

CATALOGUING THE WORLD'S ENDANGERED LANGUAGES



Edited by
Lyle Campbell and Anna Belew



CATALOGUING THE WORLD'S ENDANGERED LANGUAGES

Cataloguing the World's Endangered Languages brings together the results of the extensive and influential *Catalogue of Endangered Languages (ELCat)* project.

Based on the findings from the most extensive endangered languages research project, this is the most comprehensive source of accurate information on endangered languages. The book presents the academic and scientific findings that underpin the online *Catalogue*, located at www.endangeredlanguages.com, making it an essential companion to the website for academics and researchers working in this area.

While the online *Catalogue* displays much data from the *ELCat* project, this volume develops and emphasizes aspects of the research behind the data and includes topics of great interest in the field, not previously covered in a single volume.

Cataloguing the World's Endangered Languages is an important volume of particular interest to academics and researchers working with endangered languages.

Lyle Campbell is Professor of Linguistics at the University of Hawai'i at Mānoa and Director of the Endangered Languages Project 2009–2016.

Anna Belew is a PhD candidate at the University of Hawai'i at Mānoa and has been Project Coordinator for the Endangered Languages Project.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CATALOGUING THE WORLD'S ENDANGERED LANGUAGES

*Edited by
Lyle Campbell and Anna Belew*

First published 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

*Routledge is an imprint of the Taylor & Francis Group,
an informa business*

© 2018 selection and editorial matter, Lyle Campbell and Anna Belew,
individual chapters, the contributors

The right of Lyle Campbell and Anna Belew to be identified as the authors
of the editorial material, and of the authors for their individual chapters,
has been asserted in accordance with sections 77 and 78 of the
Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or
utilised in any form or by any electronic, mechanical, or other means, now
known or hereafter invented, including photocopying and recording, or in
any information storage or retrieval system, without permission in writing
from the publishers.

Trademark notice: Product or corporate names may be trademarks or
registered trademarks, and are used only for identification and explanation
without intent to infringe.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data

Names: Campbell, Lyle editor author. | Belew, Anna editor author. |
Endangered Languages Project (Website)

Title: Cataloguing the world's endangered languages /
edited by Lyle Campbell and Anna Belew.

Description: Milton Park, Abingdon, Oxon ; New York, NY :
Routledge, 2018. | Collected research based on information from the
Catalogue of endangered languages, available via the Endangered
Languages Project website. | Includes bibliographical references and index.

Identifiers: LCCN 2017032500 | ISBN 9781138922082 (hardback :
alk. paper) | ISBN 9781315686028 (ebook)

Subjects: LCSH: Endangered languages—Electronic information
resources. | Catalogue of endangered languages (Online)

Classification: LCC P40.5.E53 C38 2018 | DDC 400—dc23

LC record available at <https://lcn.loc.gov/2017032500>

ISBN: 978-1-138-92208-2(hbk)

ISBN: 978-1-315-68602-8(ebk)

Typeset in Times New Roman
by Florence Production Ltd, Stoodleigh, Devon, UK

Visit the companion website at www.endangeredlanguages.com

CONTENTS

<i>List of contributors</i>	vii
<i>Acknowledgements</i>	viii
1 Introduction: Why catalogue endangered languages? <i>Lyle Campbell and Anna Belew</i>	1
2 Naming convention and challenges <i>Yen-ling Chen and Lyle Campbell</i>	15
3 Language classification and cataloguing endangered languages <i>Russell Barlow and Lyle Campbell</i>	23
4 Language extinction then and now <i>Anna Belew and Sean Simpson</i>	49
5 The Language Endangerment Index <i>Nala H. Lee and John R. Van Way</i>	66
6 New knowledge produced by the <i>Catalogue of Endangered Languages</i> <i>Lyle Campbell and Eve Okura</i>	79
7 The world's endangered languages and their status <i>Anna Belew, Yen-ling Chen, Lyle Campbell, Russell Barlow, Bryn Hauk, Raina Heaton, and Stephanie Walla</i>	85

Contents

8	How the <i>Catalogue of Endangered Languages</i> serves communities whose languages are at risk <i>Raina Heaton and Sean Simpson</i>	250
9	Triage: Setting priorities for endangered language research <i>Bryn Hauk and Raina Heaton</i>	259
	<i>Index</i>	305

CONTRIBUTORS

Russell Barlow, University of Hawai‘i at Mānoa
barlowr@hawaii.edu

Anna Belew, University of Hawai‘i at Mānoa
belew@hawaii.edu

Lyle Campbell, University of Hawai‘i at Mānoa
lylecamp@hawaii.edu

Yen-ling Chen, University of Hawai‘i at Mānoa
yenling@hawaii.edu

Bryn Hauk, University of Hawai‘i at Mānoa
bhauk@hawaii.edu

Raina Heaton, University of Oklahoma
rainaheaton@gmail.com

Nala H. Lee, National University of Singapore
nala.lee@gmail.com

Eve Okura, University of Hawai‘i at Mānoa
eveokura@gmail.com

Sean Simpson, Georgetown University
sss237@georgetown.edu

John R. Van Way, University of Hawai‘i at Mānoa
jvanway@gmail.com

Stephanie Walla, Eastern Michigan University
stmawa@gmail.com

ACKNOWLEDGEMENTS

The research for the *Catalogue of Endangered Languages* project was funded by a grant from the National Science Foundation: “Collaborative Research: Endangered Languages Catalog (ELCat),” BCS-1058096 to the University of Hawai‘i at Mānoa (Principal Investigator Lyle Campbell) and BCS-1057725 to Eastern Michigan University (Principal Investigators Helen Aristar-Dry and Anthony Aristar), and by a grant from the Henry Luce Foundation: “Research on Endangered Languages in China and Mainland Southeast Asia for the *Catalogue of Endangered Languages*” (Lyle Campbell, PI.) The goals and basic organization of the *Catalogue* were established in an international workshop with some 50 specialists from around the world supported by National Science Foundation grant “Collaborative Research: Endangered Languages Information and Infrastructure Project” (NSF 0924140). The Endangered Languages Project website, where the *Catalogue of Endangered Languages* is housed, was built in partnership with Google.org.

1

INTRODUCTION

Why catalogue endangered languages?

Lyle Campbell and Anna Belew

This chapter introduces cataloguing the world's endangered languages. The principal aims of this book are to report and make known the findings of the *Catalogue of Endangered Languages* (available online at www.endangeredlanguages.com). This introductory chapter outlines the history of the *Catalogue of Endangered Languages* project, describes the *Catalogue* database, and highlights the *Catalogue*'s achievements in producing new knowledge, filling gaps in information, correcting errors, and updating information. In addition, it addresses how endangered languages are identified (together with their degree of endangerment), the causes of language endangerment, and the many reasons why loss of languages matters to us all.

Introduction: cataloguing endangered languages

Awareness of the plight of endangered languages has become more widespread in recent years, and accounts of the world's languages in crisis are not uncommon. For example, media reports of “language obituaries” – where the death¹ of a language is reported with the death of its last known speaker – are increasingly common. Some recent examples include:

Tommy George, the last fluent speaker of Awu Laya, an aboriginal language [of Australia], died August 12, 2016, at the age of 87. He had been awarded an honorary doctorate by James Cook University for help in documenting language and traditional fire management of land.

(The Australian, Aug. 13, 2016)²

The family of worldwide indigenous languages lost another fluent speaker in the death of Doris Jean Lamar McLemore. Considered the last fluent speaker of the Wichita language, she walked on August 30, 2016.

(Indian Country Today, October 6, 2016)³

Edwin Benson, last known native speaker of Mandan, died in December of 2016 at 85 years of age.

(KFYR News, December 9, 2016)⁴

Some widely reported cases are quite famous. For example, Ubykh, a Northwest Caucasian language, was reported to have “died at daybreak, Oct. 8, 1992” when its last speaker, Tevfik Esenç, passed away (Crystal 2000:2). Eyak of Alaska lost its last speaker when Marie Smith Jones died at the age of 89 on January 21, 2008.

Language obituaries such as these lend a personal touch to the alarm over language extinction and help to make the crisis of language endangerment more understandable. However, for most languages that have ceased to be spoken, we do not know who the last speaker may have been. While language obituaries such as these highlight individual cases of language death, they also help to cast a spotlight on the broader global concern of language endangerment: an endangered language, by definition, is one at risk of extinction, and many of the world’s languages are at risk of extinction, many of them in the very near future.

The endangered languages crisis is widely acknowledged as one of the pressing problems facing humanity today, posing moral, practical, and scientific issues of enormous proportions. Previously, there was no single fully reliable source dedicated to the endangered languages of the world, and a catalogue of the world’s endangered languages was called for to help address these problems. Responding to this often-repeated call, Lyle Campbell, Helen Aristar-Dry, and Anthony Aristar set out to create just such a catalogue: a comprehensive database of the endangered languages of the world.

