IntechOpen

Multilingualism and Bilingualism

Edited by Beban Sammy Chumbow

MULTILINGUALISM AND BILINGUALISM

Edited by **Beban Sammy Chumbow**

Multilingualism and Bilingualism
http://dx.doi.org/10.5772/intechopen.68609
Edited by Beban Sammy Chumbow

Contributors

Maha Hamad Aldosari, Maura Pilotti, Khadija El Alaoui, Sue Ann Lee, Irina-Ana Drobot, Julio Alonso, Diana Margarita Díaz-Mejía, Daniela Estrada, Biligsaikhan Batjargal, Daniel Enrique Támara López, Estefanía Guevara Blanquicett, Carlos Daniel Acosta-Medina, Evelyn Fogwe Chibaka, Sammy Beban Chumbow

Notice

First published in London, United Kingdom, 2018 by IntechOpen
eBook (PDF) Published by IntechOpen, 2019
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, The Shard, 25th floor, 32 London Bridge Street
London, SE19SG – United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Additional hard and PDF copies can be obtained from orders@intechopen.com

Multilingualism and Bilingualism
Edited by Beban Sammy Chumbow

p. cm.

Print ISBN 978-1-78923-226-4

Online ISBN 978-1-78923-227-1

eBook (PDF) ISBN 978-1-83881-332-1

Meet the editor

Professor Beban Sammy Chumbow, PhD, Indiana University, USA (1975), is currently the emeritus professor of Linguistics at the University of Yaoundé 1, pro-chancellor of the University of Dschang, and president of the Assembly of Academicians of the African Academy of Languages, a specialized institution of the African Union. He has published extensively in Generative Phonology and Language Policy and Planning in a multilingual setting. He is the vice president of the Cameroon Academy of Sciences (CAS), a member of the New York Academy of Sciences, and a former member of the Linguistic Society of America. He has served as the rector/vice chancellor (executive head) of four Universities and has been a visiting professor to major universities in Europe, the USA, Canada, and Africa. Among other distinguished international functions, he served as a member of the Scientific Committee of the Council for Social Science Research in Africa (CODESRIA), a member of Council of the Association of African Universities (AAU), a member of the Scientific Council of the "Agence Universitaire de la Francophonie" (AUF), and a member of Council of the Association of Commonwealth Universities (ACU).

Contents

Preface

Bilingualism and multilingualism as the object and subject of scientific inquiry is an enigmatic and intriguing phenomenon. Bilingualism is indeed phenomenal (in the etymological sense of the word), in that even though so much research is done and so much knowledge is generated on bilingualism and multilingualism, there is still so much that is unknown and so much to be known that the much we know about bilingualism may not be much after all.

Who would have thought, several decades ago, that bilingualism as the ability to use two or more languages confers on the bilingual individual cognitive advantages over the monolingual individual that makes bilingualism not only palatable but desirable and, indeed, indispensable? Today, the avalanche of empirical evidence of research findings to this effect, presented in this book, makes an assertion that bilinguals have enormous competitive advantages over monolinguals in the educational system, an incontrovertible fact.

In the Europe of yesteryears, the paradigm of "one nation one language" (the monolingual habitus) was so dominant and fashionable as a nation-building ideology that multilingualism was considered a curse, a demon that had to be exorcised. This led to the voluntary and involuntary endangerment, attrition, or death of minority languages. Today, the abundance of research findings of advantages for multilingualism or societal bilingualism has led to a rethinking of the issue. This, coupled with considerations and acknowledgement of the value of language, any language and every language as an invaluable treasure of the manifestation of the human mind and human knowledge whose loss or death impoverishes humanity immeasurably, has led today to the deconstruction of the paradigm of monolingual habitus by the European Union's language policies in favor of linguistic diversity and the preservation of rights for minority languages.

The nine chapters of this book provide elucidations of the issue of benefits of bilingualism and multilingualism and further provide original research findings on developments in the areas of psychological dimensions of bilingualism and bilingualism in information retrieval systems. Thus, there is still so much to be learned and to be known about the phenomenon of bilingualism and multilingualism to make another book, this book, relevant, worthwhile, and necessary. It is preceded by an introduction containing topics that mirror the issues discussed in the book.

Section II: Advantages and Case Studies of Bilingualism and Multilingualism

The section presents the benefits of bilingualism to the individual and the advantages of multilingualism to society and the nation followed by two case studies of multilingualism: aspects of bilingualism and multilingualism in Europe and the evaluation of achievements of bilingual policy in Columbia, South America. The section establishes the benefits of bilingualism and multilingualism, and the case studies that follow examine how, considering the phenomenon of bilingualism as a desirable "good," nations (in Europe and Latin America) accommodate what appears ostensibly as the adverse side effects of an otherwise useful and much-needed medication for the arduous but rewarding task of nation building within an ideological paradigm of linguistic and cultural pluralism.

Section III: Psycholinguistic Perspectives on Bilingualism

This section contains two original research papers that extend the frontiers of knowledge with empirical evidence on two crucial aspects of the psycholinguistics of bilingualism. The first seeks to elucidate and provide answers to important psycholinguistic questions of language acquisition. Do bilingual children possess one or two linguistic systems in the learning of their respective languages, and when do these systems emerge in the process of acquisition? The second seeks to illuminate the role of psycholinguistic identity and cognition on one hand and culture on the other hand as variables in bilinguals' perception and evaluation or assessment of their **self-efficacy,** that is, their capabilities and competence when using either language.

Section IV: Bilingualism, Information Retrieval and Access to Information

This section presents three research-based chapters with findings related to issues of information retrieval and access to information in bilinguals or in multilingual settings. "Knowledge is power," as rightly observed by the visionary, Francis Bacon several centuries ago (1597). In the present **age of knowledge economy,***knowledge is power*, and "knowledge power" rules the world because economically powerful nations in the hierarchy of development are characterized by industrialization determined by the quantity and quality of *knowledge production, knowledge dissemination, knowledge appropriation, and knowledge management* they leverage. It is now axiomatic that the present century of knowledge-based economy is also the **age of information and communication technology**, which serves as the engine and motor of the (abovementioned) pillars and determinants of knowledge economy (knowledge production, dissemination, appropriation, and management). This underscores the importance of research pertaining to information retrieval and access to information in this section of the book.

The first article contributes to solutions to the problem of determining the best method of crosslinguistic information retrieval such as retrieving information in two or more languages (e.g., English, Spanish, and French) when the query is given in one language (e.g., English). The second chapter presents research that provides answers to the complex but useful question "how does one access information from digital bases across several languages both synchronically and diachronically, i.e., in the present state of the language(s) and across historical stages"? In the third chapter, it is demonstrated that while different websites frequently use Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA), the design and successful use of a CAPTCHA obtained from analyzing handwriting samples in several languages (English, French, Spanish, and Arabic) stand as a significant innovation in Internet security and information retrieval system.

The book by its illuminating description and insightful analysis of issues of bilingualism will be of significant interest to scholars and researchers, and given the knowledge dissemination vocation of the work, in the present multilingual world, all concerned with bilingualism and multilingualism from whatever perspective will find a worthwhile interest in various components of the book.

Professor Beban Sammy Chumbow
Emeritus Professor of Linguistics
University of Yaoundé 1
President of the Assembly of Academicians
African Academy of Languages, African Union

Introduction

Chapter 1

Introductory Chapter: Relevance of Bilingualism and Multilingualism in a Multilingual World

Beban Sammy Chumbow

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.77083

1. Introduction

What is a logically possible answer to the logically possible question: "what is the relevance of multilingualism in a multilingual world?" This is an existentialist question and the first logical reaction will be to question whether the question is logical because it is not logically possible for multilingualism to be absent (and irrelevant) in a multilingual world (that is, in its own world). Indeed, multilingualism is at home in a multilingual world and cannot be challenged with respect to relevance. Ultimately, the real question is: what is the logic of asking multilingualism, what it is doing in a multilingual world or what "business" it has in a multilingual world? However, with such a question, we end up with a *logical "argumentum ad absurdum," a question whose answer leads to a logical conclusion of absurdity* for, indeed, it is self-evident that multilingualism owns the multilingual world, and its relevance is therefore equally self-evident. The apparently interesting question is, therefore, (on further examination) as foolish as asking a man what he is doing in his own home. Indeed, in our world of the twenty-first century, *multilingualism is the norm and monolingualism the exception* [1]. Besides, inherent *bilingualism* and *multilingualism*, over the years, language contact resulting from conquest, colonization, mass migration of populations due to disasters, and the impact of globalization has made all countries of the world multilingual in varying degrees. Given the symbiotic relationship between language, culture, and identity, multilingualism ipso facto implies multiculturalism and ethnolinguistic and ethnocultural identity have become inalienable rights. This imposes challenges for nation building, driven by ideals of ideological pluralism, social cohesion, and living together in diversity and so on, which have been the object and subject of various studies of bilingualism and multilingualism.

This chapter focuses on the compendium of aspects of recent original research contributions to more and new knowledge on *bilingualism*, and its societal corollary, *multilingualism* in the context of the knowledge economy and national development.

2. Advantages of bilingualism and multilingualism

The issue of advantages or benefits of bilingualism to the individual and the community is very topical; yet, it does not often benefit from a comprehensive presentation of knowledge-based information that is informed by research findings; findings that make such information authentic. The challenge is undertaken in this work, which ensures that the compendium of claims of benefits of bilingualism and multilingualism articulated are logical conclusions drawn from well-resourced research findings and not merely the fruit of a fertile imagination or wishful thinking. Presentation of the benefits of bilingualism to the individual below precedes benefits of bilingualism to the society (community, nation, or state). Assertions of benefit and advantages are followed by verifiable and verified sources.

2.1. Advantages of individual bilingualism

2.1.1. Bilingualism and cognition

Research on bilingualism and cognition leads to the empirical findings that on the whole, bilinguals (as individuals), *have more improved cognitive abilities than monolinguals*. Bilinguals "**exercise … superiority in cognitive, metalinguistic awareness, linguistic, and academic performances**" over monolinguals [2].

Bilinguals are cognitively more creative with respect to flexible thinking ability and more proactive in nature [3, 4]; bilinguals have **executive control advantage**, which stimulates the memory giving them a better capacity for recollection.

2.1.2. Bilingualism and communicative competence

Bilingualism enhances the communicative ability and communicative competence [4, 5]. The bilingual experience fosters linguistic competences in varied verbal tasks, resulting in having increased competence or **multi-competence;** that is, **added capacity.** *Thus, a bilingual that is seeking to solve a problem in one language is thought to be able to draw on the other language and related frames of mind to bring extra cognitive capacity to bear in solving a problem* [6, 7].

Besides, there is evidence that bilinguals with a multilingual competence have high self-confidence, and stronger willingness to communicate without fear of constructing non-well-formed utterances [6, 7].

2.1.3. Bilingualism and language learning

Bilingualism has significant advantages for language learning in particular and knowledge appropriation in general in that bilinguals in using more than one language, *tend to develop better language proficiency, sensitivity, and definitely more in-depth understanding and control of their mother tongue* [6]. In learning another language, bilinguals have more opportunities to *transfer knowledge and language governing rules from one language to the oth*er, making it easier for them to learn overlapping cognates in form and meaning across languages [8].

Bilinguals have t*he potentials to attain high academic standards with less effort* than monolinguals (Curtain and Dahlberg [9]), *because those who speak multiple languages tend to have a stereoscopic vision of the world from two or more perspectives, enabling them to learn reading more easily*, develop **better critical thinking abilities, *and be more flexible in their thinking.***

2.1.4. Bilingualism and dementia

Bilingualism delays aging mental disorders, dementia, or Alzheimer's disease. Studies show that there is a *significant delay of the symptoms of old age dementia recorded in bilinguals*, which on the other hand manifests faster in monolinguals. The onset of dementia in some studies is at least 5–10 years earlier in monolinguals than bilinguals [10].

2.1.5. Bilingualism and aphasia

Aphasia can be better managed in bilinguals than monolinguals because in post-stroke aphasia, for instance, therapy will involve a "transfer effect," whereby **"***the second, less dominant language can be used to transfer knowledge to the primary one, helping with rehabilitation"* [11].

2.1.6. Bilingualism and information processing

Bilinguals are better equipped for information processing as a result of the constant code-switching between languages, which provides bilinguals with potentials to switch the functionality strategies of one language to the other as the need demands. *"When the brain is constantly so exercised, it doesn't have to work as hard as the monolingual's to perform most cognitive tasks effectively"* [12].

2.1.7. Sociocultural benefits of bilingualism

There are sociocultural advantages to bilingualism and multilingualism in that multilinguals are more obviously multicultural (since language and culture are related). Bilinguals, therefore, have more **intercultural competence. Multilinguals are more open and can better** appreciate and accommodate a foreign culture than monolinguals. Such a predisposition is a factor in lessening racism, xenophobia, intolerance and enabling intercultural dialog, and social cohesion [9, 13].

2.1.8. Economic advantages of bilingualism

Bilingualism has economic advantages in that multilingualism is an investment in human capital; for knowledge of several languages can afford one the opportunity of earning the higher income and obtaining aspiring employment status in any influential knowledge society. The use of a second or third language in the workplace is associated with positive earning differentials, ranging from 3 to 8% in different Western European countries [14, 15].

2.2. Advantages of societal bilingualism

A survey of empirical findings leads to the conclusion that the opportunities and privileges of multilingual societies outnumber those of monolingual societies for many reasons. State

multilingualism governed by a policy of ethno-linguistic pluralism that consecrates societal bilingualism has a plethora of advantages for individuals that benefit the society or state as well.

2.2.1. Advantages of multilingualism and multiculturalism

Multilingualism implies multiculturalism, which as indicated above helps in strengthening social cohesion, intercultural dialog, and so on. This is possible provided, there are policies on state pluralism that provide guarantees for recognition of ethno-linguistic communities. When ethno-linguistic communities are guaranteed recognition of their identity, they tend to feel a sense of security and a sense inclusion and belonging to the nation or state [13, 16].

2.2.2. Academic advantages of multilingualism in the area of education

These are quite obvious in that, inter alia, multilingual skills, and competence open up access to a huge store of knowledge through access to a wider scope of sources and resources in the educational institutions and beyond. Thus, the lack of multilingual language skills limits researchers in their ability to engage internationally in or with their research, and in their career opportunities [10, 17].

2.2.3. Benefits of multilingualism for national security

Multilingualism has benefits for national security and conflict resolution. Garamone [18] extolls the power of multilingualism for national security and Samire [19] advises a *multilingual approach to solving conflicts in a multilingual setting*.

2.2.4. Multilingualism, governance, and policy implementation

The implementation of government policies, for instance, in health care delivery, agriculture, environmental action, and so on in multilingual societies is best carried out in the various mother tongues or "languages of the heart.," for dissemination of information and knowledge in the mother tongue **(**or better known language) leads to a greater level of comprehension and appropriation of knowledge and innovations [20].

2.3. Sample case studies of bilingualism and multilingualism

Societal multilingualism has gained acceptance as a desirable objective in the implementation of a paradigm of national development predicated on ideals of ethno-linguistic unity in diversity. **Case studies are** quite illuminating with respect to efforts to face challenges of pluralism.

2.3.1. Bilingualism and multilingualism in the European Union

The European Union's language policy seeks to consecrate and consolidate multilingualism and multiculturalism in the Union as a new-found ideological paradigm for nation building congruent with ideals of pluralism. It is known that *the European framework of reference, regulates*

and demands multilingual language competence in Europe, and to this effect, each European citizen should understand at least two languages other than the mother tongue [21–23]. Such requirements at the level of the Union impose national language policies in favor of societal multilingualism and *ipso facto,* encourage individual bilingualism in the European community of nations. **Language contact within the nation-state** may require the acquisition of a majority language by minorities and migrants who come in with a minority language. This leads to issues of **polarization** of *minority or heritage languages versus majo***rity or official languages,** *inclusions versus exclusion, insiders versus outsiders, and so on.* Language policy and planning has a duty to mediate and diffuse tension by **devising and** implementing policies that ensure revitalization, revalorization, and use of minority and heritage languages along with the major languages for development, to ensure an inclusive society. Multilingualism as a national objective, occasion's intensive foreign language learning in Europe, the rise of English as a dominant lingua franca and the advent of a thriving translation industry for building bridges of intercultural and **cross-cultural understanding.**

2.3.2. Evaluation of national bilingual policy in Columbia

The **Colombian** constitution recogn**izes the promotion of** indigenous languages and the development of Spanish-English bilingualism (or bilingualism in Spanish and another foreign language) in the school system, where Spanish is the official language, legacy o**f the colonial past.** English is adopted as a language necessary **for knowledge appropriation** and interaction with the global community for national development. The main objectives of the evaluation process are to determine to what extent the Colombian Bilingual Program is effective, efficient, and successful and ultimately, to what extent the Ministry of Education's overall goal of achieving 40% success in English (and ipso facto, Spanish**-**English bilingualism) by 2014 was attained. Detailed results show that despite the efforts conceded, the overall success rate falls below the expected target. Significant observations follow from the case study and make recommendations on challenges to be faced on the way forward to the 2025 phase of the program. The case study is quite instructive and patently useful with lessons to be learned in the enterprise of nation building in developing countries, especially in the face of challenges of multilingualism and socio-economic development, not only for what is achieved in the Colombian experience, but perhaps more importantly, for what is not achieved and why.

3. Psycholinguistic perspectives on bilingualism

Psycholinguistics perspectives of bilingualism have been an important aspect of studies in bilingualism recently, two of which are elucidated here below.

3.1. Phonetic category formation in bilingual children

How do bilingual children come to distinguish phonetic details of the two linguistic systems they are internalizing in the language acquisition process? The central issue in this respect is,

since bilingual children may acquire (learn) the two languages sequentially (L1 first then L2) or simultaneously (L1… L2 concurrently, at the same time), do these two modes of bilingualism exhibit similar or different characteristic features with respect to phonetic category formation? This preoccupation embodies the concern with whether bilingual children develop and speak the two languages from two separate systems (*ab initio*) or from one system that gets differentiated subsequently, into two identifiable systems? If the later, when do the systems differentiate and do the languages interact in the process of acquiring the characteristic features? If so, is the direction of interaction predictable, and so on? (This problematic may also involve issues of psychological reality). The research is driven by a summary of key empirical studies with evidence of the observed facts of the development of phonetic categories by bilingual children (both sequential and simultaneous cases of bilingualism). It is undertaken within the theoretical framework of phonetic categories of bilinguals; that is, the speech learning model (SML) for adult bilinguals [24] and the linguistic system model (LSM) for bilingual children [25]. The results lead to the proposal of "*The Development Model of Phonetic Category Formation,"* which is an integrated explanatory model of the findings of current research in the discipline. The model is to the effect that *detailed phonetic categories do not form across-the-board and bilingual children may invoke multi-dimensional representations of phonetic categorie*s. It goes further than the SML and the LSM to state that *phonetic category formation continues to evolve during the developmental process rather than emerge all at once in both simultaneous and sequential bilingual children*.

3.2. Bilingualism and self-perception

Linguists, anthropologist, and sociolinguists have long known that language is inextricably linked to culture and is indeed an element of culture *par excellence*. Every language encodes culture and cultural values. Thus, in the process of language acquisition, bilinguals internalize the values, norms, and concepts, including role expectations and attitudes of the culture expressed by each of the languages, they are mastering. Consequently, there is a symbiosis of language, identity and culture in the temple of the mind. How do the various cultural norms and values coexist in the bilingual, who has internalized several languages and their cultures and how can these cultures be retrieved? It is now axiomatic that *language can prime a culture,* meaning that language can activate culture in the mind, the culture with which the language is associated in the symbiosis within the mind [26]. What is the relationship between language, as the vehicle of a person's culture, and self-assessment of one's capabilities (i.e., self-efficacy) via conventional self-report measures? This is revealed in the study of bilingualism and culture in the Kingdom of Saudi Arabia (KSA), where the key languages used to prime culture (Arabic and English) pertain to cultural orientations, whose key dimensions stand in opposition to one another. Arabic represents a vertical-collectivistic culture (with high uncertainty avoidance), which places premium on such values as modesty, humility, community spirit, interdependent member of the collectivity, and so on. In contrast, English represents a vertical-individualistic culture with low uncertainty avoidance (e.g., USA), which places premium on individual values of assertiveness, independence, individual achievements, and so on. The results show that bilinguals (or multilinguals) possess two (or more) culturally

construed identities, two or more "selves"; the use of one language (Arabic) "primes," and conjures the "self" corresponding to the Arabic language, which then responds psychologically, behaviorally, and culturally in an appropriate acceptable manner for the Arabic culture and heritage. The use of English primes behavior appropriate and acceptable to the American culture (to some extent). It is thus established that not only does bilingualism imply biculturalism the different cultures of bilingualism can be primed from the individual and seen through the use of each of the languages [26, 27].

4. Bilingualism, access to and retrieval of information

In this age of knowledge economy, driven by information and communication technology, bilingualism in issues of **information retrieval (IR)** and **access to information (AI)** is crucial, especially with respect to storage and retrieval of information from databases.

4.1. Cross-language information retrieval (CLIR)

Research in the area of CLIR has focused on issues, methods, and technologies of how to retrieve information on one subject from two or more languages and databases with a query in one language. The issue studied here involves the examination and evaluation of the efficiency and effectiveness of a system of retrieval of bilingual information in English and Spanish based on semi-discrete decomposition (SDD), when the query is made in Spanish. To do so effectively, four case studies that exhibit the performance of the use of the latent semantic indexing (LSI) via SDD method for CLIR are undertaken, and the results are compared with those obtained by applying the LSI via singular value decomposition (SVD) method. This is undertaken thanks to a solid database built using the *fusion strategy* in combining documents from the Bible (Gospels) in Spanish and English. The evaluation of the innovative SDD method (LSI via SDD) shows a significantly higher performance with respect to the SVD Method. It evidences the true impact of the SDD, the ability to obtain good results, with the advantages of a higher speed and very low cost **in terms of storage space.**

4.2. Cross-lingual and cross-chronological access to information

This deals with the challenge of accessing and retrieving information from several languages with a focus on diachronic access and retrieval (retrieving information from different chronological stages of the same language (e.g., Old English, Middle English, and Modern English). Using Mongolian as an example, an experiment is set up to devise mechanisms of access to information from various stages of Mongolian; mechanisms that can be generalized. This is based on a series of measured and ordered actions: Computerized analysis of historical documents, extraction of key markers of periods and geographical information such as personal and place names with the support of a vector machine, creation of digital bases on the various stages of the language with the extracted material, encoding process of relevant information, and development of a web-based prototype system for utilizing digital editions of

historical manuscripts as scholarly tools. This approach to diachronic IR is applicable to both synchronic bilingual and multilingual IR. There are indications on how it can be applied to English diachronic data and to English and Japanese IR.

4.3. Innovative multilingual CAPTCHA

Completely automated public turing test to tell computers and humans apart (CAPTCHA) is a test used by different websites on the Internet to differentiate between humans and automated bots. Cyber-crimes and cyber insecurity are quite rampant. Consequently, IT security is a crucial aspect of information management today and is the object of research and patents. CAPTCHAs are possible because humans have abilities that machines do not have and therefore, by exploiting these abilities and capacities, tests are made to exclude interference from robots on websites. The work discussed here is a CAPTCHA realized by putting together handwriting characteristics from several human scripts in four languages (English French Arabic and Spanish). The research goes through the normative engineering stages for the production of this innovative CAPTCHA such as data gathering, algorithm technique to generate the four CAPTCHA, followed by elicitation of potential User Responses, validation of responses, experimentation, and so on. The ultimate test in the series of validation experiments is the investigation of the reaction of six bots to five hundred characters CAPTCHA in the four languages. In the English CAPTCHA images, the six OCRs (robots) failed to recognize the full text in 99% of the cases, where humans recognized them at above 80%. The 1% recognition by the bots is insignificant and is attributable to chance. This is essentially the same situation for the CAPTCHA for the other languages. These results consolidate the validation of the new multilingual CAPTCHA.

5. Conclusion

The introductory chapter speaks to the issues contained in the book in a manner as to be informative and less technical in order to ensure a rapid appraisal and perception of the originality of new dimensions to the problematic of bilingualism and multilingualism. However, it is not a substitute for the rich, valuable, and insightful information to be gained by reading the full chapters of the book. Rather, it is intended to be what it is supposed to be: A *mere introduction* to the book. The finality of an introduction is to whet the appetite sufficiently enough to say "***the best is yet to come***" and ***"the taste of the pudding is in the eating."***

Author details

Beban Sammy Chumbow

Address all correspondence to: sammybchumbow@yahoo.fr

University of Yaoundé I, Cameroon

http://dx.doi.org/10.5772/intechopen.77083

References

[1] Chumbow BS. The challenge of diversity and pluralism: The tier stratification model of language planning in a multilingual setting. In: Lopez-Varela A, editor. Social Sciences and Humanities: Applications and Theory. Vol. 1. Rijeka: Intech Publications; 2012

[2] Mohanty AK. Bilingualism and cognitive development of kond tribal children: Studies on metalinguistic hypothesis. Pharmacopsychoecologia; 1992 psycnet.apa.org

[3] Landry RG. The enhancement of figural creativity through second language learning at the elementary school level. Foreign Language Annals. 1973;**7**(1):111-115

[4] Ricciardelli LA. Two components of metalinguistic awareness: Control of linguistic processing and analysis of linguistic knowledge. Applied PsychoLinguistics. 1993;**14**(3): 349-367

[5] Dickinson D, McCabe A, Clark-Chiarelli N, Wolf A. Cross-linguistic transfer of phonological awareness in low-income Spanish and English bilingual children. Applied PsychoLinguistics. 2004;**25**:323-347

[6] Barac R, Bialystok E. Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. Child Development. 2012;**83**(2):413-422

[7] Cook VJ. Evidence for multicompetence. Language Learning. 1992;**42**(4):557-591

[8] Murphy S. Second language transfer during third language acquisition. Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics. 2003;**3**(1):1-21

[9] Curtain HA, Dahlberg CA. Languages and Children, Making the Match: New Languages for Young Learners. Allyn & Bacon; 2004

[10] Craik FIM, Bialystok E, Freedman M. Delaying the onset of Alzheimer's disease: Bilingualism as a form of cognitive reserve. Neurology. 2010;**75**(19):1726e-1729e

[11] Ellis C, Lindrooth RC, Horner J. Retrospective cost-effectiveness analysis of treatments for aphasia: An approach using experimental data. American Journal of Speech-Language Pathology. 2013;**23**(2):186-195

[12] Yehene E, Meiran N. Is there a general task switching ability? Acta Psychologica. 2007;**126**:169-195. DOI: 10.1016/j.actpsy.2006.11.007

[13] Chumbow BS. Cultural and Ethno-linguistic identity: Language and literature in african's development praxis. In: Paper Presented at the African Union First Pan African Writers Conference; Accra; March 2018

[14] Grin F, Sfreddo C, Vaillancourt F. The Economics of the Multilingual Workplace. London: Routledge; 2010

[15] Williams DR. Multiple language usage and earnings in Western Europe. International Journal of Manpower. 2011;**32**(4):372-393

[16] European Commission. Multilingualism: An Asset for Europe and a Shared Commitment, (COM 2008). Brussels: European Commission; 2008. p. 566

[17] Cummins J. Language, Power, and Pedagogy: Bilingual Children in the Crossfire. Vol. 23. Toronto: Multilingual Matters Clarendon; 2000

[18] Garamone J. Official details foreign language program's progress. American Forces Press Services, U. S. Department of Defense press release, July 2. 2010. Retrieved from http://www.defense.gov/news/newsarticle.aspx?id=59880

[19] Simire G. Developing and promoting multilingualism in public life and society in Nigeria. Language, Culture and Curriculum. 2003;**16**(2):231-144

[20] Chumbow BS. Social engineering theory: A model for the appropriation of innovations. In: Lopez-Varela A, editor. Social Sciences and Humanities: Applications and Theories. Vol. 2. Rijeka: INTECH Publications; 2012

[21] Finkbeiner C, White J. Language awareness and multilingualism: A historical overview language awareness and multilingualism. In: Encyclopedia of Language and Education. 2017. pp. 3-17

[22] Häggman J. Multilingualism and the European Union. EJM. 2010;42;191-195. Printed in Austria

[23] UNESCO. A Decade of Promoting Multilingualism in Cyberspace. Paris: UNESCO; 2015

[24] Flege JE. Second-language speech learning: Theory, findings, and problems. In: Strange W, editor. Speech Perception and Linguistic Experience: Issues in Cross-Linguistic Research. Timonium: York Press; 1995. pp. 233-272

[25] Genessee F. Early bilingual development: One language or two? Journal of Child Language. 1989;**16**:161-179

[26] Trafimow D, Silverman E, Fan RM-T, Law JSF. The effects of language and priming on the relative accessibility of the private self and the collective self. Journal of Cross-Cultural Psychology. 1997;**28**:107-123

[27] Ross M, Xun WQE, Wilson A. Language and the bicultural self. Personality and Social Psychology Bulletin. 2002;**28**(8):1040-1050

Section 2

Advantages and Case Studies of Bilingualism and Multilingualism

Chapter 2

Advantages of Bilingualism and Multilingualism: Multidimensional Research Findings

Evelyn Fogwe Chibaka

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74625

Abstract

Bilingualism and multilingualism are often perceived and considered as a problem or a major challenge to individual and/or societal development. In most instances, the only advantage recognized for the bilingual individual is the ability to use two or more languages. Beyond that, monolingualism seems more attractive, and monolinguals especially those speaking a language of wider communication seem quite content with their lot, often adopting a condescending attitude toward minority native speakers of a mother tongue who in addition have to acquire their language. Adepts of the ideology of **monolingual habitus** (one nation, one language) have tended to consider multilingualism and linguistic diversity as a curse and an obstacle to nation building. This chapter argues against the above ideology through a compendium of empirical evidence of advantages of individual bilingualism, societal multilingualism, and linguistic diversity of nations that emerge from research findings in the last several decades.

Keywords: advantages, monolingualism, bilingualism, multilingualism, linguistic diversity, multidimensional, ideology, empirical evidence, research findings

1. Introduction

It is averred in some circles that linguistic diversity evidenced by multilingualism is **a curse** because of the challenges involved in building a pluralistic nation in the face of linguistic and cultural diversity [1–10]. However, there is enough evidence that, globally, *multilingualism is the norm and monolingualism the exception* [11] and the advantages of bilingualism and multilingualism need to be identified, exploited, and used for individual growth and national development. This chapter does not belabor the fact that bilingualism or multilingualism is

more advantageous than monolingualism; it rather appraises a range of multidimensional evidences that have been established from varied research findings, highlighting the benefits of bilingualism in individuals or plurilingualism and multilingualism in different societies. It explores the phenomenon of bilingualism and multilingualism, how it can be a "blessing" and not a "curse" at both levels of its manifestation. The chapter is thus largely a compendium of empirical evidence of the advantages of individual and societal bilingualism/multilingualism and linguistic diversity of nations as established by research findings in the last several decades from studies by linguists and scholars such as Byram [5], Baker et al. [6], Bialystok et al. [7], Ewert [12], Paradowski et al. [9], Grin et al. [10], etc. Thus, our focus is not only on identifying the different advantages (a thing already partially covered in many reports) but also in bringing out the different research techniques and methodologies that have been used in arriving at the different claims or justifications of advantages for these bilinguals or multilinguals. Our discussion is presented in two major sections. The first on the "advantages of individual bilingualism/multilingualism" outlines such themes as cognitive development advantages of bilingualism; the brain of bilinguals as a neurophysiological organ; advantages for Alzheimer bilinguals; linguistic awareness, benefits of communicative ability, and competence; advantages in academic or educational performance; sociocultural, economic, and political advantages; etc. The second section handles benefits of "societal multilingualism," i.e., the advantages of multilingualism for communities and the nation. The two divides—individual and societal bilingualism/multilingualism—are linked by our discussion of how they can be a blessing to any democratic nation, ending in concluding remarks that cap the research findings.

2. Background of study

Numerous scientifically supported research activities inspired by the UNESCO [13] and UNESCO [14] policy orientation on the promotion of linguistic and cultural diversity have employed different perspectives and definitions for the concept of bilingualism and multilingualism. The term multilingualism as employed here cuts across those implied or asseverated in these reports. Thus, multilingualism is considered first, as implying bilingualism, since in order to be a multilingual, one has to be exposed to at least one situation or experience of bilingualism. It is anticipated that all the features that hold for multilingualism do so too for bilingualism. But since multilingualism involves the use of more than two languages, it manifests a more complex scenario with operational characteristics that obviously overlap with those of the bilingual setting.

Nonetheless, the definition of multilingualism is as varied as the different interrelated disciplines that identify with it and the objective(s) of the research in question. The European Commission [15], for example, defines multilingualism as "the ability of societies, institutions, groups and individuals to engage, on a regular basis, with more than one language in their day-to-day lives." This definition conscripts both multilingualism and bilingualism in the expression "… more than one language…," thus considering them as phenomena

with similar semantic content, properties, and consequences. It is this same assumption that is projected by Li [16], who looks at multilingualism in the light of "anyone who can communicate in more than one language, be it active (through speaking and writing) or passive (through listening and reading)." Li aligns with the school of thought for whom a multilingual person, group, or setting engenders *basic* proficiency in the use, speaking, or understanding of more than one language. It is the same perception that is clearly underscored by scholars like Skutnabb-Kangas and McCarty, [17] with claims that "today, the idea of perfect mastery and perfect balance of two or more languages is no longer considered a requirement for being bilingual or multilingual." Concern thus shifts to the number of languages rather than the proficiency in their usage, an idea likewise maintained by authors like Vildomec [18], McArthur [19], and Edwards [20]. These consider multilingualism as "the ability to use three or more languages either separately or in various degrees of code-mixing. Different languages are used for different purposes, competence in each varying according to such factors as register, occupation, and education." It is thus variously implied that the degree of proficiency is not essential; basic speaking and listening skills (communicative skills) of the speaker(s) are all it takes to be considered as bilingual or multilingual.

Despite different perceptions, the different definition perspectives converge on the assumption that the multilingual setting needs to have a speaker, group, nation, or activity/environment, where *two or more languages* are used for communication. None highlights considerations of the situational use/domain, function, degree of fluency, different manner, time, or place of acquisition of the second language and other languages.

All taken into consideration, the perspective adopted in this study is that of Aronin and O Laoire [21] that "plurilingualism" limits its scope to only individuals and not societal multilingualism. By implication, discussions about the different types of multilingualism, such as coordinate bilingualism, referring to person's learning of two languages in separate environments/contexts; subcoordinate bilingualism, referring to the acquisition of the second language (L2) with the help of the first language (L1); compound bilingualism, which is the learning of two languages in the same environment, time, and even context; and the different stages of acquisitions and their degrees of proficiency levels as detailed by Bassetti and Cook [22] and Baker [23], are beyond the scope of this study. They, however, constitute useful typologies of the phenomenon for those interested in conceptual details.

There are, no doubt, some drawbacks involved in the active usage of more than two languages, including negative language contact phenomena like interferences, negative transfer or overgeneralization of language rules, code-mixing, tarnishing language quality, language shift, and language endangerment. Yet, the advantages obtained from the multidimensional appreciation of multilingualism overwhelm the disadvantages, which constitutes the motivation in developing this chapter. The review of most of the empirical research and evidence(s) that capture the varied claims of multilingual advantages is herein categorized (as indicated above) into two functional units: the individual and the societal (i.e., group, institutional, or national) multilingualism.

3. Advantages of individual bilingualism and multilingualism

In this section, we appraise empirical evidence of advantages to individuals with respect to cognitive development, aging complications, linguistic awareness, communicative competence, academic or educational performance, as well as sociocultural and economic benefits.

3.1. Advantages of bilingualism in cognitive development

Asserting that bilinguals or multilingual individuals have **more improved cognitive abilities** than monolinguals is not a sentimental claim, but one substantiated by scientific experiments conducted to validate the point. The 1989 publication of Foster and Reeves [24], for example, details the use of the *Ross test* for cognitive function and the *Butterfly and Moths test instruments* for the assessment of metacognitive processes in a group of English-French bilinguals receiving instruction in French and the control group of English monolinguals. They came up with the findings that "the students who had received foreign language instruction scored higher on tasks involving evaluation which is the highest cognitive skill according to Bloom's taxonomy. The linear trend analysis showed that the students who had studied French the longest, performed the best."

That bilinguals and multilinguals have proven to be more creative and apt with respect to flexible thinking ability is now a widely accepted fact supported by Landry [25] for whom bilinguals not only have "… the ability to depart from the traditional approaches to a problem, but bilingual competence also supplies them with possible rich resources for new and different ideas." Landry's statement is a conclusion arrived after his research on the evaluation of thinking figural tasks assigned to both bilingual and monolingual groups of elementary students using both *historiometric* and *psychometric research methods.* At the end of his study, the bilinguals came out with significantly higher grades than the monolinguals. Hence, the outstanding performance realized from his experimental group buttresses his claims that the bilinguals are cognitively more creative and proactive in nature.

Another study, conducted by Mohanty [26], indicates that bilinguals "**exercise […] superiority in cognitive, linguistic, and academic performances**" over monolinguals. He reported that in "[a] series of studies involving the comparison of unilingual and balanced bilingual children, with respect to the *metalinguistic hypothesis* these studies show that the bilinguals outperform the unilinguals on a number of cognitive, linguistic, and metalinguistic tasks, even when the differences in intelligence were controlled." His interpretation of the results is incontrovertible in that, after the control of all the other obvious interfering factors and variables, the only possible variable that could be contributing to the excellent performance of the experimental group was their bilingual status.

Furthermore, on cognitive-related advantages for bilinguals, Ricciardelli [27] carried out a similar research on Italian-English bilingual and Italian monolingual children, in which he measured their metalinguistic awareness, creativity, nonverbal abilities, and reading achievement through *proficiency testing.* His report states that "Results of comparison of performance on the measures of cognitive development indicate that students who demonstrated high proficiency in both

English and Italian achieved **higher scores on the creativity, metalinguistic awareness, and reading achievement tests**." The outcome of his study is not very different from what Mohanty [26] observed with the bilinguals in his research population as stated in the above paragraph.

With the neurocognitive perspective, using *neuroimaging methodologies*, Wodniecka et al. [28] provided a comprehensive account to justify their claim that "Given that the differences observed spanned several measures of structural integrity, including enhanced gray and white matter as well as enhanced long-range connectivity in bilinguals compared to monolinguals, it has been suggested that the enhanced cognitive and neural functioning in bilinguals may rely upon this enriched neural architecture." These researchers, based on their findings, concluded that "These results consequently suggest that the **bilingual executive control advantage** does indeed extend to memory as **bilinguals were selectively advantaged in recollection as opposed to familiarity judgments.**"

Furthermore, and in the same light about the advantage in cognitive development, Rodriguez [29] investigated the effect of bilingualism on the cognitive development and linguistic performance of children at various ages living in the same cultural environment. Here, abstract thinking was measured using *verbal and nonverbal cognition test*. He reported from his findings that "The bilingual children used higher order rules more frequently than the monolingual children. The evidence seems to suggest that **bilingualism may scaffold concept formation and general mental flexibility.**" The avalanche of research evidence discussed underscores the fact that bilinguals are undoubtedly endowed with cognitive benefits as a result of their capacity to use and process two or more linguistic codes.

3.2. Bilingualism and delay of Alzheimer's disease or aging mental disorder

Studies conducted on patients with cognitive complaints (dementia or Alzheimer's disease) in a memory clinic, using the measurement of *the rate of decline in Mini-Mental State Examination (MMSE) scores*, as recorded over a span of 4 years from diagnosis date, indicated that there is a **significant delay of the symptoms of old age dementia recorded in bilinguals**, which on the other hand manifests faster in monolinguals. This prevalence of delayed Alzheimer's disease in bilinguals is also what [7] 105) describe as "strong epidemiologic evidence to suggest that older adults who maintain an active lifestyle in terms of social, mental, and physical engagement are protected to some degree against the onset of dementia. Such factors are said to contribute to cognitive reserve, which acts to compensate for the accumulation of amyloid and other brain pathologies." This finding highlights multiple advantages, including **efficient and sustainable executive functioning and cognitive control, old age mental or cognitive health guarantee, and even socioeconomic benefits** since it saves healthcare expenses for the bilinguals.

The bilingual status as a major factor enhancing the cognitive reserve of bilinguals' brain atrophy (delay) in dementia symptoms has been well articulated by Schweizer et al. [30], conforming with and confirming earlier findings by the likes of Bialystok et al. [7], Craik et al. [31], and Chertkow et al. [32]. Schweizer et al. [30] did carry out, in both bilingual and monolingual Alzheimer's patients, a significant number of linear measurements of brain atrophy from the *computed tomography (CT) scan*. After controlling other variables like level

of cognitive performance and years of education, they arrived at a finding that bilinguals did manifest an increased cognitive reserve (CR) with greater amount of brain atrophy (delay) than the monolinguals—indicating **a delay in the onset of Alzheimer disease and much better cognitive performance** than would be expected from their level of dementia disease. They further emphasize that the advantaged cognitive reserve of bilinguals serves as **a protective mechanism that helps to increase their brain's potential to cope with Alzheimer's pathology**.

3.3. Bilingualism in the management of chronic aphasia

Aphasia, defined by the US National Aphasia Association as "an impairment of the ability to use or comprehend words, usually acquired as a result of a stroke or other brain injury," is "a communication disorder resulting from a stroke or traumatic brain injury" [33]. This acute or chronic condition can be better treated in bilinguals than monolinguals. Haynes' research team carried out their study on nine chronic aphasia patients under the framework of *Outcome Measurements in Aphasia* study, where the patients were exposed to "teletherapy services that combined group therapy with one-on-one therapy sessions and online *TalkPath* language exercises." The findings were that **poststroke aphasia therapy was more responsive in bilinguals** than in monolinguals. They explained that the bilinguals were able, after stroke, to undergo a process of "transfer effect" to the primary language. Their findings go to confirm Ellis et al.'s [34] claims that "… when a person who speaks two languages experiences brain damage leading to a language condition called aphasia, the second, less dominant language can be used to transfer knowledge to the primary one, helping with rehabilitation."

3.4. Bilingualism and enhancement of linguistic awareness and communicative competence

In appreciating some of the benefits of communicative ability and communicative competence of bilinguals, Dickinson et al. [35] used the results from their investigation on "whether there is a cross-language transfer of phonological awareness" to support this added advantage for bilinguals. They engaged 123 Spanish-English 4-year-olds in a number of testing techniques such as the acronymic EPAP or *Early Phonological Awareness Profile* measurement technique [36]; the *Emergent Literacy Profile* competence assessment task, abbreviated ELP [37]; and *rhyme recognition* tasks. Their findings go to support the fact that there is a significant transfer of phonological features from a speaker's first language to the second, especially when they share some phonological entry similarities. This positive transfer is no doubt a bonus to bilinguals since **it develops their linguistic competence in other languages and broadens their content exposures of different academic or literary materials across linguistic boundaries**.

Abutalebi et al. [38] employed functional *magnetic resonance imaging techniques* in the investigation of the neural correlates of language selection processes in German-French bilingual subjects during picture naming in different monolingual and bilingual selection contexts. Their scientifically aligned evidence bolsters the fact that bilinguals have an added advantage in their capacity to switch-on one language when in use and deactivate the other during a

communication act, **helping to orientate the portion of the bilinguals' brains that stimulate linguistic awareness and subsequently improving their communication skills**.

The edge that bilinguals have over monolinguals has also been shown by the Barac and Bialystok [39] investigative research on how language, cultural background, and education can **enhance cognitive and linguistic development in bilinguals**. They experimented with a population of four groups of 6-year-olds (English monolinguals, Chinese-English bilinguals, French-English bilinguals, Spanish-English bilinguals), testing them *verbally and nonverbally with executive control tasks*. The result of their findings was that all the bilingual groups performed exceedingly better than the monolingual groups and the best performances registered on the language task were produced by the bilinguals whose language of instruction was English, the testing language. The outcome of this study led the researchers to claim that **bilingual experiences foster linguistic competences in varied verbal tasks in life**. As these sets of research studies testify, the bilinguals, eventually, in their use of more than one language, **develop better language proficiency, sensitivity, and definitely more in-depth understanding and control of their mother tongue**.

The benefits of communicative poise and competence of bilinguals have also been voiced by Cook [40]. He comments that bilinguals have "a more extensive range of affordances or interpretations providing them with a greater number of options from which to choose". This leads to a view of the **bilingual as having increased competence or multi-competence** (where multi-competence is defined as the added capacity resulting from bilingualism. This notion of multi-competence was later (1992) further elaborated on by Cook thus:

> *These subtle differences consistently suggest that people with multi-competence are not simply equivalent to two monolinguals but are a unique combination … so the multicompetence state (L1+ L2) yields more than the sum of its parts, L1 and L2. Thus, a bilingual that is seeking to solve a problem in one language is thought to be able to draw on the other language and related frames of mind to bring extra cognitive capacity to bear in solving a problem.* ([41] 557).

