

WRITING SYSTEM DEVELOPMENT AND REFORM: A PROCESS

by

Elke Karan  
Bachelor of Arts, University of North Dakota, 1996

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Arts

Grand Forks, North Dakota

August  
2006

© 2006 Elke Karan

Photographs by Mr. Daniel Sayo-Tomte

© 2005 ACATBA, Central African Republic, used by permission

Map of the Central African Republic

© SIL International, used by permission

This thesis, submitted by Elke Karan in partial fulfillment of the requirements for the Degree of Master of Arts from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

---

Dr. David J. Weber, chair

---

Dr. John M. Clifton

---

Dr. J. Albert Bickford

This thesis meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

---

Dean of the Graduate School

---

Date

PERMISSION

Title            Writing System Development and Reform: A Process  
Department    Linguistics  
Degree         Master of Arts

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Signature \_\_\_\_\_

Date \_\_\_\_\_

## TABLE OF CONTENTS

LIST OF FIGURES.....	ix
LIST OF TABLES .....	x
LIST OF ABBREVIATIONS, SYMBOLS, AND ACRONYMS .....	xii
ACKNOWLEDGEMENTS .....	xiv
ABSTRACT.....	xv
CHAPTER	
1. INTRODUCTION .....	1
2. FROM OBSCURITY INTO THE LIMELIGHT .....	4
3. TWO GLOBAL MOVEMENTS .....	8
3.1 Saving languages from death.....	8
3.2 Education for All .....	13
4. LITERATURE REVIEW .....	17
4.1 Before 1985 .....	17
4.2 1985–2006 .....	21
4.3 Specific topics .....	24
5. DEFINITIONS.....	29
5.1 Writing.....	29
5.2 Writing system.....	29

5.3	Orthography.....	31
5.4	Script.....	32
6.	WRITING SYSTEM TYPOLOGY .....	34
6.1	The notion of evolution associated with writing systems.....	34
6.2	Logographic systems .....	35
6.3	Syllabaries .....	38
6.4	Abjads.....	44
6.5	Alphabets .....	51
6.6	Alphasyllabaries .....	54
6.7	Featural system.....	57
6.8	Mixed systems .....	59
6.9	Multiple scripts for one language .....	62
7.	PRINCIPLES OF ORTHOGRAPHY DESIGN AND REFORM .....	64
7.1	Design options .....	65
7.2	Linguistic factors which influence orthographies.....	69
7.3	Synopsis.....	107
8.	NON-LINGUISTIC FACTORS .....	109
8.1	Political factors.....	109
8.2	Social and sociolinguistic factors .....	112
8.3	Educational factors .....	120
8.4	Technical factors.....	148

9. ADAPTING WRITING SYSTEMS.....	155
9.1 Roman alphabets .....	156
9.2 Cyrillic .....	162
9.3 Arabic .....	165
9.4 Abugidas.....	170
9.5 Orthography testing .....	177
10. ORTHOGRAPHY REFORM: SPECIAL CONSIDERATIONS .....	185
10.1 Motivations for reform and revision.....	186
10.2 Types of reform .....	194
10.3 Resistance to reform .....	196
11. ORTHOGRAPHY REFORM CASE STUDIES .....	204
11.1 Script replacements.....	204
11.2 Revision as part of the orthography design process .....	213
11.3 Reform of languages with a literary tradition.....	223
12. SANGO .....	238
12.1 Sango: language of wider communication and of the churches .....	238
12.2 Statistics and the spread of Sango .....	240
12.3 Classification and vocabulary base.....	241
12.4 Sango grammatical structure .....	243
12.5 The phonology of Sango.....	247
12.6 Literacy rates.....	251
12.7 Sango literature.....	251

12.8	The socio-political situation and the status of Sango .....	255
12.9	The Sango orthography before 1984 .....	263
12.10	The 1984 orthography decree .....	269
12.11	The role of SIL International .....	279
12.12	Current practices .....	285
12.13	New developments .....	301
12.14	Relating the Sango orthography 1984 reform to other case studies .....	306
13.	SUMMARY .....	310
13.1	Appropriate stakeholder involvement .....	311
13.2	Orthography as a work in progress .....	311
13.3	Weighing and balancing all the factors .....	311
13.4	Testing the writing system .....	312
13.5	Anticipating and dealing with conflict .....	313
13.6	Taking a conservative approach .....	313
13.7	Exercising patience and diplomacy .....	314
13.8	Establishing support structures .....	315
13.9	Conclusion .....	315
APPENDIX A .....		316
APPENDIX B .....		318
REFERENCES CITED .....		321

## LIST OF FIGURES

Figure	Page
1. World map: number and distribution of languages near extinction.....	9
2. German school reader using two scripts. ....	207
3. Antiqua, Sütterlin, and Latin scripts in an official document .....	209
4. Map of the Central African Republic.....	238
5. “No spitting” prohibition .....	293
6. Months of the year .....	293
7. Information Desk .....	294
8. Clubfoot information poster.....	294
9. AIDS: Call to action, Airport Road .....	295
10. AIDS: Call to action, Boyrabe District .....	296
11. AIDS: Call to action, Avenue des Martyrs .....	296
12. Protecting the unborn against AIDS .....	297
13. Nursing babies is best .....	297
14. Advertising.....	298

## LIST OF TABLES

Table	Page
1. Katakana syllables illustrating phonemic similarity but not graphic similarity ....	39
2. Sample characters from the Mende syllabary .....	40
3. Sample Roman script alphabets .....	52
4. Sample Amharic syllables.....	54
5. Sample Cree syllables .....	57
6. Tone notation in Attie .....	99
7. UNESCO symbol choice recommendations .....	130
8. Symbolization of two common African sounds, /b/ and /y/.....	160
9. The Roman alphabet expanded: miscellaneous practices in Africa.....	161
10. A selection of Cyrillic characters not used in Russian.....	164
11. Multi-letter graphemes in Abkhaz and Kabardian .....	165
12. Arabic variant letter forms of two consonants .....	167
13. Sample conjuncts from Devanagari .....	172
14. Comparison of Amharic and Tigrinya vowel values .....	176
15. Marking Tagakaulu completive aspect. ....	219
16. Sango vowels .....	247
17. Sango consonants.....	248
18. Comparison of old and new conventions.....	274

19. Impact of the change: comparison of words and text .....	275
20. Comparison of spelling conventions found in basic literacy materials.....	285
21. Comparison of spelling conventions found in post literacy materials .....	287
22. Potential homographs discriminated in Alliance Biblique orthography .....	288
23. Analysis of writing conventions used in Figures 5–8 .....	295
24. Analysis of writing conventions used in Figures 9–11 .....	296
25. Analysis of writing conventions used in Figures 12–14 .....	298

## LIST OF ABBREVIATIONS, SYMBOLS, AND ACRONYMS

[ ]	phonetic level notation
/ /	phonemic level notation
< >	orthographic notation
#	word boundary
∅	null
-	morpheme boundary; affix
+	in combination with; followed by
1	first person
2	second person
3	third person
ACATBA	Association Centrafricaine pour la Traduction de la Bible et l'Alphabétisation
AIDS	Auto Immune Deficiency Syndrome
BCE	Before Common Era (BC)
C	consonant
CAR	Central African Republic
CE	Common Era (AD)
Dr.	Doctor
DRC	Democratic Republic of Congo
ed.	editor
eds.	editors
EFA	Education for All
e.g.	example
H	high tone
i.e.	that is
indef.	indefinite
ILA	Institut de Linguistique Appliquée

ILV	Instituto Lingüístico de Verano
IPA	International Phonetic Alphabet
IRCAM	Institut Royal de la Culture Amazighe
ITA	(Pitman) Intitial Teaching Alphabet
L	low tone
L1	language first acquired
L2	language acquired after L1
LETAC	Lexiques thématiques de l’Afrique Centrale
M	mid tone
MT	mother tongue
N	noun
NEG	negative/negation
O	object
ODH	orthographic depth hypothesis
PL	plural
PNG	Papua New Guinea
poss	possessive
RPA	Romanized Popular Alphabet (for Hmong)
RTT	Recorded Text Testing
S	singular
SIL	SIL International
SM	subject marker
SVO	Subject-Verb-Object (basic word order)
syn	synonym
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children’s Fund
URL	Uniform resource locator
US	United States of America
U.S.A.	United States of America
USAID	United States Agency for International Development
USSR	Union of Soviet Socialist Republics
V	vowel

## ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. David Weber and committee members Dr. John Clifton and Dr. Albert Bickford for their help in making this work become reality. Special appreciation is due to Dr. Weber for his invaluable feedback and editing suggestions.

I would like to thank my colleagues Dr. Patricia Davis, Dr. Diana Weber, and Dr. Mary Morgan for providing the inspiration and encouragement it took for me to embrace life-long learning and pick up my studies where I had left off fifteen years earlier. I would also like to thank my supervisor, Patricia Kelley, and colleagues in the International Literacy Department, who freed me up from various responsibilities so I could concentrate on my studies.

I would like to thank my SIL colleagues who contributed to this work through their wealth of knowledge and experience, through classroom instruction and by answering questions, either face to face or via e-mail.

I am grateful to colleagues of the Association Centrafricaine pour la Traduction de la Bible et Alphabétisation (ACATBA) and SIL members of the Central Africa Group who have given of their time and labor to help me gain access to information I needed. I am especially indebted to Bertin Oundagnon-Basso, Ade Mokoko, and Christoph Müller, who made various Sango documents available to me, as I was not able to personally return to Africa at this time.

Last, but not least, I would like to thank my family for their encouragement in this endeavor, especially my husband, Mark, who cheered me on.

To the speakers, readers, and writers of the Sango language,  
and to all those involved in orthography design  
and reform at the grassroots level in any language.

## ABSTRACT

Orthography issues are complex. Although literature about writing systems has flourished in recent years, issues which preoccupy practitioners involved in orthography design or reform are rarely addressed.

This thesis provides an overview of theory relating to establishing orthographies for unwritten languages and modifying existing systems. It presents principles and recommendations which favor popular acceptance and successful implementation. Introductory chapters explain the recent increased interest in orthography, define terms, introduce pertinent literature, and give an overview of writing system typology. Remaining chapters present linguistic and non-linguistic factors which influence orthography decisions, examine writing system adaptation options, discuss orthography testing, consider motivations for and against reform, and present orthography reform case studies. The orthography development process for Sango, the national official language of the Central African Republic, is documented in detail to illustrate the complexity of the issues involved in working toward a written standard. This thesis illustrates that orthography development is a process, requiring diplomacy, dialogue and negotiation. Experience shows that involving stakeholders in the process is key to success. An orthography not used by the population spells ‘defeat’.

# CHAPTER 1

## INTRODUCTION

*Language is a natural product of the human mind...  
while writing is a deliberate product of human intellect...  
Language continually develops and changes without the conscious  
interference of its speakers, but writing can be petrified or reformed  
or adapted or adopted at will. (Peter T. Daniels 1996)*

Interest in writing systems has grown, but there are few *how-to* resources for those who design an orthography or are involved in orthography reform. Helpful, practical information is available, but it is quite scattered. Most books and articles focus on the typology, history, or description of writing systems, not on lessons learned for the benefit of languages for which orthographies are still in the conception or birthing phase, or those receiving reconstructive surgery. Knowledge about writing systems or the ability to categorize them does not adequately equip a person to participate in the shaping or remodeling of an orthography.

Several fields are impacted by orthography. Therefore not only experienced and budding linguists have an interest in writing systems, but also educators and those involved in information technology, among others.

The aim of this thesis is two-fold: First, to assemble information from a number of resources in order to provide an overview of theory relating to orthography design and reform. I will use existing literature and a variety of case studies to identify factors which

are likely to work in favor of or against successful implementation of an orthography or of an orthography reform. Secondly, I will document the orthography reform process of a language I am personally acquainted with: Sango of the Central African Republic.

Because there are many resources available describing languages and writing systems with great detail, I make no attempt to duplicate these works. Instead, I hope to judiciously select, and provide, in one place, the information needed for grassroots level orthography work. Since those involved in this type of work may not have extensive training in linguistics, and/or may not speak English as their mother tongue, I will attempt to convey the information and concepts without highly technical language. However, I will presume familiarity with terminology and symbols taught in introductory linguistics courses.

Chapter two explains how the study of written language, previously neglected by linguistics, has now gained its place. Chapter three discusses two current global movements which contribute to the increased interest in local languages and their orthographies: a concern for endangered languages and *Education for All*—a concerted effort to improve education worldwide. Chapter four gives an overview of the literature relating to writing systems and orthography design and reform. Chapter five provides basic definitions, while chapter six discusses basic writing system types in use in the world today. Understanding the differences between writing systems is foundational for orthography design and reform. Chapters seven and eight present principles of orthography design and reform. Chapter seven covers design options and discusses linguistic factors which influence orthography decisions while chapter eight discusses non-linguistic factors and highlights the conditions which contribute toward a writing

system's acceptance. Chapter nine examines ways in which writing systems have been adapted for a variety of languages. It also makes a case for orthography testing and gives an overview of types of testing that might be done. Chapter ten discusses orthography reform issues such as motivations for reform and resistance to reform. Chapter eleven presents case studies of orthography reform from around the world. These provide useful lessons. The Sango orthography reform case study, presented in chapter twelve, is central to the discussion of general principles of successful orthography reform. Chapter thirteen, in conclusion, summarizes the principles learned from the past. These should inform future efforts, benefiting languages under development.

Because I have lived and worked in Africa for extensive periods of my adult life, I have drawn heavily from African languages to illustrate points. I expect that other language development workers will find the same principles to be applicable, regardless of which corner of the world they work in. Many of the same principles would also apply to designing and implementing writing systems for signed languages, however, the area of focus of this thesis is writing systems of spoken languages.

Two appendices are included. Appendix A lists URLs for websites which deal with orthographies and writing systems of the world and related topics. Appendix B provides bibliographical information on Sango language materials which were used in the comparative analysis of orthography practices (see section 12.12.1) but do not appear in the list of references.

I present phonetic representations in square brackets, phonemes between forward slashes, and orthographic representations between angled brackets, when it is pertinent to do so for clarification.

## **CHAPTER 2**

### **FROM OBSCURITY INTO THE LIMELIGHT**

In 1916, Saussure taught his linguistics students that “the linguistic object is not both the written and the spoken forms of words, the spoken forms alone constitute the object (Saussure 1959).” This point of view made orphans of orthography design and orthography reform. The field of linguistics concerned itself almost exclusively with the structure of oral language. The field of educational psychology, concerned with reading instruction and language acquisition, operated on the premise that languages serving as the medium or subject of instruction already had a written form.

As long as educational practices in the world embraced only major world languages and well-developed national languages of wider communication, very few people concerned themselves with orthography design for minority languages. If they did, this was generally motivated by a desire to promote a particular religious faith or advance some political agenda.

But this is now history. A new era has been ushered in. International organizations, governments, linguists and educators, previously not concerned about orthography issues, are now very interested. Four factors have contributed to orthography issues gaining their rightful place in linguistic and educational discussions. These are:

1. language endangerment
2. the *Education for All* movement

3. a reading theory concerning orthographies: the orthographic depth hypothesis
4. advances in technology.

There is now a preoccupation with language death and the preservation of minority languages. Some believe that language decline and language shift can be slowed or halted through raising the status of endangered languages. Giving speakers of such languages a written form and developing literature in these languages is believed to contribute to their preservation.

Secondly, there is currently a global campaign, *Education for All*, promoted by UNESCO, the World Bank, and governmental and non-governmental institutions. Various agencies are collaborating to try to achieve universal education. It is believed that shifting to local languages as the medium of instruction for the first few years of formal education will broaden accessibility and improve the quality of education in developing nations. However, trying to introduce the use of languages into schools clearly presents a dilemma if these are still unwritten or do not have an agreed-upon written standard. This explains the current interest in orthographies.

Third, there is the orthographic depth hypothesis, a theory which predicts that the reading process is not the same for languages which have consistent sound-symbol correspondences and those which don't. The theory has been and is being tested and debated. It has caused a flurry of research projects and the penning of many an article to support or challenge it. (See 4.3.1 and 8.3.2.)

Last, but not least, technological advances in computation these last two decades have opened the door to innumerable representation options for writing languages. Just about anything is now possible in character design and typesetting. In particular, the Unicode

Consortium is hard at work to establish universal standards to permit information flow around the globe in any language (UNICODE, Inc. 1991–2006).<sup>1</sup>

Previously, the study of writing systems was frequently relegated to the fields of archaeology and anthropology or to graphology or typography. It is now viewed by many as a respected field of study in its own right.

A new reference volume on writing systems was published: Daniels and Bright (1996). Florian Coulmas has written a number of books and articles on the subject within a relatively short time span. A reputable journal, *Written Language and Literacy*, was launched in 1998 by John Benjamins Publishing Company. In 2000, the Department of Linguistics at the University of Illinois dedicated a whole volume of *Studies in the Linguistic Sciences* to “Literacy and Writing Systems in Asia.” Theses, dissertations and papers presented at linguistic conferences dealing with writing systems and orthography design and reform are on the increase. And there is a market, an eager readership, for these publications. Some universities have added “Writing systems of the world” to their course offerings,<sup>2</sup> and bibliographies and reading lists on the topics of writing systems, orthography, and spelling reform abound on the Internet.

No longer marginalized or perceived as only relating to ancient languages, the study of writing systems of living languages has become a sub-domain within linguistics.

Coulmas (1989:267) comments on the change:

---

<sup>1</sup> The Unicode Consortium was formally founded in 1991. It grew out of working group and technical committee meetings, which began in 1986 out of a concern for Chinese and Japanese character standards.

<sup>2</sup> Between 2001 and 2006, more than twenty North American universities offered courses with a study of writing system typology in the curriculum.

The past decade has seen the publication of numerous linguistic books and research papers concerned with written language, spoken language and the relations between them... [M]ainstream linguists have largely ignored the relation between spoken and written language as well as that between speech and writing. The many recent publications dealing with specific properties in its spoken and written forms are beginning to correct this oversight.

The present book is presented to the reader in the belief that this is a healthy and necessary development... In most introductory linguistics textbooks writing is relegated to a brief final chapter or an appendix. Yet, in a very real sense, students of language do not know what they are talking about and have no grasp of their subject matter before they have developed an understanding of the relationships between writing, speech and language... Since written language is what linguists usually deal with, it is imperative that they understand the complex ways in which written units relate to units of language.

## CHAPTER 3

### TWO GLOBAL MOVEMENTS

#### 3.1 Saving languages from death

##### *3.1.1 Language endangerment and language death*

“A language is endangered when its speakers are using it in fewer and fewer communicative domains and/or are ceasing to pass it on from one generation to the next” (UNESCO 2005b:11).

“A language is said to be dead when no one speaks it any more...a language is effectively dead when there is only one speaker left, with no member of the younger generation interested in learning it” (Crystal 2000:11).

Language vitality thus depends on passing a language from one generation to the next. This vitality is enhanced when non-native speakers also wish to learn it for some benefit it might bring. Healthy bilingualism results when there are positive attitudes toward both languages. But when a language suffers disdain, others are not motivated to learn it. When mother-tongue speakers choose to use it less and less in different domains, such as the marketplace or home, this threatens its survival. Language planning on a national level, favoring some languages over others, may devalue languages and cause them to be dispreferred, putting them at risk of abandonment. Bilingualism declines, resulting in a shift to a dominant, more prestigious language.

Different classification systems are in use to identify the degree of endangerment of languages. The *UNESCO Red Book on Endangered Languages*, a reference website hosted by the Department of Asian and Pacific Linguistics at Tokyo University, uses the following six categories: not endangered, potentially endangered, endangered, seriously endangered, nearly extinct, and extinct (Tokyo University 2003). Other resources frequently use the term *moribund* to refer to languages in the ‘nearly extinct’ category.

The following map indicates the number of languages near extinction in the different regions of the world, based on reports found in the *Ethnologue* (Gordon 2005).

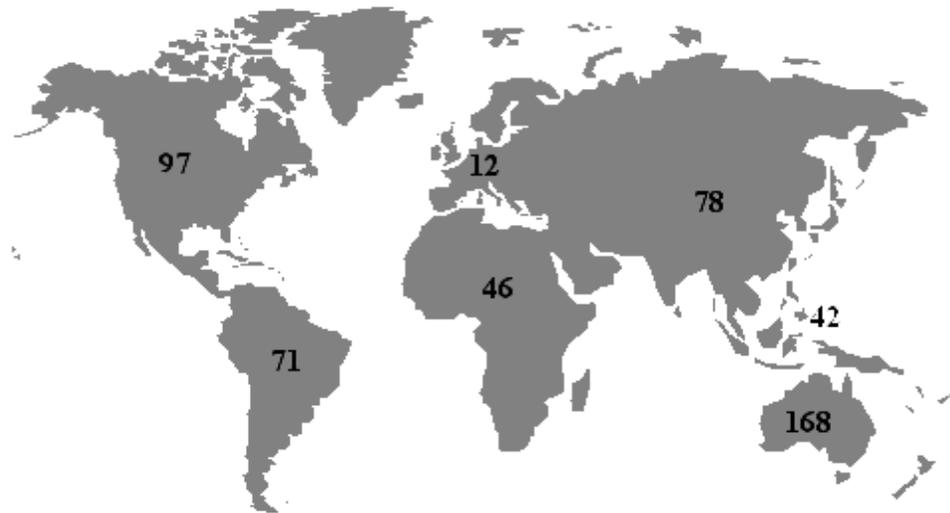


Figure 1. World map: number and distribution of languages near extinction.

The level of endangerment is not only related to the absolute number of speakers. It involves a variety of criteria such as degree of intermarriage, the degree of contact with other languages and their encroachment into the functions previously fulfilled by the endangered language. Languages of wider communication are the biggest threat to the survival of languages that have smaller numbers of speakers. “There are many

cases...where an indigenous language has come to be less used in educational, political, and other public situations, because its roles have been taken over by English, Swahili, or some other lingua franca” (Crystal 2000:21).

### *3.1.2 The need for action*

There is great variation in reporting the number of languages in the world. Crystal (2000:3) states that “most reference books published since the 1980s give a figure between 6,000 and 7,000.” With such a large number, why would linguists and others be concerned about language death and endangerment?

Crystal (2000) and others express three major concerns: One concern relates to human rights issues. If languages are endangered and dying, the question is raised whether this is due to choices made by the members of the language community, or due to oppression, resulting in speakers being denied their linguistic rights. A second concern is that languages are closely intertwined with cultural heritage and diversity and thus perceived as a precious commodity. Those who wish to safeguard the ‘intangible cultural heritage of humanity,’ out of necessity, need to be advocates for endangered languages. A third concern is the rate at which languages are dying and the magnitude of the loss this represents. Crystal (2000:19), based on assertions by several linguists who forecast the demise of 50% of the languages in this century, calculates that this translates into a language dying every two weeks. UNESCO expresses concern:

The scale at which languages are disappearing nowadays is unprecedented.

- Over 50% of the world’s estimated 6800 languages are seriously endangered.
- Only a few hundred languages are not really endangered or endangered at all.
- 96% of the world’s languages are spoken by 4% of the world’s population.

(UNESCO 2005b:11)

### 3.1.3 Action plans

Advances have been made to raise public awareness. New organizations make it their goal to support language communities who may benefit from language revitalization efforts. Linguists are increasingly committed to documenting endangered languages. This is fueled by the desire to explore the full extent of linguistic variety of features found in languages. Crystal (2000:55) laments: “with the death of each language, another source of potentially invaluable information disappears.” He indicates how important it is to develop positive attitudes on the national level toward the multiplicity of languages. “There is an urgent need for memorable ways of talking, to capture what is involved: we have to develop ear-catching metaphors—language as a ‘national treasure’, perhaps, or as a ‘cause for celebration’, or a ‘natural resource’” (Crystal 2000:98). This attempt at attitude reversal is apropos, since many governments, faced with a multiplicity of languages within their borders, are tempted to view multiple languages as a problem, i.e., a hindrance to economic development and a threat to national unity, rather than a resource. Their language policies might not be equivalent to linguistic genocide, but might encourage linguistic suicide.<sup>3</sup>

Negative attitudes of the members of a language community must be turned around as well. Crystal (2000:111) comments:

---

<sup>3</sup> The terms *linguistic genocide* and *linguistic suicide* are commonly used in the field of sociolinguistics. The first relates to language shift by imposition; the second, by choice. No one has to die; they simply are deprived of or willingly sacrifice their language. *Linguistic genocide* has a longstanding definition. *Linguistic suicide* and *linguistic euthanasia* are metaphorical extensions.

Terralingua, an organization concerned with the world's biological, cultural, and linguistic diversity, defines *linguistic genocide* as "prohibiting the use of the language of the group in daily intercourse or in schools, or the printing and circulation of publications in the language of the group."

<http://www.terralingua.org/Definitions/DLingGenocide.html> (accessed November 30, 2005)

Linguists working with endangered languages thus have a very clear task ahead of them, when they encounter negative attitudes...An early aim of intervention must be to create opportunities for the people to improve morale so that they come to think of their language with feelings of confidence, self-esteem, and pride. Only in this way will the community develop an ability from within to deal with the pressure of ongoing change.<sup>4</sup>

International organizations are declaring their intentions concerning world languages. The United Nations Organization is implementing measures to “protect, promote and preserve all languages” (UNESCO 2005c:1) and UNESCO states that it “will aim at preserving linguistic diversity as a prerequisite for the fostering of cultural diversity” (UNESCO 2005c:4). How?

Three priority lines of action guide UNESCO’s activities for the safeguarding of endangered languages

- awareness-raising of language endangerment, and of the need to safeguard linguistic diversity;
- local capacity-building for the safeguarding of endangered languages and promotion of appropriate language policies;
- mobilization of international cooperation. (UNESCO 2005b:11)

What shape does ‘safeguarding diversity’ take? It may look like this:

UNESCO’s Beijing Office has been extending its assistance to the preservation and revitalization of endangered languages of selected ethnic minority groups in China, in partnership with the Chinese Academy of Social Sciences. Research has already been completed for Uygur, Anong, She, Manchu, Lakkia and Tujia languages. In 2005, research has focused on two more languages, Hezhen and Ersu. For each language, the

---

<sup>4</sup> Crystal suggests various cultural events and activities which might be instrumental in boosting self-esteem. Among them he lists poetry reading (Crystal 2000:113). This presumes a writing system exists for the language in question, and that people already read and write in their language. This, in turn, presumes a certain level of prestige.

project consists of field recording, transcription of daily discourse and support to the introduction of mother tongue education into the formal education system.

(UNESCO 2005b:11)

It is clear that recording language, whatever the medium, serves the purposes of science through documentation, but does not—in and of itself—contribute to language preservation. If speakers themselves do not benefit, such recording may simply be academic exploitation. In research, one need not grapple with issues of orthography. But when introducing the use of a language into formal education, dealing with orthography issues becomes imperative. Pedagogical materials need to be developed; reading and writing need to be taught so that schools can effectively contribute to maintaining the language.

### **3.2 Education for All**

Local languages are not only introduced into the educational system to increase their chances of survival; they are also introduced as the medium of instruction to help students succeed educationally.

Article 26 of the Universal Declaration of Human Rights asserts that “everyone has a right to education” (United Nations 1948). But statistics in the 1980s were disappointing:

- More than 100 million children, including at least 60 million girls, have no access to primary schooling
- More than 960 million adults, two thirds of whom are women, are illiterate...
- More than 100 million children and countless adults fail to complete basic education programmes; millions more satisfy the attendance requirements but do not acquire essential knowledge and skills. (UNESCO 1990)

Basic education, in many countries, especially the least developed ones, was suffering setbacks rather than making progress. In response, the *World Declaration on Education for All* and the *Framework for Action to Meet Basic Learning Needs* were adopted by the World Conference on Education for All, held in Jomtien, Thailand in 1990. Each country was to determine what actions would be necessary to turn the situation around. One of the issues to be reconsidered was which languages should be used in education.

At a follow-up conference held in Dakar in 2000, the *Dakar Framework for Action* was drawn up. There were six goals to be achieved by 2015. Basically they related to universal primary education, gender equality and reducing adult illiteracy rates by 50%. Clearly, those who do not know a language that is already in use in education would be disadvantaged and vulnerable to continued discrimination. The second of the six stated goals was: “Ensuring that by 2015 all children, particularly girls, children in difficult circumstances and those belonging to ethnic minorities, have access to and complete, free and compulsory primary education of good quality” (UNESCO 2001).

When only the language of a former colonial power or a national language of wider communication is used for instruction, learning is hampered for those who do not come to school with age-appropriate knowledge of that language. Frustration and anxiety levels are high. Educational cost is increased as a large percentage of children repeat grades. Investment gives a low return as many drop out. This is grievous: if children leave school before grade five, they are very likely to relapse into illiteracy (UNESCO 2002:51). Dutcher (2004:ii) comments:

What do the manifestos of the international conferences say about language as a barrier to expanding educational opportunity? Very little.

...even when these children have access to schools, they are denied educational opportunity. They attend classes taught by teachers speaking, often poorly, a language the children do not understand and through which they therefore cannot learn...

In addition to the cognitive factors, there are emotional ones. Members of minority ethnic groups, whether children or adults, are empowered when their first language is used. Conversely, when the mother tongue is not used, they are made to feel awkward, inferior, and stupid. Their culture is denigrated, and the children are scared, confused and traumatized. This has long-term effects.

Desiring to remedy this situation, several nations have changed their policy regarding the use of minority languages in education. For example, Papua New Guinea (PNG), with 820 living languages (Gordon 2005), abandoned its 'English only' policy. By 1995, more than one fourth of PNG's languages had three-year initial vernacular literacy programs in their elementary schools (Litteral 1999). Implementing the new policy for all of PNG's languages presents a challenge: Each language needs to be analyzed and codified. The vocabulary needs to be expanded to include terms not necessarily needed in routine daily life, but useful in a classroom situation. PNG educators are committed to the plan. Various agencies are providing the necessary institutional support.

The implementation of bilingual education is very complex. It not only requires materials and favorable policy, but also enthusiastic teachers and the approval of parents. And to risk stating the obvious: teachers need to know the language of their students. Unfortunately, teachers are often assigned without due consideration of their linguistic resources.

Briggs (1985), in reporting on bilingual education programs in Peru and Bolivia, offers principles for successful bilingual education programs. High on her list were

parental involvement and good teacher preparation, including some training in the basic linguistics of the official language and of the language of the students.

One of the neediest areas of the world in terms of education is sub-Saharan Africa. Several African countries are embracing bilingual education, including Ethiopia, Eritrea, and Namibia (Dutcher 2004). In the private sector and with the help of non-governmental agencies, programs are carried out in several additional African countries. Africa's languages account for 30.3% of the world's total (Gordon 2005). Introducing even a small fraction of these into the formal system is daunting. Orthography is one of the most crucial issues to be addressed. UNESCO (2005a) reports that 80% of the African languages have no orthography. (I interpret this to mean there is not yet an established written standard, not that no one has ever written anything in these languages.)

## CHAPTER 4

### LITERATURE REVIEW

There are many books and articles relating to writing systems of the world. Here, I will highlight only the more important, particularly those from which a reader can extract general principles and useful lessons for orthography design and reform.

#### 4.1 Before 1985

Daniels (1996:7) comments that “In 1952, I.J. Gelb published the first *linguistically informed* study of writing...Gelb always said he intended his book to be the first, not the last, word on the theory of writing” (italics mine). Gelb’s book, like many that followed, focused on the description of ancient writing systems, on the evolution of modern writing systems, and on writing system typology. Gelb devoted some pages to the sociolinguistic factors related to writing and the importance of orthography reform. Gelb expressed his intentions clearly: “to put together certain ideas which grew out of my experience with past writings in order to see what may be learned from them in the future” (Gelb 1952:247).

Practical help for those involved in the creation of orthographies for previously unwritten languages is scarce. Some early works stand out: Pike’s *Phonemics: A Technique for Reducing Languages to Writing* (1947), a paper by Jack Berry entitled “The making of alphabets” (1958), a volume published by the United Bible Societies

(UBS), *Orthography Studies: Articles on New Writing Systems* (Smalley et al. 1964) and an article by Paul Powlison “Bases for formulating an efficient orthography” (1968).

Pike (1947) in a chapter entitled “The Formation of Practical Alphabets,” addresses linguistic as well as non-linguistic considerations for orthography design. He gives practical advice and addresses the question of community acceptability. He deals with issues such as dialects, loan words, excessive use of diacritics, and transfer to languages of wider communication. Although one may not wish to promote all of Pike’s 1947 ideas—he likely changed his mind on a few things himself—his work is still valid as a reference for orthography design or reform.

Berry presented his paper “The making of alphabets,” at the Proceedings of the 8th International Congress of Linguistics in 1958. Published in a volume dealing with sociology of language and language planning (Fishman 1968), it stood alone in its focus on orthography design. It covered the topics “the scientific principles,” “the social situation,” and “the symbols.”

*Orthography Studies* (Smalley et al. 1964) covers a wide variety of topics about which field workers need to be informed before producing literature in a previously unwritten language. This work is linguistically informed, but not so technical as to be inaccessible to non-linguists. Two of the articles, “Practical limitations to a phonemic alphabet” (Nida 1964b) and Smalley’s “How shall I write this language?” (1964b), contain valuable information and principles for orthography design.

*Die Schrift in Vergangenheit und Gegenwart* (Jensen 1958) became available in English under the title *Sign, Symbol and Script* (Jensen 1970). It is a comprehensive

scholarly reference work, but the content is purely historical and descriptive and does not offer principles for orthography design or reform.

Gudschinsky included two chapters on orthography in her *Manual of Literacy for Preiterate Peoples* (1973). She addressed topics like functional load, underdifferentiation, phonemic vs. morphophonemic representation, and orthography testing.

Fishman's *Advances in the Creation and Revision of Writing Systems* (1977), is an anthology containing both theoretical papers and case studies relating to orthography design and revision. Several of the articles deal with non-Roman script. Berry's "The making of alphabets' revisited" reviews literature published on the topic from 1958-1977 and discusses orthography design criteria presented by Smalley (1964b). It briefly touches on psycholinguistics and the reading process. The volume also includes a reprint of Venezky's 1970 "Principles for the design of practical writing systems" which relates orthography to the teaching of reading. Venezky presents ease of transfer between vernacular languages and official language orthographies as an ideal. He includes practical suggestions for orthography testing.

In 1958, Kenneth L. Pike stated that the Summer Institute of Linguistics was "working on the formation of alphabets and literacy materials for approximately 150 languages."<sup>5</sup> As SIL personnel grappled with orthography issues they published related articles in *Notes on Literacy*.<sup>6</sup> Most of these were language specific but some touched on

---

<sup>5</sup> This comment was made during the discussion time after Jack Berry presented his paper at the Proceedings of the 8th International Congress of Linguistics in 1958, as recorded by Fishman (1968:751).

<sup>6</sup> The SIL Journal *Notes on Literacy* was in circulation from 1966 until 2001. SIL International is a non-profit, non-governmental organization serving host communities through linguistic research and

general principles. Also, READ, an SIL publication produced in Papua New Guinea, published more than 25 articles on orthography topics between 1966 and 1996.

Sjoberg's "Sociocultural and linguistic factors in the development of writing systems for preliterate peoples" (1966) is still useful. This article was based on findings of a survey taken in 1964 among SIL entities worldwide. It examines sound-symbol correspondence issues and points out that the feelings of mother-tongue speakers should be a primary consideration in orthography design.

Powlison's "Bases for formulating an efficient orthography" (1968) reviews the characteristics of an efficient orthography based on the principles outline in Smalley et al. (1964), but focuses on the *functional load* of phonemic distinctions. Powlison equates a writing system's efficiency with actual usage by the community and encourages departing from phonemic notation to quite an extent, provided that communication and readability are not obstructed. He favored morphophonemic writing and simplicity in a writing system.

As various proponents for English orthography reform came and went, several publications on the topic saw the light of day. English, a case demonstrating resistance to change, provides insight into factors at work in orthography design and reform. Scragg's *A History of English Spelling* (1974) examines changes of the English language and the written standard over time. It discusses attitudes toward failed reform efforts. It gives special attention to "the invasion" of French loan words—of interest to those who are concerned with 'language purity' issues in vernacular languages. The chapter on "Power of the press" illustrates the importance and influence of print.

---

language development. Prior to 1999 the organization went by the name Summer Institute of Linguistics.

There are many case studies on orthography reform: Turkish, German, Norwegian, Dutch, to name a few. In chapters ten and eleven I will refer to some of these.

## **4.2 1985–2006**

Sampson's *Writing Systems: A Linguistic Introduction* (1985) covers linguistic concepts and the relationship between speech and writing in an introductory chapter. One chapter deals with precursors to writing, six with typology, and one with English spelling. Coverage is adequate to serve as an introduction for a beginner, but is lacking in some aspects. For example, no alphabetic system other than Roman is discussed, and much discussion is given to different typefaces, rather than to the alphabet's adaptation for different languages.

The content of Coulmas' *The Writing Systems of the World* (1989) is similar to that of Sampson (1985): It provides some theory about speech and writing and examines ancient and current writing system types. Although Coulmas also fails to examine non-Roman alphabets, his chapter on the alphabet has more comprehensive coverage in that it presents principles of alphabetic writing and a brief explanation of orthographic depth. Three chapters are of particular interest to orthography practitioners: 12. From sound to letter: creating alphabets; 13. Writing reform: conditions and implications; and 14. What writing means for linguistics.

Coulmas' *Writing Systems: An Introduction to their Linguistic Analysis* (2003) does not move far from the content of his 1989 publication. It is somewhat more theoretical and comprehensive and contains more illustrations and tables. It presents little to help practitioners, except for a brief section on augmenting the Roman alphabet.

Coulmas' *Blackwell Encyclopedia of Writing Systems* (1996) is different from his earlier publications. It is easy to use because of its alphabetically ordered entries, numerous tables, and illustrations. Although the articles are not usually exhaustive on given topics, the most important information is provided, along with suggested additional readings. While it is not a 'how-to' manual, it is a useful reference work.

Daniels and Bright's *The World's Writing Systems* (1996) is currently the most comprehensive reference book available. It covers a wide variety of topics, including the history of writing, script typology, specific scripts, adaptations of scripts, sociolinguistics, politics, shorthand, and musical notation. There are 74 different sections by almost as many authors. Although the volume was not designed as a guide for linguists needing to deal with orthography design, the case studies and articles are instructive. Of particular interest to practitioners are "Scripts invented in modern times" (Part IX), "Use and adaptation of scripts" (Part X), and "Sociolinguistics and scripts" (Part XI). The section on adaptation of scripts presents many examples from the Roman and Cyrillic alphabets, as well as from Hebrew and Arabic scripts. Several articles deal with script and orthography reform.

*Literacy: An International Handbook* (Wagner, Venezky, and Street 1999) includes two articles which relate to writing systems: Bernard's "Languages and Scripts in Contact: Historical Perspectives" and Coulmas' "Development of Orthographies."

In 2000, the Department of Linguistics at the University of Illinois at Urbana-Champaign published a volume entitled *Literacy and Writing Systems in Asia* with articles by Florian Coulmas, William Bright, Peter Daniels, Richard Salomon, and others. None of the articles present clear principles for orthography design, but the case studies

are instructive concerning orthography reform and politics. In “The Nationalization of Writing,” Coulmas, using examples of Germany, China, Japan, Korea, and republics formerly part of the USSR, shows the important link between writing system choices and identity. He looks at the question ‘Who should decide spelling?’ and explains why North and South Korea embrace different policies.

UNESCO, the main player in the *Education for All* movement, recognizes that good orthography design is a prerequisite for effective educational programs in local languages. UNESCO recently sponsored the publication of two documents that address orthography. *The Manual for Developing Literacy and Adult Education Programs in Minority Language Communities* (Malone 2004) dedicates a chapter to designing orthography for previously unwritten languages. It raises important questions, gives practical guidelines, and presents case studies from different parts of the world. The manual’s strengths are in promoting good practices in educational programs, urging maximum community involvement, and dealing with the question of sustainability. *Writing Unwritten Languages: A Guide to the Process* (Robinson and Gadelii 2003) is a practical document written for grassroots level workers. It deals with linguistic and other influencing factors, including stakeholders. It discusses the technical issues and choice of symbols and provides a brief introduction to the UNICODE standard.

*Writing Systems: A Linguistic Approach* (Rogers 2005)<sup>7</sup> is a readable textbook that provides thorough coverage of the history of writing and writing system typology. It examines some scripts very closely, with particular attention to cuneiform and Egyptian

---

<sup>7</sup> Why the phrase “linguistic approach” is in the title is a mystery.

hieroglyphs. Practical exercises are included at the end of chapters. Designing writing systems for unwritten languages is not addressed.

### **4.3 Specific topics**

Some topics within the field have received special attention in research, resulting in numerous publications.

#### *4.3.1 The link between the reading process and orthography*

In the early eighties, children's poor reading performance raised some concerns. Educational psychologists explored how much of this could be ascribed to teaching methodology and how much to individual limitations. Some suspected that the orthography (lack of consistency in the sound-symbol representation) was the problem. Much research was done relating to reading and orthographic depth, i.e., trying to establish a link between reading efficiency and the degree of abstraction inherent in an orthography. This resulted in many publications, - mostly research reports and anthologies. Two examples of the earlier works are *Perception of Print: Reading Research in Experimental Psychology* (Tzeng and Singer 1981) and *Orthographies and Reading: Perspectives from Cognitive Psychology, Neuropsychology and Linguistics* (Henderson 1984). Most researchers reported on experiments using only Roman script. By contrast, Taylor and Taylor's *The Psychology of Reading* (1983) reports on reading acquisition and the effect of orthographic depth on the reading process involving other scripts as well, specifically Chinese, Japanese and Korean.

A decade later, *Orthography, Phonology, Morphology, and Meaning* (Frost and Katz 1992) presented twenty-one articles on the question of how reading is accomplished. The various authors reported pertinent research results, many of which came from

experiments with languages other than English, including some which use non-Roman scripts. The articles consider how closely print and sound are linked and if the process of printed word recognition differs depending on the script and its orthographic depth. This is an area of ongoing research.

*Scripts and Literacy* (Taylor and Olson 1995) presents a collection of articles which “examine how scriptal differences affect such important issues as learning to read, reading achievement, word recognition, sentence and text comprehension, phonological recoding in reading, brain mechanisms, the role of metalinguistic knowledge, transfer of reading skills to other domains of activity, and levels of literacy in a culture” (Taylor and Olson 1995:5). Diverse scripts are included in the discussions. Roger’s contribution to the volume, “Optimal Orthographies,” provides more of a linguistic perspective than the other articles. The case study by Rice on orthography development for North Slavey of Canada’s Northwest Territories illustrates the complexity of the process and obstacles to standardization. Dialect differences, ethnic identity and script choice were the main issues in this case.

Within the field of Second Language Acquisition, scholars are examining the impact of differences between L1 and L2 orthographies, specifically on how students’ reading and spelling strategies and skills in a second language writing system might be impacted by the writing system they learned first. *Second Language Writing Systems*, provides “an overview of the emerging field of Second Language Writing Systems (L2WS) research” (Cook and Bassetti 2005). The editors express the opinion that “it will interest a variety of readers in different areas of psychology, education, linguistics and second language acquisition research.” This collection has little to help orthography design practitioners

beyond documenting the fairly obvious:

- (1) Ease of transfer depends on the degree of similarity between the systems; and
- (2) Educators need to carefully design curriculum and help teachers compensate for the differences between the systems.

*Handbook of Orthography and Literacy* (Joshi and Aaron 2006) focuses on literacy acquisition. Most of the 43 articles report on reading and spelling acquisition of children around the world. Some report on learning in languages with shallow orthographies, others with deep orthographies. (See 8.3.2 for a discussion of the orthographic depth hypothesis.) The majority of the articles consider languages with Latin-based alphabets, but others are also considered: the Greek and Cyrillic alphabets; Arabic and Hebrew consonantal systems; Chinese morphemic system; Kannada's alphasyllabary; and Korean's Han'gul. The Japanese mixed system is also considered, but the article focuses on literacy acquisition of bilinguals learning English. Part three of the book, "Literacy Acquisition: Instructional Perspectives," is of particular interest to educators.

#### *4.3.2 On writing tone*

Several people have addressed the issue of writing tone (Voorhoeve 1964; Buck 1973; Mfonyam 1988; Wiesemann 1989; Bird 1999a, 1999b, 2001; Yip 2002; Kisseberth and Odden 2003). Particularly interesting are the articles by Bird (1999a, 1999b, 2001), based on his research in Cameroon which indicated that marking tone exhaustively was not efficient. Voorhoeve (1964) had addressed the same topic but did not provide formal supporting evidence.

### *4.3.3 On dialects and standardization*

Complex dialect situations complicate writing system development. For instance, Germano (2004) describes the Tibetan language and its complex linguistic situation. Proponents of the multilectal approach to orthography design believe that such language communities would be best served if all speakers could be accommodated with a single orthography, instead of designing several along dialect divisions or basing an orthography on one particular (reference) dialect. Simons (1977) proposed some possible solutions to complex dialect situations. His ideas have not been put to the test. Some case studies which test the viability of taking a multilectal approach based on Simons' ideas are the following: Fine (2003) describes the complicated dialect situation of the Bouyei language of China and the multilectal experimental orthography currently being tested. Chamberlain (2004) describes the Khengkha language of Bhutan and the multilectal experimental orthography currently being tested in the Tibetan scriptal environment Germano described. No findings have yet been reported.

### *4.3.4 Practical helps for language development workers*

Pike (1947), Smalley et al. (1964), Sjoberg (1966), Berry (1968), Powlison (1968), and Gudschinsky (1973) provide "how-to" material on orthography design and reform. Among the more recent publications, Robinson and Gadelii (2003) and Malone (2004) are particularly practical and helpful for those involved in the orthography design or reform process, as well as the following three articles : (1) Roger's "Optimal Orthographies" (1995); (2) Baker's "Developing Ways of Writing Vernaculars: Problems and Solutions in a Historical Perspective" (1997); and (3) Venezky's "In search of the perfect orthography" (2004).

SIL International, because of its involvement in minority language development, has developed training modules, courses and resources on various topics related to writing system design and reform. Many of these are published in *LinguaLinks Library* (SIL 2002), an electronic resource, which is frequently updated. Modules cover topics such as factors influencing design, phonemic analysis, symbolization, orthography testing, word breaks, and others. This resource also includes all issues of *Notes on Literacy*, many of which have articles and case studies on orthography matters.

## CHAPTER 5

### DEFINITIONS

Before expanding on issues related to orthography, a treatment of terminology and writing system typology is in order.<sup>8</sup> Some terms are commonly, although inconsistently, used in the literature, and need to be defined. This is not as straightforward as one might think.

#### 5.1 Writing

Peter Daniels (1996e:3) defines *writing* as it relates to writing systems as follows (distinguishing it from the more popular uses, which commonly refer to penmanship or composition):

Rather, writing is defined as a system of more or less permanent marks used to represent an utterance in such a way that it can be recovered more or less exactly without the intervention of the utterer. By this definition, writing is bound up with language.

#### 5.2 Writing system

Coulmas (2003:35) distinguishes between two uses of the term *writing system*:

To begin with terminology, the term **writing system** as used in this book has two distinct meanings. It refers to the writing system of an individual language and to an

---

<sup>8</sup> A knowledge of basic linguistic terms and concepts will be presumed.

abstract type of writing system. In the first sense, there are as many writing systems as there are written languages, but in the second sense the number is limited to a few types, such as logographic or word writing systems, syllabic writing systems, phonetic writing systems, or variant forms thereof.

The definition for *writing system* in Daniels and Bright (1996:xliv), “signary together with an associated orthography” is not very accessible for the average reader, requiring two more definitions: that of *signary* and that of *orthography*. ‘Signary’ is not a commonly used term. Coulmas (2003:36) defines it as “the complete inventory of the basic signs of a given writing system.” Daniels and Bright (1996:xliv) point out that the term is neutral, not belonging to just one type of system. ‘Orthography’ will be defined in the next section.

Definitions for *writing system* can be found on the Internet. *WordNet* defines it as “a method of representing the sounds of a language by written or printed symbols [syn: orthography].”<sup>9</sup> This first definition covers two main aspects. First, a writing system is usually language specific. Secondly, it has to do with the symbolization of sound either by machine or by hand. What is missing from this definition is that written language is also concerned with other conventions: punctuation, capitalization, hyphenation etc.

The *Free On-line Dictionary of Computing* provides a definition for a more technically-minded audience: “the set of glyphs used for representing a given human language in written form, generally along with their conventions for use.”<sup>10</sup> The use of the words *glyphs* and *conventions* makes this a broader definition. Character encoding is not

---

<sup>9</sup> WordNet ® 2.0, © 2003 Princeton University <http://dictionary.reference.com/>

<sup>10</sup> The Free On-line Dictionary of Computing, © 1993-2005 Denis Howe <http://dictionary.reference.com/>

limited to matters of sound-symbol representation, but also encompasses punctuation, spacing, paragraph breaks and the rules that govern their usage.

### 5.3 Orthography

As suggested by WordNet's definition of *writing system*, the term *orthography* is its synonym. These two could be used interchangeably in most contexts. In reality however, they are not equivalent. Note the following three entries for orthography in the American Heritage Dictionary:<sup>11</sup>

1. The art or study of correct spelling according to established usage.
2. The aspect of language study concerned with letters and their sequences in words.
3. A method of representing a language or the sounds of language by written symbols; spelling.

It is clear that the main idea associated with the word orthography is *spelling*, i.e., letters in a word. This is a popular concept, one which precludes association with non-alphabetic scripts. A scientific definition needs to encompass the full range of written language. Such a definition unfortunately cannot also be concise. A reference work other than a dictionary may be more helpful. The entry in *The Blackwell Encyclopedia of Writing Systems* (Coulmas 1996:379–80) begins as follows:

**orthography** [Gk ὀρθός 'straight, correct' + γράφειν 'to write']

Correct spelling and that part of grammar that deals with the rules of correct spelling.

An orthography is a normative selection of the possibilities of a script for writing a particular language in a uniform and standardized way. All orthographies are language specific. As the most visible and most consciously learned linguistic subsystems,

---

<sup>11</sup> The American Heritage® Dictionary of the English Language, Fourth Edition Copyright © 2000 by Houghton Mifflin Company. <http://dictionary.reference.com/>

orthographies are often codified by official decree. In alphabetically written languages, the aspects of writing most commonly codified by means of orthographic rules are grapheme-phoneme correspondence, word division, hyphenation, capitalization, and the spelling of loan words. Punctuation is sometimes also subsumed under orthography...

In this document, I will employ *orthography* and *writing system* interchangeably, favoring *orthography* for alphabetic systems and *writing system* for non-alphabetic systems. These terms are intended to encompass all aspects of writing, including character choices for representing sound, rules for ordering, placement and shapes of these characters, diacritics, punctuation, word breaks, hyphenation, use of capitalization if appropriate, and anything else that might be regulated when establishing a written standard.

The term *orthography*, when used in this document, will not automatically include the notion of standardization, as Coulmas (2003:35) proposes. This is because language development is political, i.e., establishing a written standard is a lengthy process involving all kinds of levels of cooperation and infrastructure; and standardization is not a given, especially not for languages in the early stages of their development and history as written languages.

## **5.4 Script**

In dictionaries, the primary meaning of *script* relates to cursive handwriting as well as fonts and print types which imitate cursive writing.

In reference to writing systems, some authors and editors, including Daniels and Bright (1996:xliv) and Sampson (1985:20) employ ‘script’ and ‘writing system’ as equivalent terms. Others use these interchangeably when discussing writing system typology in general, but they do not do so when discussing a specific language’s system,

i.e., ‘orthography’. This can be confusing to readers. Coulmas (2003:35) makes a three-way distinction:

The term *script* is reserved for the graphic form of the units of a writing system. Thus for example, ‘the Croatian and Serbian writing systems are very similar, but they employ different scripts, Roman and Cyrillic, respectively.’ Some scripts are thought by their speakers to be intrinsically related to their language, while others are perceived as serving a variety of languages. The Korean, Yi, and Cambodian scripts are examples of the former, and Roman, Arabic and Devanagari exemplify the latter. The terms *writing system* and *script* are distinguished from *orthography*, which refers to the standardized variety of a given, language-specific writing system.

Thus there is agreement in the following: *script* relates to graphic representation and is not necessarily restricted to a single language. For instance, numerous languages in the world use the Roman script and many languages in the former USSR and Eastern Europe use the Cyrillic script. But some scripts are unique to -- and therefore associated with--a specific language. For example, Han’gul script (also written as Hankul), is used only for Korean, and it is not considered incorrect to refer to it as the Korean script. On the other hand, what is commonly known as the Amharic script or the Ethiopic script, is used for languages other than Amharic and not restricted to Ethiopia. Neutral terms such as *Fidel* and *Ge‘ez*, which do not link the script to a particular language or nation, are preferred.<sup>12</sup> Socio-political issues, such as the desirability of affiliation with a specific script, language, religion, or political unit, play an important role in language planning and orthography design or reform.

---

<sup>12</sup> Ge‘ez is an extinct language still used for liturgical purposes in the Coptic Church.

## **CHAPTER 6**

### **WRITING SYSTEM TYPOLOGY**

There is no shortage of resources on the topic of writing system typology. This thesis with a focus on orthography design and reform cannot provide exhaustive coverage on writing systems. The intention of this chapter is to provide an overview for novices in the field.

“Taxonomies of writing systems have usually been based, at least loosely, on the notion of representation...according to the nature of the linguistic unit that is represented by a grapheme...the minimal functional distinctive unit of any writing system”  
(Henderson 1984:15).

There are basically six different types of writing systems or scripts: logographic, syllabary, consonantal, alphabetic, alphasyllabary, and featural. Classification is not as clear-cut as one might expect. Many writers use the term ‘syllabary’ broadly, categorizing alphasyllabaries with syllabaries. A few question Korean’s Han’gul being singled out and assigned a class of its own (Salomon 2000). Daniels and Bright, who co-edited a reference volume (1996), disagree on terminology for one of the types.

#### **6.1 The notion of evolution associated with writing systems**

It has been thought that writing systems evolved along the following lines:  
Logographic systems are viewed as forerunners of syllabic systems and syllabic systems in turn are expected to evolve into alphabetic systems--the ultimate, ideal systems toward

which all writing systems evolve. Gelb (1952), who classified script into only these three types, believed in this “principle of unidirectional development.” Others have reacted against this. Coulmas (2003:198) states:

Modern scholarship has not confirmed the unidirectional theory of writing unconditionally. The real picture is more muddled, and certain aspects of Gelb’s teleological evolutionism must be rejected. Harris (1986) speaks of ‘the evolutionary fallacy’, while others have criticized Gelb’s theory as an expression of alphabetocentrism, if not Western supremacism.

Daniels (1996e:8) holds Gelb’s work responsible for many misclassifications of writing systems.

Rayner and Pollatsek (1989:45–46) point out that the quality of a writing system is not based on its origins and its classification, but on its performance for a given language:

Thus in an evolutionary sense, the alphabet is ‘fittest.’ It has won out where it has competed...

However such a conclusion must be tempered with two observations. The first is that it is not clear exactly why it is better. For example, there is no good evidence that alphabetic language can be read faster than nonalphabetic languages...

The second observation is that the alphabetic system may be fittest for languages that have adopted the system, but may be less fit for languages that have not. Thus, nonalphabetic writing stems in use today may not be anachronisms, but serve to represent those spoken languages as well as an alphabetic system could.

## **6.2 Logographic systems**

Logographic writing is not picture writing. Picture writing has been assigned the term *proto-writing* (Coulmas 1989:38). By definition, writing has to be associated with the

representation of sound to some degree, not only to meaning. Coulmas provides a concise explanation and some synonyms for logographic writing:

One way of classifying writing systems is by the level of linguistic analysis to which their basic functional units relate. Writing systems whose basic functional units are interpreted as words are known as ‘logographic’ or ‘word writing’ systems.

Alternatively, the term ‘ideographic’ is also commonly used. However, it is doubtful that there ever was a writing system that expressed ideas, as this term would seem to suggest. (2003:40-41)

There is a common misconception that logographic writing is devoid of any information concerning the pronunciation of the written symbols. This is erroneously supported by common dictionary definitions of ‘logogram’. For example:

a written symbol representing an entire spoken word without expressing its pronunciation; for example, for 4 read “four” in English, “quattro” in Italian. Also called *ideogram, logograph.*”<sup>13</sup>

and:

a single written symbol that represents an entire word or phrase without indicating its pronunciation; "7 is a logogram that is pronounced ‘seven’ in English and ‘nanatsu’ in Japanese" [syn: logograph].<sup>14</sup>

Although this claim about the absence of pronunciation clues applies to mathematical symbols and other symbols such as the ampersand (&), names of punctuation signs, or symbols and icons (those which universally encourage people to recycle or to choose the

---

<sup>13</sup> *The American Heritage® Dictionary of the English Language, Fourth Edition*  
Copyright © 2000 by Houghton Mifflin Company. <http://dictionary.reference.com/>

<sup>14</sup> *WordNet ® 2.0, © 2003 Princeton University* <http://dictionary.reference.com/>

restroom appropriate for their gender, for example), it does not apply to the whole set of the signary of a logographic system. By the above definition, a logographic system would be unmanageable, because of the sheer number of symbols it would require.

Logographic writing is most closely associated with classical Chinese, modern standard Chinese, and ancient cuneiform writing which, according to Coulmas (1996:104), was in use for 3000 years in early Mesopotamia and Asia Minor. Henderson clarifies regarding Chinese: “Chinese is usually held to be logographic, or, worse, ideographic, even though the graphemes of Chinese writing represent *morphemes*, and not words or ideas” (Henderson 1984:15-16; italics mine). The use of the term *morphemic* instead of *logographic* is increasing, since it more accurately reflects the level of representation of the Chinese writing system.

Chinese characters vary in complexity due to the number of strokes in their composition. But what contributes most to the complexity of the system is the degree of abstractness of the simple and compound characters in relation to the meaning for which they stand. Most compound Chinese characters, “90 percent or more” according to Coulmas (1996:82),<sup>15</sup> consist of a radical which provides a clue as to a meaning association, plus a phonetic component which provides a clue as to the pronunciation.

Another sound-clue strategy to aid readers is the *rebus principle*. Coulmas (1996:433) gives the following definition:

**rebus principle** Representing a word by means of the logogram of another which is phonetically similar or homophonous, for instance using the sign ☼ for ‘son.’ The rebus

---

<sup>15</sup> Hung, Tzeng and Tzeng (1992) cite a slightly more modest percentage of 85%.

principle played an important role in the development of writing as the cardinal strategy for increasing the expressive power of logographic systems.

The Chinese writing system employs several other strategies to keep the number of characters down to a manageable set, but a more detailed discussion on character types is beyond the scope of this paper.<sup>16</sup>

Because of the meaning conventionally associated with Chinese characters, Chinese literature can serve across dialect and even language boundaries. This is seen as a great advantage because one set of written materials can be used for the wider cultural community instead of being limited to a single language group. Nevertheless, what may appear to be a great advantage is an advantage only for the linguistic community who speak the dialect for which the written standard was established. In the case of modern Chinese, this would be the speakers of the Beijing dialect, also known as *Pǔtōnghuà*. Users whose speech diverges from the Beijing variety need to deal with an increased level of abstraction. They may even have to learn each symbol of the signary holistically, like an icon—an intimidating task, since the phonetic clues are of little or no value to them.

### **6.3 Syllabaries**

As the term suggests, syllabaries are writing systems whose signary is based on the syllables of a given language. The symbol inventory is larger than that of alphabetic systems, but compared to a logographic system, the symbol inventory is relatively small and thus places a lighter burden on the learner. Daniels and Bright (1996:xliv) define

---

<sup>16</sup> Taylor and Taylor (1983:33-53) and Rayner and Pollatsek (1989:46-48) can be consulted for additional information concerning Chinese character types.

syllabary as “a type of writing system whose characters denote syllables, with no deliberate graphic similarity between characters denoting phonetically similar syllables.”

By this strict definition, few writing systems qualify as syllabaries. In such a system, syllables which have a consonantal or vowel phoneme in common, such as /ba/ and /bi/ or /ti/ and /mi/, should display no deliberate graphic similarity. Sampson also holds this narrow criterion and reacts to classifications based on a broader definition: “The point needs making, because scripts are often called ‘syllabic’ which...are nothing of the kind” (1985:64). Because of the more liberal use of the term ‘syllabic’ by others, Sampson refers to the systems which fit into the narrow category as “genuinely syllabic.” (Writing systems with syllable-level representation displaying graphic similarities based on phonemic similarity are frequently referred to as *alphasyllabaries*, or, in the more recent literature, as *abugidas*. See Section 6.6.)

Some examples of syllabic systems follow. Detailed descriptions are outside of the scope of this paper.

### 6.3.1 Japanese

Hiragana and Katakana characters from Japanese fit into this ‘genuinely syllabic’ category. One would be hard pressed to find any graphic similarities between /ma/, /mi/, /me/, and /mo/ in Katakana, and that is what makes this writing system so distinctive.

(See Table 1.)

Table 1. Katakana syllables illustrating phonemic similarity but not graphic similarity

マ /ma/	ミ /mi/	メ /me/	モ /mo/
-----------	-----------	-----------	-----------

The Japanese language is actually written with a mixed system. Mixed scripts will be briefly discussed in section 6.8. Taylor and Taylor (1983, 1995) provide a thorough coverage on the Japanese writing system.

### 6.3.2 *The Mande syllabaries*

The Vai syllabary of Liberia, West Africa is one of five grouped together under ‘Mande syllabaries’. The other four are Mende (Sierra Leone), Bambara (Mali), Loma and Kpelle (Liberia). The Vai script is very much alive and in use.<sup>17</sup> Portions from the Qur’ān and the Bible which have been translated into Vai are written in this script. Mafundikwa (2004) gives good historical and descriptive information on these syllabaries. Ager’s web site (1998–2006f) reports that these West African syllabaries, excepting Vai, are “no longer used” (Loma), “eventually forgotten” (Mende), and “never achieved popular acceptance” (Kpelle). Mafundikwa (2004:70) contradicts Ager in one point: He claims that the Mende syllabary, which is the only one in the group which is written from right to left, is still used for personal correspondence and accounting.

The Mende syllabary is not “genuinely syllabic” by the strict definition because for some syllables with common phonemes, there is deliberate, planned graphic similarity.

Table 2. Sample characters from the Mende syllabary

with graphic similarity			without graphic similarity						
𞑏	𞑏̇	𞑏̈	𞑏	𞑏̇	𞑏̈	𞑏	𞑏̇	𞑏̈	
/ki/	/ka/	/ku/	/ti/	/ta/	/tu/	/kɛ/	/kɔ/	/te/	/tɔ/

---

<sup>17</sup> During the 1970s, adult male literacy rates in Vai (20.3%) exceeded literacy in Arabic (15.7%). For details on literacy acquisition in Vai, use of different scripts in the Vai community (Vai, Arabic, and English), and literacy statistics, see Scribner and Cole (1981).

For about a dozen consonants the basic Mende sign seems to have an inherent /i/, while the addition of one or two dots or a small stroke to the basic shape changes the vowel sound to either /a/ or /u/. This principle breaks down for the remaining consonants and for vowels /e/, /ɛ /, /ɔ/, and /o/. In these cases, the system is “genuinely syllabic.” (See Table 2, based on Ager 1998–2006f.)

### 6.3.3 *Bamum*

Another well-documented African syllabary is the Bamum syllabary of Cameroon, sometimes referred to as Shü-mom. It was invented in 1896 and revised over a period of thirty years by King Njoya of the Bamum people. In its original design it was “logographic, containing 465 signs. Njoya modified his script several times during his reign, each time with fewer and fewer signs, as it slowly morphed into a syllabary using the rebus principle. The final script had just 83 signs: 10 numbers and 73 syllables” (Mafundikwa 2004:87). It is reported that, when the French took control of Cameroon, “they destroyed the printing press that he [Njoya] invented, destroyed his libraries and burned many of the books he had written” (Mafundikwa 2004:83).<sup>18</sup> Eight thousand manuscripts in various forms of this script were preserved and can be found in a museum in the town of Foumban. This script, which was taught in schools in Njoya’s kingdom, has just about died out. In a press release by the United States Embassy in Cameroon in December 2004, Dr. Konrad Tuchscherer, after eight months of research, reported that as few as three people today are able to read the documents written in Bamum.

---

<sup>18</sup> Linguistic oppression is a common phenomenon. Literature burning was not restricted to French colonial rulers: Day (1985:174) reports on Americans establishing compulsory English education on Guam and prohibiting the use of the Chamorro language on school grounds and collecting and burning Chamorro dictionaries in 1922.

