea and its surrounding islands in a three thousand year embrace that is still being played out in intimate language contact with all the other languages of the region. Some of the papers here concern either Austronesian languages with significant structural resemblances to non-Austronesian languages of the region – see Exter’s paper on Wogeo – or various types of historical and typological interaction between Austronesian and non-Austronesian languages – see the papers by Reesink & Dunn and by Klamer.

For the non-Austronesian languages of Melanesia and its surrounds (excluding Australia), the collective name ‘Papuan’ has been widely used and we continue that practice here. This use, based on definition by exclusion, has hung on for want of a better term long after comparable terms like ‘Palaeosiberian’ have been abandoned, but includes upwards of forty distinct families and isolates. To get an idea of how distorting a term like this is, consider how unsatisfactory it would be to use a term like ‘Eurasian’ for the set of languages including Basque, Finnish, Georgian, Ingush, Chinese, Tamil, Cambodian, Japanese, Kurdish, Hmong, Ket, Chukchi, Burushaski and all the other non-Indo-European languages of Eurasia (where we partition off Indo-European languages only, in the same way that we partition off Austronesian languages). Yet that is arguably the level of genetic and typological diversity which we face when confronted with the full range of Papuan languages. Despite these problems, we currently have no better term, so the reader is simply cautioned to keep all these caveats in mind each time the word ‘Papuan’ is used.

Our knowledge of this exuberant linguistic cornucopia lags behind what we know about any other region of the globe. It is likely that the linguist-to-language ratio is lower here than anywhere else, and it is certain that the relative level of language documentation is lower here than anywhere else (see Hammarström and Nordhoff’s paper). The inchoate state of Papuan linguistic studies stems from many reasons. These include the recency of linguistic research in the area, the inaccessibility of many sites, the lack of relevant training organisations in the countries concerned, the fragmentation of research across national boundaries and across the academic vs missionary divide, and the general lack of prioritisation that large parts of the linguistic profession have until recently assigned to the documentation of linguistic diversity. We have put together this collection of papers as a sample of just some of the research questions, languages and approaches that currently seem particularly exciting, with the goal of raising interest in this fascinating part of the logosphere, and encouraging linguists from around the world to get involved in research in an area where there is so much just waiting to be discovered.

The present collection grows out of a conference held at the Centre for Endangered Languages Documentation (CELD) at Universitas Negeri Papua (Unipa) in Manokwari, Indonesia, in February 2010, with the support of the Australian Netherlands Research Council, the Australian National University, and the Max Planck Institute for Evolutionary Anthropology in Leipzig, all of whom we thank for their financial assistance. It is not simply a conference proceedings, however – it represents merely a selection of papers from that conference, supplemented by an additional paper to fill in gaps we thought needed coverage.

We began this introduction by continuing the well-established tradition of stating, in

an approximate and rather unquantified way, that Melanesia, and in particular the island of New Guinea, contains both the greatest concentration of linguistic diversity anywhere on earth, and the lowest level of documentation. The first paper in this volume, *The languages of Melanesia: Quantifying the level of coverage* by Hammarström & Nordhoff, adds precision to this statement by presenting relevant figures from their LangDoc database. This database aims to give comprehensive global listing of all materials existing on all languages, along with an initial, approximate metric of degree of coverage. As the authors point out, there are many shortcomings to their metric. It is relatively unambitious: a language possessing a low-quality grammar of 160 pages and no lexicon or text collection would already be placed at the highest level—well short of the modern gold standard of a Boasian trilogy supplemented by a wide variety of annotated multimedia files—and there is no measure of quality of analysis. Despite these flaws, it has the great virtue of being operationalisable and applicable to all the world's languages in a relatively automatic way, and in their paper they outline their scheme in detail as well as comparing the level of coverage for Melanesia with the rest of the world.

First, regarding the total proportion of the world's languages spoken in Melanesia, their figures count 1347 languages that are 'Melanesian' in their sense (522 Austronesian, 825 non-Austronesian) of the sub-region of Oceania extending from the Arafura Sea in the west to Fiji in the east (see figure 1 in their chapter). This is just over 20% of the world total of 6496 (living) languages on their count (see their table 7), with so-called 'Papuan languages' then making 12.7% of the world's total. In absolute terms, the number of languages in Melanesia (1,347) is almost identical to those in the whole of Eurasia (1,465), these two being surpassed only by Africa (1,986).

