

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Endangered Languages and New Technologies

At a time when many of the world's languages are at risk of extinction, the imperative to document, analyse and teach them before time runs out is very great. At this critical time new technologies such as visual and aural archiving, digitization of textual resources, electronic mapping and social media have the potential to play an integral role in language maintenance and revitalization.

Drawing on studies of endangered languages from around the world – Europe, Asia, Africa and North and South America – this volume considers how these new resources might best be applied, and the problems that they can bring. It also reassesses more traditional techniques of documentation in light of new technologies and works towards achieving a practicable synthesis of old and new methodologies.

This accessible volume will be of interest to researchers in language endangerment, language typology and linguistic anthropology, and to community members working in native language maintenance.

MARI C. JONES is Reader in French Linguistics and Language Change at the University of Cambridge and Fellow in Modern and Medieval Languages at Peterhouse, Cambridge. A highly experienced fieldworker, she has published extensively on language obsolescence and revitalization in relation to Insular and Continental Norman, Welsh and Breton. Her recent publications include *Keeping Languages Alive* (2013), edited with Sarah Ogilvie.

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Endangered Languages and New Technologies

Edited by

Mari C. Jones



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107627840

© Cambridge University Press 2015

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2015

Printed in the United Kingdom by Clays, St Ives plc

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

Library of Congress Cataloguing in Publication data

Endangered languages and new technologies / Edited by Mari C. Jones.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-107-04959-8 (hardback : alk. paper) – ISBN 978-1-107-62784-0

(paperback : alk. paper) 1. Language obsolescence – Technological innovations.

2. Language and languages – Technological innovations. I. Jones, Mari C., editor.

P40.5.L33E533 2015

417'.7 – dc23 2014032212

ISBN 978-1-107-04959-8 Hardback

ISBN 978-1-107-62784-0 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Contents

<i>List of figures</i>	page vii
<i>List of tables</i>	viii
<i>List of maps</i>	ix
<i>List of contributors</i>	x
<i>Preface</i>	xiii

MARI C. JONES

Introduction: Endangered languages in the New Multilingual Order <i>per genus et differentiam</i>	1
--	---

NICHOLAS OSTLER

Part 1 Creating new technologies for endangered languages

1 The Kiranti comparable corpus: A prototype corpus for the comparison of Kiranti languages and mythology	17
AIMÉE LAHAUSSOIS	
2 European Dialect Syntax: Towards an infrastructure for documentation and research of endangered dialects	35
SJEFF BARBIERS	
3 Keyboard layouts: Lessons from the Me'phaa and Sochiapam Chinantec designs	49
HUGH PATERSON III	
4 Rule-based machine translation for Aymara	67
MATT COLER AND PETR HOMOLA	
5 Data management and analysis for endangered languages	81
DOROTHEE BEERMANN	
6 Endangered languages, technology and learning: Immediate applications and long-term considerations	95
RUSSELL HUGO	

v

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

vi Contents

Part 2 Applying new technologies to endangered languages

- | | | |
|----|---|-----|
| 7 | Digital curation and event-driven methods at the service of endangered languages | 113 |
| | BERNARD BEL AND MÉDÉRIC GASQUET-CYRUS | |
| 8 | ‘Allant contre vent et mathée’: Jèrriais in the twenty-first century | 127 |
| | ANTHONY SCOTT WARREN AND GERAINT JENNINGS | |
| 9 | The use of new technologies in the preservation of an endangered language: The case of Frisian | 141 |
| | TJEERD DE GRAAF, COR VAN DER MEER AND
LYSBETH JONGBLOED-FABER | |
| 10 | Language description and documentation from the native speaker’s point of view: The case of the Tundra Yukaghir | 150 |
| | CECILIA ODÉ | |
| 11 | American Indian Sign Language: Documentary linguistic methodologies and technologies | 161 |
| | JEFFREY E. DAVIS | |
| | <i>Bibliography</i> | 179 |
| | <i>Index</i> | 202 |