#### 6.3.4 *Additional syllabaries*

Additional languages which use ‘genuinely syllabic’ writing systems today are Yi in China, Cherokee in the USA, and Djuka, an English-based Creole spoken in Suriname and French Guiana.<sup>19</sup>

#### 6.3.5 *The question of efficiency of syllabaries*

As seen in section 6.1, the claim of superiority of the alphabetic system has been challenged. This notion is regarded as a type of ethnocentrism. Instead of ‘superiority’ the current question now being raised is that of *suitability*. This latter term implies that, among the options of writing systems, some might work better for a given language than others. The qualifier: the structure of the language.

Although oral communication requires discriminating between sound differences which are and which are not relevant to making distinction in meaning (Jusczyk 1997), children and non-literate adults usually have a more developed notion of *word* and *syllable* than of phonemes (isolated sounds). Olson (1999:134) gives evidence that training in an alphabetic system is required to bring about phonological awareness. (See also Nagy and Anderson 1999.) Ladefoged (2000) refers to syllables as “significant units” and points out that people generally “agree on the number of syllables in the majority of words” and relates this to “peaks of prominence” in an utterance. Rogers (1995:40) states: “It has been argued as well that the syllable is a more salient psychological unit than the single segment.” Linell, commenting on the status of the segment, writes:

---

<sup>19</sup> Djuka has an alternate spelling, Ndjuká, and several alternate names. It is listed as *Aukan*, in the Ethnologue (Gordon 2005).

I have proposed that segments *may* be observed as things, which does not necessarily mean that they normally *are* regarded so... It is primarily larger sound shapes such as word forms or morphs which are observed in the sound signals...

...we may also take into consideration the numerous phonetic investigations which seem to show that syllables rather than phonemes or segments are articulatory primes or perceptual decision units. (1979:66-67; emphasis in original)

Therefore, teaching reading with a syllabic writing system may have some advantages over teaching with an alphabetic system.

The Japanese writing system, which in normal text reportedly consists of about 69% syllabic characters and 30% Kanji—the meaning based characters adopted from Chinese (Rayner and Pollatsek 1989)—is serving the Japanese people reasonably well: Japan has one of the highest literacy rates in the world.<sup>20</sup> Bernard (1999:25), however, is not willing to credit the Japanese writing system with the success of its readers: “The high literacy rate in Japan today, however, is the result neither of the introduction of Kana in the ninth century nor of the Romanized script in the seventeenth but of universal schooling through grade twelve in Kanji and Kana.” Nevertheless, the question is raised: For which type of languages are syllabaries advantageous?

Since, in a syllabary, one symbol is needed for each syllable, languages with simple syllable structures and low phoneme inventories can do quite well with a syllabic system. A language with V and CV syllable structures and a phoneme inventory of sixteen consonants and five vowels would require 85 characters. This is very manageable for learners.<sup>21</sup> If CVC is also an option, this number would increase by 80 for every

---

<sup>20</sup> 99% according to the World Fact Book, published by the USA CIA.

<sup>21</sup> The Japanese syllabary has about 75 symbols (Rogers 1995:40).

consonant that can occur in the coda, i.e., in syllable final position. For example, if /k/, /m/, /n/, and /s/ can occur in coda position, the number of characters required becomes  $5+(16 \times 5) \times 5$ , or 405. If the inventory of consonants is higher and there is additional complexity, such as CCVCC syllables (common in English: stink, blunt etc.), or vowel length, or tone which needs to be represented, the required syllable inventory may become unwieldy, imposing a heavy learning load. For such languages, the ‘genuinely syllabic’ usually gives way to an alphasyllabary or an alphabetic system for practical reasons. The syllabary with the largest character inventory currently in use is Yi. It has 819 characters. Syllables with different tones are represented by different characters (Shi 1996). According to Ager (1998-2006i), the standardized form of modern Yi was officially adopted in 1975, and has been used in schools since 1978. Some of the characters are reserved for use in loan words from Chinese and other languages.

#### **6.4 Abjads**

Abjads are commonly called ‘consonantal systems’ because they represent consonants but not vowels. They usually make provision for optional vowel marking through diacritics, but the default is to omit them. The term *abjad* is found in the more recent literature, but is not yet commonly used or necessarily the preferred term. Coulmas (2003:113) comments: “For the West Semitic languages numerous writing systems evolved, which are generally known as consonantal alphabets, although Daniels...prefers the term ‘abjad’ which is based on the beginning order of letters in Arabic (corresponding to A, B, J, D), the most widely used Semitic script in modern times.” Daniels (1996e:4) has also used the term *consonantary*, an uncommon label.

Coulmas (2003:113) comments that some expressions which refer to consonantal systems suggest that they are ‘defective alphabets’ or ‘incomplete alphabets.’<sup>22</sup> Such terms are pejorative and reflect the notion that alphabetic systems are superior.

The Hebrew, Arabic, Modern Syriac, and Tifinagh writing systems are abjads. The first three are written from right to left, with some of the characters changing shape depending on the position of their occurrence in the word. All four are said to be descendents of the ancient Phoenician writing system, an abjad also referred to as Old North Semitic (Coulmas 1996), but the origin of Tifinagh is the subject of considerable debate.

Vowel notation systems were developed for Hebrew, Arabic, and Modern Syriac. Daniels writes:

The sacred nature of the texts originally recorded in Hebrew, Syriac, and Arabic script, coupled with the need to supplement the abjad with indication of vowels—probably due to the introduction of unfamiliar foreign technical terms from languages like Greek and Persian—led scholars who used the three scripts to introduce vocalic notations that did not corrupt the consonantal text by invading the line of letters. The first script to receive this treatment was the Syriac, then the Arabic, and lastly the Hebrew (Daniels 1996d:486).

We will look more closely at each of these languages in turn.

#### *6.4.1 Arabic*

Arabic script is significant for its widespread use. This is due to the expansion of Islam and use of the Qur’ān,<sup>23</sup> and because it has become the favored script for many

---

<sup>22</sup> As recently as 6/11/2006, “defective” was used online at <http://en.wikipedia.org/wiki/Orthography> in reference to the Italian and Arabic orthographies because these do not “represent all the sounds.”

languages in Islamic nations. It is a cursive script, with many ligatures.<sup>24</sup> Some letter shapes vary depending on the letter's position in the word or the shape of the previous or following letter. Six letters are never joined to the following letter. Stevens writes:

Since pictorial representation was largely discouraged in Islam, calligraphy was the main vehicle for artistic representation. From the seventh century on, Arabic letters were shaped into an incredible variety of scripts and styles...Each letter of the Arabic script was believed to be a work of Allah, and thus calligraphy was the most potent of talismans. (Stevens 1996:244)

Arabic script is made up of 28 basic letters representing consonants. Short vowels are not represented, but three consonants do double duty and can also stand for three long vowels, namely /u:/, /i:/, and /a:/. On this basis, Bauer feels that “the common designation of the Arabic script as ‘consonantal’ is incorrect, since the long vowels are represented but consonant gemination is not” (Bauer 1996:561).

Since Arabic had more consonants than the branch of the Aramaic script from which it developed, certain letters were used for more than one consonant. To eliminate ambiguity, diacritics were added to certain letters. Unlike optional marking of vowels, these marks constitute part of the consonant letter.

---

<sup>23</sup> Arabic Script is “young” compared to most scripts. The date of the oldest document with Arabic script is dated 512 CE (Coulmas 1996:18). The first compilation of the Qur’ān was completed about 651 CE.

<sup>24</sup> A *ligature* is a character consisting of two or more letters combined into one, such as æ or œ. In Indic scripts the term *conjunct* is commonly used for ligature. The degree of resemblance between the ligature and its components can vary considerably. <http://www.thefreedictionary.com> accessed June 28, 2006).

Optional diacritics were introduced to note short vowels and consonant gemination. They are placed above or below the basic consonant. Their use was and is restricted:

They are used throughout the text only in the Qur'ān, less consistently in other authoritative religious texts, in editions of classical poetry and in textbooks for primary education, and occasionally in linguistically rather complex texts to avoid ambiguities. In book titles, letterheads, nameplates, etc., they may be used for decorative purposes. But they are virtually never applied in newspapers, ordinary books, or private documents. (Bauer 1996:562)

Jensen (1970:329) comments on the degree of vowel notation in Arabic script as compared to Hebrew: “Since the Arabic language, in contrast to the Hebrew, has always remained a living language, it did not, and does not, depend upon such accurate reproduction of the quality of the vowels as the latter; instead, the Arabs contented themselves with a rough indication.”

#### 6.4.2 *Hebrew*

From 600 BCE, Hebrew employed the Aramaic consonantal script, which later developed into the Hebrew square script. Coulmas (1996:198, 200) comments:

Both systems represented all Hebrew Cs and none of the Vs. The absence of V letters did not at first impede the comprehensibility of written texts, since the core of the Hebrew lexicon is formed by consonant roots, while Vs mainly indicate inflections which can be supplied by readers on the basis of their knowledge of the language. But Hebrew was replaced as the vernacular language of the Jews by Aramaic, and consequently knowledge of the spoken language dwindled. It became desirable, therefore, to indicate Vs in order to unambiguously represent pronunciation, especially of biblical texts.

Two systems of vowel notations were developed to assure correct public and private readings:

1. *'Plene writing'* or *'full writing'*: In this system four of the consonants were also assigned long vowel values. One represented /a:/, the other three had dual readings: /o:/ or /a:/; /o:/ or /u:/; /e:/ or /i:/. This use of the four consonantal symbols is known as *matres lectionis* or *'mothers of reading'*. This system is employed in modern Hebrew literature, newspapers, magazines, governmental publications.
2. A system of diacritics representing vowels. This is referred to as the pointed system. The pointed system is in use in the Bible, liturgy, poetry, and children's literature.

The vowel marking systems are only used in certain domains. The term *defective* is applied to Hebrew writing which does not employ *matres lectionis*, whether the text is pointed or not (Marquardt 2005). Coulmas (1996:200) comments on the pointed system:

Thus two orthographies of modern Hebrew coexist: the pointed system which specifies Vs; and the unpointed system which omits every indication of Vs, relying on context for their correct identification. Modern Hebrew is usually written without V indication. The vowelized orthography is only used for the Bible and other religious texts, children's books, and poetry.

### 6.4.3 Syriac

Classical Syriac was the first abjad to use diacritics to indicate vowels, and thus had an influence on Hebrew and Arabic. It employed two different systems over time. About 400 CE small diacritical points were used as an integral part of the system. In the 7<sup>th</sup> century, the idea of inserting vowel letters into the consonantal text was considered, but

not pursued. Some time after 700, but before 1000 CE, a unique development took place. Daniels reports: “the vowel letters of Greek could be placed alongside the Syriac consonant letters—above or below as space dictated” (Daniels 1996a:502).

Modern Syriac consists of twenty-two letters plus obligatory and optional diacritics. Like Hebrew, writing that makes use of the full set of diacritical signs is called ‘pointed’, while that which only notes consonants is called ‘unpointed’. Hoberman (1996:505) points out that “unlike Arabic and Hebrew, which are ordinarily written unpointed, Modern Syriac is nearly always fully pointed” thus behaving like a “full-fledged alphabet.”

#### 6.4.4 *Tifinagh*

The ancient Berber script, an abjad used in northwest Africa, is said to have “derived from a Semitic prototype, probably Punic” (O’Connor 1996:112). The forms are quite geometrical. Inscriptions have been found in Tunisia, Algeria and Morocco. The orientation of the script varies: right to left, left to right, top to bottom, bottom to top. The Berber script largely died out and, from 1200 CE on, Berber languages were usually written in Arabic. There is one exception: a mysterious survivor, Tifinagh. This script is “used by the Tuareg for playful purposes, for love letters, family notes, and domestic ornamentation by both men and women, often in settings where the women are not able to read Arabic” (O’Connor 1996:115). The Tuareg are a nomadic people, living in Algeria, Libya, Niger, Mali, and Burkina Faso<sup>25</sup> (Mafundikwa 2004).

Tifinagh is interesting in many ways. It has never been standardized and varies slightly from region to region. Its characters have geometric symmetry. Some letters have

---

<sup>25</sup> *Tifinigh* and *Shifinagh* are common alternate spellings for Tifinagh.

more than one form. Its orientation is not fixed. There are no spaces between words.<sup>26</sup> Literacy rates in this script are amazingly high: 2/3 of Tuareg women are reported as literate in this script in contrast to 1/3 of the men—the exact opposite of male vs. female literacy rate tendencies in the world. Literacy in Tifinagh is independent of literacy in Arabic or other languages. Several books use the script: a grammar description, a dictionary, a manual on camel disorders, a translation of *Le Petit Prince*<sup>27</sup> and portions of the Bible (Souag 2003).

Coulmas (1996:503) reports that “modern attempts at using a modified form of Tifinagh for writing the Berber languages of Morocco and Algeria have not met with success.” O’Connor (1996:115) states the same thing, but refers to a political agenda: “Recent attempts to adapt Tifinagh for serious use in the writing of other Berber languages, prompted by pan-Berber political aspirations, have failed.”

However, Mafundikwa, eight years later (2004:46), reports a change: “In January 2003, the Tifinagh alphabet was adopted over the Roman and the Arabic alphabets by the Administrative Council of the Royal Institute for Amazigh Culture [IRCAM] to teach Amazigh in Morocco. French designer and typographer Pierre di Sciullo digitized the alphabet.” Tifinagh has been taught in primary schools in Morocco since 2003. In

---

<sup>26</sup> The Tuareg language, often referred to as Tamashek has various dialects with varying spellings: Tamajaq in Niger, Tamasheq in Mali, and Tamahaq in Algeria. Grandouiller, a linguist working among the Thammallawat dialect of the Tamajaq language in Niger, reports that among that population, word breaks are normally used, as well as a punctuation mark. At times text is written in concentric circles. Various attempts have been made to add vowel notation to the traditional script, but this has not met with widespread acceptance (Christian Grandouiller, personal communication).

<sup>27</sup> *Le Petit Prince* is a children’s fantasy tale by French aviator Antoine de Saint-Exupéry. First published in 1943, it has been translated into many languages.

September 2005, IRCAM launched a comic book in Tifinagh script aimed at elementary school children to promote learning of this (their third) script (Filali-Ansary 2005).

Tifinagh has been incorporated into the UNICODE standard, and five different fonts have been designed for it. Some of these support ligatures; others don't (McCreedy 2005).

## 6.5 Alphabets

In communities where languages using an alphabetic system are spoken, small children usually encounter the term *alphabet* early in life, either in the home, or in their first year of formal schooling. It is associated with names of letters, a specific order for the letters, and with finding dictionary entries. In this sense, *alphabet* is not equivalent to *writing system*, since graphemes representing certain phonemic sounds are often not part of the alphabet. For instance: consonantal sounds written with digraphs, such as <sh> and <th> do not figure in the English alphabet and are not assigned a separate section in alphabetized reference volumes. Thus, an alphabet might be based on the individual letter inventory, or it may be based on the sounds of a language and their graphemic representation. Although English has 12 or so phonemic vowels plus vowel glides depending on the dialect, the English alphabet includes only five vowel symbols. German has a similar approach to the alphabet, not providing special sections in dictionaries for <sch>, <ei>, or <eu> nor for *umlauts* such as <ä> or <ü>. Yet in some languages, letters with diacritics are an integral part of the alphabet. Icelandic integrates its unique symbols throughout its alphabet, while Swedish lists the three vowel characters with diacritics at the end of the alphabet. (See Table 3.)

Digraphs *ch* and *ll* were previously part of the alphabet in Spanish and had their own section in dictionaries, but since 1994, based on a decision by the Association of Spanish

Language Academies, they are integrated under the *C* and *L* sections respectively (Coulmas 1996:476; Venezky 2004:150).

Table 3. Sample Roman script alphabets

English, French, German	a b c d e f g h i j k l m n o p q r s t u v w x y z
Swedish	a b c d e f g h i j k l m n o p q r s t u v x y z å ä ö
Icelandic	a á b d ð e é f g h i í j k l m n o ó p r s t u ú v x y ý þ æ ö

The word ‘alphabet’ is frequently used to refer to a language-specific signary, even when it is clear that the writing system employed is not an alphabetic system. Even Coulmas, who normally chooses terms carefully—fitting for the typology of the writing system under discussion, used “Hebrew alphabet” and “Arabic alphabet” in captions for some of his tables (1996:15, 21; 2003:119, 124). Upon encountering the term ‘alphabet design’ one should not assume that the script in question is an alphabetic system. The *Blackwell Encyclopedia of Writing Systems* has an entry for *alphabet making*: “Selecting a script and constructing an orthography for a language that never had one. It is a form of encoding the language visually so that it can be used efficiently in written communication” (Coulmas 1996:12).

So, what is distinctive about an alphabet within the typology of writing systems? Simply stated, it is “a type of writing system that denotes consonants and vowels (Daniels and Bright 1996:xxxix),” —and one might want to add the qualifier ‘consistently.’ This definition does not distinguish different ways vowels are represented, i.e., whether with letters or diacritics.

The most widely used alphabetic script is the Roman script, also frequently referred to as the Latin script. Jensen (1970:520) states: “The Latin alphabet...became the script of the western half of the Roman Empire, and through the agency of Roman Christendom it finally came to be the sole script of the majority of Europe (the only exceptions being the eastern and south-eastern areas using the Greek-Slavonic alphabet).”

The Latin script is a direct descendent from the Greek alphabet. It has been in use since the sixth century BCE and originally consisted of 21 letters adopted from Etruscan. It underwent a few modifications such as dropping some unnecessary letters, borrowing symbols, and assigning a different sound value to some letters. By the first century CE it had 23 letters, some of which were also used to represent numerals: I for 1, V for 5, X for 10, L for 50, C for 100 and D for 500. Various creative modifications and conventions introduced made the Roman script a useful tool for a variety of languages on all continents.

Other alphabetic scripts in use today are the well known ones like Greek and Cyrillic, and lesser known ones such as Armenian and Georgian in eastern Europe; Fraser, Thaana and Pollard Miao in China; Osmanya and N’koh in Africa; and Ol Cemet’ in India, among others. Ol Cemet’ is unique among Indic script, simply because it seems misplaced in India where alphasyllabaries are the norm (Zide 1996). (The Roman alphabet is also used in some areas of India due to the influence of missionaries.)

The scope of this document does not allow for expanded coverage on the history, graphical and socio-political issues of the aforementioned alphabets. Various Internet sites provide useful background information and tables with letters and numbers for each of these, as well as for some alphabets which are no longer in use. (See Appendix A.)

## 6.6 Alphasyllabaries

### 6.6.1 Characteristics

Alphasyllabaries display features of both alphabets and syllabaries. They differ from ‘genuine syllabic’ systems in that phonetic similarity is recognizable in the symbols. (See Table 4.) Vowels and consonants are noted. The distinctive characteristic of alphasyllabaries is that the basic consonant graphemes have an *inherent* vowel associated with them—often /a/. Thus each consonant symbol denotes a ‘default syllable’. If a different syllable is needed, a slight change is made, either a stroke modification or addition of a diacritic, indicating the consonant–vowel combination intended. Placement of diacritics or stroke modification is not limited to above or below the basic form; they can also occur to the right or to the left. This may result in the symbol order differing from the order of the sounds they represent in actual speech. It is possible to add more than one modification to the basic form. This type of notation works well for languages with CV syllable structures. It is sometimes referred to as the *akṣara* system.<sup>28</sup>

Table 4. Sample Amharic syllables

basic forms	modified forms		
<b>ቦ</b> /bɜ/	<b>ቦሁ</b> /bu/	<b>ቦ።</b> /bi/	<b>ቦላ</b> /ba/
<b>ተ</b> /tɜ/	<b>ተሁ</b> /tu/	<b>ተ።</b> /ti/	<b>ተላ</b> /ta/

<sup>28</sup> A CV syllable with a short inherent /a/ vowel is sometimes referred to as an *akṣara* or *akshara*, a term from Sanskrit. It is the basic unit of Indic writing systems. (Coulmas 1996:7; Salomon 1996; Bright 1999.)

For CVC or CCV structures, a mechanism to “mute” the inherent vowel associated with the basic consonant form is usually provided: either special consonant-to-consonant ligatures with a simplified or reduced form of one of the symbols, or a vowel-muting symbol, often called a *matra*.<sup>29</sup> If a language also has V-only as a syllable structure option, a ‘stand-alone’ form must be provided for vowels.

### 6.6.2 Terminology

Rather than use *alphasyllabary*, Coulmas (1996:483) opts for an alternative term, *syllabic alphabet*, defining it as: “A writing system which treats the syllable as the unit of representation, while at the same time recognizing the segment as the unit of analysis.” Additional terms used for this type of system are *neosyllabary*, *semisyllabary*, or *pseudo-alphabet*. Daniels dislikes these terms because he would like this writing system type to be viewed as a category in its own right and not as a subcategory of syllabaries or alphabets. He introduced the term *abugida* to denote this type of writing system (Daniels 1996e:4). For pedagogical purposes, the syllables are normally displayed in a chart, consonants varying vertically, and vowels varying horizontally. The term *abugida* is taken from “Ethiopic, from the first four consonants and the first four vowels of the traditional order of the script” (Daniels 1996e:4). Bright (1999:49) acknowledges the aptness of Daniel’s term, but “felt a new term was unnecessary, since ‘alphasyllabary’ was familiar in the South Asian field.”

---

<sup>29</sup> *Mātrā* generally refers to the “modifying vowel signs which are added as diacritic satellites to basic consonant graphemes in Indic scripts” (Coulmas 1996:328). Vowel silencing diacritics often have a separate name: *hasanta* in Bengali (Bagchi 1996:399); *puḷḷi* in Tamil (Steever 1996:427); and *virāma* in Devanagari (Bright 1996a:387) and in Gujarati (Mistry 1996:391). The *virāma* is generally used only at the end of words. The *puḷḷi* is more versatile. It is used in various positions in a word, useful to write consonant clusters as well as closed syllables.

### 6.6.3 *Distribution and impact*

Geographically, abugidas are limited to South and Southeast Asia and northeastern Africa. Scripts in use in Southeast Asia are quite diverse. In India and surrounding nations, the best known abugida is Devanagari. It was developed from Brahmi to write Sanskrit, but today serves many languages, most importantly, Hindi. Tibetan script was based on Devanagari. Some of the other Brahmi-descended scripts in use in India are: Gujarati, Gurmukhi, Kannada, Oriya, Telegu, and Tamil. All scripts native to India are abugidas, with the exception of one: the Ol Cemet' or Santali alphabet. Bengali of Bangladesh and Sinhala of Sri Lanka are also abugidas.

The best known African abugida is *Amharic*, also known as *Ethiopic*. (See Table 4.) Divorced from language or country affiliation, it is also referred to as *Fidel* script or *Ge'ez*, after an extinct classical language used for liturgical purposes only. The latter two terms are preferred by nationals in Eritrea.

### 6.6.4 *Canadian Syllabics*

James Evans (1801–1846), a Wesleyan missionary frustrated with trying to come up with a Latin script orthography for the Ojibwe language, turned to experimenting with a writing system based on syllabic representation. He later adapted it for Cree. Since then, this script has been modified and is used for Blackfoot, Carrier, Inuktitut, Naskapi and Slavey, as well as Ojibwe and Cree (Nichols 1996; Coulmas 1996). Now standardized, these orthographies are used for written communication and serve as a symbol of identity for these First Nations peoples.

These writing systems stand out from others. Also referred to as the *Canadian syllabaries*, or *Algonquian syllabaries*, they do not qualify under the strict definition of syllabary because there is remarkable character similarity across syllables having the

same consonant. Only a handful of symbols are used; nine were originally used to write Ojibwe. The orientation of the consonant symbols indicates the vowel. (See Table 5.) Independent vowels are written as a triangle varying in direction. Diacritics are also used. Consonants denoting closed syllables are superscripted.

Table 5. Sample Cree syllables

▽ /e:/	∨	∧	>	<	< <sup>b</sup>
△ /i/	/pe:/	/pi/	/po/	/pa/	/pak/
▷ /o/	∪	∩	⊃	⊂	⊂ <sup>a</sup>
◁ /a/	/te:/	/ti/	/to/	/ta/	/tan/

Although this system is simple and economical, it is difficult to learn. Near-uniformity in dimension, mirror-image symbols, and the abundant use of diacritics complicate learning this type of writing system. Normally, variety in the shape of symbols is desirable because this facilitates visual discrimination and thus reading fluency. Many teachers can attest to problems with *b* and *d* in Latin script, and this is not limited to dyslexic students. How much more of a challenge it would be to master a system based on symmetrical shapes and differences in orientation? Wiebelt (2004) comments that deficiency of distinctiveness affects readability of text and illustrates how, in the evolution of typography, distinctive features were augmented in Latin characters to decrease confusability and increase readability. Learners of the Algonquian syllabaries need teachers and materials which take the inherent challenges into account.

## 6.7 Featural system

‘Featural’ refers to the association between letter shapes and the points and manner of articulation of the sounds they represent. Because of this close relationship with

articulation, Coulmas (1996:195) refers to such a system as a “phonetic system of writing.” The term *featural* is sometimes used to describe the International Phonetic Alphabet since certain diacritic modifications to the basic symbols systematically represent phonetic level features, eg., dental, palatalized, apical, etc.

In spoken-language writing system typology, *featural* is only applied to a single system:<sup>30</sup> *Han’gul*, employed for writing Korean. The development of Han’gul dates back to about 1440 CE, when King Sejong, with a team of scholars who had researched various writing systems, came up with a writing system that fit the Korean language well. This undertaking was motivated by discontent with the awkward Chinese-based system in use at the time and by King Sejong’s desire to make reading and writing accessible to the entire population instead of only an elite. Han’gul is highly praised by those who study the system, including Coulmas who writes of “the invention of what many consider the most scientific and elegant writing system in general use in any country” (1996:273).

Since both consonants and vowels are noted, most agree that Han’gul is an alphabetic writing system, “the only indigenous alphabetical script of east-Asia” (Jensen 1970:211–213). Some, however, classify Han’gul with syllabaries or with alphasyllabaries for three reasons: the arrangement of symbols is non-linear; the unit of encoding is the syllable; and a syllable matrix is used to teaching reading. It displays consonant-vowel combinations, consonants varying vertically and vowels varying horizontally. Nevertheless, classifying Han’gul as anything but an alphabet is inappropriate: Phonemic similarity parallels systematic graphic similarity and there is no

---

<sup>30</sup> Most writing systems for signed languages, such as HamNoSys and Sign Writing (see [www.signwriting.org](http://www.signwriting.org)) are also featural.

inherent vowel associated with any consonantal forms. Taylor and Taylor (1983:86–87), who discuss reading instruction efficiency associated with different scripts, state:

In earlier times Hangul was taught more as an alphabet than as a syllabary. That is, children learned individual alphabet symbols and their phonemes, plus the rules for packaging them into syllable-blocks. In modern days, syllable-blocks, the actual reading units, tend to be used as teaching units...

The CV Syllable Matrix is possible because of the near perfect grapheme-phoneme correspondence...Once these 399 systematically constructed syllables are learned, a child has no trouble pronouncing any syllable string, whether familiar, unfamiliar, or nonsense. There is no need to consult a dictionary for pronunciation and spelling.

This seems like an oversimplification though, as Coulmas (2003:66) indicates that the Korean language has a rather complex syllable structure, resulting in more than 11,000 syllables, which makes a truly syllabic system impractical.

## **6.8 Mixed systems**

There are two types of mixed writing systems: those which blend the use of more than one script, such as Japanese; and those which are based on more than one level of representation: mixing morphemic with syllabic, syllabic with alphabetic, or phonemic with morphophonemic representation, for instance.

Mixing can occur by design, as was the case in Egyptian hieroglyphics, in which pictograms represented ideas at the morpheme level, while certain characters (phonograms) provided phonetic clues—usually consonants (Ritner 1996).

Mixing can also occur when a system undergoes a gradual transformation. It is a phenomenon frequently encountered in the decipherment of ancient scripts, which, over time, might change from one type into another. Cuneiform, for instance, developed from

a logographic system into a logosyllabary, and later became predominantly alphabetic (Kjeilen 2005). Old Persian script, Cuneiform's most developed form, had thirty-six phonetic characters, seven ideograms, a set of numerals, and a word divider (Testen 1996).

Two modern languages use mixed writing systems: Korean and Japanese. Both have retained some Chinese characters as an integral part of their writing.

Japanese and Korean writing systems are interesting because they make extensive use of syllabaries, and in fact either could be written totally within the respective syllabary system and be comprehensible. However, both systems are hybrids and also use characters derived from the Chinese writing system, even though neither spoken language is similar to Chinese (Rayner and Pollatsek 1989:48).<sup>31</sup>

Like the Koreans, the Japanese had no writing of their own when they encountered and began to absorb Chinese civilization in the +1 millennium. Like the Koreans—indeed under Korean tutelage—the Japanese made shift to adapt Chinese script to a language that was unrelated to, and typologically very different from, Chinese... Unlike the Koreans, the Japanese never made a clean break to a different kind of script (Sampson 1985:172).

Only North Korea has made a “clean break” from Chinese, employing Han'gul only. In South Korea, certain Chinese characters (referred to as *hancha*) have been retained, especially to write Sino-Korean words. Older South Koreans employ the mixed script to a greater extent than the younger Koreans (Sampson 1985:122).

Japanese writing is complex. About 1,850 “official” and about 1000 “unofficial” *Kanji* (Chinese origin) characters have been in use (Rayner and Pollatsek 1989:49). These

---

<sup>31</sup> As seen in sections 6.3 and 6.7, the term “syllabary” is not applicable to Korean writing.

were combined with Katakana and Hiragana syllabary characters. Katakana is used to represent loan words, while Hiragana is used to write grammatical elements, such as affixes, function words, and some content words. In a page of text one might encounter a number of different script types. Taylor and Taylor (1983:76) illustrates this with a sample line from a recipe book in which one encounters Kanji, Hiragana, Katakana, Arabic numerals, and Roman letters. Insup Taylor analyzed a newspaper article, counting occurrences of each type of character. The result was: 57% Kanji; 42% Hiragana (Taylor and Taylor 1983:68).

Reading instruction is of course complicated by the presence of multiple scripts. 88% of Japanese children learn to read Hiragana at home before going to school, and the remainder master it shortly thereafter. Thus one might erroneously conclude that “initiation into literacy is painless in Japan because of the simplicity of its Hiragana” (Taylor and Taylor 1983:63). The authors report on results from a research project:

In a large-scale cross-language study comparing reading achievements of children in three countries, Grade 1 children in Japan scored higher in sentence comprehension and oral reading of text than did Grade 1 children in Taiwan and the United States...

Learning a large number of Kanji is difficult, as many Japanese educators attest... Accordingly, both in primary and middle schools, considerable time and money are expended (*sic*) on teaching Kanji, but without satisfactory results...

Stevenson (et al., 1982)<sup>32</sup> found that whereas Japanese children outperformed Taiwanese and American children in Grade 1, the children from the three countries performed similarly by Grade 5. Beyond Grade 1, learning to read in Japan is concerned with learning more Kanji and more complex uses of Kanji (Taylor and Taylor 1983:63, 66).

---

<sup>32</sup> The reference for the original report on these studies follows:

Stevenson, H.W., J.W. Stigler, G.W. Lucker, S.-Y. Lee, Ch.-Ch. Hsu and S. Kitamura. 1982. Reading Disabilities: The case of Chinese, Japanese and English. *Child Development*, 52:1164-1181.

## 6.9 Multiple scripts for one language

In a few cases, more than one script may be in use for a single language. This may be due to one or more causes:

- a script reform only partially implemented
- deliberate accommodation for readers during a time of transition
- loyalty to a previously esteemed system
- affiliation with a group whose identity is intricately intertwined with a given script
- different scripts being relegated to different domains of usage
- different scripts are used in different nations.

The following situations illustrate some of these.

Following the dissolution of the former Soviet Union, various governments are legislating script choices for languages within their borders. Quite a few are choosing to abandon Cyrillic script. Some, like Azeri (Azerbaijani) and Uzbek, have reverted to Latin script. For Tajiki a resolution was passed to use Arabic (1989), but this has not yet been implemented. A script change takes time and effort to implement and does not happen overnight. The result is that two scripts remain in concurrent use for several languages in the former USSR (Comrie 1996).

Sixty-five years after the Latin script replaced Fraktur (or Gothic script) in Germany (Augst 1996), materials such as Bibles, hymnals, catechisms and some greeting cards continue to be printed in Fraktur to accommodate the older generation.

Although the Latin alphabet was adopted for the Somali language in 1972, some who feel strongly about their cultural identity and heritage continue to use and promote the Osmania script, an alphabet with 22 consonants and eight vowels (SomaliNet 2005).

Serbo-Croatian is written in two different scripts: Latin and Cyrillic. In the former Yugoslavia, skilled readers were expected to master both scripts, since they were in concurrent use. Now, after the political reorganization into two separate nations, Serbia and Croatia, this is no longer the case.

Alphabetic preference reflected the influence of the Slavic Orthodox church in the east, and of the Catholic Church in the west. Government policy was sensitive to regional differences... Reinterpreted in terms of the current political situation, alphabet use in the former Yugoslavia was tied to national identities and religion... Cyrillic has been eliminated from schools in Croatia (except in regions where Serbs predominate), and the use of Roman script in Serbia is diminishing. (Feldman and Barac-Cikoja 1996:769)

Latin script officially replaced Arabic script for Turkish in 1928. Arabic script is held in high esteem and continues to be used in the religious domain of Islam, but only for Arabic texts and in calligraphic ornamentation. Although modern Turkish is not written in Arabic script, old documents in Ottoman Turkish written in Arabic script remain and continue to be reprinted.

In 1946, the Cyrillic alphabet was adopted to write Mongolian. Nevertheless, in Inner Mongolia, the Uighur vertical script, a unique alphabet—the only vertical script written left to right—continues to be used (Baker 1997:121; see Kapaj 2002.).

Sindhi of India is written in two different scripts. (See 8.2.2.) Baker (1997:123) reports: “The greater part of Sindhi literature is written in the Perso-Arabic script, but several Indian educational institutions teach Sindhi in the Devanagari script. Those with a serious interest in Sindhi literature are thus obliged to master both scripts.”

## **CHAPTER 7**

### **PRINCIPLES OF ORTHOGRAPHY DESIGN AND REFORM**

Malone (2004:40) states: “Since each language group’s situation is unique, there are no specific rules for developing a writing system. There are some general guidelines, however.” Smalley (1964b:38) presented the following criteria of an adequate writing system in order of importance, as he perceived them:

- Maximum Motivation
- Maximum Representation
- Maximum Ease of Learning
- Maximum Ease of Transfer
- Maximum Reproduction

Building on Smalley’s list of criteria, Malone states that the ideal is “a writing system that:

- is acceptable to the majority of the Mother Tongue (MT) speakers of the language;
- is acceptable to the government;
- represents the sounds of the language accurately;
- is as easy as possible to learn;
- enables MT speakers to transfer between the minority and majority languages; and
- can be reproduced and printed easily” (2004:38).

Smalley and Malone describe the *desired result*. In the next few sections we will take a look at *how to obtain it*.

## 7.1 Design options

Most writing systems are descendents of previous systems from which they evolved. Some, although new and ‘original,’ were subjected to influence from existing writing systems. When designing a writing system for a language, five basic options are available. Now, however, the only advisable option is *adaptation*.

### 7.1.1 New creation

Writing system experts believe that ‘writing’ was independently invented between three to seven times (Daniels 1996b:2; Coulmas 1999:138). A theory of four is common: Mesopotamian Cuneiform in the middle east, about 3100 BCE; Egyptian Hieroglyphics in Africa, about 3000 BCE; Chinese in the far east, about 1200 BCE; and the Mayan writing system in Central America, sometime between 200 and 500 CE. Those who uphold the theory of only three inventions believe that the Egyptian Hieroglyphics were an ‘inspired creation’ developed by people who had knowledge of Cuneiform writing (Daniels 1996b:2).

### 7.1.2 Inspired creation

Inspired creation happens when a person who has been exposed to writing (but is not necessarily literate or linguistically aware) creates a novel writing system for his/her language. Bernard refers to this local type of initiative as *stimulus diffusion* (1999:26), while Daniels calls it *unsophisticated grammatogeny* (1996c:579). Examples are King Njoya’s script for Bamum in Cameroon (see 6.3.3) and Sequoyah’s script for Cherokee. The lack of formal education on the part of the creator of a script does not necessarily

preclude the effectiveness of the newly created system. Dissemination of such a writing system is fostered when its creator is a leader of the people, or if the people believe the system is divinely inspired. This was the case with Shong Lue Yang, a poor uneducated farmer who, in 1959, invented a script, *Pahawh Hmong*, for the Hmong language.<sup>33</sup> Although a romanized writing system is now more commonly used, Pahawh Hmong continues to be used in Laos and Thailand. Smalley, Vang, and Yang (1990:1) comment on that writing system: “The alphabet he produced...was fully efficient, representing all of the sound contrasts, but with a structure notably different from that of any other writing system we have been able to locate.”

### 7.1.3 *Influenced creation*

Influenced creation happens when educated individuals, from within or without the speech community, make unique and informed design decisions for a specific language. Daniels refers to this as *sophisticated grammatogeny* (1996c:579). Examples are Korean Han’gul (1446 CE) and Somali’s Osmanyā (1920 CE). These were created by leaders from within. The origin of the Sorang Sompeng abugida for Sora of India is attributed to a vision given to Mangei Gomango in 1936; Mangeir Gomango was familiar with three scripts at the time: Oriya, Telegu, and Latin (Zide 1996:613; Ager 2006e). Other writing systems were created by outsiders—frequently missionaries. We note the following examples: (1) Cyrillic, by Cyrillius and his brother Methodius from Greece, who had been commissioned to help design a writing system for Slavic languages (861 CE); (2) the Algonquian syllabaries by James Evans in Canada (1800 CE); and (3) the Fraser script for Lisu in China (1915 CE).

---

<sup>33</sup> Shong Lue was assassinated in 1971 (Smalley et al. 1990).

The influence from other writing systems may be directly reflected in the characters. For instance, the Cyrillic alphabet was based on Greek and borrowed from Coptic and Hebrew; Lisu's writing system was based on upper case Latin letters. In other cases, influence might only manifest itself in the type of system designed, the directionality, or the manner of noting vowels or tone.

Now, new inventions, excessive creativity, and mixing symbols across scripts is inadvisable for technical reasons. It is recommended that options be used from within the character sets provided by the internationally promoted Unicode standard. (See 8.4.2.)

#### *7.1.4 Adoption*

In 'adoption,' the signary and orthographic rules of a system are transferred from one language to another with little or no modification. A system which works well for one language might not serve another language well. Adopting a writing system without due consideration is inadvisable and is likely to lead to failure. It is imperative that a writing system be a good match for the structure of the language it is to serve.

Although morpheme-based characters worked well for Chinese, an isolating language, these did not serve Korean and Japanese well, since these languages are agglutinative. Writing system reforms for Korean and Japanese writing were initiated because it was evident that these languages were too different from Chinese for the logographic system to serve them well.

In francophone Africa, French spelling rules applied indiscriminately to local languages proved not only inconvenient but an encumbrance. The logic behind the rules in French did not carry over to African languages due to differences between them. Doubling <s> except word initially, and using <ou> for /u/ lengthened words

unnecessarily; writing <g> in some environments and <gu> in others, using the same digraph <ou> for both /u/ and /w/, and not marking tone even when this resulted in ambiguities, complicated the acquisition of reading and writing in local languages.

(See 12.9 for a discussion of the effect of French spelling rules on Sango.)

Tibetan script is often adopted for Tibeto-Burman languages because of ethnic and religious identity. Making changes to this highly respected system is resisted, resulting in a mismatch between the written and spoken language. People actually end up reading a proto-form of their language and having to do mental gymnastics to discover the pronunciation and the meaning of the text. This makes literacy acquisition difficult (Chamberlain 2004; see 8.3.1).

One goal of adopting the writing system of another language is ‘easy transfer’ to that language—usually a lingua franca. Coulmas comments on the ‘cost’ this may incur:

Where the phonology and other structures of a particular language differ very much from those of the major contact language, every feature that favors transferability frustrates faithful mapping. The Cyrillic orthography for Karakalpak, a Turkic language of central Asia, is a typical example. At a linguistic conference in 1954 it was condemned for its inadequacy...In spite of great differences between Russian and Karakalpak, the Cyrillic alphabet was used in a way very similar to the Russian orthography. Transferability was high, but the faithful representation of speech was low. (1989:236–237)

### *7.1.5 Adaptation*

Adaptation is similar to influenced creation: informed decisions are made for a specific language. However, with adaptation there is a little less creativity and divergence from the ‘model’ than would be the case with influenced creation. This is the approach of choice when designing a writing system for a previously unwritten language. Inherent

qualities of a proven system are carried over, unnecessary burdens done away with, and slight modifications made to tailor-fit the language at hand.

This has proven effective for many languages. When a writing system was needed for Latin, the Greek alphabet provided a good starting point. With some minor changes, the Greek alphabet served Latin well.

The Latin script, in turn, was the foundation upon which other languages built their writing system. The Latin alphabet only provided five vowels. Many languages which wished to use it needed to augment the number of symbols, especially the vowel inventory. Unfortunately, there was no uniformity in their approaches, complicating reading and spelling for multilingual individuals.

Abjads work well for Arabic and Hebrew. Consonants constitute the root of every word, and all cognates build on the same letters. Vowels are usually predictable from syntactic as well as contextual clues. This is characteristic of Semitic languages. If vowels are not noted for non-Semitic languages, writing as a means of communication could be rendered ineffective. In regions where Arabic is the script of choice, one might try an Arabic-based abjad, testing the functional load of vowels and the amount of ambiguity which results if vowels are not noted. (See 7.2.1 for a discussion on functional load.) If less than satisfactory results are obtained, the abjad could be adapted into an abugida or alphabet if this is agreeable to the local population. As with Arabic and Hebrew, special concessions likely need to be made for beginning readers.

## **7.2 Linguistic factors which influence orthographies**

Less than ideal orthographies are in use the world over. It is not wise for a linguist to enter a tug-of-war with others, insisting on an elegant, linguistically perfect orthography.

There is nothing ‘elegant’ about a halted literacy program or stockrooms of literature no one wishes to buy and read. So, ‘good, better, and best’ are rather fluid terms as they relate to orthography alternatives. At the same time, although political and sociolinguistic factors weigh heavily on orthography design, linguistic analysis must not be neglected. It provides a point of departure for good decision making.

In years past, writing was thought of as ‘representing speech’. The concern for “accurate representation of speech” arose in the late 18th century and carried over to the early 20th century (Venezky 1982:271–273; Baker 1997:94). During that era, in orthography design linguists were primarily concerned with phonemic analysis (phonemics) and one-to-one sound-symbol representation. Phonemic writing, i.e., a phoneme-based orthography, was held as an ideal by early SIL field workers, including Pike and Gudschinsky. Consistent representation at the surface level was believed to facilitate learning to read. It was viewed as advantageous since it permits the sounding out of unfamiliar words.

Although phonemic analysis is still seen as foundational, the general approach to orthography design has changed. Writing is now viewed as representing *language*; the emphasis is not only on speech and sounds (Rogers 1995:35). When making recommendations for orthography design or reform, linguists are now giving increased attention to factors such as (1) underlying form and morphophonemic processes, (2) mother-tongue speaker perception and intuition, and (3) factors which facilitate reading fluency for experienced readers, not only beginning readers.

### 7.2.1 Phonemic analysis

Jusczyk, writing about language acquisition of infants, writes, “it is not enough to know which sounds appear in words in one’s native language; one also has to discover which differences among these sounds are relevant for conveying distinctions in meaning” (1997:9). Similarly for written representation, some sound differences are significant while others are not. Not all speech sounds in a language need to be represented. By doing a phonemic analysis, a linguist seeks to discover which ones need to be written, and which do not. Smalley (1964b:37) explains this in non-technical terms:

The principle involved is this. Any language has a limited number of sound distinctions. It has a very much larger number of actual sounds than there are sound distinctions. The distinctions we call *phonemes*. A technically good writing system is one that represents each of the sound distinctions with a single symbol. It does not differentiate the nondistinctive differences of sound. A foreigner in learning the language may need some phonetic representation of the non-distinctive sounds in order to remind him of what they are, but the native speaker does not need that representation.

Anyone doing linguistic analysis and involved in making orthography decisions needs to understand the difference between *phones*, *phonemes* and *allophones*:

A *phone* is “the smallest sound unit identifiable in spoken language and for which phonetic transcription is used;” A *phoneme* is “a minimal sound unit of speech that, when contrasted with another phoneme, affects the meaning of words in a language” (Harris and Hodges 1995:183). Harris and Hodges explain that “the phoneme is an abstract concept manifested in actual speech as a phonetic variant.”

The differences between certain sounds may be significant in some languages and not in others. In English, [t] and [t<sup>h</sup>] are not considered separate phonemes but allophones of the phoneme /t/. In English, where [t] or [t<sup>h</sup>] occur is predictable; they occur in

complementary distribution. In English, aspiration does not affect a change in meaning, thus the two sounds do not need separate representation. Thai has these very same sounds, [t] and [t<sup>h</sup>], but where they occur is not predictable. Both can occur in the same linguistic environment. In Thai they are separate phonemes /t/ and /t<sup>h</sup>/, and should be given separate representation.

Allophones can result from 1) sounds being affected by their location in an utterance—i.e. word initially, word finally, intervocallically etc.; or 2) influence from neighboring sounds. Due to the physiological limitations of the vocal tract and dynamics of articulators, neighboring sounds have a tendency to assimilate. For instance, English /k/ is pronounced farther back in the mouth when it precedes or follows back vowels such as /u/ and /ɔ/ and /ʊ/ than when it precedes or follows front vowels such as /i/ and /ε/. Spectrograms clearly display these and other differences between pronunciations. A linguist in a phonological description would describe the phenomena which result in such sound differences. However, such differences are non-significant for meaning or written communication; speakers of the language, unless they have had training in phonetics, are usually not aware of these differences, i.e., they are below the speaker's level of awareness. The danger is to analyze allophones as separate phonemes and to burden mother tongue speakers with a distinction of which they are not conscious.

If two phones of a language *contrast*, either in identical or in analogous environments they are separate phonemes. An example of identical environment to establish phonemic contrast between vowels /ɪ/ and /ε/ in English is the word pair *pig* and *peg*. Such word sets are often referred to as minimal pairs. If a minimal pair cannot be found, a similar

pair, in which the linguistic environment is analogous (near identical) can be used to establish contrast.

Because loan words often behave differently when taken into a language relatively recently, it is preferable to exclude them from the analysis when establishing phonemic contrast. However, words which have assimilated to the language's sound system and are used freely by monolingual speakers need not be excluded.

When looking for distribution patterns to identify phonemes based on contrast, it is the sounds which are similar (in voicing, place of articulation or manner of articulation) which need close scrutiny. They are at times referred to as *suspicious* pairs or sets. Establishing contrast is not only required for consonants, but also for vowels. Vowels can be affected in height, front/backness, aperture (openness), lip rounding, and nasality depending on the features of the preceding or following sound.

Unconditioned *free variation* is a type of allophony. It is defined as “the ability of a sound to be substituted for another in the same context without change in meaning or function of a word” (Harris and Hodges 1995:88).

Once the phoneme inventory of a language has been determined, symbols can be assigned for each phoneme. Only one symbol is needed to represent sounds that occur in free variation; only one symbol is needed to represent allophones occurring in complementary distribution. Ideally, phonemes are assigned separate symbols, resulting in a phonemic writing system, one in which there is a one-to-one phoneme-grapheme correspondence. This makes reading and spelling as predictable as can be. There are certain conditions which would favor deviating from the one-to-one principle. These will be presented in upcoming sections.

Representing two or more phonemes by a single symbol, for instance <e> for vowels /e/, /ɛ/, and /ə/, constitutes underrepresentation, or underdifferentiation. It violates the orthography design principle of ‘accuracy’ and makes the reading process less based on sound, requiring more dependence on context. For example, in English, if /i/ and /ɪ/ were not represented in a different manner, <fit> might stand for *fit* and *feet*. Context would need to supply the clues to disambiguate between them.

Representing a phoneme with more than one symbol, for instance /k/ with <k>, <c>, <ck> and <q>, constitutes overrepresentation, or overdifferentiation. This offends the principle of ‘economy’ and requires writers to learn spelling rules. If the distribution of the graphemes is random, spellings need to be memorized.

The further a writing system is removed from one-to-one sound to symbol representation, the more opaque or abstract the system becomes. Before continuing this discussion, an explanation of *functional load* is necessary.

Functional load relates to the frequency of occurrence of a phoneme or feature and the resulting potential for confusion if it is not uniquely represented. “If many common words are distinguished on the basis of a given phonological element (phoneme, stress, or tone), that element has a high functional load. If only a few words or word pairs in a language contrast for the element in question, that element has a low functional load” (SIL 2002). Crystal (1997:161) comments: “Several criteria are used in making such quantitative judgements, such as the position within a word at which the contrast is found, and the frequency of occurrence of the words in the language.”

Underdifferentiation may not be problematic if:

1. the functional load of the phoneme or feature (such as stress, tone, length, or nasalization) is low, and
2. no homographs (words which look identical but have different meanings) result which would likely be confused in context.

Under these conditions only a handful of words would be affected, which would then need to be learned holistically. The English orthography does not distinguish between /θ/ and /ð/. Both are written as <th>, e.g., /θɪn/ <thin> and /ðɪs/ <this>. This has not caused problems for three reasons:

1. the distinction between these two sounds has a low functional load, i.e., relatively few words contain either sound.
2. the two sounds tend to occur in different types of words: /ð/ occurs almost exclusively in grammatical function words as part of a noun phrase, which are normally taught holistically (the, this, that, these, those); /θ/ occurs mainly in content words and larger phrases (thank, think, thimble, thumb, thermometer, thistle, through etc.).
3. word medially /ð/ is more frequent in common words (father, mother, brother, gather, rather, other, leather, lather; exceptions occur in less frequently occurring words, such as catholic, catheter, mathematics).

This discussion on underdifferentiation in English<sup>34</sup> was not included here to give license for such a practice when deciding on a writing system for an unwritten language.

---

<sup>34</sup> Venezky 2004:142 discusses an additional type of underdifferentiation, that of underdifferentiated phonemes occurring only in loan words.

Rather it was included as a consolation, in case it cannot be avoided, due to pressure from non-linguistic factors (discussed in Chapter 8), or when accommodating various dialects. (See 7.2.8.)

Representing a single phoneme in a variety of ways (overdifferentiation or overrepresentation) is not ideal and should be avoided. However, it does not present a problem for reading per se: it is not difficult to teach that <c, k> are both read as [k]. The difficulty comes in writing. Only under special circumstances (deadlock in negotiations or designing a writing system for multiple dialects, for example) should it be considered. Writers will have an easier time if the distribution of the graphemes is predictable from a simple rule.

### 7.2.2 *Phonological processes*

Identifying phonemes is only the first step in analyzing the phonology of a language. Under certain conditions, phonemes can change their form. For instance, in German, /d/ and /t/ are separate phonemes: /du/ ‘you’ and /tu/ ‘to do’ indicate contrast. However, /d/, evident in the plural noun [hʊndə] ‘dogs’ is realized as [t] in the singular: [hʊnt] ‘dog’. The /d/ also appears in the form indicating a female dog [hyndɪn]. This suggests that the word phonemically is /hund/ not / hunt/. The surface form differing from the underlying form is due to a devoicing rule in German: voiced consonants become devoiced word finally. Thus, [bʊnt] can mean either ‘colorful’ /bunt/ or ‘union’ /bund/. The question then is, should one write <hunt> or <hund>; <bunt> or <bund> for ‘union’?

Phonological processes can be grouped into two broad categories:

- (1) those resulting in a sound change of which native speakers are usually aware.

These are called *lexical rules* in the theory of Lexical Phonology.

(2) those of which native speakers are not usually aware. These are called *postlexical* or *phrase rules* in Lexical Phonology.

In the case of German, the devoicing rule is post-lexical. When regular post-lexical processes are at work, the general guideline is to write the underlying ‘lexical form’ in the speaker’s mind, not the surface form which resulted from postlexical rules that were applied. Therefore, this predicts that we should write <hund> for ‘dog’ and <bund> for ‘union’. Paying attention to mother tongue speakers’ intuitions will be very helpful in this respect.

We note that, although one-to-one sound symbol representation is basically desirable, an optimal orthography which writes underlying lexical forms departs, in this respect, from phonemic writing. English uses some of this type of representation, but not consistently. For instance, the plural in *hearts* and *hands* is uniformly written <s>, although the pronunciation varies: [s] and [z] respectively. Derived forms of certain verbs behave in the same manner: kick/kicks [s]; build/builds [z]. This approach preserves a single spelling for a given morpheme.

When words have irregular derivations or inflections (i.e. there are *exceptions* to one general rule) or a phonological process takes place of which mother tongue speakers *are* aware, it is preferable to write the surface form. These type of changes are said to be due to lexical rules and should be written. Some examples would be <wiyes>, not <wifes>; <mice> not <mouses>; <kept>, not <keeped>.

As illustrated, English has a mixed system of representation. Some features are represented on the phonemic level, some on the morphemic level. The prefix *in-*,

denoting negation, is not treated the same way as plural <-s>, i.e. a single form is not preserved. It is written as *in-* in ‘inconceivable’, ‘invariable’ and ‘indecent’ but *im-* in ‘impossible’ and ‘immature’; *il-* in ‘illogical’ and ‘illegal’; *ir-* in ‘irresistible’. The occurrence of each is predictable, based on the first phoneme of the stem to which it attaches. Speakers are aware of the rule, and thus the result of the rule should be written. They are aware that the rule does not apply to the negation affix *un-*: unruly, not \*urruly.

For orthography decisions, mother-tongue speaker awareness will be the most crucial criteria in deciding whether to write a surface or underlying form. Linguists may wish to use different criteria to determine whether a change-producing rule is lexical or postlexical. Additional factors to examine are whether the phonological rule producing the change refers to word-internal morphological structure, is sensitive to morpheme boundaries, applies across words or not, or is dependent on other phonological changes having taken place. These are criteria used as linguistic indicators offered by the theory of Lexical Phonology. (See Mohanan 1986, Pulleyblank 1986, Snider 2001.)

Here is a summary of the principles which have arisen from the discussion thus far:

1. Linguistic analysis must not be taken lightly.
2. Phonemes, not phones, should be written.
3. Differences which mother-tongue speakers do not perceive should not be written.
4. Sound differences perceived by mother-tongue speakers should be written.
5. Phonemic and morphophonemic analysis can provide valuable predictions about what sound differences are likely to be perceived.
6. Orthography decisions should take into account the intuitions of the speakers and their needs and preferences, and not be based on foreigners’ needs or desires.

7. Under- and overdifferentiation should generally be avoided; underdifferentiation is allowable when functional load is very light; overdifferentiation should be used only in exceptional situations.

### 7.2.3 *Retention of a morpheme's identity*

By not writing the surface form for words which have undergone postlexical rules, the image of the stem is preserved instead of having several forms. This helps experienced readers to more automatically access meaning, and thus promotes fluency.

Nida (1964b:25) advocates preserving word forms, when possible, especially when changes are affected across word boundaries:

This principle of unity of visual impression is not a warrant for regularizing the grammar of a language, thus smoothing out all kinds of anomalous forms. This principle only means that when the sounds of contiguous words affect each other in purely automatic ways, there may be some distinct advantage in preserving the basis [*sic*] form of the word than writing it in a number of different manners.

Venezky (2004:146) refers to this as the *constancy principle*, which he describes as “a principle whereby the spellings for many root morphemes remain constant even though their pronunciations may change under derivation and inflection.” He provides some examples of constancy which were introduced in the most recent (1996-2006) German orthography reform. One example was <Mop> ‘mop’, previously written with one <p>, now to be written as <Mopp> in order to preserve visual unity with the verb <moppen> ‘to mop’.

Regularity of this kind is helpful to experienced readers. Venezky (2004:157) comments: “The higher frequency words are recognized as units while the low frequency ones require attention to sub-components.”

Snider (2001:324) writes:

Mature readers read by sight...and do not take the time to sound out the words they read. For this reason a good orthography maintains a constant word-image. This helps minimize the effort it takes for a developing reader to memorize the shape of each word. By the same token, beginning readers often do sound out their words. For this reason, a good orthography bases the spelling of its words as closely as possible on the way the native speaker perceives the words to sound.

#### 7.2.4 *Complex systems*

Inherent complexity in a system can be due to a variety of factors: exotic sounds, word order, fusion of elements, differing prosody, and tone among others. These may present obstacles to quick and easy linguistic analysis and to language acquisition for outsiders.

Complexities in the system warrant description. Whether or not they need representation in the writing system needs to be determined.

The sound systems of Nilo-Saharan languages contain various complexities: Breathy vowels, vowels with advanced tongue root, vowel length, implosives, geminate consonants, preglottalized and prenasalized consonants, palatalized and labialized consonants. Complexity is added by tone and stress patterns which do not always behave in obvious patterns, and phonological processes such as vowel harmony, vowel diminishing (reduction in length and strength), nasal assimilation, nasal harmony, and elision (Gilley 2004). Because of these complexities, some mother tongue speakers “having tried, have come to believe that their language cannot be adequately written” (Hollman 2004).

Some Bantu languages display complexities in their sound system: Some have whistling fricatives, depressor consonants (consonants which cause extreme lowering of tone), clicks and unusual sequences of consonant articulations. Some have two sets of 'high vowels' instead of the normal pattern, with the second set affecting adjacent consonants differently than other vowels (Maddieson 2003). Vowel length can have a variety of causes, vowels elide, and consonants can change into unlikely forms ( e.g., in Kalanga /l/ becomes [g] before [w]). Because Bantu languages are agglutinating rather than isolating, morphophonemic processes arise at morpheme boundaries. Some sounds assimilate, some fuse, some disappear. Tone patterns change. At times, words practically become unrecognizable for all the phonological changes. Not only adjacent sounds affect each other. Some sounds trigger changes from a distance, i.e., from the next syllable (Hyman 2003).

No one trying to develop a body of literature or educational material for a previously unwritten language should be so foolish as to make orthography decisions before doing a careful linguistic analysis. As indicated, this will be more of a challenge for some languages than others. Linguists or linguistic consultants should not begin language analysis and writing system design without advance preparation. Studying the work of those who have gone before and written on a particular language or related languages, and examining the orthographies of related languages and their relative efficiency can be helpful. Testing writing options for complex features is essential.

#### *7.2.5 Tone, register, accent, and stress*

We will now consider four pitch-related phenomena: tone, register, accent, and stress.

Yip (2002:1) writes: “A language is a ‘tone language’ if the pitch of the word can change the meaning of the word. Not just its nuances, but its core meaning.” Coulmas (1996:509) defines *tone* as “distinctive pitch at the syllable level” and describes it as “a sense-discriminative feature which distinguishes lexical meanings.” It is estimated that “as many as 60–70% of the world’s languages may be tonal” (Yip 2002:17). These languages are mostly found in three areas: in East and South-East Asia, Sub-Saharan Africa, and southern Mexico. Tone systems vary in complexity from one language to the next.

Coulmas (1996:509) notes that, considering the number of tone languages that exist, “only a few writing systems have developed devices for the representation of tones.”

There are three common reasons for not marking tone:

1. Analysis of the language indicated that the language is not a tone language as first thought: tone is predictable, not phonemic: i.e., it is a redundant feature of a phonological process, or co-occurring with some other linguistic feature, and thus need not be written. For example, Munro (2003) indicates how specific vowel patterns in Valley Zapotec are uniquely associated with one of four tones: high, low, rising, or falling. Her conclusion: “thus, tone need not be specially marked.”
2. Tone was shown to have a low functional load leading to the decision to not represent it.
3. The linguist or language community may have a general resistance to noting tone, based on aesthetics, or on historical factors or on what other languages in the region are doing.

“The people can read without tone being marked” is the standard pat answer when inquiry is made as to why tone is not indicated in the writing system of a tone language.

Refusals to mark tone for reasons not founded on linguistic analysis or careful orthography testing may be just an excuse. It might be that if tone *were* marked—in some manner—those who “can read without tone being marked” might be reading and writing better, and maybe a higher percentage of people would succeed in learning to read. In sections 7.2.6 and 7.2.7 I will discuss whether or not tone needs to be represented in a given writing system, and if yes, to what extent. We will consider the effects on readers if tone is not marked when it is needed, and what may happen when it is marked excessively.

A ‘register system’ makes use of tone, but the function of tone is limited and not phonemic. Chamberlain (2004:119) alludes to a continuum of the use of tone in languages. He comments that languages related to Tibetan “display various levels of tonal development, from non-tonal to fully tonal.” Some of these have a “binary register system.” A high pitch, or tenseness in a syllable can be linked to specific phonetic features of the onset of that syllable. A different set of phonetic features is associated with low, or lax, register. Thus, register is predictable, not phonemic, and need not be marked.

Some languages belong to a subclass which are often referred to as ‘accentual languages’ or ‘accent languages.’ Japanese and some Mexican languages fall into this category. Wiesemann (1989:17) compares and contrasts accent languages with tone languages:

In accent languages, the situation is somewhat simpler than in tone languages. In tone languages, each syllable has at least one, or even a whole string of tones, which are realized on that syllable by some kind of level or glided tone. In an accent language, there is one accentuated (or high tone-carrying) syllable per word. (Occasionally there might be more than one.) Once it is determined which syllable carries the accent, the

tone of all the other syllables is predictable in relation to that accentuated syllable. Often, there is a basic form for the lexical items; a lexical accent. There might be minimal pairs, similar to the minimal pairs in tone languages. There are also grammatical changes of accent, where the accentuated syllable changes from first to second, from penult to antepenult or final, or whatever. The changes occur in the same environments as the grammatical tones. In addition, there might be special rules that accompany verb derivation, where verbs become nouns or other verbs by adding more syllables or changing the accent.

The rules of accentuation need to be researched, in order to know how to write the accent. Normally, the accentuated syllable should be indicated... If the accentuated syllable changes position, the writing rules should coincide with the pronunciation... The writing should be systematic: if accent needs to be written at all, it needs to be written whenever it is pronounced. If the research reveals that the language has neither tone nor phonemic accent, nothing special needs to be marked.

Yip (2002:4) also elaborates on the behavior and classification of accent languages:

These languages have only a small number of contrasting tones (usually only one or two), and these are sparsely distributed or even absent on some words and usually belong to specific syllables, from which they are inseparable. There is no absolute division between accent languages and tone languages, just a continuum from 'accent' to 'tone' as the number and denseness of tones increase, and they become freer to move around.

Stress, another prosodic feature, also needs to be analyzed to see if it needs to be written. As with phonological segments, analysis will indicate whether stress is phonemic or not, and if it is, if it has a high or low functional load. In English, stress has a low functional load. Marking it would be redundant information since stress is predictable on the basis of morphological information: one member of a set is a noun and the other a verb. Some examples are: *'contrast* vs. *con'trast*; *'contest* vs. *con'test*; *'object* vs. *ob'ject*; *'combine* vs. *com'bine*; etc.

When stress is predictable as in French (always on the last syllable), it need not be written. When there seems to be a regular pattern with a few exceptions as in Spanish (almost always on the penultimate syllable), stress need not be written, except for the exceptions. Dye (1988) reports on a language from Papua New Guinea, Bahinemo, in which reading fluency was greatly hindered by the failure to write stress. The decision was made to indicate word stress and sentence stress. This helped fluency and natural intonation in oral reading.

When stress and accent are phonemic, carry a significant functional load and need to be indicated, they are usually written with an acute accent over the vowel or an apostrophe or accent mark somewhere in the accented syllable.

#### 7.2.6 *Marking tone*

In Mandarin and Cantonese a single syllable can take on as many as six to twelve different meanings. Tone pattern variation causes some of the changes in meaning; some of the tone patterns also result in homonyms (Baker 1997:124-125). The traditional Chinese logographic (or morphemic) system does not indicate tone because the written representation is not closely linked to pronunciation. In the traditional Sinitic system, the visual representations of a minimal tone pair in Chinese, although acoustically differing only in tone, may look nothing alike. In an alphabetic system, however, noting tone is necessary for languages in which tone is phonemic and carries a high functional load. If tone were not marked, too much ambiguity would result. *Pinyin*, the system for writing standard Chinese (Pǔtōnghuà) with the Roman alphabet since 1958, indicates tone. Pinyin marks tone on the vowels with four different accent marks: macron for level

high tone, <ā>; acute accent for high rising tone, <á>; wedge (caron) for falling-rising tone, <ǎ>; grave accent for high falling tone, <à> (Coulmas 1996:408).<sup>35</sup>

Chinese illustrates *lexical tone* well, i.e., how tone functions to discriminate between entries in the lexicon. Another language with lexical tone is Fur, a Nilo-Saharan language spoken in Western Sudan. Kutsch Lojenga and Waag (2004) states that the lexical function of tone is quite heavy and that tone is written in the orthography. There are two level tones, high and low, and two contour tones. Low tone is unmarked. An example of a lexical minimal tone set from Fur is <kɔrɔ> ‘water’, <kɔrɔ́> ‘monkey’, and <kǎrɔ> ‘parrot’ (Kutsch Lojenga and Waag 2004:22).