A study carried out by Kessler and Quinn [42] on grade 6 students (monolingual English speaking and bilingual Spanish and English speaking groups), using both *standardized reading and verbalized test*, came up with the findings that the bilinguals outperformed the monolinguals in generating more complex hypotheses. They interpreted their findings to mean that the complexities of the language structures of the bilinguals are occasioned by **their ability to undertake convergent high thinking activity, and this relates to a property possessed by most bilinguals** from their studied population. This smart trait in bilinguals equips them with the inflected potential to be metaphorical in their expressions, as later captured by Baker's [43] claim that the **bilinguals are "creative" in nature**. Also, their communicative sensitivity, which was part of the findings of Kessler and Quinn [42], May et al. [1], confirms that the multilinguals, as a result of their diverse language experiences**, have high self-confidence and stronger willingness to communicate without fear of constructing non-well-formed utterances**.

As far as linguistic awareness, communication ability, and competences go, it can be briefly affirmed that all the authors outlined in this subsection of the chapter attest that the advantages of **being a multilingual speaker extend to other problem-solving aspects in life, besides the communicative resourcefulness and creativity privilege** they have over monolinguals.

3.5. Benefit of bilingualism to academic performance

When we talk of exposure as a better teacher, we align with Cook's [44] conviction that.

> *…a person who speaks* ***multiple languages has a stereoscopic vision of the world from two or more perspectives, enabling them to be more flexible in their thinking, and learn reading more easily****. Multilinguals, therefore, are not restricted to a single world-view, but also have a better understanding that other outlooks are possible. Indeed, this has always been seen as one of the main educational advantages of language teaching.*

This perspective of the benefits for bilinguals sees them as **persons with a better ear for listening […and] endowed with sharper memories on diverse issues in life.** It is but normal then to consider **the bilinguals as better problem solvers**, since their bilingual exposure provides them with multiple perspectives on issues at hand and **better critical thinking abilities**.

Knowledge of a second language also seems to coincide with high academic achievement. A study by Horn and Kojaku [45] shows that students who were in "rigorous" programs in high school, which included 3 years of foreign language study, were more likely to earn better grades in college and less likely to drop out. This finding is a pointer to the claim that **bilinguals have the potentials to attain high academic standards with less effort** than monolinguals. This position is reinforced by Curtain and Dahlberg [46] who assert that "… the positive impact of cultural information is significantly enhanced when that information is experienced through foreign language and accompanied by experiences in culturally authentic situations." Thus, experiences in learning a second language and learning another culture will facilitate teachers' interactions with their students' learning experience. In other words, competent teachers understand that a positive self-concept and a positive identification with one's culture are the basis for academic success.

Academic advantages for bilinguals have been affirmed by research conducted by Keshavarz and Astanch [47] with one group of Persian monolinguals and two groups of bilinguals. They used the *Controlled Productive Ability Test* in English to evaluate their knowledge and performance in English language. Their performances revealed a clear difference in marks between the bilinguals and the monolinguals as bilinguals performed significantly better than monolinguals. They suggest that the prior foreign language experience of bilinguals place them in vantage position in learning another language, as this helps to **improve their ability to learn and recall English vocabulary better than their monolingual counterparts.** This claim is also attested to by Murphy [48] and MacWhinney [49] who say that **bilinguals have more opportunities to transfer knowledge and language governing rules from one language to the other, making it easier for them to learn overlapping cognates in form and meaning across languages**—opportunities which of course monolinguals do not have. From their analysis, by virtue of the fact that bilinguals already know two languages, positive transfer and familiarity with language structures and rules of previous languages become very useful in their acquisition of another foreign language.

To this stack of evidences must be added (Nayak et al. [50]), whose study of a group of monolingual and multilingual subjects concluded that multilingual subjects performed better than monolinguals in learning the rules for syntax when exposed to formal teaching of such rules,

as well as exposure to tasks in syntax. Multilingual subjects were also more capable of structuring their strategies to the task and used a wider variety of different strategies. They concluded that **multilinguals, compared to monolinguals, have superior flexibility in switching strategies,** which is a skill that equips them to better handle concepts and rules governing the acquisition of languages and related aspects.

3.6. Bilingualism and enhancement of career opportunities

The advantages of being bilingual are more visible in today's job market than in the past, especially with the advent of globalization and the rapid rate of technological advancement reaching all nooks and crannies of the world. This is evidenced in a *survey* of 581 alumni of The American Graduate School of International Management in Glendale, Arizona, where most respondents said they had **gained a competitive advantage from their knowledge of foreign languages and other cultures.** They said that not only was **language study often a critical factor in hiring decisions and in enhancing their career paths, but that it also provided personal fulfillment, mental discipline, and cultural enlightenment** (cf. [51]).

Also, in recent years, the US government has expressed **a need for fluent speakers of languages other than English,** particularly in less commonly taught languages such as Arabic and Chinese (US General Accounting Office 2002). It is obvious that even official monolingual countries like the USA have recognized the advantages of having citizens with competence in more than one or two languages in its job market and that multilingualism is a force to reckon with and to encourage.

3.7. Bilingualism and information processing edge

There is evidence that **bilinguals are better equipped for information processing than monolinguals.** The fact that multilinguals have knowledge of at least two language systems provides them with **potentials to switch the functionality strategies of one language to the other as the need demands.** This vantage position of the bilinguals was articulated by Meiran [52] after her experiment with the use of *functional magnetic resonance imaging* (*fMRI*) in testing coactivation and inhibition in bilinguals during spoken language comprehension. According to the researcher, "It's like a stop light […] Bilinguals are always giving the green light to one language and red to another. When you have to do that all the time, you get really good at inhibiting the words you don't need." Benefits accrue because the bilingual brain is constantly activating both languages, choosing which to use and which to ignore. When the brain is constantly so exercised, **it does not have to work as hard as the monolingual's to perform most cognitive tasks effectively**.

Also, the simple ability of bilinguals to switch from one language to the other indicates that **they can, with limited effort, relate to reflexes and transfer of language rules, which is a pointer of general task-specific switching mechanism in operation**. This talent claim has been supported by Yehene and Meiran [53] who specify that "bilingual language switching may increase general switching efficiency, and might be stronger at a long *cue–target interval (CTI)*, which may better tap general switching abilities." They justify this claim with findings from their experiment on 80 Spanish-English bilinguals and 80 monolinguals engaged on a

color-shape switching task and an analogous language-switching task, varying CTI (*short vs. long*) in both tasks. At the end of the study, it was determined that the cost of task-switching among the bilinguals was smaller, more economical, and more cost-effective than that required by the monolinguals. They explain that these findings substantiate the association between bilingualism and the **switching efficiency trait, an advantage in fast and successful diversifying task engagements in the life of a bilingual,** especially with respect to information processing.

3.8. Sociocultural advantages of bilingualism

It is now well accepted that language and culture are inextricably linked because, as articulated in Chumbow [11, 54], language is the means by which people who share the same culture express their belonging to a common or shared cultural experience. Considering the fact that culture is transmitted and expressed through language, multilinguals are obviously multicultural individuals who can adapt different cultures more easily. The implication is that they are better placed to handle cultural shocks than monolinguals. This is closely related to **intercultural competence,** summarized by Curtain and Dahlberg [46] who states that "… exposure to a foreign language serves as a means of helping children to intercultural competence. The awareness of a global community can be enhanced when children have the opportunity to experience involvement with another culture through a foreign language." Here, intercultural competence is considered as the ability for successful communication with people of other cultures. This ability can exist already at a young age or be developed and improved, thanks to willpower and competence [55]. In any case, high intercultural competence in multilinguality **enables multilingual individuals to appreciate and accommodate any foreign culture easier than monolinguals.**

Cook [44] supports the sociocultural advantage by stating that "multilinguals can expand their personal horizons and — **being simultaneously insiders and outsiders** — see their own culture from a new perspective, not available to monoglots, enabling the comparison, contrast, and understanding of cultural concepts." Thus, **multilinguals can better understand and appreciate people of other countries, a factor in lessening racism, xenophobia, and intolerance**, since the learning of a new language usually brings with it a revelation of a new culture.

Related to cultural flexibility, one opportune factor for bilinguals is **their ability to excel in conflict resolution tasks.** In two experiments, using the *Flanker task test* [56], both monolinguals and bilinguals were asked to perform a flanker test task under two low-monitoring and two high-monitoring versions of congruent and incongruent trials, respectively. The result was that the bilinguals had the overall reaction time during the high-monitoring condition in which they outperformed the monolinguals. This indicates that **the bilinguals could more easily affect the monitoring processes involved in executive control during conflict condition** than their monolingual counterparts.

3.9. Economic advantages of bilingualism

It has been asserted that "language is power" and it is an invisible force that can penetrate visible social and economic boundaries [57]. Being multilingual can be considered as a form

of human capital for **it can afford one the opportunity of earning higher income and obtaining aspiring employment status in any influential society**. A research study by Di Paolo and Tansel [58] shows that in the Turkish labor market, knowledge of Russian and English as foreign languages, on the average, **brings about positive earning differentials for individuals** (20 and 10.7%, respectively). These differentials increase with the level of competence. Knowledge of French and German is also positively rewarded in the Turkish labor market, although to a lesser extent (8.4 and 8.2%, respectively). In Williams' estimation, the use of a second language in the workplace is associated with positive earning differentials, ranging from 3 to 5% in different Western European countries [59].

Since individuals make a society, the higher the number of residents with foreign language competence in a community, the more benefits that community will get from the positive attributes and affluence that come with bilingualism and multilingualism. Most of the advantages of individual bilingualism aggregate to quantifiable economic gains for the individual and the society or community. That is why some countries like Britain, America, France, and Holland which are officially monolingual realize the benefits of official bilingual or multilingual nations by virtue of intense de facto unofficial bilingualism and multilingualism attested within the nation state. These noted national advantages of multilingualism transition us into the discussion of the advantages that bilingual and multilingual states get.

4. Advantages of societal multilingualism

The advantages that any multilingual society would gain outrightly supersede that from a monolingual society in many dimensions given the aggregate advantages of individual bilingualism as seen above. However, considering societies as multilingual does not bring out the same vivid anticipation as in individual bilingualism or multilingualism. When it is societal multilingualism, the state has to install institutions and policies to legally and officially implement the different languages concerned, clearly defining their societal functional domains and so creating an environment of diglossia or multiglossia. This is explained by Fishman [60, 61] as follows: "Diglossia (Greek root for two languages, *di-glossia*), the use of two languages for different purposes *in a societal group*, is different from bilingualism (Latin roots for two languages, *bilingual*), which is the use of two languages *by individuals without societal support*." From this background, it is clear that individual bilingualism or multilingualism may or may not be officially supported. Therefore, multilingual advantages become selective and restricted to those who practice it in any society, whereas societal multilingualism is officially and legally backed with policies and structures to implement them within that confined society, putting in place language planning functional operation and implementation processes of revalorization, revitalization, instrumentalization, and intellectualization of the different languages recognized in that society (see [62–64]).

Properties of societal bilingualism or multilingualism are better judged or derived from the societies that operate on the premise of the status of official bilingualism or multilingualism such as *Belgium*, with the Dutch-speaking Flemish in the north and French-speaking Walloons in the south [65]; *Cameroon*, stemming from a union of territories of the British and

French colonial heritages; *Canada*, which has the original Protestant British and the Catholic French Quebec coming together; and *Switzerland*, where four national languages, French, German, Italian, and Romansch, are recognized from four cantons or territories ([66, 67]). It can be shown that these countries experience extra societal benefits partly due to their official multilingual-multicultural practices beyond official monolingual nations, even if they propagate unofficial linguistic heterogeneity. Some of these advantages linked to multilingual societies are manifested in the domains of economy, culture, education, security, and health. It must, however, be noted that multilingualism in most countries of the world, especially in Africa, is yet to be managed in such a way as to derive the full benefits of societal multilingualism. According to [63] "All African countries are multilingual in varying degrees; from two or three languages in Lesotho, Swaziland, Rwanda and Burundi to over 450 in Nigeria" [68]. The languages of the various ethno-linguistic communities of the nation are yet to be developed and used for development initiatives as is the case in Belgium or Switzerland for instance.

4.1. Economic and business societal advantages

Multilingualism has been attributed the strength of promoting mobility of the labor force in a single marketplace, thereby **fostering employment heights and subsequent economic growth in the society**. Thus, if any nation wants to benefit from this type of economic force that can be easily generated by the mobile labor force, it has no choice than to operate a multilingual context. It is in this same light that the European Commission [69] explains that "language skills are presented as a type of ability that contributes to economic prosperity, an asset that increases the competitiveness of European companies, and a form of human capital that can positively affect citizens' employability." In this document, the Commission argues that **multilingual skill opportunity in any society will generate a mobile workforce environment, which is a huge labor market asset for that society**. In the Council Conclusions on Language Competences to Enhance Mobility [70], language skills are presented as "an essential component of a competitive knowledge-based economy. Knowledge of foreign languages is a life-skill for all EU citizens, enabling them to enjoy both the economic and social benefits of free movement within the Union." It has also been attested that in Switzerland, skills in foreign or second languages (limited to English, French, and German) contribute to some 10% of the Swiss GDP, with English accounting for half of this percentage (cf. [71]). From all these studies and their skewed findings toward one direction, a positive direction, it is right to say that a multilingual society is exposed to more economic benefits than a monolingual one.

The Kiplinger Washington Editors in 1996 stated that in the USA, the Hispanic share of the work force would increase by 25% by 2010 and the Asian share by 50% and minorities would keep moving up the corporate ladder during the following 15 years. It asserted that managers who knew how to deal with a diverse work force would be advantaged. The analysis of this report underpins the fact that **the acquisition of a foreign language or being a bilingual is an added advantage over being monolingual because it broadens the margins of the speaker's choices in the job market.** With the globalization phenomenon and increasing advancement

of technology, a **wide range of sociocultural exposure is very much in high demand, and this quality can be easily obtained by knowing and understanding different cultures and languages**. Thus, individuals who can communicate in at least two languages **are considered as assets to the communities** in which they live and work [72].

4.2. Societal sociocultural advantage of bilingualism or multilingualism

Sociocultural integration is a vital element for developmental and general societal growth since the current rate of globalization, modernization, and technological advancement has reduced the world into a global village with a complex mixture of cultures and social values. In this vain, being multilingual and ipso facto multicultural is an added advantage to any society, as summarized in [73]:21): "**Providers of basic services (health, school, local authorities and courts) are increasingly in need of communicating with people speaking other languages** [whereas] their staff is not trained to work in languages other than their mother tongue and do not possess intercultural skills." The EC statement here attests to the growing need of multilingual status to serve as a condition for social and economic integration in any society.

The European Commission [73] further suggests that **multilingualism helps in strengthening social cohesion, intercultural dialog,** and European construction (described as *social inclusion)*, which can be defined as the process by which people resident in a given territory, regardless of their background, can achieve their full potential in life. Policies promoting equal access to (public) services and actions enabling citizens' participation in the decision-making processes that affect their lives are examples of efforts to enhance the said social inclusion. On the other hand, *social cohesion* is related to the feature of a society in which all groups have a sense of belonging, participation, inclusion, recognition, and legitimacy. This coalesces in what has been labeled *intercultural dialog*, defined by the Council of Europe as "an open and respectful exchange of views between individuals and groups belonging to different cultures that leads to a deeper understanding of the other's global perception." And these advantages of intercultural relationship in a society are further elaborated in the EC [69] which states that "Language skills facilitate intercultural dialogue because they increase the capability of EU citizens to understand the culture of other fellow Europeans (and migrants), thereby contributing to European integration" ([69] 29).

A 1990 sociolinguistic survey research conducted by Lambert and Taylor in both the USA and Canada, using *questionnaires*, came up with a significant number of persons in both societies endorsing multiculturalism over assimilation. They explain that "**Pluralism and multiculturalism may lead to a positive attitude, not only to the host and minority cultures, but to the equal validity of all cultures.** With multiculturalism at its best, certain vices like prejudice and racism… are minimized in any society" (cf. [43]: 404). This reinforces the school of thought which holds the view that a multicultural setting breeds citizens who show more respect for other people and other cultures; persons who are less stereotypical, less culturally insular, and introspective in nature, hence cultivating grounds for mutual coexistence among persons from different races, ethnicity, and linguistic backgrounds within the same nation.

The various findings from different researchers and organizations discussed above, especially the EC and EU positions, thus constitute empirical support for the fact that multilingualism comes with sociocultural benefits that can transform any community into an optimal social intercultural haven for all its citizens.

4.3. Educational advantages of societal bilingualism or multilingualism

Perhaps no other field so directly shows up the benefits of multilingualism as education and the academia. This is generated at the level of multicultural ideology of which Baker [43] writes:

> *Multiculturalism has, as one foundation,* ***the ideal of equal, harmonious, mutually tolerant existence of diverse languages, and of different religious, cultural and ethnic groups in a pluralist society****. A multicultural viewpoint is partly based on the idea that an individual can successfully hold two or more cultural identities; where it is possible to be Ukrainian and Canadian, Chinese and Malaysian, Mexican and North American. In a different sense, it is possible to be a Ukrainian-Canadian, a Chinese-Malaysian or a Mexican-North American, sometimes called the hyphenated variety…. In this sense, identities are merged; the parts become a new whole…. [and the] person becomes a more or less integrated combination of [the] parts.* ([43] 402)

The implication of Baker's ideology of multiculturalism is that individuals who have foreign language knowledge would lead a society to experience development in diverse domains, if allowed to operate in that society and to uninterruptedly use their linguistic and cultural competences.

Academic advantages as a result of multilingual competence are succinctly captured in the Moore's [74] follow-up report in *Language Matters*: "… **the lack of language skills limits researchers in their ability to engage internationally in or with their research, and in their career opportunities."** This is a well-founded empirical advantage, for a researcher who is monolingual is limited by default from accessing very important research findings of scholars of other linguistic backgrounds from his. If this is the case, there is no doubt that such researchers will be retarded in their academic career due to limited literacy and literary materials exposure.

Nieto [75] more cogently advances the advantages of multilingualism in academia, suggesting that "multicultural education will ultimately be judged by its success or otherwise in being allied to 'high quality' and 'high standards' of education. **While multicultural education may successfully increase cultural and social awareness and stimulate critical thinking skills,** *a whole curriculum* approach must also show excellence in delivering basic skills, knowledge and understandings." It is in the same light that Cummins [76] argues for an approach to "transformative pedagogy" comprising (1) education grounded in the lives of students which is (2) multicultural, antiracist, and pro-justice, (3) participatory and experiential, (4) academically rigorous with high standards of performance, and (5) culturally sensitive. He postulates that (6) students should become critical in approach, (7) enabling them to feel safe, significant, and enthusiastic to share thoughts and feelings, and (8) active in promoting social change and justice.

4.4. National security advantage of bilingualism or multilingualism

The advantage of national security in a multilingual setting is crucial to contemporary society where technology hacking in the context of a cold war is the norm, even in developing countries. It is becoming more necessary for security or defense personnel to be able to effectively interact with diverse populations in the world. The power of foreign language for national security cannot be overemphasized {as}. This is summarized by Garamone [77] as follows:

> ***Multilingualism also gives defense operations an edge in a multilingual society.*** *The United States defense department is now encouraging its defense operation personnel to study a foreign language so that* ***they can be more equipped in understanding the other culture that may be involved in the battlefield as well as in negotiations****.* ([77] 52).

The essence of encouraging multilingualism in our societies has also been strongly echoed by Simire [78]:

> *… institutions, organizations and various levels of government cannot clearly and effectively perform to the expectations of their respective communities unless they can understand and be understood by their host communities. Hence, it becomes imperative that we examine the linguistic and sociolinguistic importance of adopting a* ***multilingual approach in solving Nigeria's complex linguistic problems in public and social life at the various levels of government as well as in academic and specialized institutions,*** *in strict compliance with the country's past and current language policy.*

Simire's position is geared to seeking a long-lasting crisis-resolution strategy with the aid of the multilingual tool applications for countries like Nigeria and other African countries that operate on official multilingualism platforms.

4.5. Health sensitization flexibility advantage of societal bilingualism or multilingualism

The value of multilingualism encompasses even the specialized field of health sensitization. For even when officially monolingual, most countries have minority languages and speakers of some dialects, who form the illiterate masses of the entire population. In cases where crucial health information is disseminated only in the lone official language of that nation, the nonliterates of that official language are deprived of information in addition to their always being the less privileged and more greatly affected ones. If a society is multilingual, it stands to benefit from the flexibility of its variety of linguistic codes to reach out to all the villages and suburbs of the nation, which is a faster and more assuring approach to disseminate urgent information on health-threatening issues to a wider population, for instance. In the Singapore context, for example, Chinese vernacular languages were used in radio and television announcements, while Singlish was used in a song commissioned by the government to alert citizens to take note of hygiene precautions during the severe acute respiratory syndrome (SARS) pandemic of 2003 [79]. This happened even though Singlish was usually frowned upon and discouraged by the government. Singlish was intentionally used in this

instance on the basis of reaching out to the "less educated Singaporeans" [79]. Thus, the use of multilingualism enables all citizens to be effectively mobilized for development in their own mother tongue or *language of the heart* [80].

5. Bilingualism or multilingualism as a blessing

The avalanche of empirical findings discussed here suggests that linguistic diversity in any society should be celebrated, developed, and maintained and not combated, and this requires more than just sweet-sounding policies. Accompanying institutions, resources, facilities, and proper implementation strategies and monitoring schemes are needed to make multilingual practices effectively operational. Only proper management of multilingualism will yield expected benefits.

In considering multilingualism as a blessing and not a curse for any society, we take the standpoint of an objective evaluation of certain parameters in both multilingual and monolingual societies, adopting the *linguistic disenfranchisement rate* proposed by the European Commission (cf. [81]) in evaluating the effectiveness of multilingual regime policy for European Union states. *Linguistic disenfranchisement rate* as defined by Gazzola et al. [81] is an analytic approach to give an objective picture of the benefits any society with **a multilingual regime stands to gain and exposes its country to sustainable evolution and development for its citizens**. This designed indicator check and effectiveness evaluating tool for multilingualism advantages involves "the inputs, the outputs, and the outcomes of a language regime." The *inputs* are defined as the human, regulatory, and material means used to implement a policy (e.g., the costs of language services such as translation and interpreting), while the *outputs* are what are directly produced through the resources employed, typically, the number of pages of translated documents or the amount of hours of interpreting per year. The *outcome* is the effect of the policy on the target population. The evaluation of the effectiveness and the fairness of a language regime must be carried out on the basis of outcomes (cf. [82]: 6). From projects of this magnitude, the conclusion is that communication as "information transfer" [82] in different domains of societal networks inclusively engage the majority of the citizens of the societies and also widens the scope of progressive and global megaphone benefits at all levels of those societies. Gazzola's [82] research on the disenfranchisement rate associated with some monolingual and multilingual European countries came up with the findings that the "percentage of citizens who potentially cannot understand EU documents (e.g., legal texts, regulations, webpages, call for tenders) and oral public discussions (e.g., meetings of the European Parliament transmitted via the Internet) because they do not master any official language" is higher in monolingual societies than multilingual ones, working on the interpretation that "the lower the disenfranchisement rate, the higher the effectiveness." Thus, one could rightly hypothesize that the opportunities and privileges of multilingual societies outnumber those of monolingual ones.

6. Conclusion

Bilingualism and multilingualism provide an enormous number of advantages and opportunities at both individual and societal levels. Global and globalizing sectors including financial services, pharmaceutical, chemical, automotive, IT, and other human contact (social) services rely heavily on language skills to operate for commerce and delivery. Some of the ramifications of multilingualism benefits stretch to diverse fields of life for both the individuals and the societies that practice them. In a nutshell, the research results discussed in this chapter hold up bilingual or multilingual experiences in most instances as assets to both the individual and the society that make use of it. These advantages contribute to reinforcing UNESCO's position (arrived at on independent grounds) in favor of linguistic diversity and the consequent action of revitalization and maintenance of endangered languages [13].

Author details

Evelyn Fogwe Chibaka

Address all correspondence to: efogwe@yahoo.com

University of Buea, Cameroon

References

[1] May DR, Gilson RL, Harter LM. The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. Journal of Occupational and Organizational Psychology. 2004;**77**:11-37

[2] Gollan TH, Montoya RI, Werner G. Semantic and letter fluency in Spanish-English bilinguals. Neuropsychology. 2002;**16**:562-576

[3] Gollan TH, Bonanni MP, Montoya RI. Proper names get stuck on bilingual and monolingual speakers' tip-of-the-tongue equally often. Neuropsychology. 2005;**19**:278-287

[4] Wodniecka Z, Cepeda N. Beyond the classroom: Bilingualism. Cognitive Skills and Heath. A Journal for Language Teachers. 2007;**9**:3-8

[5] Byram M. Teaching and Assessing Intercultural Communicative Competence. Multilingual Matters; 1997

[6] Baker C, Jones SP, editors. Encyclopedia of bilingualism and bilingual education. In: Multilingual Matters. 1998

[7] Bialystok E, Craik FI, Freedman M. Bilingualism as a protection against the onset of symptoms of dementia. Neuropsychologia. 2007;**45**(2):459-464

[8] Ewert A. L1 syntactic preferences of polish adolescents in bilingual and monolingual education programmes. Morphosyntactic issues in second. Language Acquisition. 2008;**29**:47

[9] Paradowski MB, Cook VJ. Rola native speakera w nauczaniu języków obcych. Perspektywa pedagogiczna, psychologiczna i praktyczna. Kształcenie językowe w szkolnictwie wyższym. Warsaw: Academica; 2009:163-188

[10] Grin F, Gazzola M. Assessing efficiency and fairness in multilingual communication: Theory and application through indicators. In: Berthoud A-C, Grin F, Lüdi G, editors. Exploring the Dynamics of Multilingualism. Amsterdam: John Benjamins; 2013. pp. 365-386

[11] Chumbow BS. Linguistic diversity, pluralism and national development in Africa. Africa Development. 2009;**XXXIV**(2)

[12] Ewert A. Do they have different L1s. In Bilinguals' and monolinguals' grammaticality judgements. Paper delivered at the 9th Nordic Conference on Bilingualism, Joensuu. 2006

[13] UNESCO. Language Vitality and Endangerment, Ad Hoc Expert Group on Endangered Languages. Ms. Paris: UNESCO; 2003a

[14] UNESCO. Education in a Multilingual World. Paris: UNESCO Position Document; 2003b

[15] European Commission. Final report: High level group on multilingualism. Luxembourg: European Communities. 2007. Retrieved from: http://ec.europa.eu/education/policies/lang/doc/multireport_en.pdf

[16] Li D. Pragmatic socialization. In: Encyclopedia of Language and Education. US: Springer; 2008. pp. 2648-2660

[17] Skutnabb-Kangas T, McCarty T. Key concepts in bilingual education: Ideological, historical, epistemological, and empirical foundations. In: Cummins J, Hornberger N, editors. Language Policy and Political Issues in Education, Volume 1, Encyclopedia of Language and Education. 2nd ed. New York: Springer; 2008. pp. 3-17

[18] Vildomec V. Multilingualism. Leyden: A.W. Sythoff von Studnitz; 1963

[19] McArthur T, editor. The Oxford Companion to English Language. Oxford: Oxford University Press; 1992

[20] Edwards J. Multilingualism. London: Routledge; 1994

[21] Aronin L, ÓLaoire M. Exploring multilingualism in cultural contexts: Towards a notion of multilinguality. In: Hoffmann C, Ytsma J. editors, 2004:11-29

[22] Bassetti B, Cook V, editors. Language and Bilingual Cognition. Oxford, UK: Psychology Press; 2011

[23] Baker C. Foundations of bilingual education and bilingualism (Vol. 79). In: Multilingual Matters. Buffalo, Toronto: Clarendon; 2011

[24] Foster KM, Reeves CK. Foreign language in the elementary school (FLES) improves cognitive skills. FLES News. 1989;**2**(3):4

[25] Landry RG. The enhancement of figural creativity through second language learning at the elementary school level. Foreign Language Annals. 1973;**7**(1):111-115

[26] Mohanty AK. Bilingualism and cognitive development of kond tribal children: Studies on metalinguistic hypothesis. Pharmacopsychoecologia. 1992. psycnet.apa.org.

[27] Ricciardelli LA. Two components of metalinguistic awareness: Control of linguistic processing and analysis of linguistic knowledge. Applied PsychoLinguistics. 1993;**14**(3): 349-367

[28] Wodniecka Z, Craik FI, Luo L, Bialystok E. Does bilingualism help memory? Competing effects of verbal ability and executive control. International Journal of Bilingual Education and Bilingualism. 2010;**13**(5):575-595

[29] Rodriguez YG. The effects of bilingualism on cognitive development. (EdD, ProQuest Information & Learning/Temply University). Dissertation Abstracts International. 1992;**53** (4-A):1104

[30] Schweizer S, Adam H, Tim D. Extending Brain-Training to the Affective Domain: Increasing Cognitive and Affective Executive Control through Emotional Working Memory Training. httpsi://doi.org/10.1371/journal.pone.0024372

[31] Craik FIM, Bialystok E, Freedman M. Delaying the onset of Alzheimer's disease: Bilingualism as a form of cognitive reserve. Neurology. 2010;**75**(19):1726e1729

[32] Chertkow H, Whitehead V, Phillips N, Wolfson C, Atherton J, Bergman H. Multilingualism (but not always bilingualism) delays the onset of Alzheimer's disease: Evidence from a bilingual community. Alzheimer Disease and Associated Disorders. 2010;**24**(2):118e125

[33] Haynes RB, Jeffery RA, Keepanasseril A, Wilczynski NL, Navarro-Ruan MRA, et al. Methods for trials of interventions to enhance patient adherence to medication prescriptions, based on a systematic review of recent randomized trials. Clinical Research Trials. 2015;**1**:20-25. DOI: 10.15761/CRT.1000106

[34] Ellis C, Lindrooth RC, Horner J. Retrospective cost-effectiveness analysis of treatments for aphasia: an approach using experimental data. American Journal of Speech-Language Pathology. 2013: epub. PubMedID. 24105475. DOI: 10.1044/2013_AJSLP-13-0037

[35] Dickinson D, McCabe A, Clark-Chiarelli N, Wolf A. Cross-linguistic transfer of phonological awareness in low-income Spanish and English bilingual children. Applied PsychoLinguistics. 2004;**25**:323-347. http://dx.doi.org/10.1017/S0142716404001158

[36] Dickinson DK, Chaney C. Early Phonological Awareness Profile. Newton, MA: Education Development Center; 1997a

[37] Dickinson DK, Chaney C. Emergent Literacy Profile. Newton, MA: Education Development Center; 1997b

[38] Abutalebi J, Annoni JM, Zimine I, Pegna AJ, Seghier ML, Lee-Jahnke H, Lazeyras F, Cappa SF, Khateb A. Language control and lexical competition in bilinguals: An event-related fMRI study. Cerebral Cortex. 2008;**18**:1496-1505. DOI: 10.1093/cercor/bhm182

[39] Barac R, Bialystok E. Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. Child Development. 2012;**83**(2):413-422

[40] Cook VJ. The poverty-of-the-stimulus argument and multicompetence. Interlanguage Studies Bulletin (Utrecht). 1991;**7**(2):103-117

[41] Cook VJ. Evidence for multicompetence. Language Learning. 1992;**42**(4):557-591

[42] Kessler C, Quinn ME. Language minority children's linguistic and cognitive creativity. Journal of Multilingual and Multicultural Development. 1987;**8**:173-186

[43] Baker C. Foundations of bilingual education and bilingualism. (Vol. 27). In: Multilingual Matters. Buffalo, Toronto: Clarendon; 2003

[44] Cook VJ. Using the first language in the classroom. Canadian Modern Language Review. 2001;**57**(3):402-423

[45] Horn L, Kojaku LK. High School Academic Curriculum and the Persistence Path through College: Persistence and Transfer Behavior of Undergraduates 3 Years after Entering 4-Year Institutions (NCES 2001-163). Washington, DC: National Center for Education Statistics; 2001

[46] Curtain HA, Dahlberg CA. Languages and Children, Making the Match: New Languages for Young Learners. Allyn & Bacon; 2004

[47] Keshavarz MH, Astanch H. The impact of bilinguality on the learning of English vocabulary as a foreign language (L3). International Journal of Bilingual Education & Bilingualism. 2004;**7**(4):295-302. DOI: 10.1080/13670050408667814

[48] Murphy S. Second language transfer during third language acquisition. Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics. 2003;**3**(1):1-21

[49] MacWhinney B. A unified model. In: Elias NC, Robinsons P, editors. Handbook of Cognitive Linguistics and Second Language Acquisition. Mahwah, Nj: Lawrence Erlbaum Associates; 2007. pp. 341-371

[50] Nayak N, Hansen N, Krueger N, McLaughlin B. Language-learning strategies in monolingual and multilingual adults. Language Learning. 1990;**40**(2):221-244

[51] Grosse CU. The competitive advantage of foreign languages and cultural knowledge. The Modern Language Journal. 2004;**88**(3):351-373

[52] Marian V. Language-dependent memory. In HAHN, Martín; STONESS, Scott C. Proceedings of the twenty-first annual conference of the Cognitive Science Society. Lawrence Erlbaum Associates; 1999

[53] Yehene E, Meiran N. Is there a general task switching ability? Acta Psychologica. 2007;**126**:169-195. http://dx.doi.org/10.1016/j.actpsy.2006.11.007

[54] Chumbow BS. Towards a Language Planning Model for Africa. Journal of West African Languages. 1987;**17**(1):51-69

[55] Fantini AE, Arias-Galicia F, Guay D. Globalization and 21st Century Competencies: Challenges for North American Higher Education. Western Interstate Commission for Higher Education; 2001

[56] Costa A, Santesteban M, Ivanova I. How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006;**32**(5):1057-1074. DOI: 10.1037/0278-7393.32.5.1057

[57] Chumbow BS. Language and Knowledge Production in the Age of knowledge Economy. Paper presented at the World Humanities Conference, Liege, Belgium, 5-12 August 2017; 2017a

[58] Di Paolo A, Tansel A. Returns to foreign language skills in a developing country: The case of Turkey. The Journal of Development Studies. 2015;**51**(4):407-421

[59] Williams DR. Multiple language usage and earnings in Western Europe. International Journal of Manpower. 2011;**32**(4):372-393

[60] Fishman JA. Bilingualism with and without diglossia: Diglossia with and without bilingualism. Journal of Social Issues. 1967;**23**:29-38

[61] Fishman JA. Bilingualism and biculturalism as individual and as societal phenomena. Journal Multilingual and Multicultural Development. 1980;**1**:3-15

[62] Chumbow BS. The African academy of languages and the continental language development Programme. In: Christina T, Legere K, editors. North-South Contributions to Africa Languages. Koln: Rudiger, Koppe, Verlag; 2011

[63] Chumbow BS. The Challenge of Diversity and Pluralism: The Tier Stratification Model of Language Planning in a multilingual Setting. In: Lopez-Varela A. editor. Social Sciences and Humanities: Applications and Theory. Vol.1. Rijeka: INTECH Publications; 2012a

[64] Chumbow BS. Towards a framework for language Laws and Language charters in Africa. In: Claudine B et al, editors. Law, Language and the Multilingual State. University of the Free State: Sunmedia Publications; 2012c. p. 2012

[65] Lijphart A. Conflict and Coexistence in Belgium: The Dynamics of a Culturally Divided Society (No. 46). Berkeley, CA. University of California Press; 1981

[66] Billigmeier RH. A Crisis in Swiss Pluralism. Oxford: Holdan Books; 1979

[67] Klode G. Language contact and bilingualism in Switzerland. In: Paulston CB, editor. International Handbook of Bilingualism and Bilingual Education. Westport, CT: Greenwood Press; 1988

[68] Simons GF, Fennig CD, editors. Ethnologue: Languages of the World. 20th ed. Dallas, Texas: SIL International. Online version; 2017. http://www.ethnologue.com

[69] European Commission. Green Paper - Migration & mobility: challenges and opportunities for EU education systems, COM(2008) 423. Brussels: European Commission; 2008a

[70] Council of the European Union. Council conclusions on language competences to enhance mobility. Official Journal of the European Union, C. 2010;**372**:27-30

[71] Grin F, Sfreddo C, Vaillancourt F. The Economics of the Multilingual Workplace. London: Routledge; 2010

[72] Bamford KW, Mizokawa DT. Additive-bilingual (immersion) education: Cognitive and language development. Language Learning. 1991;**41**(3):413-429

[73] European Commission. Multilingualism: An Asset for Europe and a Shared Commitment, COM (2008) 566 Final. Brussels: European Commission; 2008b

[74] Moore P. Collaborative interaction in turn-taking: A comparative study of European bilingual (CLIL) and mainstream (MS) foreign language learners in early secondary education. International Journal of Bilingual Education and Bilingualism. 2011;**14**(5):531-549

[75] Nieto S. The Light in their Eyes: Creating Multicultural Learning Communities. New York: Teachers College Press. Creating and Digitizing Language Corpora. Palgrave Macmillan, London; 1999

[76] Cummins J. Language, Power, and Pedagogy: Bilingual Children in the Crossfire (Vol. 23). Multilingual Matters; 2000

[77] Garamone J. Official details foreign language program's progress. American Forces Press Services, U. S. Department of Defense press release, July 2, 2010. Retrieved from: http://www.defense.gov/news/newsarticle.aspx?id=59880

[78] Simire G. Developing and promoting multilingualism in public life and society in Nigeria. Language, Culture and Curriculum. 2003;**16**(2):231-144

[79] Wee L. 'Burdens' and 'handicaps' in Singapore's language policy: On the limits of language management. Language Policy. 2010;**9**(2):97-114

[80] Chumbow BS. Which language for African literature? Language of the head or language of the heart? Keynote at the 24th pan-African writers conference, Accra, Ghana, November 05-08, 2017. 2017b

[81] Gazzola M, Grin F. Is ELF more effective and fair than translation? An evaluation of the EU's multilingual regime. International Journal of Applied Linguistics. 2013;**23**(1):93-107

[82] Gazzola M. The Evaluation of Language Regimes: Theory and Application to Multilingual Patent Organisations. Amsterdam: John Benjamins; 2014

Chapter 3

Aspects and Dimensions of Bilingualism and Multilingualism in Europe

Irina-Ana Drobot

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72101

Abstract

This chapter aims to explore certain aspects and dimensions of bilingualism and multilingualism, with a focus on Europe. The issues analyzed are the following: languages coming into contact due to conquest or colonization, bilingualism and multilingualism as a reflection of political trends and contemporary lifestyles, official languages, and heritage languages. The field of language education is also treated, when it comes to the benefits of being bilingual and multilingual, which are also analyzed from the perspective of evolutionary psychology, with the claim that knowledge of several languages ensures survival and better living conditions. The conclusions are that bilingualism and multilingualism are a necessity and an inevitable phenomenon in today's Europe, especially due to migration and due to the need of adapting to and accepting other cultures. What is more, there is a universality of bilingualism and multilingualism throughout history.

Keywords: nationalism, culture, political science, migration, language education, language contact

1. Introduction

Today's world has made knowing several languages almost a necessity. We need foreign languages to establish business communication, to travel, and for other personal reasons, whether we wish to learn a new language for entertainment or for studying in a foreign country or to better understand a different culture. A foreign language can be chosen deliberately, or it could be necessary for someone trying to adapt to a certain community. We distinguish between **multilingualism** and **plurilingualism** as follows: the first refers to "the simultaneous presence, availability or use of several languages in a given place, territory or nation," while the

latter refers to "the simultaneous knowledge of, and acquaintance with, several languages by a given person or a given community, and the use of, or competence in, more than one language in thinking, speaking, writing and/or reading." ([1], p. 12).

Multilingualism relates to public life, as it refers to the use of multiple languages in a certain territory. Plurilingualism is related more to the personal side, although it can extend to a community.

We live in a culture where individuals often learn foreign languages as a hobby, yet this choice is not always a personal one. In some cases, even if individuals have become multilingual as a personal choice, the official language of a community can be a single one. The key to bilingualism and multilingualism is represented by language contact and interaction ([2], p. 26).

Multilingualism and bilingualism can imply accepting other cultures, getting to know them better, and adapting to them, their values, and their ways of life. Multilingualism and bilingualism mean, politically, the recognition of the way that different cultures live together and the historical basis of this symbiosis.

The issue of bilingual education dates back to first-century Rome, when there were discussions regarding which language should be taught to students: Greek or Latin ([3], pp. 109–110).

Certain countries are more closely associated with foreign language learning in the school curricula than others. According to Devlin [4], learning a foreign language is more specific to schools in Europe than to schools in America. A second foreign language is compulsory in over 20 countries in Europe, with Ireland and Scotland as exceptions, although, in Irish schools, two local languages are taught: English and Gaelic. English is the most popular choice of foreign language in European schools, followed by French, German, Spanish, and Russian, according to Devlin [4], while the United States "does not have a nationwide foreign-language mandate at any level of education." The situation in the United States is likely influenced by the fact that English has become a globalized language of understanding for and with people of other nationalities. Due to this, it is not vital for Americans to learn a second language for the sake of their professional lives.

The teaching and learning of foreign languages can be seen as a political act even in schools and language classrooms. Certain languages are spoken on a global scale, which gives one more reason for individuals to prefer these languages over others:

"Although there are around 5,000–7,000 languages worldwide, a few languages predominate in the interconnected world described above. According to the Ethnologue, 40% of the world's population speak the eight most common languages, namely Mandarin, Hindi, Spanish, English, Bengali, Portuguese, Arabic, and Russian (Gordon 2005 in Cenoz 2009, p. 1). Around 6.1% of all people speak minority languages, such as Frisian, Basque, Welsh, Irish, and Catalan in Europe." ([5], p. 12).

English, which is believed to have become a global language, is spoken officially in countries in Africa as well as Malaysia, Singapore, India, Philippines, etc. In such countries, we also encounter varieties of New Englishes, which differ in grammar, vocabulary, and pronunciation from standard British English (RP English). The varieties of New Englishes have gone through major interference from the first language of its speakers. These varieties are found in former British colonies.

It is clear that political entities such as the European Union can only encourage multilingualism. The search for a feeling of unity has always been associated with a common language. From this point of view, the European Union struggles to implement the slogan "unity through diversity" and the wish to artificially create a sense of community. The feeling of nationalism goes hand in hand with the respect for one single language: in Europe, from early modernism until the twentieth century, "a single national language" was regarded "as a guarantee of the unity and indeed the identity of a people, the language often being interpreted and praised as the purest expression of a nation's soul." The process of the standardization of a language is done through "normative prescriptions, appropriate actions for identity formation, sociocultural policies and educational policy, even measures for the unification of religious idiom by imposing specific translations of the Bible or prayer books." We could regard all these practices as part of an ideology, which has in mind a "cultural conquest" performed by means of language policy. "The same holds, of course, for variations of 'standard' language, and dialects." ([1], p. 11).

The use of a national language has been promoted through various policies, ranging from educational to sociocultural. By speaking of bilingualism and multilingualism, we automatically speak of a political pact: of two or more nations or cultures living at peace with each other, of collaboration. As Anderson [6] has taught us, a nation is, after all, an imagined community, created through common culture, values, and interests.

Before the European Union, we could establish an identity between country, culture, and language. Every country was associated with a unique language. Nowadays, the EU has declared its goal for the citizens of its member states to know several languages, which is in fact a rule: "The European Framework of Reference regulates and demands multilingual language competence in Europe. Each European citizen should understand at least two languages other than the mother tongue." ([5], p. 9). We speak nowadays of language awareness in a world that is interconnected, an awareness which is clearly necessary since we need it to communicate and to show respect toward other cultures: "Critical literacy (Fairclough 1992) has become an important democratic value in a world that has developed into an interconnected network shared by many. This requires awareness that democratic structures only exist on the surface layer as social networks are dictated by the few who own them." ([5], p. 11). Political changes have always been reflected in language policies: "Language policy has always been seen as the vehicle of nation building, and political changes usually bring language changes." ([5], p. 12).

In some countries, there are two or more official languages, for example: Canada (English and French), Belgium (Dutch, French, and German), Switzerland (German, French, Italian, and Romansh), and Singapore (English, Chinese, Malay, and Tamil). If we examine the history of those territories, we see that this attitude toward language has to do with their colonization. In Canada, French has been used since the French settlers emigrated to Canada in 1604 and English has been used in the provinces' government ever since colonization by the British. Areas where two ethnicities live together usually have two languages, such as in Romanian cities like Oradea, where both Romanian and Hungarian are spoken, or in Mulhouse, where both German and French are spoken. Areas located close to a border also often use more than one language.