Thus, the *Catalogue of Endangered Languages* (ELCat) was born. In 2012, it was made available to the public on the Endangered Languages Project (www.endangeredlanguages.com), a website dedicated to promoting endangered languages. The *Catalogue* informs users about the plight of endangered languages and encourages efforts to slow the loss. It provides information on the endangered languages of the world as a cost-free resource for the public, scholars, individuals and groups facing language loss, and funding agencies to help them make the best decisions in how to deploy limited resources most effectively (see Hauk and Heaton, Chapter 9). This website serves both as a resource for information on the endangered languages of the world, and as an arena to collaborate with others working to document, revitalize, and promote endangered languages.

This book’s main objective is to make known the results from the comprehensive *Catalogue of Endangered Languages*, the most complete resource on the endangered languages of the world. It provides up-to-date information involving language identification, speaker numbers, locations, and language classification. It identifies which languages are most in danger and what impact their loss will have. It reports the *Catalogue of Endangered Languages*’ achievements in producing new knowledge, filling gaps in information, correcting errors, and updating information (for example, on number and age of speakers, locations of languages, and degree of endangerment).

This book seizes the opportunity to present academic and scientific material that is not available on the *Catalogue of Endangered Languages* website or that is much abbreviated there. It emphasizes those aspects of endangered languages and the research on them that are of particular relevance and interest to scholars, students, and the more engaged members of the public. In addition, we hope this volume may be of use to practitioners of language documentation and conservation who are interested in global trends in language endangerment and revitalization, and in the methods used to produce this kind of data.

Language endangerment in context

Languages have been going extinct for as far back in history as we are able to see. There are many well-known cases from antiquity: Etruscan, Hittite, Sumerian, and many others. Given

this, some might ask, why the current alarm? The answer is that languages are becoming extinct today at a strikingly accelerated rate; the magnitude of language extinction is much greater now than ever before (see Belew and Simpson, Chapter 4).

The evidence of the sharp increase in language loss is seen in the history of language extinction and in the current status of languages everywhere. For example, California had some 100 American Indian languages at the time of the Gold Rush, c.1850, but only 18 are still spoken today; none of them is being learned by children through intergenerational transmission; all are highly endangered. Of the some 314 languages spoken in what is now the US and Canada when Europeans first arrived, 152 no longer have native speakers (48%). Of about 280 languages at the time of first European contact in territory now belonging to the US, only 76 are still spoken (27%). However, these figures are misleading – all of the remaining 76 are endangered, most of them critically so, and only about a dozen are still being passed on to children. Many of them will soon become dormant if language revitalization efforts are not successful. The statistics for languages no longer spoken and for those critically endangered are similarly very high in Australia, much of Latin America, and stretches of northern Eurasia; in fact, no inhabited region of the world is exempt from language endangerment. The *Catalogue of Endangered Languages* lists 427 languages as “critically endangered,” with 262 languages having ten or fewer speakers. Altogether, the *Catalogue of Endangered Languages* lists 3,138 currently endangered languages⁵; that is 45.6% of the 6,879 living languages in the world as listed by *Ethnologue* (Simons & Fennig 2017).⁶

The number of extinct language families offers another perspective on the crisis. Of the world’s c.407 independent language families (including language isolates, language families that have only one member), already 96 are extinct – no language belonging to any of these families has any remaining native speakers. This means that 24% of the linguistic diversity of the world, calculated in terms of language families, has been lost. Of all the millennia in which languages could have disappeared, two-thirds of these language families became extinct in only the last 60 years, dramatically underscoring the accelerating rate of language extinction in recent times. Worse, many more languages and language families are on the brink of losing their last native speakers and will soon follow, portending a drastic change in the world’s linguistic diversity.

What does this mean? The loss of a specific language may be likened to the loss of a single species, say the Bengal tiger or the Right whale. However, the extinction of whole families of languages is a tragedy analogous to the loss of whole branches of the animal kingdom, say to the loss of all felines or all cetaceans. Just imagine the distress of biologists attempting to understand the animal kingdom with major branches missing. Yet this is what confronts us: a staggering quarter of the linguistic diversity of the world is gone forever.

Endangered languages are often likened to endangered species in the public discourse (though we make no claims as to the validity of this comparison), but actual comparisons show that losses to the world’s linguistic diversity are even greater than losses in the diversity of biological species. Figures reported for endangered biological species vary widely, some elevated no doubt in part due to well-meaning attempts to raise public sympathy for the cause and in part due to the fact that about many species very little is known. A representative count lists as “Critically Endangered” or “Endangered” about 12% of mammals, 13% of birds, and 21% of amphibians (IUCN Red List 2017). Clearly the threat to biological species is smaller in numerical scale than that to human languages: as noted above, 46% of all living languages are currently listed as endangered (*Catalogue of Endangered Languages*) – the rate of endangerment for the most highly endangered kind of animals, the amphibians, is less than half of that of endangered languages. This underscores the seriousness of the language endangerment crisis.

How do we determine that a language is endangered?

How does a language come to have an entry in the *Catalogue of Endangered Languages*? The main criteria used by the *Catalogue* to determine whether a language is endangered are:

- The absolute number of speakers – in very general terms, the fewer the number of speakers, the less likely the language’s long-term survival.
- Intergenerational transmission – if a language is not being learned by children in the traditional way, passed on from one generation to the next, it is likely doomed to extinction unless revitalization efforts prove successful. The greater the degree of intergenerational transmission, the more likely the language’s survival.
- Trends in the number of speakers – the more rapidly the number of speakers decreases, the more endangered the language is.
- Decrease in domains of use – the more the domains in which the language is used are reduced, the greater its endangerment becomes.

The *Catalogue of Endangered Languages* reports the vitality of the world’s endangered languages based on these criteria. The *Catalogue*’s Language Endangerment Index (LEI) gives a score for the degree of endangerment of each endangered language. In general, for a language to receive an entry in the *Catalogue*, it needs to score higher than 0% on the LEI. Non-endangered languages such as French or Thai would score a 0% on all factors used by the LEI; higher percentages indicate higher levels of endangerment (for example, a language that scored 92% on the LEI would be “Critically Endangered”). See Lee and Van Way, Chapter 5, for details of how the LEI calculates levels of endangerment.

Causes of language endangerment

It bears repeating that “the factors determining obsolescence of language are non-linguistic” (Swadesh 1948:235). Numerous causal factors of language endangerment have been identified, and are sometimes grouped under higher-order categories such as the following:

- **Economic factors:** lack of economic opportunity, rapid economic transformation, shifts in work patterns, resource depletion, forced changes in subsistence patterns, communication with outside regions, resettlement, destruction of habitat, globalization, etc.
- **Political and social factors:** discrimination, repression, official language policies, level of education available, population dispersal, etc.
- **Attitudes:** attitudes of the speakers towards the languages under threat and towards the official national language(s) and dominant languages that surround them, attitudes of members of mainstream society towards minorities and their languages, prestige or stigma associated with the endangered language and dominant language(s), etc.
- **Lack of institutional support:** as represented in the roles of the languages in education, government, churches, and the media, and even recreational activities (sports events, popular culture, music, etc.).

Causes of language endangerment are complex, however, and differ widely depending upon the language(s) involved; it is impossible to provide any simple list of factors, and their effects upon a given language may vary greatly. For additional discussion of causes of language endangerment, see Belew and Simpson, Chapter 4.

Why should we care?

The endangered languages literature provides many answers to this question, and many reasons for why language endangerment matters to us all. Some of the main ones follow.

Social justice and human rights

Language loss is often not voluntary; it frequently involves violations of human rights, pushed by political or social repression, oppression, aggression, prejudice, violence, and at times by ethnic cleansing and genocide. Cases such as these are a matter of right and wrong, and that is important to everyone.

Globally, the concern for language endangerment connects with concern for various types of social injustice.

Language loss is often experienced as a crisis of social identity. For many communities, work towards language revitalization is not about language in isolation, but is part of a “larger effort to restore personal and societal wellness” (Pfeiffer and Holm 1994:35). Some scholars and community activists insist that ongoing language loss leads to damaged communities and dysfunctional behaviors. They argue that one’s psychological, social, and physical well-being is connected with one’s native language; it shapes one’s values, self-image, identity, relationships, and ultimately success in life. (Indeed, the 5th International Conference on Language Documentation and Conservation was organized around the theme of “Language and Wellbeing”; see <http://icldc5.icldc-hawaii.org>). Indigenous voices testify to the crucial role language plays in their cultural and personal identity. The following are representative:

How can I believe the foolish idea
That my language is weak and poor
If my mother’s last words
Were in Evenki?

(Alitet Nemtushkin, Evenki poet)⁷

For centuries our languages have been a reflection of those cultural distinctions that have made us who we are as a people, and in a sense have been an element of the many things that have made us strong.

(Stephen Greymorning [Arapaho] 1999:6)

Why save our languages, since they now seem to have no political, economic, or global relevance? That impression is exactly the reason why we should save our languages, because it is the spiritual relevance deeply embedded in our own languages that makes them relevant to us as American Indians today.