*Grammatical tone* serves a grammatical function, such as indicating person, number, tense, or aspect, instead of lexical function. Fur, which has lexical tone as indicated, also makes use of tone in its grammar. While <zalda> means ‘rock’, a change of tone <záldá> changes the noun into a locative ‘on the rock’. We note that the locative meaning is not necessarily linked to a specific tone pattern, but to the fact that the pattern is different from that of the basic form: <kurú> ‘hill’ vs. <kúru> ‘on the hill’. A difference in tone can also change the person and tense/aspect in the verb:

<ísó> ‘collect, 1S Past’

<iso> ‘collect, 3S Past’;

<iso> ‘collect, 1S Subjunctive’

<isó> ‘collect, 3S Subjunctive’.

---

<sup>35</sup> Although Pinyin is taught in elementary schools, it is currently used in only a limited number of domains. It is used mainly for proper nouns, such as names of people and places, and in language acquisition courses. Libraries and international organizations use it for transliteration, referencing, and archiving.

Linguists have taken different approaches to marking tone:

- a) Tone might not be noted at all. This is referred to as zero tone marking by Bird (1999a, 1999b) and also Bernard, Mbeh, and Handwerker (2002).
- b) Tone may be noted fully (exhaustively), i.e., everywhere, on every syllable.
- c) Tone may be marked minimally—in some reduced fashion, either consistently distinguishing all minimal tone pairs, or distinguishing minimal pairs only when the context would allow confusion.

In the past zero marking of tone was simply a symptom of the impact European language writing systems had on alphabetic writing systems around the world. Bird (1999b:104) comments: “Historically, the linguistic naïveté of the colonists, coupled with the local desire to conform with high-prestige colonial languages has meant the adoption of zero tone orthographies in many countries.” A reaction to this in recent years was to go to the other extreme, i.e., full marking. Bird (2001a:142) suggests that this was an identity issue, not based on linguistic research or concern for writing system efficiency, citing an example from Cameroon where people were told “they should not consider themselves to be true Cameroonians unless they wrote using tone marks, since these are distinctively Cameroonian—whereas writing without tone marks is European.” After experimenting and finding less than satisfactory reading and writing skill levels in languages which mark tone exhaustively, Bird (1999a:3) gives this exhortation: “The assertion that tone should be marked in tone languages does not have to be accepted uncritically.” Bird explains that zero tone marking might work because “there may be abundant non-tonal clues to the identity of a word or of a grammatical function.” Contextual clues and redundancy make

it possible for readers to choose the correct alternative among various possibilities. (This is similar to redundancy and context clues allowing zero marking for intonation and for stress in English). The reliability of clues as an aid depends on *where* they appear in a written text. If clues precede the spot where tonal discrimination is needed, the reader is not disadvantaged by not marking tone. If, however, clues are not available when needed (because they are yet to come and found outside the visual field), it is advisable to mark tone in some manner. Bird (1999b:93) comments regarding contextual clues:

“Orthographic ambiguity can be resolved without silent reading ahead for contextual clues if there is sufficient disambiguating information inside and to the left of the processing window.”<sup>36</sup> For any system, clues have to be in ‘what has already been seen’ or ‘in what the eye is taking in at the moment.’ Reading ahead, back and forth, works against fluency.

From the Fur grammatical tone examples, we note that for grammatical tone, context might not provide any clues for the reader to choose the correct tone pattern. It is therefore important to distinguish grammatical minimal tone pairs, more so than for lexical tone. If not, this could impede reading for meaning and fluency. Wrong interpretations result in backtracking while reading. Some errors may go unnoticed and a wrong meaning will be perceived. Bird (1999a:23) gives an example from Komo which clearly illustrates the relative importance of marking tone in the lexical and grammatical domains. Due to low functional load (28 pairs among 3000 words), lexical tone is not marked in Komo. But grammatical tone, which distinguishes completive vs. habitual aspect and person/number affixes, could not go unmarked.

---

<sup>36</sup> The phrase ‘left of the processing window’ presumes a language written from left to right.

Simply counting minimal tone pairs in a lexicon is not a reliable predictor of functional load. Powlison (1968:86) does not relate functional load to the number of lexical items that might require disambiguation in isolation, rather, he defines functional load as follows: “the amount of work that phoneme does in keeping *utterances* apart, i.e., in making them different in meaning” (emphasis mine). Several factors may reduce the confusability factor of minimal tone pairs:

1. Only one word in the set may be common, i.e., a word expected to appear frequently in texts; the other may be an uncommon word, i.e., one like a rare animal or specific plant which would rarely be found in text.
2. The words might be in different grammatical categories and thus be distinguished by different word orders (at the sentence or the phrase level), or by the affixes they bear.
3. One of the words might generally occur as part of a fixed phrase not by itself, thus distinguishing it from the other.

It is ill-advised to *not* write tone when there are numerous words which are likely to be confused (i.e., common, same grammatical category, and likely to occur in the same contextual frames).

Zero marking of tone may work if there are only a handful of such words. Schröder (2004:30) writing on Toposa, notes that tone is grammatical in Toposa rather than lexical, and is not marked. He indicates that words or utterances constructed in isolation suggest a high functional load for tone, but that in context ambiguities are rare.

In some situations minimal marking might be preferable to full marking or zero marking. The language committee for Jur Mödö, a tone language of Sudan, decided to not mark tone in order to keep reading and writing simple. However, some grammatical

minimal tone pairs needed disambiguating to avoid confusion in texts, e.g., /nì / ‘her’ and /ní/ ‘their’. The notation option chosen was as follows: <nĩ> ‘her’ and <nnĩ> ‘their’ (Persson 2004:80). This type of disambiguation is a good option, but only when the number of word sets which require disambiguating are few. One must avoid burdening writers with having to memorize spellings of a large number of words.

Marking tone fully for the sake of a few words should be avoided because this would put unnecessary burdens on readers and writers. Bird (1999b:93) reports on several experiments which compare different degrees of tone marking: “All agree that full surface tone marking is not optimal. The high tone density which results from surface tone marking imposes too great a cognitive load on readers, and they are unable to use the information conveyed by the tone marks effectively. We might call this ‘diacritic overload’.”

If tone discrimination is vital for communication, oral and written, one would expect mother tongue speakers to have a highly developed awareness of tone. It has been my experience in writer training workshops that speakers of a language with three tone levels may speak at three levels, but may have difficulties trying to distinguish mid from high tone when writing. Yip (2002:297–298) reports on studies by Harrison on language acquisition of infants.<sup>37</sup> These indicated that by the age of six to eight months, Yoruba infants had developed the ability to distinguish between high and not high, but not between mid and low tone. (Yoruba has three phonemic tone levels; infants from an English speaking environment were not detecting the pitch differences at all.) An

---

<sup>37</sup> Yip makes reference to Harrison (1998, 1999, 2000).

experiment with adult Yoruba speakers, indicated that they were sensitive to high tone, but also could not readily distinguish mid from low tone. This would be an indication that it would be unwise to write tone exhaustively.

Extensive marking might be called for when tone has a high functional load. This tends to be the case in languages with many monosyllabic words. These are more likely to yield a high proportion of minimal tone pairs than languages in which longer words are the norm. One indicator of high functional load of tone in a language would be highly developed tone awareness and discriminatory skills in mother tongue speakers, such as the ability of speakers to whistle tone patterns of individual words or ease of judging words to have the same or different tone melody. *Extensive* tone marking does not mean *full* marking. Only the tones most pertinent or significant to mother tongue speakers need to be marked. Economy in tone marking is a virtue for a system because if tone notation goes beyond what is needed, beginning readers focus on the diacritics too much and experienced readers tend to disregard the diacritics, missing essential cues. Further, writing becomes unnecessarily complex. Orthography testing by researchers, orthography designers, educators, and literature producers is the key to discovering to what degree tone needs to be marked. Bird (1999b:107) exhorts: “Rigorous testing of a variety of tone marking options should be a core part of tone orthography design.”

When tone does have a high functional load, the question is *what* should be noted: the lexical tone of the word when spoken in isolation (isolation form) or the actual tone of the utterance (shallow marking). When tone is written phonemically, it does not remain constant in a word. Just like segmental phonemes, tones interact with each other in an utterance. This may result in tone raising, lowering, leveling, changing, combining to

render a contour, etc. In addition, floating tones and intonation can also perturb tone. Voorhoeve (1964:128) reports on tone in Saramaccan of Surinam. The tone system seemed very complex. But analysis showed that mid tones (level as well as components of contours) appeared only sentence finally. They were due to intonation, and were not phonemic.

A linguist may perceive language complexity; a speech analyzer and acoustic recordings may reveal what is happening in a language. The problem is that the speaker is not aware of all that is going on; many processes are automatic. Forcing people to write surface tone when changes have occurred of which they are not aware, is like forcing them to write segments phonetically.<sup>38</sup>

Rather than discuss phonological processes here I will report on experimental results and highlight recommendations made for tone notation based on field workers' experiences.

Bird (1999a:7) reports on experiments done in the Tschang language of Cameroon:

Analysis showed that tone marking degrades reading fluency and does not help to resolve tonally ambiguous words. Experienced writers attained an accuracy score of only 83.5% in adding tone marks to a text, while inexperienced writers scored a mere 53%.

...people have not been equipped with an effective procedure for determining the correct tone marks on a written text. Since a dictionary can never list all the possible phrases, there is no end in sight for disagreements about the correct spelling of tone on phrases...

---

<sup>38</sup> The descriptivist who needs to understand issues like tone downstep, spreading, shifting, floating tones, the *obligatory contour principle* which avoids successive high tones, and rules which disfavor contour tones, would do well to read Yip (2002); if interested in African languages, Heine and Nurse (2000); and, if focusing on Bantu languages, Nurse and Philippson (2003).

The result is...accents peppering the page, burdening readers and writers with an unnecessary cognitive and mechanical load.

Yet zero tone marking is not necessarily advocated either. Nida (1964b:27) stated:

Despite the practical and scientific validity for omitting the writing of tone, stress, and length in some languages, it should be said that in too many languages such distinctions are not written when they should be. And as a result readers of such languages have to stumble and guess unnecessarily.

The constancy principle (Venezky 2004:146), discussed in 7.2.3, also relates to tone notation. It is not helpful to readers when representation of the same lexical unit is not constant but changes without a change in meaning. Linguists who are concerned about ‘efficiency’ as well as ‘linguistic representation’ have been promoting ‘constancy’ for years. Nida (1964b:25) called it the “principle of unity of visual impression.” Voorhoeve (1964:131), based on his experience with readers of Saramaccan, agrees with this principle, and repeats himself to make a point:

It can be definitely stated that the basic form of a word is more distinctive than any of the perturbed forms. The language users are always clearly conscious of the basic form of a word, but in most cases do not know what perturbed forms they use in a certain context. Seeing that we here mean by reading not mechanical reading aloud but the identifying of words and word groups, and only thereafter the pronunciation of the words aloud, I conclude that the form which is best identifiable, i.e., the basic form, is the one which should preferably be written.

... Recognisability is even more important in teaching to write... [T]he efficiency of an orthography is to my mind largely dependent on consistency of word-image... Therefore I would advocate that tonal perturbations which follow fixed rules should not be written.

Snider (2001) also strongly advocates for promoting a fixed word image in a writing system: “Mature readers read by sight, and so it is advantageous for an orthography to

maintain fixed word-images... We don't advocate writing phonetic forms for nontonal phenomena, so we shouldn't advocate writing them for tonal phenomena."

Bird (1999a:18) notes that "marking tone as it appears in isolation will be intuitively appealing and easily taught," and when reporting on tone notation experiments states, "this experiment counts as another vote in favour of maintaining fixed word images" (Bird 1999b:107). Gudschinsky (1973:123) favored this approach for Mazatec tone marking. Buck (1973:14) points out that such an approach will favor beginning readers who are still reading word for word. Bird (1999a:18) does have concerns about items that simply don't have an isolation form to appeal to, such as (1) "a floating tone" which is not associated with a specific word, but appears in certain contexts and affects other tones, and (2) grammatical particles, which need be attached to something and will never be pronounced in isolation. Bird (1999a:9, 14) warns: "we should bear in mind that tone languages differ in their use of tone. Orthographic principles that work well in one language might perform poorly in another language... The only way to be certain about issues of functional load is through actual experimentation. Anything else is pure speculation".

Bird felt that there are four acceptable ways to cut down the number of diacritics required to mark tone:<sup>39</sup>

- (i) to represent the most frequent tone by the absence of any symbol;
- (ii) to mark tone where it is maximally predictive of word or phrase level tone melodies;

---

<sup>39</sup> This summary was included in Bird's earlier version of the article available at: <http://cogprints.org/2174/index.html> . It was not included in WLL 2(1) 1999.

- (iii) to write contour tones using two vowels, avoiding ă / â in languages with no vowel length contrast; and
- (iv) to collapse tonal distinctions by writing two or more tones using the same symbol.

The first option is commonly used to reduce the density of diacritics in texts. Thus, in a language with two tone levels only one tone would be noted with diacritics; two would be noted in a three level tone language. But the choice of which tone to leave unmarked may not be arbitrary for certain languages. There seems to be an issue of markedness. Wieseemann (1995:25) reports:

Joseph Mfonyam (1990) has written an impressive dissertation about a controlled experiment...done in Cameroon. Mr. Mfonyam is a native speaker of the language (Bafut) in which he tested four ways of writing tone. The results of his testing contradicted what had been previously reported on the writing of tone. Whereas the previous assumption was that high tone should be written and low left unmarked, Mfonyam found out that the contrary is true for his language. This principle has since been applied to languages as far away as Liberia with good results. Reports are coming in of languages where tone, written on the basis of marking high tone, were difficult if not impossible to teach—a situation rectified by simply switching to marking low tone as the basic tone and making the necessary adjustments.

Writing ‘minimal tone’ is taking tone writing down as close to zero marking as possible. As mentioned before, there are two approaches: marking minimal tone pairs consistently, or marking minimal tone pairs only when the context would allow confusion. Both of these bring some challenges: For the first, one wonders if the writer is cognizant of the complete set of minimal pairs and will remember that a tone mark is called for. And as new words are created or adopted into the language, the set may expand. Consequently it would be unlikely that all writers would be up to date on the tone writing requirements (Snider 2001). For the second option, the writer has a message

clearly in mind and may not be able to imagine all the possible ambiguities a reader might encounter in the text. In any case, when marking tone “minimally” one cannot leave one of the tone levels unmarked. If a confusable word bears the tone which is not normally marked, there would be nothing there to alert the reader to pay special attention to it and make the right choice.

The intention of Bird’s research was to “discover what kind of tone marking for a given language best supports efficient reading, writing and comprehension” (Bird 1999b:84). He was surprised that people do not have more problems with zero tone marking. He reports that, clearly, reading with full, shallow notation in Tschang, in Cameroon, “is slower for marked texts, and the presence of tone marks does not reduce the amount of hesitations and repetitions” (Bird 1999b:107). These were Bird’s (1999b:104–105) conclusions:

We should not introduce contrasts into an orthography without good cause; ...

Accordingly, we default to zero marking, and experimental evidence purporting to favor a different system must first of all show that it is an improvement on zero marking.

...It would still be valuable to begin with zero marking and add tone diacritics just where readers experience problems.

### *7.2.7 Tone marking conventions*

In alphabetic systems, tone is most commonly written with diacritics. We have seen those used in Chinese Pinyin. (See 7.2.6.) The same symbols are used in Africa, but interpreted differently.

A common tone noting convention adopted for many African languages is marking low tone with grave accent, <à>; high tone with acute accent, <á>; and mid tone with the macron, <ā>. Contour tones might be indicated by use of the wedge (caron) for

rising tone, <ă>, and circumflex for falling tone, <â>; or they might be indicated by writing two vowels with differing diacritics indicating the contour, (a sequence of differing level tones). In African languages contour tones are interpreted as a sequence of two vowels with differing tones. Falling tone can be represented with <ââ> for example, if readers do not associate the double vowel with vowel length. If one of the tones is marked by the absence of diacritics, one vowel might have no accent, as in <áa>. This allows the elimination of the wedge (caron) and circumflex as diacritics, if so desired. Increased word length, resulting in longer texts, would be a consideration if contour tones occur frequently.

Tones are noted with diacritics in Vietnamese. One level is noted by the absence of diacritics, four with diacritics above the vowels, and one below. Because diacritics are also used for other purposes in this language, there is stacking of diacritics, thus visual discrimination can be a challenge, especially in small print (Ager 1998–2006h).

The danger of diacritics is that authors frequently forget to write them. Other kinds of symbolization may be seen as a more integral part of a word. For instance, some Roman based alphabets employ ‘silent’ letters to indicate tone. The romanized system for Hmong (or Miao) marks tone by adding a letter to the open CV syllables: *-b* for high, *-s* for low, *-j* for high-falling, *-v* for mid-rising, *-m* for low-falling (creaky), *-g* for mid-low, and *-d* for low-rising. One tone is left unmarked (Sloan n.d.). Iu Mien is written with characters in China, but written with a Latin-based script in Laos. Tone is written with *-v*, *-c*, *-h*, *-x*, *-z* at the end of words, which are mono-syllabic. One tone is unmarked (Finh Saeteurn: n.d.). This is like the Hmong system except that Iu Mien has closed as

well as open syllables to which these tone-indicating letters are added. To the uninitiated this may look like consonant clusters, but there are no consonant clusters in Iu Mien.

In north-east India, the Tibeto-Burman Naga group of languages have contrastive tones. Roman script is used. In one of the Naga languages, high tone is marked by doubling the vowel, while low tone is marked by vowel plus *h* (Baker 1999:125).

Tone notation with superscripted numerals at the end of syllables has been in use for several Mexican languages. Numbers 1 through 5 represent different pitches. A sequence of numbers indicates a contour tone. This notation was used in basic literacy material in Eastern Popoloca published in 1973 (ILV). The word *cat* was spelled <co<sup>2</sup>to<sup>12</sup>nchi<sup>2</sup>>. Linguists who thought that all tone contrasts should be noted saw advantages in using this system. In contrast to accent marks, it allowed for distinction between a large number of tonal melodies. Typewriters with Spanish keyboards provided few diacritics which could be used to mark tone. By using the number system, keys did not need to be modified. The expectation was that using numbers to mark tone would built sustainability into literacy programs: Any nationals who had access to a typewriter could potentially be involved in literature production.

However, writing tone with numbers ranked low on the acceptability scale. The reaction by the local population was often negative for aesthetic reasons. This type of notation was just too different from what they had seen in Spanish. It also gave the impression that the language was difficult to read (Judi Lynn Anderson, personal communication). When negative attitudes were encountered or the numbers seemed distracting rather than helpful and complicated reading instruction unnecessarily, language development workers switched to noting tone with diacritics. The Eastern

Popoloca tone orthography was revised in 1980 and literacy materials were reprinted with diacritics: <co<sup>2</sup>to<sup>12</sup>nchi<sup>2</sup>> ‘cat’ became <cotónchi>; <co<sup>2</sup>cha<sup>3</sup>pi<sup>1</sup>li<sup>1</sup>> ‘butterfly’ became <cochapílí> (ILV 1980). Comaltepec Chinantec and Copala Trique, among others, also changed to using diacritics. Where the number representation failed it was usually due to tone having a low functional load in the language with speakers not sufficiently aware of tone to mark it. Reading was less of a challenge, but writing was problematic. By contrast, the system was successful in languages where tone had a high functional load and words did not appear with multiple tone patterns depending on the environment (Cuicatec, Sochiapan Chinantec, and Tepetotutla Chinantec).<sup>40</sup> An innovative notation replacing the number notation is in use in Ozumacín Chinantec: <CV<sup>!</sup>>, <CV<sup>ʹ</sup>>, <CV<sup>ˆ</sup>>, and <CV<sup>ˊ</sup>>. These spacing modifier letters have been integrated into the Unicode standard (Unicode Consortium 2004).

The Mixtec language academy has approved an unusual combination for writing tone: acute accent is used for high tone and underscore for low tone. In response to a recent request by teachers of Copala Trique, the macron for noting low and lowest tone levels has been replaced by an underline (Barbara Hollenbach, personal communication).

In a few cases, tone has been marked by punctuation marks (SIL 2002b). For instance, tones are indicated as follows in Attie of Côte d’Ivoire:

Table 6. Tone notation in Attie

extra high	high	mid	low	extra low	mid-low falling	low-high rising	high-low falling
"CV	'CV	CV	-CV	=CV	CV-	-CV'	'CV-

<sup>40</sup> Information on tone marking in Mexican languages was graciously provided by resident linguists and literacy workers Barbara Hollenbach, Mary Hopkins, and Judi Lynn Anderson via personal communication.

Hartell (1993) lists four additional Ivorian languages which use this type of system: Dan, Muan, Wobe, and Kroumen Tépo. These have fewer tone melodies. I suspect that learnability, aesthetics, acceptability and motivation might be an issue with such a system.<sup>41</sup>

Disambiguation between tonal minimal pairs can also be accomplished by a method not based on tone. Potential homographs could be spelled in different fashions, similar to the way English disambiguates homophones *to*, *two*, and *too*. To avoid homographs in languages in which tone is phonemic but has a low functional load, one could optionally create an arbitrary (or systematic) spelling difference between them. A special segment or a diacritic may be used—not to denote a specific tone, but rather to signal “*not* the default reading, *the other one*.” This allows triggering a meaning connection for readers without tying it to tone. For instance,<sup>42</sup> tone is not written in Mixtec of Magdalena Peñasco, a Mexican language which has considerable tone changes and a floating tone. For example, /īīn/ ‘one’ and /îîin/ ‘nine’ are written <in> and <iin> respectively. In Jur Mödö, tone has a light functional load. As mentioned in Section 7.2.6, the language committee decided not to mark tone, which did not pose a problem except for two sets of minimal tone pairs which had a grammatical function. Leaving these as homographs resulted in unacceptable ambiguities. Disambiguating them by using accents was not favored since texts were already dense with diacritics indicating vowel quality. Thus, the language

---

<sup>41</sup> I have not been able to confirm if this system is currently in use.

<sup>42</sup> These examples were provided by linguistics consultants serving with SIL International. for Mixtec, Barbara Hollenbach; for Jur Mödö, Andrew Persson; for Chumburung, Keith. Snider.

committee opted for doubling the consonants in the high tone words in each set. This resulted in the following distinctions:

<ëdì> 3*S fem. O poss* vs. <ëddì> 3*PL O poss*  
<nì> 3*S to be* vs. <nñì> 2*PL to be*

In Chumburung of Ghana, a similar discrimination principle was applied. First person and third person markers were identical except for tone. A diacritic was used to distinguish these. This diacritic was not taught as a tone mark but as a ‘sense discrimination’ symbol. Arbitrary symbolization increases the degree of abstraction in the orthography in that spellings might not correspond to sound as closely as they might. However, much is gained in simplicity.

Bird (1999a:21) presents an additional notational option which he describes as “morphemic writing of grammatical tone.” He illustrates it with an orthography formerly used for Etung in Cameroon in which a four-way mood/aspect distinction corresponds to four different tone melodies. These were distinguished by writing a space, hyphen, colon, or apostrophe. The symbols used did not represent sound but were “iconic of grammatical information.” Similarly, in Sabaot of Kenya a grammatical tone difference distinguishing nominative from dative case could not be left unmarked. To disambiguate these in text, a colon was written before the subject.

Non-Roman Scripts must also take tone into consideration. The Pahaw Hmong alphabet notes tone with diacritics above the vowel.<sup>43</sup> (Ratliff 1996:620; Smalley, Vang and Yang 1990:60).

---

<sup>43</sup> Pahaw Hmong is a very unique alphabet: the consonant-vowel order in syllables is reversed. Although the alphabet is read from left to right, CV appears as VC. This somewhat undermines the

In Thai, onset consonants are associated with a default tone. Other tones are written with diacritics. (Van der Kuijp 1996:432).

If tone languages which use an abugida need to note tone, the stroke modification inventory needs to be increased to indicate vowels with different tones. The Punjabi language has three level tones, but according to Gill (1996:395) the Gurmukhi script used to write Punjabi “has no separate symbols for tones, but they generally correspond to certain consonantal signs.” In Burmese, vowel signs “contain inherent tones” (Wheatley 1996:452). Bodo, a Tibeto-Burman language group, uses the Devanagari script. The number 0 is used to indicate low tone, and the *visargah* (which looks like a colon) indicates low tone (Baker 1997:125).

The Yi syllabary denotes tone in two ways: Out of its four tones, three “are treated like other phonological distinctions in the syllabary, while one, the mid level tone, is marked as a diacritic on syllabic signs of mid rising tone. In this manner, a highly accurate notation of Yi speech syllables is achieved” (Coulmas 2003:83).

When experimenting in Cameroon with tone notation in the Dschang language, Bird not only measured people’s reading fluency and errors, but also asked what people felt about tone marking. People did not ask for zero tone marking, but they did advocate a reduction of tone marking. What is needed is language specific and needs to be determined.

### 7.2.8 *Dialects*

Dialects complicate the ideal sound-symbol correspondence. Whose speech should the written standard reflect? Does each dialect need its own orthography? Sociolinguistic

---

alphabetic principle. Readers need to process syllables or words more holistically.

and political factors bear upon whether or not one set of literature could or should serve different dialects or if each should have its own. (See 8.2.3.) In addition, there are linguistic factors which complicate the grouping of dialects. Differences in the phonology, lexicon and grammar might make it difficult to use one set of literature. Phonological differences can sometimes be overcome through accommodation in the orthography; differences in vocabulary or grammar are not so easily overcome. There are two main challenges: 1) deciding how to *group* speech varieties and 2) designing the orthography so it will meet the needs of more than one variety.

Researchers have used three basic methods to determine how closely related speech varieties are and if grouping them together is feasible: (1) standard word list comparisons; (2) historical reconstructions of the speech varieties (3) Recorded Text Testing (RTT).<sup>44</sup> These methods attempt to predict intelligibility. Results do not necessarily predict if one orthography (and set of literature) will meet the need of the various speech varieties, or if several are needed. Lexical similarity of less than 70% usually indicates that one is dealing with two separate languages (Kindell 1989). This calls for independent orthography decisions, with harmonization between them being one of the considerations. Casad (1989) states that “for those dialects demonstrating medium to high vocabulary similarity levels (60-95%), the typical case is that we still cannot predict what the degree of comprehension between those dialects will be.” Low intercomprehension between speech varieties indicates that it would not be wise to group them together; high

---

<sup>44</sup> RTT is used to determine the level of inter-comprehension (or intelligibility) between dialects. (Casad 1974; Hatfield 1994.). Various factors come into play: differences in vocabulary, grammar, and pronunciation, as well as acquired bilingualism. due to contact. (See Simons 1979.)

intercomprehension scores (90–95%), although providing no guarantee, could indicate that using literature sharing may be worth a try.

In a complex dialect situation, one might take one of the following approaches:

*dialectal*: each dialect will use a different writing system.

*unilectal*: one dialect is chosen for the written standard and all are expected to use it.

*multilectal or multidialectal*: the writing system does not represent one specific dialect, but through careful choices in symbolization is designed to extend to all.

Simons (1977) defines a multidialectal orthography as “one in which the phonologies of many dialects of a language are compared and accounted for in designing the orthography.”

A dialectal approach is costly in terms of the time and effort put into material production and complicated distribution. (In the past smaller print runs drove up the cost of production, but with current technology cost is not as much of an issue.) If ‘separate identity’ and insurmountable linguistic differences are not driving factors, a writing system should be designed that can work across the established group of variants.

When a unilectal approach is taken, one of the dialects might be chosen for the written standard because of its

- central location and resulting widespread regional comprehension
- advantage in terms of number of speakers
- elevated level of prestige.

To privilege one sector of the population by choosing it as the ‘reference dialect’ (applying a unilectal approach) may give the impression of deliberate favoritism and

offend speakers of other variants in the designated group. Much may be gained, however, by fostering agreement about the choice of a reference variant. Ideally the choice would be made with respect for the other varieties in a way that speakers do not feel that their language rights are threatened, and so all are able to learn the system. An additional option would be to take a dialectal approach for materials intended for reading instruction and for inexperienced readers, and to take a unilectal approach for literature aimed at experienced readers. Experienced readers are able to deal with much more abstraction in an orthography.

When taking a unilectal approach, decision makers sometimes resort to using a ‘dialect’ which is spoken by *no one*. Either they create something artificial or they use a local reconstruction to avoid advantaging one dialect over the others. Unfortunately, with such an approach no one is served well. All are disadvantaged since reading and writing require learning a system foreign to all. Motivation for learning such a system is often low.

Simons (1977), desiring to help linguists find solutions which “involve the least effort in learning to use the orthography for the language group as a whole,” promotes taking a multilectal approach and presents principles to apply which minimize dialect differences. Taking a multilectal approach to designing a writing system results in representing a composite of the spoken varieties. Phonological differences are overcome by:

- a) Using certain symbols with variant pronunciations. Users read the text according to their dialect. Thus writing is regulated, but not speech.
- b) Using separate symbols (overdifferentiation) where two phonemes have merged for some of the dialects and not others.

- c) Using only one symbol (underdifferentiation) although a dialect has a phonemic distinctions while others do not. Note: overdifferentiation with writing rules is preferable to underdifferentiation.

It is important to remember that just because two dialects differ in some phonetic aspect, that they do not necessarily differ in their requirements in representation. For instance, if dialect A has phoneme /s/ consistently pronounced [s], and dialect B has [s] before most vowels but [ç] before high and mid front vowels, this is of interest in the descriptive realm, but it presents no orthographic complication. Although [ç] is phonetically different, it is not a separate phoneme and does not need representation. One symbol will cover both dialects. A second example is where there has been a definite sound shift. Dialect A may have [ts] where Dialect B has [tʃ]. The same symbolization can cover both pronunciations.

When written representation becomes very much removed from spoken forms, the written words, out of necessity, become an 'iconic' type of representation. Beyer (1992) reports on the pronunciation of the word for 'hair' by speakers of a variety of Tibetan languages and dialects. Depending on where they are from, they say *skra*, *škya*, *štra*, *štra*, *kya*, *ša*, *ta* and *tša*. There is, however, only one written form for the word. Beyer (1992:18) comments:

If these Tibetans are literate and are asked to write the word they had just spoken, they will all produce the same written form, which we here transcribe as <skra>. And, if they are shown the written form <skra>, they will, again, pronounce the word differently, but they will all recognize the form and agree that – however it is pronounced – it means 'hair.'

According to Beyer, literate Tibetans can read “manuscripts more than a thousand years old.” This is because “the written form has remained unchanged; the words represented by that form has come to be—or has continued to be—different in different dialects.” This uniformity in writing appears to be advantageous; however, it comes with a price. Only a minority will be able to learn to read and write due to the difficulty and the time investment required. Beyer (1992:18) comments:

The advantage of such a uniform orthography is its transcendence of regionalism: all literate Tibetans share a single written language, however different their spoken dialects may be. The disadvantage is the divorce between the written and spoken languages, making literacy an increasingly difficult and elite accomplishment.

To accommodate the various dialects of Daai Chin of Burma, taking a multilectal approach required not writing tone. Tone, although phonemic, was not stable but differed quite a bit from dialect to dialect (Hartmann-So and Thomas 1981). Not writing tone did not pose a problem. Hartmann-So (personal communication) says:

Although I have found a lot of minimal pairs, not marking tone never seemed a problem for the people who attended literacy classes. In recent years I have tried to re-establish the tone marking in many texts, because I am writing a Descriptive Grammar of Daai Chin. That proved a problem, it seems that the function load of tone in Daai Chin is decreasing.

### **7.3 Synopsis**

There are many linguistic factors which need to be considered when designing or reforming a writing system. Smalley’s criteria of “maximum representation” needs to be balanced with simplicity for the sake of readability and ease of writing. Constancy in representation of meaningful units is helpful to readers. It is not desirable to design an

orthography which is inaccessible because only exceptional people can learn and master it. Using a single written form for a dialectally diverse group can become problematic if a large proportion of the words have to be learned by sight. When speech varieties are similar enough, a single written form can be designed for them. However, non-linguistic factors, presented in the next chapter, also need to be taken into consideration.

## CHAPTER 8

### NON-LINGUISTIC FACTORS

Various factors influence orthography design or reform, working either for or against the ‘ideal.’ Because written language represents oral language, linguists have often made the mistake of thinking that linguistic analysis holds the trump card. However, case studies indicate that non-linguistic factors often carry more weight than linguistic ones.

Ward (2002:13), reporting on developing language learning materials for the Nawat language of El Salvador, comments: “The choice of writing system may be quite complex if no writing system exists. It can also be difficult if several systems exist and the issue of which one to select will involve technical, linguistic, cultural and pragmatic factors.”

#### **8.1 Political factors**

Language development work is political. Not all languages are equal; status depends on national policy, on actual practices on the national and local levels, and on attitudes. Some nations have written policies as to which of their languages are to be developed and receive aid. By omission, the others are left to fend for themselves. Two labels are frequently used in language planning which ascribe special status: *official language* and *national language*.

*Official language* is defined as “a language approved by a government of a country as a medium of communication; especially in the conduct of governmental affairs, business, and schooling, as Hindi and English in India” (Harris and Hodges 1995:171).

*National language* is defined as “a language designated as the principal language of a country for cultural, ethnic, or political reasons” (Harris and Hodges 1995:163). This term, however, is not used consistently. Baker (1997:138) explains: “In 1953, ‘national language’ had two distinct definitions: (1) ‘officially declared to be the national language of a state’ and (2) ‘language of a nation’.” During the colonial era the label *national language* usually referred to “‘the indigenous language spoken by the majority of the population’ in contrast to the ‘official language’ of the colonial power” Baker (1997:138). However, nowadays, nations might apply either definition: For instance, in Benin, ‘national language’ simply means ‘a language of Benin’. In neighboring Togo, however, it refers to two privileged languages: Ewe and Kabiye. These are used in the newspaper, in administration, and for business. They also serve in government-sponsored adult literacy programs, along with two others: Tem and Moba. The government is not opposed to the development of additional languages, but in light of limited resources—financial and human—they choose to encourage investing in these four. French is the single ‘official language’ of Togo. There is a *French only* policy in place for the judicial system and formal education (Gadelii 2004).

“Selection among writing systems is usually political and emotional” comments Smalley (1990:162). Governments often have policies as to how languages within their borders should be written. Sometimes these allow flexibility; sometimes they are very rigid. Whatever the case, those who design or modify a writing system must be familiar with these policies.

Eira (1998:177–78) makes some thought-provoking remarks considering orthographies and politics:

The principle of the Political Discourse is that orthographies symbolise the validity or supremacy of the relevant cultural group, and that their usage must therefore reflect the nationalist/culturalist position at hand. Orthography can function to define a national/cultural group by inclusion or exclusion...

The urgency of orthography development for languages/cultures at risk is a conspicuous example of the power given to orthographies to perpetuate and validate, if not expand, a language/culture... On the other side of the coin, the sometimes violent eradication of orthographies and literary works by conquering and colonising peoples equally demonstrates the perceived power of orthography to maintain a culture.

Unification and differentiation, assimilation and repression, language/culture maintenance, and official recognition are all functions of the Political Discourse, and therefore have ramifications for ensuring lasting relevance of an orthography for a given language/culture group.

Orthographies are often fixed by decree and subject to a government agency's stamp of approval. That approval is not likely to be forthcoming if a proposed orthography does not comply with existing guidelines. For example, in 1979 in Cameroon, the *Committee for the Unification and Harmonization of the Alphabets of Cameroon Languages* drew up a unified alphabet. It serves as the standard for all languages in the country (Tadadjeu and Sadembouo 1984) and reflects the larger effort to establish the *African Reference Alphabet* to which various countries subscribe in differing degrees.

Where such policies exist, grassroots level workers may be involved in linguistic analysis and publish their findings, but the script and symbol choices will be predetermined by policy, based on the analysis. This can be helpful in situations where attempts to codify a language have resulted in lack of uniformity. Bakwerirama (2004),

a website dedicated to promoting Bakweri culture, reports that personal preferences resulted in five spellings of a single word. In response, the website staff chose to adopt the *General Alphabet for Cameroon Languages* for writing Bakweri in cyberspace.<sup>45</sup>

Abiding by national policies will contribute to the acceptance of an orthography by governmental authorities and make it easier for them to support educational efforts for the language in question.

## **8.2 Social and sociolinguistic factors**

Eira does not discredit linguistic analysis, which she refers to as the *Scientific Discourse*, but she points out that “whether or not one can propose principles for a linguistically optimal writing system, it does not at all follow that linguistic efficacy is the only or the most significant factor in the creation of orthography, defined as the accepted standard for writer/readers of the language” (1998:176).

Which factors contribute to the acceptance or rejection of a written standard? Besides national level policies, at the local level there are sociolinguistic issues to be addressed: ethnic and religious identity and tradition, for instance. Although less tangible than linguistic structure, they are powerful, and can work against a linguist’s idea of an ideal orthography. Smalley (1964b:51) urges cultural sensitivity: “If the writing system has been made with the really sensitive appreciation of the values of the native speakers of the language, they are not likely to refuse it.” Eira comments, “People are unlikely to risk their lives for a practical concern such as whether or not to underline a retroflex [ɹ].

---

<sup>45</sup> In the Ethnologue (SIL 2005), Bakweri is listed as an alternate name. The main entry name for the language is Mokpwe. This Cameroonian language has about 32,200 speakers.

However, they will on occasion risk a great deal for an orthography which stands for religious or national allegiance” (1998:174).

### *8.2.1 Community involvement in the decision making process*

In essence, orthography is about allegiance, solidarity, and identity. Emotions may be very much involved. The intended users of an orthography need to be favorably disposed toward it in order to subscribe to it. This is far less likely if the language community itself is not involved in the decision-making process. If outsiders, or some governmental agency, create (or reform) an orthography in a vacuum and then present it to the community for adoption, it is likely to be perceived as an imposition. Attempts to raise awareness after the fact will never result in the same excitement and loyalty as community involvement throughout the whole process. Although Eira recognizes that this could turn into a “painstaking, if tedious, process of inclusive community decision-making” (Eira 1998:174), she stresses the unacceptability of excluding the local people from the process: “There is a tendency...to pass off any motivation other than the scientific as ‘superstitious’ or ‘unenlightened.’ However, when people educated in the western mode assume rights of educational authority over an illiterate people, this also could be called unenlightened.”

Some of the recent publications further express the importance of community involvement. Robinson and Gadelii (2003:32) included a section on community ownership and how it may be achieved:

Language is not only a means of communication, but also a symbol of identity... When a language is in the process of being written for the first time, it evokes many different kinds of feelings. It is important that they are expressed and that the community is

vitaly involved in the process of developing a written standard. There are a number of important questions which a community will have to debate.

The authors present a list of nine questions for consideration. These relate to script choice, dialects, the relationship to other languages, spelling, the promotion of the written standard, literature production, literacy instruction, and the formation of a representative body to deal with some of these questions. The authors note:

These questions cannot be dealt with quickly or easily. Some may be settled by a deliberate process of consultation and decision-making, others may evolve slowly with a consensus gradually developing. Language issues may form part of a broader cultural or political agenda, such as indigenous people's rights or government decentralisation... (Robinson and Gadellii 2003)

Malone (2004) treats the 'Language Committee' as the main decision-making body and makes suggestions for equipping it for the task of developing a writing system for the language. She concludes:

Writing systems are not developed quickly. They take time and patience, dialogue and compromise. As soon as preliminary decisions are made, a wise Language Committee encourages as many MT speakers as possible to begin writing as much as possible in their language. It is through using the tentative writing system that people identify its strengths and weaknesses. And it is through participating in decision-making, that people make the writing system their own. (Malone 2004:45)

Smalley's list of criteria for an adequate writing system begins with "Maximum motivation for the learner" (Smalley 1964b:34). We will now examine what may foster such motivation. Basically it boils down to balancing 'desirable affiliations' with a 'desired separate identity'. What a writing system looks like—whether for large issues

like the choice of a script<sup>46</sup> or minor ones like ‘jots and tittles’—must line up with the community’s desired image and goals. Other writing systems with which the community may have had contact may influence the design of the writing system. These may include the official language, languages of wider communication, neighboring languages, and writing systems used in the past for the language at hand. The question is: Does the local community want their writing system to look like any of these or to be different? This may sound simple, but it may actually be quite complicated. The Hmong, for instance, are encouraged to use a Romanized script (the Romanized Popular Alphabet, or RPA) for technical reasons and to facilitate the transfer of reading skills to other languages. However, many cling to the Pahawh Hmong system for religious reasons and because they desire a separate cultural identity (Eira 1998; Smalley 1990). Another example is the case of the Canadian Inuit, who opted for a Cree-derived syllabary rather than a Latin-based alphabet because they wanted their writing system to symbolize their uniqueness (Coulmas 1999:139). If the chosen script is not the desired one, the community is not likely to promote its use. Rice (1995) presents the North Slavey case study, in which the community was involved in orthography discussions. The language committee expressed the desire to use a syllabic script for their language revitalization efforts; ethnic identity was important to them. This option was denied them and the Latin alphabet was to be used. It is not surprising then that there were obstacles to standardization: The “goal was not reached” and “the work of the committee is ongoing” reports Rice (1995:93). Interest in a community’s wishes must be sincere and cooperation between stakeholders authentic, otherwise language development or revitalization efforts might not advance.

---

<sup>46</sup> For a detailed discussion on script choice and motivation, see Unseth (2005).

### 8.2.2 *Identity issues*

Some scripts are closely associated with a specific religion, so religious affiliation may dictate script choice. It may even lead to the creation of something new. Coulmas illustrates the link between religion and the creation and spread of a writing system:

That the Latin alphabet has been adapted to write so many languages is a direct result of the Christianization of Europe... The distribution of the Greek alphabet and its Cyrillic extension corresponds to the realm of the Orthodox Church... Many other branches of Christianity had earlier developed their own alphabets, Coptic in Egypt, Serto in Syria, Nestorian in Iran, among them. The establishment of the Armenian and Georgian churches resulted in the creation of the Armenian and Georgian alphabets. The Arabic alphabet...now serves as many as one hundred languages in largely Islamic countries on three continents... The Chinese script spread...together with Buddhism and Confucianism... [T]he Tibetan script developed as the vehicle for the Lamaist variety of Buddhism. (2003:201-203)

Scripts, then, are symbols of religious identity. This may cause a single language community to be divided orthographically along religious lines. Serbo-Croatian, written in both Latin and Cyrillic is an example of this in Europe. (See 6.9.)

In Africa, Muslim Tigre speakers are not motivated to read literature produced in Ge‘ez script because it is the writing system used in the Ethiopian Orthodox Church and is associated with the Amharic language; they want Tigre to be written in Arabic script. Woldemikael (2003:14) writes: “Tigre-speakers have been suspicious of the attempts to write their language in Ge‘ez script... They view Amharic negatively, as a language of domination that they were forced to learn and speak until Ethiopia was defeated in 1991... [T]he demand for Arabic comes from inside of Eritrean communities... We were told that we were pushing them out of their religion.”

In India there are similar script choice issues: Hindi and Urdu could be considered one language due to a high level of mutual intelligibility. Hindi is written with the Devanagari script and Urdu is written with the Perso-Arabic script for religious (not linguistic) reasons. Hindu Sindhi speakers wanted to write their language in Devanagari, rather than in the Perso-Arabic script as prescribed in an 1853 decree. This matter had to be settled in court, resulting in both scripts being permitted and Sindhi officially becoming bi-scriptal (Coulmas 2003:232).

Socio-linguistic factors can also cause contention. Resentment toward ‘perceived oppressors’ may cause a group to distance themselves from the dominating culture by rejecting its writing system. The Tigre situation is a good illustration. Similarly, in Vietnam in 1910, as a reaction to Chinese cultural domination, the Chinese-based *Chữ’nôm* script was rejected and the Roman-based script *Chữc-ngũ* adopted. (Although *Chữc-ngũ* had originally been designed by a Jesuit missionary in the 1600s and used for Roman Catholic religious materials (Đình-Hoà 1996), the switch did not signify a change in religious loyalties.) In Turkey, on the other hand, the script change from Arabic to Roman script in 1928 was not motivated by the desire to cut off existing ties. The change was partially motivated by a desire for closer affiliation with Europe. Religious affiliation was not an issue, but the desire to establish a secular state may have been a motivating factor. The movement for standardization between African languages is propelled by two motivations: on the one hand, the desire for a separate identity from the languages of the former colonizers, and on the other hand, a sense of solidarity with other African language communities dealing with similar struggles.

A complicating factor is that there may be a generational difference in writing system preferences. Decision makers may not be the ultimate users of the system they are deciding on. Loyalties that are meaningful to an older generation may not be so for the younger generation. For this reason, it is important to have representatives of different ages on the decision-making committee. Robinson and Gadelii (2003), in a chart illustrating the consensus-building process, use the following verbs: “dialogue, negotiate, compromise, recognize, *validate*, raise awareness” (emphasis mine). It is very important to *validate* history, identity, feelings and desires. When community leaders have been heard, they may feel less of a need to take a hard stand on everything and be willing to negotiate on issues they feel less strongly about.

### 8.2.3 *Dialects*

An additional sociolinguistic factor that may affect orthography decisions relates to dialects within a speech community. The question is, can different dialects be served by one writing system and a single set of literature, or will two or more be needed?

In the past, linguists generally established language groups and subgroups on the basis of linguistic surveys: i.e., linguistic data were compared and decisions made on the basis of similarities and differences. A ‘central’ dialect would be chosen to serve as many ‘dialect’ varieties as possible. Now, sociolinguistic questionnaires are often an added component in language assessment since sociolinguistic factors may influence whether one writing system or several are needed. People’s perceptions about ‘oneness’ and ‘different-ness’ do not necessarily correlate with linguistic distance between dialects, but rather with feelings and attitudes. (See Ring 1989.) Are there conflicts between the speakers of the speech varieties? Are there demeaning attitudes toward some? Do people

agree on a variety that could serve as a standard, or do they wish the writing system to accommodate all dialects? To what extent is there contact between the speakers of the various dialects? Extensive contact may cause people to be multi-dialectal, and increase comprehension between groups. All these issues play into orthography decisions.

Case studies presented by Ring (1989), concerning three Ghanaian languages with dialectal diversity, illustrate that socio-linguistic factors carry more weight than linguistic ones. For Konkomba, unity and the desire for solidarity was so strong that, although there were communication difficulties between the dialects, leaders chose to use the same set of materials. For Mole, despite a high degree of intercomprehension between dialects, a unilectal approach did not work because of differing religious affiliations, social stigmatization, and a low level of tolerance of cultural diversity.

#### *8.2.4 Aesthetics*

For ‘maximum motivation,’ potential users of the writing system must like its look. This is, of course, influenced by exposure to other writing systems. Aesthetics relates to appearance (beauty in the eyes of the beholder), perceived difficulty of text, the acceptability of diacritics—their shape and number, the use of unusual symbols or unconventional uses of familiar ones. Phonetic symbols not part of familiar alphabets have often triggered negative reactions. Likewise, marking tone with numbers has not been well-received. Anglophone linguists and people in countries which had been colonized by the British often displayed more reticence to using diacritics than those in countries in which French was the language of education (Berry 1958:745; Baker 1997:101).

### 8.2.5 *A change in direction*

Attitudes and what is considered ‘acceptable’ can change over time. A desire for affiliation and easy transfer to a language of wider communication may have influenced decisions to go in one direction in the past, but this does not preclude the opposite sentiments from arising later. The desire for change might spring from a popular movement within the language community; it might result from a change in official policy; or it might be part of an agenda of activists (either internal or external). One example is the change away from Spanish orthographic conventions for local languages in Central and South America. (See 11.3.2 for a discussion on Quechua of Peru.)

## 8.3 Educational factors

“No writing system is going to compensate for the lack of an adequate literacy program or educational system of some kind,” warned Smalley (1964f:15). Educational infrastructure and motivation for reading and writing in the community are essential for a writing system to work. If these *are* in place, even a system that is far from ideal might be effective. English is a case in point.

About fifty years ago, Berry, one of the few linguists who wrote about orthography design at the time, addressed this question: “On what grounds is it decided that an orthography is scientifically acceptable?” (1958:738). He named linguistic, pedagogical, psychological, and typographical considerations. Under pedagogical, “economy of time and labour in learning to read and write” was the main concern.

Unfortunately, educational (or pedagogical) factors usually carry less weight with decision makers than other factors. Key clients who are to use the orthography, i.e., learners, potential authors, and teachers, usually have little or no political influence

and thus might not be consulted, nor their needs taken into consideration. Cultural and political factors and well-intentioned linguists might work against the pedagogical ideal.

In the lists of criteria for an adequate writing system by Smalley (1964b) and Malone (2004) two criteria relate to education:

- (1) learnability, and
- (2) ease of transfer to other languages.

Learnability (or teachability, depending on the perspective) relates to the inherent ease or difficulty of a system. It also relates to the human factor, i.e., the physiological and psychological processes involved in reading and writing. We will examine these factors and their importance.

### *8.3.1 Transparency of systems and learnability*

King Sejong (1397-1450 CE) believed that universal literacy in Korean was attainable by means of a writing system that fit Korean well. He did not feel that education needed to be painful or that it should take years to learn to read and write. He was pleased that the new system for writing Korean, Han'gul, was easy, and could be mastered by common people, including women and children, instead of being reserved for an elite group of scribes (Kim 2000; Kim-Renaud 2000).

Alphabetic Han'gul is easy to learn because it is logical and predictable. Arrangements of symbols into syllable blocks provides composite visual units which facilitate holistic reading of elements, aiding reading fluency for experienced readers. However, words or syllables need not be learned by rote. Word attack skills based on sound-symbol relationships (phonics) can be taught from the start. Writing does not involve guesswork.

A writing system like Korean is said to be *transparent* or *shallow*. When a writing system departs far from speech, making decoding and spelling more difficult, it is said to be *opaque* or *deep*. There is a cost associated with opaque writing systems. Smalley (1964b:32) illustrates this by comparing the relative ease and difficulty of reading acquisition in Chinese, Japanese, Vietnamese, English and Spanish. He comments:

The various words in Chinese have to be individually memorized... The cost of time spent in school is enormous... [W]ithin two years or so of normal school study a Vietnamese child can read anything which he sees in print... A Chinese child at the same point can at best pronounce a few hundred words when he sees them in print. This is a result not of the difficulty of the language but of the writing system. Children in both cases speak their mother tongue equally well.

...The English-speaking youngster learning to read his own language requires several grades of drilling in school before he can really be said to read everything he finds in print... English spelling is inconsistent or ambiguous. The Spanish-speaking youngster...in a year or two reaches the point where he can from there read anything which he meets...the Spanish rather consistently represents most of what is spoken.

Results from recent cross-language literacy acquisition studies focusing on European alphabetic writing systems provide empirical evidence that children learning to read in languages with consistent spelling systems require considerable less time and effort to do so. (See Sprenger-Charolles and Béchennec 2004, Goswami 2005.)

All writing systems can be plotted somewhere on the *orthographic depth continuum*. Inconsistencies, under- and overdifferentiation, and certain features not being represented contribute to relative ‘depth’ or ‘opacity.’ For example, an abjad is relatively opaque since it does not mark vowels, but when modified and functioning as an alphabet it can be transparent. This does not mean that alphabets are automatically transparent. English is

rather opaque because of all its inconsistencies in sound-symbol correspondences. These complicate both reading and writing. For example, [a<sup>u</sup>] is written in a variety of ways as illustrated by the underlined portions of the following words: cow, loud, bough. On the other hand, the following sets of words with identical letter groupings (underlined) are variously pronounced: {cow, low}; {loud, Bouquet}; {bough, bought, rough}.

The French orthography is also opaque but less so than English. It is quite predictable for reading, but not so for spelling: there are many silent elements and multiple spellings for identical sounds. Six to eight years of spelling instruction is included in the formal system for English and French, and yet adults are quick to admit that they have not mastered spelling.

Experienced readers are able to cope with an amazing amount of abstraction. People can make sense of text quite accurately even when handwriting is reduced to scribbles (See Taylor and Taylor 1983:194) or if letter order is reversed or scrambled, or words or letters are omitted. It is different for beginning readers; they need all the help they can get. For this reason, for languages with deep orthographies, formal schooling sometimes provides an easier system to be used temporarily for initial reading instruction. For example, vowel pointing systems are used for children learning to read in Arabic and Hebrew. The Pitman Initial Teaching Alphabet (ITA) is an augmented, regularized Roman alphabet for teaching reading in English. It was used in Great Britain, the USA, and Australia in the 1960s. It was not designed to trigger spelling reform but simply as a tool to help children learn as well as to assist those who teach English as a second language (Downing 1964). Although successful in teaching reading, the ITA was not promoted because reading material available using ITA was insufficient, and because

children often had trouble transitioning from ITA to standard English later. In France, a similar system, *Alfonic*, has proven successful in remedial instruction with dyslexic students, and is now in use in experimental classes in lower primary grades as well as in French acquisition programs (Martinet 2005). In Japan, to give children early schooling success, the transparent Katakana and Hiragana scripts are taught first before the Kanji logographic characters. Likewise, in Taiwan, alphabetic *Zhuyin-fuhao* (or *Bopomofo*), a system of Mandarin character-based phonetic symbols, is taught first (Coulmas 1996:577). In China, children are taught alphabetic Roman-based Pinyin to help them learn the sounds of characters (Taylor and Taylor 1983:37).

We conclude that transparent orthographies are easier to learn and less costly for a school system. If an orthography is opaque, extra help is needed for reading acquisition and inexperienced readers, but not necessarily for advanced readers.

The real test of an orthography is not “can it be read?” but “can and will it be written?” Berry (1958:739) wrote: “There is general agreement that phonetic ambiguity is bad and that words pronounced differently should be kept graphically apart. It is also agreed that the orthography should be such that one who knows the pronunciation of a word should be able to spell it correctly.”

If an orthography interferes with learner success, frustration levels are likely to be high, undermining motivation for reading and writing. One of Martinet’s motivations for creating *Alfonic* was to help children overcome the fear of making mistakes when writing. A transparent orthography contributes to confident writing by children and adults alike.

Thonhauser (2003) interviewed individuals in Lebanon concerning their attitudes toward writing. English, French, and Arabic, each of which have deep orthographies, are

taught in school. Unfortunately, because of the traditional focus on rules, correctness and mastery in academia, use of these languages outside of school was limited. Adults felt that writing in a language presumes mastery, and they had not attained the requisite level of skill. This is regrettable. The difficult writing system standards squelched the motivation to write. At the other end of the scale was the colloquial form of Arabic: Although it was not seen as suitable for *real* writing since it did not have an established standard and rules, this is the language individuals use freely in chat rooms and e-mails, using their intuitions for spelling.

It bears repeating here: When adapting writing systems for new languages, inherent qualities of the writing system are to be carried over, but unnecessary burdens eliminated. Simplicity in the system and motivation on the part of the language community are key. About the creation of writing systems for creoles, Baker (1997:120) writes: “There has been widespread agreement among academics that the English-lexicon Creoles require substantially phonemic spelling systems, and that the inadequacies of English spelling should not be inflicted on their speakers.”

So, how far can one stray from purely phonemic representation without impeding reading and writing success? Smalley (1964b:33) writes, “A small number of imperfections in the writing system...are not serious because the context enables the reader to adjust to the difficulties.” However, Smalley also points out that if several features are not represented in the writing system, the cumulative effect can be serious: “the reading task becomes impossible except to the extremely skillful.” For instance, in Africa, if the decision is made to not mark vowel length, nor tone, or Advanced Tongue Root in a Nilo-Saharan or Bantu language, too many ambiguities might result. Smalley

(1964b:41) warns, “Any one of those decisions might not be serious if taken alone, but the combination of all of them means that in total a very high percentage of the distinctive system of the language has been eliminated. Together they carry high functional load. The reader is put at a very great disadvantage.”

Other factors can complicate a writing system: When a system is not adequately adapted, writing conventions will be carried over which are meaningless in the target language. Chamberlain (2004:56) reports that in Dzongkha, written with Tibetan script based on a proto-language form, “when an open syllable follows a closed syllable, they often collapse into a single syllable. The initial syllable loses its coda and the onset of the second syllable becomes the coda of that word. The nucleus of the second syllable is also dropped... Dzongkha readers are taught to eliminate the penultimate written consonant and to read the final consonant as well as to eliminate the final vowel.” Thus, classical Tibetan /kɑr.mɑ/ ‘star’ is written the same in Tibetan and Dzongkha, but in Dzongkha it is pronounced /kɑm/. In English, this would be the equivalent of writing *pinto* and reading *pit* or writing *combo* and reading *cob*. This is an extreme example of a mismatch between written and spoken language. It reduces ‘accessibility’ of the writing system for the average person and thus may perpetuate elitism of a literate minority. It can also result in reading without comprehension. Chamberlain (2004:67, 71) remarks that this is common among readers of Tibetan and Dzongkha. “The main reason for learning to read was religious...the important skill in reading Tibetan scriptures is saying the words. Understanding is not necessary in order to benefit from them spiritually, but rather the forms of the words are sufficient.”

### 8.3.2 *The orthographic depth hypothesis*

About twenty-five years ago, a hypothesis was proposed in the field of educational psychology that the reading process actually differed for shallow and deep orthographies. It posited that when reading occurred with shallow orthographies, such as Serbo-Croatian or Spanish, the route taken to get from print to meaning was via the sound connection as follows: ***print input* → *phonology/sound* → *lexicon* → *meaning***

This sequence was referred to as ‘the assembled route’.

It was supposed that languages with deep orthographies, such as Chinese and Hebrew, would go the ‘direct route,’ (or ‘addressed route’), i.e., there was a direct connection between the print input and the meaning:

***print input* → *lexicon* → *meaning* → (*pronunciation, if needed*).**

Besner and Smith (1992:48) write:

The orthographic depth hypothesis in its strong form makes a very simple claim: There is no orthographic input lexicon in the minds of readers who process orthographies which consist entirely of words with consistent spelling-sound correspondences. The argument is that orthographic access to semantics and the direct mapping from orthographic input lexicon to phonological output lexicon only exists in scripts with inconsistent spelling-sound correspondences, and does so precisely because of this inconsistency.

The orthographic depth hypothesis (ODH) was supported by initial experiments and stirred considerable interest. Through follow-up experiments, however, researchers were able to demonstrate that these results were based on faulty experimental conditions in which readers of shallow orthographies were forced to use the assembled (sound connection) route instead of the direct route because too many pseudowords were

included in the tests.<sup>47</sup> Follow-up experiments showed that even with shallow orthographies, semantic fields are engaged immediately and, surprisingly, that readers of deep orthographies depended more heavily on phonological clues than expected. The notion that reading proceeds either by the ‘assembled route’ or the ‘addressed route’ has given way to the *dual route hypothesis*, also referred to as the *dual code model*. Katz and Frost (1992:81, 82) suggest that phonological processing may be the default approach to reading: “The premise that phonology is the default code is based on the fact that it is typically the code of instruction and the beginning reader receives much practice in its use.” However, this may change, depending on experience and exposure to certain high-frequency words: “[T]he search process based on a visual-orthographic representation may be rapid for the skilled reader once he or she has a well-established visual-orthographic representation in lexical memory.” Everyone would agree that frequent exposure to written words automates their recognition and thus promotes reading fluency. The old adage ‘Practice make perfect’ is valid regardless of the writing system used.<sup>48</sup>

Experience has shown that shortcomings of a writing system can be offset by quality reading and writing instruction accompanied by adequate motivation and opportunities to read and write. It is still important, however, to aim for the best possible orthography and to meet the needs of learners in orthographic considerations. In developing countries the shortage of highly skilled teachers, lack of reading material in local languages, and

---

<sup>47</sup> *Pseudowords* are also often referred to as *nonwords*. They phonologically fit a language’s sound pattern (phonotactics), but are not real words.

<sup>48</sup> A more detailed discussion on reading theory and research is beyond the scope of this paper. See Frost and Katz (1992) and Perfetti, Rieben and Fayol (1997) for ODH research premises, methodologies, results and conclusions.

absence of print in the environment makes it more difficult to counterbalance the shortcomings of an orthography. Spending eight years in school to learn how to spell is a luxury few can afford—and should not be the norm. Some languages are not likely to ever be used in the formal system. This is all the more reason to aim for an efficient orthography which is intuitive for mother-tongue speakers.

### 8.3.3 *Symbol choices, psychology and learnability*

Mattingly (1992) points out that although the visual representation that an utterance can take is practically unlimited as to its outer form, for meaning to be retrieved cognitively, constraints are necessary:

- the symbols need to be “visually discriminable”
- the orthography must correspond to linguistic realities
- the inventory of symbols needs to be limited
- there needs to be constancy in the representation of words.

We will now consider the visual impact of writing systems and how symbol choices can affect learnability. In 1951, a group of specialists made seven recommendations concerning orthographies for vernacular languages in view of their potential use in education (UNESCO 1953:62). Concerning symbol choices, recommendations 4 through 7 expressed preferences for:

4. Letters without diacritics
5. Digraphs in preference to new characters unless they [digraphs] cause ambiguity
6. Derivation of new characters from prevailing scientific usage and
7. Agreement between different languages of the region or country, especially with the national or official language.

The last point encourages harmonization between different languages so as to facilitate transfer of literacy skills between them. (See 8.3.5.)

A different group of specialists met in Niamey in 1984 with the goal of working toward harmonization across African languages. Their recommendations diverged somewhat from the 1951 recommendations. Baker (1997:115)<sup>49</sup> lists those principles for establishing West African orthographies. Three out of the six related to harmonization on the regional, national and international level. Table 7 indicates how the remaining principles compare to the 1951 international recommendations:

Table 7. UNESCO symbol choice recommendations

<i>1951 recommendations for all languages</i>	<i>1978 and 1984 recommendations for Africa</i>	<i>Comment</i>
Digraphs are to be preferred to new characters (unless they cause ambiguity)	Each phoneme shall be represented by a single unique grapheme, i.e., by a specific letter.	differing opinions
New characters are to be based on prevailing scientific usage	Letters should be maximally distinct	different criteria: normative practices vs. ease of visual discrimination
Letters without diacritics should be used	Diacritics should be avoided as far as possible	agreement, but rarely implemented

The principle of avoiding diacritics is often violated. They provide an easy solution to the need for additional symbols. The charge against them is that they are difficult for the eyes to recognize and not ‘maximally distinct’ or easily ‘discriminable’ when reading. Further, diacritics are frequently left off when writing. Despite these liabilities, world languages have used diacritics as well as digraphs effectively. Like the dot over the *i* and

---

<sup>49</sup> Baker cites Mann, Michael, and David Dalby (1987).

the crossbar of the *t*, noting diacritics needs to be automatic. They must be perceived as an integral part of the whole, even if the flow of writing needs to be broken. For writers to remember to note diacritics, these must not be used to symbolize features which are below speakers' level of awareness. German writers do not normally forget to write the *umlaut* when they write. French and Spanish omitted accents on upper case characters because typewriters could not place them above these letters. Now that computers can do so, diacritics are increasingly written on capitals as well. Baker (1997:136) sees no reason to avoid the use of diacritics. He comments: "While accepting that, in languages requiring tone-marking, diacritics should be reserved for that purpose, we can find no important practical reason for not considering their use in other circumstances." What is needed is balance and avoiding what Bird (1999b:93) calls "diacritic overload."

Although it is technically possible to superimpose more than one diacritic over vowels (e.g.,  $\tilde{e}$ ), and this is practiced in Vietnamese to accommodate its eleven vowels and six tones, Baker (1997:101, 136) points out that "the general view remains that no segment should bear more than one superimposed diacritic." Aesthetics is not the only reason for this; if information is pertinent, it needs to be marked in a salient manner. Small fonts and certain printing processes which allow slight bleeding of ink may eliminate the distinctiveness of superimposed diacritics.