Second, for their assessment of level of documentation, they lump together Papuan with all Austronesian, so their figures also include the rest of Indonesia, the Philippines, Malaysia, and so on. Drawing on these figures, they draw some striking conclusions. First, in absolute terms, Papua + Austronesian has the largest number of languages with only a wordlist to their documentation (i.e. the lowest level of documentation which they recognise). The comparison with Australia, another region where professional linguistic research is relatively recent, is salutary: over 42% of Australian languages have a grammar available, compare to half that number (20.48%) for Austronesian + Papuan. Second, in relative terms, Papua + Austronesian has the lowest proportion of languages with the highest rank of description (i.e. a grammar of 150 pages or more), the highest proportion with only a wordlist or less, and the lowest average level of documentation. Third, when Austronesian and non-Austronesian languages are compared within the above categories, the non-Austronesian languages have lower levels of documentation, making their overall documentation status even lower than that for Papua + Austronesian as a whole (see their Table 6).

In the years to come it is to be hoped that LangDoc will be extended to give more accurate metrics in a number of ways – something which will be aided to the extent that more linguists heed the authors' call to put their results in the public domain. But their paper already provides a very clear quantitative basis for our claim above that the Melanesian region – and particularly the Papuan languages within it – is far and away the most linguistically diverse part of the planet, and that conversely it suffers from the lowest level of language documentation found in any quarter of the earth. The combination of

these factors is what makes the study of Melanesian languages an enormous challenge.

Before leaving this paper, we note two important future developments. Firstly it will be crucial to link some form of comprehensive database like LangDoc to actual documents so that it is possible to inspect the actual materials listed there and gradually improve the qualitative ratings through the collective efforts of world scholarship. Secondly, it is desirable that the structure of the LangDoc database allows inspection of data at a number of different geographic levels. While their present article largely treated Melanesia as an undifferentiated whole, their discussion of one geographical variable (distance from coast) shows how more finely articulated geographical characterisations can be made – so that one can compile comparable reports for geographical regions like the Sepik, Bougainville, etc.

The staggering linguistic complexity of Melanesia creates special problems for attempts to classify languages into families and subgroups, especially for efforts that try to reduce the large number of independent maximal clades (over forty on any estimate) by grouping some of them together.

Under these circumstances, the languages of Melanesia have provided a particularly important testing-ground in recent years for new methods which aim to ‘break the time barrier’ of the classical comparative method, by drawing inferences from the signal in assemblages of typological traits rather than simply in the sound-meaning pairings of the lexicon and grammatical morphology. Though controversial and still subject to fierce critique (see references in Reesink and Dunn paper), it is likely that such methods as applied to Melanesia are here to stay, at the very least as a supplement to the comparative method. Indeed, the situation in Melanesia forces historical linguists to make a virtue of necessity by driving them to develop new methods.

The article by Reesink and Dunn gives an overview of these methods, focussing on the languages of Eastern Indonesia, spoken around the Bird’s Head area. As in their other studies, a grave problem with the method is that resemblant signals can signal either shared phylogeny or areal convergence. In the central part of their paper, they consider the case of two Papuan languages of the Bird’s Head, Hatam and Meyah, which consistently cluster with the Austronesian language Biak no matter how many ‘founding lineages’ (K values) are assumed on runs of the ‘Structure’ algorithm. In this case, then, ‘it thus appears that diffusion overrides phylogeny’, as they put it.

But they then take a further step, teasing out the fifteen typological features (out of 160 altogether) which align with phylogeny rather than areality, opposing the Papuan languages Hatam and Meyah against the Austronesian language Biak – see their table 6. Does evidence like this hold the key to refining typological-suite based models so that they can filter out areal noise to find the phylogenetic signal? Obviously, if the argumentation proceeds just from a single case, as here, it risks being merely post hoc, but on the other hand it would be possible to iterate this procedure over a number of areas and small groups.

Will iterations of this type, by filtering out the more from the less diffusible over independent cases from around the world, allow us to fine-tune an algorithm like Structure by weighting the evidentiary value of different typological characters as regards to phylogeny vs areality? This will be a crucial question over the next decades of scholarship as more extensive documentation of Melanesia’s languages provides us with more information for feeding into comparative enterprises like Reesink and Dunn’s. At the same

time, their work reminds us that, to draw maximum benefit from research like theirs, our language documentations need to ensure that matched typological data is obtained – this need not entail ‘questionnaire-style’ grammars, but feature lists like those in the Appendix to their article do lay down a basic checklist of typological points which all descriptions should make sure to cover.