(Richard Littlebear [Northern Cheyenne] 2000).

Each language still spoken is fundamental to the personal, social and – a key term in the discourse of indigenous peoples – spiritual identity of its speakers.

(Zepeda [Tohono O’odham nation] and Hill 1991:1)

I can’t stress enough the importance of retaining our tribal languages, when it comes to the core relevance or existence of our people . . . You could argue that when a tribe loses its language, it loses a piece of its inner-most being, a part of its soul or spirit . . . When it comes to native languages, the situation is simple: Use it or lose it.

(Sonny Skyhawk [Sicangu Lakota, Hollywood actor] 2012)

Human concerns

Languages are the treasure houses of information for history, literature, philosophy, art, and the wisdom and knowledge of humankind. The stories, ideas, and words contained in a language help us make sense of our own lives and of the world around us – of the human experience, and of the human condition in general. When a language goes extinct without documentation, we lose incalculable amounts of human knowledge. We illustrate this by mention of only two areas, literature and history.

Literature: As many scholars often assert, by studying literature, we learn what it means to be human. This is equally true of the oral literatures of the indigenous peoples of the world – they, too, have grappled with the complexities of their world and the problems of life, and the insights and discoveries represented in their literatures – whether written or oral – are of no less value to us all. When a language becomes extinct without documentation, taking all its oral literature, oral tradition, and oral history with it into oblivion, all of humanity is diminished.

History: We study history “to gain access to the laboratory of human experience” (Stearns 1998). Great reservoirs of historical information are contained in languages. The classification of related languages teaches us about the history of human groups and how they are related to one another, and we gain understanding of contacts and migrations, the original homelands where languages were spoken, and past cultures from the comparison of related languages and the study of language change – all irretrievably lost when a language is lost without adequate documentation.

Because languages encompass the world’s knowledge and wisdom, the loss of the literature and historical information, and much more, means loss in the potential ways of experiencing and understanding the world.

Loss of knowledge

The world’s linguistic diversity is one of humanity’s most valuable treasures. This means that the loss of the many hundreds of languages that have already become extinct is a cataclysmic intellectual disaster, on many different levels. To cite a single example, encoded in each language is knowledge about the natural and cultural world it is used in. This knowledge is often not known outside of the small speech communities where the majority of endangered languages are spoken. When a language dies without adequate documentation it takes with it this irreplaceable knowledge. It is argued that in principle, loss of such knowledge could have devastating consequences even for humankind’s very survival. Reduction of language diversity diminishes the adaptational strength of the human species because it lowers the pool of knowledge from which we can draw.

A telling example comes from the Seri (of Sonora, Mexico, with c.700 speakers). The Seri have knowledge of “eelgrass” (*Zostera marina* L.) and “eelgrass seed,” which they call *xnois*. It is “the only known grain from the sea used as a human food source . . . eelgrass has considerable potential as a general food source . . . Its cultivation would not require fresh water, pesticides, or artificial fertilizer” (Felger and Moser 1973:355–356). Seri has a whole set of vocabulary items dealing with eelgrass and its use. According to the argument, it is all too plausible to imagine a future in which some natural or human-caused disaster might compromise land-based crops, leaving human survival in jeopardy because of the loss of knowledge of alternative food sources such as the knowledge of eelgrass reflected in the Seri language. Speculation aside, it is clear that documentation of the languages of small-scale societies and of the knowledge they hold has significantly benefitted humanity.

Other examples come from medicine. It has been reported that 75% of plant-derived pharmaceuticals were discovered by examining traditional medicines, where the language of curers often played a key role (see, for example, Bierer et al. 1996). If these languages had become extinct and knowledge of the medicinal plants and their uses had been lost, all of humanity would be impoverished and human survival would be left less secure.

For these and other reasons, it can be argued that reduction of language diversity diminishes the adaptive strength of humans as a species because it lowers the pool of knowledge from which we can draw for survival.⁸

Consequences for understanding human language

A major goal of linguistics is to understand human cognition and human language capacity through the study of what is possible and impossible in human languages. Discovery of previously unknown linguistic features and traits as we document languages contributes to achieving this goal and advances knowledge of how the human mind works. Conversely, language extinction is horrendous, impeding achievement of this goal. The following example illustrates this well.

The discovery of the existence of languages with OVS (Object–Verb–Subject) and OSV (Object–Subject–Verb) basic word orders forced abandonment of a previously postulated language universal. Greenberg (1978:2) had proposed that “whenever the object precedes the verb the subject does likewise.” However, it was discovered that Hixkaryana (a Cariban language of Brazil, with only 350 speakers) has OVS basic word order, seen in the sentence:

<i>toto</i>	<i>yonoye</i>	<i>kamura</i>
man	ate	jaguar

“The jaguar ate the man.”

We now know that several languages have OVS or OSV basic word order; most of them are spoken in small communities in the Amazon. Discovery of languages with these basic word orders not only forced abandonment of the postulated universal, but also required revision of numerous other theoretical claims about language. It is all too plausible, given what has happened to indigenous languages at the hands of unscrupulous loggers, miners, and ranchers, that the few languages with these word orders could have become extinct before they were documented, leaving us forever with erroneous assumptions about what is possible and impossible in human language and how that reflects on understanding of human cognition.

Documentation of endangered languages has frequently and repeatedly demonstrated the importance of obtaining adequate descriptions of these languages; the discovery of previously unknown linguistic traits is helping linguists to comprehend the full range of what is possible and impossible in human language (for numerous examples, see Palosaari & Campbell 2011).

Peace through loss of languages?

A final reason for why we should care about language loss involves setting straight erroneous views about the role of languages in geopolitical conflicts. It is often said that if we had fewer languages, we would understand each other better and live in greater harmony. But this is far from true. That monolingualism does not guarantee nor even foster greater “understanding” is attested throughout history. As David Crystal (2000:27) reminds us, “all of the large monolingual countries of the world have had their civil wars.” It is also shown by the many recent and ongoing armed conflicts among groups speaking the same language: in Darfur, Egypt,

Iraq, Libya, Syria, Yemen, Colombia, Northern Ireland, Thailand, or the 1994 genocide in Rwanda involving Hutu and Tutsi, both speakers of the same language, Kinyarwanda. This contrasts with lack of such conflicts in relatively peaceful, officially multilingual Belgium, Canada, Finland, Luxembourg, Singapore, Switzerland, Tanzania, etc. Multilingual and multicultural countries need to recognize that national unity and understanding are not fostered by monolingualism or ethnic cleansing, but that recognition of minority language rights can bring about mutual trust, peace, and ultimately, national stability. We need to expose the erroneous assumption that people and countries cannot be both multilingual and successful, and show, rather, that there are significant benefits from multilingualism. In short, there is no evidence that fewer languages might lead to greater harmony.

A brief history of *ELCat*

In 2009, a group of prominent linguists and language activists gathered for the *Endangered Languages Information and Infrastructure Workshop* (ELIIP) at the University of Utah, funded by the US National Science Foundation. The workshop's goal was to lay the groundwork for a comprehensive digital database of information about the world's endangered languages, and for the infrastructure for disseminating and updating this fast-changing data. The workshop produced recommendations for what types of data should be collected about endangered languages, and how to structure this information in a database; recommendations for assessing linguistic vitality in a more thorough, accurate, and universally applicable way than existing endangerment metrics; and recommendations for constructing a website that would allow the public to view, search, and contribute to this database.

In 2011, the US National Science Foundation awarded funding to construct the database designed at the ELIIP meeting. The project, now called the *Catalogue of Endangered Languages* (abbreviated *ELCat*), was undertaken in collaboration between teams at the University of Hawai'i at Mānoa (UHM) Department of Linguistics and The LINGUIST List at Eastern Michigan University (EMU). UHM was responsible for providing data on the languages of Europe, the Americas, Asia, and the Pacific, while EMU was responsible for the technical infrastructure of the *Catalogue*'s database, as well as for data on the languages of Africa and Australia. The initial version of the *Catalogue* database was compiled during 2011 and early 2012 by the UHM and EMU teams, with input and supervision from the project's team of Regional Directors (see Appendix 1.1 for the list of Regional Directors).