Whether it is better to use digraphs, or to employ single, unique graphemes is debatable. Linguists might be partial to using unique symbols because the one-phoneme-one-symbol principle strikes them as more scientific, but these may not be preferred for aesthetic or technical reasons. Unique symbols often do not increase discriminability, as we can see from the sets {**p b p**} or {**n ɲ ŋ ɳ**}. Some symbols may not easily lend

themselves to form the necessary ligatures in cursive handwriting (Baker 1997:137). On the other hand, digraphs, if used too liberally, can lengthen texts by 10–30% and thus increase publishing cost (Baker 1997:134). There are additional considerations. English uses the letter *h* in four different digraphs: <ch>, <ph>, <sh>, and <th>. For some reason, orthography designers for other languages have latched on to this letter as a ‘free for all’ to symbolize a variety of features. As a single letter it is used to symbolize /h/, glottal stop, and tone; as part of a digraph it sometimes symbolizes additional consonants, aspiration, implosives, vowel length, or vowel quality (and possibly other features). There is nothing wrong with any one of these, but when the same letter is used for several purposes in a single language, the distinctiveness of each symbolization decreases and it becomes harder on the eyes to discriminate and segment words. This can interfere with fluency. In African languages, digraphs which symbolize prenasalized or coarticulated consonants can have the same effect. We note, for instance, the visual discrimination burden of the Kakɔ alphabet of Cameroon (Hartell 1993:77):

a b ɓ c d ɗ dy e ɛ f g gb gw h i j k kp kw l m mb  
n nd nj ny ŋ ŋg ŋgb ŋgw o ɔ p r s t u v w y

Teachers know that even {h, k}, {d, b, p}, {n, m}, and {v, w} of non-augmented Latin-based alphabets cause difficulties for learners. One need not avoid all digraphs or all potentially confusable symbols, but one should be aware that when orthographies are dense with similar strokes, reading instruction needs to compensate by using a well-thought out teaching order and excellent lesson content, and by providing adequate amounts of practice and review.

One might want to challenge Berry's (1958:746) statement that it is "an advantage not to disturb the traditional phonetic associations of a letter." Venturing further from the norm in digraph combinations may add more variety to the visual stimulus and thus decrease the confusability factor. Instead of eliminating certain letters from the alphabetic inventory, one might try to overcome a bias against less common combinations like <cz> and <sz>, or <kx> and <cx> which may be considered odd by some. Of course, acceptability is of the utmost importance.

Preserving a visual link between symbols based on phonetic similarity is advantageous if the mother tongue speaker makes a conscious connection between those sounds. For instance, consistency in representing phonemic aspiration or vowel length is useful if a relationship exists in speakers' minds. In that case, length as a feature could be represented in a number of ways, but *one* should be chosen and used consistently. i.e., <ah>, <eh>, <uh> etc., and not different symbolization for different long vowels. For instance, it would be bad to represent long vowels in a single language as follows: <ah>, <e: >, <ie> <o>, <uu>. But a relationship between features might not exist in people's minds. For instance, Gudschinsky (1973:129) reports her surprise at discovering that the Mazatec of Mexico made no association between individual oral vowels and their nasal counterparts. For them /a/ and /ã/ were no more related than /o/ and /u/ in English. In such cases, the rule of consistency in symbolization, although desirable from a scientific standpoint, is not as important from a pedagogical standpoint. Nida reports on two different nasal vowel notations being used in Mesquital Otomi of central Mexico (underlining for one, dieresis for the other) and comments: "It may shock our sense of consistency to indicate such nasalization in two different ways, but for the Otomi-speaker

there is no problem for he simply does not relate the nasalized vowels nor the corresponding non-nasalized vowels. For him they are entirely distinct vowels, so that any attempt to be systematic is quite beside the point” (Nida 1963b:28).

The key to facilitating visual discrimination is sufficient variety and shapes which are sufficiently distinctive. IT IS MORE DIFFICULT TO READ TEXT IN UPPER CASE than in lower case. Some may propose that this is simply due to a lack of practice. However, experiments have shown that outer contour is an important cue for readers and is used in combination with first letter recognition and word length to discriminate between words (Taylor and Taylor 1983:185–186). Outer contour is used more for short words (3-4 letters) than for long words. The ascenders and descenders of lower case contribute to word contour variation and provide visual clues aiding discrimination. Thus, in an alphabet, a combination of special symbols, digraphs, and perhaps a few letters with diacritics in an alphabet might for their diversity of shape facilitate reading acquisition and fluency. Téén from Côte d’Ivoire (Hartell 1993:142) seems to have such a variety in its alphabet:

a ä b c d e ε f g gb x h i ι j k kp l ’l m n ny ηm  
o o p ’p r s t u v v w ’w y ’y z

Digraphs can at times complicate the reading process. If the components of a digraph also exist as individual letters which can occur in sequence, the eye may treat them as a digraph at the wrong times. Processing will not result in a meaningful word, and thus provoke a regression to find another orthographic segmentation that yields an acceptable interpretation. Frequent regressions impede reading fluency. Thus it is a general principle that “two or more letters chosen to denote a single phoneme should not co-occur

elsewhere in the language where they represent individual phonemes” (Baker 1997:135). English does not abide by this principle, thus we find words such as *lighthouse*, *uphill*, and *mishap*, which may cause a beginning reader hardship. I remember as a child reading a book and being perplexed about *Nowhere Land*, which my eyes segmented as ‘Nowhere Land.’ Baker (1997:135) cites examples *hothead* and *hogshead* to illustrate the point. However, because such examples are infrequent, the violation of the above-stated principle has negligible consequences for English.

In German, this can be an issue. The digraph <ch> in the diminutive suffix *-chen*, when affixed to any noun ending in <s> could potentially be misread as trigraph <sch> which symbolizes /ʃ/. Some examples of this are:

*Haus* ‘house’ *Häuschen* ‘little house’; *Hase* ‘rabbit’ *Häschen* ‘bunny’.

In addition, the letters <s> and <t> occur in sequence in certain compounds. They should be pronounced [st], but might be misread as the digraph <st>, pronounced [ʃt]:

*e.g.*, *Hals* ‘neck’ *Tuch* ‘cloth’ *Halstuch* ‘neckerchief’.

Reading experience helps overcome such visual challenges. Nevertheless, the principle above is valid and must be taken into account. A high frequency of ambiguous consecutive occurrences of the component consonant letters may indicate that their combination as a digraph is a poor choice. For instance, a sequence should not be allowed to be ambiguous (i.e., should not allow multiple readings); for example, <tango> read as [tan.go] or [ta.ŋo] or [tã.go]; or <lambe> read as either [lam.be] or [la.mbe]. Clearly, this is not a concern for languages which do not have CVC syllable structures.

In sections 7.2 (Linguistic factors) and 7.2.6 (Tone notation) I touched on preserving ‘fixed word images’. Mattingly (1992:18) agrees with this principle, and calls “constancy

in the representation of words” a constraint. He writes: “In Chinese...though word-boundaries are absent, the form of an orthographic word is constant, or at least not subject to contextual variation. It is suggested that this is a minimal constraint that all writing systems must meet, so that words can serve as units of transcription.” If shapes of words are transcribed so as to reflect that they are context-sensitive, “the reader is...forced to process the transcription symbol by symbol, a slow and arduous procedure.”

Discriminability and constancy are not only issues for Roman-based writing systems. The more alike the strokes and overall shapes are, the harder the eye must work to find distinguishing features. Baker comments on reading Tamil: “Tamil script is particularly easy to read, in part because of its lack of ligatures. It might also be argued that its particular mix of angular and rounded characters makes them more easily distinguishable than is the case with most other scripts of the Indian subcontinent” (1997:129). There is another reason why Tamil may be relatively easy on the eye: Tamil does not require differentiating between mirror image symbols. Based on results from experiments involving non-literate adults and subjects literate in Tamil challenged by mirror image discrimination tasks, Pederson urges the avoidance of symbols which are distinctive only in their directionality when developing a writing system.

Half a century ago, Berry lamented: “We need to know much more than we do, for example, about the nature of the reading and writing acts” (1958:747). Since that time, research in the field of Educational Psychology, aided by advances in technology, has resulted in considerable knowledge about the reading process, eye movements and perception during reader-print interaction. The following descriptions, based on Taylor

and Taylor (1983:121–139) tell what happens when a person reads in a Roman-based alphabetic system.

You read, not by sweeping your eyes along a line of print, but by moving your viewpoint in a series of little jumps called *saccades*...Acuity is sharpest in the center of the visual field, the *fovea*, where the receptors are densely packed...As one reads, a target word is brought into the fovea by a saccadic jump. The eyes then fixate on the word... It is mainly during the *fixation* that a reader acquires information on the fixated word... 90% of reading time is spent in fixations...eyes occasionally jump back, or regress, to fixate on words insufficiently perceived earlier. The best fixation point from which to perceive a whole word should be near its center, and this is where a fixation tends to occur... Reading aloud requires longer and more frequent fixations than does silent reading... The more difficult the reading matter, the more fixations there are and the longer each fixation is. The eye moves ballistically to its target position... Visual sensitivity is reduced during the saccade, and the blur due to eye motion does not affect the perception of the world. Landing at the wrong position can trigger corrective eye movements. The mean length of a saccade is about eight character spaces...the saccade is influenced by the length of the word immediately to the right of the fixated word: if the word is longer, the eye tends to jump further than if it's shorter.

The parafovea...has less sharp vision than does the fovea... Information picked up in the parafovea is about gross visual features, such as length and shape of a word and blank spaces between words. Information on word length is picked up farthest into the periphery, at least 13 or 14 character positions from the fixation point.

We note that even the most experienced reader has physical constraints to deal with when interacting with text. The parafovea scouts out where fixations need to occur. Word length, word shape (contour), and space, not just segments, provide useful information to help reading success. Taylor and Taylor discuss individual differences between readers. For instance,

...the number and duration of fixations and regressions steadily decline from Grade 1 to college. The steepest decline occurs between Grades 1 and 2, and only a gentle decline occurs after Grade 5, at which time eye-movement patterns seem to have all but stabilized. ...poor readers show shorter saccades, longer fixations, and more regressions than do good readers... Ambiguous, unexpected, complex, or important information, be it semantic or syntactic, can cause regressions.

Although Taylor and Taylor make no specific connection to orthography, it can clearly contribute to ambiguity and complexity of text and thus cause regressions and interfere with a smooth reading process. When characters are not sufficiently distinctive, when diacritics and modifying strokes are not salient enough, the eye does not receive the clues it needs for smooth reading. In Roman script, diacritics are apparently more salient and visible in and above words than when written below, since readers “generally pay greater attention to the top halves of the letters they scan” (Grimes, Marwiesh, and Bauernschmidt 1964:117). This probably relates to the tops of Roman script letters being more distinctive in features than the bottoms (Nida 1964b:29).

If experienced readers are stumbling through text (as if they were beginning readers), the orthography may be contributing to the problem and, if so, ought to be modified.

Taylor and Taylor (1983:138) also comment on differences in eye movements based on the level of competency in the language being read:

Foreign languages that are not fully mastered require close examination... Chinese students in the United States (with 7-14 years of English) made 10 saccades per line of English print, with many regressions, showing that they found English reading difficult. Skilled Japanese readers...have large saccades in reading their own native text but small saccades (i.e., many fixations) in reading even easy English text.

In addition, certain scripts do not allow for long saccades: the density of information and the similarity of shapes and strokes may provoke longer fixations. Taylor and Taylor (1983:138) remark:

Chinese readers make frequent fixations...10 saccades per line, compared with English readers, who averaged 4 saccades per line... Fixation durations did not distinguish the two groups. Because less useful information can be obtained about a Chinese character than about an English word from a given distance in the periphery, more frequent fixations are required so that most characters can be seen foveally and identified. Moreover, in Chinese text, almost every character is a content morpheme and is important.

#### 8.3.4 *Word breaks*

Lack of word breaks and very long words can present a visual and processing challenge for readers. Agglutinative languages may have exceptionally long words.

Examples:

1. in Shona, a Bantu language of Africa, one may encounter *Ndakasangana naKufa* ‘Kufa and I met one another’ (Fortune 1957:160).
2. in Siksika (Blackfoot) of Canada one may encounter *iyínnakiikoiksi* ‘policemen’ or *iitsstáwaawayákiyiwa* ‘she hit him while he was down’ (Frantz and Russel 1989:120–121).
3. in Machiguenga, an Arawakan language of Peru, a learner in a basic literacy class is required to read this question: *¿Tyanirikatyo korempivagetankitsi?* ‘Who might be (berry) picking?’ (Ministerio de Educación 1989:116).

Chunking (syllable and morpheme perception within the string) and understanding such long strings of letters takes practice. If the orthography does not somehow break up the

long words, pedagogical material and easy post-literacy materials may need to provide extra help. The expansion of space between segments at morpheme boundaries or some other cue, such as hyphens between morphemes, could be used. In time, as experience is gained, readers can be weaned from such help. Nida (1964a:154) reports that for Aymará of South America, in which 8- to 10-syllable-words are common, some division of long words was helpful in the beginning stages of learning to read. However, breaking up words “contrary to their fundamental structure has resulted in great difficulty for readers...divisions of long words did not prove an advantage. Rather, the readers were confused; and as a result such divisions have been abandoned.” Concerning Bantu languages, Nida writes “In Africa there has been a great deal of discussion between those who have contended for the disjunctive or for the conjunctive method of writing Bantu languages...evidence seems to point more and more in the direction of the validity of conjunctive writing.”

We note that morphological parsing is not required for reading in agglutinative languages. With practice, eyes will know where to find roots which provide core information, syllabify words if necessary, and focus on items which are the least predictable from context. Observation and testing will reveal if strategically placed hyphens would be advantageous. Through some experimentation and formal testing, David and Judy Payne found that Ashéninka reading fluency could be improved by introducing hyphens at certain points in long Ashéninka verbs (David Weber, personal communication). Weber (1992) discovered that experienced readers found Quechua texts easier to read in narrow columns despite the frequent hyphenation that this required.

Van Dyken and Kutsch Lojenga (1993:18) indicate that it is important that word breaks correspond to the intuitions of mother-tongue speakers: “spaces are not randomly inserted at the whim of each writer. Rather they are intended to represent the intuitive knowledge of the native speaker. Mismatches between spoken and written words can confuse a beginner in his efforts to read and write.”

Since linguists often do not possess this “intuitive knowledge,” they need guidelines for “determining whether to separate or to join particular particles of speech into one or more words.” Van Dyken and Kutsch Lojenga (1993:6–17) present twelve criteria to help make such decisions:

1. Referential independence: Can the morpheme communicate meaning on its own?
2. Conceptual unity: Can each of the two items communicate on its own, but when combined is there semantic fusion? e.g., pot + hole = *pothole* (not a hole in a pot); high + chair = *highchair* (serves a special function); toad + stool = *toadstool*.  
This is definitely not a hard rule. English tends to split where German combines: *Kugelschreiber* ‘ballpoint pen’, *Bettwäsche* ‘bed linens’.  
Reduplication could go either way: *mamba mamba* or *mambamamba*; partial reduplication *mamamba* would be better written as a unit. In Shona, reduplicated words are written separately unless the reduplication results in an unpredictable semantic shift. e.g., *kare* ‘in the past’ and *kare kare* ‘long ago’ vs. *mano* ‘advice’ and *manomano* ‘deceit’ (Fortune 1972:11; Hannan 1968).
3. Minimal ambiguities: will one of the options avoid pronunciation ambiguities? *black berries* vs. *blackberries* (stress change); *man* vs. *horseman*, *policeman*; or semantic ambiguities? e.g., *green house* (a house painted green) vs. *greenhouse* (a structure for raising flowers and vegetables).

4. Mobility: Can an item be moved around in an utterance?
5. Separability: Can something be inserted between the two items?
6. Substitutability: Can the item be replaced by a pro-form?
7. Pronounceability in isolation: Can the item be spoken alone, in response to a question, or as a correction or as supplementary information?
8. Phonological unity: Is there a phonological process which treats two items as a unit? Or is there a process which might be expected to occur across a morpheme boundary but does not, thus indicating that treating the items as two words is in order? For instance, does vowel harmony or nasalization occur across the morpheme boundary, or is it blocked? Does the tone pattern treat the two parts together as one phonological word?
9. Phonological bridging: Is there a consistent elision rule? The parts should be written separately if not perceived as fused. It is best to represent them in their isolation form if speakers are conscious of what they are. For instance, in normal speech in Daai Chin, complex processes take place when there is compounding, including “vowel shortening, tone change, resyllabification, assimilation, and consonant loss” (Hartman-So and Thomas 1981). However, in very deliberate speech and in singing words take on their basic forms. Writing lexically (writing parts separately instead of joining them) was the better option.
10. Consistency: Whatever decisions are made should be consistent with the language structure. If certain categories of morphemes are written separately and others attached, there should be consistency so writers will be able to follow a simple rule.

11. Redundancy: Could joining or separating morphemes help grasp the meaning? For instance, could homographs be disambiguated by writing one set as a compound, and writing the other set separately?
12. Conflicting Criteria: When word division criteria seem to be pulling in different directions, some need to be given more weight than others. This is where testing and community preferences have the final word.

Mother tongue speakers have intuitions concerning units in their language. When determining word boundaries it may help to study what they write, analyzing their ‘splitting’ vs. ‘joining’ tendencies, and noting consistencies and inconsistencies.

### 8.3.5 *Ease of transfer between minority and majority languages*

We have examined considerations and principles related to reading acquisition and reading efficiency in a single language. We will now look at what facilitates literacy in more than one language, namely (1) harmony between writing systems of languages in a region, and (2) harmony between local languages and language(s) used in government, commerce and education. Our basic premise is that harmony will minimize the effort required to transfer reading and writing skills from one language to another.

Transfer is bidirectional. A person who has learned to read in a minority language—as an adult in a non-formal education program, or as a child in a vernacular education class—will find it relatively easy to read the official language(s), provided the writing systems are sufficiently similar.<sup>50</sup> The determining factor is the degree to which sound-symbol correspondences are shared, especially for the high frequency cases. In the same

---

<sup>50</sup> A certain degree of proficiency in the language, not only similarity in writing systems, is required to achieve comprehension—a necessary component of *true* reading.

way, if the writing systems are similar enough, individuals educated in the formal system will also be able to read and write the local language which they speak, without expending much extra effort.

In the past, ease of transfer to official languages was seen as an ideal. Local language literacy was considered a *stepping stone* to the more prestigious languages rather than an end in itself (UNESCO 1953:62; Smalley 1964b:34). Providing easy transfer or a ‘bridge’ was seen as the right thing to do. It maximized chances of minority peoples integrating into national life, thereby helping them economically, and giving them opportunities to represent themselves before government officials and to claim their rights. Berry (1958:741) wrote: “Agreement with other writing systems in use for neighboring trade or official languages is desirable where possible on pedagogical and economic grounds.” However, later Berry (1977:5) pointed out that ease of transfer may be more closely related to people’s motivation to use a particular script or orthography than to pedagogy and economy. Smalley (1964b:36) provides several examples of script choice being determined by the degree of motivational value it held for a particular linguistic group.

Pike (1947) discussed the conflicting goals of providing a system for teaching monolingual nonliterate individuals and one which would best serve as a bridge from the vernacular to the trade language. He indicated that often it was government officials that made the decisions: “The officials might decide, on the one hand, to utilize alphabets which are best for the monolinguals, or they might decide to utilize alphabets which are as close as possible to that of the national language” (Pike 1947:213).

Smalley (1964b:36) reported how missionaries in Latin America had to yield to pressures and revise their tentative orthographies to accommodate desires to have more

Spanish-like writing systems: “The influence of ‘educated’ bilinguals, the prestige of identification with Spanish culture, and the elements of transfer value have all united to make new literates want to learn a system as close as possible to the prestige language around them.” That was then; but as we shall see, things are different now. Identity and language preservation issues seem to influence orthography designers to opt for writing conventions which differ from the prestige languages. Smalley (1964b:36) urged paying attention to trends:

The principle...is that those of us who have the responsibility of preparing a writing system for any language, or of changing a writing system for any reason, adapt that writing system as much as possible to the cultural trends, to the prestige, education and political goals which are likely to win out. This I feel, is the most important consideration in a practical orthography.

When language structures differ to a great extent, concern with ease of transfer might interfere with designing an efficient orthography. Coulmas (1989:236–237) cites the Karakalpak writing system as an example. (See 7.1.4.)

As mentioned earlier, ease of transfer may come with a price: It can facilitate the transfer of loyalties to the more prestigious languages, contributing to the demise of the local language within a generation or two. Assimilation might not be the express goal; however, it might be the end result. Therefore, in recent years, ‘separate identity’ has often been valued more highly than ‘ease of transfer’.

Resentment toward past colonial masters plays a role in the trend of abandoning the orthographic systems of languages used in education during the colonial era. If such a language is currently used in secondary education, it would be advantageous to aim for balance: eliminating rules and symbols which are not beneficial, but holding on to

common symbolization where the language systems overlap. This will contribute to learning success and prove economical. Baker (1997:109) points out the advantage of using a script already in use in the region:

In 1832 an alphabetical script using the Burmese characters was proposed by US Baptist missionaries for the distantly related languages of the Karen people. The script has been in use ever since. ...the Karen can acquire literacy in both Burmese and their own language without having to master an entirely new writing system.

A related variety of Karen is spoken in Thailand. Missionaries there thought it more important to facilitate learning the national language than to symbolize unity with the Karen in Burma:

[M]issionaries have adopted the Thai script, making limited use of digraphs...in order to cope with a few Karen phonemes not occurring in Thai. The advantage claimed is that such limited opportunities for education as are open to the Karen are through the medium of Thai and thus that this orthography will help them in that respect. (Baker 1997:123)

Another type of harmonization is that of nations adopting common alphabets for all the national or vernacular languages within their borders. This may or may not affect transfer to the language of higher education. Language preservation efforts and the desire for a separate identity can fuel resistance to such a type of harmonization. For example, “all attempts by the government of India for the sake of efficiency to promote augmented Devanagari as a common script for all Indian languages have failed because no speech community is willing to accept a script other than the traditionally used script” (Coulmas 1999:140). The 'Afar speaking community of Djibuti also resisted harmonization. They did not reject a script, but rejected three of the spelling conventions of the Somali language. They opted for <x> instead of <dh> for retroflex *d*; <c> instead of <x>

for the voiceless pharyngeal fricative, and <q> instead of <c> for the voiced pharyngeal fricative (Baker 1997:117).

Even when symbols and conventions differ, research has shown that there are cross-language benefits. Gardner-Chloros (1997:209) reports: “Literacy skills are known to be highly transferable from one language to another; teachers in France have remarked that previous schooling and having learned to read and write before immigrating are decisive for successful learning in the majority language,” i.e., French.

Of course, when desiring to achieve multiple-language literacy, using the same scripts would be best, but children in Morocco are taught three different scripts in school. Asian students, whether they first learned the Japanese mixed script, Chinese characters, Korean Han’gul, or one of the many scripts on the Indian subcontinent, frequently go on to learn languages which use Latin-based systems.

Bernard (1999:27) comments on benefits across languages and across scripts:<sup>51</sup>

Multiple literacy is apparently becoming more common today... research in Morocco shows the ease with which multiple literacy can be achieved... This provides support for the interdependence thesis... learning to read in any language produces skills that are transferable to any other language, thus making it easier for children to become biliterate or multiliterate... This finding is particularly striking since French and Arabic differ radically in lexicon, syntax and script.

---

<sup>51</sup> Cross-language literacy benefits has been the topic of recent research and published articles. The reader may wish to research *Cross-language transfer* and the *Interdependence Theory* (Carlo and Royer 1999; Cummins 1979, 1984, 1989, 1991; Koda 2005; Verhoeven 1999).

These findings can offer some consolation. However, it is important to point out that they relate to situations with long-term compulsory schooling. If time, energy, efficiency and cost-effectiveness are taken into consideration, the criteria of Maximum Ease of Transfer may not be so quickly dismissed, especially if resources for implementing the orthography and promoting it for years to come are going to be limited. If speakers of minority languages are likely to be schooled in the official language of the country rather than their own, it would facilitate life-long learning if they are able to benefit from previously acquired orthographic knowledge and from print in the environment. Adopting the same capitalization and punctuation conventions could play a part in facilitating biliteracy.

## **8.4 Technical factors**

### *8.4.1 The typewriter*

The nature of technical considerations in orthography design have changed over time. Wolff (2000:337) comments on what was a primary concern for about a century:

For practical purposes of typing and printing, in particular in the pre-desktop publishing days of early orthography development for African languages, the ‘dictatorship of the typewriter’ ruled the choice of symbols and symbol combinations, and early corpus planners restricted the orthographic conventions to those symbols which were readily available on their typewriters.

Baker (1997:94) attributes certain symbol recommendations made by specialists in the past (UNESCO 1953) to a concern for “simplicity in typography” and “cost and/or difficulty of equipping printing presses and typewriters with such characters.”<sup>52</sup> Do we no

---

<sup>52</sup> I can empathize with such technical concerns: In 1978, in preparation for our first assignment in

longer need to be concerned with restricting choices of symbols based on typewriters?

Baker comments:

In an age when typewriters are fast being replaced by desktop computers, and in which printing methods are rapidly changing, simplicity in typography is perhaps becoming a less important consideration... By c. 1985, the word processor or desktop computer had largely replaced typewriters in many of the richer countries and computer technology was rapidly ousting 'hot-metal' type for printing presses... These changes herald the end of almost a century during which the limitations of typewriter keyboards have been the major constraint on orthography design. (Baker 1997:94, 132)

Nevertheless, an exchange of old for new technology is not made as rapidly in developing countries. Computer technology might be available in the capital of a country, but what if literature production needs be done on the local level? Should typewriter limitations still influence decisions? Baker (1997:132) does not think so, urging realism:

Most of the typewriters in offices around the Third World are no longer manufactured. Spare parts and ribbons will rapidly become more difficult to obtain and repairs increasingly problematic. Eventually they will have to be replaced and there is unlikely to be any realistic alternative to acquiring computer technology.

Baker (1997:133-134) encourages considering establishing permissible alternatives for special symbols and diacritics during a transition time, and provides three examples of languages which provide spelling alternatives: (1) German permits substituting <ss> for <ß>, and for umlaut letters <ä, ö, ü> one may opt for <ae, oe, ue> instead;

---

Africa, my husband and I purchased a light, portable Adler Tippa typewriter. We made a special trip to London, where a shop modified the machine, adding two keys for ε and Ϸ. These characters never typed at the right level or looked like they quite fit in with the rest. Capital Ε always looked like a backwards numeral 3, and spacing between letters was off.

(2) Senegalese languages permit <ñ> as an alternate for <ŋ>; (3) Krio of Sierra Leone permits <eh> as an alternative for <ε> and <oh> for <ɔ>.

#### 8.4.2 *The Unicode Standard:*

When the computer age was ushered in, it brought with it benefits, but also new challenges.<sup>53</sup> Computer character inventories initially only accommodated a handful of languages. Those working in non-Roman scripts and in minority languages at the time worked out ‘custom encoding’ solutions for specific languages to meet desktop publishing needs. There was a problem: there was no common standard. Constable (2003:1) comments: “there were not adequate industry standards for dealing with multilingual data, and users became accustomed to cobbling together anything that would get their information onto a printed page.”

Recognizing the need for a standard, a group of technicians began working on a universal standard for character encoding in the late 1980s. Initially the focus was on major languages of the world. Robinson and Gadelii (2003) comment that companies “concentrated on the major scripts where mass use ensures a viable market.”

Later, a group of specialists took on the challenges of communication in a multi-lingual world. They formed the *Unicode Consortium*, incorporated in 1991. It embraced an ambitious plan: to provide a framework for establishing standards to meet the needs of *all* writing systems. Good progress has been made.

---

<sup>53</sup> This discussion on Unicode is based on a Powerpoint presentation prepared by Joel Lee (2005), director of the Non-Roman Script Initiative of SIL International, for an audience with limited knowledge about computers.

To have a universal standard, all characters of each language (including each letter and variant shapes of letters, numerals, commonly used symbols such as @, #, \$, £, %, \*, =, +, etc., each punctuation mark, space, return/enter etc.) need to be referenced in a uniform way. The standard assigns a name and a code point (a number) to each. Shapes are not encoded with the character and code point; *fonts* provided by computer software take the code points and render them as different *glyphs*. Glyphs are the visual shapes displayed on the screen or printed out on a page: g vs. ġ; m vs. ṃ etc.

Those involved in orthography design and literature development need to be concerned with the following technical issues related to Unicode. Not paying attention to these may result in lost data, unexpected character behavior, and a limited capacity for sharing or archiving documents.

*1. choosing characters which are part of the Unicode standard*

When a text uses characters not part of the Unicode standard, and then is viewed on another machine, the appearance of those characters may be altered, even to the point of substituting some completely different letter or symbol. If the original font is not embedded in the document the visual appearance will not be preserved. This severely limits the sharing of documents.

Using characters *not* part of the standard complicates electronic archiving. Data needs to be archived along with a “key file” specifying which code point represents which custom character. Fonts or other display technology also need to be preserved, but this is unreasonable since these often become obsolete and unusable within 10–15 years.



- choosing the appropriate glyph renderings depending on their location of a letter in a word, such as the positional variants of *sigma* in Greek and the complex variants of Arabic characters
- the capacity to create special ligatures for consecutive characters
- the handling of changing directionality in documents containing text from various languages.

Few fonts are *smart* at this point in time. If they are not, the following may occur:

- (1) diacritics might appear at only one height looking messy in combination with ascenders or descenders or upper case characters or interfering with the dot over the <i>
- (2) diacritics might not vary in width and thus appear disproportionate to the letter they modify, or off center, sometimes touching the ascenders or descenders of adjacent letters.
- (3) letters may not have the right shape.

Thus, the following are appropriate guidelines for orthography development:

1. Choose characters already found in the Unicode standard; refrain from designing characters no one else has ever used. (Almost every imaginable option is already part of the standard. The options are practically endless thanks to ‘combining diacritical marks’ which allow not only for typical combinations such as <ë> and <ñ>, but also non-standard combinations such as <ğ>, <ġ> etc.)
2. Use only fonts which are Unicode compatible. An increasing number of such fonts are becoming available.
3. Do not mix characters belonging to different scripts.

4. Choose characters with properties and default behaviors appropriate for their intended use.
5. Do not use ‘deprecated’ characters, i.e., characters which are in the standard only to permit backward compatibility but are no longer recommended.
6. If using characters that require smart rendering, make certain that the necessary smart rendering is available in a wide selection of fonts.<sup>54</sup>

Anderson (2005) advises:

It should be considered a long-term disservice to users to saddle users with an orthography that does not work on today’s computers...

To get the easiest and quickest access to characters on computers, it is best to select from one of the over 96,000 characters already in the Unicode Standard. By using the Unicode characters as they were intended, chances are good that the devised orthography will be supported in a wide variety of off-the-shelf software and displayable with widely available fonts.

By creating an orthography with a completely new character—not in Unicode—the chances that the full orthography will be supported in readily available software is effectively zero. New characters need to go through a years-long process to be standardized, involving meetings with standards committees, followed by delays while the characters are incorporated into software and fonts... There is also a chance that a de-novo character will not be approved for standardization.

Additional information on Unicode can be found on the Internet. (See Appendix A.)

---

<sup>54</sup> Times New Roman, Ariel, Ariel Unicode MS, Lucida Sans Unicode, Verdana, Code 2000, Gentium, Charis SIL and Doulos SIL are just some of the Unicode compatible fonts available for Roman script alphabets. Charis SIL and Doulos SIL are ahead of some of the others when it comes to correct placement of diacritics and size of the inventory of characters. The Non-Roman Script Initiative of SIL International has developed several non-Roman script fonts as well, which are able to handle complex script rendering. A variety of fonts can be downloaded free of charge from <http://scripts.sil.org/FontDownloads> .

## CHAPTER 9

### ADAPTING WRITING SYSTEMS

When adapting a writing system for a new language, one faces two main challenges:

1. Finding solutions for features which differ from the language(s) in which the writing system is already used and
2. Not carrying over unnecessary burdens inherent in the system.

‘Tailoring’ methods are quite similar across scripts. There are several ways to symbolize features not provided for by the writing system being adapted:

- assigning different values to symbols not needed due to phonological differences
- combining letters to form a digraph or trigraph
- slightly modifying the appearance of an existing symbol
- adding a phonetic symbol (Latin script)
- using special ligatures
- using diacritics, underline or punctuation marks
- underdifferentiating features

Each of these options will be addressed except *underdifferentiation*, which was discussed in Chapter 7. Underdifferentiation means that a feature is not represented, or that two or more features are symbolized in the same fashion. This is not usually recommended except under special circumstances as outlined in section 7.2.8.

## 9.1 Roman alphabets

When a writing system is adapted for another language, many symbols will transfer, without discussion or dispute. If in one language speakers articulate /t/ and /d/ at the alveolar ridge, like French, and in another speakers articulate them at the teeth, like Spanish, it is no cause for concern. Orthographies are not designed for phoneticians. If both languages have only one type of *d* and one type of *t* sound, the same symbols can serve in both languages without modifying the symbol. The adaptation is in the speech: speakers pronounce these sounds slightly differently and are usually not conscious of the difference. Tuttle (1996:633), commenting on Latin script being implemented for early Romance language speech varieties in Europe, which he called “Neo-Latin vernaculars,” wrote: “Much of the new wine entered old bottles with little outward difficulty, local phonetic values being smoothly projected onto ancestral graphic correspondents.” Thus, when adapting a system, the same symbols can serve for sounds that are the same or slightly different.

In its oldest known form, the Latin alphabet had 21 letters (Jensen 1970:521). Nowadays, most Roman script alphabets comprise somewhere between 20 (Biali in Benin) and 46 letters (Duruma in, Kenya) (Hartell 1993). The number depends on whether digraphs are considered part of the alphabet or not, and whether unnecessary letters are dropped, or if they are retained for writing loan words. French, with 27 letters in its alphabet, has not dropped <w> from its alphabet because a handful of loan words from English and German contain <w>. Spanish retained <k> and <w> for writing loan words. Vietnamese uses <j>, <w>, and <z> in foreign loans, but these letters are not included in the Vietnamese alphabet. The main shortcoming of the Latin (or Roman) alphabet is that it only provided five vowels. We will look at ways it has been

adapted for languages that are quite different from Latin. There are seven ways in which the Roman alphabet has been adapted. The seventh needs to be used with caution for technical reasons.

### *9.1.1 Assigning different values to symbols*

A ‘left over letter’ can be given a different sound value. For instance, in many Central American languages <x> stands for /ʃ/. In African languages, it often stands for fricative /x/. In some Papua New Guinean languages <x> has been used to represent glottal stop. When <k> is used for /k/, <c> and <q> are ‘extra’ letters, so these may be assigned a different sound value. In fact, <c> commonly stands for /tʃ/, and <q> has often been chosen to write post-velar stop or glottal stop. Letters sometimes denote tone, such as in Hmong (Miao) (Sloan n.d.). Linguists are considering experimenting with using letters to denote tone in some African languages, for example, Kabiye and Sango. (See 7.2.7.)

### *9.1.2 Use of digraphs or trigraphs*

Combining two or three letters for one sound value is common. Fricative /ʃ/ is spelled <sh> in English, <sch> in German, <sj> in Dutch. Using digraphs for vowels expands the vowel inventory. French, for example, uses <ou> for /u/, <au> and <eau> for /o/ and <eu> for /ø/. Nasalized vowels are frequently represented by <n> following the vowel: <an>, <en>, <on>, etc. Prenasalized and co-articulated consonants are common in African languages and are usually symbolized by digraphs or trigraphs: <mb>, <mv>, <nd>, <ny>, <ng>, <ŋg>, <ngb>, etc.

### *9.1.3 Slight modification of a basic letter shape*

Modifying a letter which has a close phonetic relationship to the sound for which a symbol is needed is common. Often, symbols are simply adopted from the International Phonetic Alphabet (IPA). For instance, many African languages, represent /b/, the bilabial implosive, with the equivalent IPA symbol <ɓ>.

### *9.1.4 Use of a non-typical symbol*

Daring to be different, some languages use uncommon symbols. Icelandic, continues to use the historical symbol <þ> to write the interdental fricative; German uses <ß> for /s/. In recently developed orthographies, the representative IPA symbol is frequently adopted for sounds. For instance, <ɛ>, <ɔ>, and <ɣ> are common in African alphabets.

### *9.1.5 Use of ligatures*

The use of ligatures is not a common practice, but is an option. For example, French employs <œ> for /œ/. Likewise, not too long ago, in English, <encyclopedia> was spelled with <æ>: <encyclopædia>. In the Democratic Republic of Congo, in Mbandja and Mono, <œ> is used to represent [ə].

### *9.1.6 Use of diacritics*

The use of diacritics to expand the consonant and vowel inventory is common. German, for example, uses the umlaut; French uses acute and grave accents, circumflex, cedilla, and dieresis. There is great diversity in how diacritics are employed. They may mark vowel quality, length, nasalization, or tone, and sometimes denote an ejective or an implosive.

### *9.1.7 Use of punctuation symbols*

The apostrophe is frequently used for glottalization, glottal stops, or implosives. Colons, on occasion, mark vowel length or consonant gemination. The exclamation mark and some other punctuation symbols are used for clicks in South African Bantu language clicks. Punctuation marks have, on occasion, been used to mark tone or to distinguish homographs.

It is important to use symbols which computer programs recognize as letters, and not only as punctuation. Otherwise, technical glitches will result: spelling checks, word division, alphabetical ordering, and search functions are not likely to produce the expected result because computer programs normally treat punctuation and mathematical symbols quite differently from letters. The Unicode standard includes several pairs of characters that look alike when printed, but one behaves as a letter while the other behaves as a punctuation mark or other symbol. When keyboarding one must take care to choose the appropriate one.

### *9.1.8 Current practices in Africa*

Some nations use a reference alphabet as a standard. This means that the symbol inventory from which letters may be chosen is limited, and that all languages within the nation's borders represent sounds in the same manner. Symbols are rarely used consistently across borders, but, on occasion, governments work on harmonization of orthographies, such as Hausa in Niger and Nigeria, for example. Across languages, a single phoneme, might be symbolized in a variety of ways. For instance, Table 8, based on Hartell 1993, compares how /b/ and /ɸ/ are written in select African languages.

Table 8. Symbolization of two common African sounds, /ɓ/ and /ɣ/

<i>/b/</i>			<i>/ɣ/</i>		
<i>Country</i>	<i>Languages</i>	<i>Symbol used</i>	<i>Country</i>	<i>Language</i>	<i>Symbol used</i>
Burkina Faso, Cameroon, Guinea, Mali, Niger, Senegal	all	ɓ	Benin, Togo	all, consistently	ɣ
Chad	Lele Ngambay	'b ɓ, bb	Cameroon, Côte d'Ivoire	all, consistently	gh
Côte d'Ivoire	all, but one. Yaouré	bh b (for /b/ & /ɓ/)	Chad	Teda	ɣ
DRC	Bhele, Komo Lendu Mangbetu Ngbaka	b (bh for /b/) bb bh 'b	Kenya	Kikuyu, Kuria	g
Liberia	Bandi Bassa, Kpelle	b ɓ	Guinea	all, consistently	ɣ
Nigeria	most Engenni, Ijo Igbo	ɓ ɓ gb	Liberia	Bandi, Kpelle	ɣ
Sudan	all	'b	Mali	Tamasheq	ɣ
Uganda	Lugbara	'b	Niger	Tamajaq	gh
			Nigeria	most Angas	gh h

According to Hartell (1993, v), 32% of the world's languages are in Africa. In 1993, about a third of these had some kind of a writing system, mostly based on the Roman alphabet. Table 9, based on Hartell's *Alphabets of Africa*, presents a random selection of adaptations, indicating the country where it is (or was) in use. The sound-symbol correspondences are not in focus, but rather the variety of options and practices. All of these, and many more, are readily available for standard keyboards. One need not pioneer anything new. If a 'smart' font is used, diacritics will position themselves in the appropriate location and height in relation to the basic symbol.

Table 9. The Roman alphabet expanded: miscellaneous practices in Africa

<i>Adaptation type</i>	<i>Vowels</i>	<i>Comment</i>	<i>Consonants</i>	<i>Comment</i>
reassignment of symbol			c /ts/ h /x/	Chad Cameroon
slight modification of letter shape	ï /i/ ɪ /ɪ/ ø /ʏ/ ɥ /ɥ/ ɥ /i/	Nigeria, Cameroon Cameroon Cameroon Cameroon Cameroon	ḃ /b/ ḋ /d/ Ḍ /Ḍ/ f /f/ ɣ /ɣ/ fi /fi/ ŋ /ŋ/	Burkina Faso Cameroon Benin Togo Guinea Burkina Faso common in Africa
adding an original or IPA symbol	ʌ /ʌ/ ɛ /ɛ/ ə /ə/ ɪ /ɪ/ ɔ /ɔ/ ʊ /ʊ/	Sierra Leone common in Africa Benin, Cameroon, Chad Burkina Faso, Benin common in Africa Burkina Faso, Benin	ɾ /r/ ɣ /ɣ/ ʒ /ʒ/ ʔ /ʔ/	Burkina Faso Benin Benin Burkina Faso
use of special ligatures	æ /æ/ œ /œ/	Cameroon Cameroon		
use of diacritics	ä /ʌ/ ā /a:/ à /ã/ ë /ə/ ẹ /ɪ/ è /ɛ/ ï /i/ ö /ɥ/ ò /ɔ/ ọ /o/ ụ /u/ ü /ɥ/	Côte d'Ivoire Nigeria Senegal Côte d'Ivoire Chad Mali Côte d'Ivoire Côte d'Ivoire Mali Chad Nigeria Mali	ḃ /b/ ḥ /ɣ/ ɪ /ɪ/ ñ /ɲ/ ñ /ŋ/ ŋ /ŋ/ š /ʃ/ ş /ʃ/ ṭ /t/ ẓ /z/	Nigeria Nigeria Niger Burkina Faso Nigeria Nigeria Chad Nigeria Nigeria Nigeria
digraphs	aɔ /ɔ/ ea /æ/ ei /e/ aw /aɥ/ ay /aj/ ɔy /ɔj/	Côte d'Ivoire Côte d'Ivoire Ghana Sierra Leone Sierra Leone Sierra Leone	bb /b/ dz /dʒ/ kh /x/ ny /ɲ/ ph /β/ zh /ʒ/	Chad, CAR Ghana Kenya Benin, CAR Nigeria Cameroon
other	ḁ /ə/ ḃ /ɛ/ ḋ /ɪ/ ḏ /o/ ḑ /ɔ/	Nigeria Nigeria Ghana Ghana Nigeria	ʾb /b/ ʾd /d/ ḍ /ḍ/ h' /h/ t' /tʰ/ ṭ /t/ ' /ʔ/	Chad Chad Sudan Kenya Kenya Sudan Benin, Cameroon

## 9.2 Cyrillic

Like the Roman alphabet, Cyrillic has been adapted for a large number of languages—over 60 according to Eeste Keele Instituut (2006).<sup>55</sup> Based on upper case Greek letters and fitted for Slavonic in the 9<sup>th</sup> century, this alphabet was adopted with slight modifications for Russian. This happened in several stages: In 1708-1710 by Peter the Great; in 1735-1738 by the Academy of Sciences; in 1918, after the Russian Revolution. For Russian, there were three types of changes: the simplifications of letter forms, bringing about a closer resemblance to Roman letters; the removal of redundant letters; the addition of five letters (Coulmas 1996; Cubberley 1996).

From the late 1930s on, it was the USSR's policy to develop and promote only Cyrillic-based orthographies for languages within their borders. A compulsory, systematic changeover from Roman script to Cyrillic was put into effect.<sup>56</sup> This was to facilitate the assimilation of Russian terms into these languages and to help speakers of other languages learn Russian (Baker 1997:121).

Designing orthographies for all these languages presented quite a challenge, especially for Caucasian languages, which have an unusually large inventory of consonants (Baker 1997:108). To accommodate them, linguists resorted to the liberal use of special characters—loans from the Roman and Greek alphabets—, modified Cyrillic letters, and combined some into digraphs and trigraphs, and even a tetragraph for Kabardian. While the Russian alphabet comprises 33 letters, the alphabets of quite a few

---

<sup>55</sup> Eeste Keele Instituut is the Institute of the Estonian Language.

<sup>56</sup> This represented a second script change for many languages since a compulsory, systematic change from Arabic to Roman script had been put into effect during the 1920s.

languages within the former USSR have more than 40 letters: Kazakh has 42, eight of which are used only in Russian loans; Kabardian has 58, two of which are used only in Russian loans (Comrie 1996a). Digraphs are considered part of the alphabet for some of the languages; others do not integrate them.

Regarding alphabetic order, some languages, Kazakh and Kabardian, for instance, integrate the additional characters throughout the alphabet. Others, like Uzbek and Tajik, preserve the Russian alphabetic order and list the special letters at the end.

Regarding phonetic values, assigning a different sound to a symbol used in Russian was avoided (Baker 1997:121). The symbol was thus reserved for loan words from Russian, and ‘unlearning’ sound-symbol correlations did not become an issue when studying Russian as a second language.

Harmonization between languages within the same language family was not a priority within the former USSR. The goal apparently was assimilation and having a Russian identity—not clinging to one’s ethnic identity and affiliations. Baker (1997:121) links the lack of conformity of writing conventions between similar languages to the political agenda of the regime of the former USSR “Conventions...differed from one Turkic language to another, the political aim being to emphasize differences between these peoples and between their languages.”

Comrie (1996a) gives an overview of ways the Cyrillic alphabet has been adapted for Slavic and for non-Slavic languages. In eighteen tables he presents alphabets from five Slavic and thirteen non-Slavic languages. Based on his tables, I will list some of the special characters introduced into Cyrillic in Table 10. These illustrate how an alphabet can be expanded to accommodate a variety of languages.

Table 10. A selection of Cyrillic characters not used in Russian

<i>adaptation</i>	<i>symbol</i>		<i>phonetic value</i>	<i>languages</i>
	upper case	lower case		
addition of symbol	Ә	ә	[æ]	Kazakh, Tatar
	Һ	һ	[h]	Kazakh, Tatar
	І	і	[i]	Kazakh
	Ӣ	ӓ	[ɣ]	Abkhaz
	Ӧ	ӧ	[u]	Abkhaz
	Ӹ	ӹ	[ʦ]	Abkhaz
modification of symbol	Қ	қ	[q] ; [k]	Kazakh, Uzbek, Tajik, Chukchee; Abkhaz
	Ң	ң	[ŋ]	Kazakh, Kyrgyz, Tatar, Chukchee
	Ғ	ғ	[ɣ]	Kazakh, Uzbek, Tajik
	Ө	ө	[ø]	Kazakh, Kirghiz, Tatar
	Ү	ү	[y]	Kazakh, Kirghiz, Tatar
	Ұ	ұ	[u]	Kazakh
	Ҳ	ҳ	[h]	Uzbek, Tajik, Abkhaz
	Җ	җ	[dʒ]	Tatar
	Ҙ	ҙ	[dʒ]; [ʦʰ]	Tajik; Abkhaz
	Ү	ү	[t]	Abkhaz
	К	к	[qʰ]	Abkhaz
	З	з	[dʒ]	Abkhaz
	Ӹ	ӹ	[ʦʰ]	Abkhaz
diacritics	Ў	ў	[o]	Uzbek
	Ӗ	ӗ	[dʒ]	Moldovan
	Ӣ	ӓ	[i] / ' _#	Tajik
	Ӧ	ӧ	[o]	Tajik

Table 11, based on Comrie (1996a:717–718), lists digraphs and other letter combinations from two languages. These illustrate additional options used to adapt Cyrillic for phonologically complex systems.

A word of caution: Precedents have been set, but this does not necessarily mean that a particular choice was good and should be perpetuated and propagated. There are factors which makes some choices superior to others. These relate to desktop publishing technical issues, ease of handwriting, ease of visual discrimination, and ease of learning.

Table 11. Multi-letter graphemes in Abkhaz and Kabardian

<i>language</i>	<i>digraphs, trigraphs, tetragraph</i>
Abkhaz	гъ ъь дə ць жь жə зə кц қь кь тə тə хь хь цə цə шь шə ( ку қу гу ъу кь ху ) <sup>57</sup>
Kabardian	гу гъ гьу дж дə жь ку кІ кІу кь кьу кхь кхьу ль лІ пІ тІ фІ ху хь хь хьу цІ цІ Іу

### 9.3 Arabic

Arabic script is written right to left, except for numerals, which are written left to right. Surprisingly, after a millennium of use, “the orthography of Classical Arabic and that of Modern Standard Arabic are essentially the same” (Bauer 1996:559). How could this be? “The phonology and morphology of Modern Standard Arabic have been taken over from Classical Arabic without change. ...it is no one’s mother tongue.”

Since short vowels are rarely noted, ambiguity arises in texts. Reconstruction of the missing vowels and reading (as opposed to recitation) could only be accomplished by individuals with mastery of the language and an advanced skill level of reading. Bauer (1996:563) comments: “more than a quarter of the phonemes remain unexpressed...one can read an Arabic text correctly only if one knows the words.”

Those of the Islamic faith, in addition to revering their religious texts, also revere the Arabic writing system itself. When adapting it for a language, utmost care must be taken to show respect for its essence, form, and tradition. Because a wide variety of languages are already written in Arabic script, issues of adoption versus adaptation have frequently

---

<sup>57</sup> In Abkhaz, /y/ and /ə/ after consonants both indicate the labialization of that consonant. The digraphs noted with C+ə are considered part of the alphabet (row 1); The digraphs noted with C+y are not (row 2).

arisen. This consonantal system has proven itself quite flexible. Kaye (1006:743-744) writes:

The Arabic writing system has been and is used to write many non-Semitic languages. It is now, after the Roman alphabet, the most used segmental script in the world. ...

In some republics of the former Soviet Union, Arabic script is now once more competing with Cyrillic for writing Turkic and Iranian languages. ...

Arabic script replaced local scripts wherever it reached—notably in Iran, and among the Islamic peoples of South and Southeast Asia. Unlike the Copts of Egypt, however, these people have not given up their languages in favor of Arabic; they have absorbed many loanwords.

In general, Arabic loans retain their original spelling. Thus, when adapted for another language, letters from the Arabic signary are not usually eliminated. Loan word pronunciations do vary from language to language. Classic Arabic has some fairly unique consonants, which do not carry over into local Arabic dialects. Thus, the symbols for those sounds automatically take different phonetic values.

Many Arabic letters have four shapes: a presentation (stand alone) form plus three variant shapes that depend on where they occur in a word: final, initial, or medial position. Others have two or three variant forms. Most letters combine with ligatures, both in handwriting as well as in print; only six do not. (See Table 12.) The variation of shapes complicates learning.<sup>58</sup>

---

<sup>58</sup> Examples in Table 12 were provided by David Cross.

Few people consider that most Roman script letters also have four shapes: Block letters and cursive, each in upper and lower case. This also presents a challenge for learners.

Table 12. Arabic variant letter forms of two consonants

<i>context</i>	<i>form</i>	<i>form</i>
isolated	ج [ǧ]	ع [ʕ]
initial	جمل <i>camel</i>	عربية <i>Arabic</i>
medial	شجرة <i>tree</i>	بعد <i>after</i>
final	برنامج <i>program</i>	مع <i>with</i>

In adapting the classic Arabic script for local languages, there are certain issues that need to be addressed (Kaye 1996:745):

1. No provision is made in the traditional orthography for some very common phonemes, such as /p/, /g/, /tʃ/, /v/, /ŋ/, or /ɲ/;
2. Several sounds are represented by more than one letter (each with variant shapes). For instance: /s/, in Persian and Urdu can be written as س, ث, or ص and /z/ as ز, ذ, ض, or ظ;
3. Vowel diacritics have more than one reading, i.e., one diacritic is used for both /i/ and /e/ and another for both /o/ and /u/.
4. No provision is made for tone marking. This presents a significant challenge for languages in which tone bears a heavy functional load.

Most of these issues have been addressed as Arabic-script orthographies were designed for various languages. Kaye (1996:746) points out that “in addition to borrowing all 28 or 29 Arabic letters, new ones needed to be created, depending on the language, for the different non-Arabic phonemes.” The basic letter forms are not usually altered, but there is much liberty in modifying the system through diacritics. The most common

modification is the addition and subtraction of dots, and their repositioning in relation to the basic shape.

The Persian writing system has had a wide influence on writing systems of other languages in the Islamic world. Some symbols took on a new value. This resulted in multiple graphemes for /ʔ/, /s/ and /z/. Persian provided a way to write /p/ پ , /g/ گ, /tʃ/ چ, and /ʒ/ ژ.

Kurdish is written with Arabic script, but is a full-fledged alphabet. Vowels are written with full letter shapes, not diacritics. Only one letter was retained for /s/ and one for /z/. These still vary in shape according to their position. One consonant symbol was assigned a different value; two symbols were modified to accommodate two additional phonemes.

Kashmiri and Uyghur also use Arabic-based alphabets. Uyghur is unique in that it does not preserve the original spelling of Arabic loan words.

Pashto is spoken in Pakistan, Afghanistan, Iran, and in the United Arab Emirates. To accommodate its sound system, eight consonant symbols were created:

ټ for /t/; څ for /ts/; ځ for /dz/; ډ for /d/; ړ for /r/; ږ for /z/; بن for /ʃ/; and ښ for /n/. Although short vowels are not normally noted, this is an option. In fact, in Pakistan, a differentiation between /i/ and /e/ is being made by varying the direction of the two dots: ټ for /e/ and ټ for /i/.

Several languages in the former USSR were written in Arabic script until the government imposed a ban on its use in the 1920s (Comrie 1996). Kazakh and Kirghiz are two examples. When writing these with Arabic script, the abjad was transformed into an alphabet. Kazakh used 35 letters, and Kyrghyz 32 letters in its alphabet (Ager 1998–

2006b, 1998–2006c). Along with the Kashmiri alphabet of India (34 letters) and the Uyghur alphabet of China (35 letters), these could serve as models for language communities preferring the Arabic script. Considering the large number of languages which use Arabic script, and the geographical area they cover, the degree of consistency across languages in sound/symbol correspondence is surprisingly high.

Some languages in Africa have been written with Arabic script. In Nigeria, Hausa has been written since the 17<sup>th</sup> century in a type of Arabic script called *ajami*. It has 23 consonant symbols. Vowels are noted through a very complex pointing system (Coulmas 1996:196). Since the 19<sup>th</sup> century, a Latin based alphabet, *boko*, has been in use and has almost replaced *ajami*, which is now used only for religious texts and poetry.

Because Standard Arabic is the official language of Sudan, the government has encouraged writing local languages in Arabic script. Anuak, a cross-border language, is written in Fidel script in Ethiopia, and in Arabic script in Sudan. Baker (1997:123) comments: “Anuak has no fewer than ten vowel phonemes so cannot be adequately represented in either of these scripts without considerable modifications. It also has five consonant phonemes which have no equivalents in Arabic and which have thus required the invention of new characters.” Smalley (1964e:100) described how Arabic script was fitted for Anuak: the three Arabic long vowel symbols are used; four vowel symbols were added, and two vowels were modified with a diagonal stroke to further augment the vowel inventory. In addition, digraphs were introduced for two vowel clusters. Tone is not marked. The representation of the vowels, length, and the clusters compensate for not marking tone and for not assigning a separate symbol for a rare, breathy vowel.

Kiswahili was previously written in Arabic script, but no longer. It seems the writing system had simply been *adopted*, not *adapted*. Kaye writes: “Ottoman Turkish, or Swahili in Arabic characters, had several noticeable deficiencies in the script, perhaps moreso [*sic*] than other languages which have borrowed it. Some may claim that, in fact, these deficiencies were so severe that they led to the demise of the Arabic script for such languages” (1996:745).

These were the exception to the rule. Religious and political motivations may help an Arabic based writing system succeed, despite some shortcomings. However, it is difficult to implement and promote a writing system in a community where literacy rates are very low and reading and writing are not valued. Under such conditions the system’s survival is at risk.

## **9.4 Abugidas**

Abugidas are used mainly in South and Southeast Asia and for a few languages in East Africa. (see 6.6, Alphasyllabaries, ) These writing systems are written from left to right.

### *9.4.1 Asian abugidas*

In an abugida, each consonant-vowel sequence (in the literature often referred to by the Sanskrit term *akṣara*), is written as a unit. The basic consonant symbol has an inherent vowel. Other vowels are written by means of an obligatory diacritic. Salomon (1996:373–374) writes about the common ancestry of the large variety of Asian scripts:

Brāhmī, as developed in India and as exported to other parts of Asia in the first millennium C.E., is the ultimate source not only of all the indigenous scripts of South Asia but also of the major Southeast Asian scripts (Burmese, Thai, Lao, Khmer, etc.), of

Tibetan, and of other Central Asian scripts no longer in use. It thus constitutes one of the most important “parent” scripts of the world, rivaling Aramaic and Arabic in the number and range of its varieties and derivatives.

Here we will not consider the history nor the description of scripts, but rather examine what features of abugidas facilitate or complicate adaptation for additional languages. Devanagari, the best known abugida, is used for Hindi, Nepali, Marathi, Sanskrit, and some local minority languages in North India (Bright 1996a:384,389). Sindhi, previously written only in Perso-Arabic script, is now also written in Devanagari. (See 8.2.2.) This is evidence that there is adequate flexibility in the system to make it stretch to yet unwritten languages.

The advantages of the Devanagari abugida are:

- Transparency in the system: consonants and vowels are represented
- With one of the vowels inherent in the basic symbol, the system is economical
- Additional vowels have two forms: initial independent forms, and a diacritic representation. This allows noting V and CV syllable structures.
- Vowel diphthongs have their own representation, as do long vowels
- A vowel-canceling device is provided for the inherent vowel, permitting the representation of closed syllables (CVC) as well as consonant clusters (CCV).
- Provision is made for a large inventory of consonants, including a full set of nasal consonants.
- In some of the languages using the Devanagari script, additional sounds, not part of the system but found in loan words, are represented with minor modification to an existing symbol. For example, Hindi uses a subscript dot to effect a pronunciation change for a basic symbol.

Other Indic scripts have these same strengths, making them suitable for adaptation for additional languages. Mahapatra (1996:404) uses the term ‘adoption’: “Since a large number of tribal languages, of both the Dravidian and Munda families, are spoken within the geopolitical limits of Orissa state, many of these languages have adopted the Oriya script in writing their languages.”

In Indic writing systems, the vowel ‘silencing’ symbol is mainly added to consonants ending a closed syllable. They are not often used as part of consonant clusters. Clusters are often written as *conjuncts*: the shape of one, or both of the components, are modified and combined into one representation. Some are still very recognizable since the shapes are basically the same: one of the symbols is simply reduced in size. The positioning may not be linear. Other conjuncts are less recognizable: one or both of the consonant symbol shapes are only partially preserved in a reduced or fused form. (See Table 13 for examples.) The conjuncts add economy to the system in terms of space. They also facilitate reading if they are recognized and read as a unit instead of analyzed into the component parts.

Table 13. Sample conjuncts from Devanagari

<i>1<sup>st</sup> consonant (silenced)</i>		<i>2<sup>nd</sup> consonant (w. inherent vowel)</i>		<i>conjunct</i>	
प्	[p]	य	[ja]	प्य	[pja]
त्	[t]	क	[ka]	त्क	[tka]
द्	[d]	व	[va]	द्व	[dva]

When adapting Indic writing systems for other languages, there are several considerations. If the target language is phonologically highly similar to that of the

language the writing system has served well in the past, i.e., they have pretty much the same sound inventory, combinations and syllable structures, there is nothing to adapt.

If there are slight pronunciation differences in some of the phonemes, symbolization need not change. For example, depending on the variety of Hindi spoken, some vowel sequences are pronounced as long vowels instead of as diphthongs.

When the language being developed has fewer phonemes, some symbols become unnecessary and can be dropped from the signary. Whereas Sanskrit made a three-way distinction in sibilants [s], [ʃ], and [ç], Indic languages now usually only have a two-way distinction of [s] and [ʃ]. In Nepali all three are merged to the single sound: [s] (Bright 1996a:388-389). Tradition often forces the retention and use of the ‘extra’ akṣaras,<sup>59</sup> resulting in *overrepresentation*: three symbols with a single pronunciation. This does not complicate reading, but makes spelling a challenge.

There is another issue: “When conjunct consonants are involved, the akṣara does not actually correspond to a spoken syllable” (Bright 1996b:417). This might affect learnability. Bright offers an example from Kannada, “*rakta* ‘blood’ which would be syllabified as *rak + ta* in pronunciation, but is written with the two akṣaras corresponding to *ra + kta*. In his article on Devanagari, Bright (1996a:388) uses the following example, *sarva* ‘all’ syllabified as *sar + va* in pronunciation, but it is written with the two akṣaras

स *sa* + वै *rva*.

This type of use of conjuncts could be a disadvantage for beginning readers who sound out words.

---

<sup>59</sup> See Footnote 28 on page 54.

Fluent readers are able to tolerate considerable abstraction in a writing system, particularly if supported by formal education. But, when adapting an existing system for a target language, one might consider not carrying over all the idiosyncrasies, especially those which do not facilitate reading acquisition. Expanding the use of the vowel silencing diacritic to syllable initial clusters instead of using conjuncts may facilitate reading acquisition in the first language. Successful transfer to the more ‘prestigious’ system can be achieved by teaching the complexities as needed during second language acquisition courses.

Usually, for newly written languages, there is a close relationship between speech and writing. This is not so when there has been a long literary tradition. Speech may have changed, but the writing system may not have been modified. This can result in a high degree of abstraction, i.e., over time, a *shallow* orthography may develop into a *deep* orthography, making the acquisition of reading and writing more difficult. When adapting a writing system for another language, it is a reasonable goal to make the system user-friendly and not impose on it burdens from a lingua franca or historical proto-language. For many Indic writing systems there is a mismatch between speech and the written form, disadvantaging learners. For instance, colloquial speech and reading pronunciations are quite distinct in Kannada and Telegu, as well as in Sinhala (Bright 1996b:413; Gair 1996:409). Bengali has some irregular pronunciations due to historical changes. Vowels are noted but not pronounced (Bagchi 1996:399). Bright comments: “correspondences to the spoken sounds of the modern languages follow the lead of Sanskrit in many respects” (1996a:388). These kinds of things are not desirable for a writing system when giving it a fresh start.

Sanskrit is a classical, liturgical language used in Hindu ceremonies. There are some language revival efforts taking place but few people refer to Sanskrit as their mother tongue.<sup>60</sup> It is taught in schools as a second language, and thus has a ‘reading pronunciation.’ Bright (1996a:388) points out that in contrast to Sanskrit, in Indic languages nowadays “short *a* is normally not pronounced at the end of a word or intervocally in the environment VC\_\_VC. Thus... Sanskrit *dāsa* ‘servant’, is pronounced [da:s] in the modern languages; ...Sanskrit *upadeśa* ‘instruction’, is pronounced [upde:s].” If this is a predictable, phonological rule, of which speakers are not aware, it does not matter that a vowel silencing symbol is not used. But what may be predictable in one language may not be so in another. Doing a thorough phonological analysis is imperative for each language for which a writing system is needed. Based on that analysis, it can be discovered which processes happen automatically— beneath the speakers’ level of linguistic consciousness and not needing representation, and which processes need to be noted.

#### 9.4.2 African abugidas

*Ge‘ez* of East Africa, classified as Semitic South Ethiopic North (Gordon 2005) and which Coulmas (2003:154) refers to as the “ancestor language” of Amharic, ceased to exist as a spoken language but still serves for liturgical purposes in the Orthodox Church. This Semitic language and its script, an abugida, are both referred to as *Ethiopic*. The *Ge‘ez* or *Fidel* script (alternate names for *Ethiopic*), is the only Semitic script written from left to right. *Ge‘ez* script differs from Indic abugidas in two ways:

---

<sup>60</sup> The Ethnologue (Gordon 2005) lists 6,106 mother tongue speakers of Sanskrit, based on a 1981 census.

1. There are no symbols for independent vowels: syllables never consist of only *V*.
2. There is no vowel silencing symbol. Instead, one of the modified syllable forms has alternate values: *C+V* or *C* only.

Seven additional basic characters were needed to adapt Ge‘ez for Amharic, the official language of Ethiopia (Coulmas 2003:154). Few additional changes were needed to also adapt it for Tigrinya, spoken in Ethiopia and Eritrea. Comparing Amharic and Tigrinya teaching charts used in schools reveals that the signaries are almost identical but that the order of the characters is greatly changed. Adaptations are of the following types:

- Two basic symbols were eliminated.
- One was assigned a different value (from [h] to [x]).
- A new symbol was created by slightly modifying an existing symbol.