The next two papers each consider regions of Melanesia in which there have been complex interplays between languages belonging to quite different families, in a social environment where different types of contact appear to have played a role at different points in the past.

Marian Klamer’s *Papuan-Austronesian language contact: Alorese from an areal perspective* focuses on Alorese, an Austronesian language abutting the westernmost group of extant Papuan languages on the island of Pantar. She deduces a complex contact history comprising at least two stages played out in different locations.

The first phase, on her model, would have taken place on the island of Flores or nearby, at a time when Papuan languages were still spoken there. It is at this stage that the language ancestral to modern Lamoholot and Alor would have acquired a suite of typological features that are seen as typically ‘Papuan’ – or, more precisely, as typical of the Papuan languages of the Alor-Pantar region – including post-predicate negation, the marking of possessors, noun-locational order in locative constructions, the presence of a focus particle and the absence of a passive verb form. This ‘Papuanisation’ of proto-Lamoholot would have taken place under conditions of long-term stable contact involving preadolescents acquiring the complexities of both Papuan and Austronesian languages and melding them into a new system.

In a second phase, following the migration of Alorese speakers to Pantar and the separation this entailed from their Lamoholot cousins, a series of further changes would have taken place. Alorese contrasts drastically with Lamoholot in terms of morphological complexity. Where Lamoholot has two sets of subject affixes to the verb (prefixes for transitives, suffixes for intransitives), Alorese relies on free pronouns with all but a few frequent verbs which retain fossilised agent prefixes. And where Lamoholot has a number of derivational affixes (some productive, some lexicalised), Alorese has no derivational morphology at all – reduplication is its only productive word formation process. These differences suggest a radical process of morphological simplification in the passage from Lamoholot to Alorese. Klamer hypothesises that, in the initial stages of Alorese settlements of Pantar and Alor, Alorese-speaking men would have taken as their wives women speaking a number of different Papuan languages of the inland. Entering the speech community as adults they would have learned a simplified form of Alorese, jettisoning almost all of its morphology. The contact between Alorese and local Papuan languages, however, was neither prolonged nor consistent at this stage. The number of loanwords from local Papuan languages is relatively low (only 14 Alorese terms out of a 270 word-list have a known Papuan source) and is moreover distributed evenly across the different Papuan languages of the locality. This suggests a number of relatively weak contacts and no stable pattern of bilingual contact.

This case study illustrates a type of multi-phase contact scenario likely to have been played out between Austronesian and Papuan speakers in a number of parts of Eastern Indonesia at different phases over the last two to three millennia. The very different

outcomes of the two phases posited in Klamer's model are a salutary reminder of the social and linguistic complexity that must have been involved between two groups who would have been demographically equally poised and interdependent in many ways. At the same time, as Klamer points out, it is only a reconstruction, and we would be on much firmer ground if we were able to draw on contemporary sociolinguistic studies of the types of interaction – social and linguistic – that are occurring between groups along the Papuan-Austronesian interaction zone. As with so many of the questions raised in this issue, the time for this sort of study is running out fast, as the presence of an alternative lingua franca (e.g. Indonesian) radically alters the type of linguistic interaction between such groups.

From Nusa Tenggara we then move east to the Southern New Guinea region, the focus of Nicholas Evans' *Even more diverse than we thought: The multiplicity of Trans-Fly languages*. In contrast to the Austronesian-Papuan interactions in the preceding two articles, here the interactions are between various unrelated Papuan groups. Southern New Guinea is an intriguing zone, of great diversity, about which our level of knowledge dips even lower than the norms for elsewhere in Melanesia.

The Southern New Guinea region is essentially a nucleus of several small language families surrounded by Trans-New Guinea languages which significantly outnumber them demographically, and which at the time of first colonial documentation tended to be far more expansive and militarised than their non-TNG counterparts. It offers an excellent opportunity for historical linguistics to study the mechanisms by which Trans-New Guinea languages have expanded into areas previously characterised by greater levels of deep phylogenetic diversity.