Also in 2011, the project's principal investigators were approached by representatives from Google.org, the nonprofit arm of the technology company Google. Google.org contributed by building the web platform, in collaboration with the web design firm Vizzuality and the *ELCat* teams, and providing fixed-term funding for the maintenance and development of the project. On June 21, 2012, the Endangered Languages Project (ELP) website was launched at www.endangeredlanguages.com, featuring the initial version of the *Catalogue of Endangered Languages*. Among the other features of the ELP site, the web platform allowed the public to suggest additional information or changes to the information contained in the *Catalogue*. Over the next four years, the *ELCat* team continued adding and updating the language information in the *Catalogue* (see Appendix 1.2 for a complete list of staff, research assistants, and volunteers who have contributed to the *ELCat* project), and they continued vetting and inputting site users' suggestions for language data improvements and updates. The ability for users who have firsthand experience with a language to submit information directly, and for the *ELCat* team to reply quickly, ask for additional data, and implement these changes, is of major benefit to *ELCat*'s goal of providing the most accurate and up-to-date language information possible

(see also Heaton and Simpson, Chapter 8, for discussion of how the ELP website's functionality serves language communities). In 2013, the Henry Luce Foundation awarded *ELCat* a grant (2013–2016) supporting research on the endangered languages of China and mainland Southeast Asia (Lyle Campbell, PI), an area much in need of this concentrated effort.

In 2015, the NSF-funded period for the construction of the *ELCat* database came to a close. In order to ensure the continued maintenance and accuracy of the *Catalogue*, a workshop was held at UHM in February 2015 titled *Sustaining the Catalogue of Endangered Languages*. The personnel involved in the *ELCat* and ELP projects came together to devise a long-term governance structure for the *Catalogue*, procedures for adding and changing language data after grant funding expired, and roles and responsibilities for the continued upkeep of *ELCat*. As a result of this workshop, *ELCat*'s International Advisory Board was formed, with Dr. Bill Palmer as the initial chair. (See Appendix 1.3 for the current membership of the International Advisory Board). In 2016, Google.org's period of funding the ELP website came to a close, and hosting and technical support for the *ELCat* and ELP projects moved to the University of Hawai'i at Mānoa's Office of Information Technology and Support, where it remains today.

The *ELCat* database

Below we provide a brief description of the *ELCat* database, which may be useful to those interested in the technical aspects of the project, as well as clarifying certain aspects of how the *Catalogue*'s language data are presented.

The *Catalogue of Endangered Languages* is stored and hosted on the same servers, and within the same database, as the Endangered Languages Project. The site is written in Django (a Python-based web development framework), which encompasses both the front end (website) and back end (database), and is hosted at the University of Hawai'i at Mānoa's Office of Information Technology and Support. The integration of *ELCat* and ELP means that each language entry in the *Catalogue* generates its own webpage, where all information about the language is displayed, along with any user-submitted materials pertaining to that language. Only *ELCat* staff has access to the back end of the database, and only authorized members of the research team and International Board of Directors are able to edit or add language information directly. Users may suggest changes or additions to the *Catalogue*'s information directly through the language web pages (see Heaton and Simpson, Chapter 8), but these changes are carefully reviewed by the research team before being implemented – the *Catalogue* itself is strictly academically moderated, rather than being crowd-sourced.

ELCat is not a fieldwork initiative; it does not collect information on endangered languages directly. Rather, it compiles published and unpublished data from a wide variety of sources, including journal articles, books, news reports, conference presentations, and direct communications with scholars and community language experts. The structure of the *ELCat* database allows for the incorporation of information from multiple bibliographic sources regarding a single language. This is key to *ELCat*'s goal of providing a *comprehensive* database of language information: rather than limiting the data presented to that published in a single source, the *Catalogue*'s users can access all available information from all available sources. This is especially important in cases where conflicting information exists, or where published information has changed over time. Users are able to view all of the information from a single bibliographic source at once, and can scroll to display information from other sources. For example, the language page for Southern Selkup (www.endangeredlanguages.com/lang/8555), a language of Russia, includes information from five bibliographic sources. By scrolling through the information from different sources, users can see that a publication from 2005 provided a

count of 15–20 speakers of Southern Selkup, another source from 2013 estimated there were fewer than five speakers, and a 2015 source listed one remaining speaker. All information is presented with a citation of its original source, so users may follow up by reading the original source or make judgments about its reliability. The ability to see and compare information from multiple publications is useful for tracking changes in a language’s vitality, as well as allowing users to see if there is conflicting information about a language’s vitality and context (e.g. one source may consider a language severely endangered, while another considers it mildly endangered).

While *ELCat* strives to present all available information on a language, the design of the website necessitates designating a single source of information that will appear first when a user arrives at a language’s page. This is the “preferred” source, and is selected by the *ELCat* research team based on recency, reliability, and completeness of information. In the case of Southern Selkup, the 2015 source citing one speaker is “preferred,” both because it reflects the most recent information about the language, and because it is deemed highly reliable (a direct communication from a linguist who is conducting firsthand research on the language). All statistics and information presented in this volume are based on the “preferred” sources of information in the *ELCat* database as of April 4, 2017; because the database is updated regularly as new data becomes available, information presented here may differ from what is found in the *Catalogue* at any point in the future.

For more information about the structure and functionality of the *Catalogue* and the Endangered Languages Project, please visit www.endangeredlanguages.com/about.

The book’s chapters and how they contribute to the book’s goals

This volume aims to present the knowledge gained from the construction of the *Catalogue of Endangered Languages*: not only the data gathered about endangered languages themselves, but the lessons learned from the process of researching, compiling, and maintaining a database of information about the world’s endangered languages. While the ELP website allows users to contribute language data and materials, it is limited in its ability to present supplementary material explaining the rationale, processes, and complications involved in curating language information on a large scale. This volume therefore provides a comprehensive overview of the compilation of, and lessons learned from, the *Catalogue of Endangered Languages*. These lessons are broadly applicable: many of the issues that the *ELCat* project has addressed are also faced by researchers working with endangered languages, language workers in revitalization and revival initiatives, language activists and policymakers, learners and teachers of endangered languages, academics working with large datasets in other social sciences and humanities, and others doing language-related work.

For example, the issue of what to *call* a language is often a complicated one, and one that is probably familiar to anyone who has worked with endangered or under-documented languages. In Chapter 2, Chen and Campbell outline potential complications involved in the names representing given languages. In some cases, use of autonyms (the name of the language in the language itself, or what its speakers call it) are preferred by language communities over exonyms (names given by outsiders). However, Chen and Campbell discuss numerous complications involved in names for languages, with a description of how the *Catalogue* has chosen to resolve these complicated issues.

In Chapter 3, Barlow and Campbell provide an overview of language classification and its importance in particular to endangered languages. They discuss the challenges posed by attempts to classify under-documented or poorly attested languages, and explain *ELCat*’s

approach to contested or unclear classifications, as well as the importance of endangered languages to understanding the world's linguistic diversity, and the benefits that accurate classification can provide to speakers and language communities. In this context, they provide an updated list of the world's language families (including language isolates), which incorporates recent consensus knowledge about the classification of the languages in the *Catalogue*.

In Chapter 4, Belew and Simpson present an overview of language death, dormancy, and revival in the *Catalogue of Endangered Languages*. The concept and terminology of language "death" is discussed, and the authors outline *ELCat*'s policies regarding languages with no native speakers, including those which are undergoing revival efforts. They then discuss what is known about language extinction in the past, and how this compares to rates of global language loss today, as seen in the *Catalogue*'s data. Finally, they provide an overview of the *Catalogue*'s information on previously dormant languages that are now being revived, concluding that while language loss is currently occurring at an unprecedented pace, the growing rate of language revival is cause for cautious optimism.

ELCat's metric for measuring language endangerment is a central feature of the project, and in Chapter 5, Lee and Van Way describe the Language Endangerment Index (LEI), its rationale and workings, and its benefits as compared with other vitality-assessment tools. They discuss the need for flexibility in such a tool when dealing with a wide range of endangerment scenarios and levels of documentation. They outline the LEI's approach to accounting for all available *and* unavailable information about a language's vitality. Finally, they discuss the use of the LEI within *ELCat*, and provide a number of case studies to illustrate how the LEI may be applied to a variety of language situations.

While much of the volume's first half focuses on the *Catalogue*'s development and how the project addresses the challenges of language cataloguing, the second half of the book focuses on applications of the information compiled by the *ELCat* project. In Chapter 6, Campbell and Okura present the new knowledge gained from the *Catalogue*'s data. They provide updated calculations of the current pace of language loss, comparing it to the oft-repeated claim that "one language dies every two weeks," and they issue a corrective to this figure.

Chapter 7 presents the *Catalogue*'s full listing of the world's endangered languages, their endangerment levels, number of speakers, and locations. The lists are organized by major region, and brief overviews of language endangerment in each region are provided, along with useful references to further work on language endangerment in specific parts of the world.

While the *Catalogue* is a scholarly resource, its target audience is not limited to academics. In Chapter 8, Heaton and Simpson discuss how the *Catalogue* and the Endangered Languages Project website can serve the needs of language communities. They present the *Catalogue*'s goals for serving language stakeholders, and provide several case studies of community organizations making use of *ELCat* and ELP to support their language work. They outline some of the features of the *Catalogue* and the Endangered Languages Project that may be of additional benefit for language communities in the future.