The most important difference between the two systems lies in the value of the basic symbols: the inherent vowel for Amharic is [ɜ] but for Tigrinya it is [e]. The vowel values associated with four of the modified symbols are the same: [u], [i], [a] and [o]. The shapes associated with *C+[e]* in Amharic were assigned to *C+[ie]*. The forms associated with [ə] or ø were reassigned the Tigrinya vowel value [i]. As in Amharic this set is also used to denote *C* without an associated vowel. See Table 14 (based on Coulmas 2003:155 and Ager 1998–2006g) for a synopsis.

Table 14. Comparison of Amharic and Tigrinya vowel values

	ቦ	ቦፋ	ቦፊ	ቦፊ	ቦፊ	ቦፊ	ቦፊ
Amharic	[bɜ]	[bu:]	[bi:]	[ba:]	[be:]	[bə]; [b]	[bo:]
Tigrinya	[be]	[bu:]	[bi:]	[ba:]	[bie]	[bi]; [b]	[bo:]

The Tigrinya abugida demonstrates the flexibility that a writing system with a long tradition can afford, provided that it is not considered ‘sacred.’ Unfortunately, this script’s association with the Christian church in Ethiopia presents a major barrier to its use among language groups following the Islamic faith. Popular literacy is unlikely when the script used provokes negative reactions. Resources dedicated to publications and the language’s development might go to waste.

### **9.5 Orthography testing**

Testing needs to be an integral part of the orthography development process. Unfortunately, orthography decisions are often made top-down, at times in a ‘linguistics laboratory.’ Rules for writing language X are then presented as *de facto*, to be implemented without question, instead of being treated as something to be evaluated and improved. For example, in India, the Central Institute of Indian Languages takes this approach, which Gerbault (1997:178–179) describes:

First, a linguistic description is obtained or made. Secondly, the phonemic inventory is mapped onto the alphabet of the standard script of the language used at the state level; thirdly a trilingual dictionary of about 2,000 words is prepared. Rules for the spelling system are evolved, and primers written. The primer and dictionary are handed over to the State Education Department for use in primary schools. The basic purpose is to provide initial literacy to the child in his own tribal language and to ensure a gradual transfer to the state’s standard language and its writing system.

The use of local languages in the school system to help children toward school success is laudable, but it is worrisome that the process described by Gerbault makes no mention of *testing* the new writing system. Hopefully, provision is made for evaluating, not only the curriculum and pupils’ progress, but also the writing system. Certain types of testing

should actually be done *before* introducing a writing system into the formal system. What needs to be tested? How? Who should be involved? These are important questions now to be addressed.

### *9.5.1 What needs to be tested*

The most important issue is acceptability and motivation. Is this the script the community prefers? Do speakers find text written in the proposed orthography aesthetically pleasing? If culturally appropriate, this kind of information can be gathered through polls and interviews. A language committee, if truly representative and not dominated by a few outspoken individuals, can be the key to getting valuable and honest feedback. Observation is helpful because people's emotional reactions can be revealing. Is early literature eagerly bought up or is it frowned upon? Are there remarks which would indicate that dialect issues were not adequately addressed?

Learnability needs to be tested. There are two kinds of clients: the non-literate population and those who are literate—either in another language or in a previous or competing writing system of the same language. Testing the orthography with literate adults is, of course, easier if they are already familiar with the script. A minimal amount of instruction, explaining how the systems are alike and how they differ, should enable literate individuals to transfer their knowledge to the system being tested. Performance should be evaluated and reading errors and disfluencies with their potential causes noted. Such testing will result in anecdotal and empirical evidence for or against certain aspects of the writing system. Not only readability, but also writing ease should be evaluated. Inconsistencies in writing can provide valuable insights. The writing system should also be tested with non-literates. This can be done by teaching reading and writing in pilot

classes using the system, or through one-on-one instruction. It is important to test in all dialect areas.

Testing is not only about checking for acceptance and verifying that an orthography can be made to work. It is also about deciding between options and gathering evidence that some options are better than others. For instance: Should nasalization be represented with a diacritic or V+n? Would it be better to attach a certain prefix, or to detach it? Does representing /e/ and /ɛ/ with the same symbol make a difference in fluency and comprehension? What gives better reading results: marking only minimal tone pairs or marking all high tones on stressed syllables? etc. Which features need to be tested will be language specific. In general, concerns fall into these categories:

- grapheme choices for segments
- wordbreak issues
- phonemic vs. morphophonemic representation of elements
- under- and overdifferentiation
- representation of suprasegmentals.

Evaluating comparative efficiency cannot be done by informal testing. It requires formal testing, with a reasonable sampling across different demographics. It is labor intensive: someone needs to design the tests, prepare suitable texts and tasks, train test-givers, and interpret results.

### 9.5.2 *How to test*

Informal and formal testing often serve different purposes: Informal testing is often used to discover emotive issues and problematic symbolization, while formal testing is often used to compare efficiency of solutions.<sup>61</sup>

There is much freedom in informal testing since there is usually no pretense that this is a ‘scientific’ approach. Nevertheless it requires careful documentation with a complete description of the circumstances, observations and findings. Kosonen (2003) reports on testing done to test the readability of Chong, which uses Thai script. Under some conditions, informal testing may be adequate as a basis for making orthography decisions, but not if only a small number of people in a restricted geographical area were tested and observed.

Formal tests require tight controls. Questions to be investigated need to be prepared. Although multiple aspects of an orthography may need testing, only one should be tested at a time. A separate set of test materials is needed for each.

For a question such as “Does *not* differentiating between /e/ and /ɛ/ in writing make a difference in reading fluency and comprehension?” one might have a hypothesis concerning which will give better results, but this should not be articulated. Instead, the research question needs to be expressed as a null hypothesis. For example: “There will be no difference in reading fluency and comprehension between texts which write /e/ and /ɛ/ with separate symbols and texts which only write /e/.” A testing method, along with a performance metric (a way to measure the results) needs to then be prepared.

---

<sup>61</sup> Sections 8.5.2 and 8.5.3 are for the most part based on class notes and handouts by Dr. Steve Walter of the Graduate Institute of Linguistics, Dallas.

The following is an acceptable methodology for testing the readability of an orthography or a specific type of symbolization:

The test giver explains the purpose of the test and instructions to each subject, putting him or her at ease. (This is extremely important as individuals participating in formal testing may feel intimidated. It is not easy to convince someone that their performance is not what is being evaluated, but rather the way the language ought to be written.) The person then read four texts, one at a time, while the tester makes notes on a duplicate copy of the text. If possible, a sound recording should be made of each reading. There should actually be eight texts: two versions of each of the four, one using one orthography option, the second using the other. The two versions should be tested on an equal number of people. Texts should be of equal difficulty, and the order of the texts being read should vary from person to person to compensate for differing results due to initial nervousness and ending fatigue. The readings would be evaluated based on predetermined evaluative measures: the time used to read each text, reading errors, hesitations, restarts, etc. If ‘comprehension’ is part of the investigative question, good comprehension questions need to be prepared for each text.

Since this type of testing constitutes human research, permission should be obtained from each subject and the local authorities. The interpretation of the results needs to be carefully developed.<sup>62</sup>

### *9.5.3 Test population*

Those who participate in the test should consist of a random sampling, but with certain considerations. The participant sample should include people from the various

---

<sup>62</sup> A chi-square test may be required to prove statistical significance.

dialect areas, varying religious affiliations, both genders, and varying in age and levels of education. To include such variety, the sample must be large enough. Seventy individuals or more would be ideal. This would allow for excluding some results based on a test subject's inadequate reading competence, anxiety, or other circumstances, without hurting the validity of the test.

#### *9.5.4 Implementation as a process over time*

Orthography testing should be seen as an ongoing process. Based on initial testing results, the appropriate 'informed' decisions—or compromises—should be reached by the interested parties. Only after some testing has shown that the orthography is efficient and accepted should materials be prepared for use in schools. It would be best to do small print runs of material first and try them out in a few pilot classes before expanding to a large-scale project. Teachers need to recognize symptoms caused by orthographic shortcomings, and not be quick to blame the learners. Those learning to read can provide invaluable feedback, provided there is a way for them to give the feedback, and it is appreciated. Teachers are closest to the action, and must be given an audience.

It could be counter-productive to be hasty in elevating a new orthography to the level of a 'standard.' There are handicaps to not having a norm, but as Thonhauser (2003) points out, there can also be drawbacks to excessive focus on the code: it can be demotivating for people who wish to write but dare not do so for fear of making mistakes. Charpentier (1997:233), writing about literacy in pidgin vernaculars, comments:

Far be it from us to deny the handicaps stemming from the absence of a norm, but perhaps norms are in fact less important than we Westerners, conditioned by our very standardized written languages, might assume. After all, our own standardization is only a product of the eighteenth and nineteenth centuries; however effectively Alcuin of

York may have prescribed for Latin, for the vernaculars each scribe and each publishing house had their own conventions in Europe until printing had been established for at least two centuries....

During the last parliamentary elections in Vanuatu... unfettered by any rule, each author had freely used his imagination to render the modern or technical terms that the voters had never seen written but had heard on the radio. All the scripts were equally accessible: might it not be our European norm that becomes a handicap? By constantly prohibiting any deviation in our languages, we are put off when any disparity appears...

Interactive workshops to put an orthography development process in motion have been used successfully in Africa, in Thailand, and in Papua New Guinea (Kutsch Lojenga 1996; Person 1999; Easton 2003.). Person (1999:175) mentions that this kind of approach works toward a “sixth maximum” which Smalley (1964b) does not mention but seems to allude to: “Maximum Participation and Ownership.” (See 11.2.4.) In such a workshop people gather from different areas of the language community, discuss the sounds and symbolization options, and make preliminary decisions. A linguist usually provides some elementary linguistic knowledge about the language and principles of orthography design and facilitates the discovery process and discussions. Then preliminary decisions are made. The importance of native speaker intuition and perception becomes clear; the linguist, is handicapped in certain areas. Person (1999:184) mentions Bisu language workshop participants consistently transcribing a sound difference the linguist could not discern. (Durie (1987) stressed the importance of native speaker intuitions. Thanks to Acehese native speakers’ intuition, it became clear that the nasal/oral vowel opposition was not about nasality, but about *markedness* in various environments.)

During an orthography workshop, the participants practice writing short texts, trying their hand at implementing their own decisions. Alphabet charts or alphabet books and

simple texts are prepared and taken back to the different areas by the participants. There they are to test people's reactions and ease or difficulties encountered when learning the orthography. A follow-up workshop is usually planned to report the findings and take steps to reverse earlier decisions if so indicated. The workshop approach works well for making decisions for a single language, but multiple-language workshops have also been organized (Easton and Wroge 2002). The importance of orthography testing is stressed. A manual used for training Papua New Guinea grassroots level orthography workers instructs:

The alphabet belongs to the language group, and they can make changes if they feel they need to. They do not need someone from the outside to give them permission to make changes...During the workshop, a trial alphabet has been created. It is not a finished product. It is a 'work in progress.' (Easton and Wroge 2002:36)

The manual makes practical suggestions for testing the writing system. It does not presume sophisticated human resources for getting the task done. Nevertheless, some situations may call for a linguist's long-term involvement in the language development process. This likely depends on the complexity of the language and what human resources are available locally. Building capacity in local leaders and educators to plan their own program and make good decisions is likely to result in a better long-term return than an outsider providing solutions.

## **CHAPTER 10**

### **ORTHOGRAPHY REFORM: SPECIAL CONSIDERATIONS**

“Orthographies, particularly those written in an alphabetic script, are rarely fixed once for all time” (Baker 1997:139). “Writing systems can be altered by fiat” comments Daniels (1999:67), and Coulmas states: “It is the defining feature of a codified system that it can be altered at will” (1999:138).

Many languages around the world have implemented some kind of reform this past century. Some of the changes were rather minor, others were quite drastic. The most drastic, of course, is a script replacement. Surprisingly, the proposal of even a small change can meet with great resistance. Fishman wrote, “The greater and grander the tradition of literacy, literature, and liturgy in an orthographic community, the less likely that even minor systematic orthographic change will be freely accepted and the less likely that any orthographic change will be considered minor” (1977:xvi). This can make the change process quite painful. In this chapter, I will focus on motivations for writing-system reform and resistance to such a move.

Introducing and implementing writing-system reform is, by nature, very political. Coulmas (2000:50, 54) pointed out that “Script-reform discussions are invariably politically charged” and “scripts tend to become the focus of political controversy whenever attempts are made to change established norms.” Fishman mentions the necessity of reckoning with the stakeholders: “Orthographic change represents the

abandonment of written tradition, and as such it must cope with the gatekeepers of written tradition, the poets, priests, principals, and professors, with the institutions and symbols that they create and serve” (1977:xvi). If opinions differ and influential parties pull in different directions, an ‘orthography war,’ i.e., a long lasting struggle, can ensue.

The orthography reform process is not very different for languages only recently written. Reform is quite similar to orthography design: political, linguistic, socio-linguistic, educational, and technical factors need to be taken into consideration. Information on a written language is more likely to be readily available than for a yet unwritten language. Negative and positive experiences—whether from years of observation or from a brief testing time—can feed into the revision decisions. Additional knowledge, however, does not guarantee a simpler process. In fact, there is an encumbrance which was not there the first time around, namely, the existing system(s) to which people may have become attached. More people are likely to have an opinion concerning change than about an original creation.

## **10.1 Motivations for reform and revision**

The term ‘reform’ is more commonly applied to changes involving orthographies which have been standardized. ‘Revision’ is more commonly used for minor changes and for orthographies under development and not yet standardized.

### *10.1.1 Political Motivation*

Revolution and evolving nationhood can set the stage for drastic changes, including writing system reform. Implementing a different writing system symbolizes cutting off old affiliations and establishing a new identity. The USSR, Vietnam, and Turkey had such politically motivated writing reforms. Africa’s harmonization efforts and its

abandonment of colonial spelling traditions are partially politically motivated. Although political motivation can kick-start a writing reform, it is not adequate for implementation and long-term sustainability. These call for an encompassing action plan and commitment on the part of the political powers. The USSR government rose to the challenge, not only of providing orthographies for each of the languages within their borders, but also to effect reform and bring about harmonization of writing systems of the USSR languages: In the 1920s and early 1930s a Roman-based alphabet was put in place for all languages which were traditionally written in Arabic. A similar effort was made from 1937 on, to change over from the Roman-based alphabets to Cyrillic script alphabets (See Ager 1998-2006a). Shadrikov and Pakhomov (1999:391) point out that in the USSR, “the reform of Slavic-Cyrillic orthography was more than a purely philological phenomenon. It did away with the centuries-old elitist approach to education.” The Soviet authorities felt that illiteracy stood in the way of a strong, unified nation. They implemented a literacy drive which had dramatic results.

In the 1920s and 1930s, within the framework of the literacy drive, written languages were devised and national schools were opened for forty-eight ethnic groups... Literacy courses and schools were set up at enterprises and in the communities... From 1920 to 1940 elementary reading and writing skills had been acquired by about 60 million people. (Shadrikov and Pakhomov 1999:392)

When a change comes not only with authority, but also with opportunity, it is more likely to stand the test of time—well, at least until a new political wind blows.

With the dissolution of the USSR, the newly independent states gained new freedoms, including that of deciding for themselves which writing system to use for the languages

within their borders. Several have transitioned back to Roman script, including Azerbaijan, Uzbek, and Turkmen. Some are considering a change back to Arabic script.

### *10.1.2 Religious motivation*

A people's identity is usually wrapped up in nationality, ethnicity, and religion. If there is a mismatch between their identity and a writing system that was imposed by fiat, they may challenge the writing system and call for a reform. As already mentioned, Hindu Sindhi speakers are no longer obliged to use the Perso-Arabic script but are now free to use Devanagari (Coulmas 2003:232).

### *10.1.3 Linguistic inadequacy*

Baker (1997:139) gives the following as one of the reasons for making changes to a writing system: "The phonological analysis of the language on which the orthography was originally based may have been faulty in some respects." Such faulty analyses frequently come to light in the early stages of orthography development. Revision is not usually resisted because it is unlikely that the system has been widely disseminated. People are not likely to be attached to it and there is probably not a large body of literature printed with it. One cause for such faulty analysis may be that an outside linguist failed to recognize some of the phonemic differences. This was the cause of non-representation of retroflex *d* in the history of Fongbe in Benin. Likewise, underdifferentiation of +/- ATR (advanced tongue root) vowels or +/- breathy vowels was not uncommon in African languages. Misanalysis can also result in overrepresentation: a distinction heard and symbolized by the linguist may later prove not to be phonemic. This need not be an expatriate linguist; national linguists might also be influenced by their education in a language of wider communication.

Linguistic inadequacy can also come from a deliberate decision to not represent a certain feature. If that feature has a heavy functional load, this will, with time, prove to be a mistake and need to be rectified. A categorical refusal to mark tone in certain tone languages falls into this category.

Coulmas (1999:138) mentions another cause for inadequacy: the mismatch between the writing system and the language it is to serve. “The abstract underlying principles of the writing system were not fully understood when it was first adopted for a given language and that, therefore, the resulting orthography proved to be inadequate.” For instance, the morphemic Chinese characters worked well for non-inflecting languages, but did not serve Korean or Japanese well. Something else was needed to represent their more complex multi-morphemic word structures. Non-Semitic languages are not usually served well by an abjad. Shadrikov and Pakhomov (1999:392), reporting on script reform in the USSR associate the success of literacy efforts with the switch away from abjads:

The consonant-sound writing tradition, whereby letters stood only for consonants or semivowels, made it difficult to record and reproduce the phonetic system of ethnic languages, all of which affected the quality and accessibility of primary education. In 1930 the Turkic- and Iranian-language peoples of the USSR switched over to the vocalized-sound writing, at first on the basis of the Latin alphabet and then on the basis of the adapted Cyrillic alphabet, which gave a fresh impetus to the literacy drive and accelerated the introduction of general education.

#### *10.1.4 Sociolinguistic motivation*

Sociolinguistic conditions are not static. Baker (1999:139) points out that “ideas concerning the particular variety of the language on which the standard should be based may have altered.” Sometimes the speech variety first represented in writing was the one spoken closest to a mission station. The dialect may not be the one in which the majority

of speakers would be motivated to learn to read. Another dialect may better serve the population.

A grassroots movement might initiate a change. The Konkomba of Ghana, who previously opted to use one set of materials across dialect boundaries (Ring 1989), recently indicated that separate representation is now more desirable.<sup>63</sup>

Fine (2003) reports on orthography design efforts for the Bouyei language of China. Motivated to design an orthography which could serve all three dialects of Bouyei and a dialect of Zhuang, a related language, a mixed, multilectal approach was taken in which 40% of the words were written according to a Zhuang dialect, and 60% of the words were written according to a Bouyei dialect. There was insufficient predictability, and the effort was considered a failure, since readers encountered many difficulties. A traditional, unilectal approach was taken in the 1980s, choosing one of the dialects as the reference dialect. The writing system was based on pinyin. This also caused difficulties for one of the dialects. A minor revision was undertaken, adding two graphemes to accommodate the dialect encountering difficulties. Orthography planners were not satisfied with the unilectal approach and wished to work toward a multilectal writing system again. The goal was to choose symbols which would allow each linguistic group to associate sounds with the graphemes based on their own phonology. Such an orthography has been designed and is being tested. It is hoped that the system can extend to as many of the two million speakers of Bouyei as possible, without sacrificing ease of learning and fluency.

---

<sup>63</sup> Personal communication with Ghanaian literacy specialist Judith Bawa.

### 10.1.5 Harmonization efforts

Historically, a government may have had a *laissez-faire* attitude about how languages were written within their borders. Smalley (1964e:76) reports that mission endeavors in India among the Santali resulted in Christian literature being distributed in four different scripts: Roman, Bengali, Oriya, and Devanagari. Now however, more controls are being put into place, usually to unify representations and thus permit transfer of skills from language to language or dialect to dialect.

In Senegal, harmonized orthographies have been developed for its six national languages: Wolof, Sereer, Pulaar, Joola, Malinké, and Soninké (Baker 1997:117). It is noteworthy that although the *Alphabet of Senegalese National Languages* was fixed by decree in 1971, the document anticipated revisions. Indeed, additional resolutions and observations were added in 1975, 1980, and 1985 for specific languages. The 1971 decree stated: “Given the current state of the science of linguistics, the Government...did its best. It will fall on the linguists of the University of Dakar to continue their studies on tone and on the dialects... Then in one or two generations, this decree, if there is a felt need, can be revised in order to be completed” (Hartell 1993:245).

Baker (1997:117) reports that “recent political changes in South Africa have led to a call for new, harmonized orthographies for both the major groups of Bantu languages spoken within the Republic: Nguni (Xhosa, Ndebele, and Swati) and Tswana-Sotho.” Baker points out that although some of these languages are spoken across national borders, international harmonization is not the main concern. Some harmonization efforts do span borders: Brazil chose to follow the orthography reform movement initiated in Portugal in order to preserve a unified written form; Germany refrains from unilateral national orthography reforms so as not to disturb the existing unity between the German-

speaking nations (Augst 1988:1137). Wolff reports on efforts made on behalf of Fulfude, Hausa, and Tamasheq (amongst others). It seems that only the efforts to harmonize Hausa, involving Nigeria and Niger, have been successful. This is of particular interest since Nigeria's official international language of education is English, while that of Niger is French. The two competing orthographies in use for years were harmonized in a meeting in Niamey in 1980. The revised orthography is now serving both nations (Wolff 2000:340; 1991:25).

#### *10.1.6 Educational motivation*

Sometimes writing systems need some house-keeping. Symbols which do not serve a practical function but have been retained for sake of tradition might be eliminated. Or complexity may be reduced as, for example, in China, where the simplification of characters preoccupied the Ministry of Education for many years (DeFrancis 1977).

Coulmas (2000:55) reports on efforts in Japan: "Character limitation after the war was conceived as a measure to facilitate the acquisition of literacy and reduce the importance of character knowledge as an indicator of social status." Taylor and Taylor (1987:57) state that the list of 'official' Kanji in the 1950s numbered 1850, but that the reading population found this number inadequate. Little by little, the list expanded to about 2,000. About another one thousand 'unofficial' kanji are in use, but a mere 1,000 of these account for over 90% of the Kanji in print. These 1000 are taught in the first six years of elementary education (Taylor and Taylor 1987:64).

There are other examples. In 1996, a commission agreed that <ß> was to be excised from the German orthography, since spelling rules as to when to write <ß > and when to write <ss> were rather complex. The proposed change was to write <ss>

consistently, as was already the practice in Switzerland. The intentions were good, but this part of the reform failed. Taylor and Taylor (1987:57) states: “As everybody is aware, linguistic matters are notoriously untamable by official decree.” The German population revolted and thus <ß > remains, although its use is reduced and rules are a little more consistent.

In 1917, in the USSR, some letters were eliminated before extending the Cyrillic script to additional languages.

The alphabet was brought into the fullest possible accordance with phonetics; unpronounced letters as well as letters pronounced in the same way were removed from the alphabet. That made it possible to simplify the spelling rules drawn from literary and spoken languages closer to each other, to standardize writing, and to do away with numerous exceptions from the general rules. (Shadrikov and Pakhomov 1999:391)

### *10.1.7 Language change*

Spoken language is forever young. It adapts to new vocabulary and grammar evolves from the ever-changing use. But what ages the orthography are the phonological changes.

Language is constantly changing, while writing generally obeys tradition and does not readily respond to changes. ...a language’s overall “efficiency” tends to remain constant; but a script’s efficiency—its “goodness of fit” to its language—is maximal when it is devised, and deteriorates thereafter. (Daniels 1999:67)

When an orthography is a ‘good fit’ for a language, changes which occur in spoken language may not call for immediate revision, but when an orthography was decided upon without due consideration of linguistic and educational factors, changes in the language may be more serious, pushing the orthography beyond the threshold of effectiveness. Smalley comments: “A writing system which is inefficient and unrealistic in the first place becomes ever more hopeless through the passing of the years” (1964b:52). That is,

fewer and fewer people will be able to master it. If reform is resisted, the orthography may become accessible only to ‘exceptional’ individuals and will not serve the wider community. “Alphabetic orthographies that have been in use for any length of time tend to develop into highly complex systems with underlying rules too opaque for lay persons to understand” (Coulmas 1999:141). This holds true for non-alphabetic systems as well.

#### *10.1.8 Technical motivation*

Changes in technology over time may make previous writing conventions more difficult to handle. Masica (1996:775) reports:

About twenty years ago, Malayalam, the South Asian language with the highest percentage of literacy, replaced, at least in print, complicated old ligatures of consonant + long and short u, which varied from consonant to consonant and sorely taxed the ingenuity of keyboard designers (although they were no problem for manual typesetters, and even a convenience for handwriting), with a more linear and uniform representation of these vowels, after the consonant and separate from it.

On the other hand, new options, previously inconvenient, now become reasonable alternatives. For instance, in the past, vowels may have been underdifferentiated to keep the inventory down. Now that special symbols are much more accessible thanks to computers, decisions may be made to add symbols to the signary or to change from digraphs to a special unique symbol. The Malayalam decision to abandon certain ligatures in print could possibly be reversed when smart fonts (ones able to form the ligatures) are readily available. (Malayalam is already in the Unicode Standard.)

## **10.2 Types of reform**

There are different types of reforms. Some are quite revolutionary: they involve drastic changes to the majority language of a country or a region, and therefore receive a

lot of publicity. Others go practically unnoticed, making minor changes with little fuss. Some countries evaluate the efficiency of the orthography of their main language(s) on a regular basis and introduce small changes whenever needed. The hope is that there will be no big upheaval. The Dutch, having gone through a painful cross-border reform experience, now update their language reference books about every ten years. Language planning is a united effort, involving the Dutch, the Flemish of Belgium, and the Surinam government (Taaluniversum 2006).

Some changes may not affect segmental representation, but relate to alphabetical ordering, word breaks, punctuation, hyphenation, etc. Examples are Spanish removing <ll> and <ch> from the alphabet in 1994 and treating them as two separate letters (Venezky 2004:150) and French abandoning the hyphen in compound words: *portemonnaie*, not *porte-monnaie* (wallet).

Then there are the periodic revisions implemented early in the history of a newly written language. As new information comes to light, test results come in, people's reactions are noted, and changes are made to the writing system as it moves toward standardization. This is part of the orthography design process and goes practically unnoticed except by those producing literature and those who had been exposed to previous experimental orthographies.

Some revisions are instigated by people with a cause. Changes are introduced not because a system does not work, but for socio-political reasons. Minority groups are awarded their 'separate identity' based on decisions made by concerned outsiders. When 'advocates' use mass media to promote change, there is more publicity involved than would be otherwise. Examples of this kind are the change away from Spanish-looking

orthographies for Central American Indian languages and the change from a five vowel system to a three vowel system for Quechuan languages in South America. (See 11.3.2.)

### **10.3 Resistance to reform**

Change comes at a cost. Unless everyone is agreed that the change is the solution to a problem, there is likely to be resistance. Nida (1964b:30) notes:

It is true that a number of orthographies do need revising, for they are quite inefficient. However, we would be making a serious error if we think that they can be easily changed, or that in all instances changes in the direction of greater efficiency would be accepted. People become very much attached to systems of spelling. The very arbitrary character of any orthography seems to increase its endearment to the people. Sometimes the more awkward it is, the greater is its sentimental value. Any changes in orthography must be studied with the greatest of care, and in all instances one must base any changes upon the already existing system. One can introduce refinements of usage, but not revolutions (unless there is a cultural and/or political revolution to accompany such a change).

Human factors are rather stable across time and across cultures. What Nida expressed holds true on all continents, as much today as forty years ago. Attempts at reforming English spelling have failed. Gerbault (1997:157) comments:

Attempts at simplifying English spelling have been made, but have encountered a lot of resistance and have not brought about the expected results... Reforming alphabets or spelling systems is among the most difficult of cultural changes – the reform in Turkey in the 1920s is a noteworthy exception.

The amount of resistance may be relative to the degree of upheaval likely to result from the change. Whether one is dealing with reform or the original design of an orthography, Berry's statement holds true: "acceptance or rejection of an orthography has

little to do with its linguistic adequacy” (1977:4). I will now discuss some of the causes for resistance to change.

### *10.3.1 Tradition*

Habits do not change easily. Tradition is the ‘strong man’ to be reckoned with. Smalley (1964f:14) comments: “The attachment to a writing system may be such that educational advantages, and particularly advantages for the following generation, will be scuttled in favor of holding on to an orthographic tradition from the past.” The longer the tradition of writing, the more difficult it becomes to introduce change. English is a case in point. Also, the situation for English is more complicated because of its function as the world’s foremost international language.

Baker (1997:105–106) refers to Xhosa (a Bantu language of South Africa), with more than forty phonemes, which has a problematic orthography. Although a major reform would be beneficial, it is not likely to take place.

If it were practicable to design an entirely new orthography for Xhosa today, a more economical, more internally consistent, and perhaps more elegant orthography might be designed... However, Xhosa has been written essentially...for a century and a half and thus any proposed major changes would almost certainly be resisted. The best that might be hoped for is that minor reforms could be introduced gradually.

### *10.3.2 Changes are not in line with desired identity*

As with orthography design, orthography reform must not go contrary to people’s desires. Acceptability is very much tied to identity, so most people are reluctant to adopt a script other than the one they prefer and see as a part of their culture. Thus, when Pakistan was going to declare Urdu as the national language and thus impose the Perso-Arabic script on the population in East-Pakistan, its population felt that they had no option other

than to declare their independence in order to preserve their language and the Bengali script.

The Boro, Garo, and Khasi people do not wish to use the Assamese script “because the non-Hindu hill peoples tend to reject identification with the Hindu Assamese” (Smalley, 1964e:76). In the same vein, the promotion of augmented Devanagari in India has failed because people have separate identities and use their traditional writing systems, and put a high value on these (Coulmas 1999:140).

### *10.3.3 Perceived sanctity of the writing system itself*

When a religious faith and its sacred writings become closely identified with a specific writing system, not only the message but the writing itself may take on a ‘holy state.’ Modifying the writing system can thus be perceived as a sacrilege. Eira (1998:180) comments, “the status quo of the orthography may be regarded as sacred, i.e., as something in which change is perceivable only as degradation.”

### *10.3.4 Emotional reaction*

Even if there is no other basis, change may be resisted for purely psychological reasons. Smalley (1964d:58) comments: “People are, quite naturally, very reluctant to change from a system which has been in use for years, though on a small scale, even if one better in every respect is developed.” The Dutch orthography reform conflict, lasted about fifty years; emotions ran high: “proposals were strongly disputed...this conflict was affected by sympathy and antipathy towards modern linguistics and modern pedagogy” (Geerts, van den Broeck, and Verdoodt 1977:192).

### *10.3.5 Loss of prestige*

Individuals may wish to fight for two types of prestige: personal prestige and the prestige of a language. Fishman (1977:xvi) comments: “The introduction of a newly created writing system easily threatens to change established lines of relative advantage and disadvantage.” The elite in a society may not see popular education in a positive light. Coulmas (2000:56), discussing the introduction of Hangeul in Korea, comments, “The very idea of vernacular literacy ran counter to the communication practices of a highly stratified society in which elite literacy was a means of social control.” Those who learned Chinese characters and invested a lifetime to do so enjoyed much respect in society as a result; they were not necessarily interested in others attaining literacy by an easy road and their own knowledge becoming less valuable.

Those educated in a prestige language may oppose orthographies which depart from orthographic patterns of that prestige language. Baker (1997:119) reports on the Francophone elite opposing the Haitian Creole orthography; one individual, a maverick and proponent of French, even single-handedly promoted a rival orthography in 1947. In the Dutch orthography reform in the Netherlands and Belgium, people feared Dutch would lose status by moving away from French spelling patterns (Geerts, van den Broeck, and Verdoodt 1977:205).

### *10.3.6 Problem with the old form not perceived as acute*

With experience, readers become less and less dependent on the sound-symbol correlation of the writing system. They can depend increasingly on word recognition based on lexical memory, i.e., on “visual-orthographic representation” (Katz and Frost 1992:92). The phonetic detail might become “irrelevant to their identification” (Berry 1977:9) as long as there is “lexical” representation. Thus, reading an opaque orthography

can be quite efficient. Experienced readers would thus show little empathy for the struggles of new learners, and certainly would have no interest in orthographic changes to facilitate learning, as this would disadvantage them by invalidating the knowledge they have acquired. Smalley (1964d:59) comments: “There would, of course, be opposition from people who do not want to relearn. A pattern once learned takes on a sacredness which becomes sacrilege to violate.” Rabin (1977:155), reporting on spelling reform efforts for Hebrew in Israel, expounds:

The spelling difficulties I have enumerated earlier affect marginal and largely inarticulate groups: children, the uneducated, and new immigrants. The educated reader tends to feel that these people should make the same effort that he made himself in order to learn to read fluently, rather than causing him difficulties by changing his ingrained reading habits. Some even resent the very idea that others should have things made easier than they had themselves.

### *10.3.7 Loss of investment and shortage of resources*

Schools and publishers who have stocks of literature using a particular writing system may oppose its reform. Materials with the current orthography will be a ‘loss’ if a change were to come into effect. Schools suffer from lack of resources and school administrators fear having to retrain teachers and having to deal with new materials. Missions frequently produce much of the literature that exists in minority languages. To change the writing system would instantly date all the Bibles, catechisms, hymnbooks, tracts, and religious training materials.

Sometimes competing orthographies are in use. If those who use them are affiliated with different “camps”—whatever the nature of their affiliation (religion, denomination,

political party etc.)—giving an inch in the writing system domain may be seen as a loss of face.

### *10.3.8 Too drastic a change*

Even if there is agreement that there is a need for change, efforts may backfire if change is introduced in too liberal a measure. Tolerance of change is generally low; conservatism in proposed changes is more likely to be met with acceptance. What is perceived as ‘reasonable’ may be less than what is needed, but it pays to introduce modest changes so as not to end up in a deadlock.

Kwan-Terry and Luke (1997:279) report on past efforts to simplify Chinese characters. Some changes were quite natural because people had been prepared for them. Not so for other changes: “Occasionally the concern to reduce the number of strokes gets on top of good sense. The result is radical simplifications not acceptable to the majority.” One of the reasons for not accepting the reform was that one old symbol, the symbol for *not*, was one of the best known characters, and this made changing unacceptable. A large number of characters were simplified in 1956 and 1964, but when an additional 853 simplified characters were proposed in 1977, “the scheme was judged by most people as having gone too far” (Kwan-Terry and Luke 1997:279).

When too many changes are introduced, insecurity sets in. The literates feel semi-literate; teachers fear making mistakes. Also, a firm stand will likely be taken against changes perceived as ‘ridiculous’. For instance, although most non-Roman scripts, including Arabic, make no upper-case/lower case distinction, the idea of abandoning the use of uppercase in Roman-based orthographies is not likely to catch on. When keyboard designers met to deal with challenges of the African Reference Alphabet, they “proposed

dispensing altogether with upper-case letters, which were described by a participant...as ‘un luxe culturel et decorative’<sup>64</sup> (Baker 1997:115). It is not likely that African languages will do away with capital letters. There are psychological and educational reasons for rejecting such a move. Texts written exclusively in lower case would appear ‘inferior’ and ‘less prestigious’ even to the pre-literate who has been exposed to an international Roman-script language.

George Bernard Shaw, renowned author and playwright, critical of the English orthography, made provision in his will that someone should be commissioned to design a new writing system for English, consisting of at least forty letters. As a result, a competition was held in 1958 and an alphabet created by Kingsley Read was chosen as the winning entry. This alphabet has come to be known as the Shavian Alphabet. It consists of an original combination of lines, curves, squiggles and loops, many of which vary only in direction. Voiced and unvoiced consonant-pairs were assigned the same shape, facing in opposite directions, with the unvoiced of the pair being a “tall” letter, and the voiced counterparts descending, in part, below the line. Each letter has only one case. The punctuation marks from standard English were retained, but an additional mark was provided to indicate proper nouns. The system is phonemic, except for five of the most common grammatical function words, which have a shorthand morphemic representation: *and, for, of, the, and to* (Ager 1998–2006d). Ager reports that the alphabet “was never seriously considered as an alternative for writing English.” Why not? The whole English speaking population would be cast into a non-literate state. The Unicode Consortium recognized that “as with other attempts at spelling reform in English, the [Shavian]

---

<sup>64</sup> My translation: “a luxury of a cultural and ornamental nature.”

alphabet has met with little success.” Despite this, because “it has its advocates and users” they have made provision for it in the Unicode Standard (Unicode 1991–2006) and some fonts have been designed for it.

### *10.3.9 Too frequent a change*

When successive waves of literature in the early stages of orthography development use different conventions, some doubt may be cast on the competence of the people involved in the decision making process. For this reason it is important to refrain from producing literature on a large scale until the orthography has been tested and proven efficient.

When changes are imposed with unacceptable frequency, the good intentions and wisdom of the change agents and the governmental agency held responsible may be called into question. Initial good will on the part of the people may turn sour if they feel they are being trifled with or suspect some kind of ulterior motive on the part of the agencies involved. Too many changes will undercut motivation for reading and writing and for learning these skills.

This ‘frequent change argument’ seems not all that uncommon. Geerts, van den Broeck and Verdoodt list it as one used in the Dutch spelling conflict. They included it under the “more emotional” arguments: “This is the hundredth change in a relatively short time. Can’t we finally stop making fools of ourselves” (1977:203–4).

The discussion in this chapter seems to indicate that there may be some *do’s and don’ts* related to writing system reform. Before drawing up some general principles for maximizing the chances of successfully implementing an orthography reform, I will first present some case studies which illustrate these.

## **CHAPTER 11**

### **ORTHOGRAPHY REFORM CASE STUDIES**

Each orthography reform undertaken—or attempted—results in a story. Although conditions differ from case to case, there are common threads and lessons to be learned. Case studies support Fishman’s (1988:1646) statement that “the gains associated with the proposed revisions must authoritatively and effectively outweigh the costs for those who have already mastered and benefited from the prior system.” If benefits do not outweigh the cost on an individual or community level, successful implementation is not likely.

This chapter will present select case studies which illustrate some of the principles discussed in previous chapters.

#### **11.1 Script replacements**

##### *11.1.1 Turkish*

Latin became the official script for Turkish in 1928. Literature was previously written in classical Ottoman Turkish with Perso-Arabic script. The change was initiated and implemented by President Mustafa Kemal Atatürk (1881–1938), whom Coulmas (1989:243) describes as “a strong nationalistic leader.” The script reform reflected “changing political alignments” and is viewed as part of Atatürk’s larger “Westernization” program. (Coulmas 1996:513; 2003:203). There is general agreement that Turkey’s switch to Latin script was a success. Since the Arabic script is associated with the Islamic faith—the religion of the majority of the population in Turkey—it is

noteworthy that religious identity was not an obstacle to the script replacement. Atatürk was able to circumvent this sensitive issue: The dignity and value of the Arabic script in the religious domain was never questioned or challenged, nor were religious leaders provoked. The reform focused on national identity and on helping the Turkish language flourish. The claim was made that Turkish could not be represented as it should with Arabic script. Coulmas (1989:244) hypothesized that because of the Arabic script's long literary tradition, modifying it to accommodate Turkish vowels and other features might have resulted in more resistance than adopting a totally new script. The literacy rate was very low at the time, so few people experienced the script change as a personal setback. Coulmas (1996:515) briefly describes the Turkish alphabet: Twenty-three of the Roman alphabet letters are used, —all except <q, w, x>. The alphabet is augmented by use of modified letters and diacritics. An undotted vowel set <ı, İ> is used in addition to a dotted set: <i, İ̇>. Provision is made to note glottal stop and vowel length in words of Arabic origin.

Success of the reform was for the most part due to careful planning and preparation with the actual change happening practically overnight. Implementation took place through mass literacy campaigns and the public school system. Today, Turkey's adult literacy rate is 85% and the youth literacy rate (ages 15-24) is 96.5% (UNESCO 2002).

This case study illustrates the following:

- An authoritative, respected leader can facilitate change.
- Careful planning and due consideration for a population's identity contribute to successful implementation of change.

- A script change is easier to implement when few people are literate in the script to be supplanted.
- Acceptance is only part of the picture; implementation requires follow-up through the educational system.

### 11.1.2 Replacement of German *Fraktur* and *Kurrent* scripts

A different script tradition developed in German-speaking countries than in the rest of Europe. While *Antiqua* (Latin print) was used for most languages, German was usually written with *Fraktur* in print, and *Kurrent* in handwriting. The three scripts were in concurrent use.<sup>65</sup> Foreign language texts and German technical and scientific publications were printed in *Antiqua*; newspapers, religious materials, and novels were printed in *Fraktur*; personal communication and official records were written in *Kurrent*.

*Fraktur*, also known as *Gothic*, was mostly promoted in Germany, Switzerland, and Austria for several centuries through the use of Luther's Bible (1534) and Small Catechism (1529).<sup>66</sup> There was a movement to abolish *Fraktur* in the late 1800s, but a large majority of Germans were in favor of retaining the script. "In 1911 the German Reichstag [parliament] voted 75% in favor of this writing" and it became "the graphic symbol of a nationalistic movement" (Augst 1996:767). Augst reports that in Nazi Germany (1933–1945), the percentage of books printed in *Fraktur* increased by nearly 20%. In schools, only *Fraktur* and *Sütterlin*, a modified form of the *Kurrent* script,<sup>67</sup> were

---

<sup>65</sup> Some (Coulmas 1996) equate *Fraktur* with a typeface (Gothic Type or 'black letter') and do not consider it a separate script. However, others (Jensen 1970, Augst 1996) treat it as a script in its own right.

<sup>66</sup> Luther lived 1483–1546. (Internet Christian Library 2004)

<sup>67</sup> The *Sütterlin* form of the *Kurrent* script was designed by the Viennese graphic artist Ludwig Sütterlin (1865–1917). It was implemented as the standard script in Prussian schools in 1915, and by 1934

taught for a season, but the attitude toward these scripts changed. *Antiqua* was associated with the international scene, education, and science and thus was considered prestigious, while Fraktur was not. The German regime became concerned: their propaganda written in Fraktur could not be read by non-Germans. Thus, in the late '30s *Antiqua* and Fraktur were taught concurrently in German schools. (See Figure 2 for a page from a German primary school reader (Velhagen und Klasing 1938). The story is printed in *Antiqua*; the poem at the bottom is printed in Fraktur. )

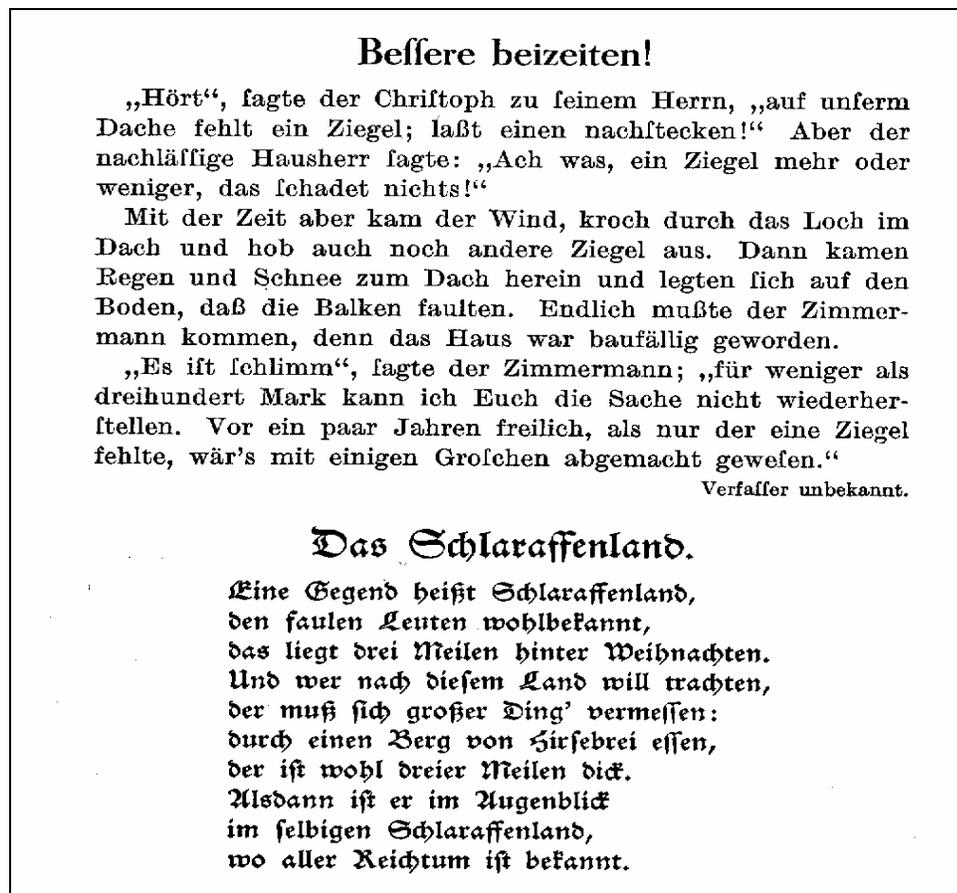


Figure 2. German school reader using two scripts.

was taught in almost all German schools. It eliminated some of the sharp angles, straight lines, and abrupt changes in direction characteristic of the classic Kurrent script, making it easier to write for young children (Walden 2006).

From 1940 on, in Nazi Germany, Fraktur was no longer looked upon with favor and the focus in schools switched to teaching Antiqua. Augst (1996) reports that Fraktur has not been taught in schools since 1945. However, it is still in use in some domains. A German Bible in Fraktur can still be purchased in 2006. In addition, Fraktur serves the same purpose as calligraphy. It is sometimes used for certificates, greeting cards (especially sympathy cards), and for ornamental purposes, such as chapter titles.

Children growing up in post-war Germany received no special help to enable them to read Fraktur. Making the adjustment was not difficult since the majority of the letters were similar to Latin script print. Tall letters were still tall; most short letters were still short. The spelling rules were the same for the most part. The main differences were that in Fraktur (1) there were additional descenders below the line; (2) there was an additional grapheme <ſ> for /s/ ; and (3) a few uppercase letters were much more divergent in shape from their lower case equivalents. Fluent readers, perceiving the overall shape of words, could adjust to reading older documents without a problem. The two scripts were about as divergent as two extremely different fonts. The situation, however, was different for the Sütterlin script.

The script replacement came with a cost for some individuals. The transition went more smoothly for some than for others. My maternal grandmother could only write in Kurrent; my father's mother, who was the same age, always wrote in Latin cursive. My parents had learned Sütterlin first, and then later *Lateinschrift* (cursive Latin). My mother's older sister never made the change-over and was always apologetic about her 'illiteracy' in *Lateinschrift*. It was not just a matter of her age, but probably also of the school district in which she was educated. One day, when I was in the 4<sup>th</sup> grade, my

teacher, Fräulein B., wrote some scribbles on the board that I could not decipher. She distributed copies of a special penmanship book. It was time for us, the young generation, to learn to read and write the *Sütterlinschrift* so we would be able to read old family records and our grandparents' letters. Nothing in this script looked familiar. It was like starting over. For the script replacement to be successful and printed resources to not lose their value, two or more generations had to be biscriptal.

During the time of transition, script mixing was common. My birth certificate bears witness to this fact. (See Figure 3; some facts have been erased.) The certificate was printed in Antiqua.

Gemeinsames leibliches Kind der Eheleute, deren Eheschließung  
auf Seite 4 und 5 beurkundet ist.

**Geburtsurkunde** E 1

1 Standesamt ..... Nr. 149/58

2 Georg Seemann

3 .....

4 ist am .....

5 in ..... geboren.

6 Vater: Dr. Hermann Karl Günz .....

7 Seemann, sonnyalif, woffufast in

8 B .....

9 Mutter: Elfriede Seemann geborne

10 Kachett, sonnyalif, woffufast in

11 B .....

12 Änderungen der Eintragung: Keine .....

Figure 3. Antiqua, Sütterlin, and Latin scripts in an official document

We note a mix of handwritten scripts:

Sütterlin <B> on line 6; Latin Cursive <B> on line 8.

Sütterlin <e> as well as Latin cursive <e> appear, often on the same lines (see line 2).

Sütterlin <h> is abundant throughout the document; three appear on line 10.

One Latin cursive <h> is also found on line 10.

This case study illustrates:

- Nationalistic movements can give impetus to a writing system change.
- The public school system is the key to a smooth transition: it is helpful to not only teach the new system but to also establish links to heritage and help overcome generational differences.
- Formal instruction may not be needed for small changes.
- A drastic change is disconcerting to those who learned a previous system: it may affect their self-image and confidence.
- Extra help is needed for changes where “transfer” of skills is limited.
- The older population should not be deprived of the literature they value. It should continue to be printed in the system they know.
- The mixing of writing systems is to be expected during the time of transition since habits do not change quickly.

### *11.1.3 Korean Han’gul*

Chinese became the official script in Korea in the 5<sup>th</sup> century and ruled there unchallenged for over a 1000 years. But because, as Kim (2000:6) expressed it, “they belong to two different linguistic stocks,” characters served well for writing Chinese but

not for Korean. King Sejong (1397–1450) made it a personal goal that a native script would be designed for Korean. Kim (2000:7) comments: “It is unequivocally clear what the king was striving for: a simple writing system for mass literacy. What Sejong devised in the process was more a revolution than an evolution.”

The writing system which resulted, Han’gul, was referred to as *The Correct Sounds for the Instruction of the People*. Many have praised this writing system and King Sejong for his wonderful achievements. Kim-Renaud (2000:13, 15) calls the writing system reform a “linguistic coup d’état” and indicates just how revolutionary Sejong’s ideas and actions were: “Compared to the passive and reading-oriented literacy of the time, Sejong’s vision was of a universal creative literacy, in which expressing one’s ideas in writing was the central issue... Sejong believed that universal literacy results from the simplicity and easy learnability of the writing system.” Up to that point *literate* was defined as ‘knowing Chinese characters.’ Sejong, by promoting popular education in the common language, was redefining literacy.

Provision was made to represent Chinese sounds not found in Korean so Chinese words could be written with Han’gul. Since a literate population would need something to read, King Sejong initiated publishing projects, including original writings, translations, and transliterations; seventeen major Buddhist works were translated. Kim-Renaud (Kim-Renaud 2000:27, 31) reports that in some of these publications pure phonemic writing increasingly gave way to morphophonemic spelling, and that letter shapes changed slightly since brushes were used as writing instruments.

The king’s “bold language planning,” which was based on both “scientific and humanistic motivation” was not appreciated by some of the influential sages in his

kingdom. These were not concerned about educating the masses. Some of them expressed their opposition openly. One intellectual stated that only barbarians would base writing on local speech (Kim-Renaud 2000:25, 35). Almost his entire cabinet opposed the king's alphabet project (Kim 2000:6). Nevertheless he persevered.

Unfortunately, a single person's vision cannot make a lasting impact if it is not shared by others. Coulmas (2000:56–57) comments: "It was a top-down reform initiated by the highest representative of the state... Yet the reform failed." How could that be? Coulmas explains: "The educated classes looked with disdain on the new system... [I]ts prestige could never match that of Chinese... Sejong was ahead of his time."

Today, however, "the creation of han'gul is celebrated as the proudest moment in Korean cultural history" (Coulmas 2000:56). Chinese affiliation and tradition had never been perceived as a threat to Korean society, thus motivation for a nationalistic movement was lacking during Sejong's time and centuries afterward. That changed when Japan took control of the Korean peninsula. Basically, Korea was 'annexed' to Japan, and was a Japanese colony from 1910-1945. The Japanese had harsh language policies, imposing their language and their script. This oppression "provided the conditions necessary to kindle the fire of linguistic nationalism... Under the Japanese, writing han'gul became a visible symbol of opposition and self-esteem... Korean in han'gul letters served as a vehicle for opposition to Japanization" (Coulmas 2000:56–57).

Thus, after the colonial period, about 500 years after King Sejong initiated his writing system reform, Han'gul became the official script of the Korean language. (See 6.8 for a discussion on differences between North Korea and South Korea's use of the script.)

This case study illustrates:

- A vision can perish with its visionary. Political power and hard work contribute toward success, but in themselves are inadequate: there has to be a general consensus and good will toward the change.
- Socio-political factors outweigh linguistic factors; a linguistically elegant writing system can go unappreciated.
- A change in the socio-political situation can set the stage for writing system reform; association with a popular agenda can be a crucial factor.
- Simplicity is desirable in a writing system.
- It is important to provide something to read for the literate population.

## **11.2 Revision as part of the orthography design process**

### *11.2.1 Shona*

Shona,<sup>68</sup> a Bantu language, is spoken by about 10.7 million people in Zimbabwe, Botswana and Zambia (Gordon 2005). At first, missionaries working in different dialects of Shona did not realize how similar the dialects were. Fortune (1972:16) reports that at one time “there were five versions of the New Testament circulating, each in its own dialectal spelling, and different from all the others.” In the 1920s, steps were taken to develop a single orthography for Shona.

Doke researched the dialects in 1929 and proposed a common alphabet for them. One group of dialects was chosen as a reference for the standard because of the number of people who spoke them, their central location, and because these had a wider variety of sounds and therefore were representative of all the other dialects.

---

<sup>68</sup> This case study is based entirely on *A Guide to Shona Spelling* (Fortune 1972).

Doke's proposed alphabet was to be implemented in 1932. It was phonemic, employed a one sound-one symbol correspondence throughout, and used 24 of the 26 letters of the Latin alphabet. These were augmented by one digraph and eight phonetic symbols, based on the Africa Alphabet, which Africanists were promoting at the time.

Doke's alphabet, known as the 'New orthography of 1932,' was not welcomed as had been expected. Readers encountered too many unfamiliar letters. Further, people were concerned with various technical issues, the most serious of which was that preparing Shona books for printing would require special typewriters and printing equipment and thus would drive up the cost of production. Doke responded by substituting Latin letter digraphs for two of the special symbols. Although the Ministry of Education prescribed the use of this orthography, it was never accepted. Various agencies and newspaper publishers simply continued to employ older spellings.

One main complaint was that the Shona orthography was too 'unlike' English and that "it would hinder the production of books and the growth of the habit of reading" (Fortune 1972:18). This discontent resulted in a committee reviewing the Shona alphabet and proposing a new orthography in 1955. It had 22 single letters and seven digraphs, none of which involved special symbols. Eight phonemes were to be symbolized by four graphemes, i.e., there was underrepresentation between four pairs of distinctive sounds. This proved to be a bad decision because it created too many ambiguities in texts. While preparing a dictionary, it was decided that each of these sounds needed to be represented. Thus in 1967, the Shona Language Committee recommended that all sounds be distinguished as in Doke's alphabet, but that this be done without using special symbols. To meet the symbol needs, three additional digraphs were introduced, as well as <n'> .

The more common sounds were to be represented by the simpler symbols; digraphs were to represent the less frequent sounds. Eight of the digraphs in the resulting alphabet involved the letter h: < bh, ch, dh, mh, nh, sh, vh, zh > .

The Minister of Education approved the language committee's recommendation in 1967, establishing the Shona standard alphabet. This alphabet has been in use since, but there is some discontent. Some are calling for revisions. Zivenge (2005) calls the Shona alphabet 'defective', raising the issues of non-representation of certain sounds and language change. Magwa (2002), focusing on dialect issues and difficult word division rules writes:

[T]he writing system...is linguistically constricting, making it difficult for speaker-writers to write the spoken language correctly. Speakers of different dialects of ChiShona experience difficulties in spelling and word division because of a defective writing system...known as the "standard" orthography that does not cater for dialectal variations.

Magwa proposes a solution, advocating a 'common alphabet' but allowing variant spellings: "[D]ifferent dialect speakers could then combine these single letters to form acceptable sound combinations for each particular dialect." He calls for more consistent word-break rules, pointing out that the ordinary speaker-writer finds the rules confusing regarding writing as one word or two words and the use of hyphenation. "One ends up not knowing when to use and when not to use the hyphen... Furthermore, there is no linguistic justification for writing the word *mesomeso* as one word, while writing *kare kare* as two words."

This case study illustrates the following:

- Linguistic analysis is a good starting point for orthography discussion, but a linguist’s ideas do not necessarily line up with people’s desires.
- People are likely to continue writing in the way they are accustomed.
- It is not wise to underdifferentiate phonemes in order to avoid using special symbols or to make a language look more like the lingua franca.
- The standardization process takes time—in this case 38 years.
- Revisions are part of the standardization process.
- A language committee can be key in working toward a unified writing system.
- Shortcomings of a writing system standard will, in time, come to light.
- Dialect issues need special consideration in writing system design.
- Word break rules, not only symbol choices, need to be addressed.

### *11.2.2 Southeastern Nochixtlan Mixtec*

Nochixtlan Mixtec<sup>69</sup> of Mexico is spoken by about 7,000 people according to a 1990 census. About 4,000 of them are reportedly monolingual (Gordon 2005).

Nochixtlan Mixtec has both lexical and grammatical tone. There are floating tones and tone sandhi. Finding a workable system for representing tone required extensive testing. The following options were tried and failed:

1. Zero marking. An unacceptable number of ambiguities occurred. The functional load of tone was too high.

---

<sup>69</sup> This case study is based on personal correspondence with SIL consultant Inga McKendry (3.20. 2006).

2. Marking one of each of the tone pairs with an acute accent, with the choice of which to mark being somewhat arbitrary. Mexican language development partners disapproved of marking low-tone verbs with an acute accent; they associated the acute accent with conventional high tone marking.
3. Marking surface tone with seven different symbols. This was too complicated and ranked low on teachability.
4. Marking tone on minimal pairs only. This did not bring about the desired consistency in writing. People did not think of words as part of a tone group.
5. Marking what the linguist hears. This did not work since mother-tongue speakers are not aware of allotones, but perceive underlying tones.

A different system is presently being tested and seems to work: Grammatical functions are marked and meaning differences taught rather than focusing on the tone associated with these functions. For example, (1) imperfect aspect is marked with an acute accent on the verb; (2) perfective aspect is marked with an underline; (3) contrafactual is marked with an acute accent (there is no confusion with imperfect due to the presence of a special prefix *ni-*); (4) irrealis is unmarked if bearing a mid tone, but marked with a grave accent when bearing a low tone. Testing continues.

This case study illustrates:

- The number of ambiguities which occur are an indicator as to whether zero marking of tone is acceptable.
- Testing is an important aspect of the orthography design process.
- Testing needs to be done for both reading and writing.

- Mother-tongue speakers cannot be expected to write features of which they are not aware.

### *11.2.3 Kalagan Tagakaulu*

Kalagan Tagakaulu,<sup>70</sup> is spoken in the Philippines, in Southern Mindanao, Sarangani Province and Davao Del Sur. The Ethnologue (Gordon 2005) reports 71,356 speakers for this language, but there may be as many as 100,000 (Bus and Jean Dawson, personal communication). About 40,000 are believed to be monolingual.

English and Filipino (Tagalog) are the two official languages of the Philippines. Filipino is spoken as a second language throughout the Philippines and serves as the national lingua franca, but it is not used as extensively in the area where Tagakaulu is spoken. There the language of wider communication is Visayan. Visayan is a written language, but serves mostly for oral communication.

SIL linguists became involved in Tagakaulu language development in 1953. A preliminary phonemic and grammar analysis were completed during the early stages of the program. An alphabet book and a primer were printed in a tentative orthography. The orthography scored high in acceptability and in readability, even though it had not adequately taken stress and length into account.

Based on recommendations by a linguistics consultant, some orthography changes were implemented in the late 1960s. These changes were not motivated by observed shortcomings of the orthography; neither were they motivated by a request by the population, learner, teacher or general user; instead they were based on the consultant's

---

<sup>70</sup> The information for this case study was provided by Bus and Jean Dawson (SIL consultants) during an interview in February 2006.

feelings that it would be better to conform to the spelling patterns of the regional and/or national language.

The orthography changes in Tagakaulu were based on Visayan. The two languages differed in some important aspects, which became an issue for the revised orthography:

1. Completive aspect is marked in Tagakaulu verbs by palatalizing the first consonant. This was indicated in the first orthography by C+y. Since Visayan consonants were never followed by <y>, it was decided to write palatalization with C+i instead. This was problematic since Tagakaulu had CVV syllable structures. It was difficult to determine if a CVV sequence should be interpreted as C<sup>j</sup>V or a true CVV. (See Table 15.)

Table 15. Marking Tagakaulu completive aspect.

	verb (incompletive aspect)		verb (completive aspect)	
phase 1	lumawat	<i>go out</i>	lyumawat	<i>went out</i>
	bumasa	<i>read</i>	byumasa	<i>read (past)</i>
phase 2 (problematic)	lumawat	<i>go out</i>	liumawat	<i>went out</i>
	bumasa	<i>read</i>	biumasa	<i>read (past)</i>

2. Visayan did not normally symbolize word-final glottal stops, therefore, it was decided that Tagakaulu word-final glottal stops would no longer be written. This decision was not expected to be problematic because dropping them did not result in a large number of homographs. However, the glottal stop is a common consonant in Tagakaulu. To not write it in one of its positions undermined teachability and reading fluency.

Because readability was severely affected, these orthography decisions were reversed—to everyone’s relief. In addition, to further enhance readability, vowel length was to be

represented from then on. Doubling the vowels to mark vowel length was tried but did not work. Readers, literate in Filipino, would stumble in their reading of Tagakaulu; restarts were frequent. The reason for this is as follows: In Filipino, two consecutive vowels represented vowels belonging to two syllables, separated by a glottal stop, i.e., <aa> stood for [aʔa].

It was therefore decided to mark vowel length with a macron, (<ā> for /a:/ for example), and with a tie bar if vowel length resulted across morpheme boundaries from a vowel sequence ( <āā> for {\_a+a\_} for example).

This orthography, implemented in the third phase of experimentation, has served for seven years without difficulties. Despite the differences between the writing systems, i.e., writing six vowels, marking length and glottal stops in Tagakaulu, those who learn to read in Tagakaulu are able to transfer their skills easily to reading Filipino and Visayan.

This case study illustrates the following:

- Linguistic factors, readers' needs, and ease of transfer should all be taken into consideration.
- Readability should not be sacrificed.
- Testing an orthography is essential.
- Willingness to reverse detrimental decisions can make the difference in language development efforts.
- The orthography design process can be seen as a sequence of stages proceeding by trial and error.

### *11.2.4 Ngiti*

Ngiti<sup>71</sup> is a Central-Sudanic language spoken in the Democratic Republic of Congo. There are about 100,000 speakers (Gordon 2005).

Ngiti church leaders decided to devise an alphabet for their language in the hope of eventually translating the Bible into Ngiti. After printing a hymnal it became evident that the alphabet they had designed was difficult to read. They asked for outside help in devising a better solution. Dr. Constance Kutsch Lojenga responded to the request for help. Instead of simply offering suggestions, she used a community participatory approach to discovering the optimal orthography. At first between ten and fifteen volunteers participated in the work sessions, which were held about three times a year. Gradually this number increased as more and more highly educated individuals started taking an interest.

The participatory research involved the mother tongue speakers in data gathering and phonemic analysis. They discovered that the reason the five-vowel alphabet they had designed did not work was that Ngiti had nine contrastive vowels. (These are due to a difference in the high and mid vowels between advanced and non-advanced tongue root positions.) In addition, when analyzing the consonants, it became evident that although /b/ and /β/ had been differentiated, in practice, the writing of <b> and <βh> had been inconsistent.

The group decided on a 'tentative orthography'. Since the original attempt to write Ngiti had been based on Kiswahili, the consultant encouraged mapping sounds that were identical in the two languages to the same graphemes. This went along with the group's

---

<sup>71</sup> This case study is based on an article by Constance Kutsch Lojenga (1996).

intuitions and desires, and allowed for maximum transfer to and from Kiswahili. The group discussed options and decided on symbols for sounds not found in Kiswahili.

Kutsch Lojenga (1996) stated: “This method helped the speakers to gradually get used to them and to respond positively or negatively early on.” She felt that this approach to linguistic analysis had many benefits, including the following:

1. The analysis went faster and was of a better quality than would have been possible if a linguist had to come up with solutions with input from just a few speakers.
2. Participants were trained on the job and were equipped to be effective participants in the language development process.
3. The participatory approach enhanced community ownership of the orthography.
4. The discovery of the language as a group created excitement about its richness, uniqueness, and affirmed the speakers in their identity.

Kutsch Lojenga contrasts this with the traditional approach to orthography design:

A ‘Western’ way of handling the decision-making process for an orthography might be to present a neatly worked-out proposal, to explain the reasons behind the choices, discuss, reason, and come to a conclusion as to what would be the best solution. However, things are more likely to turn out positively if the people concerned have ‘experienced’ the written language development over a period of time, if their awareness has been raised and they have been fully involved in the discovery process.