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Figures

0.1	Content languages for websites	page 8
0.2	Percentage of internet users by language	9
1.1	The Integral Text View	28
1.2	The Similarity View	29
1.3	Concordance results for the English term ‘to die’	30
3.1	First stroke of a two-keystroke character, highlighting the diacritic in OSX	56
3.2	Visual feedback loop	57
3.3	Conceptualization of a graphical element	59
4.1	A comparison of tree structures in English and Aymara	70
4.2	F-structure and dependency tree for <i>naja-w aka ut uta-ča-ya:t-xa</i>	75
4.3	F-structure and dependency tree for <i>huma-n-x hiw-i-w kimsa č'ijar p^hisi-ma-xa</i>	76
4.4	Constituent structure for <i>naja-x kułaka-ma-r uñh-t-wa</i>	77
4.5	F-structure for <i>naja-x kułaka-ma-r uñh-t-wa</i>	77
5.1	Interlinear Glossed Text seen from within the TypeCraft Editor	85
5.2	The TypeCraft data model	87
5.3	External and internal convertibility of linguistic data	89
5.4	A XLFG5 c-structure parse of the Akan sentence: <i>Papa no kyereε mmɔfra no kwa</i> ‘the man showed the children the way’	92
5.5	HPSG syntactic and semantic parsing from IGT data	93
6.1	Example work flow diagram with a central coordinator	108
9.1	The Dictionary Room at the Frisian Academy	146
9.2	The original word cards	147
9.3	The Dictionaries of the Frisian Language	147
9.4	The online version of the Dictionary of the Frisian Language	147
10.1	A TY classroom	153
10.2	Alayi	158
11.1	Sustainability model for AISL documentation and revitalization	172

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Tables

0.1 An overview of language technologies	<i>page</i> 10
3.1 Options available to encode ñ in the Me'phaa and Chinantec layouts	60
3.2 Options available to encode á in the Me'phaa and Chinantec layouts	61
3.3 The conceptual construction of ⟨ 𐄀 ⟩ in Unicode	61
4.1 Information used in DTs	72
4.2 C-structures, f-structures and DTs in a parallel Aymara-English corpus of around 200 sentences	73
5.1 Tool-supported linguistic work flow, output, and impact	88
6.1 A summary of the fundamental questions for endangered language learning technology projects	104

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Maps

0.1 GDP – growth or recession 2007	<i>page</i> 4
0.2 Official zones for the world’s major languages in 2010	7
1.1 Map of the Kiranti area	20
6.1 Map of Washington State as a distribution model example	107
10.1 The Tundra Yukaghir speech community	152
11.1 Great Plains cultural and geographic area	163

Cambridge University Press
978-1-107-04959-8 - Endangered Languages and New Technologies
Edited by Mari C. Jones
Frontmatter
[More information](#)

Contributors

MARI C. JONES

Reader in French Linguistics and Language Change, Peterhouse and
Department of French, University of Cambridge, United Kingdom

SJEF BARBIERS

Senior researcher, Meertens Institute and Professor of Dutch Variation
Linguistics, the University of Utrecht, the Netherlands

DOROTHEE BEERMANN

Professor, Norwegian University of Science and
Technology, Norway

BERNARD BEL

Ingénieur de recherche, University of Aix-Marseille, France

MATT COLER

Researcher, INCAS³, the Netherlands

JEFFREY E. DAVIS

Professor, Educational Interpreter Programme, University of
Tennessee, USA

MÉDÉRIC GASQUET-CYRUS

Lecturer, University of Aix-Marseille, France

TJEERD DE GRAAF

Research Associate, Fryske Akademy, the Netherlands

PETR HOMOLA

Researcher, Codesign, s.r.o, Czech Republic

RUSSELL HUGO

Researcher, University of Washington, USA

GERAINT JENNINGS

Assistant Language Support Officer, L'Office du Jèrriais, Jersey,
Channel Islands

x

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

List of contributors

xi

LYSBETH JONGBLOED-FABER

Researcher, Fryske Akademy, the Netherlands

AIMÉE LAHAUSSOIS

Ingénieur de recherche, CNRS/University of Paris (VII, Denis Diderot), France

CECILIA ODÉ

Researcher, University of Amsterdam, the Netherlands

NICHOLAS OSTLER

President of the Foundation for Endangered Languages, School of Oriental and African Studies, London, United Kingdom

HUGH PATERSON III

Researcher, SIL International and University of North Dakota, USA

ANTHONY SCOTT WARREN

Senior Language Support Officer, L'Office du Jèrriais, Jersey, Channel Islands

COR VAN DER MEER

Project coordinator, Fryske Akademy, the Netherlands

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Preface

Mari C. Jones

At a time when UNESCO deems 43 per cent of the world's 6,000 languages at risk of extinction, the imperative to record and analyse these linguistic varieties while they are still spoken has scarcely been greater. The ever-increasing availability of new technologies, from visual and aural archiving to digitization of textual resources and electronic mapping, have the potential to revolutionize the documentation, analysis and revitalization of endangered languages for the linguist and indigenous community alike.