Concluding the volume is Chapter 9, in which Hauk and Heaton address the issue of "triage," how to utilize the information contained in the *Catalogue of Endangered Languages* to prioritize language documentation work. One of the *Catalogue*'s objectives is to help granting bodies, policymakers, and community organizations make informed decisions about how to deploy limited resources for language documentation, by providing accurate information about endangered languages. Hauk and Heaton discuss a number of potential criteria for setting priorities in language documentation, such as degree of current endangerment, typological uniqueness, and amount of existing documentation, as well as extralinguistic factors such as the language's accessibility to outside researchers.

Conclusions

This book, *Cataloguing the World's Endangered Languages*, aims to present the results of the *Catalogue of Endangered Languages* Project and to make its findings better-known to its varied audience, to scholars, members of communities whose languages are endangered, funding agencies, teachers and students, the media, language policy decision makers, and the general public. The *Catalogue of Endangered Languages* is the most comprehensive resource on the endangered languages of the world, and this book reports its achievements in producing new knowledge, filling gaps in information, correcting errors, and updating information: for example, on number of speakers, degree of endangerment, and much more. The book goes beyond this, presenting also academic and scientific information about endangered languages that is not available on the website that hosts the *Catalogue of Endangered Languages* (at www.endangeredlanguages.com) or is not sufficiently elaborated there.

There was previously no single fully reliable source dedicated exclusively to the endangered languages of the world; the *Catalogue of Endangered Languages* (ELCat) meets its objective of filling this gap. The *Catalogue* and this volume have broader goals, as well: they are dedicated to promoting endangered languages, to informing users and readers about the plight of endangered languages, and to encouraging efforts to turn the tide and reduce language loss. It is hoped that this book will contribute to these ends and prove to be a useful resource for all with interest in and concern for the languages of the world, their diversity, and for the language endangerment crisis facing us all.

Notes

- 1 Terms such as “death” and “extinction” are still common in the linguistic literature and popular media, and used here with the caveat that “death” of a language need not be permanent; see Belew and Simpson, Chapter 4, for more on the terminology of language “death.”
- 2 www.theaustralian.com.au/news/nation/language-lost-with-the-passing-of-great-elder-tommy-george/news-story/3fca836f8f19e249e18437bc2732415f
- 3 <https://indiancountrymedianetwork.com/news/native-news/doris-mclemore-last-fluent-wichita-speaker-walks-on/>
- 4 www.kfyrvtv.com/content/news/Edwin-Benson-last-known-fluent-speaker-of-Mandan-passes-away-at-85-405723515.html
- 5 The total number of languages listed in the *Catalogue of Endangered Languages* is 3,394, but this includes 256 dormant or awakening languages, that is, languages with no known native speakers.
- 6 *Ethnologue*'s total number of languages is 7,099; adjusted for the 220 with no known speakers and one constructed language, their number of living languages is 6,879.
- 7 www.unesco.org/new/en/culture/themes/endangered-languages/faq-on-endangered-languages/
- 8 In principle it is possible for a society to give up its language and shift to another, and find ways to talk about this sort of knowledge in the new language. What we observe, however, in case after case, is that when a language is not passed on to the next generation, the knowledge of the natural and cultural world encoded in that language fails to be transmitted as well.

References

- Bierer, Donald E. & Carlson, Thomas J. & King, Steven R. 1996. *Shaman pharmaceuticals: Integrating indigenous knowledge, tropical medicinal plants, medicine, modern science, and reciprocity into a novel drug discovery approach*. Saluda, NC: Network Science Corporation.
- Crystal, David. 2000. *Language death*. Cambridge: Cambridge University Press.
- Felger, Richard & Moser, Mary Beck. 1973. Eelgrass (*Zostera marina* L.) in the Gulf of California. *Science* 181(4097). 355–56.
- Greenberg, Joseph H. 1978. Introduction. In Greenberg, Joseph H. & Ferguson, Charles A. & Moravcsik, Edith A. (eds.), *Universals of human language*, vol. 2, 1–6. Palo Alto, CA: Stanford University Press.

- Greymorning, Stephen. 1999. Running the gauntlet of an indigenous language program. In Reyhner, Jon & Cantoni, Gina & St. Clair, Robert N. & Yazzie, Evangeline Parsons, *Revitalizing indigenous languages*, 6–16. Flagstaff, AZ: Northern Arizona University.
- International Union for Conservation of Nature and Natural Resources. 2017. *Summary statistics, Table 3a*. (www.iucnredlist.org/about/summary-statistics#Tables_3_4)
- Littlebear, Richard. 1999. Some rare and radical ideas for keeping Indigenous languages alive. In Reyhner, Jon & Cantoni, Gina & St. Clair, Robert N. & Yazzie, Evangeline Parsons, *Revitalizing indigenous languages*, 1–5. Flagstaff, AZ: Northern Arizona University.
- Palosaari, Naomi & Campbell, Lyle. 2011. Structural aspects of language endangerment. In Austin, Peter K. & Sallabank, Julia (eds.), *The Cambridge handbook of endangered languages*, 100–119. Cambridge: Cambridge University Press.
- Pfeiffer, Anita & Holm, Wayne. 1994. Laanaa Nisin: Diné education in the year 2004. *Journal of Navajo Education* 11. 35–43.
- Simons, Gary F. & Fennig, Charles D. 2017. *Ethnologue: Languages of the world, twentieth edition*. (www.ethnologue.com)
- Skyhawk, Sonny. 2012. Why should we keep tribal languages alive? *Indian Country*, April 6, 2012. (<http://indiancountrytodaymedianetwork.com/2012/04/06/why-should-we-keep-tribal-languages-alive-99182>)
- Stearns, Peter. 1998. Why study history? ([www.historians.org/about-aha-and-membership/aha-history-and-archives/archives/why-study-history-\(1998\)](http://www.historians.org/about-aha-and-membership/aha-history-and-archives/archives/why-study-history-(1998)))
- Swadesh, Morris. 1948. Sociologic notes on obsolescent languages. *International Journal of American Linguistics* 14. 226–235.
- Zepeda, Ofelia & Hill, Jane. 1991. The condition of Native American languages in the United States. In Robins, R. H. & Uhlenbeck, E. M. (eds.), *Endangered languages*, 135–155. Oxford: Berg.

Appendix 1.1 Regional Directors for the *Catalogue of Endangered Languages*, 2011–2015

Region ¹	Director(s)
Africa	Matthias Brenzinger, University of Cape Town
Australia	Claire Bower, Yale University
Caucasus	Alice C. Harris, University of Massachusetts Amherst
East Asia	David Bradley, La Trobe University
Europe	Brian Joseph, Ohio State University
Indonesia	I. Wayan Arka, Australian National University
Mexico, Central America, and the Caribbean	Lyle Campbell, University of Hawai‘i at Mānoa
Near East and Central Asia	Habib Borjian, Columbia University Charles C. Häberl, Rutgers University
North America	Keren Rice, University of Toronto Lyle Campbell, University of Hawai‘i at Mānoa
Northern and Central Eurasia	Juha Janhunen, University of Helsinki
Pacific	Bill Palmer, University of Newcastle
Romance languages	Martin Maiden, Oxford University
Sign languages	James Woodward, University of Hawai‘i at Mānoa
South America	Willem F. Adelaar, Leiden University Lyle Campbell, University of Hawai‘i at Mānoa
South Asia	Gregory Anderson, Living Tongues Institute for Endangered Languages
Southeast Asia	David Solnit

¹While the *Catalogue* is divided into 12 primary geographic regions, some specific language families (e.g. Romance) or sub-areas (e.g. Indonesia) have supplementary supervision by additional Directors.

Appendix 1.2 Staff, research assistants, and volunteers for the *Catalogue of Endangered Languages* project

<i>Eastern Michigan University</i>	<i>University of Hawai‘i at Mānoa</i>
Anthony Aristar (Co-Principal Investigator, 2011–2013)	Lyle Campbell (Principal Investigator and project director, 2011–2016)
Helen Aristar-Dry (Co-Principal Investigator, 2011–2013)	Gary Holton (Catalogue Director, 2016–Present)
Verónica Grondona (Principal Investigator, 2013–2015)	Carolina Aragon (volunteer)
Anna Belew (project manager, 2011–2013)	Russell Barlow (RA)
Amy Brunett (RA)	Anna Belew (RA, co-project coordinator)
Jacob Collard (RA)	Yen-ling Chen (RA)
Lesley Dennison (RA)	Katie Butler Gao (RA)
Kristen Dunkinson (RA)	Bryn Hauk (RA, co-project coordinator)
Bryn Hauk (project manager, 2013–2014)	Raina Heaton (RA)
Uliana Kazagasheva (RA)	Joelle Kirtley (volunteer)
Lwin Moe (lead programmer)	Nemanja Komar (RA)
Marcus Nero (assistant project manager, 2015–2016)	Nala Huiying Lee (RA)
Sara Oldaugh (assistant project manager, 2015–2016)	Stephanie Locke (volunteer)
Stephanie Walla (project manager, 2014–2015)	Colleen O’Brien (volunteer)
Brent Woo (RA)	Christianne Ono (RA)
	Melody Ann Ross (volunteer)
	Sean Simpson (RA, project coordinator)
	Alexander D. Smith (RA, project coordinator)
	Kaori Ueki (volunteer)
	John Van Way (RA, project coordinator)