Nonetheless, it is clear that all languages of the region share a number of typological characteristics – to the extent that some languages, which have been classified as TNG, like Marind, pattern typologically with other Southern New Guinea languages (as well as some languages further afield, including Yeli-Dnye and Inanwatan – see Reesink and Dunn, Fig. 1, as well as discussion in footnote 4 of Evans' article.) This suggests that, even if the presence of TNG languages in Southern New Guinea results from expansion at the expense of other groups, there must have been enough stable long-term bi- or multilingualism for significant linguistic convergence to occur. The languages of the Southern New Guinea exhibit high levels of morphological complexity allied with a host of highly unusual typological features, and Evans' paper gives short sketches of two neighbouring but unrelated languages – Nen and Idi – focussing particularly on their complex verbal and case morphology. He shows that, despite the presence of some convergent features and widespread bilingualism and contact between the speakers of these two languages, there are major differences in how they organise their grammars. (Note in passing that the Reesink et al 2009 sample did not include any language from the Pahoturi River family which Idi belongs to, so it is not clear how far it would fare on their typological profile).

The Southern New Guinea case – rooted as it is in a system of marriage by sister-exchange which favours comparable demographics, interdependence, and intermarriage and multilingualism between neighbouring groups – is a clear case of how prolonged language contact can lead to areal patterns characterised by shared complexification. In illustration of this, Evans considers the way different languages in Southern New Guinea derive a three-valued number contrast. (singular, dual, plural). This is something found in virtually every language in Southern New Guinea except Marind (some languages have an

additional trial or paucal). But the exact route by which such systems are derived varies significantly from language to language. This suggests that whatever series of pathways leads to shared areal features of this type is a long and tortuous one, probably based on the slow patchwork emergence of grammatical solutions to particular semantic targets shared across languages of the region.

Staying in Southern New Guinea and sticking to the topic of grammatical number, Wayan Arka's paper *Projecting morphology and agreement in Marori, an isolate of southern New Guinea* examines issues of how to represent systems of 'constructed' grammatical number featurally, focussing on the TNG-level isolate Marori and other languages with comparable phenomena.

Marori, in line with the South New Guinea pattern described in Evans' article, constructs a three-valued number system by combining two binary values in a system of distributed exponence. However, the base features used to derive this result are different: where Nen crosses a singular vs non-singular with a dual vs non-dual distinction, Marori crosses a singular vs non-singular with a plural vs non-plural distinction. Arka's article shows how the unification of number values in Marori morphology can be derived within a model in which the features are hierarchically structured, in different ways in different languages. Thus where Marori treats dual as the number value that is neither singular nor plural – and hence relegates the dual to a derived category – the Nen feature structure builds in dual as a primary specified feature, but treats plural as a derived category that is neither singular nor dual. This model is an elegant illustration of how some cross-linguistic variability in feature structure can be built into a robust overall architecture – the presence of an overall feature structure, and of a primary singular vs non-singular cut, remain constant, but the internal makeup of the non-singular subspace differs as between Marori and Nen. The availability of differing feature architectures then makes it possible to model the differences between languages with similar sets of contrasts, but derived in different ways, within a formalism like LFG – we refer the reader to that chapter for the formal details.

The difficulties involved in lining up language-specific descriptive categories with comparative concepts are nicely illustrated, from a different theoretical perspective, in Mats Exter's article *'Realis' and 'Irrealis' in Wogeo: A valid category?* Recall that, in Reesink and Dunn's article, one of their questions (50/87, as listed in their appendix) is 'Is a distinction between realis/irrealis mood available as a morphological choice (1: present, 0: absent)?' But how do we decide what 'realis/irrealis' actually means? Wogeo offers interesting difficulties in answering this question.

Wogeo is a 'mood-dominated' Austronesian language, spoken off the north coast of PNG, with a complex verbal morphology including six prefixal and eight suffixal slots. A basic opposition is between the 'realis' and 'irrealis' forms of the pronominal prefix, illustrated by a pair like *o-lako* 'I go, I went' vs *go-lako* 'I must go, I want to go, I will go (now)'. If this were all there was to the opposition the characterisation would be fairly straightforward, but once we consider more semantically precise combinations problems arise. Wogeo has additional prefixal combinations expressing such meanings as future, tentative ('try doing X'), counterfactual ('would have done X'), proximal imperfective ('am/was doing X, nearby') and distal imperfective ('am/was doing X (further away)'), which are followed by either the realis or irrealis prefix, e.g. *m-o-lako* [FUT-1sg.realis-go] 'I will go, I can go, I may go'.