In this digital age, the linguist can now collect high-quality data and transform them via a range of sophisticated and powerful tools and applications far more accurately and quickly than ever before. Fewer researchers and resources are now needed for most pieces of analysis, and work can be disseminated and shared at the click of a mouse. For revitalizers working 'on the ground', new technologies make it easy for teachers to reach large audiences, for authentic learning material to be developed and reproduced conveniently and cheaply and for a wealth of linguistic resources to be fully accessible to all members of the speech community.

However, it is also clear that with these new technologies come new challenges. Technology can often distance the 'producer' from the 'user', resulting in a possible loss of valuable momentum. Indeed, at times it seems that the speaker risks becoming more and more 'invisible', with the success of revitalization movements being measured by the quantity of material generated rather than by the number of subscribers. Put another way, what is the point of manufacturing linguistic resources (computer programmes, blogs, online grammar tools and so forth) if no one actually logs on? How successfully can one revitalize a language when face-to-face communication is increasingly replaced by a face-to-computer context? It is also pertinent to ask how long can a language be kept alive if its social aspect is removed through the apparent separation of its linguistic apparatus and cultural content? Surely the transmission of cultural capital to the next generation is one of the main motivations for preserving and revitalizing endangered languages.

This volume considers how new resources might best be applied, and the problems that they can bring, reassessing more traditional techniques in light

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

xiv Preface

of new technologies and working towards achieving a practicable synthesis of old and new methodologies. It debates how researchers and indigenous communities might seek to enhance the functionality of new technological resources in order to advance their application beyond mere superficialities.

The volume opens with an introductory essay by Nicholas Ostler, President of the Foundation for Endangered Languages, who argues that, as technology revolutionizes attitudes to foreign languages and feeds on itself to bootstrap more sophisticated systems, it seems increasingly likely that it will integrate endangered languages into the world community of linguistic systems. The remaining eleven chapters then divide into two parts. Part 1, 'Developing new technologies' explores the way in which new technologies are currently being created by theoretical and field linguists for the description and analysis of endangered languages on four different continents (Europe, Africa, Asia and South America). The section begins with Aimée Lahaussais' study of how aligned corpora can contribute to the documentation of the syntax and lexis of endangered languages. She focuses on a prototype comparable corpus which is being created with data from Khaling, Thulung and Koyi, three languages of the Kiranti group spoken in Eastern Nepal.

Sjef Barbiers highlights the methodological, technological and linguistic requirements for building and maintaining a network of linked databases to collect and document data, with specific reference to the European Dialect Syntax project. Hugh Paterson III explores keyboard design as a modern means of codifying languages and discusses issues which have arisen in this context for two different established orthographies elaborated for endangered varieties in Mexico. Matt Coler and Petr Homola demonstrate how machine translation is being developed as a useful language-learning tool for Aymara, an endangered language of the Andes. Dorothee Beermann discusses how Interlinear Glossed Text can be used for the exchange and reuse of linguistic research data, which is illustrated using data from Akan and Ga, two languages spoken in Ghana. The section ends with Russell Hugo's chapter, which examines the practical and linguistic considerations of technology use for indigenous language revitalization programmes in North America's Pacific North-west, asking whether it is always necessary to 'tailor-make' new software for each endangered language.

Part 2, 'Applying new technologies', considers how new technologies are being used 'on the ground' by revitalization movements in France, the Channel Islands, the Netherlands, Siberia and the USA. Médéric Gasquet-Cyrus and Bernard Bel focus on the dialects spoken in the northern French Alps. They examine the concept of digital curation and question the very purpose of collecting electronic data if these are merely destined to remain 'frozen' in an archive. This is followed by two studies which demonstrate the support that can be provided for language maintenance. The case of Jèrriais is examined by Anthony Scott Warren and Geraint Jennings and that of Frisian by Tjeerd de

Cambridge University Press

978-1-107-04959-8 - Endangered Languages and New Technologies

Edited by Mari C. Jones

Frontmatter

[More information](#)

Graaf, Cor van der Meer and Lysbeth Jongbloed-Faber. Cecilia Odé introduces a new angle to the debate by describing the ways in which new technologies can benefit the speech community that lies at the centre of a field study, highlighting each field linguist's moral duty to aid revitalization within that community by using the data collected to give back pedagogical and other resources. She illustrates this from her work with the Tundra Yukagirs of northern Siberia. Jeffrey E. Davis' chapter, which concludes the volume, extends the debate to non-verbal varieties and describes the way in which modern technology is also being used to document and revitalize endangered American Indian signed languages. Unless otherwise stated, all websites referenced in this volume were last accessed in January 2014.

Finally, and on a personal note, this volume is dedicated to my parents, Philip and Eirwen Jones. Diolch o waelod calon am eich cariad ac am eich cefnogaeth.