Appendix 1.3 Membership of the *Catalogue of Endangered Languages* International Board of Directors, 2015–present

Bill Palmer (International Board of Directors Chair), University of Newcastle
Gary Holton (<i>Catalogue</i> Director), University of Hawai‘i at Mānoa
Willem Adelaar, Leiden University (Regional Director for South America)
Greg Anderson, Living Tongues Institute for Endangered Languages (Regional Director for South Asia)
Habib Borjian, Columbia University (Regional Director for Near East)
David Bradley, LaTrobe University (Regional Director for East Asia)
Matthias Brenzinger, University of Cape Town (Regional Director for Africa)
Lyle Campbell, University of Hawai‘i at Mānoa (Regional Director for the Americas)
Verónica Grondona, Eastern Michigan University
Tracey Herbert, First Peoples’ Cultural Council
Brian Joseph, The Ohio State University (Regional Director for Europe)
Mary Linn, Smithsonian Institute
Keren Rice, University of Toronto (Regional Director for North America)
David Solnit (Regional Director for East and Southeast Asia)

2

NAMING CONVENTIONS AND CHALLENGES

Yen-ling Chen and Lyle Campbell

How languages should be named presents numerous challenges for the *Catalogue of Endangered Languages*. This chapter identifies the difficulties and explains how these are dealt with and how language names are handled in the *Catalogue*. A guiding principle with important consequences for naming languages is to make the *Catalogue of Endangered Languages* maximally useful and maximally accessible to the widest possible range of users and audiences, recognizing their different needs and respecting their preferences.

Introduction

This chapter is about the language names in the *Catalogue of Endangered Languages*, and it explains how decisions about language names are handled in the *Catalogue*.

The *Catalogue* has identified 3,394 endangered languages. All of them are catalogued with a name, and some with several alternative names, which may be either autonyms or exonyms. (Not all of them have a corresponding ISO 639-3 code, but the ISO code is also provided wherever one exists for a given language.) An autonym is the name used by a group of people to refer to themselves or their language. In contrast, an exonym is a name given by outsiders to refer to a group of people or their language; it is not the name that a group uses to refer to itself or its language.

Note that different speech communities that speak a single language may employ different names to refer to themselves or their language, especially when different varieties of the language are involved. The naming conventions used in the *Catalogue of Endangered Languages* attempt to represent all the salient names by which a language is (or has ever been) known. Therefore, these names can be either autonyms or exonyms. The name by which a language is most widely known in the literature is typically adopted as the primary name (that is, the name that appears first and most prominently in a language page's heading) in the *Catalogue*. However, where community members¹ express interest in displaying a different name as the primary name, the *Catalogue of Endangered Languages* attempts where possible to display these preferred names as the primary identifier for the language. For the purposes of the *Catalogue*, it is not so important which of the alternate names may be listed first, and alternative names are generally not listed in any specific order; what the *Catalogue* holds as

most important is the possibility for users to search for and find the language by any of its alternate names.

It is worth mentioning that the *Catalogue of Endangered Language* is presented on the up-to-date, open-access website of the Endangered Languages Project (ELP), at www.endangeredlanguages.com. Hence, the searchability of ELP is an important criterion for being a user-friendly website. Primary names of all the language entries in the *Catalogue* are represented in large font. Alternative names provided in *Catalogue of Endangered Languages* are represented in the “also known as” line on the website. Language names written in non-Latin scripts are also included in the database wherever available, and are searchable on the site. However, given the inability of mainstream web browsers to display certain characters, names written in scripts that do not have Unicode-compatible fonts are not eligible for inclusion in *ELCat*.

The *Catalogue of Endangered Languages* seeks the best solution to issues involving language names based on some general principles. The decision for what entities are given an independent entry as a separate language in the *Catalogue* is based on strictly linguistic criteria (i.e. mutual intelligibility). Experience has made it clear that the *Catalogue*’s decisions must be free of political and other non-linguistic agendas – the *Catalogue* will not include (or delete) an entry in order, for example, to justify aggression or repression, to gain access to government funding, to get government or other recognition, to provide support for political independence, to underwrite land rights claims, to attempt to justify petitions for extra representatives in governing bodies, to seek exemption from restrictions that apply to others, etc. Decisions in unclear cases in the *Catalogue* were reached based on standard linguistic criteria, in consultation with the Regional Directors, community members, and others. In cases where there is significant disagreement, the different opinions and the reasons behind them, where known, should be represented in public comments in the language entries of the *Catalogue of Endangered Languages*.

The problem of a single name applied to two or more languages

One difficulty is that it is not uncommon for a single name to apply to two or more separate languages. For example, the name “Tuha” applies to both Dukha [no ISO 639-3] and Uighur Uryangkhai [no ISO 639-3], two distinct Turkic languages. “Nanai” can refer to Nanai [gld],² Kilen [no ISO 639-3], and Kili [no ISO 639-3], which are separate languages all belonging to the Tungusic language family. Also, sometimes languages within the same language family, although in different subgroups, bear names that are similar – for instance, Mulam [mlm] and Mulao [giu], both of which are Tai-Kadai languages, and Lizu [no ISO 639-3] and Lisu [lis], both of which are Tibeto-Burman, and Kadu [zkd] and Kaduo [ktp], both being Tibeto-Burman as well. These often cause confusion and misidentification. There are also languages with the same name but in different parts of the world. For example, there are four languages named “Mono.” Such cases are distinguished by placing an indication of the location in parentheses after the name, for example in this case, “Mono (Cameroon)” [mru] and “Mono (Democratic Republic of the Congo)” [mnh] (both Niger-Congo languages) vs. “Mono (Solomon Islands)” [mte] (an Austronesian language) vs. “Mono (United States)” [mnr] (a Uto-Aztecan language). A particular name can be associated with different languages in the same region, for example, “Solon” is associated with Evenki [evn], Ongkor Solon [no ISO 639-3], and Solon [no ISO 639-3], three different Tungusic languages, and “Khamnigan” is associated with Khamnigan Ewenki [no ISO 639-3], which is Tungusic, and Khamnigan Mongol [no ISO 639-3], which is Mongolic.

The problem of what to call a language is not, of course, confined to endangered languages. For example, linguistically speaking, there is no single language within the Chinese/Sinitic subgroup (of the Sino-Tibetan family) named “Chinese.” Chinese is a generic term for a large grouping of languages. Nowadays, the term “Chinese” is typically associated with Mandarin (the language of the bureaucrats, originally from Sanskrit *mantrin* “counselor”), or more precisely with the Mandarin (官話) subgroup, which includes Standard Mandarin, Pekingese, Sichuanese, Shandonghua, etc., of the Sinitic branch. Because speakers of Mandarin have been politically dominant in the history of China, Standard Chinese [cmn], also known as *Putonghua* (普通話), literally “Common Language,” or *Kuoyu* (國語), literally “National Language,” has been selected as the only official language in China and Taiwan. Using “Chinese” as equivalent to “Standard Mandarin” leads people for political reasons to consider other Sinitic languages as not independent languages but as mere dialects³ of Mandarin (see Chapter 3 for discussion of the problem of distinguishing dialects from distinct languages). In the *Catalogue*, “Chinese” is used as a generic term for the Sinitic branch or things written in Chinese characters (Traditional Chinese or Simplified Chinese), or in reference to an unspecified Chinese language; it is never used to refer only to Standard Mandarin.

In the following sections, four major challenges for the naming conventions used in the *Catalogue of Endangered Languages* are addressed.

The problem of a multiple names applied to a single language

The first challenge is that of multiple names associated with a single language, and how these are to be represented in the *Catalogue of Endangered Languages*. A single language often has multiple names. Given that the *Catalogue* seeks to provide the most accurate and comprehensive information on endangered languages, the inclusion of multiple language names cannot and should not be avoided, both conventional names (exonyms, xenonyms) and community-recommended names, i.e. self-designations, community-preferred names, or the name of the language in the language itself (autonyms, endonyms). Briefly, a xenonym is nearly the same as an exonym; it is the name of a language given by outsiders, “foreigners,” not the name the language community itself calls their language. An endonym, in contrast, is essentially the same as an autonym, the name members of a language group call their language.

For example, Nivacle [cag] is also widely known by the names Ashluslay and Chulupí (each with several variant spellings), but has also been called at one time or another by several other names: Etehua, Guentusé, Mathlela, Sogciagay, Sowa, and Suhin (most of these with multiple spellings) (see Campbell et al. in preparation; Fabre 2016). “Nivacle” is the most recent and now the most common of these names, an autonym, from the native word *niwakle* “man, person, Nivacle person, Nivacle language.” “Chulupí” was very common as the name for the language and people, but is now being abandoned because *chulupí* means “cockroach” in the Spanish of Bolivia and parts of Paraguay, and that is taken as negative (see below for the problem of offensive names). Many languages of South America, for example, have multiple names; often several names for the same language (see Campbell 2012 for numerous examples).