Language is always part of a certain cultural identity and of the symbols and practices associated with nationalism and patriotism. In some official bilingual areas, some persons can

refuse to answer in a language that they do not consider to be their own. This raises the issues of cultural and political tensions in areas where two cultures are supposed to coexist peacefully. By recognizing two or more official languages in a territory, we recognize the strong impact of past historical events, such as conquests. In its turn, the use of a certain language indicates the belonging of someone to a certain culture and cultural identity.

Using two different languages can raise the issue of belonging to two different cultures, which can have an impact on personal identity. The novels by Hanif Kureishi (1990–2017), for example, feature characters who are half-Indian and half-British and often explore these themes. The same identity issues often affect immigrants, who come from cultures and speak languages other than those of the country in which they now live.

On the other hand, promoting a certain culture requires promoting its language. We find ourselves drawn by exotic languages, such as Japanese, whose cultures we wish to discover. Through the inclusion of Spanish dramas in Romanian television programs, the respective language and culture are promoted.

Learning different languages is an aspect of politics, and it is done through various manifestations: through wars and physical, armed conflict, or imposed through culture.

2. Languages in contact

Historical events cause languages to come into contact, as the result of invasion, colonization, annexation, resettlement, etc., and also, as is the case nowadays, for establishing business partnerships. As a result, a new language needs to be learned by people who "may become refugees, either in a new place or in their homeland." What is more, as a consequence of an invasion, "the indigenous population may have to learn the invader's language in order to prosper." (Wei 2013).

One such example is that of migrants in our contemporary world. This situation applies both to refugees and to other persons who leave their own country to find temporary work or to emigrate entirely. They have no other choice but to come into contact with another language and culture; they will need to adapt and learn the language of the country. They may join communities of their own people, if there are any in the respective country, but often, in order to find better conditions for life and work, they will need to adopt the language of the country they have moved to.

Whole populations may be forced to move due to natural disasters and find themselves faced with new languages. As a result of people's resettlement, language contact occurs. An example of such a situation is the following: "Some of the Irish and Chinese resettlements in North America were the result of natural disasters." (Wei 2013).

Religious issues can also cause people to move to other countries and deal with other languages. Religious oppression can cause people to leave the respective country and move to another country and then they will have to learn the language of the new country. "Russian speakers in Israel are a case in point." [2].

The wish to learn a culture's language out of passion and personal empathy "with a particular ethnic, cultural, or social group" is not excluded. The main issue, however, remains the economic aspect that triggers migration in search for better living conditions. Accordingly, "Economic factors account for most of the linguistic diversity of the United States and an increasing proportion of the bilingualism and multilingualism in present-day Europe." [2].

Nowadays, the movement from one country to another has become possible due to many reasons. One such example is the politics related to migration, which has been allowed among countries in Europe. When poor countries join the European Union, this allows their citizens, Romanians, for example, to move to work and live in other European countries with better economic situations.

According to Hebden [7], "Multilingualism is vital for an inclusive EU." Hebden [7] believes that the current mobility provided by the EU offers great opportunities, such as the possibility "to move to another country for study or retirement." However, attention should be paid to the "new challenges of inclusion within a host community." ([7], p. 2).

The freedom to move across Europe allows citizens to feel equal with one another, as this sort of migration is so common that it has become almost a lifestyle in our contemporary world. Media advertising of classes focused on foreign languages and cultures, some even given free online, has advertised this lifestyle as something usual nowadays. In the past, moving to another country was a more isolated phenomenon and the person felt alone in the process; nowadays, the cultural shock is minimized through all the information made available worldwide and for free on the Internet. Even traveling has become something very usual as it has become simpler than it was in the past. We get lots of advertisements praising the reduced prices for flights and hotels. The plane has become, through low cost offers, a very common means of transportation. Working abroad has become something common, especially for citizens of poorer countries of the EU, such as Romania. The business of foreign language teaching seems to be supported by the EU, as "Multilingualism: an asset for Europe and a shared commitment" was in fact a strategy adopted by the European Commission in 2008. Regional and minority languages were included in the strategy. Afterwards, "In November 2008, the Council adopted for the first time a resolution on multilingualism. They recognised the need to widen the range of languages offered in the education system." [8].

In what way do these movements influence languages? Varieties of New English, such as Malaysian English, Philippine English, Indian English, and others, borrow ways of pronunciation, grammar structures, and words from the languages spoken natively in Malaysia, the Philippines, and India. Even English as lingua franca in European countries has led to borrowings; young people especially use English words and expressions mixed into discussions in their mother language. IT terms are often borrowed from English, but so are terms that already exist in the other languages. The reactions to this phenomenon have been varied: while some countries still accept this linguistic borrowing, other countries such as France have violently opposed this practice and have argued for the need to keep the purity of the country's language. Countries such as Hungary use their language only, avoiding borrowing at all costs; they go as far as translating proper names belonging to other nations into Hungarian.

3. Bilingualism and multilingualism: a reflection of political trends and contemporary lifestyles

The current discourse on bilingualism and multilingualism reflects the political issues of the time and the political situation in various countries. We speak of language policies, of requiring that certain languages be studied in schools. In Romanian schools, minority students have the chance to study the mother tongue of their ethnic group instead of Romanian literature and culture. Such a practice shows that they are accepted and not forcibly assimilated by the culture of the country they live in. However, they also need to speak Romanian as they need to communicate with people belonging to this culture.

Belgium is another example. It is a federal state, divided into three regions: the Flemish (where Dutch is used), the bilingual region (which is bilingual), and the Walloon region (where French and German are used). Yet, it manages to preserve a sense of unity that makes it a model for the European Union, according to http://www.eutrio.be/structure-de-letat-federal-belge [9].

Immigrants and refugees are often multilingual, due to their living in multiethnic home countries as well as having to learn the language of the country they moved to ([1], p. 12).

As a reflection of contemporary lifestyles, the use of several languages has led to the creation of a new domain of linguistics, **forensic linguistics:** "The study of bi- and multilingualism, together with that of social media, represents new frontiers in the field that has come to be known as 'forensic linguistics.'" ([10], p. 671). This field investigates the way online criminals talk: "In a world connected by social media and globalization, the role of the study of multilingualism in forensic linguistics is increasing rapidly." ([10], p. 671). The trend of multilingualism has affected all aspects of our lifestyle, on- and off-line.

Of course, there is the case of the European Union, discussed by Sokol [11]. He draws attention to "the political role of language," as much more than a simple tool for communication. According to him, unity of religion and political unity, mean unity of language, as he offers an example from history: "Political and religious unification went hand in hand with linguistic unification, as occurred in the Albigensian Wars in France, the English conquest of Ireland and Wales or in the Reconquista in Spain." ([11], p. 185).The English language strives to be a common language of understanding among different countries in the European Union. This was needed as "Modern society as well as democratic politics is based on communication among citizens," and as "Everybody knows that Europe, and particularly Central and Eastern Europe, is divided by a wide variety of languages into linguistic groups, whose members are not able to communicate with each others."

According to Mkhize and Hlongwa [12], "Historically, higher education in South Africa and Africa in general has relied on foreign languages; this has become a basis for social discrimination and inequality." The authors claim that English is seen as the language of the elites, while indigenous languages are associated with an inferior status, which is a clear example of how a conqueror's language can be imposed as the elite language on the conquered population.

In today's world, we try to show respect to several languages spoken in the same territory, to respect the language and culture of minority groups, and, at the same time, to bring them together, offering a sense of unity through a common language, a lingua franca. Brexit [13], when Great Britain voted to leave the European Union, affected the translation industry in Great Britain. This was because "The EU allows the free movement of people across the continent which necessitates a lot of public spending on language services," (https://www.empowerlingua.com/brexit-will-affect-translation-industry/) which will no longer be necessary if Great Britain establishes business relationships with America and China, where the English language is used. However, despite these claims, Brexit does not represent any danger: free movement in Europe will continue, and English will still be the lingua franca for a good number of years at least. There could be in fact an advantage to Brexit: more interpreters and translators will be needed, in case English will no longer be the lingua franca. Since once the majority uses English as a lingua franca, there is no need for many translators and interpreters. Once English is no longer a lingua franca, other languages will be needed, and more interpreters and translators, as once Romanian is your native language and you can use no lingua franca, an interpreter will be needed to help you understand someone who speaks Spanish. The interpreter will need to use Romanian; otherwise, he would have had to use English, the lingua franca, and not be forced to know Romanian.

While multilingualism allows for opportunities for the flourishing of the translation industry, there is also a search for a lingua franca. For certain countries, it is English. In the case of immigrants, the lingua franca becomes the language of the country they have migrated to, as they need to communicate with the inhabitants of the respective country. In any country, there can also be immigrants of other nations, and they may need to establish business communication with them as well.

3.1. Causes and cases of multilingualism: official languages and heritage languages

In every country, there is one, but often more, official language(s), constituting the majority language, and there are also languages spoken by minorities, which can be official languages or not. The situation of immigration has raised the subject of heritage languages, defined by Valdés [14] as the immigrants' languages. The term refers to languages that are not English and is defined as follows: "In the United States the term 'heritage language' refers to a language learned at home that is different from the dominant language of the community." ([15], p. 452).

Of course, a heritage language can have different status depending on the way it is used and on the number of its speakers: for instance, Spanish is, in the United States, a second language as well as "the most widely studied foreign language in middle schools and high schools. […] And because Spanish is spoken by Latino immigrants, it is also a heritage language." Other languages, such as Romanian, Polish, and Finnish, are, in the United States, "heritage languages whose distribution is highly restricted." ([16], p. 170).

Other heritage languages are Native American languages in the USA and Canada. McCarthy discusses the situation of these languages as follows: "Most Native American languages are

no longer acquired as a first language by children. They are nonetheless languages of identity and heritage, and in this sense can and should be considered mother tongues." ([17], p. 201). These are endangered languages that need to be protected. The Native American Languages act was passed in Congress in the USA in 1990, and as a result, we find nowadays Native American immersion programs for learners.

The heritage speakers will also have to learn the standard, official language of the country they live in and they will be at least bilingual. The issue of heritage speakers has appeared due to the possibilities of mobility in today's world.

3.2. Benefits of being bilingual and multilingual

The theories showing the importance and benefits of being bilingual encourage learning at least one additional language. This practice supports the political need to set up the current "unity in diversity" and to legitimize the official bilingualism and multilingualism in countries where there are two or more official languages. At the same time, the practice is encouraged as there are professional opportunities for translators and interpreters in our contemporary world.

The benefits of being bilingual come from the need to adapt to and function in certain cities. For instance, Hong Kong is a bilingual city. Its inhabitants use both English and Cantonese. The Hong Kong Basic Law (Article 9) and the Official Languages Ordinance (Chapter 5) have established that English and Chinese are Hong Kong's official languages.

Leaving the issue of cultural adaptation aside, there are political reasons for promoting bilingualism and even multilingualism through psychological theories. Linguistic studies have proved that we can take cognitive distance from an issue in a second language. Multilingualism is promoted for political purposes, in order to help people accept the concept of unity in diversity. Persons speaking two or more languages could be regarded as more diplomatic and more likely to be on good terms with other cultures, which is in the interest of the European Union. The European Union promotes the study of languages so that its citizens will find it natural to live on good terms with other cultures.

The benefits of being bilingual (enhancing cognitive abilities, task-switching, adjustment to "C" of 7-month-old children and "less cognitive decline" for seniors) are also constantly advertised through scientific research [18].

The cognitive benefits of better attention and task-switching among bilinguals can only encourage and motivate learners to learn a second language. What is more, we are told that in today's world bilinguals outnumber monolinguals, this being another challenge we set up for ourselves, to be like the majority, and not to be left out. What is more, a definition of being bilingual is under debate, as "Language proficiency can vary from having some conversational fluency in one language, to being fully versed in reading, writing, and speaking two languages." ([19], p. 292).

Nobody wishes to be an outsider in any social group. Scientific research constantly promotes the benefits of knowing several languages and people are beginning to expect that school or work will require that they learn a second, third, or even fourth language.

The issue of multilingualism is tied to modern democracies, and it is presented as such in the description of courses. For instance, part of the description of a Bachelor in Minorities and Multilingualism at the University of Groningen claims that the main challenge of democracies nowadays is to minimize the tensions caused by minorities, as they fight for "independence from encompassing majority states. At the same time migratory movements around the globe have created all kinds of new minorities like for instance Kurds in Germany or Turks in the Netherlands." The main problem "is to find a stable balance between unity and diversity. How to deal with ethnic and cultural differences, that is the main question." (http://minorities.nl/courses/bachelor/ [20])

The topic of foreign language learning cannot be separated from its cultural and political component. Students of foreign languages cannot be without knowledge of political issues. Contemporary political situations cannot be ignored by university curricula.

At the same time, since language makes up a culture, and a feeling of nationality, in what way are we supposed to look at a second and third and even fourth language? Are they a threat or are they an act of reconciliation and understanding among cultures? How peaceful is today's world? We know that there are areas of the world still engaged in conflict and we also talk of cyberwars. If you know a second language, do you distance yourself from your own emotional expression when you use it, and if so, does this have an effect of creating a distance from your own culture? In what way? Do you become more objective or do you lose your identification with it?

Immigration could be seen as something usual in today's world. It has been shown that "Interestingly, attitudes towards immigration in Japan become more positive the more fluent a person is in English, suggesting that boosting English education may help to make the Japanese more accepting of immigration." (http://www.vdare.com/posts/the-english-language-as-a-trojan-horse-in-japan [21]) Through a second language, politics could simply attempt to encourage people to adapt to the world they will have to live in from now on. Romanian newspaper authors have drawn attention to the fact that the use of English for communication among young Romanians has made them more open to accepting immigration, and they have taken a psychological distance from their own culture. It is known that young people from Romania wish to emigrate in search of a better future. Does the use of language lead to the wish to immigrate or does the wish to live abroad aid in learning English? This is still a matter of debate and personal perspective.

3.3. Individual vs. societal bilingualism

Once an individual speaks two or more languages, we deal with individual bilingualism. In this case, we could have to do with a personal preference or decision. When a country or community uses two or more languages, the phenomenon is called societal bilingualism.

Romaine [15] wishes to establish the distinction between individual and societal bilingualism, but she believes that this distinction is difficult to make. Certain communities can be larger and others smaller. Sometimes, certain individuals are part of a small community, but they are also part of larger communities of speakers of a certain language. A community can be formed of two or more individuals, according to Romaine [15]. However, if we look at the

number of languages and countries, we find it is impossible. There is a "concentration of the world's 6,900 languages into about 200 countries," which "means that there are over 30 times as many languages as there are countries, or in other words, that bilingualism or multilingualism is present" ([15], p. 448).

Thus, contrary to the understanding we got in the introduction of this paper, European countries are not entirely monolingual. We do make this mental association, however. We do associate Spain with the Spanish language, France with the French language, and so on. Yet, these countries have minority ethno-linguistic communities that speak a language of their own in addition to the national language. Thus, for instance, Great Britain has the 'Welsh,' France has the "Bretons,' and Spain has 'Catalan' as indigenous ethno-linguistic minorities. They also have more or less large populations of immigrants speaking other languages. The European Union expects every member country to take into consideration the language needs of minorities and immigrants. In their turn, immigrants and minorities have no choice but to learn standard languages used in the countries where they now reside, and used by those with whom they wish to interact.

Romaine raises the question of the existence of nation-states, which by definition recognize several languages. This is because within the nation-states there are several groups, and each and every one uses different languages. Not all languages are official languages, used in institutions and schools. Certain groups are more powerful and this leads to their language becoming the used one ([15], p. 449).

Due to the contemporary organization of the world, it is impossible not to speak of bilingualism and multilingualism. This naturally follows the structure of states and nation-states, as well as the connections among communities.

3.4. Evolutionary psychology: knowledge of several languages for survival

Multilingualism and bilingualism could be, in certain cases, ways of keeping culturally significant languages alive. For instance, the French try to preserve the Breton language by using it in haiku poetry contests and, of course, by speaking it in a restricted community. (NB: This is not a French government initiative but a resistance initiative of Bretons who want their language recognized by France aided by the multilingualism policy of the EU).The connection between language and culture is integral to the history of the respective country. By preserving languages through bilingualism and multilingualism, they try to preserve their civilization. Even if English is a lingua franca, it does not attempt to replace the languages spoken historically in the countries in Europe. Civilizations could be defined as a combination between a particular view of the world and history. According to Mozaffari [22], civilizations are formed by "an explicit *world vision*," meaning "a set of cultural system, an ideology or a religion" and by "a *historical system*," meaning "a coherent political, military and economic system."

Since language is part of a country's and people's national identity, it is natural to regard it as part of history. Language is inseparable from history as it was used to create a nation's cultural products and identity. World vision refers to the values of a culture—which are, in turn, established and legitimized through literature and recorded history, especially in legends of local heroes. The hero could be a character in either literary or historical works, and the model of the

hero varies according to the passage of time. When literary works and historical documents are studied in schools, they impose certain values on which to build a strong national identity. This national identity is used to keep unity among the members of the community. In this way, they will protect their country and the model of the hero will motivate them to succeed.

Bilingualism and multilingualism are promoted in order to encourage adaptation to today's world, together with preservation of national identity. Evolutionary psychology explains bilingualism and multilingualism as a survival tactic for certain communities and of smaller groupings of migrants that seek to adapt and be accepted. Within this approach, "knowledge and principles from evolutionary biology are put to use in research on the structure of the human mind" (http://www.cep.ucsb.edu/primer.html [23]).

Certain countries function due to the respect shown to a multilingual community. One issue regarding the study of bilingualism from an evolutionary perspective has been the fact that "Intuitively, learning two languages is harder than learning one, yet bilingualism is prevalent in the world." [24]. However, the conclusion is the following: bilingualism is a result of the interaction among multilingual communities. It "is seen as a secondary linguistic ability - a sort of by-product." ([24], p. 1).

The common element of bilingualism and multilingualism is the need for adaptation and communication with those with whom we live. This is confirmed by Sternberg and Christiansen [25], who refer to multilingualism as something normal in the world. In India, there are "twenty-two official languages, and only 18% of the population is a native Hindi speaker. Half of the population of sub-Saharan Africa is bilingual as well."

Crowd psychology may also play a role. Once an individual is part of a crowd, and the crowd behaves in a certain way, the individual believes that this is a universal type of behavior, to which he will have to adapt. The larger the crowd, the larger the feeling of its universal behavior, and the individual's loss of responsibility [26, 27].

Once bilingualism and multilingualism are the official rule, individuals have no choice but to adapt to the crowd that has adopted the rules. If they do not conform to the crowd, they will be left out. Romaine [15] supports this claim: "Bilingualism and multilingualism are normal, unremarkable necessities of everyday life for the majority of the world's population." This does not apply only to communication related to business necessities. Students need to learn foreign languages in order to pass their exams, graduate, and then be able to apply for a job to make their living.

Certain countries have been, historically, built by immigrants, such as the USA. They were "voluntary and involuntary immigrants" ([28], p. 467).

In this case, "The sociolinguistic profile of speakers of non-English languages in the USA is a direct reflection of the country's constantly ongoing history as a nation of immigrants. Such speakers are overwhelmingly either immigrants themselves or the children of immigrants." ([28], pp. 466–467).

This could be the reason why Anderson sustained his theory of a nation as an imagined community. Since language is not common, and several languages are officially accepted,

there was a need for something common to legitimize the holding together of the American community. Group cohesiveness is an evolutionary necessity; otherwise, there would not have been any security of the community during external attacks.

3.4.1. Foreign languages

3.4.1.1. Foreign language teaching

EU has started a campaign advertising the benefits of foreign language learning with a political purpose: to keep its citizens united. Everyone wishing to learn a language can find the method that fits him or her best. If you open any Internet page, it is impossible not to find advertising regarding a wondrous method of language learning. Some promise extremely fast results; the characters used for advertising are persons that have been trying for years to learn a certain foreign language, but with no results. They needed it for their jobs and, due to the fast method, they managed to get the promotion they had been wanting for years. While skepticism regarding such methods is advisable, we cannot help but notice the campaign led by EU to unite citizens under a European identity. Raising the curiosity of its citizens toward different languages and cultures is part of the campaign. As many citizens would feel that they are losing their national identity due to constant traveling and migration, EU tells them that their language and culture will still matter and will be learned, respected and enjoyed by the other Europeans. Online correspondence courses such as EUROCOR include courses that will help their learners get a diploma to ease their access to the job they want. The courses include foreign language learning and ease the access for their learners to understand another world. In this way, every nation will feel proud that they are studied by other Europeans. It will also make them more aware of their touristic potential. Many European countries, such as France, Austria, Hungary, and others, rely on the tourist industry.

According to Hebden [7], studying a second foreign language has a political purpose: to make citizens feel included in EU. She states:

"increasing bilingual education in schools could enhance language exposure and therefore make inclusion easier, while also equipping learners with key assets for mobility without sacrificing on time spent teaching other subjects."

As not everyone can learn a second or third language in his or her family, one solution is to learn it in school. The benefit of knowing foreign languages is, according to Frijhoff et al. [1], pp. 10–11, that it brings different people together, through bringing enrichment of the vocabulary and of the new perspectives brought on reality and through the empathy with other cultures and communities.

When historical circumstances and official rules bring two or more nations together, they have no choice but to live together and communicate through learning each other's language. In this way, the functioning of the respective state is ensured.

Frijhoff et al. ([1], pp. 10–11) underline the political role of language and the political aspect of imposing it in language teaching. At the same time, they draw attention to the fact that sometimes certain languages can be in political competition with one another. They are tools used to impose "another culture, a different ideology, a new world view."

The language competition and conflict could be illustrated by cases of bilingual cities, such as the Romanian-Hungarian ones in Romania, where certain people traveling there who only speak Romanian report that sometimes Hungarian speakers who also know Romanian refuse to answer using the Romanian language.

Keeping these types of conflict under control is done through rules and through crowd psychology. Ensuring unity and respect helps keep possible armed conflicts at a distance. At the same time, multilingualism and bilingualism suggest a type of collaboration for several nations to survive together.

3.4.1.2. Language and culture

By studying foreign languages, people can minimize the culture shock they experience when moving to another place. Knowing the language of the foreign country can help them adapt more easily, especially since foreign language courses nowadays include a cultural component. Foreign language teaching methodologies debate on the inclusion of culture in foreign language curricula. According to Kramsch [29], the debate is still in progress.

The truth is that it was impossible for foreign language teaching to be separated from the culture of the respective language. However, some textbooks could have included more focus on the linguistic structures and less focus, if any at all, on the cultural background. Conversation guides, which only focus on traveling phrases or basic needs, do not include cultural aspects. Manuals for beginners in Romanian schools in the 1990s did not include cultural aspects, or they were restricted to fragments from French and English authors. They were using a different strategy: making learners adapt the language to their needs and relate it to everyday life. The awareness of cultural aspects was given later, when they would read about cultural and historical monuments such as Big Ben and the Eiffel Tower. In this way, the respective culture was felt as one's own, and then the cultural aspects were gradually introduced. Nowadays, the tendency is to introduce cultural aspects right away. The purpose is the same, to make the learner familiar with a different culture. As EU citizens, we are supposed to be already familiar with the respect and need for multilingualism. The benefit of knowing several foreign languages has reached the level of communities, not of individuals as in the past.

Foreign language teachers ask themselves the following question: "how can we develop in the learners an intercultural competence that would shortchange neither their own culture nor the target culture, but would make them into cultural mediators in a globalized world?" ([29], p. 57). This question is related to an issue that has preoccupied the members of EU ever since the beginning: will the possibility of movement make people feel less tied to their own culture and identity? EU promotes cultural values and tries to preserve the cultural and national identity of its countries in order to preserve the tourist industry, which in its turn generates income. The purpose is not only political but also economical, helping countries maintain their status and budget.

According to Hebden [7], the issue of the EU policy toward foreign language is the following: European citizens should be able to use two foreign languages. This was established in 2002 by the heads of states of the European Union, as knowing foreign languages can "make people more employable and build bridges between different cultures, leading to a more inclusive society."

Translation has prospered, creating a bridge between cultures. Foreign language learning methodologies have also come to include communication studies.

Perhaps in this way the differences between minority and majority languages would be minimized. This is one of the reasons why nowadays there is an active promotion of learning foreign languages, in schools, in universities, or within business companies and even as hobby courses.

3.4.2. Multilingualism in cyberspace

A UNESCO brochure of 2015 draws attention to issues of multilingualism in cyberspace, focusing on the fact that the Internet should be multilingual; otherwise, minorities will not have access to quality resources and communication. This would lead them not to be able to be included in current debates and online communities, "which are critical elements in support of sustainable development." ([30], p. 2).

The Internet works similarly to the real world, dividing its users into communities. They either look for their communities or have to adapt, using the lingua franca, English.

The Internet, being multilingual, helps preserve rare languages: "interest in the role that a multilingual Internet can play in the preservation, revitalization and promotion of languages" ([30], p. 6).

The Internet itself has become part and parcel of our everyday lives. Communication via Internet has become a necessary and a compulsory skill in every working environment. We need to adapt to it just like it needs to adapt to the contemporary world.

3.4.3. Yiddish as multilingualism

An example of adaptation to multilingualism is that of the Jewish people. They have been multilingual throughout history, as their communities have spread throughout Europe. Yiddish itself is a language containing elements of several languages. It was called "a fusion language" by Weinreich [31], since it can bring into one single word three elements, from German, Hebrew, and Russian, as shown in the example of the word "*shlimezalnik* (an unfortunate person)": "*schlimmazlnik* (German, Hebrew, and Russian) and the general European "doctor" gets a Hebrew plural: *doktoyrim*." (http://www.yivoencyclopedia.org/article.aspx/Language/Multilingualism [32])

As a multilingual language, Yiddish has been a connection to both Jewish culture and European culture. Its speakers "could easily adapt to speaking German (Yiddish minus Hebrew and Slavic words) and Yiddish speakers revived modern Hebrew." (http://www.yivoencyclopedia.org/article.aspx/Language/Multilingualism [32]).

The Jewish people have, historically, been largely 'nomadic,' both through choice and for the sake of escaping persecution. They can be regarded as an example of multilingualism, just like the Yiddish language. In other circumstances, problems may arise related to identity issues. The Jewish people have sought to minimize these potential issues by preserving their sense of culture and identity wherever they lived and traveled by preserving the Yiddish language, while other nationalities, which immigrated but did not maintain a single cultural language, faced identity issues.

Mok and Morris [33] draw attention to the persons who are not only bilingual, but also have to identify with two different cultures. Jewish people have tried to minimize these problems by choosing Yiddish as common language.

The contact between two different languages and cultures could lead to negative experiences, such as that of not knowing where home is. Literature has described such experiences; for instance, Chimamanda Ngozi-Adichie in *Americanah* portrays a girl who wishes to return home to Nigeria after living for a while in the US, the country where she hoped her dreams would come true. Due to their many diasporas and persecutions, Jewish people have historically adapted to living in several cultures. They have needed to make their home in various different cultures. Ashkenazi Jews especially, through their use of Yiddish, prove that the phenomenon of knowing several languages and living in at least two cultures has existed since early history.

3.4.4. The universality of bilingualism and multilingualism through history

Any history of multilingualism originates in the myth of Babel. The myth of Babel preserves the hope for a universal language that would help different nations understand each other ([1], p. 7).

We have noted the use of English as lingua franca today. However, there may be places where members of the older generations cannot speak and understand English. There is also a distinction between languages used at various levels in communities, from vernacular to regional and national, as they are "owned by specific social groups and cultural communities." ([1], p. 7).

Wright [34] underlines the way different language groups come into contact and the inevitability of bilingual and multilingual education. What is more, he gives examples from history, showing that bilingual education dates back to the age before Christ, to the ancient Greece and Rome.

Schendl [35] believes that multilingualism has been commonplace in societies throughout history and that monolingualism is an impossibility, or "a marginal and unwanted phenomenon." He also states that "There is strong evidence that this was even more true of earlier historical stages, including medieval Europe."

This could happen as part of the consequences of historical conquests and conflict. There is a conflict between conquerors, expressing their conquest through imposing their language, and nationalists, who oppose themselves to this type of domination. Through language policy, a certain perspective, with values, social order, and cultural discourse, is imposed. In Europe, using a certain language can be seen as a sign of resistance, when minorities and oppressed groups use their native language. This act can be seen as a political way of freedom from domination of a certain language and culture ([1], p. 11).

The global spread of English has been accepted as normal. This is because it has become very useful to know English and people have understood the advantages. Even advertising can be done in English in certain non-Anglophone cultures; examples are Russia and China, where there are laws that "regulate the use of English in global advertising." English is used creatively in the advertising industry, and sometimes in France, they put the French translation of the English words for explanation ([10], p. 594).

English has become part of the culture of countries that are not even Anglophone and that have not been under British rule.

4. Conclusions

This paper has dealt with several aspects and dimensions of bilingualism and multilingualism with a focus on Europe. It has examined political circumstances that lead to whole communities and even countries needing to use several languages besides the official ones. A country almost never uses only one language, as we may expect. Circumstances such as conquests and migrations lead to different communities living together. The use of English as a lingua franca will remain despite fears related to the event of Brexit. Multilingualism in Europe is, however, another reality. The social mobility leads to a need of multilingualism. Once we learn at least another language, we could sympathize with other people and cultures. At the same time, identity issues could arise, as national belonging goes hand in hand with using a certain language, a majority language. Switching within various countries and communities can lead to the experience of culture shock. Despite these issues, migration will continue, as will learning other languages, as it has become a matter of survival and better living conditions. The need for communication and cohesion of groups of various cultures has led to today's promotion of multilingualism. There is, obviously, a contradiction, since cohesion of a nation has been created throughout history not through multilingualism and through different cultures living together, but through a majority language and one single culture, with its products.

The teaching and studying of foreign language, a domain belonging to philology has, apparently, no relation to politics. Yet, perhaps not so surprisingly, we notice how much philology and politics are related, as politics create the rules for teaching and studying at all levels, including university.

The research regarding the bilingualism and multilingualism is encouraged by EU. It is a topic that is currently discussed and that will be analyzed in the future as well. As multilingualism is a contemporary issue, it will constantly raise problems and academic research will attempt to analyze them and provide solutions.

Author details

Irina-Ana Drobot

Address all correspondence to: anadrobot@yahoo.com

Technical University of Civil Engineering Bucharest, Romania

References

[1] Frijhoff W, Escale M-CK, Sanchez-Summerer K, editors. Multilingualism, Nationhood, and Cultural Identity. Amsterdam, Netherlands: Amsterdam University Press; 2017

[2] Wei L. In: Bhatia TK, Ritchie WC, editors. Conceptual and Methodological Issues in Bilingualism and Multilingualism Research. John Wiley & Sons, Ltd, Chichester, UK, 2013

[3] Butler YG. Bilingualism/multilingualism and second-language. In: Bhatia TK, Ritchie WC, editors. Acquisition, The Handbook of Bilingualism and Multilingualism. Malden, MA: Wiley-Blackwell, 2013

[4] Devlin K. Learning a Foreign Language a 'must' in Europe, Not So in America. 2015. http://www.pewresearch.org/fact-tank/2015/07/13/learning-a-foreign-language-a-must-in-europe-not-so-in-america/

[5] Finkbeiner C, White J. Language awareness and multilingualism: A historical overview, Language Awareness and Multilingualism. In: Encyclopedia of Language and Education. 2017. pp. 3-17. Switzerland: Springer International Publishing. https://link.springer.com/referenceworkentry/10.1007/978-3-319-02240-6_1

[6] Anderson B. Imagined Communities. London and New York: Verso; 2006

[7] Hebden S. Multilingualism Is Vital for an Inclusive EU – Researchers. 2016. https://horizon-magazine.eu/article/multilingualism-vital-inclusive-eu-researchers_en.html

[8] Häggman J. Multilingualism and the European Union. Europäisches Journal für Minderheitenfragen. 2010;**2**(4):191-195. DOI: 10.1007/s12241-010-0063-y Springer-Verlag 2010, Printed in Austria

[9] Belgian Presidency of the Council of the European Union, 2010. Available online at: http://www.eutrio.be/structure-de-letat-federal-belge

[10] Bhatia TK, Ritchie WC, editors. The Handbook of Bilingualism and Multilingualism. 2nd ed. London: Blackwell Publishing Ltd; 2013

[11] Sokol, Jan. European Multilingualism and its Political Consequences. 2014. http://www.jansokol.cz/2014/03/european-multilingualism-and-its-political-consequences/

[12] Mkhize N, Hlongwa N. African languages, indigenous knowledge systems (IKS), and the transformation of the humanities and social sciences in higher education. Alternation. 2014;**21**(2):10-37

[13] How Brexit Will Affect Translation Industry. 2016. Available online at: https://www.empowerlingua.com/brexitwill-affect-translation-industry/

[14] Valdés G. The teaching of heritage languages: An introduction for Slavic teaching professionals. In: Kagan O, Rifkin B, editors. The Learning and Teaching of Slavic Languages and Cultures. Bloomington, IN: Slavica; 2000:375-403

[15] Romaine S. The bilingual and multilingual community. In: Bhatia TK, Ritchie WC, editors. The Handbook of Bilingualism and Multilingualism. Chichester, UK: John Wiley & Sons, Ltd; 2013

[16] Montrul S. Bilingualism and the heritage language speaker. In: Bhatia TK, Ritchie WC, editors. The Handbook of Bilingualism and Multilingualism. Chichester, UK: John Wiley & Sons, Ltd; 2013

[17] McCarthy TL. Native American languages as heritage mother tongues. In: Language, Culture and Curriculum. Vol. 21(3). London: Taylor Francis; 2008 online

[18] Marian V, Shook A. The cognitive benefits of being bilingual. Cerebrum. Rockville, Bethesda, MD: US National Library of Medicine; 2012;**2012** Sep-Oct:13. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3583091/

[19] Martinovic I, Altarriba J. In: Bhatia TK, Ritchie WC, editors. The Handbook of Bilingualism and Multilingualism. 2nd ed. Blackwell Publishing Ltd; 2013

[20] Rifkin B, editor. Learning and Teaching of Slavic Languages and Cultures: Bachelor Minorities & Multilingualism. Groningen: University of Groningen; 2017. http://minorities.nl/courses/bachelor

[21] Sailer, Steve. The English Language as a Trojan Horse in Japan, 2017. Available online at: http://www.vdare.com/posts/theenglish-language-as-a-trojan-horse-in-japan

[22] Mozaffari M. Can a declined civilization be re-constructed? Islamic civilization or civilized Islam? In: Chan S, Mandaville P, Bleiker R editors. The Zen of International Relations. London: Palgrave Macmillan; 2001. http://www.Islam.West/Reaction.Inter, first draft

[23] Cosmides L, Tooby J. Evolutionary Psychology: A Primer, University of California, Santa Barbara, 1997. Available online at: http://www.cep.ucsb.edu/primer.html

[24] Roberts SG. An Evolutionary Approach to Bilingualism, a Thesis Submitted for the degree of Doctor of Philosophy. School of Philosophy, Psychology and Language Sciences, Edinburg: University of Edinburgh; 2013, October 2012

[25] Sternberg DA, Christiansen MH. The implications of bilingualism and multilingualism for potential evolved language Mechanisms. In Cangelosi A, Smith A, Smith K, editors. Proceedings of the Sixth International Conference on the Evolution of Language. London: World Scientific Publishing; 2006. pp. 333-340. cnl.psych.cornell.edu/pubs/2006-sc-EvoLang.pdf

[26] Greenberg MS. Corsini Encyclopedia of Psychology. Chichester, UK: John Wiley and Sons, Ltd. 2010

[27] Toch H. Psychology of crowds revisited. Contemporary Psychology. 1988;**33**(11):954

[28] Fishman JA. In: Bhatia TK, Ritchie WC, editors. Language Maintenance, Language Shift, and Reversing Language Shift. Malden, MA: Wiley-Blackwell, 2013

[29] Kramsch C. Culture in foreign language teaching. Iranian Journal of Language Teaching Research. 2013;**1**(1):57-78

[30] UNESCO. A Decade of Promoting Multilingualism in Cyberspace. Available online at: https://www.linkedin.com/pulse/decade-promoting-multilingualism-cyberspace-irm-garda, 2015

[31] Weinreich M. History of the Yiddish Language. Translated by S. Noble New Haven: Yale University Press; 2008

[32] Yivo Encyclopedia of Jews in Eastern Europe, Language Multilingualism, Yivo Institute for Jewish Research, 2010. Available online at: http://www.yivoencyclopedia.org/article.aspx/Language/Multilingualism

[33] Mok A, Morris MW. Managing two Cultural Identities. The Malleability of Bicultural Identity Integration as a Function of Induced Global or Local Processing. 2011. http://journals.sagepub.com/doi/abs/10.1177/0146167211426438

[34] Wright WE. In: Bhatia TK, Ritchie WC, editors. Bilingual Education. Malden, MA: Blackwell Publishing, Ltd., 2013

[35] Schendl H. Multilingualism, code- s witching, and language contact in historical sociolinguistics. In: Hernández-Campoy JM, Conde-Silvestre JC, editors. The Handbook of Historical Sociolinguistics. 1st ed. Blackwell Publishing Ltd. 2012

Chapter 4

Achievements of a Bilingual Policy: The Colombian Journey

Julio C. Alonso Cifuentes, Diana M. Díaz Mejía and Daniela Estrada Nates

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72114

Abstract

The Colombian journey to become a country with an educational system that develops bilingual skills in students began in 1979. More recently, Colombia National Ministry of Education presented Colombia's National Bilingualism Program 2004–2019, a policy that recognized for the first time that mastering English was a matter of advantages in competitiveness for the country. Two important goals were established under this policy for 2014. First, 40% of secondary graduates should achieve at least an intermediate level—B1 or threshold or independent user following the Common European Framework of Reference for Languages (CEFRL). Second, 20% of Bachelor degree (under) graduates should achieve at least a vantage or upper intermediate—B2 level. A descriptive approach was applied to the data and it was found that the neither goal was met. The impact of the different programs implemented should be evaluated, so goals proposed for 2025 can be achieved.

Keywords: bilingualism public policy, Colombia, language test, higher education and secondary education

1. Introduction

Colombia is a multilingual and multicultural country. The government recognizes the coexistence of different native languages and Spanish in the same territory, as well as the importance of teaching a foreign language, especially English. Spanish is Colombia's most widely spoken official language, which is present in all high schools´ levels and university curriculums, and the Ministry of National Education (MEN) has set standards for its teaching as well as measures of students' performance at the end of high school and college.

The MEN has stated main objectives for language's instruction (Spanish as a first language) that high-school graduates should be able to communicate in oral and written form, transmit information, represent reality, express feelings and esthetic properties, be responsible citizens and give meaning to their existence.

The MEN has also issued standards to suggest what needs to be taught in every grade. Although the suggested curriculum states that Spanish needs to be taught based on five factors (text production, text comprehension and interpretation, literature, communication media, and other symbolic systems and ethics of communication), the mandatory exit standardized exams only test critical reading. The oral component and literature as fundamental aspects of language are not being assessed.

MEN has found that reading and writing are being taught more systematically and efficiently, whereas the teaching of literature and oral skills have been less methodical and rigorous [1]. MEN also found that narrative texts were the types of texts mostly used by teachers. This means that all the other kinds of texts have been taught less. Despite the fact that Spanish being the primary official language, more research needs to be done to understand Spanish's learning processes and teaching in high schools and colleges.

Whereas there are basic standards of what students should be able to do at the end of each grade in Spanish, and MEN fully describes the competencies, there is nothing remotely similar for Native Languages. The MEN has not developed a curriculum to be taught nor has it included a standardized test to measure the competencies as it does in Spanish (and English). There is a route to be followed by communities to present proposals for the construction of an education program for ethnic groups. Although the mechanism exists, by 2014 only 30% of towns had submitted projects and 20% had implemented the proposals to improve the quality of their ethnic education models [2].

The Ministry of Culture has been in charge of promoting native languages, but has done so from a cultural perspective and not from an educational one. There are at least four different types of native tongues that are recognized: Amerindian languages (around 68 languages) from indigenous populations, Romani spoken by the gypsies, creole languages from African-American descent groups in San Basilio de Palenque—near Cartagena—and in San Andres Island [3]. The Colombian government, through the Ministry of Culture, has made significant efforts to acknowledge, protect, and develop native languages through Law 1381 of 2010, also known as the Law of Native Languages. Although the law seeks to promote these languages and it states that education should be bilingual where necessary, the Ministry of Culture does not ensure such educational processes. In charge of bilingual education is the Ministry of National Education (MEN), which in 2013 undertook a modification of the current Law of Education, adding items related to bilingualism. Broadly, the law states that all public schools should teach at least one foreign language and that this should not interfere with the teaching of Spanish or native languages. The law also explains that English should be considered the most important foreign language.

The Ministry of Culture has funded research to find out what is the state of each native language (how many people use it, understand it, and how are they). This Ministry has also

http://dx.doi.org/10.5772/intechopen.72114

promoted and tried to preserve native language by translating into many of those languages, the peace agreements (the process that Colombia has been living for a few years now). These initiatives hope to keep languages alive, promoting its use, and keeping people who use them informed of important political, economic, and social issues. Nonetheless, since these actions for information dissemination do not come accompanied by a pedagogical strategy to promote education in these languages or even make sure that they are being taught in schools, preservation, and promotion of the native languages have not been happening.

Colombian Government and society have not developed strategies to follow up on the application of the Law of Native Languages. There are, however, explicit efforts and policies to improve English proficiency levels within the Colombian population mainly for economic reasons. Colombian policymakers have made clear different reasons to implement a successful bilingual (English) program. First, there is substantial evidence of a positive and statistically significant relationship between English proficiency and critical economic variables. For example, [4] found a relationship between higher levels of English proficiency in population and positive economic growth. Several studies [5–7] provide evidence on the relationship of English proficiency levels of the population and international trade development. In the same vein, [5] documented the relationship between people's English proficiency levels and a higher level of integration with the global economy and a greater level of competitiveness.

Second, the Colombian Government wants to implement strategies and programs for internationalization of its economy. One condition for these to be successful is to guarantee that national citizens dominate the mastery of English, the language of international businesses, so they can truly take advantage of the opportunities of an internationalized economy. For example, to benefit from the activities carried out through ProColombia (Tourism, Foreign Investment, and Exports Promotion Agency) such as international fairs, business agendas, macro business rounds, and other promotional activities [8], it is necessary that Colombian entrepreneurs must be fluent in English.

Third, the Colombian government has implemented the Productive Transformation Program (PTP), which has identified key sectors that should propel Colombia's economic competitiveness, increase employment and productivity [9]. Some of these sectors have a close relationship with Colombian labor force English proficiency. For example, one of the selected sectors is the business processes outsourcing (BPO), which can offer accounting services among other online worldwide services. However, to reach the English-speaking market, Colombian accountants need to exhibit a good level of English to communicate with clients. Belkaoui [10] documented that the communication problem that occurs between accountants from different countries, who do not speak the same language, affects the success of the accounting process.

Another selected sector in the PTP is tourism. The Colombian government has adopted a strategy to position the country as a destination for Health and Welfare Tourism; it hopes to achieve a share between 20 and 30% of the medical tourism market by 2032. One of the strategies proposed to meet this goal is to increase health professionals' English proficiency level [11]. Additionally, Colombia was the second country recommend by **Lonely Travel** as a holiday destination because of its natural diversity, number of national parks and cultural events,

historic attractions, and for the warm hospitality of Colombians [12]. Throughout 2016, United States was the main country of origin of the tourists, with 498.960 people and a share of 24% of the total tourists that visited the country [13]. Thus, to boost the tourism sector, Colombia needs a higher English proficiency level in its labor force related to this economic sector.

Furthermore, there are numerous reasons for the implementation of a successful bilingual (English) program other than the economic ones, many of which are also aligned with the Colombian government interests. The academic literature has documented cognitive, cultural, and social benefits of being bilingual. Bialystok, from the Department of Psychology at York University in Canada, has found that bilinguals have better control of visual attention; they show advantages in various tasks where executive function is required; they have a wider range of expressive vocabulary and faster access to words as well as improved flexibility in task switching [14]. Lazaruk [15] found that bilingualism was associated with "heightened mental flexibility and creative thinking skills, enhanced metalinguistic awareness, and greater communicative sensitivity." Also, it has been proven that bilingualism can slow the effects of old age, such as the effects of dementia or Alzheimer's. Additional to the individual benefits of being a bilingual, there are social benefits. Bilinguals are believed to be more empathic, open-minded, and tolerant to difference.