For some of these, the choice of prefix makes sense in terms of normally-characterised properties of this opposition, e.g. the realis is used with the two imperfective series. But for others, notably the counterfactual, it is the realis series that is chosen rather than the expected irrealis. Exter then goes on to consider what such cases mean for the overall enterprise of trying to define terms like realis and irrealis in cross-linguistic terms. He ends up arguing against the usefulness of a term with as broad a range as the realis-irrealis contrast – which, if accepted, raises the possibility that typological comparisons may be more successful if they work at much more semantically-specified levels where cross-linguistic comparison can be more precise.

The next two papers examine the embedding of language in its sociocultural and psychological contexts.

Darja Hoenigman's paper *From mountain talk to hidden Talk: Continuity and change in Awiakay registers* examines the diachronic sociolinguistics of special registers in Awiakay, a language of East Sepik province, and in the process throws a fascinating light on how ideologies of the need for linguistic difference intersect with high levels of metalinguistic awareness to drive a dynamic of lexical innovation. Particularly noteworthy is the continuity – in terms of utilising special registers – that holds in the face of significant change – in the form of Christian strictures against the ongoing use of some traditional registers.

Traditionally, Awiakay people used a special register, known as 'mountain talk', to protect themselves from mountain spirits when travelling up into mountain regions; this involved the substitution or avoidance of a number of lexical items. The arrival of Christianity has arrested the use of 'mountain talk', with the recognition it gives to the power of pagan spirits, and knowledge and use of this traditional register is in decline. But at the same time, another special register has come into use, *kay menda* or 'hidden talk'. Travelling outside the village to regional centres such as Wewak, especially when it is for commercial purposes which leaves the travellers vulnerable to theft and predation, is regarded as a risky business and speaking a language impenetrable to outsiders provides good security.

Though Awiakay is traditionally spoken in just one village, and would therefore normally have been incomprehensible to outsiders, the recent arrival of Tok Pisin loanwords creates chink in the armor of linguistic impenetrability. It is precisely these loanwords which get replaced in *kay menda*, through ingenious native coinages some of which have already won full acceptance in the community and others of which still include rival coinages.

Hoenigman's paper includes subtitled video footage of a journey from the village to the regional centre, during which we can witness the camouflaging processes of 'hidden talk' at work, as well as watching the rehearsal and induction of less experienced members of the party while travelling towards the destination. This is of interest not just for the topic of special registers, but more generally for our understanding of how at least some of the processes of linguistic diversification in Melanesia are driven along by very conscious and negotiated processes of change aimed at differentiating one's language from that of other groups.

Her paper concludes by surveying the parallels and differences between the new register of hidden talk and the fading old register of 'mountain talk'. Both are used in unfamiliar,

perilous territory where one goes to obtain valued items, encountering dangerous entities (mountain spirits before, rascals now) and dangers (sickness before, robbery and theft now), preventing these dangers through judicious out-of-the-ordinary language use, and predominantly involving men who are the ones travelling to the dangerous destinations. The most interesting difference, on her comparison, has to do with who is held to have created the special register. In the case of mountain talk this is attributed to ‘mountain spirits’ deep in the past, whereas in the case of ‘hidden talk’ the process of creation is still taking place, involves contemporary Awiakay individuals, and is therefore a process that is amenable to direct research on such questions as how rival innovations are selected between, which items are chosen for camouflaging, and how changes are propagated from innovative individuals to the community.

Michael Frank’s paper, *Cross-cultural differences in representations and routines for exact number*, leads us from the known diversity of Melanesian languages to the presumed but untested cognitive diversity this subtends. Beller & Bender (2008), whom he quotes in his article, observe that ‘there may be no other domain in the field of cognitive sciences where it is so obvious that language (i.e., the verbal numeration system) affects cognition (i.e., mental arithmetic).’ Combining this with the likelihood that Melanesian diversity in numeral systems (Lean 1992) is perhaps even greater, in relative and absolute terms, than in other aspects of the language systems, we have here a fascinating domain for the investigation of how linguistic diversity shapes cognitive diversity – as well as how cultural practices like different counting routines themselves select for the emergence of different types of numeral system.