This case study illustrates:

- Linguistic analysis is foundational.
- Community involvement can result in community ownership.

### 11.3 Reform of languages with a literary tradition

Change is not easily introduced into writing systems of languages which have a long literary tradition. I already mentioned the Dutch orthography reform, which resulted in a long drawn-out conflict. It was very costly, in time and energy as it dragged on for about fifty years. (See Sections 10.2 and 10.3.) The following case studies are well documented in a variety of resources. I will not go into details, but rather summarize challenges and issues and highlight some patterns.

#### *11.3.1 Chinese and minority languages in China*

China has been working hard at modernization. The question is: does being a modern nation state and participating in the ‘global village’ require sacrificing a unique writing system? Is giving up valued traditions equated with progress? Some seem to think so.

Wong (1960:A35) wrote the following almost 50 years ago:

If the Communists have their way in China, the age-old characters of the Chinese language will finally join the Egyptian and Mayan hieroglyphics and the more recently buried Vietnamese ideographs in oblivion... [T]he Communists’ effort to remold the nation includes a drive for drastic changes in a language ill suited to science and technology, the education of the masses, to the communications of a directed economy, to their international purposes.

Wong reports that the traditional Wade-Giles romanization system for Chinese, which he labels “antiquated and artificial,” was being replaced by Pinyin. “In August 1959 Shanghai merchants were ordered to use Pinyin in their signs. (Wong 1960:A51). The military reportedly was using Pinyin effectively: “Illiterate recruits for communication battalions from Kwangtung province and the Kwangsi Chuang Autonomous Region are

said to have learned Pinyin in three or four months with the help of teaching aids” (Wong 1960:A49).

Forty years later, in 2000, the Library of Congress announced that, “Beginning October 1, American libraries will join the international community in using Pinyin as the standard romanization scheme for Chinese characters. The Wade-Giles romanization system, followed in American libraries for the last century, will no longer be used” (Research Library Group 2000). Working toward this new standard called for a tremendous amount of work so that older works could be accessed effectively.

But what of Chinese characters? Zhou (2001:39) reports on language planning goals of the 1950s: “The long-term goal for Chinese writing system reform was to replace Chinese characters with a romanized system, while the immediate goals were to use a romanized system as a pronunciation tool and to simplify Chinese characters.” In 1958 a plan was made public which expressed a “unified language policy” for the nation (Zhou 2001:41). It outlined principles for the creation and reform of writing systems for minority languages in China. The Roman alphabet was to be used as the standard and cross-language harmonization was highly valued. This, however, meant giving up other traditional systems, which provoked opposition. Emotive labels like *bourgeois linguistics* were tossed about; the time was not right for reform. Zhou reports that frequent language policy changes (swinging from assimilation to accommodation) resulted in chaos. There were “as many as two new writing systems, and two to three versions of the old writing system” for a single minority language, which interfered with education and literacy efforts and resulted in very high illiteracy rates. Between the lines we read that Zhou might not think that “direct involvement by governments in minority language

maintenance and development” is a good idea. He talks of “a painful lesson” and exhorts the reader: “The creation of a language policy requires long-term consideration, and a writing system reform requires caution upon caution before it is carried out. A reform may be reversible, but its consequence is not. This is true not only for politically motivated reforms of writing systems, but also for linguistically motivated reforms” (Zhou 2001:60).

Chinese characters continue to be central to Chinese education and communication. The pinyin phonetic alphabet is taught first in school, but the heavy emphasis on characters beginning in grade three and the neglect of pinyin in higher grades can result in *hui-sheng* ‘return to unfamiliarity’ (DeFrancis 1977:139). An inverse phenomenon involving loss of mastery of characters seems to be occurring amongst those who are dependent on computers. Typing Pinyin automatically renders keystrokes into characters; the technician’s memory is not taxed and it is feared that this may be eroding handwriting skills and mastery of characters (Lee 2001).

Although some political leaders have been in favor of taking drastic measures in reform, i.e., a possible script replacement for Chinese, this has not been the expressed desire of the majority. DeFrancis (1977:139) wrote: “Again and again in the literature on the reform movement the point is emphasized that the Phonetic Alphabet is not intended to replace characters—certainly not now, though perhaps sometime in the future, and then only after much preparatory ‘study’ and ‘experimentation’. It appears that little official encouragement has been given to specific preparatory work.” Rather, the focus of writing system reform has been on the simplification of characters (Kwan-Terry and Luke 1997).

Attempts to promote the use of simplified characters have been only partially successful.  
(See 10.3.8.)

This case study illustrates:

- Even dynamic leaders cannot single-handedly introduce reform and make it work.
- Policy changes happen.
- Frequent changes in policy can be detrimental.
- Government involvement can at times be harmful.
- Literacy in several scripts can be achieved, but skills must be kept up in each.

### *11.3.2 Quechua of Peru*

Quechua is a family of languages of the Andean region, members of which are spoken in Bolivia, Peru, Ecuador, Northern Chile, Argentina, and Southern Colombia.

The Ethnologue (Gordon 2005) lists 32 different Quechua languages/dialects for Peru alone. The number of speakers of these dialects vary greatly, ranging from 250 speakers of Pacaraos Quechua (ethnic population 900) to 1.5 million Cusco Quechua speakers, of which 20-33% are monolingual.

Due to the low status of the Quechua language, speakers were formerly ashamed to speak it. Parents discouraged their children from speaking Quechua, insisting that they speak Spanish. Motivation for Quechua literacy is generally very low. Hornberger (1998:392) comments on the Quechua situation:

Despite their large numbers, the oppression and exclusion of Quechua speakers has been a constant fact of Peruvian society. Thus, it was of great significance when, in 1975, as part of a larger programme of socially progressive reforms undertaken by the Revolutionary Government of Juan Velasco Alvarado, Quechua was declared an official language co-equal with Spanish.

Quechua revitalization efforts are taking place in some of the languages, resulting in some Quechua speakers having a more positive attitude toward their language. Quechua is taught in some rural Peruvian schools, promoted through some adult literacy efforts, and broadcasts on the radio in a few places.

Because of the high rate of shift to Spanish among Quechua speakers, those involved in revitalization efforts believed that the orthography which had been in use for Quechua needed changing. Certain Spanish writing conventions were to be eradicated. This related to no longer using <c> and <q>, replacing <j> with <h> and, more importantly, to the issue of how many vowels should be written in Quechua. On this latter issue, some Peruvian linguists and members of the Quechua Language Academy ended up in two opposing camps.

Both camps had the best interest of the Quechua population at heart. They agreed “that everything possible should be done to stop the abandonment of Quechua by its speakers. All wish Quechua to be revitalized by means of writing, and by broadening its functional domain to include education, legal transactions, mass media, and so on” (Weber 2006:3–4). However there was disagreement as to what would facilitate and what would impede this from happening. Weber (2006:3) states that “the debate has become intense, with acrimonious exchanges in national and provincial newspapers, as well as in academic publications.” Those involved in the debate came at the orthography issues from different angles:

*1. an ideological standpoint:* The Quechua language must be liberated “from its linguistic subjugation to Spanish” and the Quechua people must be liberated “from their social subjugation to Western culture.” Those who wish to promote the use of

five vowels are labeled as “assimilationists” (Weber 2006). The promoters of a three-vowel writing system focus on establishing a separate identity for the Quechua people and on the unification of Quechua groups. They promote writing only three vowels in Quechua: < a, i, u >. They believe that writing system uniformity will contribute to social unity.

*2. an educational standpoint:* The Quechua orthography must be easy to learn for the Quechua community in order to promote Quechua literacy. Writing Quechua with three vowels is representative of proto-Quechua and just a few dialects, but not all the currently spoken Quechua languages. Five vowels have been written in the Quechua orthography since the beginning of colonization. Some see no reason to abandon this orthography which has served effectively.

Neutral terms for the two camps are the “tri-vocalistas” and the “penta-vocalistas.”

Linguist Nancy Hornberger, having lived and worked in Peru for a decade, takes a personal and professional interest in the Quechua situation. She sums up the situation: “With more than ten million speakers and numerous local varieties, the unification and standardisation of Quechua/Quichua has been a complicated politically charged and lengthy process...the process is far from complete, and multiple unresolved issues remain, at both the national and local levels” (Hornberger 1998: 390).<sup>72</sup> She points out that the university linguists often do not speak Quechua, while those of the Quechua Language Academy are all speakers of the language.

Weber, a professional linguist who has learned and done research in Quechua and been actively involved in promoting the language through literacy and education, writes:

---

<sup>72</sup> Varieties of Quechua in Ecuador are referred to as Quichua.

“It might be expected that a debate of this kind would be resolved on the basis of linguistic, pedagogical, sociolinguistic, psycholinguistic, typographic, or historical criteria, taking into account – of course – the opinions of Quechua speakers. However, that has not been the case” (2006:3). Weber recognized that as a foreigner his opinion is not likely going to be valued. But since it supports “an opinion generally held by the Quechua people themselves” he set out to show that there is “considerable scientific justification” for writing Quechua with five vowels. Weber believes that others have not come to the same conclusion because “their ideology has caused them to lose objectivity regarding the relationship between Quechua and Spanish” (Weber 2006:x).

Weber argues that basing Quechua spelling on a proto-form of Quechua and forcing speakers of certain dialects to read and write an opaque orthography and memorize spelling rules (or the spellings of individual words), will not contribute positively to promoting Quechua. He agrees that the mid-vowels [e] and [o], which were excluded from the writing system, are allophones in some cases in some dialects. However, he cites the analyses of various linguists for many Quechua languages, which show that the mid-vowels are phonemes in these languages. Weber claims that different dialects developed along different lines and that it is therefore not reasonable to expect spelling harmony between them all. He points out that those who promote the use of only three vowels insist on ‘minimal pairs’ to establish contrast but that this is not a requirement in sound linguistic analysis: contrast can also be established with ‘similar pairs,’ and he provides these. He further points out that the categorical dismissal of all words of Spanish origin from the phonological analysis is not required. Loan words which have become adapted

and are integrated into the language so that even monolingual speakers use them are allowable and should be taken into consideration.

Weber claims that the imposition of standards by “small special interest groups” is counter productive to promoting the interests of the Quechua people. He reports that the implementation of the orthography (adopted in 1985) was problematic. Local departments of education wrote their own resolutions in response; resistance was countered with additional resolutions. Finally a law was passed in 1990 that gave the Quechua Language Academy and its regional affiliates the authority over Quechua orthography.<sup>73</sup> Weber sees this as “a positive step because the Academies are more representative of and sensitive to the needs and attitudes of the Quechua people” (Weber 2006:102). Weber believes that by allowing the Quechua people freedom to write according to regional differences, they can “bring their opinions to bear on the emergence of a standard.” He concludes:

Linguistic unification does not necessarily lead to social unity... A pan-dialectal orthography is not more economical than multiple orthographies; it merely shifts the costs from the state to the Quechua people... Cultural and political unification motivates the linguistic standardization, not the other way around, and government regulation plays virtually no role in the process... Quechua people have valid intuitions and opinions about orthography and these should be respected” (Weber 2006:103).

Weber further states that the idea that Hispanic orthography and loan words perpetuate oppression is misguided. Facilitating literacy in Quechua would promote the Quechua language and culture. Using an orthography that is opaque and that people are not motivated to use is likely to have the opposite effect. Since most Quechua speakers

---

<sup>73</sup> The text of this law, Ley No. 25260 Crean la “Academia Mayor de la Lengua Quechua,” is included in an appendix in Weber 2006.

who learn to read and write learn to read in Spanish before Quechua, “making Quechua orthography different from Spanish puts an additional burden on them. To learn to read and write Quechua...requires them to *unlearn* – or otherwise overcome – patterns learned for Spanish” (Weber 2006:119).

In the past, since groups were far from reaching a consensus, matters were brought to a vote. Hornberger (1998:336) comments:

The sequence of meetings and conflicting decisions in the three versus five vowel debate demonstrates just how unsatisfactory voting was as a means of reaching agreement. As in most political debates, it seemed that there were always procedural grounds on which one or another of the interest groups might challenge the validity of decisions taken, and the cycle of new meetings and different votes was potentially endless.

The ultimate measure is this: “To the extent that Quechua peoples can write *in their own Quechua languages*, their cultures will be vindicated” (Weber 2006:119).

This case study demonstrates:

- Orthography design and reform are political in nature.
- It is difficult to remain neutral on orthography issues.
- Opposing ideologies tend to hamper cooperation and the appreciation of the work of others.
- Energies and resources may be wasted when orthography struggles ensue.
- Those who are to be served (the speakers, readers, and writers of the language) may end up suffering the consequences of orthography struggles.

### *11.3.3 English*

English has resisted many reform efforts. Rogers (2005:196–197) lists the objections to reform:

The amount of literature in the current English orthography is enormous. Most people, certainly university students, would have to learn to read both systems for at least 50–75 years; ...the political likelihood of persuading all English speaking countries and publishers to use a single revised system is unlikely. Even if the political will for revision existed, it is not unimaginable that a hodgepodge of new standards would emerge... There is a danger of fragmenting a stable system. At present, there is no viable movement to reform English spelling.

Noah Webster, however, successfully introduced some reforms over 200 years ago. Coulmas (1989:252) credits this success to the timing of the project: “It is no coincidence that this happened at the time of the American independence movement. The waves of the new patriotism went high, and everything that underlined the differences with the colonial motherland had a good chance of being welcomed.” Coulmas also points out that Webster had called for a much more radical reform, one which would make the spelling of English “more systematic and easier to learn” but the more drastic changes did not find favor. Coulmas (1989:254) comments: “American spelling is no significant improvement over the English norm, but it is what Webster wanted it to be: a symbol of a separate national identity. The net benefit is very questionable because there are two standards now instead of one, which hardly simplifies matters.”

Educators are usually in favor of spelling reform when a writing system is difficult to teach. However, for English, reading specialists are not in favor of phonemic spelling. Rogers (2005:196) states: “Evidence from psychology suggests that some of the so-called irregularities of English actually serve to facilitate reading, especially for the experienced

reader.” The general attitude is that the system is working, and so it is better not to change it. Dialect issues would also be a major concern. It would be near impossible to come up with a writing system which might satisfy all speakers of English.

The English case study illustrates the following for well established writing systems:

- Arguments against reform will likely outnumber motivations for reform.
- A nationalistic movement might contribute to the acceptance of a reformed writing system.
- Conservative reforms are more likely to be implemented than radical ones.
- The cost of education and the question of ease of teaching and learning are not usually weighted as heavily in the discussions. It is easier to stick with the status quo.

#### *11.3.4 French*

French spelling has been quite stable over the last four hundred years. Almost half of the words out of a 20,000 corpus found in a dictionary published in 1694 have the same spelling today (Baddeley 1995). The responsibility of establishing spelling norms and dealing with reform issues lies with the *Académie Française*, which also publishes French dictionaries. In 1740 and 1762 drastic changes were introduced, changing the spellings of about half of the words:

(1) <j> and <v>, not previously part of the French alphabet, were introduced to eliminate many homographs resulting from <i> and <u> doing double duty, as vowels and consonants.

(2) Accents, already in use by printers, were adopted by the Academy.

(3) Many silent consonants were dropped.

Some spelling changes proposed over the years did not meet with popular acceptance and were abandoned (Ball 1999:276). A spelling reform proposed in 1990 in response to requests by primary school teachers triggered much public debate. Ball reports:

The Conseil supérieur de la langue française...having obtained approval from the French Academy and from organizations in Belgium and Quebec, published proposals for revised spelling conventions to be introduced into the school curriculum over the following few years.

Considering the anarchic character of French orthography, the changes were modest—even the official name for them was simply *rectifications*...

Nevertheless a vocal anti-reform movement sprang up, including novelists, proof-readers..."

and the list went on. The result was that

the Academy decided to reconsider the matter. Early in 1991, it ruled that, rather than being imposed by Ministry of Education decree, the changes should merely be recommended and the final decision left to 'usage'. Given public conservatism in these matters, this amounted to declaring the reform null and void. Larousse, Robert and other dictionaries incorporated very few of the new spellings (and then only as variants), they have been disregarded by the press and publishers, and even school teachers are ill-informed about them.

Thus about 2,000 words have possible alternate spellings. If the new form is used, it is not considered an error. New spelling guidelines for the *orthographe retouchée* (touched-up orthography) are posted on the web. The slogan *Le français se modernise* is used to instill positive attitudes toward the change (AIROÉ 2006). An important factor is that *Microsoft's* software for French spelling check and thesaurus incorporate the new spelling recommendations. It is likely that this will be instrumental in the recommendations

becoming the norm. The new French orthography is being referred to as *renouvo* by those who wish to see it promoted. An association for the promotion of *renouvo* has been set up with offices in France, Switzerland, Belgium, and Canada.

This case study illustrates:

- Final decisions concerning spelling norms rest with the users.
- Even governmental institutions need to be sensitive to popular opinion.
- Computer technology can be a deciding factor in orthography implementation.

### 11.3.5 German

Coulmas (2000:50–52) reports on a spelling reform attempt for German which turned into a national conflict. More than a decade of research and preparatory committee work involving representation from Austria, Germany, and Switzerland preceded the announcement of “a very moderate reform proposal.” Only 32 of the total number of words learned in the first four years of schooling would be affected. Twenty-eight of these related to simplifying the use of <ß>, i.e., changing over to the digraph <ss> instead. The reform was announced in 1996 and was to be implemented starting in 1998. Public outcry was not a short-lived dilemma: the decision was challenged in the courts. Especially parents objected to the reform. They saw no compelling reason why their children should spell differently than they themselves had been taught. Twelve regional courts had to rule on the matter. “Seven decided in its favor, five against it” (Coulmas 2000:51). The matter had to be referred to the highest court, which allowed the reform to pass. But the matter was far from settled and public dispute continued for years to come. The issue was not about how to spell certain words, but something more encompassing: “Who has the right to alter the spelling rules of German?... Can a reform be enacted by

ministerial decree or does it need parliamentary approval? Who is the master of the written language?” (Coulmas 2000:51). Coulmas notes that previously, German orthographic conventions had “evolved without much official guidance.” Johnson (2005) gives a detailed historical account of German spelling reform issues, expounding on the ideological differences concerning ‘writing.’ One issue is “the assumption that there is only one correct way of writing” (Johnson 2005:7); another issue is that orthography is not usually viewed as a political tool, or gatekeeping device, but as a “technical accomplishment whose primary function is little more than ‘the reduction of speech to writing’ .”

Although many newspapers and magazines switched to the new system, some, including *Der Spiegel*, later reverted to the old spelling system. Some schools had begun teaching the new rules, but abandoned doing so due to the heated debate. Dictionary producers were caught in the middle.

A special council was set up for discussing German orthography issues. This group worked out compromises and produced recommendations which were then adopted in March 2006 (Giersberg 2006). The modified spelling reform will be in effect as of August 1, 2006 (in time for the new school year), after which use of the old system will be considered a “mistake” in schools. A grace period of one year is allotted before students are penalized for such errors. Private citizens and publishers have their choice of spelling system but all government publications and schools will abide by the new (modified) regulations. Newspapers have indicated that they would abide by the modified rules.

The idea that the unique symbol <ß> might be eliminated was an emotional issue for Germans, who have some attachment to it— despite the difficult spelling rules which governed its usage. Perhaps this relates to German identity. (Austrians use <ß> as well; Swiss do not, but use the digraph <ss> instead.) Thus the symbol lives on. The rules have been simplified to make its use more consistent.

This case study illustrates:

- People are attached to the writing systems and symbols to which they are accustomed.
- Various stakeholders (including parents) want a voice.
- Even a moderate reform can incite heated debates.
- It may be wise to decide early on: *Who* is to establish writing norms?
- Setting up an official body to discuss orthography issues, weigh different factors, and give guidance in language planning could be advantageous.
- Top-down decisions are not always welcome.

## CHAPTER 12

### SANGO

#### 12.1 Sango: language of wider communication and of the churches

Sango, at times spelled Sangho, is the lingua franca of the Central African Republic (CAR). The US Department of State (2005) reports the CAR population to be 3.9 million, based on 2004 figures.

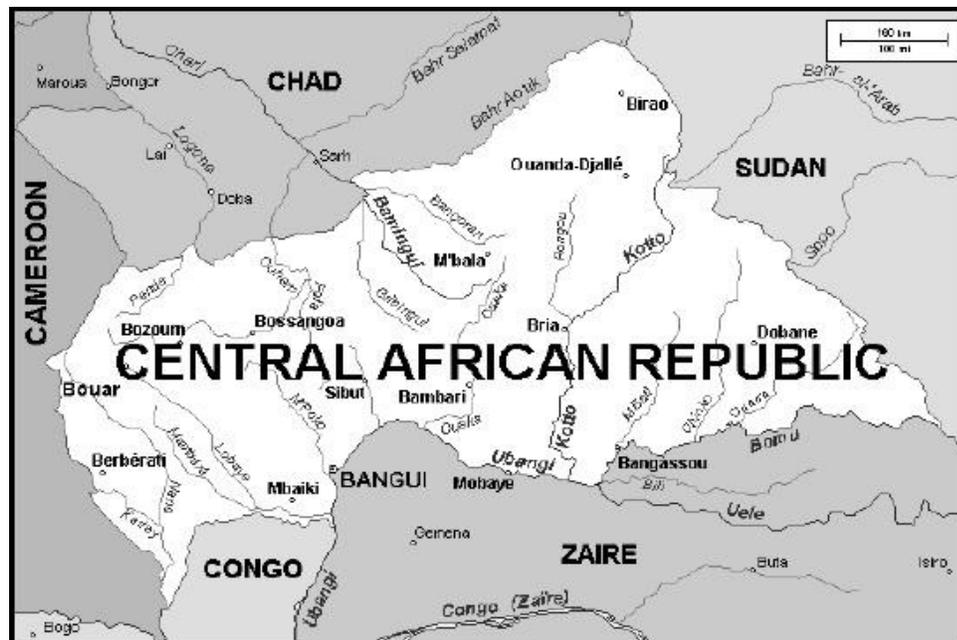


Figure 4. Map of the Central African Republic<sup>74</sup>

<sup>74</sup> Note: Zaire, which borders on the CAR, was renamed the Democratic Republic of Congo in 1997.

The US Department of State (2005) also states that more than 80 ethnic groups are represented in the CAR. The Ethnologue (Gordon 2005) reports that 69 living languages are spoken in the CAR. Le Page (1997:60) comments:

Although they have been described, and used by missionaries from time to time, none of the African vernaculars of the CAR has a sufficiently large population base to appear viable as a language of wider communication within the country. That role is filled in spoken use by Sango, originally a trade or market Central African pidgin which seems to have come into existence in the latter part of the nineteenth century, no doubt partly as a result of increasing urbanization... It was never used by the colonial French government; instead it became associated with the independence movement.

Karan (2001:5–6), based on Samarin (1955:256; 1982) and Kalck (1980:119), reports that one variety of Sango, *Sango riverain*, is said to have been the language of wider communication employed by traders along the Ubangi River before colonization of the area, which began in 1889 when the first European settlement was set up by the French. The region became the territory of Oubangui-Chari five years later, and in 1910 was integrated into French Equatorial Africa.<sup>75</sup> Central African recruits in the French army, in need of a means of inter-ethnic communication, promoted the spread of Sango in the region during the colonial period.

The Roman Catholic Church began work in the region, Oubangui-Chari, in 1894. They first used local languages in their work, but later abandoned that approach, using Sango instead (Koyt 1994:14). Two Protestant missions, the Baptist Mid-Missions and the Grace Brethren, began their work in the CAR in 1921 (SIM 2003). In view of the multiplicity of languages which complicated their work, they opted to use Sango as the

---

<sup>75</sup> Central African Republic, retrieved April 24, 2006 at <http://www.historycentral.com/nationbynation>

church language. Thus, besides being considered ‘soldier speech,’ Sango became associated with Christianity (Karan 2001:5). For a time, the ability to read in Sango was a prerequisite for baptism in one of the protestant denominations. That requirement was later lifted. Various missions began work in the CAR over the years. Some of these used local languages. For example, the Lutherans focused their work on the Gbaya (Northwest) language, which was also spoken across the border in Cameroon. SIM International did some work in Zande, also spoken in Sudan. But they were the exceptions. Swedish Baptists, Swiss Pentecostals, Jehovah Witnesses, Seventh Day Adventists and other arrivals, used Sango as their primary language for oral communication and in literature. Unlike most CAR languages, whose development was neglected, Sango benefited from the labor, funds, and printing presses of the missions. In the 1960s and 70s, evangelism efforts were particularly fruitful with the result that “the CAR has Africa’s highest percentage of Evangelicals” (Johnstone and Mandryk 2001). Church exposed attendees to oral Sango through songs, sermons, announcements, and public reading of translated Scripture portions (as these became available). The Sango Bible—Old and New Testaments combined—was first published in 1966 (Alliance Biblique Universelle). There followed a volume of hymns, *Bia ti Sepala Nzapa* (Presse Biblique Baptiste 1968). Christian instructional material for different ages was made available for teaching and self-instruction for those who could read.

## **12.2 Statistics and the spread of Sango**

The Ethnologue (Gordon 2005) indicates that, according to a 1988 census, Sango was spoken by about 404,000 people as their first language, with 350,000 of these speakers residing in the CAR. It further indicates that another 1.6 million speak Sango as a second

language. These numbers have been rapidly increasing. This is due to acquired bilingualism in Sango as well as a shift to Sango from minority languages, fostered by ethnically mixed marriages and market, school yard and other contact situations. Karan (2001:84) gives an interesting anecdote of a Peace Corps director's observation concerning the amazing increase in Sango oral proficiency among village women over the span of twelve years. Her previous experience had been that she could only converse in Sango with the men in the location of her former assignment.

### **12.3 Classification and vocabulary base**

Karan (2001:5) states that Sango is based on a "reduced form" of a Ubangian language and "thus it can be genetically classified as Eastern (or Ubangian) of Adamawa-Eastern Niger-Congo." The Ethnologue (Gordon 2005), simply states "Creole, Ngbandi based."

There is disagreement as to whether or not Sango should be classified as a creole. With about two generations of mother tongue speakers of Sango, mostly living in towns, it is definitely not a pidgin. It is natural to refer to it as a creole, since a large percentage of its vocabulary are loan words from French, and the remainder from Central African local languages and African languages of wider communication. Although a large percentage of lexical entries may have their roots in French,<sup>76</sup> however, in normal speech this is not as evident as their frequency is much lower than the vocabulary from Ngbandi, a Ubangian language from the Democratic Republic of Congo.<sup>77</sup> Thornell (1995),

---

<sup>76</sup> The Ethnologue (Gordon 2005) states that the lexical similarity with French is 51%. This seems high and needs further investigation.

<sup>77</sup> The Democratic Republic of Congo has had several name changes. It was formerly known as the

discussing fieldwork by Charles Taber and William Samarin, states that the textual frequency of French based lexemes is only 6.8%. She also reports that Samarin found that when restricting research to core vocabulary items on the 100-word Swadesh list, 78% of the words are from Ngbandi. Its geographical proximity to its lexifier language gives cause for not labeling Sango as a typical creole (Thornell 1995).

Central African linguist Diki-Kidiri (2002) objects strongly to the reference of Sango as a creole.

The expansion of sängö did not happen without considerable losses to the original ngbändi vocabulary and partial replacement of indigenous words by several loan words, nor without significant transformation of its syntax. This led to its being classified as ngbändi-based pidgin or creole by William J. Samarin (1967, page 17, §1)...founded neither on a comparative study of linguistic data, nor even on a rigorous theoretical approach... Although now proved to be unfounded, the classification of sängö as a pidgin or creole continues to be represented ad infinitum on the Internet, like a sound wave which takes a long time to fade away.

The ability to speak Sango was useful for communication within the nation, but was not ascribed much prestige. While speaking Sango, the educated elite would frequently code switch and code mix with French. This would mistakenly give the impression that Sango was impoverished and ill-suited for communication. Thornell (1995; 1997) did extensive research on the lexicon of Sango. She studied the decrease in the number of French loan words employed in Biblical texts. There was a general trend—actually a directive, to “purify” Sango of French. The team working on a revision of the New Testament made an effort to comply. Comparing the 1966 Gospel of Mark (Alliance

---

Belgian Congo and also as Zaire.

Bible Universelle) with the 1986 version (Société Biblique), Thornell (1995:5) found that the occurrence of French items was reduced from 15.1% to 1%, i.e., from 115 items to 24 items.

According to Thornell (1995:46), “serious engineering” of Sango vocabulary began about 1970. La Commission Nationale pour l’Étude du Sango (The Commission for Sango Studies) had been set up in 1965. The Institut de Linguistique Appliquée (Institute of Applied Linguistics), or ILA for short, was created in 1975. These two bodies were entrusted with language planning. ILA became responsible for the ‘modernization’ of Sango, i.e., the expansion of terminology, so it could serve in all domains. ILA, like similar agencies in other francophone African nations, worked on the elaboration of a series of lexicons (*Lexiques thématiques de l’Afrique Centrale*, or LETAC). They employed a three-volume French resource designed to facilitate the task (ACCT 1980). Tolerance for French loans or cognates is very low in the thematic *lexiques* published to date. Borrowing from African languages seems more acceptable (Thornell 1997:46). Neologisms (new words) are also being coined, or phrases are chosen to express the desired concepts.

#### 12.4 Sango grammatical structure

Sango is an isolating language. It has few morphological features of interest and a relatively simple grammatical structure. It is an SVO language. The basic noun phrase order is quite similar to French and English, as illustrated by example 1 below.

1.    mbênĭ kêtê mōlengê                    *a small child*  
      mbênĭ    kêtê    mōlengê  
      indef.    small    child

I will limit further linguistic discussions to features which affect Sango orthography.<sup>78</sup> For the noun phrase I will discuss the plural *â-*, the derivational suffix *-ngö*, and compounding. The only important item to consider in the verb phrase is the subject marker *a-*.

#### 12.4.1 *The plural*

The proclitic *â-* ‘plural’ precedes noun phrases. Thus it can attach to the noun, an adjective or an indefinite determiner, as shown in examples 2–4.

- |    |   |                                     |
|----|---|-------------------------------------|
| 2. | <p><i>âzo</i><br/> <i>â-zo</i><br/>         PL-person</p>   | <i>people</i>                       |
| 3. | <p><i>âkötä zo</i><br/> <i>â-kötä zo</i><br/>         PL-big person</p>                               | <i>important people/dignitaries</i> |
| 4. | <p><i>âmbênĩ kêtê môlengê</i><br/> <i>â-mbênĩ kêtê môlengê</i><br/>         PL-indef. small child</p> | <i>some small children</i>          |

Some speakers tend to attach the *â-* to more than one item in the noun phrase, as shown in example 5, but this is not the norm.

- |    |   |                                     |
|----|---|-------------------------------------|
| 5. | <p><i>âkötä âzo</i><br/> <i>â-kötä â-zo</i><br/>         PL-big PL-person</p> | <i>important people/dignitaries</i> |
|----|---|-------------------------------------|

---

<sup>78</sup> I will employ the official Sango orthography as decreed in 1984, unless I am citing from a source which did not use this decree as a reference or if the phonetic or phonemic transcription is needed to make a point. Diacritics denote tone. The circumflex denotes high tone, dieresis mid tone. Low tone is unmarked. The 1984 official orthography is discussed in detail in Section 12.10.

### 12.4.2 The derivational affix -ngö

The suffix *-ngö* changes verbs into nouns. The tone of the verb to which it is affixed becomes uniformly mid, represented by dieresis. This is illustrated in examples 6–10. This tone change provides redundant information, but takes on importance if tone is marked exhaustively in the orthography.

6. kono	<i>to grow, be big</i>	könöngö	<i>size</i>
7. wa	<i>to advise</i>	wängö	<i>advice</i>
8. kîri	<i>to return, repeat</i>	kîrîngö	<i>return</i>
9. hôle	<i>to dry</i>	hölëngö susu	<i>dried fish</i>
10. sî	<i>to arrive</i>	sî sîngö	<i>completely full</i>

The derived forms sometimes serve in a modifier role as illustrated by examples 9–10.

### 12.4.3 Genitive constructions

In genitive constructions the head noun is followed by a prepositional phrase, as illustrated in examples 11–13 below.

11.	këkë tî wâ	<i>firewood</i>
	këkë tî wâ	
	tree of fire	
12.	yângâ tî da	<i>door</i>
	yângâ tî da	
	mouth of house	
13.	dû tî ngû	<i>water hole, well</i>
	dû tî ngû	
	hole of water	

The language, however, is in flux. Written and read by only a small portion of the population, spoken Sango tends to change more quickly than languages where a standard

has been established and maintained through writing (see Augst 1988:1136). There is a trend to drop *tî* from a number of expressions, resulting in an increased number of compounds, such as *yângâda* and *dûngü* for door and well. (Compare with 12 and 13 above.)<sup>79</sup> The components of the compound do not necessarily keep their isolation tone, as observed by comparing *dû tî ngû* with *dûngü* ‘well’. As new compounds are introduced into the language, several questions are raised: should they be written as one word? (At present these are frequently, but not consistently written as single words.) Should hyphenation be used? or Should the components be written separately, reducing the number of long words? Would tone changes be written, or the ‘look’ of the roots be preserved? These are issues lexicographers need to deal with.

#### 12.4.4 The prefix ‘a-’ in the verb phrase

There is a verbal prefix *a-*, which Samarin (1970) refers to as *marque du sujet* ‘subject marker’. It does not encode any feature such as gender, number, animacy or grammatical role. It does not vary in form and is present whenever the subject is a simple noun or noun phrase (14) or an undefined entity (15–16). It is absent when the subject is a pronoun (17), or when the pronoun is left implicit, such as in imperative constructions (18).

14. âmôlengê tî lo agä                      *his children came*  
     â-môlengê tî lo a-gä  
     PL-child    of 3S SM-come

15. adü lo                                      *he was born (literally: someone bore him)*  
     a-dü                      lo  
     SM-give.birth 3S

---

<sup>79</sup> William Samarin has been researching this phenomenon (personal communication).

16. alîngbi *it is enough*  
 a-lîngbi  
 SM-suffice/can
17. lo löndö *he rose*  
 lo Ølöndö  
 3S rise
18. löndö mo Gä *get up and come (here)*  
 Ølöndö mo Øgä  
 rise 2S come

The orthography question that needed to be addressed for the verbal prefix *a-* is whether or not to write it together with the verb.

## 12.5 The phonology of Sango

### 12.5.1 Vowels

Sango has seven oral and five nasal vowels as follows:

Table 16. Sango vowels

	Oral vowels		Nasal vowels	
	<i>Front</i>	<i>Back</i>	<i>Front</i>	<i>Back</i>
<i>High</i>	i	u	ĩ	ũ
	e	o		
	ɛ	ɔ	ẽ	õ
<i>Low</i>	a		ã	

### 12.5.2 Consonants

The 26 consonantal phonemes represented in the Sango official orthography are written as they appear in Table 17, except /j/ and /nj/ are written <y> and <ny>.

Additional consonants (palatal affricates) appear in loan words and in certain dialects.

Dialect differences are evident in variation in the prenasalized consonants ([mv]~[v]; [mb]~[ngb]; [mb]~[b]), affrication, free variation of [l] and [r] word medially, and of [h] and [ʔ] word initially. [mv] occurs in only a few words. Samarin (1970) doubts that it is a phoneme.

Table 17. Sango consonants

	<i>Bilabial</i>	<i>Labio-dental</i>	<i>Dental</i>	<i>Alveolar</i>	<i>Palatal</i>	<i>Velar</i>	<i>Labio-velar</i>
Obstruents:	p b		t d			k g	kp gb
Fricatives:		f v		s z			
Prenasalized	mb	mv	nd	nz	nj	ng	ngb
Nasals	m		n				
Approximants				l r	j	h	w

### 12.5.3 Syllable structure

Sango syllable structure is almost exclusively CV. Labialization and palatalization of certain consonants occur. Consecutive vowels are rare, but do occur. For those who don't know Sango, The convention of symbolizing labialization with <u> instead of <w> and palatalization with <Ci> instead of <Cy> creates the illusion that the CVV syllable structure is more frequent than it is.

Most words are mono- or bisyllabic. Trisyllabic words are less frequent. Four-syllable words are the result of reduplication (19) or compounding (20–21). If writing conventions for reduplicated and compound forms lean toward 'splitting' rather than 'joining,' long words are practically eliminated.

	<i>joining option</i>	<i>splitting option</i>	<i>translation</i>
19.	kêtêkêtê	kêtê kêtê kêtê kêtê small small	<i>tiny; tiny bit</i>

20.	dambëti	da mbëti	<i>school</i>
		da mbëti	
		house book/paper	
21.	walikundû	wa likundû	<i>sorcerer</i>
		wa likundû	
		person sorcery	

#### 12.5.4 Tone

Sango is a tonal language with three level tones: high, mid and low. Contour tones are infrequent, except in French loan words. In single syllable words of French origin the pattern is high-low falling (22–23); in longer words of French origin, all but the last syllable carries a low tone. The vowel in the last syllable tends to have extra length and descending tone. Final syllables of nouns most frequently bear mid-low falling tone while verbs bear high-low falling tone. (See 24–29 below.) There are some exceptions to these patterns.

	Sango (official 1984 orthography)	French cognate	English	
22.	bâan	H-L falling	banc	<i>bench</i>
23.	bôn	H-L falling	bon	<i>IOU, debt</i>
24.	balëe	L M-L falling	balai	<i>broom</i>
25.	motôo	L M-L falling	moto	<i>motorcycle</i>
26.	ananäa	L L M-L falling	ananas	<i>pineapple</i>
27.	votêe	L H-L falling	voter	<i>to vote, elect</i>
28.	mankêe	L H-L falling	manquer	<i>to miss</i>
29.	aretêe	L L H-L falling	arrêter	<i>to stop</i>

Tone has a low functional load, functioning mainly in the lexicon. A search in the Bouquiaux (et al. 1978) dictionary would turn up a surprising number of minimal tone pairs or sets. But many of these include obscure words. Among the more common words,

the members of minimal pairs/sets often belong to different grammatical categories, helping discriminate meaning, for example: <dü> ‘give birth’ and <dû> ‘hole’. Other contextual clues help disambiguate as well. For instance, <dü> is a transitive verb. Common expressions are <dü pärä> ‘lay an egg’ and <dü molênge> ‘give birth to a child’.

Tone patterns of words in isolation can vary slightly from person to person. This was evident when in the final editing stages of the SIL orthographic lexicon (1995), some tones were still debated by individual speakers, including linguists from the University of Bangui Applied Linguistics Institute (ILA).

The mother tongue of someone who speaks Sango as a second language may have quite a bit of influence on that person’s idiolect. It may affect tone, vowel quality, number of nasalized vowels, as well as consonants occurring in free variation.<sup>80</sup>

Loan word pronunciations vary quite a bit. Thornell (1995:3) links the variation to levels of education: “Phonological integration does not seem to be the norm for speakers proficient in French...pronunciation of French incorporations tends to be a function of education, in that the speakers of little formal education had the highest rate of phonological adaptation to Sango.” The question then is: whose speech should the written standard represent?

---

<sup>80</sup> When I was first learning Sango in 1990, a person correctly identified the ethnic group of my tutor by the way I was speaking.

## **12.6 Literacy rates**

There is a disparity between various reports on literacy rates in the CAR. The UNESCO Institute for Statistics (2006) list a 48.6% literacy rate for adults (15 and older), with a significant difference between the literacy rate among men (64.8%) and women (33.5%). The net enrollement rate in primary school is 53.3% for girls and 77.9% for boys (UNESCO 2006).

## **12.7 Sango literature**

Concerning written materials available in Sango, Le Page (1997:60) pointed out that, at the time of his research, about 1992, “apart from religious literature, however, there is very little written in it.” Fasold (1997:257), based on research by others, states:

Sango remains used mostly for spoken communication. Little reading is done in this language, and very little writing. All formal education is in French. The few experimental classes in Sango that were launched in the late 1970s have had no impact on the course given to education. Today the written production in Sango continues to be limited to religious materials. A few informal education booklets in Sango are also produced by Christian missions. A newspaper in Sango was created in 1974 and disappeared in 1981.

Fasold seems to interpret the fact that circulation of the Sango newspaper ended as an indication of a lack of interest in reading Sango, commenting, “The current (1994) daily newspaper is printed entirely in French” (Fasold 1997:257). But in the late 1990s, the French newspaper also ceased to exist. No one interprets this as indicating a lack of interest in reading French! Many enterprises failed due to financial mismanagement. Others failed because potential clientele was in such a crisis that they spent no money on anything other than the bare necessities. Gerbault (1994:75) indicates that the Sango

newspaper, *Linga*, was launched in 1976 with the help of UNESCO and “published for almost five years, selling at about 5,000 copies.” The project was probably intended to become self supporting, and may not have achieved that goal. On May 4, 2006, a bilingual government-sponsored weekly newspaper *Sêse tî mo* was launched (Agence Centrafrique Presse 2006).<sup>81</sup>

Experimental classes using Sango as the language of instruction in school while teaching French as a foreign language were just that: experiments that ended. (See 12.8.3.) Le Page (1997:58) expressed it well: “Experimental work runs out of steam for the usual reasons: lack of adequate preparation and training; lack of sustained support; parental worries about children missing out on French-medium examinations and the passport to a non-menial job.” To the chagrin of those who hoped that Sango would be reintroduced into the schools, much of the materials used in the 70s, which might have been able to serve again with minor revisions, were lost during political upheavals in the early 90s.

Despite various setbacks, the amount of literature available in Sango has greatly increased since 1990—with and without religious content, as discussed in the next few sections.

### *12.7.1 Expatriate and national efforts*

Some of the missions had linguist members on site. Dr. William Samarin prepared a Grammar of Sango in English (1963); the French equivalent was published in 1970. He oversaw the production of a Sango-English/English-Sango dictionary (Taber 1965), out of which Taber’s research on French loan words in Sango grew. Samarin continued his

---

<sup>81</sup> *Sêse tî mo* means ‘your land.’ The paper is sold at 400 CFA (about 80 cents).

research in Sango over the years, with a particular focus on how the language is changing. (See Samarin and Walker 1994.)

Sango did not have a written standard, but was written in a variety of ways. UNESCO recruited linguist Luc Bouquiaux to help the National Commission of Sango Language Studies take on the challenge of coming up with a written standard. These efforts resulted in a comprehensive dictionary (Bouquiaux et al. 1978), but the orthography used in the book never served as a standard (Thornell 1994:36).

Various sets of basic literacy materials have been available. I have personally looked at primers developed by ten different agencies. The missions were most prolific in the production of Sango literature. Four had their own printing facilities. Post-literacy material they produced included booklets on health issues, agriculture, and civics.

The Peace Corps published a comic book, *Mabôko na Mabôko*,<sup>82</sup> which explained their development activities in the CAR (Debold 1986). US AID included written Sango instructions with the heavily subsidized condoms promoted in the fight against AIDS (1992–1994). Health agencies designed survey questionnaires and community health materials in Sango. An impressive set of agricultural books were produced by ILA researchers<sup>83</sup> for an international agency which sponsored a functional literacy and development project in the small northern town of Bouca. (Diki-Kidiri 1998:238 lists eight titles for 1992–1993.) Unfortunately the material did not have a wide distribution,

---

<sup>82</sup> *Mabôko na mabôko*, literally ‘hand in hand,’ means ‘together’ —in activity not just in presence. The phrase is used to express cooperation.

<sup>83</sup> ILA functions as the department of linguistics at l’Université de Bangui, and is the agency responsible for research on Central African languages.

serving only the population participating in the project during the years the project was funded.

SIL International, in the CAR since 1990, developed basic literacy material and a variety of motivational and instructive post-literacy materials. After 2000, a partnership agency, ACATBA,<sup>84</sup> continued literacy and post-literacy efforts in Sango using both SIL materials and their own publications.

The Société Biblique en Centrafrique, the national chapter of the United Bible Society, has repeatedly revised and published Sango Bible portions and extracts. The revision of the Old Testament is in progress.

#### *12.7.2 The Institute of Applied Linguistics*

The Institute of Applied Linguistics (ILA) was created as a department of the University of Bangui in 1975. It usually has about fifteen researchers serving in the following domains: sociolinguistics, research and description, terminology, language planning, literacy, instructional applications, and the development and promotion of Sango. As a research agency, it does not normally produce Sango literature, but rather focuses on the development of Sango terminology and the promotion of the official orthography. In that vein, it has produced reference documents and a series of thematic lexicons. These cover a variety of domains such as economics, health, education, linguistics, animal husbandry, the judicial system, and information technology. Central African linguist Marcel Diki-Kidiri, who resides in France, has served as consultant to ILA on some of the publications. In addition he has produced a spelling dictionary (1998)

---

<sup>84</sup> ACATBA stands for l'Association Centrafricaine pour la Traduction de la Bible et l'Alphabétisation.

and a book which urges the population to be involved in civic life (1982). The majority of his publications are in French. Christian Dagnan has written the first Sango novel (1998).

## **12.8 The socio-political situation and the status of Sango**

### *12.8.1 Inherited attitudes*

To understand the language and literacy situation in the CAR, one must have an understanding of colonial and post-colonial Africa—francophone Africa in particular. Consider the following comment from Haugen (1985:11):

The English generally made it a policy to introduce vernacular education at the lower levels and gradually introduce English as the student progressed. The French have almost everywhere disregarded local vernaculars entirely and given all education in French... Or to put it more bluntly: the English were tolerant of native tongues but unwilling to accept their speakers as social equals. The French were willing to receive natives of all colors into the French community provided they gave up their identity and learned French.

French officials expressed disdain for all languages other than French. This was given expression through their 'French only' educational policies and in extreme cases, the destruction of literature where local languages flourished. (See 6.3.3 on Bamum.) Crystal (2000:55) comments, "Early colonial contempt for subjugated peoples was automatically transferred to their languages" contrasting it with healthy bilingualism: "a state in which two languages are seen as complementary, not in competition—fulfilling different roles, with each language being seen in a rewarding light" (Crystal 2000:81). Central African languages did not enjoy such a state under French colonial rule, nor since. Local languages have been seen as primitive and worthless. Sango did not fare much better, but

was tolerated and promoted because it (1) served a utilitarian role for communication and (2) was considered a symbol of national unity.

### *12.8.2 The political history of Sango*

The Central African Republic was a French colony, known first as Oubangui-Chari, and later as French Equatorial Africa. It gained independence in 1960. In 1964, through a constitutional law, French was made the official language and Sango the national language. In 1965 “a commission to study Sango (Commission Nationale pour l’Étude de la Langue Sango) was established under the Minister of Education (Ministère de l’Education Nationale). In 1974, a national pedagogical institute was created. This institute was given the task of preparing the introduction of Sango into the formal and non-formal education systems” (Koyt 1995:6 as cited in Karan 2001:6–7 ).

Gerbault (1994:72) states that “there have been several attempts at promoting the use of Sango in writing and especially in formal education. However, there have only been so far limited experiments in education in Sango.” Sango was in fact introduced into the schools for a few years in the late 70s, but only in 1984 was it declared to be an acceptable language of the schools alongside of French (Fasold 1997: 257). Gerbault (1994:75) reports that “in spite of the official policy stating that Sango was to be introduced in formal education, the educational system continues to function entirely in French”. In the capital, the Department of Non-Formal Education taught literacy to adults in French; Sango literacy instruction in Sango was restricted to the rural areas.

The official orthography was established in 1984, and efforts to promote the new code began. In 1991, a constitutional change elevated Sango to the status of *official language*

alongside of French (Koyt 1995:7). From then on, the term *national language* was applied to all languages native to the CAR.

But attitudes do not change overnight: In 1992 SIL was looking for distribution outlets for Sango story books they had developed. In view of the difficult national economic situation, I, as SIL literacy coordinator at the time, asked a bookshop in downtown Bangui to take some on consignment. A shopper was present while I was trying to come to an agreement with the shop manager. In my naïve enthusiasm I showed the books to the customer, expecting a positive response. His “Sango is for illiterates!” was a blow to my positive expectations for all efforts in Sango. Crystal (2000:147) comments: “Linguists...often remark on how emotionally stressful this sociopolitical context can be. Traditionally, there is nothing in a linguist’s training which prepares for it.”

### *12.8.3 Sango in the school system*

As indicated in 12.8.2, a pilot program, employing Sango as the medium of instruction for the first few grades with about 120 classes, began in 1976. It was abandoned in 1980 (Koyt 1994:21; Diki-Kidiri 1994:28). Karan (2001:7) wrote: “The program failed, and for years acted as a deterrent to primary and/or bilingual education in Sango.”

Although it was a failure in many ways, academically it had been a success. Children learned to read quickly and enjoyed learning in general. Why was it labeled a failure? First of all, there was no long term financial support for the program. Experimental programs funded from the outside rarely survive once the funding ends. Secondly, the attitude that only French is worthy to be used in education and for literary purposes had

been absorbed by the general population. Teachers were not convinced that using Sango was a good idea. Most of all, parents objected. “Many parents were reluctant to allow their children to enter or continue in such classes because they felt that by being in such a bilingual classroom their children would be disfavored and disadvantaged in learning French, and thus be disadvantaged and disfavored in life” (Karan 2001:7).

Basically, parents had not been sensitized to the benefits of education in ‘Sango first.’ Crystal (2000:136) highlights some of these:

The role of a school in developing a child’s use of its mother-tongue is now well understood, following several decades of research and debate in educational linguistics... The school setting provides an increasingly widening range of opportunities for children to listen and speak, as they learn to cope with the demands of the curriculum... It gives the opportunity to engage with literacy...which will open the doors to new worlds... [I]f careful planning has managed to give the indigenous language a formal place alongside the dominant language, the result can be a huge increase in the pupils’ self-confidence.

In the late 90s there were hopes that Sango would again be given its place in the schools. In 1997, ILA personnel and SIL members served on a planning committee and participated in a UNESCO-financed workshop preparing for such an eventuality.<sup>85</sup>

A guest speaker came from Rwanda. Papers were presented and there was lively debate. Working groups established learning goals for beginning grades. Recommendations for action were drawn up and read at the formal closing ceremony, but there was no follow-up.

---

<sup>85</sup> This four-day workshop was called *Atelier Sous-Régional d’Élaboration du Plan d’Introduction du Sango dans l’Enseignement et de Prototypes de Matériels Didactiques*.

#### *12.8.4 Motivation for literacy*

Carrington (1997:82) comments: “The status of national or official language does not necessarily guarantee that the relevant vernacular will become an accepted medium for literate activity.” This can be due to lack of official action, but it may also be due to personal choice. When there are no negative social consequences for being a non-reader, one cannot expect a person to make the effort, financial sacrifice, and time investment to become literate. Without extrinsic motivation, all depends on the individual’s intrinsic motivation. When an economic crisis and unemployment perturb a nation, education in *any* language is devalued, whether official, national, or vernacular, and yes, even in French—a respected international language. Fasold (1997:258–60) paints a dismal picture of Sango in the CAR, regarding reading and writing in general:

There is no need for the written media in traditional Central African modes of communication... [T]he use of literacy is still reserved to a fraction of the population... In Bangui and other cities, access to employment is not given by literacy in Sango—nor does any amount of school education, incidentally, guarantee employment; small jobs that can be secured often do not require reading or writing skills. Employment as a civil servant requires only oral competence in Sango, along with competence in spoken and written French... Sango is used nationwide in specific domains, but these domains are not those that have long been associated with literacy and/or formal education. The conceptual association is between writing and French, not between writing and Sango... In the past twenty years, there have been in the CAR, as in other countries, meetings, decisions, and committees created, aiming at developing and standardizing Sango, at developing literacy in Sango, and at introducing the language in formal education. Programmes have collapsed because there has been no continued commitment to expressed objectives. When actual action was called for, no effort was made towards reaching the goal... There is an obvious lack of determination on the part of project directors; individual initiatives are practically absent... [T]here is no organized pressure

for change from local linguists and educators, or other educated persons, upon the political authority...

Another side of this political coin is that, in terms of funds for education, the CAR, like most other francophone countries of Africa, depends largely on the French government or other francophone funding agencies. Promoting *francophonie* has not at any time meant promoting national languages.

I would not personally be quite as negative as Fasold, especially concerning the efforts that have been made. Workers involved in language development have been very committed, despite lack of outside support and the irregular payment of salaries. Frequent shuffling of personnel at various levels, including the Ministry of Education, has interfered with continuity in various undertakings.

Motivation for learning to read Sango is not altogether absent. It stems from two main desires:

1. self-improvement, accompanied by a hope for an increase in self-esteem
2. the desire to read the Sango Bible, both for personal enrichment and to better serve in the local church.

Adults do, in fact, sign up for Sango literacy classes when the opportunity arises. However, various factors frequently prevent facilitators and learners from seeing the program through to the end: unrest in the country, illness, economic activities, etc. Between 1992 and 2004, more than 2000 adults completed literacy classes under the supervision of SIL International and ACATBA.<sup>86</sup> Other agencies (religious, governmental and non-governmental) had similar programs. Combined results on the national level

---

<sup>86</sup> Statistics are based on SIL International Africa Area Literacy Reports. These literacy classes enrolled men and women who did not know how to read or write in any language, and equipped them to read independently. Facilitators were trained, but not paid. Approximately 80% of the learners were women.

could either be viewed as a mere ‘drop in the bucket,’ or as evidence that reading and writing Sango is valued after all, at least in the Christian community. This community is a substantial part of the population. The US Department of State (2005) reports 25% of the population to be Catholic and 25% to be Protestant. Atlapedia Online (1993–2005) cites higher percentages: 33% Roman Catholic and 50% Protestant.

In addition, there seems to be a new wind blowing in *La Francophonie*: In 1988, “of the 15 former French colonies, only four were using one or more African languages in their primary education” (Albaugh 2005). Since 1995, twelve countries have changed their educational policy and begun or expanded experiments in mother tongue education. Albaugh, researching this shift, indicates that this is not a matter of activists from within Africa finally having gotten an audience, but that it is actually the result of “an altered strategy in the government of its former colonizer.” Reports published between 1989 and 2000 painted a dismal picture of the impact of French education in Africa. The result: “the rhetoric of la Francophonie in favor of African languages has become stronger, and actions have begun to follow” (Albaugh 2005). The motivation does not appear to be people’s linguistic rights to use their own languages, but rather “the right of all to French.” Commitment to changing to a different educational strategy is backed up by funding. Albaugh comments: “What may appear an independent decision on the part of African governments to implement a program that includes local languages in education, is likely an anticipation of significant support from la Francophonie and from France.” But not all languages get help, only vehicular languages. Sango would qualify.

### *12.8.5 Efforts paying off*

Pasch (1994:85), who studied the use of Sango and French in formal and informal situations, comments that “Sango has not been able to fulfill the same roles as French... the domains where the two languages are used are in almost complementary distribution.” Yet Sango has indeed gained some respect over time. Since it was made the *national* official language in 1991, an increasing number of formal speeches were made at official events in Sango rather than in French. Even when international dignitaries have been present, speeches made in Sango have not necessarily been interpreted for the benefit of those who did not know the language.

Time given to Sango programming on TV and the radio increased. Fasold (1997:258) reports that 65% of radio programming is in Sango. He adds: “The national radio has been the only official body with an internal language policy (training of translators, systematic use of French-free Sango, more uniform pronunciation, etc.) in agreement with the guidelines established by the higher political authorities.” Since speeches and radio programs are not normally impromptu, one presumes that Sango programs were written before being aired.

Fasold (1997:259) also gives an encouraging glimpse of efforts by the Department of Agriculture: “Efforts are being made in two directions: developing Sango terms for the dissemination of oral information, and teaching reading to those involved in agriculture, enabling them to read technical information in order to improve their practical skills and production.”

The First International Conference on Sango was held in Cologne, Germany, in September 1992. The focus was on language development and standardization. There, “Diki-Kidiri emphasized that, while Sango will in time take over occupational functions,

that will not be possible until the lexicon has been expanded with the terminologies for modern occupations and standardized” (Le Page 1997:61).

ILA has been rising to the challenge of expanding the Sango vocabulary for various specialized domains. Marcel Diki-Kidiri has been developing specialized Sango vocabulary in the information technology domain and maintains a Sango website, an expression of his personal commitment to making Sango one of the languages *used* (not just talked about) on the Internet. He objects to languages only serving as ‘language objects’ in a “virtual museum intended to inform future generations about what the languages of humanity were like” (2002).

These efforts are laudable. However, unless Sango is introduced into the schools and some mass literacy campaigns are organized, the status quo is likely to continue: The use of French will dominate except in the religious domain, and Sango will continue to be used mostly for oral communication.

## **12.9 The Sango orthography before 1984**

The first written literature in Sango was the result of Christian missions activity in the capital and outlying areas. Since the administrative and educational language of the colony was French, Sango was written with French symbols and conventions as much as possible. This had several advantages:

- easy transfer between the two languages, i.e., a person, bilingual in French and Sango who had learned to read in one of these, could—without extra effort—also be able to read the other
- uniformity of spelling for place names and loan words

- a ‘proper’ look for Sango: i.e., acceptable to the French colonial administration, and motivational for literacy in Sango
- avoidance of special characters that might cause technical difficulties.

The Catholic and the Protestant orthography practices differed slightly. Diki-Kidiri (2005) states that one of the missions producing written Sango early on, Baptist Mid-Mission (BMM), had an internal set of orthographic conventions by 1920.<sup>87</sup> This mission and the Grace Brethren Mission cooperated in a project with the United Bible Societies to translate the Bible; this resulted in the publication of the Sango Bible, *Mbeti ti Nzapa* (Alliance Biblique Universelle 1966). This version, and the hymnal, *Bia ti Sepala Nzapa* (Presse Biblique 1968), found in many homes and used in churches today, were (and still are) the most influential materials to promote these initial writing conventions of Sango. While the orthography was highly functional, there were some deficiencies. These will be presented in the following subsections.

### *12.9.1 Problems related to dialects*

There were some dialectal issues with the orthography used in religious materials:

1. Although Sango has two phonemic liquids /r/ and /l/, in some dialects free variation occurs between them in word medial position in a restricted number of words. The symbol <l> was chosen for writing these words, as illustrated by examples (30) and (31).

---

<sup>87</sup> According to their website, the history of the BMM work in the CAR began January 1, 1921, when six of their missionaries traveled inland to set up three different mission stations (Baptist Mid-Missions 2006).

	in hymnal	pronunciation variation	translation
30.	voulou	[vùlù]~[vùrù]	<i>white</i>
31.	télé	[tèlé]~[tèré]	<i>body</i>

2. Vowels vary in quite a few common words, as illustrated by (32–34). One vowel was chosen over the other in writing these words, possibly based on the variety spoken by the literature production team.

	in hymnal	pronunciation variation	translation
32.	sepala	[sèpàlà]~[sèpèlà]	<i>praise, honor</i>
33.	tambela	[tàmbéla]~[tàmbúla]	<i>walk</i>
34.	mbito	[mbìtò]~[mbètò]	<i>fear</i>

Although church documents provided a standard for writing Sango, they did not influence spoken Sango. Regional variation between [r] and [l] and different vowels persist.

### 12.9.2 Underdifferentiation of /o/ and /ɔ/

Sango has open as well as closed *e* and *o*. Vowel harmony assimilation rules preclude the co-occurrence of closed and open vowels in the same word, i.e., [e] can co-occur with [e] but not with [ɛ] and vice-versa; [o] can co-occur with [o] but not with [ɔ] and vice-versa; [e] and [o] can occur in the same word as can [ɛ] with [ɔ], but not [ɛ] with [o] nor [e] with [ɔ]. The first Sango publications underdifferentiated between [o] and [ɔ], writing both with <o>. There was one exception: [sɔ̃] ‘to save’ which was written <sau>. This permitted graphic similarity with the French loan *Sauveur* ‘Savior’ and differentiated between /sɔ̃/ ‘to save’ and the determiner /só/ ‘this/that’, and thus

eliminating the potential homograph <so> which would have resulted since tone was unmarked.

### 12.9.3 Representation of /e/ and /ɛ/

The symbolization of [e] and [ɛ] was problematic. In French, <é> (acute accent) represents /e/, while <è> (grave accent) represents /ɛ/. In Sango protestant literature, usage of acute and grave accents was haphazard. Phoneme-symbol correspondence was unreliable. This did not pose a problem for reading, since readers in general have a propensity for ignoring diacritics, but it did complicate writing in a consistent manner. Two-syllable words containing two <e> 's, whether open or closed, were written with a grave accent on the first and an acute accent on the second, as illustrated by (35–38).

	orthography in 1966 Bible	pronunciation (accents indicate tone)	translation
35.	tèné	[tènè] or [tēnē]	<i>say; word or speech</i>
36.	pèpé	[pēpē]	<i>not</i> (negation)
37.	ndèké	[ndèkè]	<i>bird</i>
38.	pèlé	[pélē] or [pérē]	<i>grass</i>

Since vowel harmony dictated that all <e> 's in a word were obligatorily identical, the accents served a purely decorative purpose, affecting a French look. For words with a single open or closed <e>, sometimes the accents corresponded to French rules, sometimes not. This can be seen by comparing (32–33), repeated here, in which <e> bears no accent and represents /e/, with (39–42), which highlight some inconsistencies in representation.

	<i>orthography in 1966 Bible</i>	<i>pronunciation (accents indicate tone)</i>	<i>translation</i>
(32.)	sepala	[sèpàlà]	<i>praise, honor</i>
(33.)	tambela	[tàmbélà]	<i>walk</i>
39.	fade	[fàdē]	<i>soon, just (aspect)</i>
40.	sessé	[sésè]	<i>earth, ground</i>
41.	kouè	[kwè]	<i>all</i>
42.	pèko	[pèkó]	<i>back</i>

#### 12.9.4 Other vowels and /w/

The vowels [a] and [i] were represented by <a> and <i> respectively; they did not bear diacritics as they occasionally do in French.

Faithful to the French spelling system, Sango [u] was written <ou>. In French, simple <u> symbolizes the high front rounded vowel [y]. The word *Alléluia*, with /u/ written as <u>, not <ou>, was the exception in both French and Sango.

In French, the letter *u* is also employed as a silent part of the digraph <gu> before <e> and <i> to represent /ge/ and /gi/, as in *guerre* [gɛr] ‘war’ and *guillotine* [gijɔtin]. Before other vowels a simple <g> is written for /g/, as in *garage* [garaʒ] ‘garage’ and *gorille* [gɔrij] ‘gorilla’. In the absence of the <u>, <g> before <i> or <e> is pronounced [ʒ], as in *garage*, *général* and *girafe*. This practice of representing /g/ in two manners was adopted for Sango, as illustrated by

- 43. *ga* [gà] ‘come’
- 44. *go* [gò] ‘throat, voice’
- 45. *gui* [gì] ‘search’
- 46. *lègué* [légē] ‘path, way’.

The consonant <w> is rare in French, occurring only in a few loan words such as *whisky, weekend, talkie-walkie* etc. In proper nouns such as place names, the French traditionally used <ou> to represent /w/ as, for example, in *Quagadougou* /wagadugu/, the capital of Burkina Faso. In Sango, /w/ is not as infrequent as in French. Thus the digraph <ou> used for both /u/ and /w/ lengthened quite a few words, and made some of them appear indomitable like *ououssouououssou* ‘disorder’.

#### 12.9.5 Tone

Tone was not marked. A handful of minimal pairs were discriminated through the use of an apostrophe following the vowel or a circumflex over the vowel.

#### 12.9.6 Additional conventions

The plural marker *a-* (tone not indicated) was consistently attached to the following word. If a word began with a vowel—usually a French loan word—a hyphen was used, for example: *a-ancien* (elders) and *a-église* (churches). When *a-* was attached to a proper noun, upper case was preserved for the stem, but not applied to the affix: *aGentil* (Gentiles) and *aCorinthien* (the Corinthians).