The Colombian government has not been unaware of the benefits of bilingualism for almost four decades; the first public policy was designed in 1979 to strengthen English proficiency. From that year until now, MEN has promoted different public policies to increase the English language level of high school students´ and teachers'. In 2005, during the program "Educational Revolution," the MEN presented Colombia's National Bilingualism Program 2004–2019, a policy that recognized for the first time that mastering English was a matter of advantages in competitiveness for the country. As a part of that initiative, the MEN developed the Program for Strengthening the Development of Foreign Language Communication Skills (PFDCLE). Under the PFDCLE, the Colombian Government established two important goals for 2014. First, 40% of secondary graduates should achieve at least an intermediate level—B1, threshold, or independent user following the Common European Framework of Reference for Languages (CEFRL). Second, 20% of Bachelor degree (under) graduates should achieve at least a vantage or upper intermediate—B2 level.

To establish whether the goals of English Language proficiency levels are met, the government applies the SABER 11[1] test for high school students and SABER PRO test for students finishing their university program. These tests include a section that measures the level of English in reading comprehension and language use. In 2007, the test's results were aligned with the standards proposed by the Common European Framework of Reference for Languages (CEFRL). Since this framework is the most used worldwide, the government decided, initially to use it to assess the level of the students and teachers, and afterward as a guideline for teaching processes.

[1]Saber 11 and Saber PRO are part of the set of standardized test used in Colombia that seeks to assure the quality of education in the country. These sets of tests are called Pruebas Saber (Knowledge Tests) and are taken by students in third grade (SABER 3), fifth grade (SABER 5), ninth grade (SABER 9), before graduation in eleventh grade (SABER 11) and at the end of professional programs at University (SABER PRO).

http://dx.doi.org/10.5772/intechopen.72114

This decision has made it possible for researchers, investors, educators, government entities, among others, to know the level of English in a standard that can be compared with and equal to international standards. The CEFRL describes what language users should be able to do at the different stages without being language specific. Therefore, it does not consider aspects related to specific languages, like English grammar or French pronunciation, but the communicative tasks that a person can develop in their learning process. Additionally, it describes different communicative functions at the various stages.

The PRUEBA SABER[1] presents its results using this standard, so there is an international framework to interpret them. Results classify Colombian students according to the CEFRL in the following levels: A1 Breakthrough (beginner user), A2 Waystage (basic user), B1 Threshold (pre-intermediate), and B+, which it is supposed to include the higher levels: B2 Vantage (intermediate), C1 Effective operational proficiency, and C2 Mastery. Although the CEFRL refers in general to the ability to communicate considering the four skills (writing, reading, listening, and speaking) and the standards for teaching English issued by the Ministry of National Education do so too, the Colombian tests only measure reading comprehension and language use.[2]

The goal of this chapter is to determine whether students met the goals established for 2014. In particular, the aim is to determine if graduates from high school and undergraduate university programs accomplished the goal of bilingualism for 2014. To do that, this chapter describes different achievements accomplished in the main cities of Colombia (Bogotá, Medellín, Cali, Cartagena, Barranquilla, and others) and compares the types of institutions (private vs. public). Additionally, it aims to contribute to the discussion of how to implement a bilingual program in a developing country.

This chapter is divided into three sections. Section 1 covers the history of Colombian Bilingualism Policy (history) from 1979 to 2015, Section 2 focuses on presenting and analyzing the results for graduates from high school and university undergraduate programs, and Section 3 contains some final remarks.

2. Colombian bilingualism policy

For over 40 years, the Colombian Policy Makers have designed policies to strengthen English language proficiency and teaching in the country. In 1979, the Colombian Government, through the MEN, issued a decree making the teaching of English compulsory in secondary education. The teaching of English became mandatory for grades 6 and 7 and the teaching

[2]This does not necessarily mean that the schools should only teach reading comprehension, but it certainly indicates to schools what they will be measured on, and focus their efforts on. There is a difference in what is stated in the Laws and what is taught in the schools. The conception of bilingualism behind the policy advocates for a communicative approach, where students can use the language in oral and written form to communicate at different levels. It does not pretend for students to develop mastery in the foreign language similar to the one they have in their first language, but it does hope for students to be able to start communicating in English, in addition to Spanish. The reality of what is expected and what is taught is not comparable.

of French mandatory for grades 10 and 11. The decree let schools choose whether to teach English or French in grades 8 and 9 [16].

In 1982, the MEN, supported by the British Council and the Centro Colombo Americano, created the program named "The English Syllabus" (TES). TES introduced a universal English syllabus for grades 6 and 9 and 10 and 11. Under this program, the teaching of English remained mandatory only for grades 6 and 7. According to Valencia (2006), TES was the first time that policy makers established reading comprehension as the main long-term objective in the teaching of English in Colombia [17].

This program was followed by the project "Colombian Framework for English" (COFE) carried by the MEN between 1991 and 1996. As documented by USMA (2009), this project was part of a broader bilateral cooperation program between Colombia and the UK [18]. The MEN designed this project to improve the training of English teachers.

The COFE project came at the same time of a great institutional change in Colombia. In 1991, a new National Constitution was adopted. The new constitution decentralized some responsibilities, the expenditure on education being one of them. Under the new constitution, each local government (Departments and Municipalities) had a budget for, including but not limited to, security, education, planning, and development as well as the monitoring and evaluation of existing programs.

In 1994, the Colombian Parliament voted the General Education Law (GEL), which was developed after the new constitution. The GEL regulated the supply of the public service of education by private and public institutions. The GEL also introduced the notion of school autonomy, created school governance, and gave school communities the ability to define their curriculum and pedagogical processes within a general set of guidelines included in it [19]. In this new scenario, a national mandatory English Syllabus was part of the past.

The GEL included the need for a foreign language as a compulsory subject from third grade onwards. In Article 21, the policy makers make it clear that the objective of this foreign language instruction was the development of elements of conversation and reading in at least one foreign language. This law was the first sign of recognition of the importance of English proficiency for Colombian economic growth and its inclusion in a globalized economy [20].

In 1999, the MEN published the "Curricular Guidelines for Foreign languages" (Lineamientos Curriculares Lenguas Extranjeras) [21] and in 2004, they launched the "National Bilingual Program" (NBP), which was initially designed to run from 2004 to 2019. This program was explicitly designed to improve Colombian human capital to facilitate economic development. It recognized the need of English proficiency in Colombian work force for economic growth and development. The NBP adopted the CEFRL, produced the tools to evaluate students (PRUEBA SABER) and developed projects for public school teachers and standards for the teaching of English. This program was the first in Colombia that established measurable goals. For example, it stated that by 2019, all high school graduates should be at B1 Threshold (pre-intermediate) level, while all university graduates should be at B2 Vantage (intermediate) level.

http://dx.doi.org/10.5772/intechopen.72114

In 2010, the MEN transformed the NBP in the Foreign Languages Competencies Development Program (PFDCLE), which aimed to train "citizens who were able to communicate in English with internationally comparable standards, to insert the country in universal communication processes, in global economy with cultural openness." The PFDCLE kept the goals established in the NBP for 2019, but it added intermediate goals for the period 2010–2014. The goals for 2014 were: 40% of high school graduates should achieve at least B1 Threshold (pre-intermediate) level, while 20% university graduates should be at B2 Vantage (intermediate) level. The PFDCLE established four priorities: in-service teacher training and coaching, pedagogical support, evaluation and assessment, and institutional capacity building.

In 2013, the Colombian government issued the Bilingualism Law (Law 1651) that modifies the GEL making it clear that the new goal of English instruction is to "develop communication skills so students can read, understand, write, listen, speak and express themselves correctly in a foreign language" [22].

Under the PFDCLE, the MEN continued to develop programs for public school teachers, helped local governments to adopt local bilingual public policies, designed an educational curriculum and content for upper-secondary schools, and continued the assessment with SABER tests. The MEN has also promoted local programs to extend school hours for primary and secondary public schools as well as access to digital resources for students and teachers.

3. Results

The analysis carried out in this section employs the public databases containing the results of a graduation-required standardized test for high school and university undergraduate students: Saber 11 and Saber Pro. These tests are applied by the Colombian Institute for the Evaluation of Education (ICFES) and are comparable in their English component. As indicated above, this test classifies student's English reading comprehension skills according to the CEFRL in the following levels: A1 Breakthrough (beginner user), A2 Waystage (basic user), B1 Threshold (pre-intermediate), and B+, which it is supposed to include the higher levels: B2 Vantage (high intermediate), C1 Effective operational proficiency, and C2 Mastery.

Colombia has a centralized administrative structure (political organization), with 1118 municipalities grouped in 32 departments. Bogota serves as its capital. The five biggest cities are Bogota, Cali, Medellin, Barranquilla, and Cartagena. These five cities account for about 30% of all the people living in Colombia, with 16% of the Colombian population living in Bogota. The rest of the chapter shows results for the national level and for these five municipalities.

3.1. Secondary Education: Saber 11

Table 1 shows the main statistics of the variables employed in this subsection. The main source of information is Saber 11 2014 test provided by the Colombian Institute for Educational Evaluation. Information recorded by the Ministry of National Education and the National

City	Sector	Num. of students	Num. of schools	Mean students per school	SD students per school	Min. students per school	Max. students per school
Barranquilla	Private	6288	265	23.7	33.9	1	241
Barranquilla	Public	10,837	274	39.6	44.6	1	207
Bogota	Private	53,651	1405	38.2	66.4	1	1255
Bogota	Public	52,812	998	52.9	85.3	1	566
Cali	Private	16,330	594	27.5	40	1	275
Cali	Public	13,488	424	31.8	75.5	1	780
Cartagena	Private	4702	155	30.3	52	1	485
Cartagena	Public	9470	176	53.8	72.4	1	332
Medellin	Private	15,023	371	40.5	95.9	1	929
Medellin	Public	18,770	492	38.2	69.1	1	916
Other cities	Private	84,171	2944	28.6	37.1	1	474
Other cities	Public	331,797	5961	55.7	61.8	1	682

Source: Authors' calculations.

Table 1. Descriptive statistics higher education, Saber 11 2014.

Administrative Department of Statistics (DANE) about bilingual programs implemented per city and its characteristics is used.

About 32% of the country's high school graduates come from the five biggest Colombian cities, 17% of students that presented Saber 11 in 2014 were from Bogota, 5% were from Cali, another 5% were from Medellin, 3% from Barranquilla, and 2% from Cartagena, and the rest were from other towns and cities. The participation of private high school graduates is uneven across cities. For example, in Bogota, it was around 50% in 2014, while in the rest of Colombia was 20% at most. In the towns such as Cali, it was 55% and in Medellin (was) 44%, Barranquilla 37%, and Cartagena 33%.

The goal proposed by the Ministry of National Education of 40% of secondary education graduates achieving an intermediate level – B1 or independent user following the Common European Framework of Reference for Languages (CEFRL), was not achieved in 2014. Only 41,477 (6.7% of 617,536) students achieved an intermediate level (B1) or independent level (B+) (see **Figure 1**). This striking result is a signal that either the goal was very ambitious or there is a structural problem in the way English is taught in secondary education, or both. Plea for the first reason, is that the English test taken by English teachers in 2013 showed that only 43% achieved a B+ level (MEN, 2014). Therefore, it is quite difficult to expect 40% of students to achieve a B1 or B+ level. Unluckily, no information for basic education is available for doing furthers comparisons, which could give a better idea if young children are achieving the expected English level for their age, and if not, how severe is this deficiency.

http://dx.doi.org/10.5772/intechopen.72114

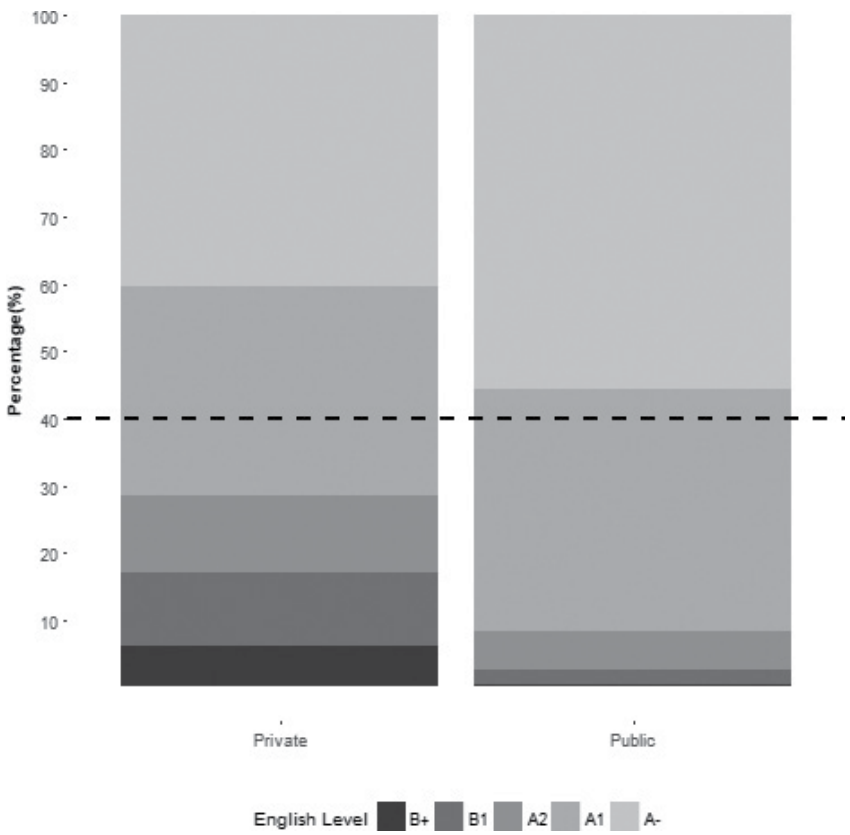

Figure 1. English levels in private and public schools, 2014.

Besides, from the goal not being reached, the information also raises a concern regarding the huge gap between the achievements made by students from private schools vs. students from public schools: 17 vs. 2.5% achieved B1 or B+, and 11.6% students from private schools achieved A2 vs. 5.8% from public schools (see **Figure 1**). Since more than two-thirds of secondary students study in public schools (71% or 437,350 students in 2014 according to Saber 11 2014), it generates many doubts as to the efficiency of the actions implemented during the Program for Strengthening the Development of Foreign Language Communication Skills (PFDCLE). Some of the strategies sought to train 9500 teachers, certifying 94 Secretaries of Education in language and methodology; they intended to implement a pedagogical program for learning English called "English, Please!" in 9th, 10th, and 11th grade in public schools, among others. Policy makers designed all these actions directed toward public schools, and they did not formally evaluate the effectiveness of these measures.

Figure 2 shows the results per city. No city achieved the government's goal. On the one hand, the cities that presented the highest percentage of students that classified in B1 and B+ were Bogota (14.3% of 106,950 students) and Barranquilla (12.6% of 17,227 students), while the city with the lowest percentage was Cartagena with 6.8%. On the other hand, almost 50% of students

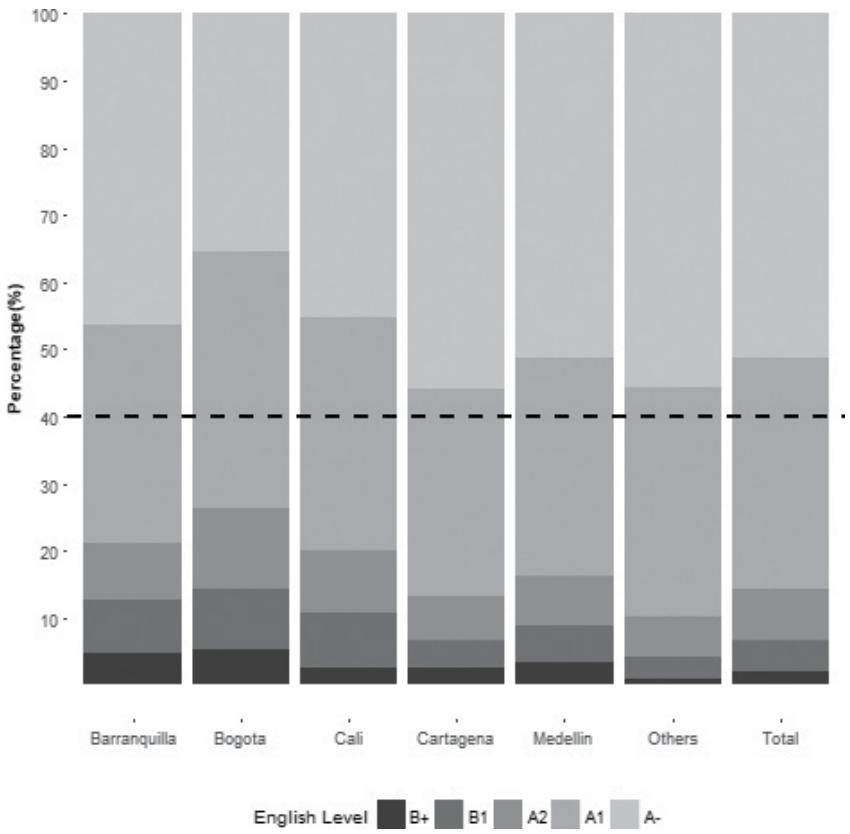

Figure 2. English levels in secondary education by main cities, others cities, and total.

of each city, except Bogota, had an A-level, this means that they have not developed the communicative competencies of a basic user, such as understanding simple phrases, asking or giving basic personal information, among others. Barranquilla's results might be related with public policy implemented during 2012–2015 "Barranquilla Bilingue" [23, 24].

Analyzing results from all cities and school sectors at the same time, it is clear that more students from private schools achieved a B1 or B+ level. However, the percentage of students in these levels was distant from the goal proposed by the government (40%) (see **Figure 3**). Barranquilla had 26.6% of students from private schools in B1 and B+, followed by Bogota with 24.5%, Cali with 17.0%, Cartagena 15.9%, and others cities with 11.9%, that does not appear to be a great difference between cities in the achievements attained by students from public schools. Cartagena is an exception because only 1.4% (133 in 9479) of students achieved B1 or B+ compared to the 4% of students who attained an intermediate level in the other main cities considered (Bogota, Barranquilla, Cali, and Medellin). Even though for the whole country, English proficiency is essential, it is of great concern that Cartagena achieved such low percentage in B1 and B+, since this city hosts one of the largest ocean terminals of the country

http://dx.doi.org/10.5772/intechopen.72114

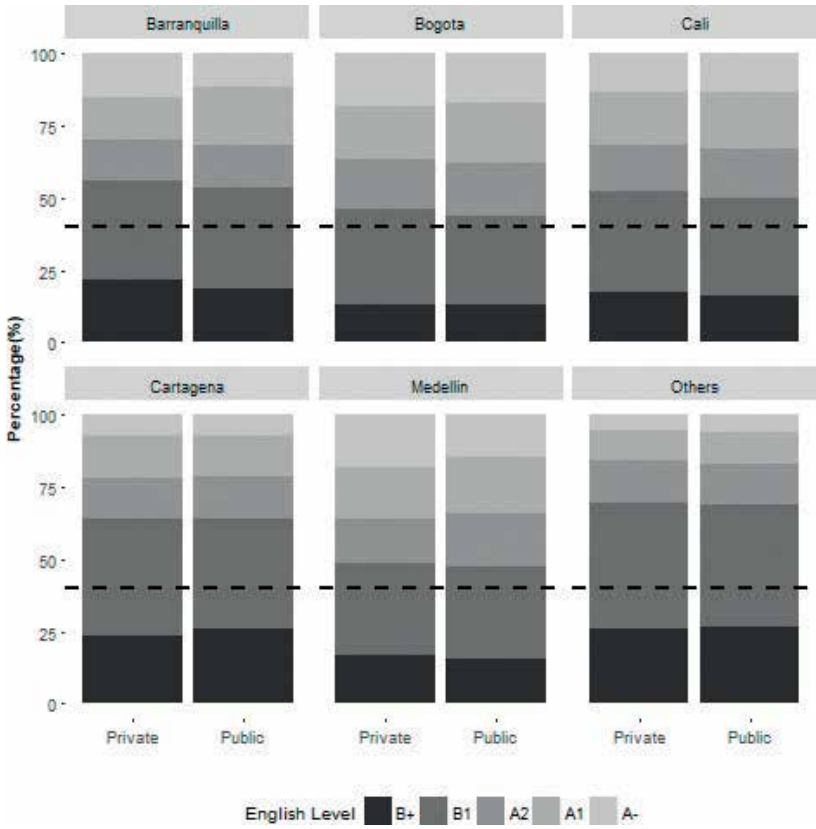

Figure 3. English levels in secondary education by main cities and others cities and type of school.

and it is the most famous tourist destination of the country. This city is a key for international trade operations and tourism in Colombia.

3.2. Higher Education: Saber Pro

Table 2 shows the main statistics of the variables from Saber Pro 2014 test provided by ICFES. It indicates that 30% of students that presented Saber Pro in 2014 were from Bogota, 7% were from Medellin, 5% were from Cali, 4% from Barranquilla, and 3% from Cartagena, and the rest were from other cities. In other words, 49% of university graduates come from the five biggest cities in the country. The participation of higher education graduates from private institutions is larger, even more than in secondary education. Bogota had the highest participation of private colleges and universities, and Cartagena had the lowest with 58%; in other cities of the country, the percentage is around 46%. Alternatively, the number of higher education institutions available for studying is higher in Bogota than in any of the other main and small cities (see **Figure 4**).

City	Sector	Num. of students	Num. of college	Mean students per college	SD students per college	Min. students per college	Max. students per college
Barranquilla	Private	6403	45	142.3	330.1	1	1321
Barranquilla	Public	2024	18	112.4	434.6	1	1853
Bogota	Private	49,520	108	458.5	771.2	1	4127
Bogota	Public	13,130	49	268	630.9	1	3015
Cali	Private	7766	53	146.5	345.9	1	1892
Cali	Public	3253	26	125.1	358	1	1823
Cartagena	Private	3828	47	81.4	153.9	1	662
Cartagena	Public	2770	22	125.9	407.6	1	1908
Medellin	Private	9632	56	172	296.8	1	1226
Medellin	Public	5522	34	162.4	390.6	1	1919
Other cities	Private	49,863	152	328	493.3	1	3240
Other cities	Public	58,072	67	866.7	1178.3	4	7292

Source: Authors' calculations.

Table 2. Descriptive statistics Saber 11 2014.

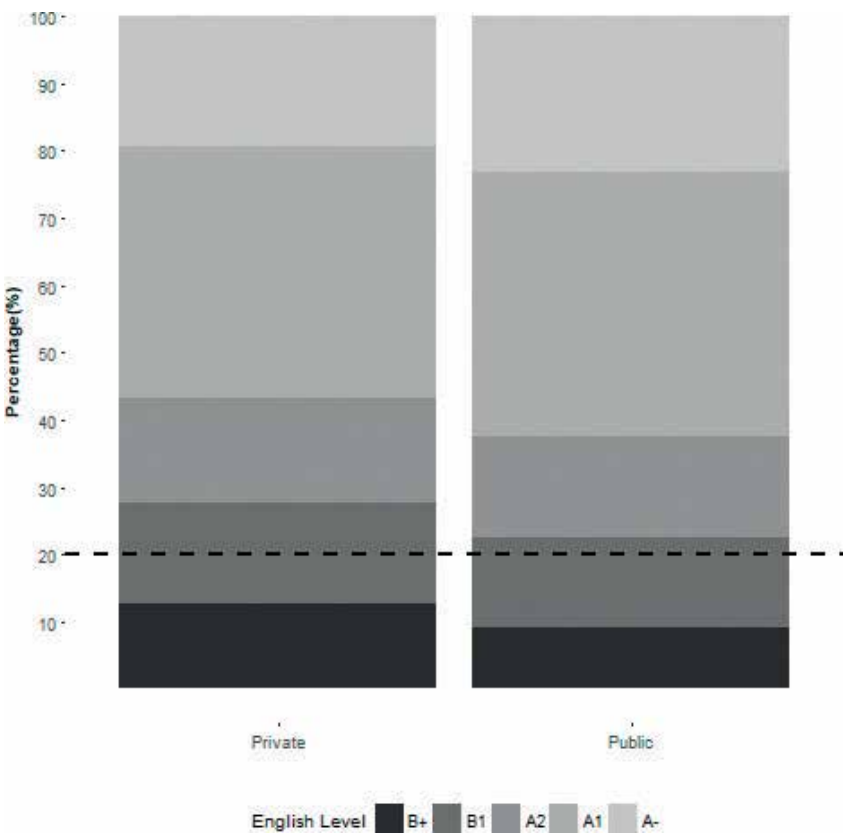

Figure 4. English levels in private and public universities, 2014.

http://dx.doi.org/10.5772/intechopen.72114

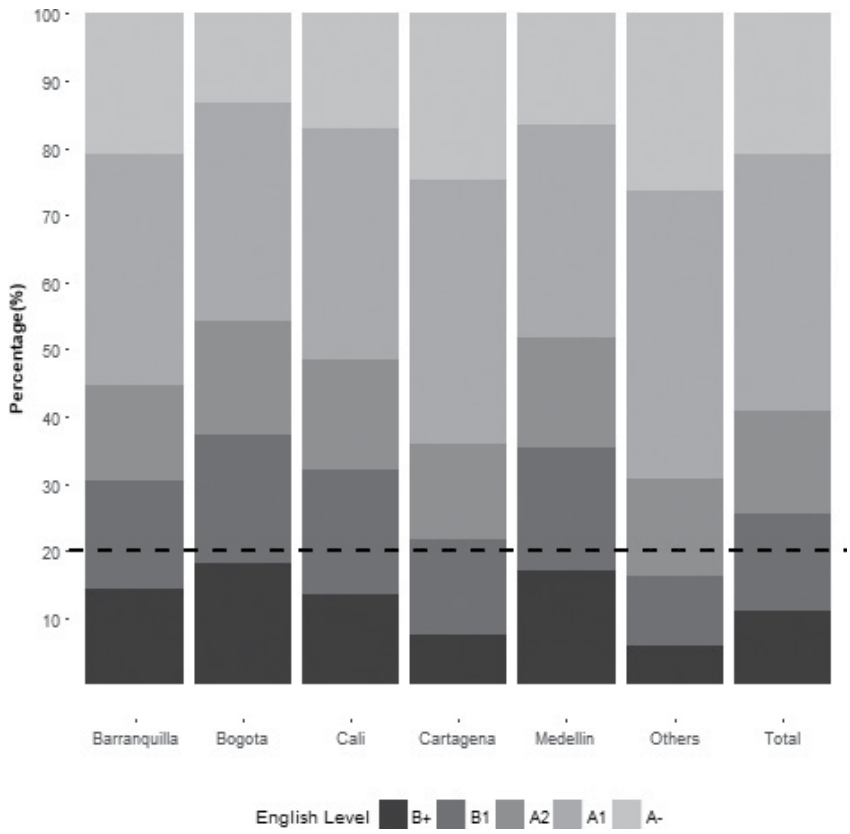

Figure 5. English levels in higher education by main cities, others cities and total.

Results by cities show the same pattern: there was not a city, which achieved the goal. The city that was closer to reaching the objective was Bogota (18.1%), and then Medellin (17.0%), followed by Barranquilla (14.2%), Cali (13.4%), and at finally, Cartagena with only 7.5% (see **Figure 5**). This value is almost 10 percentage points below the percentage of students from Bogota that attained a B+ in the English module of Saber Pro 2014. This result evidences the vast difference in education outcomes that exists among Colombian cities. On the other hand, cities with the highest percentage of students in A- were Cartagena (24.7%) and Barranquilla (20.9%). Again, as argued before, these results are troublesome for the case of Cartagena.

Figure 6 shows that more students from the private sector achieved the B+ level. However, the percentage of students in these levels is far from reaching the goal proposed by the government (20%), except for students from private colleges in Bogota (18.4%) and Medellin (18.4%). Besides that, a relevant difference between the results achieved by persons who studied in the main cities and those who studied in others cities can be seen in **Figure 6**. The average of students that attained a B+ level in main cities was 13.7% and in other cities, it was 5.9%. The gap remains when comparing the outcomes from private and public institutions: 14.5% of students from private institutions from main cities achieved a B+ level and in other cities 5.5%. In public institutions, the average in main cities was 12.9% and in other cities 6.2%. It is also interesting to note that in Cali and other cities, the public institutions are the ones that have a higher percentage of students in B+.

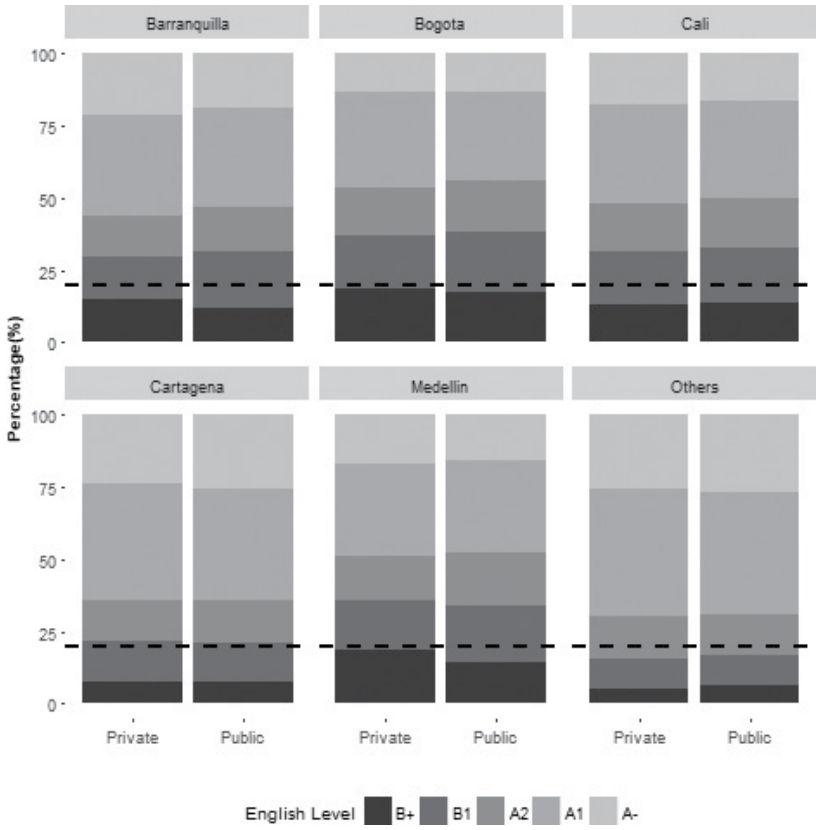

Figure 6. English levels in higher education by main cities and others cities and type of school.

4. Final observations

The Colombian journey to become a country with an educational system that "develops communication skills in order for students to be able to read, understand, write, listen, speak, and express themselves correctly in a foreign language" [22] began in 1979. Thirty-five years later, in 2004, the MEN set long-term quantitative goals to be achieved in 2019. In 2010, the MEN established intermediate goals for 2014 that could serve as a sign of how close the country was (of) to having all high school graduates at B1 Threshold (pre-intermediate) level, while all university graduates were supposed to be at a B2 Vantage (high intermediate) level.

The results presented here show that the intermediate goals were not met. By 2014, only 6.7% of the high school graduates had achieved a B1 Threshold (pre-intermediate). The goal was 40%. In other words, only 16.8% of the target was reached. On the other hand, 11.1% of university graduates were at a B2 Vantage (intermediate) level. The goal was 20%. In this case, the goal was met in only 55.5%.

What about the English Level in neighboring countries? No assessments have been made in Latin American countries that are comparable to the Colombian context; besides most

http://dx.doi.org/10.5772/intechopen.72114

of the neighboring countries have only recently proposed an English Policy. For example, Chile proposed the National English Strategy 2014–2030 [25], with the English Open Doors Program (PIAP); Peru's English policy has the Gates to the World program (2015–2021) [26] and Uruguay has the English CEIBAL PLAN since 2012. In all three countries, the policies have focused on improving students' performance in English at the initial (primary and secondary) levels of education.

Colombian results show that it is was not feasible to reach the targets set for 2019. Colombia still needs a labor force that can integrate into a globalized economy. The MEN established in 2015 a new program to assess the situation: "National English Program (2015–2025), Colombia Very Well" [27].

The National English Program (NEP) established challenging goals for 2025 for the secondary education: 50% of high school graduates should achieve at least a B1 Threshold (pre-intermediate) level. While for higher education, the goals for 2025 are similar to the ones that were not met: 30% of university graduates should be at a B1 Threshold (pre-intermediate) and 25% at a B2 Vantage (high intermediate). In this case, the MEN did not set intermediate goals, which proved to be useful to reformulate the public policy.

Although these 2025 goals do not appear to be ambitious, the recent history and results show that Colombian government and society should change their policies drastically to reach them.

During 2016 and 2017, the MEN has carried out different programs under the NEP, such as:

- English courses and incentives plans for short internships abroad for teachers from 120 targeted educational institutes [28].
- Accompaniment by foreign native English trainers to educational institutions under the methodology of co-teaching [28].
- Creation of English curriculum, which aims at helping structure English learning during the transition and primary education [28].
- Deliver the "Colombia Bilingual English Kit" (22,000 in total) to official schools of the country. This kit is a pedagogical tool aimed at strengthening methodologies and classroom practices for teaching English. It contains the Basic English Language Rights (DBA) for grades 6–11, the suggested curriculum and pedagogical principles for teaching English, and the basic standards for foreign language competence in English [29].

The impact of these programs should be evaluated. We presented in this chapter a macro view of the policy, but it is clear that to reach the goals set for 2025, further research to evaluate and formulate specific programs is needed.

Author details

Julio C. Alonso Cifuentes*, Diana M. Díaz Mejía and Daniela Estrada Nates

*Address all correspondence to: jcalonso@icesi.edu.co

Cienfi and Universidad Icesi, Cali, Colombia

References

[1] MEN. Política Pública Educativa para los Grupos Étnicos en Colombia [Internet]. 2013. Available from: https://www.mineducacion.gov.co/1759/articles-342571_Atencioneducativa_grupos_etnicos.ppt [Accessed: 2017-08-28]

[2] Fundación Compartir. ¿Cómo enseñan los maestros colombianos en el área de lengua castellana? Análisis de las propuestas del Premio Compartir al Maestro [Internet]. 2015. Available from: http://www.fundaciontelefonica.co/educacion/investigaciones/como-ensenan-los-maestros-colombianos-en-el-area-de-lengua-castellana/ [Accessed: 2017-08-28]

[3] Landaburu J. Las lenguas indígenas de Colombia [Internet]. 2017. Available from: http://www.banrepcultural.org/blaavirtual/antropologia/lengua/clas2.htm [Accessed: 2017-06-01]

[4] Lee CG. English language and economic growth: Cross-country empirical evidence. Journal of Economic and Social Studies. 2012;**2**(1):5

[5] Hejazi W, Ma J. Gravity, the English language and international business. Multinational Business Review. 2011;**19**(2):152-167

[6] Ku H, Zussman A. Lingua Franca: The role of English in international trade. Journal of Economic Behavior & Organization, 2010;**75**(2):250-260

[7] Fidrmuc J, Fidrmuc J. Foreign languages and trade. (CESifo Working Paper Series No. 4670). 2014. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2411407

[8] Proexport. Proexport, aliado en la internacionalización de Bogotá [Power Point]. 2011, noviembre 29. Available from: http://www.mincit.gov.co/minindustria/loader.php?lServicio=Documentos&lFuncion=verPdf&id=62338&name=Presentacion-Bogota-ZFProexport.pdf&prefijo=file [Accessed: 2017-02-21]

[9] MinComercio, Industria y Turismo. Programa de Transformación Productiva [Internet]. n.d. Available from: https://www.ptp.com.co/contenido/contenido.aspx?catID=607&conID=1 [Accessed: 2017-02-21]

[10] Belkaoui A. Accounting and language. Journal of Accounting Literature. 1989;**8**(1)

[11] McKinsey & Company. Desarrollando sectores de clase mundial en Colombia [Internet]. 2009. Available from: https://www.ptp.com.co/documentos/Plan%20de%20Negocios%20Turismo%20de%20Salud.pdf [Accessed: 2017-02-22]

[12] Abu L. ¿Por qué Colombia está en casi todas las listas internacionales de los mejores destinos para 2017? CNN. [Internet]; 2017, January 27. Available from: http://cnnespanol.cnn.com/2017/01/27/por-que-colombia-esta-en-casi-todas-las-listas-internacionales-de-los-mejores-destinos-para-2017/ [Accessed: 2017-03-18]

[13] MinCit Citur. Estadísticas Nacionales – Tráfico Áreo. [Data statistics]. 2017. Available from: http://www.citur.gov.co/estadisticas/df_lleg_pax_inter/all/48 [Accessed: 2017-04-15]

[14] Bialystok E. Reshaping the mind: The benefits of bilingualism. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale. 2011;**65**(4):229

[15] Lazaruk W. Linguistic, academic, and cognitive benefits of French immersion. Canadian Modern Language Review. 2007;**63**(5):605-627

[16] de Mejía AM. Bilingual education in Colombia: Towards an integrated perspective. International Journal of Bilingual Education and Bilingualism. 2004;**7**(5):381-397. http://doi.org/10.1080/13670050408667821

[17] Valencia S. Literacy practices, texts, and talk around texts: English language teaching developments in Colombia. Colombian Applied Linguistics Journal. 2006;**8**:7-37

[18] Usma JA. Education and language policy in Colombia: Exploring processes of inclusion, exclusion, and stratification in times of global reform. PROFILE Issues in Teachers' Professional Development. 2009;**11**(1):123-141

[19] Ocampo JF. La educación colombiana: historia, realidades y retos. Cooperativa Editorial Magisterio. 2002

[20] Rey de Castro R, García D. Landmark Review of the Use of Teaching and Learning of English in Latin America – Colombia. London: British Council; 1997

[21] MEN. Lineamientos Curriculares para Idiomas Extranjeros. 1999

[22] Law 1651, 2013. Available from: http://wsp.presidencia.gov.co/Normativa/Leyes/Documents/2013/LEY%201651%20DEL%2012%20DE%20JULIO%20DE%202013.pdf [Accessed: 2017-02-21]

[23] Sala de prensa. Barranquilla será la ciudad bilingüe de Colombia. Alcaldía de Barranquilla. 2015, January 30. Available from: http://www.barranquilla.gov.co/sala-prensa/5773-barranquilla-sera-la-ciudad-bilinguee-de-colombia [Accessed: 2017-03-4]

[24] Redacción. Barranquilla es la ciudad con mayor potencial bilingüe. El Heraldo. [Internet]. 2012, October 30. Available from: https://www.elheraldo.co/local/barranquilla-es-la-ciudad-con-mayor-potencial-bilingue-87435 [Accessed: 2017-02-21]

[25] Ministerio de Economía, Fomento y Turismo Chile. Chile: Estrategia Nacional de Inglés 2014-2030. 2012, October 30. Available from: http://www.economia.gob.cl/wp-content/uploads/2014/03/140307-Presentacion-CS-lanzamiento-ENI-v3.pdf [Accessed: 2017-04-28]

[26] Ministerio de Educación Perú. Gobierno emprende nueva política de inglés para que más peruanos accedan a mejores oportunidades [Internet]. 2015. Available from: http://www.minedu.gob.pe/n/noticia.php?id=34374 [Accessed: 2017-02-21]

[27] MEN. Programa Nacional de inglés 2015-2025 [Internet]. 2014. Available from: http://www.colombiaaprende.edu.co/html/micrositios/1752/articles-343287_recurso_1.pdf [Accessed: 2017-02-21]

[28] Colombia Aprende. Colombia Bilingüe: Metas para el 2018 [Internet]. n.d. Available from: http://www.colombiaaprende.edu.co/html/micrositios/1752/w3-article-315516.html [Accessed: 2017-02-21]

[29] MEN. Colombia fortalecerá los planes de enseñanza en inglés en colegios oficiales del país [Internet]. 2016. Available from: http://www.mineducacion.gov.co/1759/w3-article-355910.html [Accessed: 2017-03-03]

Section 3

Psycholinguistic Perspectives on Bilingualism

Chapter 5

A Psycholinguistic Perspective on Development of Phonetic Category Formation in Bilingual Children

Sue Ann S. Lee

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72112

Abstract

This chapter reviews theories and research about phonetic category formation in bilingual children. Investigating phonetic categories provides us with a way to answer one of the longstanding theoretical issues in bilingualism, that is, whether bilingual children possess one versus two linguistic systems in the learning of their respective languages. In this chapter, theoretical backgrounds of phonetic categories in bilingual adults and children are reviewed. Then, empirical evidence showing phonetic categories in bilingual children is summarized. Finally, a development model of phonetic category formation in simultaneous and sequential bilingual children is proposed. Based on the model, detailed phonetic categories do not form across-the-board and bilingual children may invoke multi-dimensional representations of phonetic categories.

Keywords: phonetic category, stop consonants, vowels, acoustics, bilingual children

1. Introduction

Over the last 30 years, the organization of phonetic systems in bilingual speakers has been extensively examined (see [1] for more information). Phonetic category formation refers to the processes by which bilingual or second language leaners come to distinguish phonetic details of share phonemes in each language [2]. Most studies dealt with adult speakers who learn a second language (L2) after they have fully acquired their first language (L1). These studies mainly focused on how the influence of one language on the other depends on the learner's age of exposure to the L2. The current chapter deals with phonetic characteristics of sound produced by bilingual children. Bilingual children are different from adult bilinguals or L2 learners in that their language systems of two languages continue to develop during developmental processes. Thus, it is of interest to examine bilingual children to understand how

phonetic categories develop and are organized across languages and how L1 and L2 systems interact with each other. Two further questions are raised in investigating phonetic categories in bilingual children. First, currently available studies examining phonetic development in bilingual children mainly focused on stop consonant production; limited evidence regarding whether phonetic characteristics of other categories (e.g., vowels) are similar to stops is available. Second, children who are exposed to two languages are either simultaneous or sequential bilingual. Whether phonetic categories of simultaneous bilingual and sequential bilingual children show similar characteristics is not well-examined either. In order to address these questions, first, I outline the currently dominant theoretical models of phonetic category formation in bilinguals. Then, a comprehensive review of existing literature of phonetic categories in bilingual children is provided. Finally, a proposed model of development of phonetic category formation is formulated. Directions for future research on phonetic category formation in bilingual children are also suggested.

2. Theories of bilingual speech acquisition

2.1. Speech learning model for phonetic systems in adult bilingual and L2 speakers

Flege [1] offers the Speech Learning Model (SLM) to account for how individuals learn to produce the vowels and consonants of their L2. The aim of the model is to explain production limitations of experienced L2 learners focusing on more perceptual aspects of learners rather than their motoric constraints. Thus, SLM posits that difficulty to produce a certain L2 phoneme is attributed to a perceptual limitation to discern the sound, and not to production difficulty. A basic assumption in Flege's model is that phonetic elements of the L1 and L2 are related to each other at the level of allophones, and the language-specific aspects of speech sounds are formed in long-term representation called phonetic categories. Since perception plays an important role in the establishment of phonetic categories, if bilingual speakers are able to perceive phonetic differences between L1 and L2, then a new phonetic category can be established for the L2 sound. The likelihood of establishing a new category is further increased by the degree of dissimilarity between an L2 sound and it's closest related L2 sound [1, 3].

Flege further hypothesized that a single phonetic category is used to process similar L1 and L2 sounds due to equivalent classification. If sounds in the L1 and L2 are perceptually linked, then their perceived similarities may block category formation by what Flege refers to as the "mechanism of equivalence classification" [4]. For this process, phonetic category assimilation may occur. Flege and Eefting [5] examined the voice-onset-time (VOT) values of Spanish and English stop consonants as produced by Spanish-English bilinguals. They noted that Spanish-English bilinguals produced stop consonants in English with VOT values resembling those seen in Spanish, suggesting that phonetic category formation was blocked due to similarities in stop consonant production. Flege also predicts two circumstances in which bilingual productions may differ from that of monolinguals: a bilingual's category formation is deflected away from the L1 category in order to maintain phonetic contrast between categories sharing a common L1-L2 phonological space; as a result, representations produced by bilinguals are based on

features different from monolinguals. Bohn and Flege [6] investigated the production of German and English vowels by adult German learners of English. They note that these bilingual speakers produced vowels in such a way that they were able to maintain contrasts within the individual's phonological space. During this process, phonetic category dissimilation may take place.

2.2. Linguistic system models in bilingual children

While SLM was developed to explain adult bilingual and L2 learners, the linguistic system model is a theory regarding language acquisition in bilingual children. The focus of this theory is whether bilingual children develop one or two linguistic systems in the learning of their respective languages. The one-system model, known as the Unitary Language System (ULS), was originally hypothesized by Volterra and Taeschner [7] and the two-system model, also known as the Dual Language System (DLS) hypothesis, was posited by Genesee [8]. Under the ULS model, during early language development, bilingual children would take received input from both languages and combine the information into a single language system. As the language acquisition process continues, bilingual children develop more advanced linguistic skills, and undergo a differentiation process. It is during this process that these children distinguish between languages and achieve bilingual status. The DLS hypothesis stands as an alternative to the ULS hypothesis. The DLS hypothesis posits that children establish two separate linguistic systems from the beginning of the language acquisition process. Under this model, children receive dual language input and separate this information into two distinct language systems. These children do not undergo a period in which their linguistic systems are merged. They have separate linguistics systems from the onset of the acquisition processes; thus children are always considered to be bilinguals under the DLS hypothesis. Since the ULS and DLS hypotheses are significant for understanding bilingual children, more detailed information on each hypothesis is discussed.