The problem of mismatches between exonyms and autonyms

The second challenge is that the names for languages in the academic or popular literature and the names preferred by speakers of the languages often do not match. Some language group members have, in recent times, strongly advocated adoption of their autonyms, and a good

refers to a particular variety (dialect), have in this way attempted to get recognition for the dialect as a separate language, often for political motives.

One example is a user-submitted recommendation that the *Catalogue* change the current primary name of the language identified by code [str] from “Senćoten” to “Northern Straits Salish,” since Senćoten is the name of just one of several dialects of Northern Straits Salish. In addition to Saanich (Senćoten), other dialects involved include Samish, Lummi, Songish, Sooke, and Semiahmoo, and motivation for the recommended name change was to encourage the people associated with these other dialects also to feel included.

Jejueo [ije] (제주어; 濟州語) “Jeju Language” has several names in the literature, i.e. Jejimal (제주말) or Jejutmal (제웃말) “Jeju Speech,” Jeju Saturi (제주 사투리) “Jeju dialect,” Saturi (사투리) “dialect,” Jeju Bangeon (제주 방언; 濟州方言) or Jejudo Bangeon (제주도 방언; 濟州島方言) “Jeju dialect,” and Jeju Jiyeokeo (제주 지역어; 濟州地域語) “Jeju regional language” (Sejung Yang, personal communication 2016). Taking language preservation and revitalization into consideration, Jejueo (제주어) “Jeju language” was chosen by O’Grady et al. (2014) when applying for an ISO 639-3 code for the language, and Jejueo was adopted as its primary name in the *Catalogue*. The local government in Jeju also uses the name Jejueo (제주어) in official documents.

Pazeh-Kaxabu [uun] serves as another example where different speakers of the same language prefer different names and have separate group identities. Pazeh-Kaxabu, a critically endangered Austronesian language in Taiwan, has two dialects, namely Pazeh and Kaxabu. The Pazeh variety has received more attention in the academic literature since the late 1800s (see Ino 1897; Ogawa 1923; Asai 1937), and study of this variety has been taken as representative of the whole language (see Blust 1999; Lin 2000; Li & Tsuchida 2001; 2002, among others). The Kaxabu variety, on the other hand, was first mentioned in Ferrell (1970), together with Pazeh. Ferrell (1970:73) mentioned that the term Pazeh “ripe, matured” was probably derived from the Chinese term 熟番, literally “cooked savages,” which was used to refer to those Sinicized aborigines in Taiwan, whereas Kaxabu is a place name, the region where the Kaxabu live.⁴ After the death of the last fluent speaker of the Pazeh dialect, Tata Pan Jin-yu/Phoann Kim-giok (潘金玉) in 2010, Ethnologue mistakenly classified Pazeh-Kaxabu as extinct (Lewis et al. 2014, 2015). However, there remain 12 fluent elderly speakers of the Kaxabu dialect, with the oldest being 92 in 2016 (Tatauwan Kaxabu a han, n.d.; Elizabeth Zeitoun, personal communication 2016).⁵ Given that the name “Pazeh,” representing the Pazeh group, has occurred more frequently in the academic literature but that the Kaxabu group still has fluent speakers and they prefer “Kaxabu,” the name Pazeh-Kaxabu is used in the *Catalogue*. (Speakers prefer the spelling “Kaxabu” instead of “Kahabu.”)

Non-linguistic issues associated with language names

The fourth challenge is how to handle the socio-political concerns often associated with language names. It is not uncommon for pejorative names to be given to languages by outsiders. Despite Shakespeare’s opinion that “that which we call a rose by any other name would smell as sweet,” many do not find these names sweet but offensive, and may want the offensive names eliminated. Nevertheless, some languages are best known by names that some now consider offensive, and it is for these names that many users of the *Catalogue* will search in order to find these languages. Therefore, for the *Catalogue* to be effective, searchable, and comprehensive, the names cannot be removed altogether; however, whenever a name is known to be offensive, the *Catalogue* includes it with an explanation of the issue, along with a recommendation that this name should not be used. For example, the Nadahup languages of Brazil

and Colombia were previously known as “Makú” languages, from an Arawakan exonym meaning “without speech.” The term Makú is now strongly dispreferred, and the *Catalogue*’s entries for the Nadahup languages alert readers to this fact.

The Berber languages that are spoken in northern Africa (a branch of Afro-Asiatic) present another example. The term “Berber” is an exonym meaning “barbarous” (cf. Greek *Barbaroi* and Latin *barbarous*, Brett & Fentress 1997); now many prefer to call these languages “Tamazight” or “Amazigh” languages, based on an autonym from the root *mazigh*, meaning “free man” or “noble man.” In a similar case, the names “Lolo(ish)” (used in the English literature) and “Yi” (used in the Chinese literature) both refer to the same Tibeto-Burman branch of Sino-Tibetan.⁶ Bradley (2004) has proposed “Ngwi” to replace “Lolo(ish)” as the name for that subgroup, since *Lolo* is pejorative: both characters in the Chinese spelling of the name “Lolo” have the dog radical, 獬 *luo* and 獯 *luo*. *Ngwi* originally meant “silver” but now is used as a widespread autonym within the Ngwi subgroup, which included Sa’nguie [ysy], Samei [smh], Nosu [iii], Sani [ysn], Hani [hni], Akha [ahk], Lisu [lis], Lipo [lpo], etc. Chinese scholars have also recommended the use of 保 *luo*, which contains the human radical, as a replacement for 獬 *luo* and 獯 *luo*, both with the dog radical. As for the Chinese name “Yi,” Bradley (2004) explains that the government of the People’s Republic of China have adopted the new Chinese spelling, *Yi* (彝), to refer to the Yi Nationality, replacing the old name *Yi* (夷), a derogatory term used in Chinese documents during the Ming and Qing dynasties to refer to certain minority groups in the South.

Despite the fact that the etymologies of Berber and Lolo/Yi involve derogatory terms, given the widespread use of Berber and Lolo(ish)/Yi in the literature, both names are employed in the *Catalogue*, especially as the names for subgroups, although also with the newer proposals as alternative names.

The exonym/autonym distinction is not always clear, and an autonym may change so much that it ends up becoming an exonym. Wichí [mzh, wlv, mtp] is a Matacoan language of Argentina and Bolivia. The name previously used for the language, “Mataco,” originally was not offensive, but some were uncomfortable with it because it sounded to them like Spanish *matar* “to kill.” As a result, the name was changed to *Wichí*, in principle an autonym, but actually closer to an exonym; it is based on the Wichí word /wik’i/, which is the name for the group (an ethnonym). The Wichí name for their language is *lhamtes* “the language” or *wik’i lhamtes* “the language of the Wichí” (Terraza 2008:8; Fabre 2016).

The names “preferred by” some speakers sometimes are not generally accepted or recognized by most members of the community, but rather are favored (even invented) for particular non-linguistic purposes. For example, some felt that *Kaqchikel* (Cakchiquel) [cak] (a Mayan language of Guatemala) was not a true autonym and suggested instead *Kaqchikel ch’ab’al*, whereas the true autonym is *qach’ab’al* (literally “our language”), and the truest autonym would be [qač’aβil], the phonetic pronunciation of what is spelled *qach’ab’al*. Sometimes there is also disagreement about which spelling of a name to use, or which script or orthography – *Kaqchikel* vs. *Cakchiquel* is a good example.

A mismatch between names for linguistic groups and for ethnic groups is not uncommon. It is not uncommon for speakers to adopt the ethnonym given by local authorities as the name for their language. Speakers of Lakkia [lbc] and Bunu [bwn, buh] are classified as belonging to the Yao (瑶) Nationality in China (Ratliff 2010:3), and consider themselves Yao/Mienic speakers, although Lakkia linguistically belongs to the Tai-Kadai family and Bunu linguistically is Hmongic. In cases such as this, both the linguistic and ethnic names are given in the *Catalogue* when clarification is needed. Following the above-mentioned case, to avoid

such confusion, the term Hmong-Mien is adopted in the *Catalogue* for the language family, instead of Miao-Yao (苗瑶), which is often associated with ethnic, not linguistic classification in China.

In sum, naming is not purely linguistically driven, as illustrated by the case studies under discussion. Choices about language names often need to be adjusted accordingly, with socio-political factors taken into consideration. Also, it is important to note that the classification and categorization of one's linguistic community and one's ethnicity might not always be a one-to-one match.

Conclusion

As seen here, decisions about how to list the names of languages in the *Catalogue of Endangered Languages* must confront several challenges. The overarching principles that *ELCat* follows are to make the *Catalogue* as user-friendly and comprehensive as possible, to facilitate the ability to search for and find languages, and to respect the communities whose languages are involved.