This ‘joining’ convention was not applied to the subject marker *a-*. If the subject was unspecified, and in passive constructions, *a-* was written as a separate word. If the ‘actor’ of an event was God, the subject marker was capitalized. To understand the rule as to when the subject marker needed to be written separately required solid training in grammar (see 47–49); knowledge as to when to use upper case required theological insight (see 50). I will highlight the subject markers in bold.<sup>88</sup>

---

<sup>88</sup> These four examples are excerpts from Mbeti ti Nzapa (Alliance Biblique Universelle 1966). The references are as follows: #52. Proverbs 23:27; #53. Proverbs 23:30; #54. Proverbs 23:31. #55. John 3:7

47. ouali ti pitan ayèké caniveau so ali *a prostitute is like a deep pit*  
 prostitute SM.be ditch which SM.deep
48. ti tara vin so a bian ya ni *to taste mixed wine*  
 to taste wine which SM change in it (to taste wine someone mixed)
49. moba vin pèpé tonga na a yèké bingba, *don't look at wine when it is*  
 2S lookwine NEG when SM be red *red and sparkles in the cup*  
 tonga na a za na ya kopo  
 when SM shine in cup
50. a lingbi A kiri A dou i mbèni *it is necessary that once more*  
 SM must SM return SM give.birth 2PL again *someone (God) give birth to*  
*you. (You must be born again.)*

The 1966 Bible and 1968 hymnal used words of French origin such as *merci* ‘thanks’, *lumière* ‘light’, *commandement* ‘commandment’, *gloire* ‘glory’, and *liberté* ‘liberty’, among others, proper nouns like *Calvaire* and *Nazareth*, and personal names like *Jésus*, *Moïse*, and *Abraham*. All these were written exactly as they are written in French.

The use of uppercase for nouns with special religious significance tended to be more generous than in French practices: All titles, pronouns and agreement markers referring to God or Jesus, or having a close association with these, were capitalized: *Berger* (shepherd), *Sauveur* (savior), and *Ndia* (law), *Temple ti Nzapa* (temple of God), *Melengué* (child; i.e., the infant and 12 year-old Jesus) and *Lo* (he), among others.

### 12.10 The 1984 orthography decree

With the plan to introduce Sango into the schools, it was important to have a written standard. In Africa, there was a growing reaction against spelling as in colonial languages. There was a movement to instead work toward more uniform conventions

across African languages. UNESCO-sponsored meetings of African linguists had taken place in Bamako in 1966, in Cotonou in 1975, and in Niamey in 1978. The objective: the preparation of an African Reference Alphabet. With relationships strained between the CAR and France, local linguists made plans to join this pan-African movement and create a new writing system for Sango. An independent identity would be gained, but at the expense of easy transition between Sango and French. Expatriate producers of literature were not consulted; they were only informed after the fact.

Initial efforts to come up with a standardized orthography included a proposal by Diki-Kidiri in 1977: All of the Sango vowels were to be represented and tone was to be marked. Phonetic symbols were used to note the open vowels. These proposed conventions did not find favor. In Bouquiaux's (et al. 1978) dictionary which followed shortly after, tone was not marked in the orthography, only in the phonetic transcriptions following the entries. All vowels were represented: not with linguistic symbols, but with French conventions instead of linguistic symbols: /e/ as <é>; /ɛ/ as <è>; /o/ as <ô> and /ɔ/ as <o>. I suspect that the reason this orthography was never implemented either was that it was 'too French' and not African enough. ILA linguists returned to the drawing board. The orthography they wanted to promote was ready in 1983.

#### *12.10.1 Content of 'Décret N° 84.025'*

On January 28, 1984, President André Kolingba signed Decree 84.025, which declared that the official orthography would be, from then on, the one described in the accompanying document *Code de l'orthographe officielle du sango*. That document had been prepared by a team of linguists associated with ILA. The decree mentions as its basis:

- the CAR constitution (1981) and decree 65.022 (1965), which had established the special commission for studying Sango
- recommendations made by the *Séminaire National sur la Réforme du système éducatif de la République Centrafricaine*
- a report of the Ministry of Education, subscribed to by the Council of Ministers.

The document *Code de l'orthographe officielle du sango* was written up in six parts.

Part one presented the Sango alphabet consisting of the following 22 letters:

A B D E F G H I K L M N O P R S T U V W Y Z

This was followed by the comment that letters C, J, Q, and X may be used in proper names and in foreign language quotations, but were not an integral part of the Sango alphabet. Part one also included a name for each letter.

Part two consisted of a chart of consonant symbols, distributed according to their points and manners of articulation. For example, at the intersection of *Palatales* (palatals) and *Continues* (continuants) one found *y*; at the intersection of *Sonores* (voiced) and *Vélaires Glottale* (velar glottals) one found *g*. (No sample words or text were provided). Alternate letters were provided for the spelling of a few loan words: < `b > in addition to < b >, < ty > in addition to < s >, and < dy > in addition to < z >. This provision was intended for loans which had not yet fully integrated into the phonological system of Sango. (Again, no examples were offered.)

Part three addressed the writing of vowels. As in part two, linguistic terms were employed to describe the vowels in terms of position and *degré d'aperture* (relative height). Nasal vowels were included; these were to be written by *n* following the vowel, as in French.

Part four dealt with tone. It stated how the three level tones should be written: low tone was simply V, without diacritics; mid tone with dieresis over the vowel as in <ä>, <ë>, etc.; and high tone with circumflex over the vowel, <â>, <ê> etc. In cursive, writing high tone with acute accent was presented as an option. The three contour tones were to be written as composite tones by doubling the vowels and indicating the rising or falling pattern. For example, <aâ> would indicate low-high rising, <âa> high-low falling, and <äa> mid-low falling.

Part five addressed the indication of intonation by way of punctuation marks.

Part six addressed several topics:

1. Digraphs must never be divided by a hyphen.
2. The plural prefix *â-* was to always be attached to the noun. (Other parts of speech to which the plural might attach were not mentioned.)
3. The definite article *nî* was to be written separately from the noun it modified, but was to be affixed in adverbial expression. (This would mean writing *na ndo nî* ‘in that place’, but *na ndônî* ‘in addition’) A list of locative and time phrase examples were provided, such as *na pekônî* ‘afterward’.
4. The subject marker *a-* was to be affixed to the following verb.

No directives were given for writing French loan words, compound nouns, and words for which there were dialect variations. But the document did mention the forthcoming publication of an orthographic dictionary, which would conform to the Code. The *Dictionnaire orthographique du sāngö* (Diki-Kidiri 1998) was published fourteen years later.

### 12.10.2 *Changes from the status quo*

Few consonants were affected. The open and closed vowels /e/ and /ɛ/, and /o/ and /ɔ/ were not to be differentiated. Vowel /u/ was to be written as <u>, abandoning the French digraph <ou>.

The most drastic change was the insistence by ILA linguists that tone should be marked exhaustively. High tone, which according to Diki-Kidiri (2005) makes up 38% of the total, was to be marked with the circumflex (^); mid tone, which made up 18%, with dieresis (¨). These symbols were chosen because both were available on French typewriter keyboards as dead keys; in contrast, accents were usually only available as fixed keys on <e> and <a>. Low tone, the most frequent tone at 44%, was to be indicated by the absence of diacritics above the vowel, resulting in the least possible number of diacritics overall.

Although dialect issues were not addressed in the decree, ILA expressed certain preferences: For words where [l] and [r] varied, <r> was to be written, not <l> as was the former practice. Affricates were not to be represented. These preferences unofficially treated the dialect of the capital (Bangui) as the reference dialect.

Three French spelling conventions were to be abandoned:

1. <s> was not to be doubled in word medial position; the rationale: <s> would never be interpreted as [z] in Sango as it was in French.
2. /g/ was to be consistently written as <g>. The practice of writing <gu> before <e> and <i> was to be abandoned; the rationale: <g> would never be interpreted as [ʒ] as it was in French; Sango does not have a phonemic /ʒ/. (The sound [ʒ] exist, but as a dialectal variation of [z].)

Similarly, the digraph <ng> was to be consistently written <ng>. No <u> was to be inserted before <e> and <i> as was the practice previously.

3. /w/ and labialization were no longer to be written as <ou>; /w/ was to be written as <w> and labialization as <u> following the consonant.

Overall there were not many changes. (See Table 18 for a summary.) Only two consonants were replaced: <gn> with <ny> and <ou> with <w>. Other conventions had already been in use. The consonant symbols of the official orthography correspond to the characters of the International Phonetic Alphabet with the following exceptions:

<y> stands for [j], <ng> for [ŋg], <ngb> for [ŋgb], and <ny> for [ɲ].

Table 18. Comparison of old and new conventions

unaffected by '84 decree:	a b d f gb h i k kp l m mb n nd p t v y
vowel conventions:	formerly: e é è o au ou new: e o u
consonant replacements:	formerly: gn ou new: ny w
French rules abandoned:	formerly: #s; VssV g{a,o,u}; gu{e,i} ng{a,o,u}; ngu{e,i} new: s g ng
unilectal representation	formerly: mv, v mb, mgb, ngb r, l dj, j, z new: mv ngb r z

Although the changes were not drastic, printed text nevertheless ended up with quite a different look. (See Table 19.) This was due to (1) the high frequency of occurrence of <ny> and <w>, (2) the replacement of <ou> with <u>; (3) the elimination of accents; (4) no more doubling of <s>; and (5) the introduction of diacritics for noting tone.

Table 19. Impact of the change: comparison of words and text

Based on French and the 1966 Bible	According to ILA's interpretation of the 1984 decree	English translation
soussou, assoussou	susu, âsusu	<i>fish (sg.), fish (pl.)</i>
aouou	awü	<i>numerous, plentiful</i>
guigui	gîgî	<i>life, outdoors, nature</i>
Lo lango na tëlé ti guèlé ti Boaz juska na nda-pelèlé, na lo loundou kodjoni si jo alingbi ba là ti mbèni jo.	Lo längö na terê tî gerê tî Boaz zusûka na ndâpêrê, na lo löndö kôzonî sî zo alîngbi bâa lê ti mbênî zo. <sup>89</sup>	<i>She lay at Boaz' feet until morning, and got up before anyone could recognize (see face of) a person. (Ruth 4:14)</i>

### 12.10.3 Dissemination efforts and its effects

ILA distributed copies of the decree and orthography document to the agencies who were at the time producing literature of any kind in Sango. As new agencies arrived over the years, they were also presented with copies. However, the language of the document was problematic. It had originally been prepared for linguists, not for a wider audience. Since the vocabulary employed was technical and no sample text was provided, many who received a copy did not understand it and filed it away. The general population was not necessarily aware of the official orthography. There were some negative reactions from the literate community and literature publishers. The matter of writing tone met with

<sup>89</sup> The text on the left is a citation from the 1966 Bible. The middle column does not exist in any publication. It is only given here to illustrate the potential impact of the orthography decree.

the most resistance. People felt it was excessive and an unnecessary burden to readers and writers. Le Page (1997:60) noted:

There is considerable disagreement as to the best orthography, particularly between those who, like Diki-Kidiri, want to mark tone both to avoid ambiguity between homophones and to mark the ‘Africanness’ of the language in its written form, and those who feel that the context almost always disambiguates homophones and that marking tone creates unnecessary difficulties for the learner.

ILA was the ‘orthography authority’ but did not have the right to regulate publications. As a service, ILA personnel read and edited documents for other organizations, bringing them into conformity with the ’84 decree. Some clients accepted the editorial changes, but others disregarded or only partially implemented the changes, viewing them merely as recommendations or suggestions. Most agencies refused to mark tone.

ILA made it a priority to train the CAR communications personnel of the national radio and television stations. Efforts to purge French from Sango were quite successful. But promotion of the official orthography lagged behind other language planning efforts. ILA linguists had little direct impact on how Sango was being written by the general population or those producing literature: (1) Radio and television personnel were not involved in print media, so their knowledge only benefited themselves; (2) ILA’s thematic lexicons did not have a wide distribution; (3) ILA was not usually directly involved in adult education; and (4) written Sango had no part in formal education, except at university level.

Fasold (1997:260) felt that government agencies were not proactive enough in the implementation of the new orthography. He comments, “The official orthography, with

the current laissez-faire policy, and with the restricted domains of use of written Sango, cannot be expected to spread in the immediate future.” What Fasold refers to as “laissez-faire policy” could be interpreted differently. First, there was a lack of financial resources to make much of a concerted effort to promote the official orthography. Secondly, since the first reactions to the orthography had not been very positive, the lack of proactive measures could have been an act of diplomacy and caution.

The Peace Corps’ *Mabôko na Mabôko* (Debold 1986) was one of the first books printed with the new standard, including tone. Father Saulnier, a Catholic priest, also ventured out and compiled a diglot French-Sango book on medicinal plants (Saulnier 1995), applying the new standard, including tone. (According to Diki-Kidiri (1998:237), Saulnier also produced an orthographic dictionary in 1994.) ILA had direct impact on *Projet Bouca*, a functional literacy program, since two of their associates developed the pedagogical material. The eight books were in line with the ’84 decree, including tone notation.

Agencies who worked directly with Central African churches contributed the most to the promotion of the official orthography. Most consumers of Sango literature were found in the churches. It helped that the churches had the most stable infrastructure in the country, relatively unperturbed by the political turmoil of 1996–97 and 2001. While a few missionaries might have hoped that the new orthography would go away, most accepted it.

Swedish missionaries, along with their staff linguist Christina Thornell, had no objections to the official orthography, except for noting tone. Unlike missions that had arrived much earlier, they had not invested in the French-based orthography. Together

with African colleagues they developed materials of all sorts with the official orthography, including basic literacy material. However they did not mark tone.<sup>90</sup>

The Grace Brethren Mission revised some of their literature, introducing one or two changes at a time, when reprinting existing publications. They also revised their primer for their literacy classes. Mission personnel and the literature production committee felt torn between adopting the new orthography and accommodating their reading clientele: pastors and others who came through the ranks of French education and were accustomed to reading the old ‘standard’ found in the Bible. With each work going to press, old or new, they pondered how to write certain words. They came up with their own solutions for some dilemmas. For example, they observed that when <gu> was used for /g<sup>w</sup>/ when preceding <i> or <e>, some readers paid no attention to the <u> marking labialization since in French /g/ is written <gu> in this environment. Consequently, the literature production team experimented by writing labialization with C+*w*. Thus, literature produced by this mission sometimes promoted the new orthography, sometimes the old, and sometimes chose something else as a compromise.

The non-denominational para-church organization Emmaus, known for producing Bible correspondence courses in many languages, initiated a Bible distance learning program in Sango.<sup>91</sup> The official orthography was adopted, except that tone was not marked. The courses were surprisingly popular, considering that most clients had no previous experience with this orthography and that there was a fee charged for lessons.

---

<sup>90</sup> This mission, *Mission Baptiste Suedoise*, was later renamed *Interact*. The literature they helped develop was published under the *Église Baptiste de l'Ouest* (EBO).

<sup>91</sup> Emmaus Correspondence School was established in 1942 in Canada. Today, Courses are available in 125 languages in 105 countries. Seven course titles are available in Sango (Emmaus 1997–2006).

Grand'Eury, having surveyed community development programs and literature produced in the CAR in the early 90s, had a positive outlook concerning the implementation of the official orthography, commenting:

These past two or three years, most of the churches (protestant and catholic) have been progressively adopting the official orthography in their new publications, and thanks to computers, they are able to produce an increasing number of small booklets in Sango, dealing with health issues, agriculture etc. (Grand'Eury 1994:64; translation mine).<sup>92</sup>

### **12.11 The role of SIL International**

Central African linguists heard about SIL's work in other African countries. Of particular interest was SIL's part in introducing the use of local languages into Cameroonian schools. At the initiative of ILA personnel, SIL received an invitation to work in the CAR. A contract between the CAR government and SIL was signed in 1990. SIL is not often involved with languages of wider communication, but the government clearly wanted SIL involved. Concerning Sango, the contract referred to SIL having a part in establishing the orthography, preparing literacy material and pedagogical material for the introduction of Sango into the schools as well as transition material from mother tongue languages to Sango.

Various leaders of churches requested help from SIL in the area of adult basic literacy instruction. ILA's interest was in the promotion of the official orthography. In response, SIL implemented five strategies:

---

<sup>92</sup> Original text: "depuis deux-trois ans, la majorité des Eglises (protestantes et catholique) adopte progressivement dans leurs nouveaux écrits l'orthographe officielle et grace à l'informatique, elles sont en mesure de produire en sango de plus en plus de petits livrets divers qui traitent de la santé, l'agriculture, etc."

1. Interpreting the '84 decree for the general public
2. Testing the official orthography with non-literate adults
3. Testing the orthography with adults literate in French
4. Disseminating literature using the official orthography
5. Harmonizing orthographies being developed for other national languages with the official Sango orthography.

#### *12.11.1 Interpreting the '84 decree for non-linguists*

In 1991, the English-speaking missions community requested that SIL explain the *Code de l'orthographe officielle du sango* at a special meeting. In response, I prepared a draft document in English comparing the old and new (now official) orthographies. It explained changes and provided sample words. It could have been an effective document, but since I was inexperienced in orthography matters at the time, I had unfortunately used the words “right” and “wrong” in the document, resulting in a very negative reaction. Subsequently I prepared an equivalent six-page document in French, *Lire et Écrire le Sango — Sans Difficulté* (Reading and Writing Sango With Ease), using the terms *ancienne* (former) and *officielle* (official). This document was used during literacy teacher training and writers workshops, and distributed to Francophone agencies upon request.

I prepared a second document: a self-instruction booklet in French (SIL 1992). Answers to exercises were provided in the back. It served as a tool for transitioning from French to the official Sango orthography. The former Sango orthography conventions were not included in the discussion.

### *12.11.2 Testing the official orthography with non-literate adults*

I led a team who prepared basic literacy material using the official orthography. We trained volunteer teachers who taught a few pilot classes. It was a very modest beginning because, on one hand, few people were interested in participating in an orthography experiment, and, on the other hand, past adult literacy endeavors had promised much and rarely delivered (Grand'Eury 1994:64). Despite this, by using a proven methodology and following up on classes, the results were good. After a few graduates from the program were able to read publicly, teach Sunday School classes in church, and even help children with their French homework, people took notice. SIL received requests for additional classes. The program expanded from Bangui to outlying provinces. Sponsorship remained with the churches or local community. SIL provided teacher training, supervision, and subsidized materials.

It was clear: the orthography was learner-friendly. Students as old as 70 learned to read. It also became clear that, although tone markings may have enhanced the development of word attack skills at first, learners were not dependent on these marks when reading. All reading material used in the classes had tone marked except materials produced by the national Bible Society. The switch from full tone marking to no tone marking did not perturb learners nor experienced readers.

### *12.11.3 Testing the orthography with adults literate in French*

SIL tested the orthography with adults literate in French. One type of testing was done by observing people's emotional reactions. For example, a vegetable seller, seeing a book with the official orthography, asked: "What language is that?" And when told it was the

new way of writing Sango, she commented: “It looks like Lingala.”<sup>93</sup> When given the opportunity to read a few sentences, she noted: “It’s easy.”

Basically, the young generation was very open to Sango’s new look. Older church leaders were less enthusiastic about the change. They reacted negatively to some conventions that differed from those of French.

The second type of informal testing came from observing the teachers. Once trained, they had no difficulty in writing any of the primer material, but their tone awareness had not been heightened. When asked to write original texts, experienced teachers mastered the segments, but not the tone notation.

A third type of testing occurred in writers’ training courses. In two different courses, each with eighteen participants or more, individuals wrote original compositions. They read these to each other and gave feedback on how the stories might be improved. The content was excellent: stories were creative, funny or heart-wrenching, true or fictional. Spelling increasingly conformed to the official standard, but most participants wrote their stories without tone and then submitted them to someone who could write tone ‘correctly’—the ILA linguists and church choir directors. Others had trouble. In order for them to mark tone correctly, they would have had to consult a dictionary or memorize the tone pattern for every word or to say each syllable out loud and compare it to a standard high, mid or low tone syllable. That was certainly not desirable. The requirement to write tone made writing more difficult, created stress, made writers self-conscious, and reduced potential writers’ motivation.

---

<sup>93</sup> Lingala is the Lingua Franca used in the Republic of Congo and the western part of the Democratic Republic of Congo (formerly Zaire).

#### *12.11.4 Dissemination of literature employing the official orthography*

Little literature was available using the official orthography. Since opposition to the official orthography was partly due to a lack of exposure and a psychological reaction, in 1992–1994 and 1997–2000 SIL made it a priority to produce and to help others produce literature. The population's cash flow problems kept sales of most titles at a minimum, but still some bestsellers were in demand: small religious tracts, health booklets, a book on marriage. The new titles served to supply new readers in the literacy program with reading material. Those who were used to reading the official orthography without tone notation welcomed the literature despite the tone, but basically ignored the diacritics.

Since the promised orthographic dictionary had not yet made its appearance, SIL produced one in 1995 with about two thousand entries of the most common words and idiomatic expressions. Since Sango had been elevated to official language in 1991, SIL anticipated that expatriates would make use of such a dictionary, so included French, English and German equivalents. ILA thoroughly checked the Sango entries.

#### *12.11.5 Harmonization of other orthographies with Sango*

SIL was involved in language development work in other CAR languages. After doing the analysis and preparing a phonological description, an orthography proposal would be submitted to ILA. Whenever possible, the languages were written like Sango. Some exceptions were made in symbolization: /e/ and /ɛ/, /o/ and /ɔ/ were usually each assigned a symbol instead of representing two phonemes by a single symbol. Tone, when marked, was marked as in Sango. This allowed bilingual Aka pygmies, who had learned to read in Yaka (a Bantu language with two tone levels, high and low), to pick up a Sango book and read it with understanding without further instruction. Due to low functional load, few local languages needed extensive tone marking.

### *12.11.6 SIL's recommendations concerning the orthography*

In 2000, after ten years of SIL work in the CAR, the management of its Sango literacy program was officially transferred to the Central African organization ACATBA. Before this occurred, however, I wrote up observations and recommendations concerning the Sango orthography (Karan 1998) based on my experiences from my years of involvement. SIL submitted this set of recommendations to ILA and the Ministry of Education in a report.

The report was positive in regards to the implementation of the graphemes of the official orthography, but it provided evidence that exhaustive tone marking was problematic and counterproductive to the promotion of Sango as a written language. It recommended that ILA reconsider its position on exhaustive tone marking. It urged the promotion of a practical orthography employing minimal tone marking for general literature. Exhaustive tone marking was to be reserved for linguistic works and language acquisition materials aimed at expatriates. Since the reading of new technical terminology was a concern, it recommended providing pronunciation helps in parentheses, where necessary. It encouraged dialogue and compromise.

There was no official response from ILA or the Ministry of Education. Despite making these recommendations, SIL did not abandon tone notation. This was not a matter of conviction, but respect for the contractual partnership with ILA. ACATBA continues to use materials with exhaustive tone marking, both their own new titles as well as those published by SIL during the experimental phase of the program. Sunday school material developed under ACATBA supervision, however, uses minimal tone marking, in line with conventions used in the revised New Testament (Alliance Biblique en Centrafrique 2001; Bertin Oundagnon-Basso, personal communication).

## 12.12 Current practices

At first glance, one might be tempted to describe the Sango situation as chaotic, but, in fact, great progress has been made toward a unified standard. I have compiled some sample literature which is representative of the current practices in writing Sango.

### 12.12.1 Orthography in the literature

I will present an analysis of actual writing. Table 20 presents practices based on basic literacy materials designed to teach reading. It compares past and present practices.

Table 20. Comparison of spelling conventions found in basic literacy materials

<i>PBB '81, '00 MEF</i>	<i>Min. de l' Edu. Nat. J&amp;S&amp;A</i>	<i>'84 official SIL, UNICEF</i>	<i>MEF rev.</i>	<i>EBO '86, '89</i>	<i>W.B.T.S. '00</i>	<i>Ed. Non- Formelle</i>	<i>C.F.A.R.</i>
e, é, è	e, é, è	e	e	e	e	e	e
ou	ou	u	u	u	u	u	u
g, gu	g, gu	g	g	g	g	g	g
ng, ngu	ng, ngu	ng	ng	ng	ng	-	ng
mb, mgb, ngb	ngb	ngb	ngb	ngb	ngb	-	ngb
gn	ny	ny	-	ny	ny	-	ny
nz, ndj	nz	nz	nz	nz	nz	-	nz
r, l	r, l	r	r	r, l	r, l	r, l	r
s, ss	s, ss	s	s	s	s	s	s
dj, j, z	z	z	z	z	z	-	z
ou	w	w	w	w	w	w	w
C+ou (labializ.)	C +w	C +u	C +u	C +u	C +u	C +u	C +u
a-, a	a	a-	a-	a	a-, a	a-	a-
min. tone/ no tone	no tone	all tone	no tone	no tone	intro to tone	all tone	no tone

For the sake of space, only changes resulting from the 1984 decree are included. The first column represents the French-based conventions found in the 1966 Bible and earliest primers and literature. One column indicates the official orthography conventions employed by ILA, SIL, ACATBA, and agencies for whom ILA personnel developed

materials. The remaining columns represent works by different agencies. (See Appendix B for document details.) There is an amazing degree of conformity to the official orthography in basic literacy materials across agencies. The writing system of 1984 is easy for learners and instructors thus there is little hesitation to use it in literacy programs. The biggest divergence occurs in tone marking and in whether the subject marker *a-* is attached or not.

Other literature is a different matter. The clients are not just those who came through the ranks of adult literacy classes, but also those who have had years of formal education. Many are loyal to the French-based spelling system. Literature producers are faced with a dilemma: to whom should they cater in their literature?

Table 21 compares orthographic practices in literature aimed at experienced readers. Not all texts used contained the item being compared, resulting in blank fields. Clearly, in post-literacy materials, like in basic literacy materials, efforts have been made by publishers to work toward the official orthography. For instance, other than in reprints of old literature, <s> is not doubled when not word initial; <g> and <ng> are no longer followed by <u> when preceding <e> or <i>. But, as with the basic literacy materials, there is divergence in the treatment of the subject marker, labialized consonants, and tone notation.

There are also differences in some other areas not addressed by the decree. For example:

1. treatment of contracted words (such as *ködörö*, for which the second <ö> is elided)

2. writing “silent h” before verbs where [h] is not evident in most dialects
3. writing of French loan words
4. treatment of the demonstrative /ní/ and adverbial expressions containing /ní/.  
(At times it is attached; at times it is written separately.)

Table 21. Comparison of spelling conventions found in post literacy materials

<i>ABU '66 BMM '82, '83</i>	<i>'84 official ILA, SIL, Pr. Bouca</i>	<i>ABU '95, '01, '02</i>	<i>C.E.C.A. '94, 98</i>	<i>MEF (n.d)</i>	<i>EBO '86, 88, '89</i>	<i>S.E J.P. C. '91</i>
e, é, è	e	e	e	e	e	e
ou	u	u	u	u	u	u
g, gu	g	g	g	g	g	g
ng, ngu	ng	ng	ng	ng	ng	ng
mb, mgb, ngb	ngb	ngb	ngb	ngb	ngb	ngb
gn	ny	ny	ny	ny	ny	
nz, ndj	nz	nz	nz	nz	nz	
r, l	r	r, l	r, l	r, l	r, l	r, l
s, ss	s	s	s	s	s	s
dj, j, z	z	z	z	z	z	z
ou (labializ.)	w	w	w	w	w	w
C +ou	C +u	C +u	C +u/w	C +w	C +u/w	C +u
a-, a	a-	a-, a	a	a-	a-, a	a
min. tone/ no tone	all tone	min. tone	no tone	no tone	no tone	no tone
kodro	ködörö	kodro	kodro		kodoro	kodro
hinga	hînga	hinga	inga	hinga	inga	inga
honda	hûnda	honda	unda	honda	honda, unda	unda <sup>94</sup>

<sup>94</sup> The first sound in *hînga* “to know” and *hûnda* “to ask” is not /h/ but /ʔ/.

If zero-tone marking negatively affected comprehension, more agencies and individuals would probably decide to mark tone. The 1966 Bible differentiated between a few minimal pairs. In recent Scripture publications, some minimal pairs are marked.

Table 22 provides a partial list.

Table 22. Potential homographs discriminated in Alliance Biblique orthography

Potential Homographs	English equivalents	1966 Bible	2005 New Testament
/só/, /sō/, /sḍ/	<i>this, save, hurt</i>	so, sau, so'	so, soo, so <sup>95</sup>
/lé/, /lḗ/, /lḙ/	<i>fruit, eye, face</i>	lé', lé, lè	le, le, le
/tènè/, /tḗnḗ/, /tḙnḙ/	<i>say, word, stone</i>	tènḙ, tènḙ, ténḙ	tene, tene, tḙne
/dà/, /dḗ/	<i>house, there</i>	da', da	da, daa
/k <sup>w</sup> á/, /k <sup>w</sup> ā/, /k <sup>w</sup> à/	<i>corpse, hair, work</i>	koua', koua, (koussala)	kuā, kua, kua
/fà/, /fá/	<i>show, cut/kill</i>	fa, fa'	fa, fā,

### 12.12.2 Orthography as practiced by individuals

When SIL began work in the CAR in 1990, it was evident that few Central Africans were aware of the existence of a decree concerning the orthography. Few people felt a need to write in Sango at all. When they did, they usually wrote in a spelling system that came naturally: the one to which they had the most exposure, i.e., French. This section will look at some samples of writing by different individuals: letters, study notes, a religious tract, an election flyer. These had been discarded. Since individuals writing in Sango are usually part of a church community and would most likely base their practices on French orthography conventions and on the most distributed book in Sango, the 1966

---

<sup>95</sup> As noted in 12.10.1, consecutive identical vowels with differing diacritics indicate a contour tone. In the absence of diacritics, the two-vowel sequence in itself serves to distinguish between words. Underdifferentiation of the open and closed vowels, not only the absence of tone marking, can result in homographs.

Bible, I will treat this as the expected norm and comment on divergences from that system.

1. *A two-page handwritten personal letter from a women's group leader in the north of the country to her counterpart in the capital:* This letter (dated July 25 2005) is written with the system found in pre-'84 church literature. There is one exception: the word for 'family' /sēwā/ is not written as < séoua > as one would expect. Instead it is written < séwa >. We note the use of < w > instead of < ou > for /w/ as would have been expected. In the Sango Bible, *séoua* was not used; the French loan word *famille* had been used instead.

2. *A half-page letter from a man living in the north of the country to his elder brother in the capital, upon hearing the news about the death of their father:* This letter is not dated, but was written in 2005. Again, French spelling rules are applied. But, there is some divergence from the 1966 Bible standard. For instance, accents in words containing two < e > did not follow the *C è C é* pattern, as was the case in the Bible. Instead, these were changed to acute accents on several words. This notation was also not consistent with French spelling practices. In French an acute accent would normally be found on /e/ and a grave accent on /ɛ/. This distinction was not made in this letter: the < e > s in *téné*, *légué*, and *téré* all bear the same accent, although vowel quality differs in these words. The pronunciation would be /tēnē/ 'words', /légē/ 'way' and /tèré/ 'body' respectively. In addition, the writer wrote certain words with < r >: < sara > 'make, do' and < téré > 'body'. In the 1966 Bible these two were always written with < l >.

3. *Five pages of hand-written Bible study notes by a young man, age 25, on financial accountability (dated 2.19.05):* These notes, by nature of the theme, are full of French

words, which are spelled as such: *caisse* (cash box), *recette* (receipt), *tiroir* (drawer), *armoire* (closet), *poche* (pocket), etc. Sango words are spelled according to the 1966 Bible spelling standard, but there were inconsistencies: ‘person’ was spelled the new (1984) way, <zo>, eleven times, and only once the old (1966) way <jo>. ‘Good’, is written the new way <nzoni> twice and the old way <ndjoni> twice. The verb ‘make’ is written once with <r>, <sara> (new), and five times with <l>, <sala> (old). This person has definitely had exposure to the official orthography, which has influenced his writing habits: the question word ‘what?’ and ‘theft’ are written with the new standard <nye> and <nzi>, not <gne> and <ndji> as would have been expected. Although the pages are full of the French digraph <ou> in most words, there are five occurrences of /u/ written as <u>. The plural affix, the subject marker, and even the derivational affix <-ngo> are randomly attached or written separately. No attempt was made to indicate tone.

4. *A half-sheet election publicity flyer (April 2005)*: This bilingual Sango and French flyer has very little text, just a few slogans. What is unusual is that *BE-AFRICA*, the name of the CAR in Sango is written with <c> which is not part of the Sango alphabet. One would have expected <k> in Sango; in French the name of the continent is written as *Afrique*. ‘Body’ *TERE* is written with <r>, not with <l> as in the 1966 Bible. Since the flyer is printed completely in upper case, no diacritics are marked anywhere (following the French convention during the era of the typewriter).<sup>96</sup>

---

<sup>96</sup> The most striking thing about this flyer has nothing to do with spelling. It says: “One country, one people, one language.” The French equivalent actually says “*UNE SEULE LANGUE*,” literally, ‘a single language’—a strange phrase for a bilingual flyer which seeks votes from all ethnic groups. ‘A common

5. *A folded religious, illustrated tract, translated into Sango by a pastor (1989)*: More than 30% of this text consists of citations from the 1966 Bible. One would expect those citations to provide the spelling standard for the whole tract. Surprisingly, the pastor took the liberty to modify the spelling of three words, even in the citations. *Zo*, *zia*, and *nzoni* are used instead of *jo* ‘person’, *djia* ‘let’ and *djoni* ‘good’. Although <ss> was kept in word medial positions, they were changed to a single <s> when a morpheme break was involved. For example in the Bible one finds *assala* ‘to do, make’ (subject marker *a-* attached to *sala*, ‘do’) and *ouassiokpari* ‘sinner’ (compound noun consisting of *oua* ‘person’ + *siokpari* ‘sin’ ). This individual wrote these words simply *asala* and *ouasiokpari*. Additional examples were *asoro* and *ouasaungo*. There was some inconsistency: in two instances a hyphen was used: *oua-siokpari*.

6. *Printed devotional notes for choir, same church, two dates (October 2004 and June 2005)*: Comparing the two sets of notes, it is clear that they were not drawn up by the same person. The person who drew up the more recent document is much more consistent in following the older, 1966 Bible spelling conventions. Two items are noteworthy: The author changed the direction of the accents so they would line up with French practices: acute for [e], and grave for [ɛ]. Secondly, the plural prefix *a-* is attached to all Sango words, but it is not attached to French loans, such as *a instrument* ‘instruments’ and *a choriste* ‘singers’. There is inconsistency with one word spelled two ways: *aresponsable* and *a responsable* ‘leaders’.

---

language’ might have been a better choice.

The author of the older document integrated a few of the official orthography conventions. For instance *zo* was used instead of *jo* ‘person’, *nzoni* instead of *ndjoni* ‘good’, *zia* instead of *djia* ‘let’, and <w> was employed to denote the consonant instead of <ou> in words such as *wamabe* ‘believer’, *mawa* ‘sorrow’, and *sewa* ‘family’. But <ou> continued to be used to denote labialization, such as in the word *koua* ‘work’.

### 12.12.3 *Print in the environment*

Public signage increasingly employs Sango. Although literature is increasingly lining up with the official orthography, this is not the case for signs and posters found around town. Not subjected to an editorial process, these are more likely to reflect the spelling practices of individuals. One must not fault the creators of the signs. As long as the general public is not educated concerning the official orthography, one cannot expect conformity. Public signage is, of course, much more visible than personal correspondence or study notes. Some signs found in public places in the capital city of Bangui will be presented here, along with a discussion as to how their spellings compare with the old and the official orthography standards.

The three signs in Figures 5, 6, and 7 are examples of 100% adherence to the old, French-based spelling system. Accents are not used with upper case. Figure 8 is an example of a complete switch to the official orthography—but tone markings are absent. This might be due to the use of upper case. These four signs were posted at hospitals and clinics, the most common locale for Sango signage according to Gerbault (1994:75).

DEFENSE DE CRACHER  
PAR TERRE MO KOU  
SORO NA SESSE PEP

Figure 5. "No spitting" prohibition

NUMS DES MOIS EN SANGO

GNEGNE	= Janvier
FOLONDIGUI	= Février
MBANGO	= Mars
NGOUBE	= Avril
MBELAHO	= Mai
FONDO	= Juin
LENGOUA	= Juillet
KOUKOUROU	= Août
MVOUKA	= Septembre
NGBERERE	= Octobre
NABANDROU	= Novembre
KAKAOUKA	= Décembre.

Figure 6. Months of the year



Figure 7. Information Desk

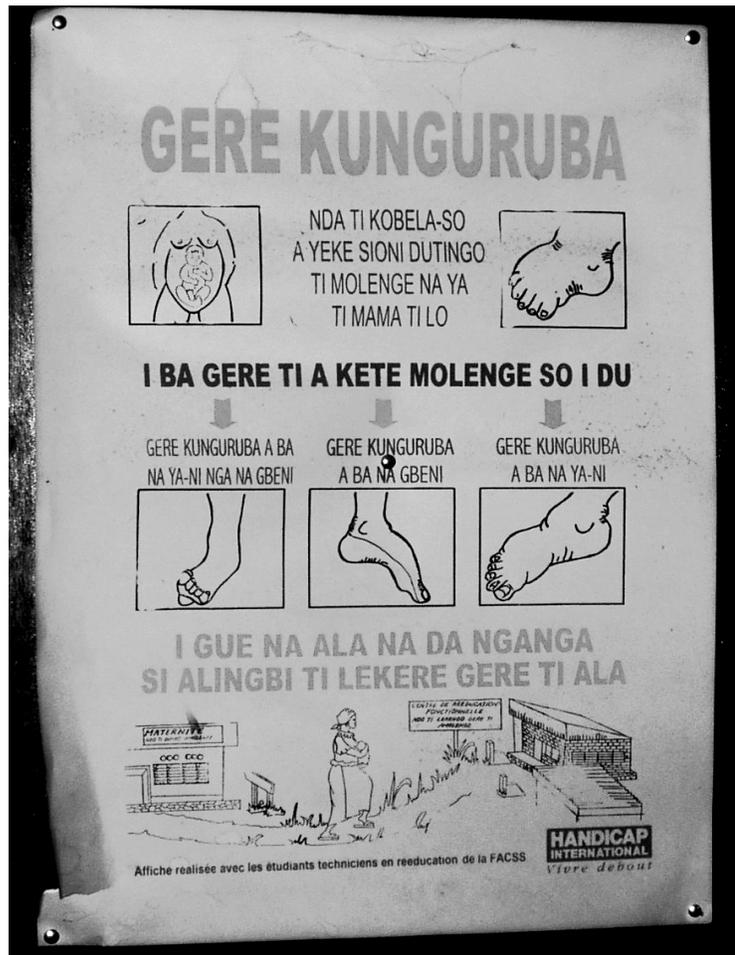


Figure 8. Clubfoot information poster

Table 23 indicates which orthography was applied in each of these signs and posters.

Table 23. Analysis of writing conventions used in Figures 5–8

Figure	words in sign	comment on system used	in former system would be	in official orthography would be	English equivalent
5.	kou soro sesse	old old	sessé	kü soro sêse	<i>spit ground</i>
6.	gnegne folondigui lengoua koukourou	old old old old		nyenye folondingi lêngua kûkûrû	<i>January February July August</i>
7.	hounda gué	old old		hûnda ge	<i>ask here</i>
8.	gere molenge gue	official, minus tone official, minus tone official	guelé melèngué gouè	gerê môlengê	<i>foot, leg child go</i>

The three signs in Figures 9, 10, and 11 were posted in three different locations. Although posted by the same agency, their spelling is not consistent.



Figure 9. AIDS: Call to action, Airport Road



Figure 10. AIDS: Call to action, Boyrabe District



Figure 11. AIDS: Call to action, Avenue des Martyrs

Table 24 contrasts actual words from these last three signs with what they would look like in the former and official writing system.

Table 24. Analysis of writing conventions used in Figures 9–11

Figure	words in sign	comment on system used	in former system would be	in official orthography would be	English equivalent
9.	kue nzoni	new, no tone new, no tone	kouè ndjoni	kûê nzönî	<i>all</i> <i>good</i>
10.	koue	old, no accents	kouè	kûê	<i>all</i>
11.	lege mossoro a mai	new, no tone old either, no diacritics	lègué mossoro amaï	lêgè mosoro amâi	<i>way</i> <i>wealth</i> <i>flourish</i>

Mixing the old system with the new is the current norm. The signs in Figures 12, 13, and 14, placed by different agencies, testify to this. Those writing the texts are making an effort to communicate in written Sango. It is doubtful that sign makers are trying to promote one system over another. Table 25 illustrates the different spelling choices made.

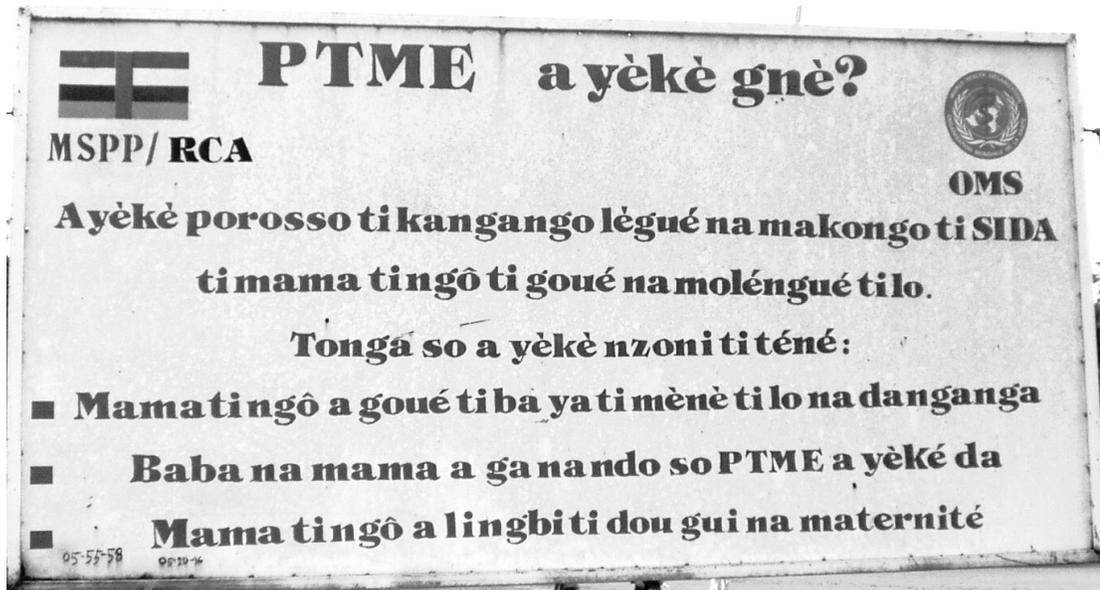


Figure 12. Protecting the unborn against AIDS



Figure 13. Nursing babies is best



Figure 14. Advertising

Table 25. Analysis of writing conventions used in Figures 12–14

Figure	words in sign	comment on system used	in former system would be	in official orthography would be	English equivalent
12.	gnè goué yèkè dou gui nzoni	old old, accent flipped old, accent flipped old old new, without tone	gnè gouè yèké dou gui ndjoni	nye gue yeke dü gï nzönî	<i>what?</i> <i>go</i> <i>be</i> <i>give birth</i> <i>only</i> <i>good</i>
13.	gui ngu nzoni molengue	old new, without tone new, without tone old, vowel changed	gui ngou ndjoni melèngué	gï ngû nzönî môlengê	<i>only</i> <i>water, liquid</i> <i>good</i> <i>child</i>
14.	mou légé zo téré	old original new original	mou lègué jo télé	mû lêgë zo têre	<i>take/give</i> <i>way</i> <i>person</i> <i>body</i>

We note from the photos that some of the signage is not recent. Some billboards have served for years. Since the dates of creation of the signs and posters is an unknown variable, it is not clear if adoption of some of the official orthography conventions correlates with time (has been progressive) or with the amount of exposure to the new conventions by the individuals who were involved in the posting of the signage. However, the variance in spelling practices indicates that this is a time of transition and that people are not totally unaware of the official orthography. To bring greater uniformity to the way Sango is written, more awareness-raising is needed.

#### *12.12.4 Sango in cyberspace*

Sango is not only a language which is described on the Internet, but, as Diki-Kidiri hoped, some resources in Sango are available on the Internet. Diki-Kidiri himself maintains a bilingual Sango-French website with several pages, including information about the Central African Republic, a bibliography of publications in Sango and a dictionary, among other resources. Lessons for learning Sango are available, but for these one must subscribe. Until the end of 2005, this website maintained by Diki-Kidiri used the official orthography as it was proposed in 1984 and which was used in his 1998 dictionary. Since January 2006, he has applied different spelling rules. This orthography, designed by Diki-Kidiri, will be discussed in section 12.13.

Two elaborate websites using Sango are maintained by C. Mabada-Mabaye in association with *Ministère de la Nouvelle Alliance* (M.N.A.), a religious organization in France. One of the sites uses the old (French and 1966 Bible-based) spelling conventions for the most part and has mostly religious content, while the second uses the official (1984) orthography and states its purpose as the “defense and promotion of Sango, the

official national language of the Central African Republic” (translation mine). This second site marks tone consistently on every syllable, as ILA linguists would.

The Watch Tower Bible and Tract Society (2006) maintains a website containing pages with texts translated into 270 plus languages, including Sango. Nine different pages provide background information on Jehovah Witnesses, dated 2000. In addition, sixteen lessons, dated 1996, teach Jehovah Witness doctrines. The 1984 official orthography is applied except for tone marking. Tone is marked on a few words to differentiate minimal tone pairs, even if these are not likely to cause ambiguity in context, for instance: *mênë* ‘blood’ vs. *mene* ‘swallow’. The site states that Bible verses, unless stated otherwise, are cited from the 1966 Bible (Alliance Biblique Universelle), “me na fini lege so asû na Sango” (but in the new way that Sango is [now] written).

Other than these sites, there is little available on the web that is written in Sango. One text, the Sango translation of *The Seven Signs of Jesus* is available on the Grace Brethren Mission website (2005). This document, to be used by church workers, uses the official orthography for the most part, except that tone is not marked and labialization is marked with <w> not <u> as prescribed in the decree. A few French spellings are retained.

The Sango version of the Lord’s prayer in Sango is posted on the web as part of a collection of this text in 1353 languages and dialects. The Sango version is listed in the pre-1984 French-based script (Christus Rex 1994).

Some Central Africans have chosen catchy Sango headers for their personal or organizations’ websites, but the remaining content is in French. The content of *Sangonet*, a site set up for potential visitors to the CAR, is also exclusively in French.

Wikipedia has reserved a page for Sango language entries and welcomes visitors: *Gä nzönê na ndö tî lêndo tî Wikipedia na yângâ tî ködörö tî Bêafrîka!* (Welcome to the Wikipedia site in Central African Republic's national language!) There are no articles to date. The invitation is open: "If you know this language, replace this text and start creating your encyclopedia!"

### **12.13 New developments**

In early 2006, Marcel Diki-Kidiri made a visit to the Central African Republic and met with ILA linguists to discuss a new orthography option. He presented a paper which explained the changes and the motivation behind them. Meetings were called and some of the stakeholders were asked to attend. ILA, ACATBA, and the United Bible Societies were represented at the meeting. They were urged to test this new orthography. Soon thereafter, all written texts on the website maintained by Diki-Kidiri conformed to the new rules.

#### *12.13.1 Motivation for the new proposal*

As reasons for the change, Diki-Kidiri stated that although it had been 21 years since the official Sango orthography was settled by decree, hesitation to use it persisted and solutions had not yet been found to get people to use it. He stated that marking tone on only some of the words (based on potential confusion) is an unsatisfactory practice for legal documents, pedagogical material, news media, or novels. He thus proposes to implement a strategy which reduces the number of diacritics required for writing Sango, in the hope of removing the stumbling block to acceptance.

### 12.13.2 Proposed changes

Diki-Kidiri proposes that the letter <j> is to be included in the alphabet, raising the number of letters to 23. But <c>, <q>, and <x> would still be excluded. To accommodate loan words and dialect variation, three additional digraphs would be added to the repertoire: <tj>, <dj>, and <sh>.

Diki-Kidiri proposed three additional tone writing rules; he calls these “une notation contextuelle des tons” (contextual tone notation). The word *contextual* does not refer to semantic context, but to the linguistic environment. Because the rules are only three in number, Diki-Kidiri refers to them as “petites retouches” (touch-ups). These changes would, however, affect a great number of words. The rules are as follows:

1. for mono-syllabic words with CVV structure (or C<sup>w</sup>V or C<sup>y</sup>V depending on one’s analysis), to reduce tone marking, <w> is to replace <u> and <o> and <y> is to replace <i> and <e> when the tone pattern is HH or MM. Since <w> and <y> would not bear tone diacritics, this would reduce the number of diacritics. For VV sequences with a LL or modulated tone pattern, <u>, <o>, <e> and <i> would be unaffected and tone marked as previously. Some changes affected would be:

e.g., <kûâ> → <kwâ>; <kûê> → <kwê>; <bîâ> → <byâ>; but, <hîo> remains <hîo>.

2. a) for words with CVCVCV(CV) structures, if the tone is High on each syllable, tone is to be marked only on the first syllable. The consonant of the second syllable would be doubled if not a digraph, and if a digraph, be preceded by a silent <h> to indicate the High tone is stable across the word:

e.g., <mbâsâmbâlâ> → <mbâssambala>; <mbîrîmbîrî> → <mbîrrimbiri>;

but <yângângû> → <yâhngangu> because of the digraph.

b) HH, HL, and LL(L)(L) tone pattern words would remain unaffected:

< sêngê > would remain < sêngê >; < lîngbi > would remain < lîngbi > ,

c) Words bearing low tone, regardless of word length, would be unaffected,

i.e., there would be no diacritics: e.g., < da >; < gapa >; < girisa >;

< govoroma > etc.

d) For marking mid-mid tone in CVCV(CV) words, the same two rules apply as

with a sequence of high tone: if the second consonant is a digraph, < h > is to be

inserted at the end of the first syllable, indicating a stable tone; if the second

consonant is not a digraph, the consonant is to be doubled to indicate the stable

tone. Neither of these rules are to be applied in words with palatalization or

labialization where a < y > or < w > replacement has already eliminated the need

for one of the diacritics. Mid tone is no longer to be marked in any of those

constructions, even in the presence of the derivational affix, which was previously

written *-ngö*.

e.g., < sêwä > → < sewwa >; < îri > → < irri >; < diköngö > → < dikkongo >;

but due to digraph, < löndö > → < lohndo >; < dëngö > → < dehngo >.

3. when the derivational affix /-ngö/ is added to stems with a CVV structure,

whether they qualified for the < w > or < y > substitution or not, tone is not to be

marked with any diacritic, but the same rule (doubling the consonant or h-

insertion is to be applied.

e.g., (bâa + -ngö) < bängö > → < baahngo >; (mâi + -ngö) < mäingö > → < mayhngo >.

Word division rules are to be as follows: after the < h > or between the doubled consonant:

< baah-ngo >; < mayh-ngo >; < sew-wa >; < ir-ri >.

There is more to the proposal. These rules suffice to raise concern. Ease of writing would be sacrificed under this system. Simplicity and systematicity do not seem to have a high priority. With such complicated rules Sango literacy efforts would be undermined. It would be possible to teach reading using this orthography. However, it is not likely that the average person could master spelling by these abstract rules requiring a high level of tone awareness (mid-tone in particular), which has been shown to be absent in most of the population. The question is not if a writing system can be taught, but what the optimal writing system would be. Usually the key is simplicity.

### *12.13.3 Discussion of the proposed new tone marking conventions*

Diki-Kidiri's proposal is based on the premise that the refusal to adopt the orthography decreed in 1984 is due to the number of diacritics it requires. Thus it focuses on a single criteria: the reduction of diacritics. In his document Diki-Kidiri (2006) provides several texts to show the effect of the new rules. Based on one of the texts, he calculates that this would reduce the number of diacritics by 10.74%. It is not likely that a 10.74% reduction of diacritics would adequately compensate for the degree of complication introduced by the proposed spelling rules.

The issues which needed addressing were (1) simplifying writing and (2) improving acceptability. The original objections did not relate to the presence of diacritics, but to writing tone in general, since

- a) tone has a low functional load,
- b) few people are conscious enough of tone to mark it, and
- c) tone pronunciations vary in context as well as from person to person and regionally.

None of these issues are addressed by this proposed change of tone marking conventions.

Writers who needed help from ILA personnel or choir directors to mark tone for them in the past would fare no better with this proposed system.

The proposed system is contrary to good principles of orthography design and reform in several respects:

- (1) ease of learning and teaching: The proposed rules do not seem to be intuitive or easy. A single sound has more than one representation (for example: < m<sup>â</sup>i > [m<sup>á</sup>ì] ‘grow’ vs. < mayhngo > [m<sup>ā</sup>iŋɔ̄] ‘growth’; < koto > [kòtò] ‘scratch’ vs. < mbetti > [mb<sup>ē</sup>tì] ‘book’). Word/morpheme images are not preserved and the same feature (repeating mid tone) is indicated in two different ways (< korrongo > ‘perforation’; < hihngango > ‘knowledge’), unrelated to other mid tone notation using dieresis. Rules are based on tone awareness, which has been shown to be weak among the Central African population.
- (2) aesthetics: diacritics are still present and the high frequency of doubled letters and presence of *h* gives texts an unusual look.
- (3) harmony with other languages: the change would not foster transfer to French or to and from any other CAR language for which orthographies have already been established.
- (4) respect for past investments: such a change nullifies all previous progress in working toward a standard, and would make all previous Sango materials outdated, including ILA’s thematic lexicons which have been produced over the years. This would represent a setback in efforts made and financial loss, and possibly could result in motivational loss as well.
- (5) community involvement: speakers of Sango and those involved in literacy efforts have not expressed a need for this type of change. Further, it does not address the recommendations they have made previously.

(6) simplicity and economy: words are lengthened unnecessarily by the addition of *h* and doubling of consonants; word division rules become more complex.

(7) excessive frequency of change: this can instill a lack of confidence in the agencies involved in language development decisions. This would be the fourth proposed change:

- a) Diki-Kidiri 1977, using IPA symbols for open vowels;
- b) Bouquiaux et al., using diacritics to differentiate open and close vowels, but tone not indicated;
- c) the orthography as decreed in 1984;
- d) the orthography as proposed in 2006 by Diki-Kidiri.

The new discussion on the Sango orthography raises the following questions:

What status did the 1984 decree have?

Can *anyone* propose changes? If so, what is the process?

Interestingly, the 1984 decree prescribed how tone should be written in Sango but it did not mention to what degree it should be written, leaving the issue open to interpretation and discussion. An addendum to the decree would appear unnecessary if a compromise could be reached concerning tone notation. It will be interesting to follow developments concerning the Sango orthography now that another option is on the table.

## **12.14 Relating the Sango orthography 1984 reform to other case studies**

### *12.14.1 Observations*

Twenty-one years is a relatively short time for introducing a change and implementing a written standard. Implementation *is* happening for the Sango official orthography because of the good will of the various publishers of Sango literature and an openness

among the consumers to accept change. Personal writing habits of older citizens are not likely to change, but governmental and international organizations could be held accountable and be expected to abide by the standard, especially in public signage. Mixing norms during a time of transition is normal. Not depriving people of literature using the system they are comfortable with, and allowing freedom of choice for publishing material for such clients helps ease the way for the change: It avoids upheaval during the time of transition, thus reducing resistance.

In the orthography of 1984 several factors came into play. Political, linguistic, and technical factors were weighted heavily, but educational factors were not given adequate consideration. Unfortunately, the orthography decisions were made by a small minority, mostly linguists. Testing of the orthography and seeking the input of stakeholders were neglected. Since the orthography was presented as a ready-made product, there seemed to be no room for dialogue and compromise. Little awareness raising took place. Fasold (1997:257) reports that in a 1988 survey, only 28% of the interviewed people “were able to recognize the official orthography, and many were not even aware that an ‘official’ orthography existed.” This is not surprising because the survey was taken only four years after the decree was published. Diffusion, not imposition, seemed to be the strategy, and that takes time.

Although people may not have eagerly welcomed the new system, resistance was not strong. Initial negative reactions gave way to embracing spelling simplifications as the benefits became clear. The move away from the French system was seen as part of gaining an independent ‘national identity.’ Since the new system was not forced on people, time allowed them to gradually overcome their initial emotional reactions. As the

generation attached to the former system passes away, the following generation comes into leadership. Differences between the French and Sango writing systems do not preclude mastery of both systems. Bi-literacy in French and Sango can be the expected norm if adequate investment is made to give the population access to quality education in both. The attitude that “Sango is for illiterates” need not be perpetuated.

#### *12.14.2 Action plans for implementation of the standard*

Regarding the promotion of the desired written standard, it would help if ILA, various Ministries, and governmental, non-governmental, and international agencies were more proactive. Increasing Sango print in the environment would help raise awareness of the new standard. (Koyt 1995:9 made reference to such action plans. Unfortunately, political upheavals interfered with their implementation.) Sango classes could be introduced in the public school system—if not in the lower grades, then for an hour or two per week for those who already read and write French. This need not be costly. Television could be instrumental in the city in sensitizing viewers to the official orthography. Story writing competitions could be held with appropriate publicity and prizes to motivate participation. Dictionaries need to be more accessible. A bilingual newspaper, if the content is of general interest and it is affordable for the majority, could help. Grand'Eury (1994:65) pointed out that non-governmental organizations are generally not aware of what other organizations are doing or of materials which already exist. Dialogue and cross-fertilization would be helpful. And last, but not least: It may be wise to reconsider full tone notation since experiments (Bird 1999a; 1999b) have indicated that this is sometimes not the optimal choice for African languages and experience in the CAR specifically has shown that the average Sango speaker struggles with writing tone.

### *12.14.3 Lessons learned from the Sango reform case study*

- It takes time to implement a writing system reform.
- Negative reactions might abate, given time.
- If there is goodwill among producers of literature, the transition can go smoothly without authoritative measures being taken.
- Awareness-raising must be done if the new system is to be implemented; material for awareness-raising must be prepared for a variety of audiences.
- ‘Right’ and ‘wrong’ are inappropriate terms when referring to writing systems in concurrent use.
- Parents and teachers need to be convinced of the benefits of using a particular language in school.
- Steps can be taken to increase the status of a language and foster positive attitudes toward it.
- During the time of transition, documents and public signage may display a mix of several writing systems.
- Through the use of a writing system, its strengths and weakness will come to light. Observation and testing will reveal what may need changing to maximize its effectiveness.
- If only ‘specialists’ can write a language, the community is not well served.

## **CHAPTER 13**

### **SUMMARY**

The process of designing a writing system and that of reforming one are very similar. The same factors come into play. However, the reform process is often more complicated: competing writing systems may exist; attachment to an existing system may limit reform options; and more people may want a say in the matter. In both writing system design and reform, it is possible to make mistakes and it is easy to offend.

This thesis has provided an overview of theory and case studies related to establishing a writing system for unwritten languages and modifying existing systems. Certain principles seem to foster popular acceptance and successful implementation. Orthography practitioners will do well to take note of these principles, summarized in 13.1–13.8:

1. Involve the stakeholders
2. Approach the writing system as a work in progress
3. Weight all the factors appropriately
4. Test the adequacy of the writing system and revise as necessary
5. Anticipate conflicting views
6. Be conservative
7. Exercise patience and diplomacy
8. Establish networks and support structures

### **13.1 Appropriate stakeholder involvement**

The following matters concerning involvement, roles and constraints need to be settled early on in the writing system design or reform process: Who will decide on orthography issues? What role will the government play? How much flexibility do their policies afford (assuming there are relevant policies)? If linguists are involved, what is to be their role (research, observation, consulting, intervention)? One of the most important questions is: How will the various stakeholders be given voice?

In the past, decisions were often made top-down. Now, in different parts of the world, an interactive consensus approach with key players consulting and dialoguing is proving to be more effective. A committee can be instrumental in the successful implementation of a writing system, especially if members display good negotiation and people skills.

An individual or a small interest group should not introduce change alone; it is best to respond to the desires of a larger sector of the population.

### **13.2 Orthography as a work in progress**

Treating an orthography as a work in progress instead of as a product (or trophy!) makes *time* less of a factor. It allows more flexibility and freedom to try something for a while. Using terms like ‘tentative,’ ‘provisionary’ and ‘experimental’ when referring to the orthography under development may help avert negative reactions and prepare the way for consequent revisions. A way to give and receive feedback from users during the early stage of implementation needs to be built into the process.

### **13.3 Weighing and balancing all the factors**

Linguistic facts that result from thorough analysis, such as an inventory of phonemes, and a morphophonemic analysis provide a starting point for orthography decisions.

Linguistic factors need to be balanced with other factors: political, socio-linguistic, educational, technical. Considerable attention should be given to the desire of future users. Identity issues are important. Choosing the wrong script or symbol, for example, can be very costly if it results in low motivation or resistance.

The writing system needs to be in line with mother tongue speaker intuition and preferences. Special consideration needs to be given to the relative ease and difficulty of learning and teaching the writing system. Proficiency in writing, not only in reading, should be the goal.

#### **13.4 Testing the writing system**

The process of designing and implementing a writing system and working toward its standardization is best conducted in stages. A planned, extended trial period can serve as a window of opportunity: By integrating formative evaluation and scheduled reviews and revisions into the process, the writing system can be modified over time. It is necessary to ensure that ‘the new’ is an improvement over ‘the old.’ Testing is vital. Teachers need to be given adequate opportunities to communicate their observations and experiences with the writing system during the trial period. Willingness to retract bad decisions is essential. Although revising a writing system to improve it is generally a positive step, one should avoid multiple successive revisions. Frequent or continual tinkering with a writing system can undermine people’s confidence in using the system, as well as confidence in the decision-makers.

### **13.5 Anticipating and dealing with conflict**

An error often committed is to expect an orthography to result from a neat scientific formula. Knowing better, Smalley (1964f:13–14) wrote:

Writing systems, after all, are cultural phenomena, used by people with feelings and emotions, with prejudices and fear. In fact it is surprising how much heat a controversy over orthographies can produce...

The ways in which these emotional factors will enter into the acceptance of a writing system in areas where new systems are being devised, remains one of the great critical problems in this field. Some workers have paid altogether too little attention to the problem, with occasional lamentable results.

Orthographies are a vulnerable part of language development. Conflict is more the norm than the exception. Those who do language development and face orthographic issues should anticipate some measure of conflict. They need to take proactive measures to avoid and resolve it, as well as be prepared to face it. Public relations work, clear communication, and cooperative planning can help avoid misunderstanding and mistrust.

Documenting discussions and decisions, as well as the rationale behind decisions, keeps track of the history of the writing system development or reform. It also provides a platform for accountability: People are more likely to follow-up on concerns and action plans which have been recorded. Observations and test results, formal and informal, require careful documentation since these may play a role in future decisions.

### **13.6 Taking a conservative approach**

In general, it is better to err on the side of conservatism, making small changes to what people are accustomed to, especially for the most frequent words. Baker (1997:141)

commented: “We...suspect that the different orthographic preferences which people have are in large measure determined by the system(s) to which they are accustomed.”

Only when there is major political reform or revolution is there likely to be enough impetus to implement major orthography changes. LePage commented: “The argument most strongly urged in favour of a revolutionary as opposed to an evolutionary approach to social change is that the matrix of the old social order must be smashed so that it does not persist in the institutions and stereotypes of the new” (1997:55). If revolution is not the current political agenda, modest revisions to a writing system are more likely to be accepted than major reforms. An incremental introduction of changes causes less upheaval for the literate population.

### **13.7 Exercising patience and diplomacy**

Implementing change takes time. It is unwise to rush the establishment of a standard; writing systems can evolve naturally. Initial negative reactions to a proposal might—with time—abate. The socio-political climate needs to be conducive to the change.

Fishman’s (1988:1647) comments indicate that patience is key during a writing system reform process:

The inertia of older readers and writers is particularly difficult to overcome, given that no dramatic change in function identity or reward structure is available to foster conformity with the modified system...a longish period of drift may result with the final outcome depending on natural processes of generational replacement...The young become the natural, unideologized carriers of modification, having been socialized accordingly by the schools, media, armed force, etc. whereas the old remain with the conventions to which they are accustomed. A three generational lag may transpire before modifications are generally implemented.

Even when there are strong motivations for change, one does well not to antagonize the sector who approves the status quo. It may be necessary to ease the transition for those who have to pay the price when the change is implemented. The formal education system is an effective venue for implementing a new or revised writing system. Public signage can contribute to acceptance of a system. Uniformity of spelling is desirable, but is not an absolute requirement.

### **13.8 Establishing support structures**

An orthography is to serve the language community. Providing opportunities to learn and use it is imperative. Schools are an effective venue for this for the younger generation. Corpus planning involves the development of various resources. Literature which interests adult members in the community will contribute to their motivation to learn and use the system. As individuals benefit from the system they are likely to develop positive attitudes toward it. Various agencies can play a part.

### **13.9 Conclusion**

Developing a writing system or reforming an existing one is a process. It calls for awareness of all the factors and sensitivity. The issues are complex. Acceptance of a writing system by governmental agencies, producers of print media, and the language community can be achieved. Orthography practitioners will do well to learn from history and apply the principles which favor acceptance and successful implementation. It is important to realize that potential users have the final word. Involving them in the process from the start is only logical.