Frank’s paper does not in itself begin the exciting project of investigating how Melanesian diversity in numeral systems produces (or doesn’t) significant differences in cognition. Rather, its goal is to clarify the relations between, on the one hand, how exact number is represented linguistically or through other types of representational tool such as the the Mental Abacus, and numerical cognition on the other. Frank adduces experimental evidence that linguistic systems in the form of numerals, but also non-linguistic systems in the form of the Mental Abacus, both provide a ‘cognitive technology’ enabling the online encoding and manipulation of quantity information. Frank shows that cultural exposure alone does not scaffold the manipulation of exact number, that the lack of exact numeral terms in a language impacts negatively on arithmetical manipulations, and that it is not enough to possess a language with appropriate terms but that one must also be able to access it online in order to successfully carry out arithmetical calculations.

Different numeral systems, such as the use of different bases (2, 5, 6, 10, 20) can be expected to furnish different cognitive strategies in this sense; and so would different numeral sets for different types of counted objects. This suggests that collaborative research on the impact of numeral systems on numerical cognition will yield rich results. Yet the challenges of investigating this interaction are great, and require a type of collaboration between linguists and psychologists that have been all too rare so far:

The data that lead to this conclusion could not have been gathered by the standard methods of cognitive psychology, nor by the standard methods of field linguistics. Many of the results cited here come from carefully controlled studies performed in the field with populations that possess culturally, linguistically, or cognitively

interesting numerical representations. This generalization suggests the benefits of psycholinguistic fieldwork that combines experimental design with cross-cultural or cross-linguistic populations (Frank, this volume:234).

In fact, the relevant numerical systems are rather fragile, in some cases significantly more so than other parts of the language system: ‘there are many cases where a language is not endangered, or not particularly endangered, but whose numeral systems are endangered’ (Comrie 2005). Oksapmin is a salutary Melanesian case, investigated by Saxe (1982) and then Saxe & Esmonde (2005). Though the language is still healthy (Loughnane 2007) its distinctive base-27 body count system is giving way to an English-style decimal system in a modern setting where counting is most commonly applied to money. This great fragility of numeral systems means that there is an exceedingly narrow time window for carrying out the sort of collaborative work on the impact of numeral systems on numerical cognition which is outlined in Frank’s article.

The many fascinating questions and research thrown up by the preceding articles – and even more so, the future research which we hope they will stimulate – generate an enormous amount of primary data as fieldworkers of a range of interests and nationalities record materials on numerous language varieties, in increasingly data-rich formats. But endangered data is a problem we need to take almost as seriously as endangered languages themselves – field notes and recordings may end up lost, uncatalogued, unlocatable, or degenerate on old tapes or other materials. Equally problematic are questions about where data should be housed and who should get access to it. Modern digital archives are giving us the power to address these issues in an efficient way. They have the potential to preserve huge amounts of data far into the future while allowing them to be accessible to researchers and community members from all locations. It is with the design and running of one such archive, Paradisec, that the last article, by Nicholas Thieberger and Linda Barwick, is concerned: *The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC): A resource for Melanesian linguistics*.

PARADISEC was established in Australia in 2003, by a team of researchers led by Thieberger and Barwick. It was born as a response to the challenges set out in the preceding paragraph, from an awareness that a vast body of hard-won field data was at risk of vanishing altogether – partly as a result of poor facilities in local archives (e.g. lack of air-conditioning to maintain tapes in good condition), partly through technological changes (e.g. the disappearance of machines able to read old recordings on wax cylinders, wire-recorders etc.), partly through a lack of emphasis in the field of linguistics on the primacy of documentation as opposed to theoretical debate or grammar-writing, and partly through the reluctance of individuals to make their materials available to others until they had analysed them themselves – a moment which sometimes gets overtaken by Alzheimer’s or death.

Part of their article is devoted to showing how PARADISEC works, in terms of equipment setup, backup, access, workflow and data ingestion. But Thieberger and Barwick also discuss a number of other important concerns raised in discussions of where archives like PARADISEC fit in a region characterised by such vast discrepancies between countries in terms of living standards, technology, access to digital data, and the potential value of local information. A central issue is the moral tension centred on the foreseeable

and unforeseeable uses of archived information (e.g. in establishing clan ownership of lands, rights to royalties etc.) which require graduated levels of access, on the philosophy that there should be general commitment to permanent archiving, for future safety's sake, but that communities and researchers should be supplied with technologies for regulating access where this is warranted.