2.2.1. Unitary Language System hypothesis

Under the ULS hypothesis, Volterra and Taeschner [7] claim that between infancy and the age of 3, children progress through three stages in order to become bilingual. The first stage of language acquisition in bilingual children shares many similarities with the language development of monolingual children. As children receive language input from both languages, they organize the information into one system. Volterra and Taeschner provide evidence for this by noting a lack of translation equivalents during the early stages of language development. Children receiving dual language input appear to avoid learning words in both languages that share the same meaning. Volterra and Taeschner developed three stages based on a study conducted with two Italian-German bilingual sisters and from data taken from Leopold [9]. Speech samples from these three subjects were taken between 1 year and 2 months (1.2) and 3.9. Their parents indicated that they used the one parent, one language policy and thus only spoke to the children in their native languages. Data obtained during their study seemed to suggest that children do in fact learn translation equivalents between languages. Volterra and Taeschner refuted this idea by suggesting that word meanings have contextual ties which influence the child's use of a word; thus they would not be considered

a translation equivalent. During the second stage of language acquisition, the child is able to differentiate between the lexicons of each language but still continues to apply the same syntactic rules to both languages. Evidence for this stage of language development in bilingual children is seen in the presence of translation equivalents. The child's language now indicates that he or she has words in both languages with equivalent or corresponding meanings. The presence of translation equivalents indicates that the child is able to distinguish lexical items of one language from the other, and sort them by language. Despite the distinction made between lexical items of each language, the incorporation of grammatical components from one language into the other continues to suggest a unified system. During the third stage, language acquisition in the bilingual child is complete. Both the lexical and syntactic linguistic systems are differentiated. Volterra and Taeschner found that the children from the Italian-German study and the Leopold study distinguished and applied the appropriate syntactic rules of each respective language as early as 3.9. It is at this stage that children become bilingual.

2.2.2. Dual Language System hypothesis

Paradis and Genesee [10] argue that bilingual children may acquire separate linguistic systems, and pose an additional option to the DLS by further categorizing these systems into autonomous (no interaction between the two language systems) and interdependent (interaction between the linguistic systems). If these linguistic systems are formed autonomously, then we would expect acquisition of each language by a bilingual to mimic that of a monolingual speaker of each respective language. However, if the two linguistics systems interact during language acquisition, we would expect to see three processes such as transfer, acceleration or delay in each language. Transfer occurs when bilinguals incorporate grammatical elements of one language into the grammar of another language. Acceleration happens when grammatical properties occur earlier in bilinguals than in typical developing monolinguals. Delay is the process in which the burden of simultaneous language acquisition causes delays in the grammatical development of bilinguals when compared to monolinguals.

3. Empirical evidence of phonetic category formation in bilingual children

This section reviews research examining phonetic categories of bilingual children. Investigating phonetic categories in bilingual children started in early 1980s and continued until now, although limited studies have been conducted. This section only reviews studies examining normally developing bilingual children and adolescents focusing on speech production. If any bilingual study examined only one language without addressing the theoretical question (e.g., one vs. two systems or how one language influences the other language), the study is not included in this review [e.g., 11, 12]. After reviewing studies that met the inclusion criteria, the studies are summarized based on the following aspects such as languages, sound category, age of bilingual children, type of bilingual children, etc. (see **Table 1**).

Languages	Ages (years; months)	Types of bilingual	Number of bilingual participants	Types of study	Sound category	Monolingual control group	Two systems
Spanish English							
Konefal and Fokes [22]	4, 7, 10	Sequential	3	Case study	Stop	No	Unknown
Deuchar and Clark [25]	1.7, 1.11, 2.3	Simultaneous	1	Case study	Stop	No, parents' input speech	Yes
Yavas [36]	2nd graders	Sequential	10	Group	Voiceless stop	No	Yes
Fabiano-Smith and Bunta [40]	3	Sequential/ simultaneous	8	Group	Voiceless stop	Yes	No
Muru and Lee [41]	5–6, 10	Sequential	32	Group	Stop	No	Yes
Korean English							
Baker and Trofimovich [35]	10, 16 and adults	Sequential	40	Group	Vowel	Yes	Yes
Lee and Iverson [19]	5, 10	Simultaneous	30	Group	Stop	Yes	Yes
Lee and Iverson [2]	5,10	Simultaneous	40	Group	Vowel	Yes	Yes
Lee [43]	3	Sequential	12	Group	Stop	Yes	Yes
Lee and Iverson [38]	3	Simultaneous	12	Group	Stop/vowel	Yes	Yes/no
Japanese English							
Johnson and Wilson [27]	2.10, 4.8	Simultaneous	2	Case study	Stop	No, parents' input speech	Yes/no
Harada [37]	6, 8, 10	Sequential	15	Group	Voiceless stops	Yes	Yes
French English							
Watson [23]	5, 6, 8, 10	Simultaneous	20	Case study	Stop	No	Yes
Mack [24]	10	Simultaneous	1	Case study	Stop	Yes	Yes
Chinese English							
Yang et al. [45]	3.7–5.3	Sequential	1	Case study	Vowel	No	Unknown
Yang and Fox [46]	5	Sequential	15	Group	Vowel	Yes	Yes
Arabic English							
Khattab [26]	5, 7, 10	Simultaneous	3	Case study	Stop	Yes	Yes/no

Languages	Ages (years; months)	Types of bilingual	Number of bilingual participants	Types of study	Sound category	Monolingual control group	Two systems
German English							
Whitworth [32]	9.11, 12.5	Simultaneous	2	Case study	Stop	No	Yes
German Spanish							
Kehoe et al. [31]	2–3	Simultaneous	4	Case study	Stop	Yes	Yes/no
Dutch English							
Simon [44]	3	Sequential	1	Case study	Stop	No	Unknown

Table 1. Summary of empirical evidence of phonetic category formation.

It is necessary to define bilingual children before each study is discussed. Bilingual children are commonly categorized into simultaneous and sequential [13], but the ages at which each group is categorized vary depending on the researchers. For example, Padilla and Lindholm [14] apply the term bilingual speaker to individuals who have simultaneously acquired two languages, and have generally received an equal amount of exposure and input from each language from birth. Genesee et al. [15] apply the term to individuals who have been exposed to their L2 within the first year of life while McLaughlin [16] and Hamers and Blanc [17] considered simultaneous bilingual children as having acquired the L2 before the L1 is established. Based on the Padilla and Lindholm as well as Genesee et al.'s definitions, certain bilingual children such as Korean-English bilinguals are always categorized as sequential or consecutive because most children in that language started to be consistently exposed to English only when they were enrolled into English-speaking daycare centres, preschools or kindergartens (unless one of the parents is English-speaking) [18]. This may lead to considerable heterogeneity of sequential bilingual children [13]. Lee and Iverson [19] argued that it is necessary to identify when L1 is established in order to determine bilingual status as simultaneous versus sequential. In other words, the identification should be based on a solid developmental milestone rather than an arbitrary age. In this chapter, following Hammers and Blanc, I consider simultaneous bilingual children as those who first learned L1, and then L2, before 5–6 years of age because a child's sound system is not fully developed until 7 years of age [20]. Although a child is exposed to L2 before age 5 or 6, he or she should be exposed to L1 and L2 for a substantial period to become a simultaneous bilingual. If any study tests 3-year-old bilingual children who had been exposed to L2 for less than 1–2 years, these children are considered as sequential bilinguals.

To my knowledge, the earliest studies examining stop production in bilingual children were conducted by Bond et al. [21] and Konefal and Fokes [22]. Two of three children in Konefal and Fokes were also included in Bone et al., when they were young. Thus, only Konefal and Fokes's results are discussed here. Konefal and Fokes examined three female Spanish-English children who were born in a Spanish-speaking country and moved to the US. These children were 4, 7 and 10 years of age. It is not certain about the duration of

http://dx.doi.org/10.5772/intechopen.72112

English language exposure, but these children had been exposed to English for approximately 3 years. Both English and Spanish stops were examined. English and Spanish languages have both voiced and voiceless stops, but the acoustic features (e.g., VOT) are different between the languages. VOT refers to the temporal interval between the release of stop closure and the onset of voicing of a following vowel. English voiced stops are produced with short lag VOT whereas Spanish voiced stops are produced with voicing lead. English voiceless stops are produced with long lag VOT while Spanish voiced stops are produced with short lag VOT. Since the 10-year-old girl had a language disorder, only results of the other two children are discussed here. The authors found that the 4- and 7-year-old children produced English and Spanish voiced stops and voiceless stops differently. The 7-year-old child was able to produce Spanish voiced stops with voicing lead, but the 4-year-old child was not able to. These studies mainly focused on comparing between normal and disordered children without direct comparisons between English and Spanish phonetic categories. It is not certain whether the bilingual children distinguished stops across English and Spanish.

Watson [23] examined stop consonants of 5-, 6-, 8- and 10-year-old French-English bilingual children. Five children for each age group were recruited in this study. Compared to Bone et al. and Konefal and Fokes, all children were well-balanced between English and French. Similar to Spanish, French stops are produced with either voicing lead (voiced stops) or short lag (voiceless stops). Watson found that the 5-year-old children had established voiced and voiceless contrast for only English, but not for French. Three of the five children did not show voiced and voiceless contrast. However, by the age of 6, the bilingual children developed voiced and voiceless distinction for each language. Watson also reported that VOT values decrease as age increases in bilingual children. Due to the small number participants per age group (five children), no statistical analyses were conducted. Although Watson examined both English and French within each bilingual child, he did not systematically compare English with French stop categories. The main interest was whether bilingual children demonstrate voiced and voiceless distinction in each language. Furthermore, bilingual children stop productions were not compared with those of monolingual English- or French-speaking children. Regardless of these limitations, Watson concluded that bilingual children can and do master two separate patterns.

In the 1990s, limited studies were still made to examine bilingual children's production characteristics. Unlike previous studies, however, these studies employed control data from monolingual counterparts or input speech to compare bilingual children's speech. Mack [24] examined stops produced by a 10-year-old French-English bilingual child and a monolingual English- or French-speaking child. Her question was to investigate the extent to which the two languages of a bilingual are interdependent or influence each other. Mack found that the French-English bilingual child produced English voiced stops similarly as compared to the English monolingual child; however, French voiced stops produced by the bilingual child were different from the monolingual French child. The French voiced stops were produced with short lag VOT like English stops, exhibiting transfer from English into French. In terms of voiceless stops, the bilingual child's English voiceless stops were produced with much longer VOT than the monolingual child, but within a normal range. Although the author

did not specify the mechanism for the longer VOT in this child, the longer VOT may be explained as a dissimilation effect to maximize different voiceless stops between English and French. The bilingual child's French voiceless stops were produced with a longer VOT than the French monolingual child; but its VOT values were not within a normal range. Mack claimed that the bilingual child showed some degree of independence between the phonetic systems of his two languages in that the bilingual child demonstrated a distinction between the VOTs of his English and French voiceless stops; but there was also evidence of L2 language influence on L1.

Deuchar and Clark [25] examined a younger bilingual child in order to investigate early acquisition of the voicing contrast in the child's two languages. This child was exposed to both English and Spanish relatively equally from birth by a Spanish-speaking father and an English-Spanish bilingual mother. Deuchar and Clark collected VOT measurements of utterance-initial stops in both English and Spanish productions made at three ages 4 months apart, which corresponded to the following ages: 1.7, 1.11 and 2.3. This study differed from previous research in that it also analysed the data that served as the Spanish and English input for the child, thus allowing for an additional layer of comparative analysis not typically seen in other studies. The authors found a lack of a voicing system in both English and Spanish at age 1.11, the establishment of a clear voicing system in English at age 2.3 but only the beginnings of a similar system in Spanish. The Spanish data did not reflect the caregivers' voicing contrasts but rather progression towards an English-adult speaker voicing contrast. Interestingly, an analysis of the parent's productions in Spanish revealed that the lag measurements were comparable to those of the child at age 2.3. When English and Spanish stops were compared within the child, voiceless stops were significantly different from each other; but voiced stops were not by 2.3. Deuchar and Clark claim that "at least, there is not a single, unified English/Spanish system" ([25], p. 363). The child may acquire English stop pairs earlier than Spanish because of the greater differences in the lag between English voiced and voiceless stops. Although they included speech input as a comparison, age equivalent monolingual children's data are still needed to fully understand bilingual child's phonetic category formation.

In the 2000s, more and more researchers examined phonetic category development in bilingual children. Small case studies were mainly conducted during an initial period; however, a relatively larger number of bilingual children followed. Unlike previous studies, studies during this era examined vowels in addition to stops. Khattab [26] tested three English-Arabic bilingual children (aged 5, 7 and 10) and age equivalent English- or Arabic-speaking children. The three bilingual children were siblings and raised in a city in the UK. Both parents were native Arabic speakers. Arabic was spoken to the children at home, but all three bilingual children were English-dominant. Arabic stops fall into two categories: stops with voicing lead and stops with short lag, similar to Spanish and French. The author found that the 5-year-old bilingual child only distinguished voiceless stops across languages, but she produced similar VOT for Arabic and English voiced stops. Arabic voiced stops were produced with short lag, instead of voicing lead. The other older children had acquired distinct VOT patterns for both voiced and voiceless stops, but the patterns did not always mirror those of their monolingual

http://dx.doi.org/10.5772/intechopen.72112

counterparts. The oldest child failed to produce the Arabic voiced stops with voicing lead VOT, suggesting that an interaction effect of English on Arabic.

Another small scale study examining different language users was conducted by Johnson and Wilson [27]. They examined two Japanese-English bilingual children whose ages were 2.10 and 4.8. They were sisters that lived in a bilingual family in Japan. When the children were 2.11 and 1.1, they moved into Canada. Both children had been exposed to a relatively equal amount of English and Japanese at home based on the one parent and one language principle. Both Japanese and English stops were examined using VOT. Japanese stops are similar to Spanish, French and Arabic in that voiced stops are produced with voicing lead whereas voiceless stops are produced with short lag VOT. Similar to Deuchar and Clark [25], parents' input speech was collected for comparison as well as VOT values from existing literature ([28], for English) and Homma ([29, 30] for Japanese). The authors found that both children differentiated voiced and voiceless stops for each language. English voiced stops were produced with short lag whereas English voiceless stops were produced with long lag. None of the bilingual children produced Japanese voiced stops with voicing lead. Both bilingual children produced Japanese voiceless stops with long lag, which may be an influence from English. In short, the younger child produced similar English and Japanese stops for either voiced or voiceless; the older child produced English voiceless stops with longer VOT than Japanese voiceless stops. Although the authors did not specify the underlying mechanism for the longer VOT, it may be considered as a dissimilation process to maximize English and Japanese voiceless stop categories.

Kehoe et al. [31] examined another language group of bilingual children, that is, four Spanish-German bilingual children aged 2.0–3.0. Voicing contrast and VOT values between German and Spanish are similar to those of English and Spanish. The bilingual children's VOT production was compared to three German children and to previous literature findings in Spanish. They found three patterns of VOT development. First, two bilingual children showed delay in the phonetic realization of voicing. These children did not acquire German voicing contrasts; Second, one child showed a transfer effect that he produced German voiced stops with voicing lead (Spanish-like) whereas he produced Spanish voiceless stops with long lag VOT (German-like); third, one child did not demonstrate any cross-language influence. By age 3, none of the German-Spanish bilingual children acquired Spanish voiced stops. In terms of cross-languages, two children distinguished German and Spanish voiceless stops; however, the other two children did not make such distinctions.

While previous studies mainly focused on stop productions, limited studies started investigating vowel production in bilingual children. Whitworth [32] examined vowel length and VOT acquisition in two German-English bilingual children, aged 9.11 and 12.5. Both children were exposed to both languages from birth based on the one parent, one language approach. The mother only spoke German whereas the father only spoke English to the children. English was the language used while the children attended schools and communicated with their friends. Thus, these children were English-dominant. The 9-year-old child possessed an English accent when he spoke German while the 12-year-old's German is

native-like with a northern standard German accent. German and English are produced with short lag VOT for voiced and long lag VOT for voiceless stops with a small difference in VOT values within each category. The author found that the younger child distinguished German and English voiceless stops, but not voiced stops, whereas the older child differentiated both voiced and voiceless stops across two languages. The author argued that the results seem to support two linguistic hypotheses. However, the VOT patterns these children showed were different from English and German. For example, although the younger child distinguished English and German voiceless stops, English VOT was shorter than German VOT, which shows the reverse pattern. Similarly, the older child produced longer VOT for German voiced stops than for the English voiced stop, which also appears to be a reverse pattern. One of the major criticisms of this VOT study was that the author did not control the place of articulation. It is well known that VOT for velar stops are produced with longest VOT for both voiced and voiceless stops. However, the author did not provide any information on how many tokens were included regarding place of articulation. In addition to VOT, the length of tense and lax vowels was also examined in Whitworth's study. According to the author, German lax vowels are approximately half as long of German tense vowel [33]. English tense vowels are one-third longer than English lax vowels [34]. Both children produced English tense vowels significantly longer than German tense vowels; however, they did not differentiate English and German lax vowels. Although the author claimed that the younger child did distinguish them, the difference was marginal ($p < .06$).

So far, most studies were limited to case studies with a small number of participants involved. A larger group study was conducted by Baker and Trofimovich [35] to investigate how the phonetic vowel representation would be similar or different between long and short exposure duration for each age group. In this study, Baker and Trofimovich included four groups of Korean-English bilingual speakers. All participants were born in Korea and moved to the US at various ages. Two groups were adults with either shorter (M = 0.9 year) or longer (M = 6.9 years) exposure duration to English. The other two groups were older children. One of the children's group was aged 10.2 years with 1.3 years of exposure duration; the other group was aged 16.9 years with 8 years of exposure duration to English. The authors found that the earlier the exposure to two languages, the more likely a bilingual will produce distinct acoustic realization of L1 and L2 sounds. For example, bilingual children with longer exposure duration distinguished English /ɪ/ from /i/, /æ/ from /ɛ/ and /u/ from /ʊ/ better than children with shorter exposure duration. They also found an L2 transfer effect on L1 in that the Korean /u/ was centralized in bilingual children with longer exposure. Age equivalent monolingual children were also recruited to compare with bilingual children with early exposure. They found that these bilingual children produced English /i/, /u/ and /ɛ/ similarly as monolingual children but they differed from monolingual children in their production of English /ɪ/, /ʊ/ and /æ/. They produced these vowels with higher F1 vowels than monolingual children. The authors suggested that the bilingual children with longer exposure duration demonstrated some evidence of L1 vowel influence on L2 vowels. Baker and Trofimovich conducted the first well-designed group study to provide important findings for how bilinguals organized their phonetic systems and the complex interactions between L1 and L2. However, these

http://dx.doi.org/10.5772/intechopen.72112

bilingual children started to be exposed to English at a later age, it is not certain whether language influence patterns appear in young bilingual children who acquire both languages at a young age.

Yavas [36] conducted a study where he examined older aged Spanish-English sequential bilingual children (10 2nd graders). These children were monolingual Spanish-speaking until age 5 in Florida, US; then started learning English in kindergarten and had been exposed to English 2–3 years. Unlike previous studies, Yavas used mixed sentences to elicit Spanish and English stops. For example, "Pon el papel on the table". Only voiceless stops were elicited in both languages. Yavas did not conduct any statistical analysis; only a qualitative description for each individual child was addressed. The author reported that Spanish-English consecutive bilingual children's data supported that heterogeneity of bilinguals. One bilingual child's stop production was similar to monolinguals; this child manifested a totally separate system for English or Spanish. Four of the bilingual children showed a separate system for both languages with variations. For instance, one child produced bilabial English stops with shorter VOT, but with an acceptable range and the other child differentiated one place, but not the other places. Only one child did not differentiate two systems at all. Yavas concluded that the bilingual children showed unique and specific linguistic patterns. Yavas collected Spanish stops from mixed sentences while English stops with English only sentences. It is not certain whether such method leads to accurate production results. Also, Yavas examined only voiceless stops for older age children. It would be more useful if both voiced and voiceless stops were examined. In fact, whether voiced Spanish stops are influenced by English would be of interest.

Harada [37] examined VOT produced by 15 English-Japanese bilingual children in a Japanese immersion program in the US. The bilingual children were from grade 1 (age 6), grade 3 (age 8) or grade 5 (age 10). The children's primary language is English, but they started to learn Japanese after enrolling in the immersion program. Thus, these children are categorized as sequential or consecutive bilinguals. This study also included 5 English-Japanese bilingual adults, 10 monolingual Japanese children, 5 monolingual Japanese adults and 5 monolingual English adults. However, no monolingual English-speaking children were included. Also, five English-Japanese bilingual teachers in the immersion program participated. Only English and Japanese voiceless stops were examined. Harada found that the bilingual children produced Japanese voiceless stops with significantly longer VOT values than the monolingual Japanese children and the immersion teachers. Within the comparison, the bilingual children's Japanese stops were produced with significantly shorter VOT than English voiceless stops. These results indicated that the bilingual children make a phonetic distinction between Japanese and English although their VOT values are different from monolinguals.

In the 2010s, more comprehensive studies examining phonetic category formation have been conducted. Each study employed a relatively large number of children, and compared bilingual children's speech with that of monolingual counterparts. Also, recent studies examined a variety of bilingual language groups such as Korean-English, Chinese-English or Dutch-English bilingual children. In addition, these studies made attempts to evaluate SLM in bilingual children.

Lee and Iverson [2, 19, 38] conducted a series of studies examining phonetic category formation in Korean-English bilingual children. First, Lee and Iverson [19] examined the phonetic representation of Korean and English stops produced by 5- and 10-year-old Korean-English bilingual children. The bilingual children's stop productions were compared to age equivalent English- and Korean-speaking children. They had two research questions. First, when do Korean-English bilingual children establish fully independent phonetic systems for each language? Second, what kind of mechanisms (assimilation or dissimilation) do bilingual children employ in their development process? Each age or language group was compared of 15 children; a total of 90 children participated in this study. Investigating Korean-English bilingual children was of interest because Korean stops show a three-way laryngeal contrast and are distinguished by vowel-onset fundamental frequency (hereafter fo) in addition to VOT [39]. Unlike bilingual children whose languages have only voiced and voiceless distinctions for stop category, Korean-English bilingual children may have difficulty in differentiating phonetic categories of stops due to its complexities.

Lee and Iverson [19] reported that Korean-English bilingual children were able to make phoneme distinctions within each language. Both age groups of bilingual children clearly produced all English and Korean stops. When the authors compared English and Korean stops produced by 10-year-old Korean-English bilingual children, it was found that all possible comparisons were significantly different in terms of either VOT or fo values, indicating that 10-year-old Korean-English bilingual children established fully distinctive stop categories across two languages as monolingual English- or Korean-speaking children did. However, 5-year-old bilingual children did not distinguish stop categories across languages when they fall in the same VOT regions although these stop pairs were fully distinctive in monolingual children. For example, English voiced and Korean fortis stops are produced with short lag VOT. When compared, Korean fortis were produced with significantly higher fo values than English voiced stops. Similarly, English voiceless and Korean lenis and aspirated stops are produced with long lag VOT. Korean lenis stops are produced with lower fo than English voiceless whereas Korean aspirated stops are produced with longer VOT than English voiceless stops. These stop pairs were significantly different between the two 5-year-old monolingual groups, but not by 5-year-old children.

When the stop production was compared between bilingual and monolingual children, it was found that 10-year-old bilingual children showed longer VOT for Korean lenis and aspirated stops than monolingual Korean children. The bilingual children produced shorter VOTs for English stops than monolingual English-speaking children. The bilingual children also showed different fo values than monolingual children. They produced lower fo for Korean aspirated stops. These results were interpreted that Korean-English bilingual children employed both assimilation and dissimilation depending on age. Dissimilation took place by producing VOT longer than monolingual children in order to maximally distinguish all stops within a long lag region. Although a merged category was not found, 10-year-old bilingual children produced lower fo, indicating that lower fo in English may influence their fo for Korean stops.

Fabiano-Smith and Bunta [40] examined Spanish and English voiceless stops produced by eight 3–4-year-old Spanish-English bilingual children. Some bilingual children had recently arrived in the US, while the parents of other bilingual children had grown up in an English-speaking community. Regardless, the bilingual children attended a bilingual preschool where both languages were used and the language of the classroom alternated each day. Thus, both simultaneous bilingual and child L2 learners were included in this study. The bilingual speech was compared to eight monolingual Spanish or eight English-speaking children. Only bilabial and velar voiceless stops were examined. The authors found that although English or Spanish VOT values were significantly different between monolingual English- and monolingual Spanish-speaking children, these values were not significantly different in bilingual children. In terms of between group comparisons, English VOT values produced by bilingual children were significantly different from monolingual English-speaking children; however, their VOT values of Spanish were not different from monolingual Spanish-speaking children. The authors suggested that the results of this study provide evidence to support Flege's claim, that is, equivalent classification that L1 may trigger assimilation of the L2 segmental category. This study provided important information that monolingual English- and Spanish-speaking children did distinguish voiceless stops across languages. However, since the authors did not test whether monolingual children distinguish English and Spanish voiced stops, it would be more comprehensive if they tested voiced stops in their study.

Muru and Lee [41] examined VOT produced by 5–6-year-old and 10-year-old Spanish-English bilingual children. These children were raised in a Spanish-speaking home and started to learn English at English-speaking daycare centres. Thus, these children were categorized as sequential bilingual children. The authors did not include monolingual counterparts. Thus, only Spanish and English VOT values produced by bilingual children were compared. The authors found that the 5–6-year-old Spanish-English bilingual children only made a distinction between English and Spanish for voiceless stops, but not for voiced stops. On the other hand, 10-year-old Spanish-English bilingual children were able to distinguish both voiced and voiceless stops across English and Spanish. One exception was that no significant difference was found between English voiced and Spanish voiced for velar place of articulation. This study was a good extension of Fabiano-Smith and Bunta's study in that older aged Spanish-English bilingual children were examined. It seems that phonetic category formation is not established between English and Spanish at 3 years of age; however, their phonetic representation develops as they grow older and distinctive phonetic categories for voiceless first evolve at 5 years of age. Finally, phonetic category formation for English and Spanish stops is established at 10 years of age. These results were similar to Korean-English bilingual children, confirming that phonetic category formation is fully established by 10 years of age, but not 5 years of age.

In another study, Lee and Iverson [2] examined English and Korean vowels produced by 5- and 10-year-old Korean-English bilingual children. In their previous study [19], Lee and Iverson found that phonetic category for stops were established in 10-year-old Korean-English bilingual children, but not in 5-year-old children. The goal of this study was to determine when

phonetic category formation takes place for vowel production. Is it similar to stop production? The same cohort of Korean-English bilingual children in Lee and Iverson's stop study participated in this vowel study. Thus, all children characteristics were the same. Unlike Baker and Trofimovich's study that involved Korean-English bilingual children who learned English after they fully acquired Korean, the Korean-English bilingual children in Lee and Iverson's study had been exposed to both English and Korean for at least 2 years (5-year-olds) and 5 years (10-year-olds). First and second formant frequencies (F1 and F2) were measured. When bilingual and monolingual children were compared, English vowels were similar between the two groups except for a few vowels whereas F2 values of Korean vowels /u/ and /o/ were significantly higher in bilingual children than in monolingual Korean children, indicating English language influence on Korean vowels. When English and Korean vowels were compared within bilingual children, these vowels were grouped into four groups: high-front /i, ɪ, e/; non high-front /ɛ, æ/; high-back /u, ʊ, o/ and non high-back /ʌ, ɑ, ɔ/. The results showed that F1 and F2 values for high-front vowels were distinguished based on F2 values except for Korean /i/ and English /i/. In terms of non high-front, English /ɛ/ and Korean /ɛ/ were similar to each other; but they were different from English /æ/. All high-back vowels were produced fully distinctively; none of the F1 and F2 values overlapped each other. F1 or F2 values of non high-back vowels were also significantly different except for English /ɔ/ and Korean /ʌ/. The authors claimed that detailed phonetic categories across languages are not formed holistically in an across-the-board fashion. In other words, vowel acquisition is typically earlier than stop acquisition in monolingual children. Phonetic category formation also takes place in vowels earlier than stops in bilingual children. The authors also found little evidence regarding assimilation and dissimilation. The higher F2 of Korean /u/ was interpreted as evidence of assimilation that the centralized English /u/ influences the Korean /u/. This finding parallels the findings of Flege [42] with higher F2 than is characteristic of native French. The authors also found evidence of dissimilation in that 10-year-old Korean-English bilingual children produced the vowel /æ/ with higher F1 than found among monolingual English-speaking children. Korean-English bilingual children may exaggeratedly lower the tongue in the production of /æ/ to maximally distinguish it from the vowel /ɛ/.

Recently, Lee and Iverson [38] examined when phonetic categories of stops emerge in 3-year-old Korean-English bilingual children and whether phonetic category formation takes place similarly between two different sound categories. The bilingual children were exposed to both Korean and English languages from birth to 18 months. The authors examined both English and Korean stops as well as front vowels produced by 12 bilingual, 15 monolingual Korean and 15 monolingual English-speaking children. VOT and fo values of English and Korean stops and F1 and F2 values for English and Korean vowels were measured. The study found that monolingual and bilingual children produced English or Korean vowel phonemes distinctively. When English and Korean were compared, both monolingual and bilingual children did not distinguish any stop categories within the same VOT region; neither VOT nor fo was different across English and Korean stops. However, the bilingual and monolingual children produced stops differently in that the bilingual children produced higher fo values for English voiceless stops. While stops were not produced distinctively by both monolingual and bilingual children, both groups produced English and Korean vowels significantly

http://dx.doi.org/10.5772/intechopen.72112

differently for Korean /i/ and English /ɪ/ pairs as well as Korean /ɛ/ and English /æ/ pairs. When English vowels were compared between monolingual and bilingual children, no group differences were found in either language, indicating that Korean and English vowels produced by the 3-year-old bilingual children were similar to monolingual children. The authors concluded that phonetic categories in 3-year-old children develop without much interaction between the two languages in simultaneous bilingual children exposed to two languages at an early age.

Lee [38] further examined VOT values produced by 3-year-old sequential Korean-English bilingual children. These children had been exposed to both languages for only 6–8 months. They had very limited English language abilities when the study was conducted. The author found that these bilingual children showed some evidence to distinguish English and Korean stops in that English voiced and Korean fortis stops were produced differently. Korean fortis stops were produced with higher fo than English voiced stops. However, the Korean lenis and aspirated and English stops were not significantly different from each other. Since these children fully acquired Korean stops when they were exposed to English, the phonetic distinction between Korean fortis and English voiced stops may be salient to these children. Although the sequential Korean-English bilingual children distinguished English voiced and Korean fortis stops, the other consonants were not distinguished from each other, suggesting that these children did not fully acquire phonetic category formation in stop production. Since the author did not compare the bilingual child with monolingual counterparts, it is not certain whether sequential bilingual children's stop production is similar or different from monolingual children.

Another study examining a 3-year-old sequential bilingual child was done by Simon [44] and Yang [45]. Simon reported a longitudinal case study examining the acquisition of English and Dutch stops. Dutch voiced and voiceless stops are produced with voicing lead and short lag, respectively that is similar to Spanish. The first recording was made 3 months after his exposure to English until 4.0 in 11 sessions. The author found the bilingual child successfully mastered the English contrast within a 7-month period, but the child's L1 system showed changes. The percentage of Dutch voiced stops produced with voicing lead decreased 30% at the end of session, suggesting the influence from L2 on L1. Yang longitudinally examined a Chinese-English bilingual child's vowel production for a 20-month period. Recording began when the child started to attend an English language preschool at age 3.7. Approximately, one recording session was made each month until 5.2. The author found three phases of vowel development. During the initial phase, several broad L1 categories are clustered near the three L1 corner vowels (/i, u, a/). Then, the child began to contrast among individual vowels in L2 with great production variation. Finally, the child's vowel system was stabilized and reduced within category variation. Acoustic vowel space of English and Chinese was compared during the period. While Chinese vowel space was relatively stable, the child's English vowel space showed substantial changes in both size and shape. Because these two studies did not compare stop or vowel segments between the two languages, it is not certain whether these children showed distinctive phonetic categories across languages.

Yang and Fox [46] further examined Chinese and English vowels produced by 5–6-year-old Chinese-English bilingual children as a group. Fifteen bilingual children participated; the children were divided into two groups depending on their English language proficiency. The authors found that although no significant difference of vowel formant frequencies among three groups (monolingual English and two bilingual children), bilingual children with low English proficiency showed greater variation and slight positional changes. Furthermore, the bilingual children with high English proficiency showed better separation among the vowel categories, similar to that of the monolingual English-speaking children whereas the bilingual children with low English proficiency showed great overlaps for most vowel pairs than the other groups. In addition, shared vowels of English and Chinese were compared. The authors reported that no significant difference was found for English and Chinese /i/ by monolinguals, but the other shared vowels were fully separated from each other. Two groups of bilingual children showed similar production patterns. The authors concluded that L2 vowel systems in the bilingual children with low English proficiency were strongly influenced by their L1. The bilingual children produced L2 vowels in a near-native manner, but some L2 features were transferred to L1 vowels, suggesting an assimilation process taking place during L1 acquisition. **Table 1** shows a summary of empirical evidence of phonetic category formation.

4. Developmental model of phonetic category formation in bilingual children

Based on the findings of previous work on phonetic category formation, I propose a model called "development model of phonetic category formation" in both simultaneous and sequential bilingual children. In this model, I argue that phonetic category formation continues to evolve during the developmental process rather than emerge all at once in both types of bilingual children. **Figures 1** or **2** shows a schematic representation of phonetic category development in terms of stop and vowel categories for either simultaneous or sequential bilingual children. The direction of arrows shows the language transfer effect. As can be seen in **Figure 1**, in simultaneous bilingual children at 3–4-years of age, phonetic categories for L1 and L2 stops are not distinguished at all regardless of language types. Thus, two circles representing L1 and L2 overlap each other. The size of the circle denotes the development of a stop system in each language. Whether the stop system of each language is fully developed or not depends on the sound system of each language. For example, 3-year-old simultaneous Korean-English bilingual children were able to produce both English and Korean stop phonemes distinctively within a language [38] whereas Spanish-English [25], Japanese-English [27] or Spanish-German [31] bilingual children were not able to produce Spanish or Japanese voiced stops which fall in voicing lead category. This finding was similar to previous research reporting that monolingual children have difficulty to acquire voicing lead stops [47]. The language influence effect also varied depending on languages. While Korean-English bilingual children did not show much interaction effects, bidirectional interaction [31] or unidirectional an influence of L2 on L1 [27]. At 5–6-years of age, phonetic categories for stops across language remains constant. Lee

http://dx.doi.org/10.5772/intechopen.72112

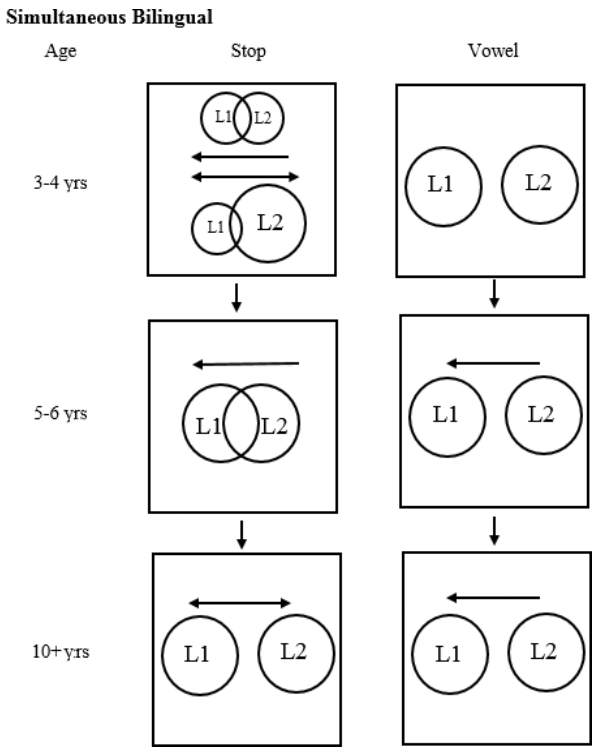

Figure 1. Developmental model of phonetic category formation in simultaneous bilingual children. *Note*: The size of the circle indicates the phoneme development within each language. Arrow direction denotes the influence of one language on the other. Both unidirectional and bidirectional influences are indicated when both claims are reported.

and Iverson [19] reported that Korean-English bilingual children did not distinguish English and Korean stops across languages; neither Khattab [26] nor Watson [23] reported that bilingual children distinguished stop categories at five years of age. These children still failed to produce voiced stops with lead voicing if any language has voicing lead stops. An L2 influence on L1 still exists at this age [19]. Phonetic category for stops, however, is fully established at age 10 or older in simultaneous bilingual children. It was also noted that interaction effects between L1 and L2 take place at these ages. The interaction direction may be unidirectional in that L2 influences L1 [19, 24, 26] or bidirectional [32]. It is not certain why Whitworth found a bidirectional influence with these children. Further studies need to verify this aspect.

While phonetic category for stops is not fully established until 10 years of age, that of vowels seems developed earlier than stops. At 3–4-years of age, simultaneous bilingual children produced vowels of both languages distinctively with limited interaction effects. The fully separated vowel systems remain the same at 5- and 10-years of age; however, the L2 language started to have an influence on the L1 vowel system at 5-years of age. The developmental model of vowel category formation heavily relies on Korean-English bilingual children [2, 38]. Since limited evidence is available on vowel production produced by simultaneous bilingual children, further studies are warranted to verify this observation.

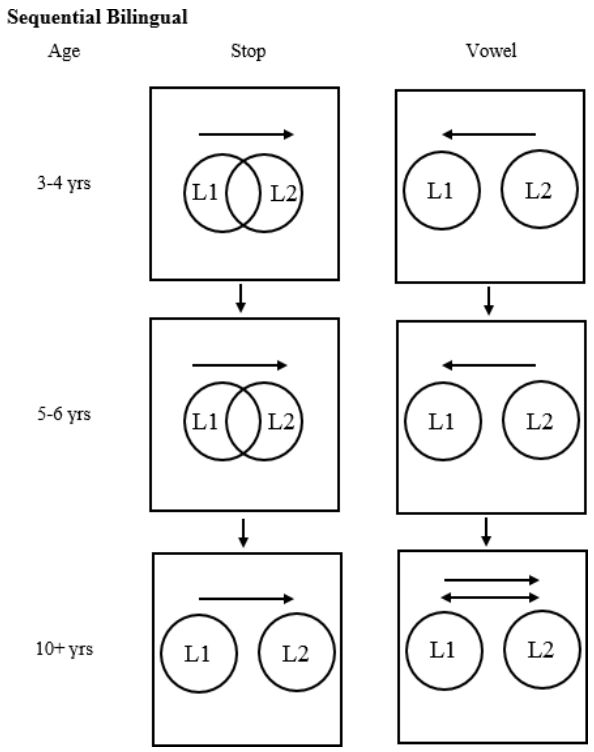

Figure 2. Developmental model of phonetic category formation in sequential bilingual children. *Note*: The size of the circle indicates the phoneme development within each language. Arrow direction denotes the influence of one language on the other. Both unidirectional and bidirectional influences are indicated when both claims are reported.

It seems that phonetic category formation in sequential bilingual children develops similarly with simultaneous bilingual children; but some differences are also observed. At 3–4-years of age, sequential bilingual children did not manifest fully distinctive phonetic category for stops, similar to simultaneous bilingual children. While no transfer effect was observed in simultaneous bilingual children, a language transfer effect appears in that there was a strong effect of L1 on L2 language in sequential bilingual children. Similar to simultaneous bilingual children, sequential bilingual children did not manifest distinctive phonetic categories for stops at 5–6-years of age. Although voiceless stops were distinguished from each other, voiced stops across languages remains undistinguished by this age. Similar to 3–4-years of age, a unidirectional L1 influence on L2 exists during this age [40, 44]. Phonetic category formation; however, is fully acquired at 10 years of age or older in sequential bilingual children. There was also L1 influence on L2 during this age [37]. It is interesting to observe that L1 influence on L2 on stops in sequential bilingual children because L2 typically influences L1 in simultaneous bilingual children. It is not certain why this happens. It may be due to the fact that sequential bilingual children fully develop a stop system of their L1; thus, it may affect stops of L2, which is not fully developed yet.

http://dx.doi.org/10.5772/intechopen.72112

Vowel category formation in sequential bilingual children also showed a similar pattern as compared to simultaneous bilingual children. At 3–4-years of age, a sequential bilingual child showed separation of two vowel systems after short exposure duration to L2, suggesting that this child tends to distinguish two systems although there was an influence of L2 on L1 [45]. However, this finding was based on a single bilingual child without direct comparisons between the two languages. Further studies are warranted to confirm their findings. At 5–6-years of age, sequential bilingual children continue to manifest two systems. L1 production is also influenced by L2 at this age [46]. The distinctive vowel categories remain in separation at 10 years of age. Unlike younger aged sequential bilingual children, 10-year-old sequential bilingual children showed either bidirectional influence for children with longer exposure duration or L1 influence on L2 for shorter exposure duration [35]. L1 influence on L2' vowels were not observed in research with bilingual children, but the effect is commonly found in adult L2 learners. These differences may suggest that phonetic category formation and the effect of interaction between L1 and L2 may be different between child and adult bilingual speakers. In short, phonetic category formation in bilingual children is established progressively using multi-dimensional representations for each sound category, and continues to evolve in the developmental process. Interaction between L1 and L2 varied depending on types of bilingualism.

5. Limitations and directions for future research

The developmental model proposed in this chapter is based on current empirical evidence. Some research studies are a single case study without employing rigorous statistical analysis. Thus, this model should continue to develop based on more empirical findings in the future. Future studies should consider following aspects when phonetic category formation is examined in bilingual children. First, more group studies are expected in the future. Among 20 studies examining phonetic category formation in bilingual children, only half of the studies employed group comparisons. In order to lead to a more solid theoretical model of phonetic category formation, findings should be based on group studies. Second, when studies examine phonetic category formation in bilingual children, it is necessary to employ monolingual control groups of each language. Without understanding the phonetic development of monolingual children, it is not certain whether such a pattern shown in bilingual children is a natural developmental consequence or a bilingual effect. For example, several studies reported that bilingual children whose stops are produced with voicing lead often produced voiced stops with short lag VOT instead of voicing lead. It is not certain whether such production is attributed to the fact that these children acquire two languages or one language influences the other. Third, compared to studies examining stops, vowel studies are relatively limited. Only vowels produced by Korean-English and Chinese-English are currently available. In addition, no fricative or other consonantal study has been conducted. Thus, future studies are warranted to examine vowels and other consonants in simultaneous or sequential bilingual children. Fourth, although recent studies examined more diverse

bilingual languages, still limited bilingual languages have been studied. Some bilingual languages are similar in that stops are categorized as either voiced or voiceless. Only Koreans, whose stop systems are different from other languages, were examined. Future studies may examine more simple or complex stop or vowel systems in order to fully understand how bilingual children manifest distinctive phonetic categories when they are in different language systems.