## APPENDIX A

### Useful URLs

Nations of the world background information:

<http://www.historycentral.com/nationbynation>

<http://www.nationsencyclopedia.com>

<http://www.cia.gov/cia/publications/factbook>

Languages of the world:

<http://www.ethnologue.com>

Language endangerment:

<http://www.tooyoo.l.u-tokyo.ac.jp/Redbook/>

<http://www.ogmios.org/home.htm>

<http://www.sil.org/sociolx/ndg-lg-home.html>

Writing systems of the world:

<http://www.omniglot.com>

<http://www.krysstal.com/writing.html>

UNESCO resources:

<http://unesdoc.unesco.org/ulis>

Education statistics:

[http://portal.unesco.org/en/ev.php-url\\_id=33383&url\\_do=do\\_topic&url\\_section=201.html](http://portal.unesco.org/en/ev.php-url_id=33383&url_do=do_topic&url_section=201.html)

Technical Issues, Unicode and Fonts:

<http://www.unicode.org>

<http://scripts.sil.org>

<http://www.alanwood.net/unicode>

<http://www.travelphrases.info/fonts.html>

Sango language websites:

<http://sango.free.fr/>

<http://sango.ti.laso.free.fr/>

<http://fini.sango.free.fr/>

## APPENDIX B

Bibliographical information on the Sango printed matter used in *Section 12.12.1, Orthography in the literature*, comparing Sango writing practices:

**Primers:** (represented in Table 20)

Centre de Formation Agricole et Rurale. (n.d.) *E manda sango*. Bangui.

Centre de Formation Agricole et Rurale. (n.d.) *Zingo*. Bossangoa.

Église Baptiste de l'Ouest. 1986-1989. *Mandango ye: Syllabaire en sango*. (55 lessons in a series of four booklets).

Mission Évangélique des Frères. (n.d.). *Guingo Lègué ti Diko Mbèti ti Nzapa na Yanga ti Sango: Apprenant à lire la Bible en Sango*. (2 booklets).

(n.d.). *Gingo Lege Ti Diko Mbeti ti Nzapa na Yanga ti Sango: Apprenons à lire la Bible en Sango*. (revised edition).

Ndokoyo, Jean. *Kêtê Mbèti tî Manda Sängö*. Direction de l'Éducation Non-formelle.

Presse Biblique Baptiste. 1981. *Bama*. (1<sup>st</sup> book in a series; reprinted in 2000).

1982. *Assita*. (2<sup>nd</sup> book in the series; reprinted in 2000)

1982. *Kossi*. (3<sup>rd</sup> book in the series; reprinted in 2000)

Service National d'Alphabétisation e d'Éducation des Adultes. n.d. (pre-1990). *Manuel d'Alphabétisation*. Ministère de l'Éducation Nationale de la Jeunesse, des Sports et des Arts.

Société Internationale de Linguistique. 1993. *Âla gã ë manda tî dîko mbèti*. (2<sup>nd</sup> and 3<sup>rd</sup> book in the series.)

UNICEF-Education RCA. (n.d.). *Bûku tî dîko: Yongôro ngoi – Da ôko*.

Watch Tower Bible and Tract Society of Pennsylvania. 2000. *I Sala Ngangu Kue ti diko na ti sù mbeti*.

**Reading Material:** (represented in Table 21)

- Alliance Biblique en Centrafrique. 2001a. *Jonas: Wayanga-Nzapa so akpe yongoro na le ti Nzapa*.
- 2001b. *Mbeti ti Nzapa*. Fini Mbuki.
- 2001c. *Nzoni Tene*. Oku mbeti.
2002. *David: Kpengba maseka-koli*.
- Alliance Biblique Universelle. 1966. *Mbeti ti Nzapa*. La Sainte Bible en sango. Bangui:Le Foyer de la Bible.
- Conférence Épiscopale CentrAfrique. 1998. *E sala mbeto pepe, e duti tä wamabe na ya ti sorongo azo*.
- Église Baptiste de l'Ouest (n.d.). *Mbeti ti Wafango ye ti Kozo buku ti dikongo ye*. Livre du maître en sango. (teacher's manual for reading readiness lessons).
- 1986-1989. *Mbeti ti Wafango ye: Livre du maître*. (teacher's manuals to accompany the 55 literacy lessons).
- Église Baptiste de l'Ouest et Commission de Lutte contre le SIDA et les M.S.T. (n.d.) *E kanga lege na kobela ti SIDA*. Gamboula.
- Grace Brethren International Missions. (n.d.) *Le na Le na Jesus*. (translation of the Seven Signs of Jesus.)
- Koyamafou, Gédéon, et C. Yangue-Boyfini. 1993. *Fängö yäkä fi kârâkö*. Bouca: Projet de développement rural Bouca, volet alphabétisation fonctionnelle.
- Mololi, André, Gédéon Koyamofou, et Julien Kouanga, en collaboration avec Marcel Diki-Kidiri. 2005. *Lexique de suite bureautique*. Français-sango et sango-français. Bangui: Institut de Linguistique Appliquée, Université de Bangui.
- Paze-Lekyssane, Mathurin-Lazare, Jean-Ignace Manengou, and Chantal Gaudin. 1994. *Fango tene ti Nzapa na Be Afrika: Akete Molenge*. Commission nationale de catéchèse. Conférence Episcopale Centrafricaine.
- Poinsenet, Marie-Dominique. 1991. *Dunia ti awamokondo Thérèse na Jean ti Krua*. Translated by P. Charles Cencio. Bouar: Séminaire Enfant Jésus des Pères Carmes.

Presse Biblique Baptiste. 1968. *Bia ti Sepala Nzapa*. (5th printing, 1983).

1981. *Bama*. Mbèti ti Maître (1st book in the series).

1982. *Assita*. Mbèti ti Maître (2<sup>nd</sup> book in the series)

1982. *Kossi*. Mbèti ti Maître (3<sup>rd</sup> book in the series)

Société Internationale de Linguistique. 1993. *Lêgë tî kâi sasa na yângâ da*. Diarrhée:  
comment la traiter à la maison.

Société Biblique Centrafricaine 1995. *Buku ti Nzapa*. Fini Mbuki.

## REFERENCES CITED

- ACCT. 1980. Lexiques thématiques de l'Afrique Centrale: carnets d'enquête. (3 volumes). Paris.
- Agence Centrafrique Presse 2006. Le gouvernement a son hebdo.  
[http://www.webzinemaker.com/admi/m7/page.php3?num\\_web=41362&rubr=3&id=299405](http://www.webzinemaker.com/admi/m7/page.php3?num_web=41362&rubr=3&id=299405) (accessed June 19, 2006).
- Ager, Simon. 1998–2006a. Cyrillic alphabet. *Omniglot: Writing Systems and Languages of the World*. <http://www.omniglot.com/writing/cyrillic.htm> (accessed April 17, 2006).
- 1998–2006b. Kazakh. *Omniglot: Writing Systems and Languages of the World*.  
<http://www.omniglot.com/writing/kazakh.htm> (accessed April 17, 2006).
- 1998–2006c. Kyrgyz. *Omniglot: Writing Systems and Languages of the World*.  
<http://www.omniglot.com/writing/kirghiz.htm> (accessed April 17, 2006).
- 1998–2006d. Shavian alphabet. *Omniglot: Writing Systems and Languages of the World*.  
<http://www.omniglot.com/writing/shavian.htm> (accessed April 17, 2006).
- 1998–2006e. Sorang Sompeng script. *Omniglot: Writing Systems and Languages of the World*.  
<http://www.omniglot.com/writing/sorangsonpeng.htm> (accessed April 17, 2006).
- 1998–2006f. Syllabaries. *Omniglot: Writing Systems and Languages of the World*.  
<http://www.omniglot.com/writing/syllabaries.htm> (accessed April 19, 2006).
- 1998–2006g. Tigrinya alphabet and pronunciation. *Omniglot: Writing Systems and Languages of the World*.  
<http://www.omniglot.com/writing/tigrinya.htm> (accessed February 27, 2006).
- 1998–2006h. Vietnamese. *Omniglot: Writing Systems and Languages of the World*.  
<http://www.omniglot.com/writing/vietnamese.htm> (accessed July 1, 2006).
- 1998–2006i. Yi syllabary. *Omniglot: Writing Systems and Languages of the World*.  
<http://www.omniglot.com/writing/yi.htm> (accessed April 17, 2006).
- AIROÉ 2006. L'orthographe recommandée. <http://www.orthographe-recommandee.info/> (accessed June 15, 2006).
- Albaugh, Ericka. 2004. The Colonial Image Reversed: A New Politics of Language in African States. Paper presented to the Society of Comparative Research 2004 Graduate Student Retreat, University of California, San Diego, May 14–15, 2004.  
<http://socsci2.ucsd.edu/~aronatas/scrretreat/Albaugh.Ericka.doc> (accessed July 2 2006).
- Alliance Biblique en Centrafrique. 2001. *Mbeti ti Nzapa: Fini Mbuki*. (Le Nouveau Testament en Sango Courant, nouvelle version révisée). Bangui.
- Alliance Biblique Universelle 1966. *Mbeti ti Nzapa*. (La Sainte Bible en Sango). Foyer de la Bible.

- Anderson, Deborah. 2005. Recommendations for Creating New Orthographies. Unicode Technical Note 19. Unicode Inc.  
<http://www.unicode.org/notes/tn19/> (accessed June 19, 2006).
- Atlapedia Online. 1993–2005. *Countries A to Z: Central African Republic*. Latimer Clarke Corporation Pty. Ltd.  
<http://www.atlapedia.com/online/countries/cenafri.htm> (accessed Dec. 29, 2005).
- Augst, Gerhard. 1988. Schreibweise (Orthographie). In: Ulrich Ammon, Norbert Dittmar, and Klaus J. Mattheier (eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society (2)*. New York: Walter de Gruyter, 1134–1144.
1996. Germany: Script and Politics. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 765–768.
- Baddeley, Susan. 1995. The 'Dictionnaire historique de l'orthographe française': A landmark in historical spelling research. *Journal of the Simplified Spelling Society*. J 19(2), 13–17.  
<http://www.spellingsociety.org/journals/j19/french.php> (accessed June 23, 2006).
- Bagchi, Tista. 1996. Bengali Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 399–403.
- Baker, Philip. 1997. Developing Ways of Writing Vernaculars: Problems and Solutions in a Historical Perspective. In: Andrée Tabouret-Keller, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.), *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press, 93–141.
- Bakwerirama. 2004. Bakwerirama Adopts Orthography Standard. Bakwerirama: Spotlight on Bakweri Culture and Society. [http://www.bakweri.org/2004/03/bakwerirama\\_ado.html](http://www.bakweri.org/2004/03/bakwerirama_ado.html) (accessed January 25, 2006).
- Ball, Rodney. 1999. Spelling reform in France and Germany: attitudes and reactions. *Current Issues in Language and Society* 6(3and 4), 276–280.
- Baptist Mid-Missions. 2006. Central African Republic.  
<http://www.bmm.org/BMM/WhereWeServe/Africa/centralafricanrepublic.htm> (accessed June 19, 2006).
- Bauer, Thomas. 1996. Arabic Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 559–564.
- Bernard, H. Russell. 1999. Languages and Scripts in Contact: Historical Perspectives. In: Daniel Wagner, Richard L. Venezky and Brian V. Street (eds.), *Literacy: An International Handbook*. Boulder, Colorado: Westview Press, 22–28.
- Bernard, H. Russell, George N. Mbeh, and W. Penn Handwerker. 2002. Does marking tone make tone languages easier to read? *Human Organization* 61(4), 339–349.  
[http://www.findarticles.com/p/articles/mi\\_qa3800/is\\_200201/ai\\_n9025895](http://www.findarticles.com/p/articles/mi_qa3800/is_200201/ai_n9025895) (accessed July 6, 2006).
- Berry, Jack. 1958. The Making of Alphabets. Originally presented as a paper at the Proceedings of the 8th International Congress of Linguistics. In: Joshua A. Fishman (ed.), *Readings in the Sociology of Language*. 1968. The Hague: Mouton, 737–747.

1977. 'The Making of Alphabets' revisited. In: Joshua. A. Fishman (ed.), *Advances in the Creation and Revision of Writing Systems*. The Hague: Mouton, 3–16.
- Besner, Derek, and Marilyn Chapnik Smith. 1992. Basic Processes in Reading: Is the Orthographic Depth Hypothesis Sinking? In: Ram Frost and Leonard Katz (eds.), *Orthography, Phonology, Morphology, and Meaning*. Advances in Psychology 94. Amsterdam: North-Holland, 45–66.
- Beyer, Stephan V. 1992. *The Classical Tibetan Language*. Suny Series in Buddhist Studies. Albany, New York: State University of New York Press.
- Bird, Steven. 1999a. Strategies for representing tone in African language systems. *Written Language and Literacy* 2(1), 1–44.
- 1999b. When marking tone reduces fluency: An orthography experiment in Cameroon. *Language and Speech* 42(1), 83–115.
2001. Orthography and identity in Cameroon. *Written Language and Literacy* 4(2), 131–162.
- Bouquiaux, Luc, en collaboration avec Jean-Marie Kobozo, Maracel Diki-Kidiri, Jacqueline Vallet, et Anne Bhaghel. 1978. Dictionnaire Sango-Français et Lexique Français-Sango. Paris: SELAF.
- Briggs, Lucy T. 1985. Bilingual Education in Peru and Bolivia. In: Nessa Wolfson and Joan Manes (eds.), *Language of Inequality*. Contributions to the Sociology of Language (36). Berlin: Mouton Publishers, 297–310.
- Bright, William. 1996a. The Devanagari Script. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 384–390.
- 1996b. Kannada and Telugu Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 413–419.
1999. A matter of typology: Alphasyllabaries and abugidas. *Written Language and Literacy* 2(1), 45–55.
- Brown, R. McKenna (ed.). 2002. *Endangered Languages and Their Literatures*. Proceedings of the Sixth FEL Conference in Antigua, Guatemala, 8–10 August, 2002. Bath, England: Foundation for Endangered Languages.
- Buck, Marjorie J. 1973. Lesson in reading tone. *Notes on Literacy* 15, 9–15.
- Carlo, María S., and James M. Royer. 1999. Cross-language Transfer of Reading Skills. In: Daniel Wagner, Richard L. Venezky and Brian V. Street (eds.), *Literacy: An International Handbook*. Boulder, Colorado: Westview Press, 148–154.
- Carrington, Lawrence D. 1997. Social Contexts Conducive to the Vernacularization of Literacy. In: Andrée Tabouret-Keller, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.), *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press, 82–92.
- Casad, Eugene. 1974. Dialect intelligibility testing. Norman, Oklahoma: Summer Institute of Linguistics, University of Oklahoma.

1989. State of the art: Dialect survey 15 years later. In: Gloria E. Kindell (ed.), *Proceedings of the Summer Institute of Linguistics International Language Assessment Conference*, Horsleys Green, 23–31 May 1989. Dallas: Summer Institute of Linguistics.
- Chamberlain, Bradford L. 2004. *The Khengkha Orthography: Developing a Language in the Tibetan Scriptal Environment*, MA Thesis. Dallas: GIAL.
- Cheng, Chin-Chuan. 2000. Frequently-used Chinese Characters and Language Cognition. *Studies in the Linguistic Sciences* 30(1), 108–117.
- Christus Rex. 1994. Convent of Pater Noster. <http://198.62.75.1/www1/pater/index.html> (accessed May 24, 2006).
- Comrie, Bernard. 1996a. Adaptations of the Cyrillic Alphabet. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 700–726.
- 1996b. Script Reform in and after the Soviet Union. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 781–784.
- Constable, Peter. 2003. Understanding characters, keystrokes, codepoints and glyphs: Encoding and working with multilingual text. Dallas: SIL International. [http://scripts.sil.org/cms/scripts/page.php?site\\_id=nrsi&item\\_id=IWS-Chapter02](http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=IWS-Chapter02) (accessed June 21, 2006).
- Cook, Vivian, and Benedetta Bassetti (eds.). 2005. *Second Language Writing Systems*. Second Language Acquisition 11. Clevedon, England: Multilingual Matters Ltd.
- Coulmas, Florian. 1989. *The Writing Systems of the World*. Oxford: Blackwell.
1996. *The Blackwell Encyclopedia of Writing Systems*. Oxford: Blackwell.
1999. Development of Orthographies. In: Daniel Wagner, Richard L. Venezky and Brian V. Street (eds.), *Literacy: An International Handbook*. Boulder, Colorado: Westview Press, 137–142.
2000. The Nationalization of Writing. *Studies in the Linguistic Sciences* 30(1), 47–59.
2003. *Writing Systems: An Introduction to their Linguistic Analysis*. Cambridge: Cambridge University Press.
- Crystal, David. 1997. *A Dictionary of Linguistics and Phonetics*. 4<sup>th</sup> edition. Cambridge, Mass: Blackwell Publishers.
2000. *Language Death*. Cambridge: Cambridge University Press.
- Cubberley, Paul. 1996. The Slavic Alphabets. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 346–355.
- Cummins, J. 1979. Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research* 49, 222–251.
1984. Wanted: A Theoretical Frame for Relating Language Proficiency to Academic Achievement among Bilingual Students. In: C. Rivera (ed.), *Language Proficiency and Academic Achievement*. Clevedon: Multilingual Matters, 2–19.
1989. Language and literacy acquisition in bilingual contexts. *Journal of Multilingual and Multicultural Development* 10 (1), 17–31.

1991. Interdependence of first- and second-language proficiency in bilingual children. In: E. Bialystok (ed.), *Language Processing in Bilingual Children*. New York: Cambridge University Press, 70–89.
- Dagnan, Christian Roger. 1998. Längö tí Fütängö Âwakua tí Letäa (Mâleyömbö): Nzö abe awe. Bangui: Davungä Bûku tí Bêafrîka (BBA).
- Daniels, Peter. 1996a. Aramaic Scripts for Aramaic Languages. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 499–504.
- 1996b. Grammatology, Introduction to Part I of *The World's Writing Systems*, Peter Daniels and William Bright (eds.). New York: Oxford University Press, 1–2.
- 1996c. The Invention of Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 579–586.
- 1996d. Middle Eastern Writing Systems: Introduction to Part VIII. Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 485–486.
- 1996e. The Study of Writing Systems. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 3–17.
2001. Writing Systems. In: Mark Aronoff and Janie Rees-Miller (eds.), *The Handbook of Linguistics*. Malden, Massachusetts: Blackwell Publishers, 43–80.
- Daniels, Peter, and William Bright (eds.). 1996. *The World's Writing Systems*. New York: Oxford University Press.
- Day, Richard R. 1985. The Ultimate Inequality: Linguistic Genocide. In: Nessa Wolfson and Joan Manes (eds.), *Language of Inequality*. Contributions to the Sociology of Language (36). Berlin: Mouton Publishers 163–181.
- Debold, Kathleen Joan. 1986. Mabôko na Mabôko. Bangui: Peace Corps in Central African Republic.
- DeFrancis, John. 1977. Language and Script Reform in China. In: Joshua. A. Fishman (ed.), *Advances in the Creation and Revision of Writing Systems*. The Hague: Mouton, 121–148.
- Diki-Kidiri, Marcel. 1982. Kua tí Ködörö: Mbëti tí Hînga na Sêndâlêkëngö- Ködörö) / Le devoir national: Initiation à l'instruction civique. Paris: SELAF-ACCT.
1994. Developper le sango en tant que langue officielle. In: Helma Pasch (ed.), *Sango: The National Official Language of the Central African Republic*. Proceedings of the Colloquium "The Status and Uses of Sango in the Central African Republic," Cologne, September 3–4, 1992. Köln: Rüdiger Köppe Verlag, 25–29.
1998. *Dictionnaire Orthographique du Sängö*. Bangui: Bûkû tí Bêafrîka (BBA).
2002. Towards real linguistic and cultural diversity in cyberspace. Communication and Information: Points of View Archives. UNESCO. [http://portal.unesco.org/ci/en/ev.php-url\\_id=6628&url\\_do=do\\_topic&url\\_section=201.html](http://portal.unesco.org/ci/en/ev.php-url_id=6628&url_do=do_topic&url_section=201.html) (accessed June 15, 2006).
2006. Pour une notation contextuelle des tons en Sängö. Working paper.

- Dinh-Hoà, Nguyễn. 1996. Vietnamese. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 691–695.
- Downing, John. 1964. *The Initial Teaching Alphabet*. London: Cassell & Company.
- Durie, Mark. 1987. The Orthographic Representation of Nasal Vowels in Acehnese. In: Philip A. Luelsdorff (ed.), *Orthography and Phonology*. Philadelphia: John Benjamins, 131–150.
- Dutcher, Nadine. 2004. *Expanding Educational Opportunity in Linguistically Diverse Societies*. Washington, DC: Center for Applied Linguistics.
- Dye, Sally F. 1988. New literates reading aloud for audience comprehension: The Bahinemo case. *Notes on Literacy* 53, 41–44.
- Easton, Catherine. 2003. Alphabet Design Workshops in PNG: A Community-Based Approach to Orthography Development. Paper presented at the Language Development Conference, Institute of Language and Culture for Rural Development, Hahidol University, Bangkok. [http://www.sil.org/asia/ldc/parallel\\_presentations.html](http://www.sil.org/asia/ldc/parallel_presentations.html) (accessed June 14, 2006).
- Easton, Catherine, and Diane Wroge. 2002. *Manual for Alphabet Design through Community Interaction*. Ukarumpa, Papua New Guinea: SIL.
- Eeste Keele Instituut. 2006. Letter database: languages, character sets, names; Data on languages using cyrillic script. <http://www.eki.ee/letter/languages/cyrillic.html> (accessed Feb. 16, 2006)
- Eira, Christina. 1998. Authority and discourse: Towards a model of orthography selection. *Written Language and Literacy* 1(2), 171–224.
- Emejulu, James D., and Yolande Nzang-Bie. 1999. Sociolinguistic Aspects of Initial Language Development. In: Stephen L. Walter, Patricia M. Davis, and Ron Morren (eds. and comps.), *Explaining Multilingual Education: Information on some tough questions*. Working papers, Dallas: SIL, 38–64.
- Emmaus Correspondence School (1997–2006). ECS Emmaus Correspondence School – Home. <http://www.emmaus.name/ewc/index.php> (accessed June 21, 2006).
- Fasold, Ralph W. 1997. Motivations and Attitudes Influencing Vernacular Literacy: Four African Assessments. In: Andrée Tabouret-Keller, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.), *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press, 246–270.
- Feldman, Laurie Beth, and Dragana Barac-Cikoja. 1996. Serbo-Croatian: A Biscrptal Language. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 769–772.
- Filali-Ansary, Houda. 2005. Amazigh kids get role model. Morocco Times. 8. July, 2005. <http://www.moroccotimes.com/paper/article.asp?idr=11&id=7957> (accessed June. 15, 2006).
- Fine, Cathryn. 2003. Extendibility in Bouyei Orthography Design: A Multilectal Approach. M.A. Thesis. Dallas:GIAL.
- Finh Saeteurn, Fahm. n.d. About the Iu-Mien: Language. On Line Lessons. <http://mienh.net/language/tones.html> (accessed June 16, 2006).

- Fishman, Joshua A. (ed.). 1968. *Readings in the Sociology of Language*. The Hague: Mouton.
- (ed.). 1977. *Advances in the Creation and Revision of Writing Systems* The Hague: Mouton.
1988. The Development and Reform of Writing Systems. In: Ulrich Ammon, Norbert Dittmar, and Klaus J. Mattheier (eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society* (2). New York: Walter de Gruyter, 1643–1650.
- Fortune, George. 1957. *Elements of Shona*. London, Longmans, Green and Co., Ltd.
1972. *A Guide to Shona Spelling*. Salisbury, Rhodesia: Longman Rhodesia.
- Frantz, Donald G., and Norma Jean Russel. 1989. *Blackfoot Dictionary of Stems, Roots, and Affixes*. Toronto: University of Toronto Press.
- Frost, R., and S. Bentin. 1992. Reading Consonants and Guessing Vowels: Visual Word Recognition in Hebrew Orthography. In: R. Frost and L. Katz (eds.), *Orthography, Phonology, Morphology, and Meaning*. Advances in Psychology 94. Amsterdam: North-Holland, 27–44.
- Frost, R., and L. Katz (eds.). 1992. *Orthography, Phonology, Morphology, and Meaning*. Advances in Psychology 94. Amsterdam: North-Holland.
- Gadelii, Karl E. 2004. Annotated Statistics on Linguistic Policies and Practices in Africa. (revised) Paper presented at the Intergovernmental Conference on Linguistic Policies in Africa, Harare March 17–21, 1997 organized by UNESCO.  
<http://www.african.gu.se/downloads/gadeliiireport.pdf> (accessed June 15, 2006).
- Gair, James W. 1996. Sinhala Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 408–412.
- Gardner-Chloros, Penelope. 1997. Vernacular Literacy in New Minority Settings in Europe. In: Andrée Tabouret-Keller, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.), *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press, 189–221.
- Geerts, G., J. van den Broeck, and A. Verdoodt 1977. Successes and Failures in Dutch Spelling Reform. In: : Joshua. A. Fishman (ed.), *Advances in the Creation and Revision of Writing Systems*. The Hague: Mouton, 189–245.
- Gelb, I.J. 1952. *A Study of Writing*. (2nd ed. 1963). Chicago: University of Chicago Press.
- Gerbault, Jeannine. 1994. The Development of Sango Literacy in Central African Society. In: Helma Pasch (ed.), *Sango: The National Official Language of the Central African Republic*. Proceedings of the Colloquium “The Status and Uses of Sango in the Central African Republic,” Cologne, September 3–4, 1992. Köln: Rüdiger Köppe Verlag, 71–84.
1997. Pedagogical Aspects of Vernacular Literacy. In: Andrée Tabouret-Keller, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.), *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press, 142–185.
- Germano, David. 2004. Guide to Tibetan Orthography. The Tibetan and Himalayan digital library <http://iris.lib.virginia.edu/tibet/xml/showEssay.php?xml=/education/tllr/xml/orthography.xml&l=d1e209> (accessed June 15, 2006).

- Giersberg, Dagmar. 2006. Goethe-Institut Dossier The German Language. The End of the Debate. <http://www.goethe.de/kue/lit/dos/dds/en630493.htm> (accessed June 16, 2006).
- Gill, Harjeet Singh. 1996. The Gurmukhi Script. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 395–398.
- Gilley, Leoma C. (ed.). 2004. *Occasional Papers in the Study of Sudanese Languages 9*. Entebbe, Uganda: SIL-Sudan.
- Gordon, Raymond G., Jr. (ed.), 2005. *Ethnologue: Languages of the World*, Fifteenth edition. Dallas, Texas: SIL International. <http://www.ethnologue.com>
- Goswami, Usha. 2005. Synthetic phonics and learning to read: a cross-language perspective. *Educational Psychology in Practice*. 21(4), 273–282.
- Grace Brethren International Missions. 2005. The Seven Signs of Jesus. Resources Downloads. Strategic Alliance for Leadership Training. <http://www.gbim.org/home/resources/downloads/7%20signs%20SANGO.doc> (accessed May 24, 2006).
- Grand'Eury, Sylvie. 1994. Le Sango au service de l'animation des structures communautaires rurales et urbaines. In: Helma Pasch (ed.), *Sango: The National Official Language of the Central African Republic*. Proceedings of the Colloquium "The Status and Uses of Sango in the Central African Republic," Cologne, September 3–4, 1992. Köln: Rüdiger Köppe Verlag, 57–69.
- Grimes, Joseph E., Augustus B. Marvieh, and Amy Bauernschmidt. 1964. Several Kru Orthographies. *Orthography Studies: Articles on New Writing Systems*. Helps for Translators 6. London: United Bible Societies, 114–119.
- Gudschinsky, Sarah. 1973. *A Manual of Literacy for Preliterate Peoples*. Ukarumpa, PNG: SIL.
- Haile, Getatchew. 1996. Ethiopic Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 569–576.
- Hannan, M., S. J. 1968. *Standard Shona Dictionary*. London: Macmillan and Co Ltd.
- Harris, Theodore L. and Richard E. Hodges (eds.). 1995. *The Literacy Dictionary: The vocabulary of Reading and Writing*. Newark, Delaware: International Reading Ass'n.
- Harrison, P.A. 1998. Yoruba babies and unchained melody. *UCL Working Papers in Phonetics* 10, 33–52.
1999. The acquisition of Phonology in the First Year of Life. PhD Dissertation, University College London.
2000. Acquiring the phonology of lexical tone in infancy. *Lingua* 110, 581–616.
- Hartell, Rhonda L. 1993 (ed). *Alphabets of Africa*. Dakar: UNESCO-Dakar Regional Office and Summer Institute of Linguistics.
- Hartmann-So, Helga, and David Thomas. 1981. Morphemic Writing in Daai Chin. *Notes on Literacy* 36, 30–31.
- Hatfield, Deborah H. 1974. Lecture notes. Multilingual Societies. *LinguaLinks Library* version 5. 2002. Dallas: SIL International. (CD-ROM) at: Glossary of sociolinguistic terms/Glossary (Sociolinguistics)/What is a recorded text test (RTT)?

- Haugen, Einar. 1985. The Language of Imperialism: Unity or Pluralism? In: Nessa Wolfson and Joan Manes (eds.), *Language of Inequality*. Contributions to the Sociology of Language (36). Berlin: Mouton Publishers, 3–17.
- Heine, Bernd, and Derek Nurse (eds.). 2000. *African Languages: An Introduction*. Cambridge: Cambridge University Press.
- Henderson, Leslie. 1984. Writing Systems and Reading Processes. In: Leslie Henderson (ed), *Orthographies and Reading: Perspectives from Cognitive Psychology, Neuropsychology and Linguistics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 11–24.
- 1984 (ed). *Orthographies and Reading: Perspectives from Cognitive Psychology, Neuropsychology and Linguistics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Hoberman, Robert D. 1996. Modern Aramaic. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 504–510.
- Hollman, John. 2004. Preface to Leoma C. Gilley (ed.), *Occasional Papers in the Study of Sudanese Languages*. 9. Entebbe, Uganda: SIL-Sudan.
- Hornberger, Nancy H., and Kendall A. King. 1998. Authenticity and unification in Quechua language planning. *Language, Culture and Curriculum* 11 (3), 390–410.  
<http://www.multilingual-matters.net/lcc/011/0390/lcc0110390.pdf>  
 (accessed June 16, 2006).
- Hung, D. L., O.J. Tzeng, and A.K. Tzeng. 1992. Automatic Activation of Linguistic Information in Chinese Character Recognition. In: R. Frost and L. Katz (eds.), *Orthography, phonology, morphology, and meaning*. Amsterdam: North Holland, 119–130.
- Hyman, Larry M. 2003. Segmental Phonology. In: Derek Nurse and Gérard Philippson (eds.). *The Bantu Languages*. New York: Routledge, (42–58).
- ILV. 1973. Si<sup>2</sup>nche<sup>2</sup>xro<sup>2</sup>an<sup>2</sup>ni<sup>2</sup> Nque<sup>2</sup>he<sup>2</sup> Na<sup>1</sup>xrjon<sup>1</sup>. (alphabet book). Instituto Lingüístico de Verano: Mexico.
1980. Sinchecuáxinni tí xrooni imá náxrjón nixja: Abecedario en Popoloca en el idioma popoloca del oriente. Mexico: Instituto Lingüístico de Verano.
- Internet Christian Library. 2004. Selected Works of Martin Luther. Project Wittenberg. ILCnet. <http://www.iclnet.org/pub/resources/text/wittenberg/wittenberg-luther.html>  
 (accessed June 23, 2006).
- Jensen, Hans. 1970. *Sign, Symbol and Script: An Account of Mans Efforts to Write*. Third Revised and Enlarged Edition, translated from the German by George Unwin (original title: Die Schrift in Vergangenheit und Gegenwart. 1958.) London: George Allen and Unwin.
- Johnson, Sally. 2005. *Spelling Trouble? Language, Ideology and the Reform of German Orthography*. Clevedon: Multilingual Matters.
- Johnstone, Patrick, and Jason Mandryk. 2001. *Operation World*. (6th edition). Waynesboro, Georgia: Paternoster USA.
- Joshi, R. Malatesha, and P.G. Aaron (eds.). 2006. *Handbook of Orthography and Literacy*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

- Jusczyk, Peter W. 1997. *The Discovery of Spoken Language*. Language, Speech and Communication. Cambridge, Mass.: MIT Press.
- Kalck, Pierre. 1980. *Historical Dictionary of the Central African Republic* (Translated by Thomas O'Toole). Metuchen, New Jersey: The Scarecrow Press.
- Kapaj, Luigi. 2002. Writing Mongol in Uighur Script. <http://www.viahistoria.com/SilverHorde/main.html?research/UighurScript.html> (accessed June 21, 2006).
- Karan, Elke. 1991 Lire et écrire le Sango – sans difficulté. Bangui: SIL in cooperation with ILA.
1998. Quelques observations faites par la SIL: Recommendations de la SIL. Bangui: SIL.
- Karan, Mark E. 2001. *The Dynamics of Sango Language Spread*. Dallas: SIL International.
- Katz, Leonard, and Ram Frost. 1992. The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. In: Ram Frost and Leonard Katz (eds.), *Orthography, Phonology, Morphology, and Meaning*. Advances in Psychology 94. Amsterdam: North-Holland, 67–84.
- Kaye, Alan S. 1996. Adaptations of Arabic Script. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 743–762.
- Kim, Chin W. (ed.). 2000. *Studies in the Linguistic Sciences* 30(1).
- Kindell, Gloria E. (ed.) 1989. Language assessment criteria: Conference recommendations. Proceedings of the Summer Institute of Linguistics International Language Assessment Conference, Horsleys Green, 23–31 May 1989 Dallas: Summer Institute of Linguistics.
- Kisseberth, Charles, and David Odden. 2003. Tone. In: Derek Nurse and Gérard Philippon (eds.), *The Bantu Languages*. New York: Routledge, 59–70.
- Kjeilen, Toren. 1996–2005. Cuneiform writing. Encyclopaedia of the Orient. LexicOrient. <http://i-cias.com/e.o/cuneiform.htm> (accessed July 8, 2006).
- Koda, Keiko. 2005. Learning to Read Across Writing Systems: Transfer, Metalinguistic Awareness, and Second-language Reading Development. In: Vivian Cook and Benedetta Bassetti (eds.), *Second Language Writing Systems*. Second Language Acquisition 11. Clevedon: Multilingual Matters, 311–334.
- Kosonen, Kimmo. 2003. Testing Chong orthography in Thailand. Paper presented at the Language Development Conference, Institute of Language and Culture for Rural Development, Mahidol University, Bangkok. [http://www.sil.org/asia/lcd/parallel\\_presentations.html](http://www.sil.org/asia/lcd/parallel_presentations.html) (accessed June 16, 2006).
- Koyt, Michel Marie. 1994. La situation du sango en République Centrafricaine. In: Helma Pasch (ed.), *Sango: The National Official Language of the Central African Republic*. Proceedings of the Colloquium “The Status and Uses of Sango in the Central African Republic,” Cologne, September 3–4, 1992. Köln: Rüdiger Köppe Verlag, 13–23.
1995. La terminologie dans le contexte de l'aménagement linguistique de la République Centrafricaine. Terminologies nouvelles. Rint 14, 6–9 <http://www.cfwb.be/franca/termin/charger/rint14.pdf> (accessed May 23, 2006).

- Kutsch Lojenga, Constance. 1996. Participatory research in linguistics. *Notes on Linguistics* 73, 13–27.
- Kutsch Lojenga, Constance, and Christine Waag. 2004. The Sounds and Tones of Fur. In: Leoma C. Gilley (ed.), *Occasional Papers in the Study of Sudanese Languages*. 9. Entebbe, Uganda: SIL-Sudan, 1–25.
- Kwan-Terry, Anna, and K.K. Luke. 1997. Tradition, Trial, and Error: Standard and Vernacular Literacy in China, Hong Kong, Singapore, and Malaysia. In: Andrée Tabouret-Keller, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.), *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press, 271–315.
- Ladefoged, Peter. 2000. *A Course in Phonetics* (4th edition). Boston: Thomson Wadsworth.
- Le Page, R.B. 1997. Political and Economic Aspects of Vernacular Literacy. In: Andrée Tabouret-Keller, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.), *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press, 23–81.
- Lee, Jennifer. 2001. Where the PC is mightier than the pen. *New York Times*. February 1, 2001. ProQuest Historical Newspapers, G1.
- Lee, Joel. 2005. Unicode in SIL. Powerpoint presentation. Dallas: Non-Roman Script Initiative, SIL International.
- Linell, Per. 1979. *Psychological Reality in Phonology: A Theoretical Study*. Cambridge Studies in Linguistics 25. Cambridge University Press.
- Litteral, Robert. 1999. Four Decades of Language Policy in Papua New Guinea: The Move Towards the Vernacular. SIL Electronic Working Papers 1999–001. First presented to the annual meeting of the Linguistic Society of Papua New Guinea, Madang 1995 <http://www.sil.org/silewp/1999/001/silewp1999-001.html> (accessed January 2006).
- Maddieson, Ian. 2003. The Sounds of the Bantu Languages. In: Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*. New York: Routledge, 15–41.
- Mafundikwa, Saki. 2004. *Afrikan Alphabets: The Story of Writing in Afrika*. West New York, New Jersey: Mark Batty Publisher.
- Magwa, Wiseman. 2002. The Shona writing system: an analysis of its problems and possible solutions. *Zambezia* 29(1). <http://archive.lib.msu.edu/DMC/African%20Journals/pdfs/Journal%20of%20the%20University%20of%20Zimbabwe/vol29n1/juz029001002.pdf>. (accessed July 8, 2006).
- Mahapatra, B.P. 1996. Oriya Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 404–407.
- Malone, Susan. 2004. *Manual for developing Literacy and Adult Education Programs in Minority Language Communities*. Bangkok: UNESCO. <http://unesdoc.unesco.org/images/0013/001351/135164e.pdf> (accessed June 16, 2006).
- Mann, Michael, and David Dalby. 1987. *A Thesaurus of African Languages: A Classified and Annotated Inventory of the Spoken Languages of Africa with an Appendix on Their Written Representation*. London: Hans Zell Publishers.

- Marquardt, Susanne. Transliteration und Retrieval: Zur Problematik des Auffindens Hebräischsprachiger Medien in Online-Katalogen. Berliner Handreichungen zur Bibliothekswissenschaft 157. Berlin: Institut für Bibliothekswissenschaft, Humboldt-Universität zu Berlin. <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h157/h157.pdf> (accessed July 6, 2006).
- Martinet, Jeanne. 2005. Je parle, donc j'écris: Appui à l'alphabétisation par les TIC. Liège Demain. [http://www.liegedemain.be/projets/projets\\_jeparledoncjecris.html](http://www.liegedemain.be/projets/projets_jeparledoncjecris.html) (accessed June 13, 2006).
- Masica, Colin P. 1996. South Asia: Coexistence of Scripts. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 773–776.
- Mattingly, Ignatius G. 1992. Linguistic Awareness and Orthographic Form. In: Ram Frost and Leonard Katz (eds.), *Orthography, Phonology, Morphology, and Meaning*. Advances in Psychology 94. Amsterdam: North-Holland, 11–26.
- McCreedy, David Roger. 2005. Tifinagh Unicode Fonts. *David McCreedy's Gallery of Unicode Fonts*. [http://www.travelphrases.info/gallery/Fonts\\_Tifinagh.html](http://www.travelphrases.info/gallery/Fonts_Tifinagh.html) (accessed June 12, 2006).
- Mfonyam, Joseph Ngwa. 1988. Tone in Orthography: The Case of Bafut and Related Languages. Ph.D. dissertation. Yaoundé, Cameroon: University of Yaoundé.
- Ministerio de Educación. 1989. *Pishiti (La Pinsha) Libro de Lectura y Escritura No 4 Machiguenga con traducción al castellano*. Lima: Republica Peruana, Ministerio de Educación, Educación Bilingüe de la Selva.
- Mistry, P.J. 1996. Gujarati Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 391–394.
- Mohanam, K.P. 1986. *The Theory of Lexical Phonology*. Studies in Natural Language and Linguistic Theory. Dordrecht: Reidel Publishing Company.
- Munro, Pamela. 2003. Orthography and Language Preservation. Paper presented at the 2003 meeting of the American Anthropological Association, Chicago, Nov. 19–23, 2003.
- Nagy, William E., and Richard C. Anderson. 1999. Metalinguistic Awareness and Literacy Acquisition in Different Languages. In: Daniel Wagner, Richard L. Venezky and Brian V. Street (eds.), *Literacy: An International Handbook*. Boulder, Colorado: Westview Press, 155–160.
- Nichols, John D. 1996. The Cree Syllabary. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 599–611.
- Nida, Eugene. 1964a. Orthographic Problems in Yipounou. *Orthography Studies: Articles on New Writing Systems*. Helps for Translators 6. London: United Bible Societies, 148–155.
- 1964b. Practical Limitations to a Phonemic Alphabet. *Orthography Studies: Articles on New Writing Systems*. Helps for Translators 6. London: United Bible Societies, 22–30.
- Nurse, Derek, and Gérard Philippson (eds.). 2003. *The Bantu Languages*. New York: Routledge.
- O'Connor, M. 1996. The Berber Scripts. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 112–116, 119.

- Olson, David R. 1999. Literacy and Language Development. In: Daniel Wagner, Richard L. Venezky and Brian V. Street (eds.), *Literacy: An International Handbook*. Boulder, Colorado: Westview Press, 132–136.
- Pasch, Helma. 1994. Status and Role of Sango in the Central African Republic as Contrasted to French. In: Helma Pasch (ed.), *Sango: The National Official Language of the Central African Republic*. Proceedings of the Colloquium “The Status and Uses of Sango in the Central African Republic,” Cologne, September 3–4, 1992. Köln: Rüdiger Köppe Verlag, 85–97.
- Pasch, Helma (ed.). 1994. *Sango: The National Official Language of the Central African Republic*. Proceedings of the Colloquium “The Status and Uses of Sango in the Central African Republic,” Cologne, September 3–4, 1992. Köln: Rüdiger Köppe Verlag.
- Pederson, Eric. 2003. Mirror-image discrimination among nonliterate, monoliterate and biliterate Tamil subjects. *Written Language and Literacy* 6(1), 71–91.
- Perfetti, Charles A., Laurence Rieben, and Michel Fayol (eds.). 1997. *Learning to Spell*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Person, Kirk R. 1999. Writing Bisu: A Community-Based Approach to Orthography Development. In: Graham W. Thurgood (ed.), *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*. Arizona State University.
- Persson, Janet. 2004. Bongo-Bangirmi Languages in Sudan. In: Leoma C. Gilley (ed.), *Occasional Papers in the Study of Sudanese Languages*. 9. Entebbe, Uganda: SIL-Sudan, 77–84.
- Pike, Kenneth L. 1947. Phonemics: A Technique for Reducing Languages to Writing. Ann Arbor: University of Michigan Press.
- Powlison, Paul S. 1968. Bases for formulating an efficient orthography. *The Bible Translator* 19(2), 74–91.
- Presse Biblique Baptiste. 1968. *Bia ti Sepala Nzapa*. Sibut, République Centrafricaine.
- Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Studies in Natural Language and Linguistic Theory 4. Dordrecht: Reidel Publishing Company.
- Rabin, Chaim. 1971. Spelling Reform-Israel 1968. In: Joan Rubin and Bjorn Jernudd (eds.), *Can Language Be Planned? Sociolinguistic Theory and Practice for Developing Nations*. Honolulu: University Press of Hawaii, 95–121. Reprinted in Joshua. A. Fishman (ed.), 1977. *Advances in the Creation and Revision of Writing Systems*. The Hague: Mouton, 149–76.
- Ratliff, Martha. 1996. The Pahawh Hmong Script. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 619–624.
- Rayner, Keith, and Alexander Pollatsek. 1989. *The Psychology of Reading*. Englewood Cliffs, New Jersey: Prentice Hall.
- République Centrafricaine. 1984. Décret n° 84.025 Fixant l'orthographe officielle du Sango. Bangui: Présidence du Comité Militaire de Redressement National.

1990. Accord de Coopération entre le gouvernement de la République Centrafricaine et la Société Internationale de Linguistique. Bangui: le Haut Commissaire à la Présidence de la République Chargé de la Recherche Scientifique et Technique.
- Research Library Group. 2000. Library of Congress, Other US Libraries Join International Community on Use of Pinyin. RLG News Release, 10/2000.  
<http://www.rlg.org/legacy/pr/pr2000-pinyin.html> (accessed June 14, 2006).
- Rice, Keren D. 1995. Developing Orthographies: The Athapaskan Languages of the Northwest Territories, Canada. In: Insup Taylor and David R. Olson (eds.), *Scripts and Literacy: Reading and Learning to Read Alphabets, Syllabaries and Characters*. Dordrecht, Netherlands: Kluwer Academic Publishers, 77–94.
- Ring, Andrew. 1989. Three Case Studies Involving Dialect Standardization Strategies in Northern Ghana. In: Gloria E. Kindell (ed.), Proceedings of the Summer Institute of Linguistics International Language Assessment Conference, Horsleys Green, 23–31 May 1989. Dallas: Summer Institute of Linguistics, 281–87.
- Ritner, Robert K. 1996. Egyptian Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 73–84.
- Robinson, Clinton and Karl Gadelii. 2003. *Writing Unwritten Languages: A Guide to the Process*. UNESCO. (full text available: [http://portal.unesco.org/education/en/ev.php-url\\_id=30738&url\\_do=do\\_printpage&url\\_section=201.html](http://portal.unesco.org/education/en/ev.php-url_id=30738&url_do=do_printpage&url_section=201.html) (accessed May 25, 2006).
- Rogers, Henry. 1995. Optimal Orthographies. In: Insup Taylor and David R. Olson (eds.), *Scripts and Literacy: Reading and Learning to Read Alphabets, Syllabaries and Characters*. Dordrecht, Netherlands: Kluwer Academic Publishers, 31–43.
2005. *Writing Systems: A Linguistic Approach*. Oxford &, Maine: Blackwell Publishing.
- Salomon, Richard G. 1996. Brahmi and Kharoshthi. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 373–383.
2000. Typological Observations on the Indic Script Group and its Relationship to Other Alphasyllabaries *Studies in the Linguistic Sciences* 30(1), 87–103.
- Samarin, William J. 1955. Sango, an African Lingua Franca. *Word* 11(2), 254–367.
1963. *A Grammar of Sango*. Hartford, Connecticut: Hartford Seminary Foundation.
1964. Questions and Orthography in Sango. *Orthography Studies: Articles on New Writing Systems*. Helps for Translators 6. London: United Bible Societies, 161–164.
1967. *A Grammar of Sango*. The Hague: Mouton.
1970. Sango: *Langue de l'Afrique Centrale*. Leiden, Netherlands: E.J. Brill.
1982. Colonization and pidginization on the Ubangi River. *Journal of African Languages and Linguistics* 4, 1–42.
- Samarin, William J., and James A. Walker. Verb-Marking in Sango Predicate Chains. In: Helma Pasch (ed.), *Sango: The National Official Language of the Central African Republic*. Proceedings of the Colloquium “The Status and Uses of Sango in the Central African Republic,” Cologne, September 3–4, 1992. Köln: Rüdiger Köppe Verlag, 113–128.

- Sampson, Geoffrey. 1985. *Writing Systems: A Linguistic Introduction*. Stanford: Stanford University Press.
- Saulnier, Pierre. 1995. *Plantes Médicinales et Soins en Centrafrique*, 4<sup>e</sup> edition. Bangui: Centre Pastoral.
- Saussure, Ferdinand de. 1972. *Cours de linguistique générale*. Édition critique préparée par Tullio de Mauro. Paris: Payot.
- Saussure, Ferdinand de. 1959. *Course in General Linguistics*. Translated by Wade Baskin. New York: The Philosophical Library. Revised 1974 edition, with foreword by Jonathan Culler.
- Schröder, Martin C. 2004. Issues in Toposa Orthography. In: Leoma C. Gilley (ed.), *Occasional Papers in the Study of Sudanese Languages*. 9. Entebbe, Uganda: SIL-Sudan, 27–37.
- Scragg, D.G. 1974. *A History of English Spelling*. Manchester: Manchester University Press.
- Scribner, Sylvia, and Michael Cole. 1981. *The Psychology of Literacy*. Cambridge: Harvard University Press.
- Senner, Wayne M. 1996. Germanic Languages. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 642–651.
- Shadrikov, V., and N. Pakhomov. 1999. Literacy in Russia and the Former USSR. In: Daniel Wagner, Richard L. Venezky and Brian V. Street (eds.), *Literacy: An International Handbook*. Boulder, Colorado: Westview Press, 391–395.
- Shi, Dingxu. 1996. The Yi Script. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 239–243.
- SIL International. 2002a. *LinguaLinks Library*. Version 5.0. Dallas.
- 2002b. Symbolizing tone with punctuation marks. *LinguaLinks Library* version 5. Dallas: SIL International. (CD-ROM) at: Develop an orthography/ Orthography Development / Options for representing tone in an orthography.
- 2002c. What is functional load? *LinguaLinks Library* version 5. Dallas: SIL International. (CD-ROM) at: Reference materials/Glossary of literacy terms/Glossary (Literacy): F.
- SIL International in CAR. 1992. *Yê sô mo lîngbi tî hînga: Cahier d'auto-instruction pour apprendre à lire et à écrire l'orthographe officielle du sango*. Bangui.
1995. *Kêtê Bakarî tî Sängö: Petit Dictionnaire Sango*. Bangui: SIL, in cooperation with ILA.
- SIM. 2003. SIM Country Profile: Central African Republic. <http://sim.org/country.asp?fun=1&CID=2&ver=print> (accessed June 21, 2006).
- Simons, Gary F. 1977. Principles of Multidialectal Orthography Design. In: R. Loving and Gary Simons (eds.), *Workpapers in Papua New Guinea Languages 21*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics, 325–342.
1979. Language variation and limits to communication. Technical Reprint No. 3. Ithaca, New York: Department of Modern Languages and Linguistics, Cornell University.

- Sjoberg, Andrée F. 1966. Socio-cultural and Linguistic Factors in the Development of Writing Systems for Pre-literate Peoples. In: William Bright (ed.), *Sociolinguistics. Proceedings of the UCLA Sociolinguistics Conference 1964 (Janua Linguarum XX)*. The Hague: Mouton, 260–276.
- Sloan, Cliff. *The Hmong Language: An Oral Memory*. Washington State Arts Commission. <http://www.arts.wa.gov/progfa/AsianFest/Hmong/fahmong4.html> (accessed June 16, 2006).
- Smalley, William. 1964a. Dialect and Orthography in Gipende. *Orthography Studies: Articles on New Writing Systems. Helps for Translators 6*. London: United Bible Societies, 138–144.
- 1964b. How Shall I Write This Language? *Orthography Studies: Articles on New Writing Systems. Helps for Translators 6*. London: United Bible Societies, 31–52.
- 1964c. Orthography Conference for French West Africa. *Orthography Studies: Articles on New Writing Systems. Helps for Translators 6*. London: United Bible Societies, 120–126.
- 1964d. A Problem in Orthography Preparation. *Orthography Studies: Articles on New Writing Systems. Helps for Translators 6*. London: United Bible Societies, 53–59.
- 1964e. The Use of Non-Roman Script for New Languages. *Orthography Studies: Articles on New Writing Systems. Helps for Translators 6*. London: United Bible Societies, 71–107.
- 1964f. Writing Systems and Their Characteristics. *Orthography Studies: Articles on New Writing Systems. Helps for Translators 6*. London: United Bible Societies, 1–17.
- Smalley, William, et al. 1964. *Orthography Studies: Articles on New Writing Systems. Helps for Translators 6*, London: United Bible Societies.
- Smalley, William, Chia Koua Vang, and Gnia Yee Yang. 1990. *Mother of Writing: The Origin and Development of a Hmong Messianic Script*. Chicago: The University of Chicago Press.
- Snider, Keith. 2001. Linguistic Factors in Orthography Design. In: Ngessimo M. Mutaka and Sammy B. Chumbow (eds.), *Research Mate in African Linguistics: Focus on Cameroon*. Köln: Rüdiger Köppe Verlag, 323–332.
- Société Biblique Centrafricaine. 1986. *Nzoni Tene so Marke asu na Mbeti*. (L’Evangile de Mark en langue Sango). Bangui.
1995. *Buku ti Nzapa: Fini Mbuki*. (Le Nouveau Testament en Sango Courant). Bangui.
- SomaliNet. 2005. Interview with Michael Everson. [http://somalinet.com/library/osmaniya/?show=micheal\\_int](http://somalinet.com/library/osmaniya/?show=micheal_int) (accessed June 16, 2006).
- Souag, Lameen. 2003. Writing Berber Languages: A Quick Summary. <http://www.geocities.com/lameens/tifinagh/> (accessed June 15, 2003; no longer available).
- Sprenger-Charolles, Liliane, and Danielle Béchenec. 2004. Variability and invariance in learning alphabetic orthographies: From linguistic description to psycholinguistic processing. *Written Language and Literacy* 7(1), 9–33.

- Steever, Sanford, B. Tamil Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 426–430.
- Stevens, John. 1996. Asian Calligraphy. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 244–251.
- Taber, Charles R. 1965. *A Dictionary of Sango*. Hartford, Connecticut: Hartford Seminary Foundation.
- Tabouret-Keller, Andrée, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.). 1997. *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press.
- Tadadjeu, Maurice, and Etienne Sadembouo. 1984. *General Alphabet of Cameroon Languages*. Bilingual version. Propelca 1. Yaoundé: University of Yaoundé.
- Taylor, Insup, and David R. Olson (eds.). 1995. *Scripts and Literacy: Reading and Learning to Read Alphabets, Syllabaries and Characters*. Neuropsychology and Cognition 7. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Taylor, Insup, and M. Martin Taylor. 1983. *The Psychology of Reading*. New York: Academic Press.
1995. *Writing and Literacy in Chinese, Korean and Japanese*. Amsterdam: John Benjamins.
- Testen, David D. 1996. Old Persian Cuneiform. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 134–137.
- Thornell, Christina. 1994 Reflections on the 1984 Sango Orthography Decree. In: Helma Pasch (ed.), *Sango: The National Official Language of the Central African Republic*. Proceedings of the Colloquium “The Status and Uses of Sango in the Central African Republic,” Cologne, September 3–4, 1992. Köln: Rüdiger Köppe Verlag, 34–55.
1995. Established French Loanwords in Sango – a Pilot Study. Lund University Department of Linguistics. Working Papers 44, 159–177.  
<http://www.ling.lu.se/disseminations/pdf/44/Thornell.pdf> (accessed June 24, 2006).
1997. *The Sango Language and Its Lexicon* (Sêndâ-yângâ tî sângö). Travaux de l'Institut de Linguistique de Lund 32. Lund, Sweden: Lund University Press.
- Thonhauser, Ingo. 2003. Written language but easily to use! Perceptions of continuity and discontinuity between written/oral modes in the Lebanese context of biliteracy and diglossia. *Written Language and Literacy* 6(1), 93–110.
- Tokyo University Department of Asian and Pacific Linguistics. 2003. UNESCO Redbook of Endangered Languages.  
<http://www.tooyoo.l.u-tokyo.ac.jp/Redbook/> (accessed June 16, 2006).
- Tuchscherer, Konrad. 2004. African Script Facing Extinction: Case of the A-ka-u-ku and Shümom of the Bamum. Press Release December 14, 2004. Yaoundé, Cameroon: Embassy of the United States. [http://yaounde.usembassy.gov/december\\_14\\_2004.html](http://yaounde.usembassy.gov/december_14_2004.html) (accessed June 16, 2006).
- Tuttle, Edward. 1996. Romance Languages. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 633–642.

- Tzeng, Ovid J.L., and Harry Singer (eds.). 1981. *Perception of Print: Reading Research in Experimental Psychology*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- UNESCO. 1953. *The Use of Vernacular Languages in Education*. Monograph on Fundamental Education VIII. Paris: United Nations Educational, Scientific and Cultural Organization. <http://unesdoc.unesco.org/images/0000/000028/002897eb.pdf> (accessed March 24, 2006).
1990. *World Declaration on Education for All and Framework for Action to Meet Basic Learning Needs*. World Conference on Education for All: Meeting Basic Learning Needs. Jomtien, Thailand. March 5–9, 1990. <http://unesdoc.unesco.org/images/0012/001275/127583e.pdf> (accessed May 10, 2006).
2001. *Education for All: Information Kit*. <http://unesdoc.unesco.org/images/0012/001241/124166e.pdf> (accessed June 16, 2006).
2002. *Education for All: Is the World on Track? EFA Global Monitoring Report 2002*. Paris: UNESCO.
- 2005a. *Endangered Languages: UNESCO Culture Sector* [http://portal.unesco.org/culture/en/ev.php-url\\_id=8270&url\\_do=do\\_topic&url\\_section=201.html](http://portal.unesco.org/culture/en/ev.php-url_id=8270&url_do=do_topic&url_section=201.html) (accessed November 29, 2005).
- 2005b. *Promoting the Convention for the Safeguarding of the Intangible Cultural Heritage: Information Kit*. <http://unesdoc.unesco.org/images/0014/001412/141247e.pdf> (accessed June 16, 2006).
- 2005c. *Report by the Director-General on an overall strategy for an intersectoral programme for languages at UNESCO*. Executive Board, 171<sup>st</sup> session. <http://unesdoc.unesco.org/images/0013/001387/138768e.pdf> (accessed June 16, 2006).
2006. *Literacy for Life: Education for All Global Monitoring Report 2006*. Paris: UNESCO. [http://portal.unesco.org/education/en/ev.php-url\\_id=43366&url\\_do=do\\_topic&url\\_section=201.html](http://portal.unesco.org/education/en/ev.php-url_id=43366&url_do=do_topic&url_section=201.html) (accessed July 8 2006).
- UNESCO Institute for Statistics. 2006. *Adult (15+) Literacy Rates and Illiterate Population by Country and by Gender for 2000–2004*. [http://www.uis.unesco.org/TEMPLATE/html/Exceltables/education/Literacy\\_National\\_April2006.xls](http://www.uis.unesco.org/TEMPLATE/html/Exceltables/education/Literacy_National_April2006.xls) (accessed June 27, 2006).
- Unicode Consortium 1991–2006a. *History of Unicode*. <http://www.unicode.org/history/index.html> (accessed Feb. 14, 2006).
- 1991–2006b. *The Unicode Standard, Version 4.0.0 on-line version, defined by: Unicode Consortium 2003*. *The Unicode Standard, Version 4.0*. Boston, Massachusetts: Addison-Wesley. <http://www.unicode.org/versions/Unicode4.0.0/pdf> (accessed June 16, 2006).
2004. *Approved minutes of the UTC 101/L2 198 Joint Meeting*. Cupertino, CA, November 15–18, 2004. <http://www.unicode.org/consortium/utc-minutes/UTC-101-200411.html> (accessed July 8, 2006).
- United Nations. 1948. *Universal Declaration of Human Rights*. General Assembly resolution 217 A (III). <http://www.un.org/Overview/rights.html> (accessed June 16, 2006).
- United States Department of State (2005). *Background Note: Central African Republic*. Bureau of African Affairs. <http://www.state.gov/r/pa/ei/bgn/4007.htm> (accessed June 16, 2006).

- Unseth, Peter. 2005. Sociolinguistic parallels between choosing scripts and languages. *Written Language and Literacy* 8(1), 19–42.
- Van Dyken, Julia R., and Constance Kutsch Lojenga. 1993. Word Boundaries: Key Factors in Orthography Development. In: Rhonda L. Hartell (ed.), *Alphabets of Africa*. Dakar: UNESCO-Dakar Regional Office and Summer Institute of Linguistics, 3–20.
- Velhagen und Klasing Verlag. 1938. Deutsches Lesebuch für Deutsche Volksschulen. Zweite Auflage, IX. Bielefeld: Verlag von Velhagen und Klasing.
- Venezky, Richard L. 1977. Principles for the Design of Practical Writing Systems. In: Joshua. A. Fishman (ed.), *Advances in the Creation and Revision of Writing Systems*. The Hague: Mouton, 37–54. Originally published in 1970, *Anthropological Linguistics* 12(7), 256–270.
1982. Linguistics and/or Reading or is Applied Linguistics a Caveat Emptor Technology? In: William Frawley (ed.), *Linguistics and Literacy*. New York: Plenum Press, 269–283.
2004. In search of the perfect orthography. *Written Language and Literacy* 7(2), 139–163.
- Verhoeven, Ludo. 1999. Second Language Reading. In: Daniel Wagner, Richard L. Venezky and Brian V. Street (eds.), *Literacy: An International Handbook*. Boulder, Colorado: Westview Press, 143–147.
- Voorhoeve, J. 1964. Some Problems in Writing Tone. *Orthography Studies: Articles on New Writing Systems*. Helps for Translators 6. London: United Bible Societies, 127–131.
- Wagner, Daniel, Richard L. Venezky, and Brian V. Street (eds.), 1999. *Literacy: An International Handbook*. Boulder, Colorado: Westview Press.
- Walden Font Company 2006. On the History of the Old German Script. <http://www.waldenfont.com/content.asp?contentpageID=8> (accessed June 16, 2006).
- Ward, Monica. 2002. The Pedagogical and Linguistic Issues Involved in the Production of EL Materials – a Case Study of Nawat. In: R. McKenna Brown (ed.). *Endangered Languages and Their Literatures*. Proceedings of the Sixth FEL Conference, Foundation for Endangered Languages. Bath, England, 13–18.
- Watch Tower Bible and Tract Society of Pennsylvania (2006). 265 Languages. Jehovah's Witnesses official website. <http://www.jehovantodistajat.fi/languages/languages.htm> (accessed May 24, 2006).
- Weber, David John. 1992. Syllable-based hyphenation. *Notes on Linguistics* 58, 4–17.
2006. *Writing Quechua: The Case for a Hispanic Orthography*. UCLA Latin American Studies 87.
- Wheatley, Julian K. 1996. Burmese Writing. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 450–456.
- Wiebelt, Alexandra. 2004. Do symmetrical letter pairs affect readability? A cross-linguistic examination of writing systems with specific references to the runes. *Written Language and Literacy* 7(2), 275–304.
- Wiesemann, Ursula. 1989. Orthography matters. *Notes on Literacy* 57, 14–21.
1995. Tone orthography and pedagogy. *Notes on Literacy* 21(3), 25–31.

- Woldemikael, Tekle M. 2003. Language, Education, and Public Policy in Eritrea. *African Studies Review*.  
[http://www.findarticles.com/p/articles/mi\\_qa4106/is\\_200304/ai\\_n9219176](http://www.findarticles.com/p/articles/mi_qa4106/is_200304/ai_n9219176)  
 (accessed June 16, 2006).
- Wolff, Ekkehard H. 1991. Standardization and Varieties of Written Hausa. In: *Standardization of National Languages*. Symposium on Language Standardization, February 2–3, 1991. Hamburg: University of Hamburg and the Unesco Institute for Education.  
[http://www.unesco.org/education/pdf/35\\_57.pdf](http://www.unesco.org/education/pdf/35_57.pdf) (accessed June 16, 2006).
2000. Language and Society. In: Bernd Heine and Derek Nurse (eds.), *African Languages: An Introduction*. Cambridge: Cambridge University Press, 298–347.
- Wolfson, Nessa and Joan Manes (eds.). 1985. *Language of Inequality*. Contributions to the Sociology of Language 36. Berlin: Mouton Publishers.
- Wong, J. N. 1961. The Progress of Pinyin. *Stud. Intel.* V5:1-A35–A52. Released 1993.  
[http://www.cia.gov/csi/kent\\_csi/docs/v05i1a06p\\_0001.htm](http://www.cia.gov/csi/kent_csi/docs/v05i1a06p_0001.htm) (accessed June 14, 2006).
- Yip, Moira. 2002. *Tone*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Zhou, Minglang. 2001. Language policy and reforms of writing systems for minority languages in China. *Written Language and Literacy* 4(1), 31–65.
- Zide, Norman. 1996. Scripts for Munda Languages. In: Peter Daniels and William Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press, 612–618.
- Zivenge, William. 2005. The effects of Shona language change on monolingual lexicography: The need for a revised alphabet. Prepublication  
[http://ir.uz.ac.zw:8080/dspace/bitstream/uzlib/51/1/Zivenge\\_the\\_effects\\_of\\_shona\\_language\\_change.pdf](http://ir.uz.ac.zw:8080/dspace/bitstream/uzlib/51/1/Zivenge_the_effects_of_shona_language_change.pdf) (accessed July 8, 2006).