Looking in the other direction, the potential for harnessing the collective knowledge of various kinds of expert through cumulative annotation of archived material by different archive users at different locations is a goal that has great potential to galvanise a more collective and interdisciplinary approach to adding commentary and interpretation to primary material through time.

The nine contributions we have outlined can do no more than give a tantalising glimpse of the challenges raised by the languages of Melanesia – for linguists and scholars in allied fields, but also for educators, communication technologists, and development agencies wanting to focus on local knowledge and expertise. Most importantly, this is a challenge of utmost interest to community members wanting to maintain the intellectual wealth held in their linguistic heritage. For them, collaborative work with linguists and others can offer new ways of integrating that heritage with other sorts of language products such as orthographies, dictionaries, grammars, text collections, and digital ethno-encyclopaedias.

For even a fraction of these challenges to be met, many things must happen. We need to attract a new generation of adventurous and capable young scholars to work in this fascinating, diverse and hospitable part of the world. We need to build capacity among local linguists and language workers in the countries where these languages are spoken, so as to reverse the drastic current imbalance between where Melanesian languages are spoken and where future researchers can receive advanced training in how to study them. CELD, the Centre for Endangered Language Documentation in Manokwari, which hosted the conference where most of the papers here were presented, is a promising step in this direction.

There need to be many other developments like this, and international funding agencies need to be convinced that language diversity is a resource, not a handicap. This is particularly relevant at a juncture when key sources of international research funding over the last two decades (the Volkswagenstiftung's DoBeS program, the Hans Rausing Endangered Languages Program, the NWO *Bedreigde Talen* program and the ESF EuroBABEL program) are drawing to a close, or have already. There is vast potential in such new approaches as BOLD or Basic Oral Language Documentation (<http://www.boldpng.info/iwlp>) and mobile-phone based crowd-sourcing to assist the data-gathering process. But the need for long-term traditional fieldwork drawing on the knowledge of linguists who learn the languages and cultures on-site will remain fundamental. Finally, while there will always be some divergence of interest between missionary organisations and academically-motivated researchers, the vast extent of missionary enterprises through Melanesia means that the potential for fruitful collaborative work is vast, given goodwill on both sides. An important recent initiative is the reestablishment of the journal *Language and Linguistics in Melanesia*, now as an open-access on-line journal (<http://www.langlxmelanesia.com/>), as a forum for publishing peer-reviewed research and book reviews on the languages of Melanesia.

As can be seen from these considerations, and the fact that almost every paper in this collection is an early step in a new research path, the study of Melanesia's languages offers abundant opportunities to make new discoveries. We hope that in the collection of papers gathered here you will find material that invites you into an engaged and diverse international community of scholars dedicated to advancing our understanding of a linguistic territory that is arguably the least charted on earth.

REFERENCES

- Beller, S. & A. Bender. 2008. The limits of counting: Numerical cognition between evolution and culture. *Science* 319. 213-215.
- Comrie, Bernard. 2005. Endangered numeral systems. In Jan Wohlgemuth & Tyko Dirksmeyer (eds.), *Bedrohte Vielfalt: Aspekte des Sprach(en)tods [Endangered Diversity: Aspects of Language Death]*, 203-230. Berlin: Weissensee Verlag.
- Frank, Michael C. This volume. Cross-cultural differences in representations and routines for exact number.
- Haspelmath, Martin. 2009. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3). 663-687.
- Lean, G. 1992. Counting systems of Papua New Guinea and Oceania. Lae: Papua New Guinea University of Technology doctoral dissertation.
- Loughnane, Robyn. 2007. A grammar of Oksapmin. Melbourne: University of Melbourne PhD thesis.
- Saxe, G. B., 1982. Developing forms of arithmetical thought among the Oksapmin of Papua New Guinea. *Developmental Psychology* 18(4). 583-594.
- Saxe, G. B. & Esmonde, I. 2005. Studying cognition in flux: A historical treatment of *fu* in the shifting structure of Oksapmin mathematics. *Mind, Culture, and Activity* 12(3/4). 171-225.

Nicholas Evans
nicholas.evans@anu.edu.au

Marian Klamer
M.A.F.Klamer@hum.leidenuniv.nl