Author details

Sue Ann S. Lee

Address all correspondence to: sueann.lee@ttuhsc.edu

Texas Tech University Health Sciences Center, Lubbock, TX, USA

References

[1] Flege JE. Second-language speech learning: Theory, findings, and problems. In: Strange W, editor. Speech Perception and Linguistic Experience: Issues in Cross-Linguistic Research. Timonium: York Press; 1995. pp. 233-272

[2] Lee SAS, Iverson GK. Vowel category formation in Korean-English bilingual children. Journal of Speech, Language, and Hearing Research. 2012;**55**:1449-1462

[3] Yeni-Komshian GH, Flege JE, Liu S. Pronunciation proficiency in the first and second languages of Korean-English bilinguals. Bilingualism: Language and Cognition. 2000; **3**:131-149

[4] Flege JE. English vowel productions by Dutch talkers: More evidence for the "similar" vs. "new" distinction. In: Second Language Speech. Structure and Process. Berlin: Mouton de Gruyter; 1996. pp. 11-52

[5] Flege JE, Eefting W. The production and perception of English stops by Spanish speakers of English. Journal of Phonetics. 1987;**15**:67-83

[6] Bohn O, Flege JE. The production of new and similar vowels by adult German learners of English. Study in Second Language Acquisition. 1992;**14**:131-158

[7] Volterra V, Taeschner T. The acquisition and the development of language by bilingual children. Journal of Child Language. 1978;**5**:331-326

[8] Genessee F. Early bilingual development: One language or two? Journal of Child Language. 1989;**16**:161-179

[9] Leopold WF. Speech Development of a Bilingual Child: A Linguist's Record. Vol. 2. NY: AMS Press; 1970

[10] Paradis J, Genessee F. Syntactic acquisition in bilingual children: Autonomous or interdependent? Studies in Second Language Acquisition. 1996;**18**:1-25

[11] Tsukada K, Birdsong D, Bialystok E, Mack M, Sung H, Flege J. A developmental study of English vowel production and perception by native Korean adults and children. Journal of Phonetics. 2005;**33**:263-290

[12] Oh GE, Guion-Anderson S, Aoyama K, Flege JE, Akahane-Yamada R, Yamada T. A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. Journal of Phonetics. 2011;**13**:156-167

[13] Goldstein B. Bilingual Language Development & Disorders in Spanish-English Speakers. Baltimore: Brookes Publishing Co.; 2004

[14] Padilla A, Lindholm K. Child bilingualism: The same old issues revisited. In: Martinez J, Mendoza R, editors. Chicano Psychology. Orlando, FL: Academic Press; 1984. pp. 369-408

[15] Genesee F, Paradis J, Crago M. Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning. Baltimore, MD: Brookes; 2004

[16] McLaughlin B. Second-Language Acquisition in Childhood. Mahwah, NJ: Lawrence Earlbaum Associates; 1978

[17] Hamers JF, Blanc MHA. Bilinguality and Bilingualism. Cambridge: Cambridge University Press; 2000

[18] Shin SJ. Developing in Two Languages: Korean Children in America. Tonawanda NY: Multilingual Matters Ltd.; 2005

[19] Lee SAS, Iverson GK. Stop consonant productions of Korean-English bilingual children. Bilingualism: Language and Cognition. 2011;**15**(2):275-287

[20] Banthal J, Bankson N, Flipsen P. Articulation and Phonological Disorders: Speech Sound Disorders in Children. 6th ed. Engle wood Cliffs, NJ: Pretice-Hall; 2009

[21] Bond ZS, Eddey JE, Bermejo JJ. VOT del Español to English: Comparison of a language-disordered and normal child. Journal of Phonetics. 1980;**8**:287-291

[22] Konefal JA, Fokes J. Voice onset time: The development of Spanish/English distinction in normal and language disordered children. Journal of Phonetics. 1981;**9**:437-444

[23] Watson I. The production of VOT contrasts in English-French bilingual children. Cambridge Papers in Phonetics and Experimental. Linguistics. 1982;**1**:1-17

[24] Mack M. Phonetic transfer in a French-English bilingual child. In: Nelde PH, editor. Language Attitude and Language Conflict. Dümmler: Bonn Germany; 1990

[25] Deuchar M, Clark A. Early bilingual acquisition of the voicing contrast in English and Spanish. Journal of Phonetics. 1996;**24**:351-365

[26] Khattab G. VOT production in English and Arabic bilingual and monolingual children. Leeds Working Papers in Linguistics and Phonetics. 2000;**8**:95-122

[27] Johnson CE, Wilson IL. Phonetic evidence for early language differentiation: Research issues and some preliminary data. International Journal of Bilingualism. 2002;**6**:271-289

[28] Lisker L, Abramson A. A cross language study of voicing in initial stops. Acoustic Measurements. 1964;**20**:384-422

[29] Homma Y. Voice onset time in Japanese stops. Onsei Gakkai Kaiho. 1980;**163**:7-9

[30] Homma Y. Durational relationship between Japanese stops and vowels. Journal of Phonetics. 1981;**9**:273-281

[31] Kehoe MM, Lleo C, Rakow M. Voice onset time in bilingual German-Spanish children. Bilingualism: Language and Cognition. 2004;**7**:71-88

[32] Whitworth N. Acquisition of VOT, and vowel length by English-German bilinguals: A pilot study. Leeds Working Papers in Linguistics and Phonetics. 2000;**8**:229-235

[33] Antoniadis Z, Strube HW. Untersuchungen zur spezifischen Daur deutscher Vokale. Phonetica. 1984;**41**:72-87

[34] House AS. On vowel duration in English. Journal of the Acoustical Society of America. 1961;**33**:1174-1178

[35] Baker W, Trofimovich P. Interaction of native- and second-language vowel system(s) in early and late bilinguals. Language and Speech. 2005;**48**:1-27

[36] Yavas M. Voice Onset Time Patterns in Bilingual Phonological Development. Themes in Clinical Linguistics. Mahwah, NJ: Lawrence Erlbaum Associates; 2002. pp. 341-350

[37] Harada T. The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. International Review of Applied Linguistics in Language Teaching. 2007;**45**(4):353-378

[38] Lee SAS, Iverson G. The emergence of phonetic categories in Korean-English bilingual children. Journal of Child Language. 2017;**44**:1485-1515

[39] Cho T, Jun S, Ladefoged P. Acoustic and aerodynamic correlates of Korean stops and fricatives. Journal of Phonetics. 2002;**30**:193-228

[40] Fabiano-Smith L, Bunta F. Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. Clinical Linguistics & Phonetics. 2012;**26**(2):148-163

[41] Muru AM, Lee SAS. Development of phonetic categories of stop consonants in Spanish-English bilingual children. Clinical Archives of Communication Disorders. 2017;**2**(1): 60-68

[42] Flege JE. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. Journal of Phonetics. 1987;**15**:47-65

[43] Lee SA. Phonetic categories of English and Korean stops in 3-year-old sequential Korean-English bilingual children. Clinical Archives of Communication Disorders. 2016;**1**(1):3-10

[44] Simon E. Child L2 development: A longitudinal case study on Voice Onset Times in word-initial stops. Journal of Child Language. 2010;**37**:159-173

[45] Yang J, Fox RA, Jacewicz E. Vowel development in an emergent Mandarin-English bilingual child: A longitudinal study. Journal of Child Language. 2015;**42**:1125-1145

[46] Yang J, Fox RA. L1-L2 interactions of vowel systems in bilingual Mandarin-English children. Journal of Phonetics. 2017;**65**:60-76

[47] Macken MA, Barton D. The acquisition of native voice onset time in English-learning children. Journal of the Acoustical Society of America. 1980;**1241**:1180-1191

Chapter 6

Bilingualism and Self-Perception: Self-Efficacy through the Veil of Two Languages

Maura A. E. Pilotti and Khadija El Alaoui

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72740

Abstract

This chapter is concerned with the relationship between language, as the vehicle of a person's culture, and self-assessment of one's capabilities (i.e., self-efficacy) via conventional self-report measures. It relies on the assumption that a language "is 'a veil' over the reality of the culture in which it is used, involving an agreement of its users about what there is to be seen and how it should be seen". Thus, the information weighted and integrated into judgments of one's self-efficacy is filtered through, and thus it is shaped by cultural schemas which are elicited by the language used to formulate such judgments. Evidence that supports this viewpoint is reviewed.

Keywords: culture, self-efficacy, identity, cognition, bilingualism

1. Introduction

The research discussed in this chapter focuses on the construct of self-efficacy, which Bandura [1, 2] has defined as the belief in one's capabilities to initiate, manage and execute a variety of actions with the purpose of attaining desired goals. Self-efficacy beliefs are known to be key determinants of people's current and future behavior. For instance, individuals who possess high self-efficacy compared with those who have low self-efficacy tend to undertake challenging activities, are persistent, devote substantial effort to initiated activities, and experience fewer adverse emotional reactions if difficulties are encountered [1, 3, 4]. Self-efficacy beliefs also play an important role in shaping people's self-regulatory processes, such as goal setting and self-monitoring [5, 6].

The authors' research stems from the recognition that although the construct of self-efficacy exists in a variety of cultures, the information that people disclose about their self-efficacy is likely to be shaped by not only the self-report instrument used to gather information about such

a construct, but also the language in which it is written. Current work relies on two key assumptions: (a) a language carries its culture, including its unique denotations, connotations, prescriptions, and reactions [7, 8]; and (b) although translations of expressions used in measurement scales to refer to a person's capabilities are considered linguistic equivalents, they may not be culturally equivalent. Thus, a bilingual speaker's responses to a scale, which gathers information about his/her beliefs in personal capabilities, can unintentionally adjust to the language in which the scale is written and, consequently, to the unique denotations, connotations, prescriptions, and reactions of its accompanying culture. This particular adjustment performed by the bilingual speaker is likely to induce different answers to ostensibly the same statements presented in his/her two languages, because each statement remains conceptually unlike the other.

Supporting evidence is provided by the authors' empirical studies of Arabic-English bilinguals who estimated their capabilities through a self-efficacy scale either written in English or translated into Arabic. Evidence is explained by relying on a social constructionist framework [9, 10], according to which different languages, such as Arabic and English, can elicit different culturally oriented selves who are nevertheless connected to a single speaker [11]. Specifically, a solipsistic self, which is assumed to be prompted by English, is a self that is construed as autonomous and attuned to internal goals, thoughts, and motives. It is thus that of a person who is inclined to see himself/herself as the only one that really matters. Instead, a contextualized self, which is assumed to be evoked by Arabic, is a self that is interdependent at its core. It is the self of a person who recognizes that in-group goals take priority over personal goals. Does the bilingual speaker who is the vehicle of these two cultures indeed modulate his/her estimates of self-efficacy to adjust to the perspective of the self, either solipsistic or contextual, which is active at any given time? Section 3, which is devoted to our research methodology and findings, addresses this question. A narrative of the context that both motivated and permitted the research is outlined first.

2. The selected context: The Kingdom of Saudi Arabia (KSA)

KSA is currently a country where the clash between traditions and Western influences, in the form of values and norms, is simmering so powerfully at the surface of an ostensibly unruffled social and cultural system that its clatters are impossible to ignore. For instance, a cacophony of voices rarely heard before is manifest in the upsurge of social media use. At the same time, the ensuing impassioned conversations remain concealed in people's smart phones and computers as overt public gatherings remain unfeasible or tightly regulated [12–15]. Similarly, women may wear abayas and hijabs when visiting mixed-gender locales, such as shops and restaurants, but underneath, they often wear jeans, t-shirts and other Western-style garments. Most importantly, among the large contingent of young people who live in KSA, English is quickly moving from merely being the language of business transactions to being the language through which education is imparted and friends are made [13, 16]. If the use of a foreign language increases within a social group, not only its perceived utility is likely to be more widely recognized, but also people's contact with, understanding of, and internalization of the culture that the language smuggles in tend to be greatly enhanced. Thus, signs of change in KSA may

http://dx.doi.org/10.5772/intechopen.72740

appear to the untrained observer as whispers rather than rumbles, but they are as palpable and persistent as the well-known elaborateness of Arabs' verbal communication patterns [17].

According to scholarly narratives, a few key dimensions can be used to differentiate the traditional culture of KSA from the culture of English-speaking countries, such as the USA, including vertical collectivism and uncertainty avoidance [18, 19]. From the vantage point of a person's identity within the group to which he/she belongs, collectivism refers to viewing oneself as related to others in mutual interdependent relationships. When an interdependent self-shapes people's interactions with each other, associations within the group to which one belongs and its goals are likely to be highly valued. In contrast to collectivism, individualism refers to a viewpoint of a person's identity whereby one sees himself/herself as largely independent of the in-group (i.e., self-reliant). Accordingly, one's personal goals supersede those of the group.

If collectivism dictates preservation of in-group social bonds, thereby making one's actions expressly guided and highly regulated by group norms, individualism dictates independence and self-interest, thereby making relationships time-bound contracts based on a balance of costs and benefits. In either context, the dimension of verticality refers to the mere fact that the collective to which one belongs is recognized as hierarchically organized, and, within it, diversity and inequality are tolerated. The most visible by-product of verticality in collectivism is service to or even sacrifice for the in-group, whereas in individualism it is competition. Interestingly, the traditional culture of KSA has been labeled as being vertical collectivistic with a high level of uncertainty avoidance. Namely, it is a culture that views change as well as the ambiguity and uncertainty that tend to accompany it with apprehension. As such, it is a culture that has high regard for the preservation of the status quo through conformity to group norms and values. In contrast, the culture of the West, embodied by the USA, has been defined as being vertical individualistic with low uncertainty avoidance. Of course, it is important to bear in mind that the aforementioned distinctions are starkly dichotomous, but, in reality, the boundaries imposed by conceptual categories are often blurred.

3. The research context: questions, methodology and findings

Young people in KSA are uniquely positioned to embody both cultural orientations as Arabic-English bilingualism has gradually permeated educational institutions, social networks, values and informal norms across the country. Indeed, Murphy [13] noted that two KSAs are visible even to casual, untrained observers, one driven by religious traditions and tribal commitments, and one that imitates manners and follows ideas of the Western world, often epitomized by the English language spoken in the USA. Similarly to water spilled on a table, the impact of bilingualism and its offshoot, biculturalism, is not uniform. At one time, it may touch some objects while it leaves others untouched, whereas at another time, a different pattern emerges. It can even soak some of these objects while it merely pats the surface of others, depending on the contextual factors present. For instance, if a bilingual speaker's communicative behavior is considered, context may involve the setting where an event occurs, the situation faced, and the nature of the people present [see 9, 20–22]. Thus, the key issue is not so much the extent to which largely opposite cultural orientations shape, broadly speaking, the mind of KSA nationals,

but rather how specific areas of thought, emotion, and behavior, at a given point in time, are shaped by each orientation, thereby expressing temporary dominance over the other.

The methodological approach taken by our interdisciplinary research group is to attempt to address this issue by first selecting a key construct of human existence (e.g., self-efficacy), and then determining the role played by each cultural orientation in it among KSA youth. The key premise upon which this methodological approach relies is two-fold: First, besides the properties possessed by human language that make it a unique communication system among those used by other species (e.g., recursion), one property that is particularly notable in every-day life is language's ability to be the vehicle of the culture held by the collective who makes use of it. As such, it permits each member to transmit and share key aspects of that culture, such as norms and values, to express a common identity, as well as to selectively perceive and interpret events in and outside the collective to which one belongs. A suitable metaphor to describe the latter is that a language "is 'a veil' over the reality of the culture in which it is used, involving an agreement of its users about what there is to be seen and how it should be seen" [23, p. 89]. If a language, as a communication system within a people, ensures the transmission and sharing of a culture, its use can be expected to prime (i.e., make available in one's mind) the selected culture. Second, when bilingual individuals rely on one of the languages they possess as the current mode of communication, they are likely to experience and exhibit a distinctive self which is the expression of the culture associated with their linguistic selection.

Important to note is that the premises of our interdisciplinary research and its findings neither espouse linguistic determinism nor blindly adopt linguistic relativism. They merely rely on the idea that language can prime (activate in one's mind) a culture. Priming refers to the incidental (i.e., automatic) activation of knowledge by the current environment [24]. Priming has been assessed through a variety of materials (e.g., spoken and written words, pictures, drawings, and sounds) and tasks (e.g., stem completion, fragment completion, and lexical decision) in which accuracy and speed are usually the criteria used to measure participants' responses. Of course, a prime is not equivalent to a hypnotist's command which can trigger in a person's mind a precise thought and a correspondingly precise action. Priming heavily relies not only on specific knowledge of objects and events in one's environment as well as knowledge of the range of potential responses to such objects and events, but also on pre-existing associations between contents of the environment and one's knowledge. Although its impact is much more subtle and heavily dependent on previously acquired knowledge, it has been reported to occur in a variety of domains of human action and thought where the presentation of a stimulus (e.g., the word "furniture") can change a person's response to a subsequent stimulus (e.g., the speed and accuracy at which he/she identifies "chair" as a word from a list of both words and non-words; [25]). For instance, the presentation of stimuli related to either the concept of rudeness or the concept of politeness can lead research participants to behave in ways consistent with the activated concept [24]. Similarly, the presentation of stimuli related to the concept of old age can steer young adults to adopt a slower walking speed than that of young adults exposed to neutral stimuli [24]. Priming effects involve not only content, but also form as in the case of syntactic priming. Namely, participants' exposure to a sentence with a particular syntactic construction can affect the subsequent processing of an otherwise unrelated sentence with the same structure [26]. Of course, the influence of priming on observable

behavior is simply the epiphenomenon of its impact on cognition. Thus, it is not surprising that the presentation of stimuli related to hostility can make research participants more likely not only to behave in a hostile manner, but also to perceive hostility in others [27, 28].

In sum, evidence of priming indicates that knowledge, including schemas and other organized mental databases, can become activated, and thus accessible to a person, from events unfolding in his/her environment. Furthermore, evidence shows that one's tendency to behave in a manner consistent with a schema is increased when the latter is activated. Grounded in this knowledge, the primary aim of our research is to examine whether priming occurs by means of the language in which instruments of self-assessment are written. Yet, even if it is agreed that a culture shapes human cognition and behavior, can two cultures naturally coexist in a bilingual speaker as two separate but interacting selves? It is reasonable to assume that in the process of learning a language, bilingual individuals interiorize the values, norms, and concepts, including role expectations and attitudes, of the culture expressed by each of the languages they are mastering [29]. Furthermore, because each universe of linguistic and cultural knowledge, habits and skills has its unique functional utility, often linked to particular settings and motives [30], both universes can not only co-exist, but also be employed strategically in the service of one's goals. Thus, bilingual speakers can be primed by either of the two languages they possess. Namely, at any given time and under proper environmental conditions, they can be expected to express in their cognition and behavior the unique features of the culture linked to the language in use.

Obviously, in research pertaining to bilingualism, comparability of results of different cultural schemas rests, first and foremost, on translations that maintain the original meaning and intent of the source-language in which the selected assessment instrument was written (i.e., fidelity). Moreover, translations must remain comprehensible and culturally relevant to individuals for whom the receptor-language is the native language (i.e., dynamic equivalence; [29, 31]). Yet, the prerequisite to assess the role of language as a prime for culture is that the two languages (e.g., Arabic and English) must activate somewhat unique cultural schemas. Furthermore, to attribute such a role to language, each schema must shape the person's responses to the selected instrument *in accordance with* the norms and values of the activated schema. Consistency between the expected content of the activated cultural schema and participants' responses is critical since behavioral differences as a function of language can be disputed as merely arising from a deficient translation.

Among the many psychological constructs that define human existence, self-efficacy emerges as the ideal testing ground for assessing the role of language as a prime for culture, especially if the selected languages pertain to cultural orientations whose key dimensions stand in opposition to one another. For the purpose of our research, self-efficacy is selected as the object of study not only because of the central function that this psychological construct plays in human cognition and behavior, but also because of scant cross-cultural research on self-efficacy beliefs that includes the Arab world. According to Bandura [1], general self-efficacy beliefs refer to cognitions about one's competence to initiate, manage and execute a variety of actions across domains with the purpose of attaining sought-after goals. Self-efficacy, which depends on one's ability to self-reflect, is a fundamental aspect of human agency that captures the essence of most human actions. Namely, actions are, by and large, motivated and guided by the conviction

that they produce effects, and thus that a person is able to exercise some measure of control over his/her own functioning as well as over events in his/her life [2]. Self-efficacy beliefs are the product of a complex self-appraisal process which requires that the person selects, weights, and integrates information about his/her performance from four key sources: (a) direct experiences of success and failure; (b) vicarious experiences through the observation and consideration of the behavior of models; (c) feedback from respected others; and (d) emotional and physical reactions exhibited while dealing with a variety of circumstances.

Culture can affect self-efficacy beliefs by guiding the selection of, the weight attributed to, and the integration of the information about the self obtained from each of the sources that are used to nurse such beliefs across the lifespan [32]. In a vertical-collectivistic culture with high uncertainty avoidance (e.g., KSA), one may expect a heavily structured educational system and family environment where children's self-appraisal largely depends on demonstrating required competencies to the collective. Self-appraisal is also influenced by feedback from authority figures (i.e., teachers and parents) whose experience and wisdom are neither to be questioned nor contradicted. Accordingly, children are perceived as passive vessels of the culture that they are required to assimilate and replicate [13]. Thus, the goal is not only to reach performance indicators that are recognized by the collective, one of which is the quality of rote memorization [33], but also to ensure that goal attainment is at the service of the collective. In contrast, in a vertical-individualistic culture with low uncertainty avoidance (e.g., USA), one may expect an educational system that is child-focused rather than teacher-focused, as well as a family environment where children's self-appraisal is more subjective, and less dependent on authority figures, since the judgment and knowledge of the latter can be questioned. In this type of culture, ambiguity does not tend to be perceived as a threat, but rather as a challenge.

In vertical individualism, competition may be expected to be tangible, often entailing harsh comparisons with peers, broadcasting successes, and understating or hiding failures. In vertical collectivism, on the other hand, performance differences are likely to be accepted if they serve the collective of interdependent selves, but are rejected if they are viewed as self-serving. Thus, although in the latter, the goal is to reach recognized performance indicators, broadcasting one's achievements as personal successes is likely to be perceived as contrary to serving and sacrificing for the collective. In KSA, the value of modesty emerges from its collectivistic framework and has had as its primary promoter an educational system that emphasizes memorization at the expense of exploration of ideas [34–36]. As a result, a unique interpretation of the value of modesty has been practiced, which by confounding humility with conformity, is likely to render overt expressions of self-confidence objectionable.

If vertical collectivism values modesty and vertical individualism does not [37], how will Arabic-English bilingual young adults who are natives of KSA respond to a scale that measures self-efficacy which, by definition, requires respondents to overtly estimate their own capabilities? Will each of the two languages in which the scale can be written prime distinct cultural schemas and thus lead to reports of self-efficacy consistent with the contents of the primed schemas? To answer these questions in our preliminary studies[1], we examined the self-efficacy reports of Arabic-English

[1]Portions of these data have been published elsewhere: El Alaoui, K., Mulhem, H., Pilotti, M. A. E., Amir, S., & Tallouzi, E. (2017). Arabic-English bilingual speakers' reactions to the statements of the new general self-efficacy scale. The International Journal of Learner Diversity and Identities, 24(2), 21–38.

bilingual speakers. All are female college students ($n = 627$) who attend a private university in the Eastern Province of KSA (age range: 18–25). Arabic is participants' first language, whereas English is their second language, mostly learned and practiced at school and among peers.

Participants' mean age of acquisition of the second language was 5.71 ($SD = 2.86$), and mean proficiency, measured by the Shipley vocabulary test [38], was 66.80% ($SD = 16.17$). For university admission, participants had demonstrated English language proficiency through standardized English proficiency tests (i.e., IELTS, Aptis, or TOEFL). The Shipley test was selected for its good psychometric properties [39], and demonstrated ease of administration in educational, cognitive, and clinical research. It served as a current measure of second language proficiency. As bilingual speakers, participants were expected to have developed two cultural selves, one linked to the communal cohesion of the KSA cultural tradition and one linked to the individualistic values produced by the forces of industrialization and economic mobility of the Eastern Province of KSA [40]. As KSA nationals, participants were considered carriers of the value of modesty, which is a key aspect of vertical collectivism, but a trait discouraged by vertical individualism. As females in KSA, the value of modesty was expected to be highly relevant to their cultural schema of vertical collectivism, because females in KSA are the primary recipients of onerous restrictions of movement, thought and affect, which are purportedly intended to ensure and guard their virtues [41–44].

The New General Self-Efficacy (NGSE) scale, developed by Chen, Gully, and Eden [45], was selected over other self-report measures for (a) its focus on assessing a general sense of mastery that is not associated with a particular situation, domain or behavior, (b) its desirable psychometric properties [see 45], (c) brevity and thus ease of administration, and (d) the absence of data gathered with this tool in KSA. The NGSE tool includes eight general statements of performance capabilities, each one containing the pronoun *I*. It required participants to report the extent of their agreement with each statement on a 5-point Likert scale from *strongly disagree* to *strongly agree*. The Arabic translation of the NGSE measure was based on a consensus model whose goals was to obtain a culturally sensitive adaptation of the scale which maintained the meaning and intent of the original English version [29, 31, 46]. Participants were randomly assigned to either the Arabic or English version of the NGSE measure. Their task was to complete it along with a vocabulary test [38] as well as to answer a few demographic questions.

At the time of the administration, existing evidence of potential differences in self-efficacy between respondents from individualistic and collectivistic cultures was unclear. For instance, Scholz et al. [46] found lower scores for Japanese participants, but not for participants from other collectivistic cultures. Wu [47] using the same data found no relationship between self-efficacy and degree of individualism and collectivism. However, the assessment instrument of general-self-efficacy was administered in the participants' first language. Thus, the extent to which languages can activate distinctive cultural schemas in the same respondents had not been tested. Furthermore, to our knowledge, no evidence existed from a vertical collectivistic culture, such as that of KSA, where religion (i.e., Islam) has wide-ranging prescriptive norms on modesty and on serving the collective.

In our preliminary studies, significant lower self-efficacy scores were obtained in the Arabic condition ($M = 2.18$; $SD = .46$) than in the English condition ($M = 3.87$; $SD = .53$; $F(1, 625) = 1825.43$, $MSE = .24$, $p < .05$, *partial Eta*$^2 = .75$ (See **Figure 1**). The language in which the assessment tool

was administered was not linked to group differences in either vocabulary scores or age of acquisition, $F \leq 2.48$, *ns*. Similarly, no significant relationships, as measured by Pearson correlation coefficients, were found between self-efficacy scores and either vocabulary scores or age of acquisition.

If Arabic indeed activates the cultural schema of vertical collectivism, in which modesty is relevant, whereas English activates the cultural schema of vertical individualism, in which modesty is discouraged, then a direct test of participants' attitude toward modesty should produce results consistent with the revealed self-efficacy pattern. To test this hypothesis, a sample of Arabic-English bilingual females from the same population described above completed a measure of modesty, a general self-efficacy measure (n = 204), or both (n = 196). Instructions and measuring instruments were consistently presented to each participant in either Arabic or English. Random assignment to language condition was again utilized. As in all studies, each participant was also asked to complete an English vocabulary test [38], and respond to a few questions regarding the time and mode of second language acquisition. For this sample, the mean age of acquisition of the second language was 7.85 (SD = 3.83), and mean proficiency, measured by the Shipley vocabulary test [38], was 67.69% (SD = 20.83). The NGSE scale of Chen et al. [45] was again adopted to estimate self-efficacy. The Modest Responding scale [48, 49] was selected to assess participants' value of modesty because it distinguished one's internal disposition toward modesty (e.g., "I prefer to keep my accomplishments to myself than talk about them"), or disinclination toward modesty/propensity to brag (e.g., "When people tell me about one of their successes, I like to tell them about one of mine"), from one's view of modest responses as being socially desirable (e.g., "I believe it's impolite

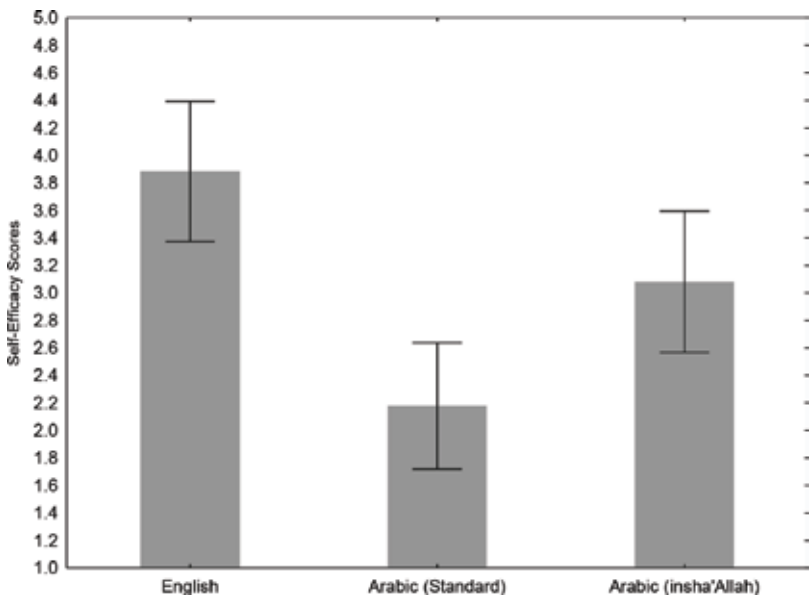

Figure 1. Mean self-efficacy scores and standard deviations of the NGSE scale administered in English, Arabic, or Arabic with the qualifier *insha'Allah* added to each of the eight statements.

http://dx.doi.org/10.5772/intechopen.72740

to talk excessively about one's achievement, even if they are outstanding"). Each statement of the measurement instrument was to be answered on a 7-point Likert scale that ranged from *strongly disagree* to *strongly agree*.

Language primed both self-efficacy and desire to be perceived as modest, and did so in a manner consistent with our predictions. Specifically, Arabic compared with English led to lower self-evaluations of mastery, $F(1, 202) = 581.27$, $MSE = .34$, $p < .001$, *partial Eta*$^2 = .742$ ($M = 2.00$; $SD = .57$ and $M = 3.97$, $SD = .59$, respectively), and higher desire to be perceived as modest by others, $F(1, 194) = 7.42$, $MSE = .80$, $p = .007$, *partial Eta*$^2 = .037$ ($M = 4.68$, $SD = .87$ and $M = 4.33$; $SD = .92$, respectively; see **Figure 2**). Inclination toward modesty ($M = 4.02$; $SD = .81$), disinclination toward modesty ($M = 4.50$; $SD = .81$), English vocabulary scores and age of acquisition were not significantly different between language conditions, $Fs < 1$, *ns*. The findings that young female students exhibited lower self-efficacy along with a desire to be perceived as modest are consistent with those of earlier studies where the salience of a role or norm induced behavior consistent with the activated role or norm [50–55]. They are also consistent with the view [see 37] that the value of modesty, serving as a self-restraint, is integral to the cultural schema of vertical collectivism (as activated by Arabic), and inconsistent with the cultural schema of vertical individualism (as activated by English). Actually, the latter has been often mentioned as promoting self-enhancement, which refers to the tendency of one's self-perceptions to be self-serving and more positive than those produced by the judgment of others, such as peers and independent observers [56, 57]. Most importantly, what seems to matter in the cultural schema of vertical collectivism is not a person's disposition to be modest, an internal attribute, but rather the perceptions that others have of the person as someone who upholds modesty.

Self-enhancement and modesty can be seen as opposite devices to nurture a solipsistic self-perception and a communal (i.e., other-focused) self-perception, respectively. If self-enhancement is associated with vertical individualism, where self-interest and competition

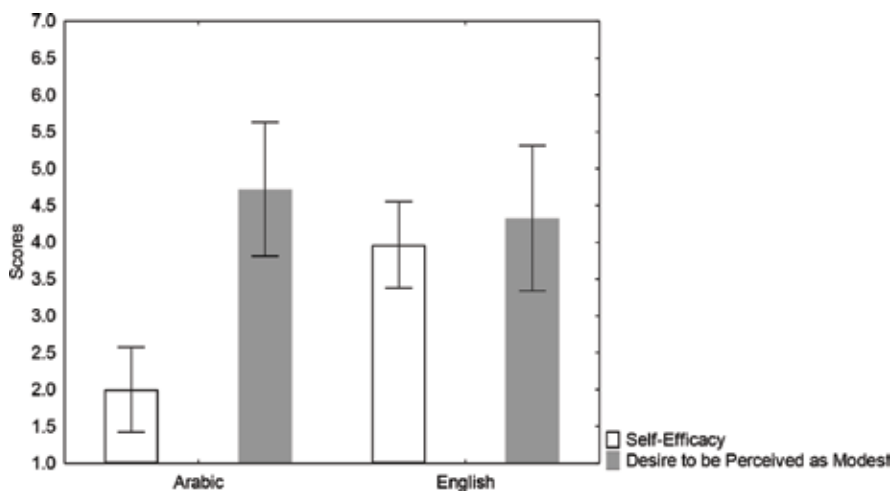

Figure 2. Mean self-efficacy scores, mean desire to be perceived as modest, and standard deviations as a function of the language in which the NGSE scale (range: 1–5) and the Modest Responding scale (range: 1–7) were administered.

reign, and modesty is nurtured by vertical collectivism, where service to and sacrifice for one's community are valued, one may ask how these two self-orientations can co-exist in the bilingual speaker. Perhaps, the most parsimonious account of the co-existence of opposite viewpoints is that they serve as presentation tactics through which to facilitate one's interactions with others in different contexts. Although presentation tactics can be seen as producing changes that merely scratch the surface of one's identity, the extent to which their impact seeps from the surface to the very core of that identity is to be further investigated. In fact, in our research, Arabic was associated with not only enhanced concern to be seen as modest, but also with reports of lower self-efficacy. Furthermore, since modesty emerged mostly as a concern for the views that others have of oneself, it is reasonable to ask whether the impact of this outward orientation is to make people who have interiorized vertical collectivism particularly susceptible to expressions that mitigate the boastful tone implicit in the NGSE scale. To find a preliminary answer to this question, a change to the Arabic translation of the NGSE measure was introduced based on careful consideration of the pragmatics of the expression *insha'Allah* إن شاء الله.

Insha'Allah, is a common expression that saturates social exchanges of KSA natives and of Arabs in general, which is habitually used to denote that events in one's life are ultimately in the hands of a higher power [58]. According to Nazzal [59], in addition to confirming one's cultural and religious identity, the expression has three pragmatic functions, which serve to mitigate the intended meaning of a (a) rejection of a request, (b) prediction of future events, and (c) implicit acceptance of a statement or course of action. We examined the extent to which the latter function of insha'Allah can indeed counteract participants' discomfort with self-enhancement and, at the same time, preserve the value of modesty (as a self-restraint for public consumption and thus simply as a self-presentation tactic). To do so, insha'Allah was added to each of the eight statements of the Arabic version of the NGSE instrument (n = 136). Not surprisingly, it had the effect of enhancing the self-efficacy reports of students pooled from the same population of the earlier studies (n = 329; insha'Allah version: M = 3.08; SD = .51; original version: M = 2.18; SD = .46, respectively), $F(1, 463) = 346.73$, MSE = .23, $p < .001$, *partial Eta*^{2} = .428 (see **Figure 1**). Although the self-efficacy level of students given the insha'Allah Arabic version of the NGSE measure improved, it was still statistically lower than the level obtained by a random sample of students (n = 293) given the English version of the NGSE in the earlier studies, $F(1, 427) = 228.76$, MSE = .26, $p < .001$, *partial Eta*^{2} = .349 (original version: M = 3.88; SD = .51, respectively). Thus, the addition of insha'Allah attenuated, but did not eliminate the restrictions about self-enhancement that are an integral part of the vertical collectivistic culture of KSA.

Important to note is that modesty often appears to be a different object of study, depending on the vantage point of the researchers who study it. For instance, psycholinguists have focused on politeness in communication practices, personality psychologists have attempted to determine whether modesty is a human disposition, whereas social psychologists have preferred to conceptualize it as a self-presentation tactic relevant to impression management [see 60]. Our finding that participants desired to be seen as modest supports the latter viewpoint, best exemplified by the definition of Cialdini and de Nicholas [61] that modesty is "the under-representation of one's positive traits, contributions, expectations, or accomplishments" (p. 626). Yet, because

http://dx.doi.org/10.5772/intechopen.72740

evidence of modesty was elicited by the Arabic language relative to the English language, modesty can be said to be a marker of vertical collectivism, but not of vertical individualism where self-enhancement tactics, to a certain extent, may be necessary for one's survival in a competitive environment.

4. Conclusions

In sum, languages are not only a sign of group membership, and a medium through which social relationships are established, maintained or even interrupted, but also a medium through which the contents of cultures, including norms and values, are expressed and transmitted inside and outside a collective as well as interiorized or rejected by the members of that collective [62]. As such, a language embodies a culture which can shape in its image the realities that its speakers perceive in everyday life (including self-perceptions of attributes, dispositions, intentions and actions). Thus, merely using a language can activate its corresponding culture which then influences speakers' thoughts and actions, especially when thoughts and actions directly concern the self (e.g., self-efficacy reports). Our findings are consistent with those of studies showing that bilinguals' responses, including self-reports and perceptions, memory records, and emotional responses, vary with the language used to elicit such responses [see 63–69]. In particular, the research discussed in this chapter suggests that bilingual speakers can activate, through the use of Arabic, a cultural schema (i.e., vertical collectivism) that is opposite in nature to the schema activated by English (i.e., vertical individualism). Thus, bilingual speakers can be said to possess not only two languages, but also two cultures, which like code-switching (i.e., the use of more than one language in the same conversation; [70]), can be exercised strategically as self-presentation tactics and practices may dictate [10]. Yet, it is reasonable to ask whether the two culturally construed identities that bilinguals possess will merge into a complex, multifaceted whole or sharpen their distinctions as experience with the two languages increases. The answer to this question may come from data about the plasticity of the brain [71–73], which are yet to be conclusive. Nevertheless, if the interactions between two interiorized cultures, whose values and norms mostly conflict with each other, resemble those exhibited by two languages, each culture is unlikely to lose its distinctiveness. In the everyday life of a young adult, the coexistence of two languages may be effortful, produce interference and errors, and yield less-than-optimal performance on comprehension and production tasks. However, in old age, bilingualism has been shown to enhance executive functioning and even protect the latter against age-related declines [74]. Thus, it is entirely possible that the effects of biculturalism on the human brain will be found to resemble those currently attributed to bilingualism.

Rarely, bilingualism is fully balanced (i.e., speakers who are equally and entirely fluent in both languages). Most bilingualism is unbalanced since the first language (L1), whose acquisition has usually occurred early in life, is more dominant than the second language (L2). Studies have shown that in highly proficient bilinguals, largely overlapping brain areas are activated by the two languages, even in spite of differences in age of acquisition [75–77]. Our participants' use and knowledge of Arabic and English can be described as illustrating unbalanced

bilingualism with Arabic being the dominant language, learned at an early stage in life and spoken in a wide variety of contexts, and English being more likely to be linked to formal education and peers' social networks. Thus, it is interesting that even unbalanced bilingualism permits the activation of opposing cultural schemas, each related to one of the languages that the speaker is currently using. It is reasonable to expect languages to preserve their functional utility and coexist as largely distinct entities in the speakers' minds if each language is attached to cultural norms and values relevant to the speaker's community. It is the task of future research to determine the extent to which the same pattern may be obtained with less proficient bilinguals. Similarly, it is an open question whether our findings may generalize to older participants, males, participants whose financial resources are less plentiful, and those who live in countries of the Arab world where industrialization has been accompanied by a more visible relaxation of traditional norms. For instance, it may be of interest to determine whether differences in the extent to which young adults of several Arab countries are connected to their families financially (i.e., monetary reliance), emotionally (i.e., need for family approval), and functionally (i.e., sharing of daily routines; [78]) translate into differences in self-efficacy beliefs, either general or related to specific domains of knowledge and practice. Lastly, important to note is that cultural differences, which are observed with explicit self-appraisal measures, tend to be less likely to be observed with implicit measures, such as the Implicit Association Test [79]. Thus, results reflected in explicit self-appraisal measures, such as ours, may be merely inflated by self-presentation practices and concerns [80] or be the undiluted expression of such practices and concerns. Methodological issues besides the explicit versus implicit nature of the assessment measure will also need to be considered in future research (e.g., comparisons involving bilingual individuals or different groups of monolingual individuals).

Acknowledgements

The research mentioned in this chapter is based on discussions conducted by an interdisciplinary group of researchers, including besides the authors, Dr. Siddiqua Aamir, Ms. Khadijah Aldabbagh, Ms. Mona M. A. al-Mqbas, Ms. Houda Debbabi, Dr. Amani K. Hamdan, Dr. Huda Mulhem, and Mrs. Ebtesam Ahmad Tallouzi. We thank all collaborators for sharing their valuable insights into the complexities of the phenomena that we have explored empirically. We also thank the students who provided reliable data as well as helpful commentaries and observations that comprise the indispensable pillars of our inquiry into the human mind in diverse cultures.

Notes

Maura A. E. Pilotti and Khadija El Alaoui, College of Sciences and Human Studies, Prince Mohammad Bin Fahd University, P.O. Box 1664, Al Khobar 31952, Kingdom of Saudi Arabia. Correspondence concerning this article should be addressed to either author.