In following these principles, *ELCat* attempts to list all the salient names by which a language is or has been known. *ELCat* identifies where possible and displays the names preferred by the majority of community members (representing the language as a whole rather than any particular variety of the language or subgroup of the overall community). *ELCat* identifies, where known, names that are considered offensive and recommends against their use. *ELCat* also displays both external and conventional names (exonyms, xenonyms) and community-recommended names (self-designations, community-preferred names, or the name of the language in the language itself – autonyms, endonyms).

Notes

- 1 By community members, we here loosely mean speakers, people whose heritage language is involved, or people who are otherwise stakeholders in decisions about a language.
- 2 Throughout this volume, three-letter codes in square brackets following language names represent language codes in the ISO 639-3 standard. Brackets containing multiple codes, e.g. [abc, xyz] indicate that the *Catalogue* considers two or more ISO 639-3-designated varieties to belong to a single language, one with multiple ISO codes.
- 3 The term 方言 *fangyan* used in the Chinese academic literature is often mistranslated as “dialect.” However, *fangyan*, literally “a speech variety of a place,” should often be regarded as a vernacular or a language, instead of a dialect of Mandarin.
- 4 Ferrell (1970:73) states that “Pazeh informants at present no longer recognize the derivation of *pazeh* and state that it is ‘just our tribal name.’ Kahabu is the name of the area where the Kahabu villages are now located, and may possibly be attributed to a different ethnic group living in the area when the Pazeh-Kahabu moved into the P’uli Basin in the early 1800s.”
- 5 *Ethnologue* (Simons & Fennig 2017) calls this Kulon-Pazeh and notes: “2 [speakers] (Jen-kuei Li 2013). Last speaker of the Pazeh dialect died in 2010. Only known speakers use Kaxabu dialect (Jen-kuei Li 2013).” However, their source, Paul Jen-kuei Li (2013), does not mention that there are only two remaining speakers – the population of the Pazeh-Kaxabu is also larger than two.
- 6 The name of the Tai-Kadai language family is also controversial, although not derogatory. Ostapirat (2000:19–20) points out that Tai-Kadai is homophonous with “Tai ladder” in Thai, the official language of Thailand. He further explains that “The term /thai kadai/ has often elicited smiles or funny looks from non-linguists (sometimes from linguists as well!) when they first hear it.” For this reason, he proposed the use of “Kra-Dai” where both morphemes are autonyms (see Ostapirat 2000 for more detail). Both “Tai-Kadai” and “Kra-Dai” have been in use in the academic literature.

References

- Asai, Erin (淺井惠倫). 1937. *Basai*. (Field notes.)
- Blust, Robert. 1999. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38(2). 321–365.
- Bradley, David. 2004. *Endangered Central Ngwi languages of northwestern Yunnan*. (Paper presented at the 37th International Conference on Sino-Tibetan Languages and Linguistics, Lund University, Sweden, October 2004.)
- Brett, Michael & Fentress, Elizabeth. 1997. *The Berbers*. Oxford and Cambridge: Blackwell.
- Campbell, Lyle. 2012. The classification of South American indigenous languages. In Campbell, Lyle & Grondona, Verónica (eds.), *The indigenous languages of South America: A comprehensive guide*, 59–166. Berlin: de Gruyter.
- Campbell, Lyle & Díaz, Luis & Ángel, Fernando. In preparation. *Nivacle Grammar*.
- Fabre, Alain. 2016. *Gramática de la lengua nivacle*. Munich: LINCOM.
- Ferrell, Raleigh. 1970. The Pazeh-Kahabu language. 考古人類學刊 [*Journal of Archaeology and Anthropology*] 31/32. 73–97.
- Fu, Changzhong (符昌忠). 2011. 來語概況 [On the Lai language]. 民族語文 [*Minority Languages of China*] 2011(3). 67–81.
- Fu, Changzhong (符昌忠). 2012. 來語的系屬問題 [The genetic position of the Lai language]. 民族語文 [*Minority Languages of China*] 2012(1). 51–55.
- Fu, Zhennan (符鎮南). 1990. 黎語的方言島—那斗話 [A dialect island of Li - Natouhua]. 民族語文 [*Minority Languages of China*] 1990(4). 14–18.
- Ino, Kanori (伊能嘉矩). 1897. 番語集 [Pazeh language]. (Field notes.)
- Lewis, M. Paul & Simons, Gary F. & Fennig, Charles D. (eds.). 2014. *Ethnologue: Languages of the world, seventeenth edition*. Dallas, Texas: SIL International. (www.ethnologue.com/17/)
- Lewis, M. Paul & Simons, Gary F. & Fennig, Charles D. (eds.). 2015. *Ethnologue: Languages of the world, eighteenth edition*. Dallas, Texas: SIL International. (www.ethnologue.com/18/)
- Li, Paul Jen-kuei. 2013. The last text of the last Pazih speaker. In Guangshun, Cao & Chappell, Hilary & Djamouri, Redouane & Wiebusch, Thekla (eds.), 《綜古述今 鉤深取極》 [*Breaking down the barriers: Interdisciplinary studies in Chinese linguistics and beyond*], 1135–1147. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Li, Paul Jen-kuei (李壬癸) & Tsuchida, Shigeru (土田滋). 2001. *Pazih dictionary*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Li, Paul Jen-kuei (李壬癸) & Tsuchida, Shigeru (土田滋). 2002. *Pazih texts and songs*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Lin, Ying-chin (林英津). 2000. 巴則海語 [*On Pazeh*]. 2000. Taipei: 遠流出版社 (Yuan-Liou Publishing Co., Ltd.).
- Norquest, Peter. 2007. *A phonological system of Proto-Hlai*. University of Arizona. (Doctoral dissertation.)
- Ogawa, Naoyoshi (小川尚義). 1923. 台灣の蕃語に就て [On the Formosan aboriginal languages]. 台灣時報 [*Taiwan Jiho*] 49. 6–23.
- O'Grady, William & Yang, Changyong & Yang, Sejung. 2014. *ISO 639-3 language code request for Jeju*. (www-01.sil.org/iso639-3/cr_files/2014-004_je.pdf)
- Ostapirat, Weera. 2008. The Hlai language. In Diller, Anthony V. N. & Edmondson, Jerold A. & Luo, Yongxian (eds.), *The Tai-Kadai languages*, 623–652. London and New York: Routledge.
- Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics.
- Simons, Gary F. & Fennig, Charles D. 2017. *Ethnologue: Languages of the world, twentieth edition*. (www.ethnologue.com)
- Tatauan Kaxabu a ahan (噶哈巫語全球資訊網) [The Kaxabu]. (n.d.). (<http://kaxabu.weebly.com/221342170424043354862448925391.html>)
- Teng, Stacy Fang-Ching. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Pacific linguistics. 595. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Teng, Stacy Fang-ching. 2009. Case syncretism in Puyuma. *Language and Linguistics* 10(4). 819–844.
- Terraza, Jimena. 2008. *Gramática del wichi: Fonología y morfosintaxis*. Montreal: Université du Québec à Montréal. (Doctoral dissertation.)

3

LANGUAGE CLASSIFICATION AND CATALOGUING ENDANGERED LANGUAGES

Russell Barlow and Lyle Campbell

Introduction

A major contribution to result from the creation of the *Catalogue of Endangered Languages* (*ELCat*) is a more definitive understanding of the language families of the world and their classification. Surprisingly, most people, including most linguists, have little idea how many independent language families exist. In this chapter we answer the following questions: why is an accurate classification of languages important for endangered languages?; how many language families are there?; why is it so difficult to determine this?; and how have language endangerment and extinction affected the linguistic diversity of the world?

Why is an accurate classification of the world's language families so important for endangered languages?

To understand the linguistic diversity of the world it is necessary to have an accurate understanding of the classification of the world's languages. Unfortunately, the rapid loss of endangered languages is quickly reducing that diversity. The loss of any language constitutes a monumental loss of scientific information and cultural knowledge; however, the extinction of a whole family of languages is a tragedy far greater in magnitude. Yet loss not only of individual languages but of whole families of languages is what we must contend with.

Since languages are dying at a rapid rate, and since the resources available to academics and activists are limited, sadly, it is highly probable that not every language will be documented or have the chance to see revitalization efforts before disappearing. Scholars and granting agencies alike need an accurate classification of the world's language in order to make judicious decisions about how to invest efforts and where to deploy limited funds in order to achieve the maximum positive results. Although there are a number of factors – social, economic, and political, as well as linguistic – that can enter into decisions about which languages deserve priority in language documentation (see Hauk and Heaton, Chapter 9), language affiliation is an extremely important one, and is the topic of this chapter. A central goal of linguistics is to comprehend the full range of what is possible and impossible in human languages, charting the extent to which natural languages can and do differ while revealing any linguistic universals or general typological tendencies that exist among them. Therefore, it is important to document