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.
Endangered Languages and New Technologies

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

Includes bibliographical references and index.
P40.5.L33E533 2015 417.7 – dc23 2014032212


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.
### Contents

*List of figures*  
*List of tables*  
*List of maps*  
*List of contributors*  
*Preface*

**MARI C. JONES**

Introduction: Endangered languages in the New Multilingual Order *per genus et differentiam*  
**NICHOLAS OSTLER**  

17

**AIMÉE LAHAUSSOIS**

2 European Dialect Syntax: Towards an infrastructure for documentation and research of endangered dialects  
**SIEF BARBIERS**  

49

3 Keyboard layouts: Lessons from the Me'phaa and Sochiapam Chinantec designs  
**HUGH PATERSON III**

67

4 Rule-based machine translation for Aymara  
**MATT COLER AND PETR HOMOLA**

81

5 Data management and analysis for endangered languages  
**DOROTHEE BEERMANN**

95

6 Endangered languages, technology and learning: Immediate applications and long-term considerations  
**RUSSELL HUGO**
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’: Jerriais 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

Bibliography 179
Index 202
Figures

0.1 Content languages for websites
0.2 Percentage of internet users by language
1.1 The Integral Text View
1.2 The Similarity View
1.3 Concordance results for the English term ‘to die’
3.1 First stroke of a two-keystroke character, highlighting the diacritic in OSX
3.2 Visual feedback loop
3.3 Conceptualization of a graphical element
4.1 A comparison of tree structures in English and Aymara
4.2 F-structure and dependency tree for naja-w aka ut uta-čya:t-xa
4.3 F-structure and dependency tree for huma-n-x hiw-i-w kimsa č’ijar pʰisi-ma-xa
4.4 Constituent structure for naja-x kušaka-ma-r uīh-t-wa
4.5 F-structure for naja-x kušaka-ma-r uīh-t-wa
5.1 Interlinear Glossed Text seen from within the TypeCraft Editor
5.2 The TypeCraft data model
5.3 External and internal convertibility of linguistic data
5.4 A XLFG5 c-structure parse of the Akan sentence: Papa no kyereesa mmnqra no kwa ‘the man showed the children the way’
5.5 HPSG syntactic and semantic parsing from IGT data
6.1 Example work flow diagram with a central coordinator
9.1 The Dictionary Room at the Frisian Academy
9.2 The original word cards
9.3 The Dictionaries of the Frisian Language
9.4 The online version of the Dictionary of the Frisian Language
10.1 A TY classroom
10.2 Alayi
11.1 Sustainability model for AISL documentation and revitalization
# Tables

0.1 An overview of language technologies  

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Options available to encode ŋ in the Me'phaa and Chinantec layouts</td>
<td>60</td>
</tr>
<tr>
<td>3.2</td>
<td>Options available to encode á in the Me'phaa and Chinantec layouts</td>
<td>61</td>
</tr>
<tr>
<td>3.3</td>
<td>The conceptual construction of ⟨ í ⟩ in Unicode</td>
<td>61</td>
</tr>
<tr>
<td>4.1</td>
<td>Information used in DTs</td>
<td>72</td>
</tr>
<tr>
<td>4.2</td>
<td>C-structures, f-structures and DTs in a parallel Aymara-English corpus of around 200 sentences</td>
<td>73</td>
</tr>
<tr>
<td>5.1</td>
<td>Tool-supported linguistic work flow, output, and impact</td>
<td>88</td>
</tr>
<tr>
<td>6.1</td>
<td>A summary of the fundamental questions for endangered language learning technology projects</td>
<td>104</td>
</tr>
</tbody>
</table>
Maps

0.1 GDP – growth or recession 2007  page 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
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