Author details

Maura A. E. Pilotti* and Khadija El Alaoui

*Address all correspondence to: maura.pilotti@gmail.com

Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia

References

[1] Bandura A. Self-Efficacy: The Exercise of Control. New York: Freeman & Company; 1997

[2] Bandura A. Social cognitive theory: An agentic perspective. Annual Review of Psychology. 2001;**52**(1):1-26. DOI: 10.1146/annurev.psych.52.1.1

[3] Pajares F. Current directions in self-efficacy research. Advances in Motivation and Achievement. 1997;**10**(149):1-49

[4] Pajares F, Kranzler J. Self-efficacy beliefs and general mental ability in mathematical problem-solving. Contemporary Educational Psychology. 1995;**20**:426-443. DOI: 10.1006/ceps.1995.1029

[5] Pajares F. Gender and perceived self-efficacy in self-regulated learning. Theory Into Practice. 2002;**41**(2):116-125. DOI: 10.1207/s15430421tip4102_8

[6] Zimmerman BJ, Bandura A. Impact of self-regulatory influences on writing course attainment. American Educational Research Journal. 1994;**31**:845-862. DOI: 10.3102/00028312031004845

[7] Müller M. What's in a word? Problematizing translation between languages. Area. 2007;**39**(2):206-213. DOI: 10.1111/j.1475-4762.2007.00731.x

[8] Ochs E, Schieffelin BB. The theory of language socialization. The Handbook of Language Socialization. 2011;**71**:1-21. DOI: 10.1002/9781444342901.ch1

[9] Hong YY, Morris MW, Chiu CY, Benet-Martinez V. Multicultural minds: A dynamic constructivist approach to culture and cognition. American Psychologist. 2000;**55**(7):709-720. DOI: 10.1037/0003-066X.55.7.709

[10] Pavlenko A, editor. Bilingual Minds: Emotional Experience, Expression, and Representation. Clevedon: Multilingual Matters; 2006

[11] Triandis HC, Chan DKS, Bhawuk DP, Iwao S, Sinha JB. Multimethod probes of allocentrism and idiocentrism. International Journal of Psychology. 1995;**30**(4):461-480. DOI: 10.1080/00207599508246580

[12] Askool SS. The use of social media in Arab countries: A case of Saudi Arabia. In: Cordeiro J, Krempels KH, editors. Web Information Systems and Technologies. Berlin: Springer-Verlag Berlin Heidelberg; 2013. pp. 201-219

[13] Murphy C. A kingdom's Future: Saudi Arabia Through the Eyes of Its Twentysomethings. Washington, DC: Wilson Centre; 2013

[14] Sait SM, Al-Tawil KM. Impact of internet usage in Saudi Arabia: A social perspective. International Journal of Information Technology and Web Engineering (IJITWE). 2007;**2**(2):81-115. DOI: 10.4018/jitwe.2007040104

[15] Simsim MT. Internet usage and user preferences in Saudi Arabia. Journal of King Saud University-Engineering Sciences. 2011;**23**(2):101-107. DOI: 10.1016/j.jksues.2011.03.006

[16] Prokop M. Saudi Arabia: The politics of education. International Affairs. 2003;**79**(1):77-89. DOI: 10.1111/1468-2346.00296

[17] Feghali E. Arab cultural communication patterns. International Journal of Intercultural Relations. 1997;**21**(3):345-378. DOI: 10.1016/S0147-1767(97)00005-9

[18] Hofstede G. Culture's Consequences: International Differences in Work-Related Values. Beverly Hills: Sage; 1980

[19] Singelis TM, Triandis HC, Bhawuk DP, Gelfand MJ. Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. Cross-Cultural Research. 1995;**29**(3):240-275. DOI: 10.1177/106939719502900302

[20] Chen SX, Bond MH. Explaining language priming effects: Further evidence for ethnic affirmation among Chinese-English bilinguals. Journal of Language and Social Psychology. 2007;**26**(4):398-406. DOI: 10.1177/0261927X07306984

[21] Hong YY, Chiu CY, Kung TM. Bringing culture out in front: Effects of cultural meaning system activation on social cognition. Progress in Asian Social Psychology. 1997;**1**:135-146

[22] Kagan J. Five Constraints on Predicting Behavior. Cambridge, MA: MIT Press; 2017

[23] Kim YY. Communication and Cross-Cultural Adaptation: An Integrative Theory. Philadelphia: Multilingual Matters; 1988

[24] Bargh JA, Chen M, Burrows L. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. Journal of Personality and Social Psychology. 1996;**71**(2):230. DOI: 10.1037/0022-3514.71.2.230

[25] Schacter DL, Chiu CYP, Ochsner KN. Implicit memory: A selective review. Annual Review of Neuroscience. 1993;**16**(1):159-182. DOI: 10.1146/annurev.ne.16.030193.001111

[26] Branigan HP, Pickering MJ, Liversedge SP, Stewart AJ, Urbach TP. Syntactic priming: Investigating the mental representation of language. Journal of Psycholinguistic Research. 1995;**24**(6):489-506. DOI: 10.1007/BF02143163

[27] Bargh JA, Pietromonaco P. Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. Journal of Personality and Social Psychology. 1982;**43**(3):437-449. DOI: 10.1037/0022-3514.43.3.437

[28] Carver CS, Ganellen RJ, Froming WJ, Chambers W. Modeling: An analysis in terms of category accessibility. Journal of Experimental Social Psychology. 1983;**19**(5):403-421. DOI: 10.1016/0022-1031(83)90019-7

[29] Sperber AD. Translation and validation of study instruments for cross-cultural research. Gastroenterology. 2004;**126**:S124-S128. DOI: 10.1053/j.gastro.2003.10.016

[30] Auer P, editor. Code-Switching in Conversation: Language, Interaction and Identity. London: Routledge; 2013

[31] Nida E. Principles of correspondence. In: Venuti L, editor. The Translation Studies Reader. New York: Routledge; 2012. pp. 1126-1140

[32] Oettingen G. Cross-cultural perspectives on self-efficacy. In: Bandura A, editor. Self -Efficacy in Changing Societies. New York: Cambridge University Press; 1995. pp. 149-176

[33] Yusuf M. Memorization as a learning style: A balance approach to academic excellence. OIDA International Journal of Sustainable Development. 2010;**1**(6):49-58

[34] Alamer SM. Challenges facing gifted students in Saudi Arabia. Research on Humanities and Social Sciences. 2014;**4**(24):107-112

[35] al-Hashimi AT. Methods of teaching Islamic education. Damascus, Syria: ar-Rissalat. [in Arabic]; 2000

[36] Rugh WA. Education in Saudi Arabia: Choices and constraints. Middle East Policy. 2002;**9**(2):40. DOI: 10.1111/1475-4967.00056

[37] Kurman J. Why is self-enhancement low in certain collectivistic cultures. Journal of Cross Cultural Psychology. 2003;**34**:496-510. DOI: 10.1177/0022022103256474

[38] Shipley WC. A self-administering scale for measuring intellectual impairment and deterioration. Journal of Psychology. 1940;**9**:371-377. DOI: 10.1080/00223980.1940.9917704

[39] Goodman JT, Streiner DL, Woodward CA. Test-retest reliability of the Shipley-Institute of Living Scale: Practice effects or random variation. Psychological Reports. 1974; **35**(1):351-354

[40] Ali AJ. The Islamic work ethic in Arabia. The Journal of Psychology. 2001;**126**(5):507-519. DOI: 10.1080/00223980.1992.10543384

[41] al-Hashimi MA. The Ideal Muslimah: The True Islamic Personality of the Muslim Woman. Riyadh: International Islamic Publishing House; 1996

[42] al-Olayan FS, Karande K. A content analysis of magazine advertisements from the United States and the Arab world. Journal of Advertising. 2000;**29**(3):69-82. DOI: 10.1080/00913367.2000.10673618

[43] Luqmani M, Yavas U, Quraeshi Z. Advertising in Saudi Arabia: Content and regulation. International Marketing Review. 1989;**6**(1). DOI: 10.1108/EUM0000000001503

[44] Syed J. An historical perspective on Islamic modesty and its implications for female employment. Equality, Diversity and Inclusion: An International Journal. 2010;**29**(2):150-166. DOI: 10.1108/02610151011024475

[45] Chen G, Gully SM, Eden D. Validation of a new general self-efficacy scale. Organizational Research Methods. 2001;**4**(1):62-83. DOI: 10.1177/109442810141004

[46] Scholz U, Doña BG, Sud S, Schwarzer R. Is general self-efficacy a universal construct? Psychometric findings from 25 countries. European Journal of Psychological Assessment. 2002;**18**(3):242. DOI: 10.1027//1015-5759.18.3.242

[47] Wu CH. Factor analysis of the general self-efficacy scale and its relationship with individualism/collectivism among twenty-five countries: Application of multilevel confirmatory factor analysis. Personality and Individual Differences. 2009;**46**(7):699-703. DOI: 10.1016/j.paid.2009.01.025

[48] Cialdini RB, Wosinska W, Dabul AJ, Whetstone-Dion R, Heszen I. When social role salience leads to social role rejection: Modest self-presentation among women and men in two cultures. Personality and Social Psychology Bulletin. 1998;**24**(5):473-481. DOI: 10.1177/0146167298245003

[49] Whetstone MR, Okun MA, Cialdini RB. The Modest Responding Scale. Paper Presented at the Convention of the American Psychological Society; August 1992; San Diego

[50] Berkowitz L, Daniels LR. Affecting the salience of the social responsibility norm. Journal of Abnormal and Social Psychology. 1964;**66**:275-281. DOI: 10.1037/h0040164

[51] Cialdini RB, Reno RR, Kallgren CA. A focus theory of normative conduct. Journal of Personality and Social Psychology. 1990;**58**:1015-1026. DOI: 10.1037/0022-3514.58.6.1015

[52] Deaux K, Major B. Putting gender into context: An interactive model of gender related behavior. Psychological Review. 1987;**94**(3):369. DOI: 10.1037/0033-295X.94.3.369

[53] Kramer RM, Goldman L. Helping the group or helping yourself? Social motives and group identity in resource dilemmas. In: Schroeder DA, editor. Social Dilemmas: Perspectives on Individuals and Groups. Westport: Praeger; 1995. pp. 49-67

[54] Messick DM. Equality as a decision heuristic. In: Mellers BA, Baron J, editors. Psychological Perspectives on Justice: Theory and Applications. New York: Cambridge University Press; 1993. pp. 51-68

[55] Turner JC, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS. Rediscovering the Social Group: A Self-Categorization Theory. Oxford: Basil Blackwell; 1987

[56] John OP, Robins RW. Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. Journal of Personality and Social Psychology. 1994;**66**(1):206. DOI: 10.1037/0022-3514.66.1.206

[57] Taylor SE, Brown JD. Illusion and well-being: A social psychological perspective on mental health. Psychological Bulletin. 1988;**103**(2):193. DOI: 10.1037/0033-2909.103.2.193

[58] Mehawesh MI, Jaradat AA. Inshallah: Extensive flouting of Grice's maxim of quality. Asian Social Science. 2015;**11**(4):319. DOI: 10.5539/ass.v11n4p319

[59] Nazzal AR. The pragmatic functions of the recitation of Qur'anic verses by Muslims in their oral genre: The case of Insha'Allah, 'God's willing'. Pragmatics. 2005;**15**(2/3):251-273. DOI: 10.1075/prag.15.2-3.05naz

[60] Chen S, Bond MH, Chan B, Tang D, Buchtel EE. Behavioral manifestations of modesty. Journal of Cross-Cultural Psychology. 2009;**40**(4):603-626. DOI: 10.1177/0022022108330992

[61] Cialdini RB, de Nicholas ME. Self-presentation by association. Journal of Personality and Social Psychology. 1989;**57**(4):626. doi:10.1037/0022-3514.57.4.626

[62] Saville-Troike M. Cross-cultural communication in the classroom. In: Alatis JE, editor. Current Issues in Bilingual Education. Washington, DC: Georgetown University Press; 1980. pp. 348-355

[63] Bond MH. How language variation affects inter-cultural differentiation of values by Hong Kong bilinguals. Journal of Language and Social Psychology. 1983;**2**(1):57-66. DOI: 10.1177/0261927X8300200104

[64] Koven M. Two languages in the self/the self in two languages: French Portuguese bilinguals' verbal enactments and experiences of self in narrative discourse. Ethos. 1998; **26**(4):410-455. DOI: 10.1525/eth.1998.26.4.410

[65] Kuroda Y, Hayashi C, Suzuki T. The role of language in cross national surveys: American and Japanese respondents. Applied Scholastic Models and Data Analysis. 1986;**2**:43-59. DOI: 10.1002/asm.3150020105

[66] Marian V, Kaushanskaya M. Self-construal and emotion in bicultural bilinguals. Journal of Memory and Language. 2004;**51**:190-201. DOI: 10.1016/j.jml.2004.04.003

[67] Panayiotou A. Switching codes, switching code: Bilinguals' emotional responses in English and Greek. Journal of Multilingual and Multicultural Development. 2004; **25**(2/3):124-139. DOI: 10.1080/01434630408666525

[68] Ross M, Xun WQE, Wilson A. Language and the bicultural self. Personality and Social Psychology Bulletin. 2002;**28**(8):1040-1050. DOI: 10.1177/01461672022811003

[69] Trafimow D, Silverman E, Fan RM-T, Law JSF. The effects of language and priming on the relative accessibility of the private self and the collective self. Journal of Cross-Cultural Psychology. 1997;**28**:107-123. DOI: 10.1177/0022022197281007

[70] Milroy L, Muysken P. One Speaker, two Languages: Cross-Disciplinary Perspectives on Code-Switching. Cambridge, UK: Cambridge University Press; 1995

[71] Bialystok E. The bilingual adaptation: How minds accommodate experience. Psychological Bulletin. 2017;**143**(3):233. DOI: 10.1037/bul0000099

[72] Chiao JY et al. Dynamic cultural influences on neural representations of the self. Journal of Cognitive Neuroscience. 2010;**22**(1):1-11. DOI: 10.1162/jocn.2009.21192

[73] Costa A, Sebastián-Gallés N. How does the bilingual experience sculpt the brain? Nature reviews. Neuroscience. 2014;**15**(5):336-345. DOI: 10.1038/nrn3709

[74] Bialystok E, Craik FI, Luk G. Bilingualism: Consequences for mind and brain. Trends in Cognitive Sciences. 2012;**16**(4):240-250. DOI: 10.1016/j.tics.2012.03.001

[75] Abutalebi J, Cappa SF, Perani D. The bilingual brain as revealed by functional neuroimaging. Bilingualism: Language and Cognition. 2001;**4**(02):179-190. DOI: 10.1017/S136672890100027X

[76] Crinion J, at al. Language control in the bilingual brain. Science. 2006;**312**(5779):1537-1540. DOI: 10.1126/science.1127761

[77] Perani D et al. The bilingual brain. Proficiency and age of acquisition of the second language. Brain. 1998;**121**(10):1841-1852

[78] Dwairy M, Achoui M, Abouserie R, Farah A. Adolescent-family connectedness among Arabs a second cross-regional research study. Journal of Cross-Cultural Psychology. 2006;**37**(3):248-261. DOI: 10.1177/0022022106286923

[79] Greenwald AG, McGhee DE, Schwartz JL. Measuring individual differences in implicit cognition: The implicit association test. Journal of Personality and Social Psychology. 1998;**74**(6):1464-1480. DOI: 10.1037/0022-3514.74.6.1464

[80] Yamaguchi S et al. Apparent universality of positive implicit self-esteem. Psychological Science. 2007;**18**(6):498-500. DOI: 10.1111/j.1467-9280.2007.01928.x

Section 4

Bilingualism, Information Retrieval and Access to Information

Chapter 7

Cross - Language Information Retrieval Using Two Methods: LSI via SDD and LSI via SVD

Daniel Enrique Tamara Lopez,
Estefania Julieth Guevara Blanquicett and
Carlos Daniel Acosta Medina

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74171

Abstract

This chapter presents a method for the recovery of bilingual information based on semidiscrete matrix decomposition (SDD); that is, the problem of retrieving information in two languages, Spanish and English, is studied when the queries are made only in Spanish. In it, four case studies that exhibit the performance of the use of the latent semantic index (LSI) via SDD method for cross-language information retrieval (CLIR) are displayed. Concurrently, these results are compared with those obtained by applying LSI via singular value decomposition (SVD). All experiments were performed from a bilingual database, built from the gospels of the Bible, which combines documents in Spanish and English. For this, a fusion strategy was used that increases the size of the database by 10%. It was found that in terms of errors, the methods are comparable, since equal results were obtained in 58.3% of the queries made. In addition, the methods presented a success rate of at least 65% in the task of retrieving relevant information in the two languages considered.

Keywords: information retrieval, latent semantic indexing, semidiscrete decomposition, singular value decomposition, cross-language

1. Introduction

The retrieval of information (IR) is focused on the problem of finding information that is relevant for a specific query. It is common that in many fields of research such as medicine, theology, international law, mathematics, among others, there is a need to retrieve relevant information from databases that have documents in multiple languages, which makes reference to cross-language information retrieval (CLIR), whose objective is to identify useful

information in the same and other languages than the language of the queries. For example, a user could ask a question in language X (source language) to find documents in languages X, Y, Z (target languages).

Many methods have been used in IR, among them the vector model, which interprets queries and documents as vectors and information retrieval, is based on operations among them. Latent semantic indexing (LSI) is an IR method based on the vector model, which replaces a term document matrix with a sum of matrices of a particular structure. In this sense, the singular value decomposition (SVD), QR and ULV factorizations, and the semidiscrete decomposition (SDD) have been used in LSI to IR. The SDD has shown benefits in saving storage of large databases, but has not been tested in CLIR. This chapter examines and evaluates a method for bilingual retrieving information (Spanish-English) based on semidiscrete decomposition (SDD), which retrieves relevant information in both languages when the query is made in Spanish. In addition, are presented four case studies that show the performance of the LSI via SDD method for CLIR and the results are compared with those obtained by applying the LSI via SVD method. To do this, a database was built combining documents (Bible Gospels) in Spanish and English.

2. Vector model

One of the most common methods in text mining for automatic indexing is the vector model. In it, every document and any need for information or query is encoded as a vector whose components reflect the importance of a particular term in its meaning or semantics.

2.1. Matrix term document

A database containing n documents described by m terms is represented as a matrix $A_{m\times n}$ called *matrix term document*, where the element a_{ij} denotes the *weight* of term i in document j. A natural choice for the components of vector document is a function of the frequency with which each term occurs in it, that is to say, $a_{ij} = f_{ij}$, where f_{ij} is the number of times the term i appears in the document j. There are more sophisticated schemes such as those given in [1, 2] that may lead to better results, and in general, as is done in [2], the procedure is to define the inputs from A to

$$a_{ij} = l_{ij} g_i d_j, \tag{1}$$

where l_{ij} is the *local weight* of term i in document j, g_i is the *overall weight* of term i in the collection of documents, and d_j is the *component standardization*, which specifies whether the columns of A (that is, the documents) are normalized or not. Local and global weights are applied to increase or decrease the importance of terms within or between documents. In [1], they explain the weight scheme for terms that are recommended to use depending on the characteristics of the document collection. In the language of the vector model, the columns

and rows of A are called document vectors and term vectors, respectively. Because each vector document contains only a small part of the totality of terms that describe the entire collection of documents, normally, the term document matrix is *sparse*; i.e., most of its entries are zero.

2.2. Latent semantic indexing

Latent semantic indexing (LSI) [3–5], also called latent semantic analysis (LSA) is an automatic indexing method based on the semantics of documents, which attempts to overcome the two main problems that have the traditional indexing schemes of lexical coincidences: *polysemy* and *synonymy*. The first has to do with a word that can have multiple meanings, and therefore, the words of a query may not coincide in meaning with those of the documents; the second means that several terms can have the same meaning and hence the words used in queries can match nonrelevant documents.

LSI is based on the assumption that there is some latent semantic structure underlying data that is corrupted by the variety of words used [4], but this semantic structure can be discovered and enhanced by approximating the matrix term document by a summation of matrices of particular structure, for example, by a low rank approximation obtained by some matrix decomposition.

2.3. Queries and measures of performance

In the vector model, the queries are also seen as vectors and then match a query q means finding in the column space of A (the subspace generated by the vectors documents) the documents a_j that are most similar to her in meaning. In [2], they explain that it is possible to associate a weight scheme with a query and have

$$q_k = l_k g_{k'} \tag{2}$$

where q_k is the kth input of q, and l_k and g_k are the components of local and global weight, respectively. The documents considered as relevant are those that are geometrically closer to the query according to some measure, and often the cosine of the angle between the query vector and each of the document vectors is used as a measure of similarity, so that the largest values correspond to the most relevant documents. Then, a_j is recovered if

$$\cos\left(\theta\left(q, a_j\right)\right) = \frac{q^t a_j}{\|q\|_2 \|a\|_j} > tol, \tag{3}$$

where tol is a predefined tolerance. Another commonly used measure of similarity is the *dot product* between query vector and each document vector that is computed as

$$p = q^t A, \tag{4}$$

where the ith entry of p represents the score of the document i. Thus, the documents can be organized from major to minor, by relevance to the consultation, according to their score.

In [6], they describe that a good process of matching queries is when the intersection between the set of documents retrieved and the set of relevant documents is as large as possible and the number of irrelevant documents recovered is small. In this way, to measure the performance of an information retrieval system, one must evaluate the ability of the system to retrieve relevant information (*recall*) and to reduce irrelevant information (*precision*).

Other measures frequently used to evaluate the quality of an IR system are the *pseudoprecision, average pseudoprecision,* and *mean average pseudoprecision (MAP).* Let r_i be the number of relevant documents retrieved up to position i in the sorted list of documents:

- The recall for the ith document from the list, R_i, is the ratio of relevant documents seen so far, that is, $R_i = r_i/r_n$, where r_n is the amount of relevant documents retrieved.
- The precision for the i-th document, P_i, is the proportion of documents up to position i that are relevant to a given query, that is, $P_i = r_i/i$.
- The *pseudoprecision* for a level of recall x, $\tilde{P}(x)$, is defined by

$$\tilde{P}(x) = \max P_i, \quad where\ x \leq \frac{r_i}{r_n}, i = 1, 2, \cdots, n. \tag{5}$$

- The *average pseudoprecision* for a single query is defined by

$$P_{av} = \frac{1}{n}\sum_{i=0}^{n-1} \tilde{P}\left(\frac{i}{n-1}\right). \tag{6}$$

- The *mean average pseudoprecision (MAP)*, used to evaluate yield in a set of queries, is defined by

$$MAP = \frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{n}\sum_{i=0}^{n-1} \tilde{P}\left(\frac{i}{n-1}\right)\right], \tag{7}$$

where M is the number of queries.

3. Latent semantic indexing via singular value decomposition

In this section, we initially present the singular value decomposition (SVD) and two theorems that show how the SVD gives useful information about the structure of a matrix. Also, it is explained why these theorems are important for IR and in particular for LSI. Subsequently, the LSI method based on the SVD is exposed.

3.1. Singular value decomposition

The SVD of a matrix $A_{m\times n}$, with $m \geq n$, is a factorization of the form

$$A = U\begin{pmatrix}\Sigma\\0\end{pmatrix}V^t, \tag{8}$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices whose columns are called, respectively, singular vectors to the left and to the right of A, and $\Sigma \in \mathbb{R}^{n\times n}$ is a diagonal matrix that contains the singular values $\sigma_1 \geq \sigma_2 \geq \sigma_n \geq, \cdots, \geq 0$ of A in decreasing order within its diagonal. This factorization exists for any matrix A, and numerical linear algebra texts commonly include it in their content [7, 8]. Methods to calculate the SVD of *dens*e and *sparse* matrices are well documented [1, 6, 7].

The following two theorems show how the SVD reveals important information about the structure of a matrix.

Theorem 1. Let $A_{m\times n}$, where without loss of generality $m \geq n$, $A = U\Sigma V^t$, the *SVD* of A and $\sigma_1 \geq \sigma_2 \geq, \cdots, \geq \sigma_r > \sigma_{r+1} =, \cdots, = 0$. If *R (A)* and *N (A)* denote the column space and null space of A, respectively, and if $U = [u_1\ u_2, \cdots, u_m]$ and $V = [v_1\ v_2, \cdots, v_n]$, then,

- *rank (A) = r*
- $R(A) = span\{u_1,\ u_2, \ldots, u_r\}$
- $N(A) = span\{v_{r+1},\ v_{r+2}, \ldots, v_n\}$
- $R\left(A^t\right) = span\{v_1,\ v_2, \ldots, v_r\}$
- $N\left(A^t\right) = span\{u_{r+1},\ u_{r+2}, \ldots, u_n\}$

$$A = \sum_{i=1}^{r} u_i\sigma_i v_i^t$$

Proof: See [6–8].

The theorem reveals that the SVD gives orthogonal bases for the four fundamental subspaces associated with a matrix and, in particular, in the context of term document matrices, the second part indicates that generating the semantic content of a database does not require to use all document vectors but a subset of the singular vectors to the left corresponding to the range of the matrix. The sum of the last part of the theorem is usually called *expansion in singular values* of A.

Theorem 2 (Eckart and Young). Suppose $A \in \mathbb{R}^{m\times n}$ has rank $r > k$. Then,

$$\min_{\text{rank}(B)} ||A - B||_f^2 = ||A - A_k||_f^2, \tag{9}$$

where

$$A_k = \sum_{i=1}^{k} u_i\sigma_i v_i^t := U_k\Sigma_k V_k^t. \tag{10}$$

Proof: See [6–8].

In this case, the theorem states that A_k is the matrix of rank k closest to A. The U_k columns live in the semantic space and are used to approximate the documents. As is known, truncated SVD is useful for "*eliminating noise*" present in an array and therefore, in the case of matrices representing a database, to remove term-document associations that are obscuring the real meaning of it.

3.2. LSI via SVD

As mentioned earlier, LSI is an IR method based on the vector model that approximates a document term matrix by a sum of the matrices of particular structure. In this regard, according to Theorem 2, LSI via SVD uses the singular value decomposition to obtain a k-rank approximation of the original document term matrix $A_{m\times n}$ in order to eliminate the noise present in it and project the m terms, n documents, and queries in a k-dimensional space, where $k \ll \min(m,n)$. It is important to keep in mind that document term matrices are commonly *well conditional*, that is, their singular values have no gaps and do not decay rapidly to zero, so a suitable k to truncate the SVD cannot be estimated [6]; experimentally, it has been concluded that for very large databases, k is taken between [100; 300] [3].

As in the vector model, in LSI, it is possible to match a query q through operations between vectors, namely, the cosine of the angle or the product point between the query vector and the document vectors. In this case, we calculate

$$p = \tilde{q}^t \tilde{A}, \tag{11}$$

where $\tilde{q} = U_k^t q$, $\tilde{A} = \Sigma_k V_k^t$ and $A_k = \mathrm{U_k}\Sigma_k V_k^t$ is the approximation of rank k of A obtained from the truncated SVD. Therefore, the recovered documents will be those corresponding to the largest components of p.

4. Latent semantic indexing via semidiscrete matrix decomposition

In this section, we present the semidiscrete decomposition (SDD) of a matrix and the method LSI via SDD. For more details, see [2, 9].

4.1. Semidiscrete decomposition

A semidiscrete decomposition (SDD) expresses a matrix as a weighted sum of outer products formed by vectors whose inputs are taken from the set $\mathrm{S} = \{-1, 0, 1\}$ that is given as

$$A_k = \underbrace{[x_1\ x_2 \cdots x_k]}_{X_k} \underbrace{\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_k \end{bmatrix}}_{D_k} \underbrace{\begin{bmatrix} y_1^t \\ y_2^t \\ \vdots \\ y_k^t \end{bmatrix}}_{Y_k^t} = \sum_{i=1}^{k} d_i x_i y_i^t, \tag{12}$$

where the $x_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}^n$ are formed by elements of the set $S = \{-1, 0, 1\}$, and d_i is a positive scalar, called the *i*-th SDD value. The matrix A_k is called semidiscrete decomposition of rank k (or SDD k-term). The algorithm that allows to calculate the SDD of a matrix and some of its properties, for example, its convergence, is described in [9].

4.2. LSI via SDD

The truncated SVD produces the best approximation of range k for a matrix; however, generally, even for a very low range approximation, more storage is required than the original matrix if it is *sparse*, that is, if the majority of its components are zero. As document term matrices that correspond to real databases are commonly large and sparse, using truncated SVD can be extremely expensive in terms of storage. It is for the above, that to save space (and consultation time) in [2], they propose the SDD as an alternative of SVD in LSI.

In this sense, LSI via SDD consists of replacing the term document matrix by an approximation that allows, as they sign in [10], to identify the clusters that form the documents present in databases and at the same time save a considerable amount of storage with respect to other factorizations. In [2], they show that for equal values of k, the SVD requires about 32 times more storage than the SDD. Specifically, LSI via SDD consists of approximating the document term matrix by a sum of exterior products of rank 1 such as in the SVD, but whose vectors consist only of elements of the set $S = \{-1, 0, 1\}$. For more details on LSI via SDD, see [2, 11].

To match the queries with the documents using LSI via SDD, we proceed in the same way as in LSI via SVD, that is, by calculating the product

$$p = \tilde{q}^t \tilde{A}, \tag{13}$$

where $\tilde{q} = X_k^t q$, and $\tilde{A} = D_k Y_k^t$. Relevant documents are those that correspond to the largest components of p.

5. The CLIR problem

Different documents can contain information that is conceptually the same without having to use similar words. People when they make a query in an IR system, for example, a search engine such as Google, do so by concept, and the words they use in it generally do not match those of the relevant documents. In this way, the main objective of CLIR, which is the retrieval of relevant information in the same and other languages to the queries, is highly affected because they must be compared terms of different languages.

To address this situation, databases have been created between languages, which are collections of documents that combine low percentages of languages and for its construction, it is necessary to take into account two concepts of close relationship with CLIR: *parallel aligned corpus* and *fusion strategies*.

5.1. Parallel aligned corpus

A *parallel text* is a text accompanied by its translations in other languages. Large collections of parallel text are called *parallel corpus*. In order to use a parallel corpus correctly, it is necessary to align the original text with its (your) translation (translations), that is, you must identify the phrases or words in the original text with their corresponding translations in the other languages. This is known as the *parallel aligned corpus*.

As stated by Kolda et al. in [12], perhaps the biggest decision to make when implementing LSI multilanguage is which parallel aligned corpus to use. In this work, we have adopted the Bible as ours and reasons for this are: (i) it is probably the most translated book in the whole world, which allows us to have many translations of the same documents, (ii) given its presentation by chapters and verses, its parallel alignment is facilitated, (iii) if we refer to the Gospels (Matthew, Mark, Luke, John), it is easy to identify facts related to the life of Jesus and thus recognize relevant documents for queries made in this context.

5.2. Fusion strategies

The central purpose in CLIR is to develop tools that allow the terms of query to coincide with those of documents that describe the same or similar meaning, even if they are in different languages [13]. The goal is the construction of parallel aligned corpus using the languages of the documents, which can be done, for the case of two languages, for example, taking portions of documents in a certain language and adding them to the corresponding documents of the other language. This is called a *fusion strategy*. Related works are [14, 15].

This work seeks to recover relevant documents in Spanish and English when queries are made in Spanish using fusion strategies, which combine approximately 10% of documents. The central idea behind each fusion used is to take a specific amount of verses in a certain language and add them to the corresponding verses of the other language.

6. Study cases

Four case studies are developed with the intention of evaluating the performance of two methods of LSI, LSI via SVD and LSI via SDD, applied in CLIR. The first one identifies the LSI model that allows obtaining the best results in terms of the mean average pseudoprecision (MAP). For this case, two English translations of the Gospels were used: *The King James* and *New Living Bible*. In the second case, we start with the model previously chosen to develop two experiments that involve two fusion strategies that combine small portions of the Gospels in the English-Spanish languages, using the *King James* and *Reina Valera 1966* versions. In Cases 3 and 4, computational comparisons are made between the LSI methods and their performance are analyzed when the collection of documents increases, respectively.

In all cases, the documents used to consist of a group of verses that form a story, verses which were taken from the *New International* version of the Bible, which organizes the verses by

stories and gives each one a title. The queries used are those given in **Table 1**, which describe parables and miracles in the life of Jesus. **Table 2** shows the biblical quotation where each of these queries is located, that is, the documents that are relevant.

6.1. Case 1: identification of the LSI model

An *LSI model* is the set of parameters that are considered in the application of the latent semantic index method, that is, the local and global weight schemes, the number of factors (k), the fusion strategies, etc., that are chosen for performing recovery experiments. A four-letter string has been used to differentiate LSI models. The first three indicate the local weight, the global weight and the use of standardization in the matrix term document, respectively, and the last corresponds to the query matrix and refers to the local weight of the terms. In this way, the nomenclature *fex.l*, for example, means that in the matrix term document, the frequency (f) for the local weight and an entropy value (e) (see [1]) for the global weight of the terms were used; besides, the columns of the matrix document term were not normalized (x) and a logarithmic value (l) was used for the local weight of the query terms.

Query		Query		Query	
1	El bautizo de Jesús	5	Niño epiléptico curado	9	Vino nuevo viejo odres
2	Impuesto al Cesar	6	La alimentación a cinco mil	10	El sembrador y la tierra
3	Limpieza al templo	7	La higuera maldita	11	Grano de mostaza
4	Entrada a Jerusalén	8	Tela nueva vestido Viejo	12	Higuera

Table 1. Queries used for the case studies.

Query	Matthew	Mark	Luke	John
1	Mt 3:13-17	Mc 1:9-11	Lc 3:21-23	Jn 1:29-39
2	Mt 22:15-22	Mc 12:13-17	Lc 20:20-26	
3	Mt 21:12-13	Mc 11:12-14		Jn 2:14-22
4	Mt 21:1-11	Mc 11:1-10	Lc 19:29-44	Jn 12:12-19
5	Mt 17:14-18	Mc 9:17-29	Lc 9:38-43	
6	Mt 14:15-21	Mc 6:35-44	Lc 9:12-17	Jn 6:5-13
7	Mt 21:18-22	Mc 11:12-14,20-25		
8	Mt 9:16	Mc 2:21	Lc 5:36	
9	Mt 9:17	Mc 2:22	Lc 5:37-38	
10	Mt 13:3-8,18-23	Mc 4:3-8,14-20	Lc 8:5-8,11-15	
11	Mt 13:31-32	Mc 4:30-32	Lc 13:18-19	
12	Mt 24:32-35	Mc 13:28-29	Lc 21:29-31	

Table 2. Location of queries in the gospels.

In this case, different LSI models are tested and the best one is determined from the MAP for $k = 100$. **Table 3** reports the results.

It is observed that with both methods, that is, LSI via SVD and LSI via SDD, the highest values of the MAP, marked in bold, were achieved with the models *len.f*, *len.b*, and *len.l*. This means that the *log-entropi* scheme is the one with the best performance and that the local weight of the terms in the query matrix does not affect the quality of a recovery. For this reason, in all subsequent experiments, only the *len.f* model will be used in both methods.

6.2. Case 2: fusion strategies

Two experiments are developed that involve merging the documents in English with their corresponding versions in Spanish. In each one, a different fusion strategy is used and 20 documents are retrieved by query. For each query in each experiment, an analysis of the selection of the k is made in order to establish a margin for the choice of the same. The errors obtained are illustrated in terms of the average of pseudoprecision and tables that give details of what was recovered in each query are shown. The errors were calculated with the formula

$$Error = 1 - \frac{1}{n}\sum_{i=0}^{n-1} \tilde{P}\left(\frac{i}{n-1}\right). \tag{14}$$

The database has 670 documents, of which, considering the 12 queries, 72 are relevant, that is, only 10.74% of the collection is relevant. The amount of storage required is 0.375 MB.

6.2.1. Experiment 1

The fusion strategy increases the size of the database by approximately 10% and consisted of *taking a single verse from the beginning of each document in Spanish and adding it to the end of the corresponding first verse in English*. **Table 4** illustrates the structure of the database and one of its documents.

In **Figure 1**, we show the graphs that relate the k to the error levels for each of the queries. It is observed that the error curves give clues for the selection of a k in almost all the queries. In the

LSI model	SVD	SDD	LSI model	SVD	SDD
cxn.f	0.6948	0.6440	**len.l**	**0.7591**	**0.7035**
fxn.f	0.6762	0.5920	lin.f	0.7375	0.6811
lxn.f	0.7444	0.6276	lin.b	0.7375	0.6811
lxn.b	0.7444	0.6276	lin.l	0.7375	0.6811
lxn.l	0.7444	0.6276	lpn.f	0.7058	0.6957
len.f	**0.7591**	**0.7035**	lpn.b	0.7058	0.6957
len.b	**0.7591**	**0.7035**	lpn.l	0.7058	0.6957

Table 3. MAP for different LSI models.

http://dx.doi.org/10.5772/intechopen.74171

Gospels	Doc. English	Doc. Spanish	Example of a database document
Matthew	Eng + Spa	Spanish	And seeing the multitudes, he went up into a mountain: and when he was set, his disciples came unto him: **Viendo la multitud, subio al monte; y sentandose, vinieron a el sus discipulos.** And he opened his mouth, and taught them, saying,
Mark	Eng + Spa	Spanish	
Luke	Eng + Spa	Spanish	
John	Eng + Spa	Spanish	

Table 4. Fusion scheme of the database (left) and example of a document of it (right).

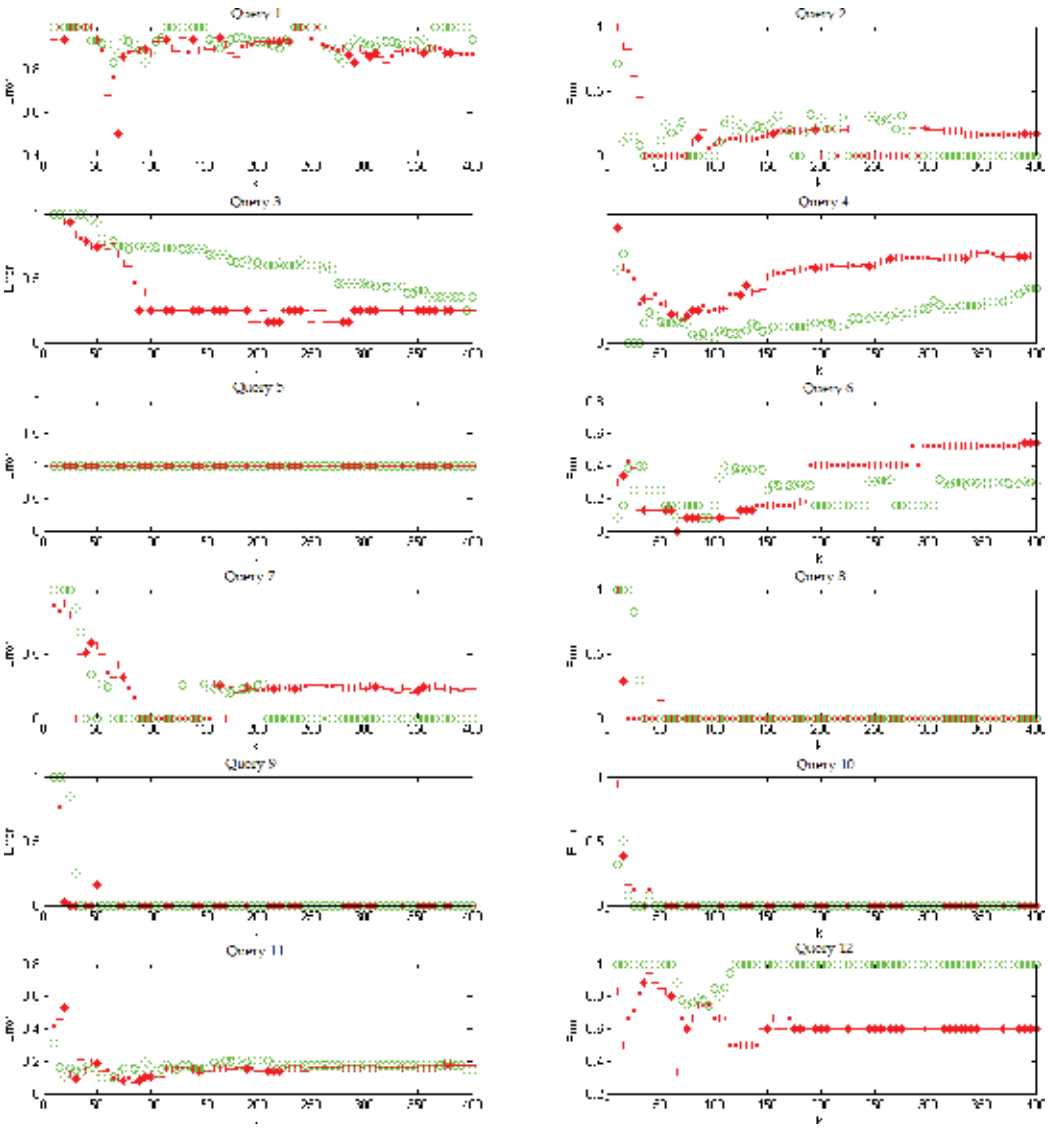

Figure 1. Errors versus k for each query. The asterisks and circles indicate the methods LSI via SVD and LSI via SDD, respectively.

first one, for example, it is observed that $k = 70$ would be the *optimum k* for LSI via SVD. In Query 5, the two methods completely failed with a 100% error; in Queries 8, 9, and 10, the errors are approximately zero in almost all values of k. There is good behavior of the methods in Queries 2, 4, 6, 7, 8, 9, 10, and 11 in many values of k, in particular, for some less than 110. In Queries 1 and 12, the best results were also in these values. It is also observed that there are usually many local minima, which makes it difficult to automate the choice of k through some parameter selection algorithm.

In **Table 5**, we show for each query two values of k and the corresponding errors. The first, called *optimal k*, indicates the smallest k for which the smallest error was obtained. The other represents the same but considers $k < 110$. It is noted that in all queries, except q3, the *optimum k* matches the *selected k*, which leads us to think about the possibility of reducing the domain of choice of k when considering the *k selected* in an interval of lower amplitude. **Table 6** shows analogous results to those in **Table 6** considering the values of the *k selected* in the interval [70, 100].

	SVD		SDD		SVD		SDD		SVD		SDD
q1	$^{k}opt\ Err$	70	65	**q5**	$^{k}opt\ Err$	10	10	**q9**	$^{k}opt\ Err$	25	35
	$^{k}sel\ Err$	0.5	0.833		$^{k}sel\ Err$	1	1		$^{k}sel\ Err$	0	0
		70	65			10	10			25	35
		0.5	0.833			1	1			0	0
q2	$^{k}opt\ Err$	35	35	**q6**	$^{k}opt\ Err$	65	10	**q10**	$^{k}opt\ Err$	30	25
	$^{k}sel\ Err$	0	0		$^{k}sel\ Err$	0	0.078		$^{k}sel\ Err$	0	0
		35	35			65	10			30	25
		0	0			0	0.078			0	0
q3	$^{k}opt\ Err$	195	395	**q7**	$^{k}opt\ Err$	30	40	**q11**	$^{k}opt\ Err$	85	55
	$^{k}sel\ Err$	0.157	0.25		$^{k}sel\ Err$	0	0		$^{k}sel\ Err$	0.065	0.097
		90	80			30	40			85	55
		0.25	0.727			0	0			0.065	0.097
q4	$^{k}opt\ Err$	70	20	**q8**	$^{k}opt\ Err$	20	35	**q12**	$^{k}opt\ Err$	65	95
	$^{k}sel\ Err$	0.173	0		$^{k}sel\ Err$	0	0		$^{k}sel\ Err$	0.33	0.74
		70	20			20	35			65	95
		0.173	0			0	0			0.33	0.74

Table 5. Fusion 1. Query errors for the optimal k and the selected k.

	SVD		SDD		SVD		SDD		SVD	SDD
q1	$^{k}sel\ Error$	70	95	**q5**	$^{k}sel\ Error$	70	70	**q9**	$^{k}sel\ Error$	70
		0.5	0.83			1	1			0.5
q2	$^{k}sel\ Error$	70	75	**q6**	$^{k}sel\ Error$	70	90	**q10**	$^{k}sel\ Error$	70
		0	0			0.07	0.07			0
q3	$^{k}sel\ Error$	90	80	**q7**	$^{k}sel\ Error$	90	70	**q11**	$^{k}sel\ Error$	90
		0.25	0.72			0	0			0.25
q4	$^{k}sel\ Error$	70	100	**q8**	$^{k}sel\ Error$	70	70	**q12**	$^{k}sel\ Error$	70
		0.17	0.03			0	0			0.17

Table 6. Fusion 1. Errors per query for the selected k in [70, 100].

Here, it is appreciated that in all queries, except for q3 and q12, the *selected k* increased; however, the errors in q1, q2, q5, q6, q7, q8, q9, and q10 remained the same. In q4 and q11, the errors increased from 0 to 3% and from 9 to 12%, respectively.

From the above, it is concluded that the performance of the LSI methods subtly deteriorated when considering k in such interval, since in 10 of the 12 queries, the errors were maintained and in only two they increased in small percentages. The main contribution of these tables is to have identified a small range for the choice of the parameter k.

6.2.2. Experiment 2

The fusion strategy used in this case also combines a single verse, that is, 10% of the documents, but unlike the previous experiment, it takes verses in English and adds them to the corresponding verses in Spanish and vice versa. The structure of the database is illustrated in **Table 7**.

Figure 2 illustrates the error curves for each query by increasing the parameter k.

Again it is observed that at q5, an error of 100% was obtained. In q2, q6, q7, q8, q9, and q10, the methods reached errors close to zero in some values of k. In q1 and q11, only LSI via SDD obtained errors close to that value. Again, the existence of many local minimums in the error levels of each query is highlighted. Information on the *optimal k*, the *selected k*, the *k selected in the interval* [70, 100] (denoted by k_{sel1} and k_{sel2}, respectively) and the corresponding errors is given in **Table 8**.

We find that in q2, q5, q6, q7, q8, q9, q10, q11, and q12, the errors for k_{opt}, k_{sel1}, and k_{sel2} did not change when using LSI via SVD, that is, for this group of nine queries, the *optimal k* lies in the interval [70, 100]. With LSI via SDD, for this same group of queries, except for q2 and q11, the k_{opt} is also obtained in that interval; in q3, the errors increased when the selection interval of

		Chapters	
Matthew	Eng + Spa	1–14	Spanish
	English	15–28	Spa + Eng
Mark	Eng + Spa	1–8	Spanish
	English	9–16	Spa + Eng
Luke	Eng + Spa	1–12	Spanish
	English	13–24	Spa + Eng
John	Eng + Spa	1–10	Spanish
	English	11–21	Spa + Eng

Table 7. Scheme of the database for the structure of the merger 2.

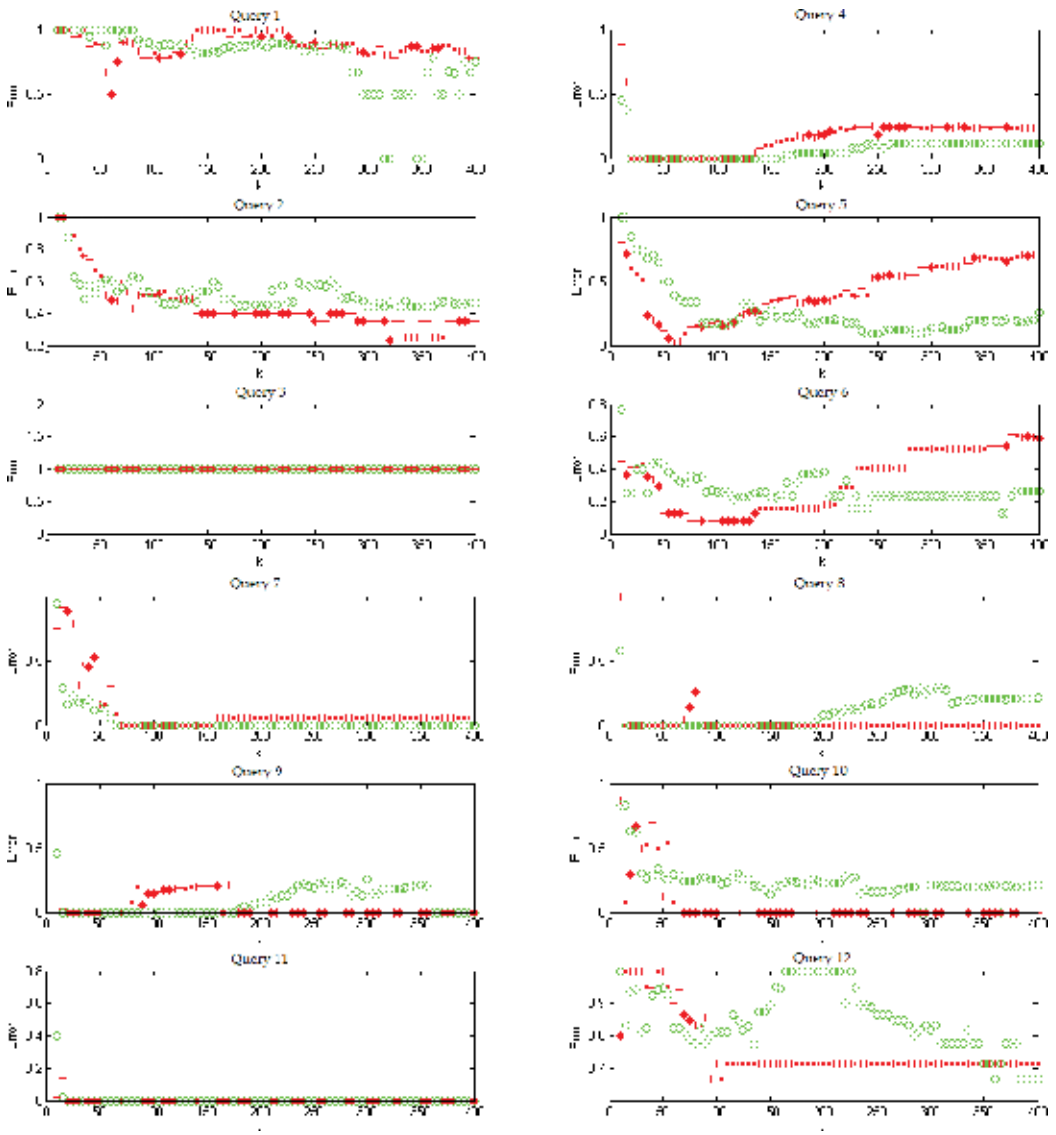

Figure 2. Errors versus *k* for each query. The asterisks and circles indicate the methods LSI via SVD and LSI via SDD, respectively.

the *k* was reduced; in q4, for the k_{sel1} and the k_{sel2}, the errors were equal but superior to the corresponding ones of the k_{opt}.

In this way, considering the two experiments, it is concluded that the results in terms of the Eq. (14) to calculate the errors in the recoveries favor the *Fusion 1* since when considering *k* in the interval [70, 100], LSI methods obtained minor errors compared to those found with *Fusion 2*. Therefore, in the following cases, only *Fusion 1* will be used in order to continue with the comparison of LSI methods.

	SVD		SDD		SVD		SDD		SVD		SDD
	^{k}opt	55	355		^{k}opt	10	10		^{k}opt	25	15
	Err	0.75	0		*Err*	1	1		*Err*	0	0
q1	$^{k}sel1$ *Err*1	55 0.75	65 0.87	**q5**	$^{k}sel1$ *Err*1	10 1	10 1	**q9**	$^{k}sel1$ *Err*1	25 0	15 0
	$^{k}sel2$	70	70		$^{k}sel2$	70	70		$^{k}sel2$	70	70
	*Err*2	0.83	0.89		*Err*2	1	1		*Err*2	0	0
	^{k}opt	35	15		^{k}opt	60	85		^{k}opt	25	30
	Err	0	0		*Err*	0	0		*Err*	0	0
q2	$^{k}sel1$ *Err*1	35 0	15 0	**q6**	$^{k}sel1$ *Err*1	60 0	85 0	**q10**	$^{k}sel1$ *Err*1	25 0	30 0
	$^{k}sel2$	70	80		$^{k}sel2$	70	85		$^{k}sel2$	70	70
	*Err*2	0	0.16		*Err*2	0	0		*Err*2	0	0
	^{k}opt	130	150		^{k}opt	90	40		^{k}opt	75	20
	Err	0.33	0.33		*Err*	0	0		*Err*	0.06	0
q3	$^{k}sel1$ *Err*1	110 0.49	105 0.55	**q7**	$^{k}sel1$ *Err*1	90 0	40 0	**q11**	$^{k}sel1$ *Err*1	75 0.06	20 0
	$^{k}sel2$	100	75		$^{k}sel2$	90	70		$^{k}sel2$	75	85
	*Err*2	0.56	0.61		*Err*2	0	0		*Err*2	0.06	0.25
	^{k}opt	115	150		^{k}opt	25	15		^{k}opt	25	90
	Err	0.3	0.16		*Err*	0	0		*Err*	0.33	0.5
q4	$^{k}sel1$ *Err*1	80 0.34	70 0.21	**q8**	$^{k}sel1$ *Err*1	25 0	15 0	**q12**	$^{k}sel1$ *Err*1	25 0.33	90 0.5
	$^{k}sel2$	80	70		$^{k}sel2$	70	70		$^{k}sel2$	75	90
	*Err*2	0.34	0.21		*Err*2	0	0		*Err*2	0.33	0.5

Table 8. Fusion 2. Errors per query for the optimal *k*, the selected *k*, and the selected *k* in [70, 100].

6.3. Case 3: computational comparison of LSI models

The results shown in the experiments of the second case study do not consider the efficiency of the IR systems, that is, the time of the LSI methods, the amount of storage required by each of them, the ability to quickly obtain relevant documents, and the relationship between these aspects. For this reason, in this case, computational results are presented that allow the LSI methods to be compared in such aspects. All tests were performed on a computer with *Intel (R) Core (TM) I5–3230 CPU @ 2.60 Hz* and with 6 GB of *RAM*. In **Figures 3** and **4**, the results obtained by SVD and SDD have been marked with an asterisk (*) and a circle (°), respectively.

Figure 3 illustrates the size in *megabytes* (MB) of the SVD and SDD decompositions for various values of *k*. Clearly, it is observed that in all the *k*, the SDD consumes much less space than the SVD. For *k = 400*, for example, the SVD occupies a space of 22.325 MB, while the SDD 0.875 MB, that is, there is a saving of 21.45 MB. In addition, in the lower part, the time used,

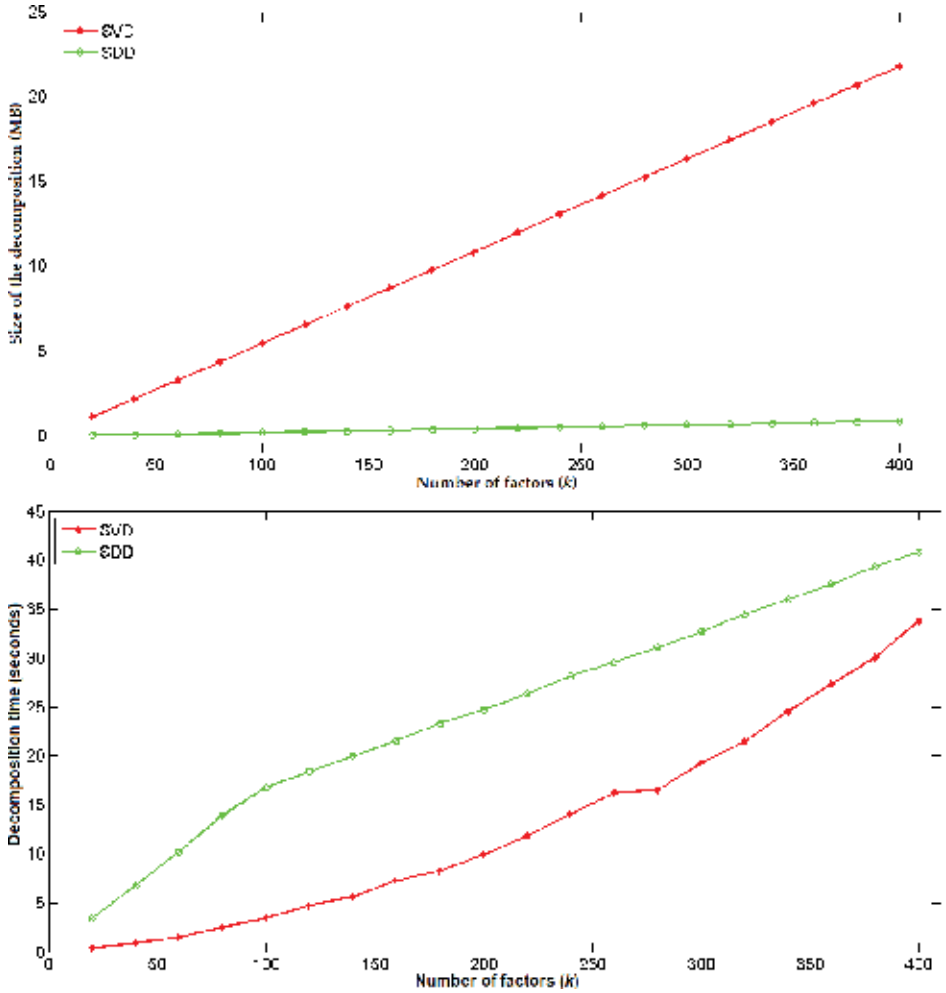

Figure 3. Size in megabytes (above) and execution time in seconds of the decompositions (below), as a function of the number of factors (*k*).

in seconds, to obtain each decomposition is shown. For the presented *k*, it is evident that in the SVD, less time is used. However, the amount of seconds used by each algorithm to build the matrices of the SVD and SDD factorizations is small, because even for high values of *k*, the recorded time is approximately 40 seconds.

Finally, in **Figure 4**, the MAP is presented as a function of the time of the LSI method, calculated using the formula *Time LSI methods = Decomposition Time + Query time*, and the amount of storage required by each one. In the graph, on the left, there are 20 asterisks corresponding to the values $k = 20, 40, \ldots, 400$ and 20 circles related to the same values of *k*. The second asterisk (corresponding to $k = 40$), for example, means that LSI via SVD required approximately 1.06 seconds to reach a quality of 0.5850, while the second circle shows that LSI via SDD at 9.21 seconds reached a MAP of 0.6062.

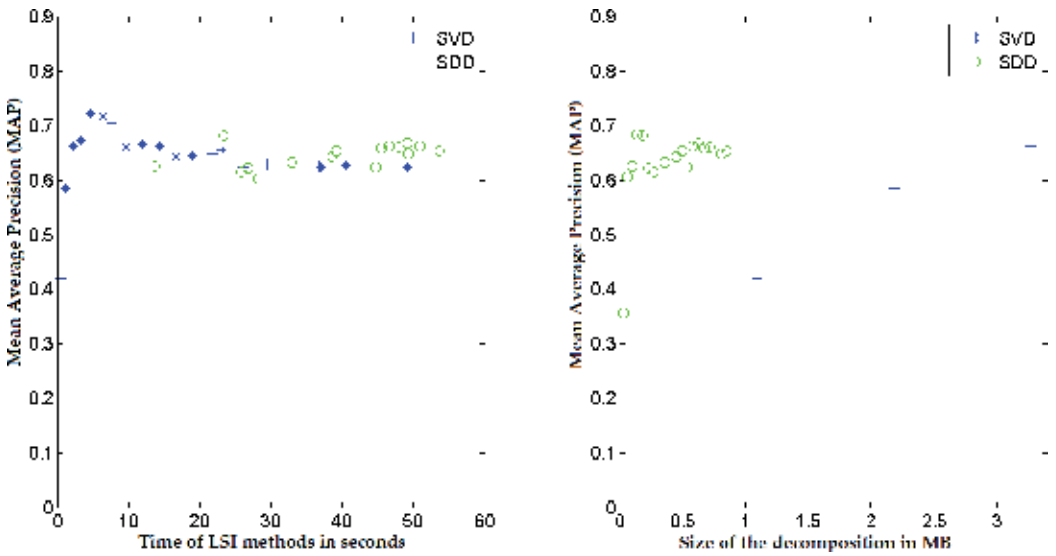

Figure 4. MAP versus time of LSI methods (left) and size of decompositions (right).

It is observed that the SVD-based method reaches its highest score, 0.7227, in $k = 100$ (at 4.7 seconds), and that the other one does it in $k = 80$ (at 18.9 seconds) with a value of 0.6838. It should be noted that the qualities of the methods, from $k = 40$ for LSI via SDD, were very close. On the right side, the size of the decompositions is crossed with the MAP. The last circle (for $k = 400$) means that a quality of 0.6539 was reached with just 0.85 MB; in turn, the first asterisk illustrates that with 1.09 MB, there was a MAP of only 0.4196. It is also observed that there are only three asterisks for SVD, and it is because the rest surpasses the scope of the graph. Likewise, it is highlighted that the highest score for the SDD-based method required only 0.14 MB of storage (approximately one-third of the weight of the matrix term document), while LSI via SVD required 5.6 MB of storage (approximately 15 times the weight of the matrix term document) to achieve its best performance. This time LSI via SDD widely outperformed the other method.

6.4. Case 4: adding documents to the data base

So far, we have only studied the LSI methods when you have a fixed document collection. In practice, it often happens that these collections are dynamic, that is, that new documents are added or that some existing ones are deleted. In this case study, the performance of the LSI methods is analyzed when different amounts of documents are added to the database. For this, the average pseudoprecision (see Eq. (9)) is used as a measure of quality to make an analysis by query and the MAP to generalize the study to all of them.

Specifically, 20 and 88 documents were added to the initial collection of 670 documents in order to obtain two new databases with 690 and 758 documents, which have 75 and 78 relevant documents, respectively. For these three collections, the results of **Figure 5** and **Table 9** are presented.

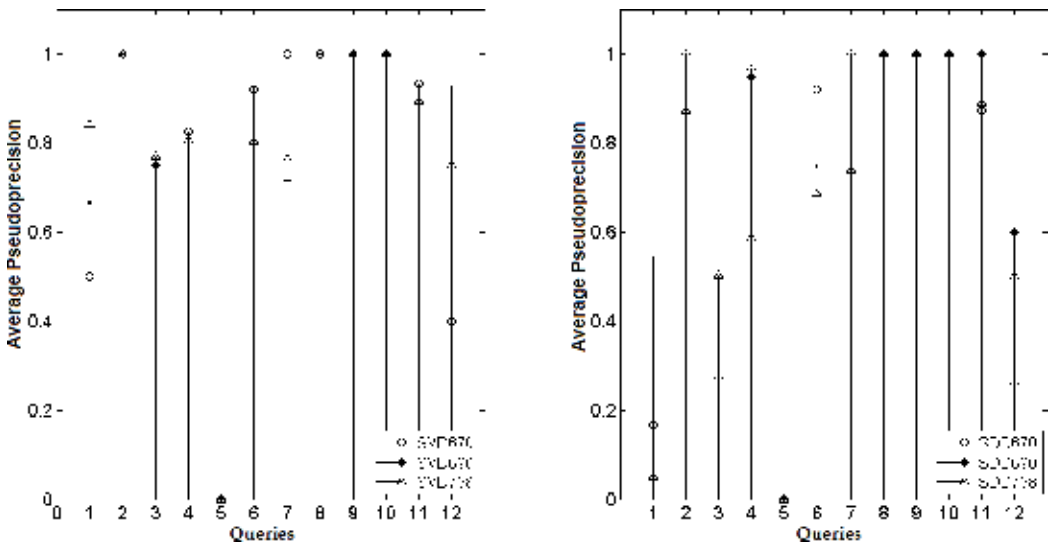

Figure 5. Comparison of the LSI methods with respect to the average pseudoprecision when adding documents to the database.

Documents			**MAP**	
Existing	Additions	% de Doc. relevant	SVD	SDD
670	—	10.74	0.7776	0.7046
670	20	10.86	0.8059	0.7428
670	88	10.29	0.8022	0.6523

Table 9. Percentage of relevant documents and MAP for the different databases.

In this figure, the average pseudoprecision obtained with the *k* in **Table 7** is shown for each of the 12 queries and for each database. The results of the LSI via SVD method are shown on the left and on the right are those corresponding to LSI via SDD. Methods for collections with 670, 690, and 758 documents have been labeled with a circle, an asterisk, and a triangle, respectively. It is observed that in the two methods, the average of pseudoprecision for Query 5 is 0 and for Queries 8, 9, and 10, it is 1. In Query 2, LSI via SVD also had a score of 1, while LSI via SDD did so only for 670 documents. In the rest of the queries, there are averages that go up or down as documents are added to the database. In Query 12, for example, it is noted that the highest score is for 690 documents, decreases when the collection is increased to 758 and decreases again when there are barely 670 documents. From this, it is concluded that there is no direct or inverse relationship between the average pseudoprecision and the number of documents in the database.

On the other hand, in order to evaluate the performance of the methods considering all the queries, in **Table 9**, the MAP obtained in each database is presented when again using the *k* of the **Table 6**.

http://dx.doi.org/10.5772/intechopen.74171

As a first observation, it is highlighted that MAP levels are higher, in all databases, when LSI is used via SVD. However, the six scores shown give evidence of the good performance of the two methods considering the low percentages of relevant documents in each collection, since all the success rate exceeds 65%. On the other hand, it is emphasized that there seems to be a direct relationship between the percentage of relevant documents and MAP values with the SDD-based method, that is, the higher the percentage of relevant documents, the greater the MAP.

7. Conclusions

The LSI method originally used the singular values decomposition (SVD) for the benefits that it has in terms of data representation in spaces of reduced dimension and other properties with respect to data filtering. This makes the SVD a powerful tool in IR and in CLIR. The semidiscrete decomposition (SDD), of which few investigations have been developed, has been successfully used in IR, and this research has shown that it is also useful in CLIR and that it is also comparable with the standard approach used by the SVD. Evidence of this is that

- In Case 2 for Fusion 1, the errors for the k selected in the interval [70, 100] were the same in 7 of the 12 queries and in the rest, they differ at most by 47%.
- When in Case 3 efficiencies were evaluated, in aspects, particularly one method surpassed the other. Specifically, when the MAP measure was related to the time of the methods, LSI via SVD was imposed because it requires fewer seconds to reach its highest performance; in contrast, when analyzing the MAP and the amount of storage, LSI via SDD showed a significantly higher performance. In this aspect, it is emphasized that with SDD, only one-third of the weight of the original matrix was needed to reach its highest performance; with SVD, on the other hand, it required almost 15 times the weight of the matrix term document to achieve such value. This is the true impact of the SDD, the ability to obtain good results at a very low cost in terms of storage.
- The MAP quality measure evaluates the performance of an IR method considering a set of queries, so that the higher this value, the better the method's performance will have been. In the fourth case study, when the performance of the methods was considered by increasing the number of documents in the database, higher performance was obtained when using the SVD since higher MAP values were found. However, with both methods, satisfactory results were obtained, because when conducting a search in Spanish, you can retrieve relevant documents in this language and in English with a success rate of at least 65%.

Therefore, it is concluded that although the LSI via SVD method has been widely used and is a powerful tool in CLIR, the LSI via SDD method results in an important and innovative alternative in information recovery tasks, since, in addition to achieving results comparable to those of the other method in the task of retrieving relevant information in multiple languages after consulting only one, and also has the benefit of saving large amounts of space when huge databases are stored.

Author details

Daniel Enrique Tamara Lopez*, Estefania Julieth Guevara Blanquicett and Carlos Daniel Acosta Medina

*Address all correspondence to: detamaral@unal.edu.co

Universidad Nacional de Colombia, Manizales, Colombia

References

[1] Berry M, Browne M. Understanding Search Engines: Mathematical Modeling and Text Retrieval. 2nd ed. Knoxville, Tennessee: SIAM; 2005. p. 135. DOI: 10.1137/1.9780898718164

[2] Kolda T, O' Leary D. A semidiscrete matrix decomposition for latent semantic indexing in information retrieval. ACM Transactions on Information Systems (TOIS). 1998;(4):322-346. DOI: 10.1145/291128.291131

[3] Berry M, Drmac Z, Jessup E. Matrices, vector spaces, and information retrieval. SIAM Review. 1999;**41**(2):335-362. DOI: 10.1137/S0036144598347035

[4] Berry M, Dumais S, O'brien G. Using linear algebra for intelligent information retrieval. SIAM Review. 1995;**37**(4):573-595. DOI: 10.1137/1037127

[5] Deerwester S, Dumais S, Landauer T, Furnas G, Harshman R. Indexing by latent semantic analysis. Journal of the Association for Information Science and Technology. 1990;**41**(6): 391-407. DOI: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

[6] Lars E. Matrix methods in data mining and pattern recognition. Linköping, Sweden: SIAM; 2007. p. 234. DOI: 10.1137/1.9780898718867

[7] Golub G, Van Loan C. Matrix Computations. 4th ed. Baltimore: Johns Hopkins University Press; 2012. 780 p. DOI: 10.1137/1028073

[8] Ipsen I. Numerical matrix analysis: Linear systems and least squares. Raleigh, North Carolina: SIAM; 2009. p. 140. DOI: 10.1137/1.9780898717686

[9] Kolda T, O' Leary D. Algorithm 805:Computation and uses of the semidiscrete matrix decomposition. ACM Transactions on Mathematical Software (TOMS). 2000;**26**(3):415-435. DOI: 10.1.1.41.7987

[10] McConnell S, Skillicorn D. Outlier detection using semi-discrete decomposition. Department of Computing and Information Science. Kingston, Ontario, Canada: Queen's University; 2001. DOI:10.1.1.8.7634

[11] Kolda T, O'Leary D. Latent semantic indexing via a semi-discrete matrix decomposition. In: Cybenco G, O'Leary D, Rissanen J, editors. The Mathematics of Information Coding,

Extraction and Distribution. College Park, USA: Springer, University of Maryland; 1999. p. 73-80. DOI:10.1007/978-1-4612-1524-0_5

[12] Kolda T, Chew P, Abdelali A, Bader B. Crosslanguage information retrieval using PARAFAC2. In: Knowledge Discovery And Data Mining. Albuquerque, NM: ACM; 2007. p. 143-152

[13] Jian-Yun N. Cross-language information retrieval. Morgan & Claypool: Quebec, Canadá; 2010. p. 125. DOI: 10.2200/S00266ED1V01Y201005HLT008

[14] Berry M, Young P. Using latent semantic indexing for multilanguage information retrieval. Computers and the Humanities. 1995;**29**(6):413-429. DOI: 10.1007/BF01829874

[15] Dumais S, Letsche T, Littman M, Landauer T. Automatic cross-language information retrieval using latent semantic indexing. AAAI spring symposium on cross-language text and speech retrieval. 1997;**15**:21. DOI:10.1.1.533.9602

Chapter 8

Cross-Lingual and Cross-Chronological Information Access to Multilingual Historical Documents

Biligsaikhan Batjargal

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72421

Abstract

In this chapter, we present our work in realizing information access across different languages and periods. Nowadays, digital collections of historical documents have to handle materials written in many different languages in different time periods. Even in a particular language, there are significant differences over time in terms of grammar, vocabulary and script. Our goal is to develop a method to access digital collections in a wide range of periods from ancient to modern. We introduce an information extraction method for digitized ancient Mongolian historical manuscripts for reducing labour-intensive analysis. The proposed method performs computerized analysis on Mongolian historical documents. Named entities such as personal names and place names are extracted by employing support vector machine. The extracted named entities are utilized to create a digital edition that reflects an ancient Mongolian historical manuscript written in traditional Mongolian script. The Text Encoding Initiative guidelines are adopted to encode the named entities, transcriptions and interpretations of ancient words. A web-based prototype system is developed for utilizing digital editions of ancient Mongolian historical manuscripts as scholarly tools. The proposed prototype has the capability to display and search traditional Mongolian text and its transliteration in Latin letters along with the highlighted named entities and the scanned images of the source manuscript.

Keywords: historical documents, multilingual databases, information access, information retrieval, digital edition

1. Introduction

As historical materials are increasingly being digitally preserved, multilingual materials concerning a diversity of languages and historical periods have been made available to the public

on the Internet. Recently, a number of large-scale digital library projects have been launched, e.g., Europeana, World Digital Library, HathiTrust and Google Book Search. These websites make multilingual materials covering various languages and historical periods available to the public.

There are various technical challenges, however, in implementing universal integrated access to these digital collections due to this great diversity, and difficulties occur in accessing these information sources, mainly due to the diversity of languages. Even within the same language, considerable differences exist in grammar, vocabulary and script depending on the historical period, and this is the primary cause of the difficulties in implementing universal information access. Thus, this chapter presents our approach to providing cross-lingual and cross-chronological access to historical documents that account for evolution of languages over periods ranging from ancient to modern. Particularly, in this chapter, we introduce our approach in providing cross-lingual and cross-chronological information access to historical materials in a less-researched language such as ancient Mongolian.

In Section 2, we discuss the current situation of digitized ancient historical materials written in ancient Mongolian and the challenges in providing universal information access to them in the digital era. Then, our proposed method for cross-lingual and cross-chronological information access to ancient Mongolian historical materials is discussed in Section 3. Finally, in Section 4, we discuss the future prospects of this research.

2. Ancient Mongolian manuscripts

This section briefly explains certain characteristics of Mongolian manuscripts and current situation of digitized ancient historical materials written in ancient Mongolian and challenges they present in the digital era.

2.1. A brief introduction of Mongolian manuscripts

Mongolian historical documents have been written in numerous scripts, i.e., the traditional Mongolian script, Square or Phags-pa script, Soyombo script and Horizontal square script [1]. Among them, the traditional Mongolian script is the most popular and longest-surviving script for over 800 years and has better supports with the computer systems recently since its integration to the Unicode Standard [2] in September 1999. On the 20th of June, 2017, the Soyombo and Horizontal square scripts (a.k.a. Zanabazar scripts) were standardized in the most recent version of the Unicode Standard [3]. However, this research focuses on the traditional Mongolian script because of its popularity, availability of digital texts and improved supports at the computers.

In 1946, Mongolia has made language reforms to eliminate a difference between written and spoken Mongolian language, and the Cyrillic script was adapted to Mongolian. The spelling of modern Mongolian in the Cyrillic alphabet was based on the pronunciations in the Khalkha dialect, the largest Mongol ethnic group [4, 5]. Such a radical change separated the Mongolian people from their historical archives written in traditional Mongolian script. Manuscripts in traditional Mongolian script preserve the ancient writing, while modern Mongolian reflects

http://dx.doi.org/10.5772/intechopen.72421

the unique pronunciations in modern dialects. Understanding historical documents in traditional Mongolian script is becoming as equally important a consideration for Mongolians as modern Mongolian in Cyrillic script. However, reading traditional Mongolian documents by using literacy in modern Mongolian is not a simple task. Traditional Mongolian is a distinct dialect with grammar different from that of modern Mongolian. The traditional Mongolian script is written vertically, from top to bottom, in columns advancing from left to right. This script has four derivative scripts: Todo or Clear, Manchu, Vaghintara and Sibe (Xibe) script. The Todo script was used by the Oirats and Kalmyks, and the Manchu script was a writing system in the Qing dynasty. The Sibe script is used in Xinjiang, in the northwest of China. The Vaghintara script was used by the Buryats.

Moreover, the circumstances that the manuscript passed through a process of copying or reprinting with possible alterations, corrections and unintended errors makes researchers wonder which ancient spelling is correct or what the ancient word originally meant. Scholars had been pointing out from time to time that copies could not meet the requirements of scholars who want to study them as a source material [6]. Moreover, various different commentaries, transcriptions, annotations and interpretations have been suggested by humanities researchers. Besides, manuscripts are vulnerable to degradations and might have lacunas, physical damages or missing parts, which require costly reconstructions of the original text.

In general, there are two main demands from both users and researchers for making ancient Mongolian manuscripts usable in this digital era. Firstly, a digital representation that explains a given manuscript in a modern language is helpful for users who want to read, search and browse ancient Mongolian manuscripts. Secondly, in the field of humanities, getting knowledge by analysing various historical documents is an important task. There are increasing demands from Mongolian humanities researchers to perform text analysis at massive scale with prompt and accurate results. Having a digital representation that fully reflects a given manuscript is an awaited demand for researchers who want to study it as a scholarly source using a computer.

Nevertheless, computerized text analysis of Mongolian historical documents has not been done due to the lack of natural language processing (NLP) tools that can handle ancient Mongolian. Such demands have encouraged us to introduce our approaches in providing universal information access to ancient Mongolian historical documents.

2.2. Ancient Mongolian manuscripts in the digital age

To the best of our knowledge, there are a small number of digital texts of ancient Mongolian manuscripts. A few ancient Mongolian historical manuscripts including (1) 'Qad-un ündüsün-ü quriyangγui altan tobči neretü sudur' (the Altan Tobchi or the Golden Summary: Short history of the Origins of the Khans) (written in 1604) a.k.a. 'Little' Altan Tobchi and (2) the 'Asaraγči neretü-yin teüke' or 'Asragch nėrtĭĭn tüükh' (the Story of Asragch) (written in 1677), which were written in traditional Mongolian script, have been converted to digital texts and made publicly available through the traditional Mongolian script digital library (TMSDL) [7]. **Figure 1** shows a folio of the 'Little' Altan Tobchi in the TMSDL with keywords' highlights.

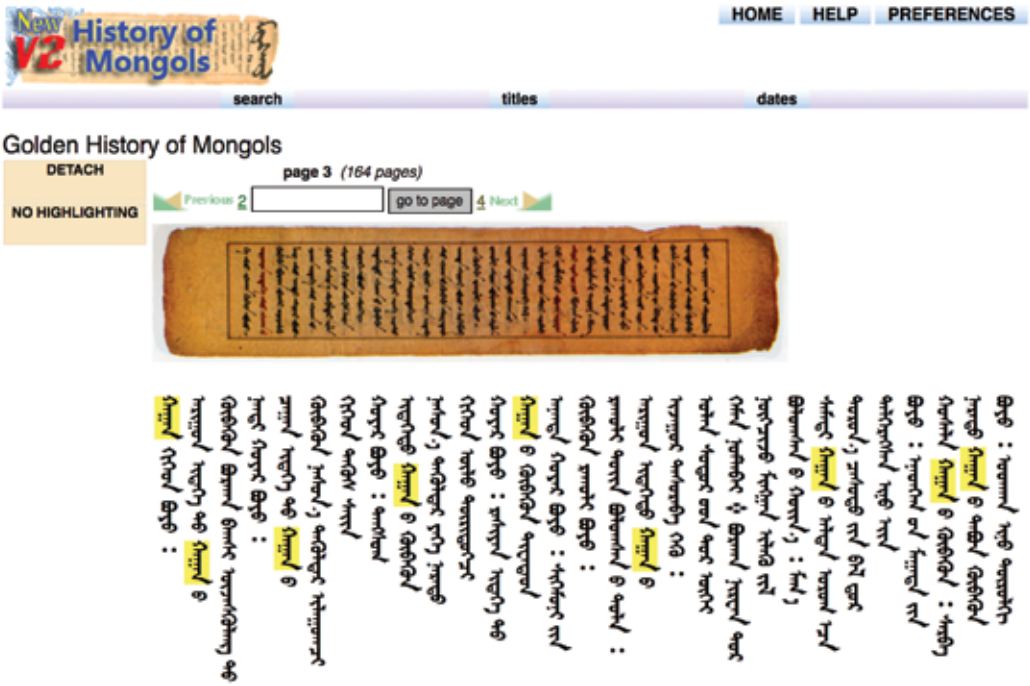

Figure 1. A folio of the 'little' Altan Tobchi in the TMSDL with keywords' highlights.

TMSDL can be used to access and retrieve the historical manuscripts written in traditional Mongolian script using a query in modern Mongolian (Cyrillic). The research achievements, as well as the experiences obtained from the development of the TMSDL, have motivated us to share further research results in developing methods to providing cross-lingual and cross-chronological information access to ancient Mongolian historical documents.

Certainly, there has been a little research on text mining for Mongolian language, and none of the research has considered text mining on ancient Mongolian historical documents due to the lack of research in those areas. Because of the notable difference between mediaeval Mongolian and modern Mongolian, the existing NLP tools, which were designed on modern Mongolian, do not perform well on ancient Mongolian texts. Therefore, further computerized analyses of ancient Mongolian historical documents are necessary.

3. Information access to Mongolian historical documents

In the recent years, the needs for utilizing digital representations and proving access to historical documents encouraged the development of various tools for transcribing, annotating and publishing of historical manuscripts. In order to provide computer technology-driven solutions to

http://dx.doi.org/10.5772/intechopen.72421

solve the facing challenges of Mongolian humanities scholarship as well as to benefit the recent achievements in the digital humanities worldwide, it is necessary to analyse the requirements of Mongolian historical documents for digital tools.

In this section, we describe our methods for implementing integrated access to historical documents that are capable of coping with linguistic transformations from ancient times to the present. First, we propose an information extraction method for digitized ancient Mongolian historical documents. The proposed method extracts named entities from historical manuscripts by utilizing machine learning techniques. Results will be utilized for building digital text representations that encode named entities, the possible alterations, corrections, errors and interpretations of ancient Mongolian words in a modern language. In the later sections, we discuss how to develop a digital edition of Mongolian historical documents by considering various features and requirements of Mongolian manuscripts.

3.1. Information extraction from ancient Mongolian documents

This section discusses an information extraction method for digitized ancient Mongolian documents by using the features of traditional Mongolian script. Named entities such as personal names and place names are extracted automatically from digitized text of ancient Mongolian documents by employing support vector machine (SVM) for aiming to reduce the labour-intensive analysis on historical text. Information extraction, named entity extraction (NEE) and tagging or annotations are able to turn plain text into structured data for analysis or effective use, via NLP applications and analytical methods. State-of-the-art NEE systems for English produce near-human performance to extract named entities [8]. However, there has been little research on text mining or NEE for Mongolian language, and none of the research has considered text mining on ancient Mongolian historical documents due to the lack of research in those areas. Therefore, proposing an information extraction method for ancient historical documents in traditional Mongolian script is crucial.

3.1.1. The proposed approach

The flowchart in **Figure 2** shows an overview of the main steps and components of the proposed approach. The proposed approach starts with preprocessing tasks where an ancient Mongolian corpus gets tokenized, each token gets annotated and gold standard annotations are prepared for inputting into SVM for learning. The proposed method learns the extraction rules of personal names from annotated training corpora and then extracts personal names from ancient Mongolian texts by using SVM. The following sections explain the main three components: (1) pre-processing, (2) annotating and (3) named entity extraction.

3.1.1.1. Preprocessing step

The first step is to divide digitized ancient Mongolian plain text of into tokens. This is necessary because we want to mark up each token in the next tasks. A token is quite often a word delimited by space, but there exist some unique features for traditional Mongolian script. For

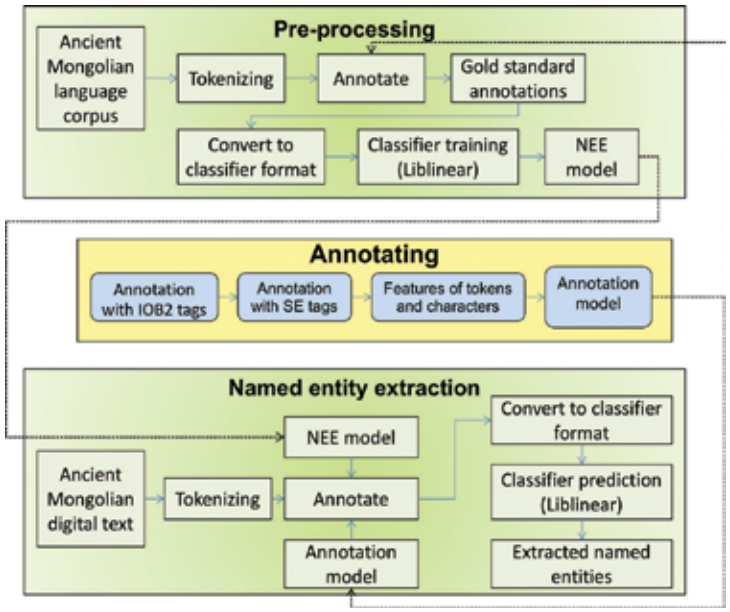

Figure 2. Overview of the main steps and components of the proposed approach.

instance, in traditional Mongolian script, certain words with a final vowel letter 'a' or 'e' are separated visually from the preceding consonant by a narrow gap. Moreover, some suffixes are visually separated from the stem of a word or from other suffixes. However, the 'a' or 'e' is an integral part of the word stem, as well as any attached suffixes are considered to be an integral part of the word as a whole. In Unicode, control characters Mongolian Vowel Separator (MVS) and narrow no-break space (NNBSP) handle the behaviour of Mongolian suffixes and vowels 'a'/'e' in the end of a word [2]. This information can be used as a feature in SVM. Other features are discussed in Section 3.1.1.3.

The next step is to annotate tokens and prepare gold standard annotations. Because of the lack of NLP tools and part of speech data for ancient Mongolian manuscripts, we first annotate all the personal names in the 'Little' Altan Tobchi using the manually compiled personal names' indices (lists of personal names) obtained from the 'Qad-un ündüsün quriyangγui altan tobči-Textological Study' [9]. After converting to a format that is suitable for a linear classifier, we input that data into the classifier for training, which returns a probability matrix (i.e., a model). The classifier is trained with gold standard annotations of tokens with known classes (i.e., personal names). The classifier calculates weights for each feature in correlation to each class. This can be seen as a probability of an object belonging to a certain class (i.e., personal names) when having those specific characteristics. These weights are saved in a probability matrix (i.e., NEE model), which will be used for classifying unseen named entities in the next steps.

3.1.1.2. Annotating step

In this step, each token of digitized ancient Mongolian manuscript will be annotated with the correct tag. We use the IOB2 [10] format for tagging tokens. 'B' tag indicates the beginning of a

http://dx.doi.org/10.5772/intechopen.72421

personal name, and 'I' tag indicates the tokens inside a personal name. 'O' tag indicates other tokens not belong to personal names. An example of the IOB2 annotation of the text in traditional Mongolian script can be seen in **Table 1**.

Because of some unique features of traditional Mongolian script, we also use 'Start/End' (SE) chunk tag set [11], which represents the character position in a word, along with the IOB2

Token	Transliteration	IOB2 tag
	ejen	O
	naiiman	O
	sirɣ-a_yi	O
	ögelen	B
	eke_dür_iyen	I
	abču	O
	irebei	O
	iregsen_ü	O
	qoyin-a	O
	qan	O
	yeke	O
	oru	O
	saɣubai.	O
	tengri_eče	O
	jayaɣ-a_bar	O

Token	Transliteration	IOB2 tag
	törügsen	O
	temüjin	B
	činggis	B
	qaɣan	I
	buyu	O

Table 1. An example of the IOB2 annotation of personal names in traditional Mongolian script text.

tags. 'S' tag is attached to the first character of each word including the personal names and 'E' tag to the last character. Therefore, each token will include the (1) IOB2 tag and (2) SE tag. SE tags are useful when there is a difference in word boundary between the test data and trained data [11, 12]. Particularly, an approach based on SE tags could improve the SVM prediction when there is no stemmer for traditional Mongolian. After attaching the IOB2 and SE tags to each token, we extract the features for chunking that will be used to learn the rules of personal name extraction. The features, i.e., characteristics of a token are explained in the next section.

3.1.1.3. Named entity extraction step

In this step, the proposed approach had to find the personal names in ancient Mongolian digitized texts. This method conducts the classification and grouping of tokens by SVM. The classifier in the SVM calculates a probability of a token belonging to personal names by inputting the extracted features to SVM. The features of a token might be possible clues to the proposed approach of whether or not this token is a named entity. In other words, we need some features to distinguish personal names.

We consider the following features of traditional Mongolian script for distinguishing personal names.

- **Preceding information of the current token**: If the preceding token is generational or dynastic information, an inherited or lifetime title of nobility, or a traditional descriptive phrase, it could indicate that current token is a personal name.
- **Beginning of a sentence**: For example, subjects or personal names are often at the beginning of a sentence.

- **Suffix**: In traditional Mongolian script, many living being and humankind proper names take only certain plural suffixes such as nar or ner and possessive suffixes [13].
- **Special non-word boundaries:** In traditional Mongolian script, some suffixes are visually separated from the stem of a word or from other suffixes, although they are an integral part of the word. Moreover, in some words with a final vowel letter 'a' or 'e', final vowel letters 'a' and 'e' are separated visually from the preceding consonant by a narrow gap although they are an integral part of the word stem.
- **End of token or special word delimiters**: A token is usually a word delimited by space, but there exist some unique features in traditional Mongolian script.
- **Information of the preceding and following tokens**: We also extract a feature by looking at the context of the current, preceding and succeeding IOB2 annotations (currently, the window stretches from C_{n-2} *to* C_{n+2}) as visualized in **Table 2**. Such a feature could correct mislabelled IOB2 annotations.

The final task in this step is to extract the personal names, which have the proper names' markups, from the ancient Mongolian digital text.

3.1.2. Performance of extracting named entities from Mongolian historical documents

The proposed method [14] is capable of extracting proper nouns from digitized text of ancient Mongolian manuscripts with 0.6993, 0.5679 and 0.6268 of precision, recall and F-measure, respectively, when utilizing a SVM tool LIBLINEAR with the L2-regularized L2-loss support vector classification (dual) solver [15].

When conducting experiments in extracting personal names from traditional Mongolian historical documents, we utilized digitized text of a chronological book of ancient Mongolian kings and the Mongol Empire—'Little' Altan Tobchi—which was made using bamboo pen xylograph technique as the experimental corpus. The 'Little' Altan Tobchi consists of 164 pages that contain approximately 16,200 words. The average number of words is 100 per page, with the longest one having 115 words and the shortest one 75 words. Precision, recall and F-measure were calculated by the fivefold cross-validation for extracting personal names.

Manually annotated named entities, extracted named entities [14], manually compiled scholar's commentaries and interpretations [9], as well as digital texts of ancient Mongolian manuscripts [7], will be utilized for building a digital edition of ancient Mongolian manuscripts. The next sections discuss how to develop a digital edition of Mongolian historical documents by describing some features and requirements of Mongolian manuscripts.

Tokens	W_{n-3}	W_{n-2}	W_{n-1}	W_n	W_{n+1}	W_{n+2}	W_{n+3}
IOB2 tags	C_{n-3}	C_{n-2}	C_{n-1}	C_n	C_{n+1}	C_{n+2}	C_{n+3}

Table 2. A feature of the preceding and following two tokens.

3.2. Making a web-based system by utilizing research outcomes

The past achievements in developing the TMSDL and the research outcomes of extracting named entities from Mongolian historical text allow us to create a digital representation that reflects ancient Mongolian historical manuscripts. This section covers our development in creating a web-based prototype system, which browses ancient Mongolian historical manuscripts.

3.2.1. A digital edition of Mongolian manuscripts

We utilized Edition Visualization Technology (EVT) for creating and browsing a digital edition of Mongolian manuscripts, which is encoded according to the Text Encoding Initiative (TEI) XML schemas and guidelines [16]. The named entities including the historical figures and place names are explicitly encoded using the TEI guidelines along with the additional data such as editorial markup, various commentaries, transcriptions and interpretations that have been suggested by researchers [9], etc., [17]. Well-known historical figures including generational or dynastic information, an inherited or lifetime title of nobility, or a traditional descriptive phrase or nickname are also marked. In the proposed digital edition, Unicode is chosen at the character level, and TEI P5 is applied on higher levels. As shown in **Figures 3** and **4**, all the personal names and place names in the 'Little' Altan Tobchi are visualized and highlighted in both transliteration and traditional Mongolian text. Image-to-text feature can

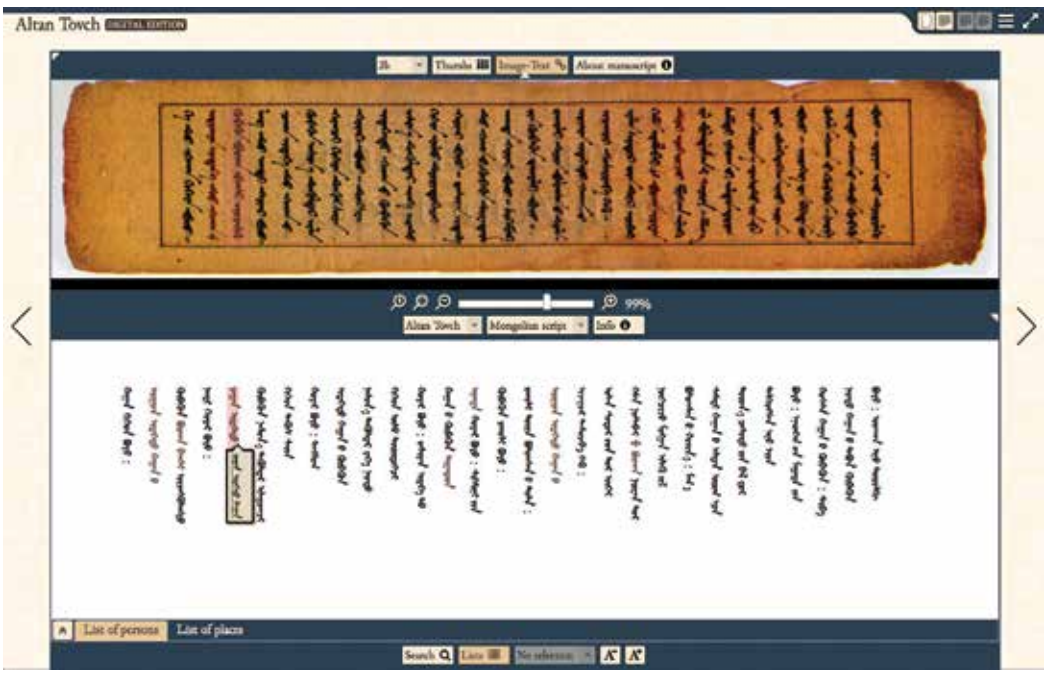

Figure 3. A digital edition with image-to-text link and personal names' highlights.

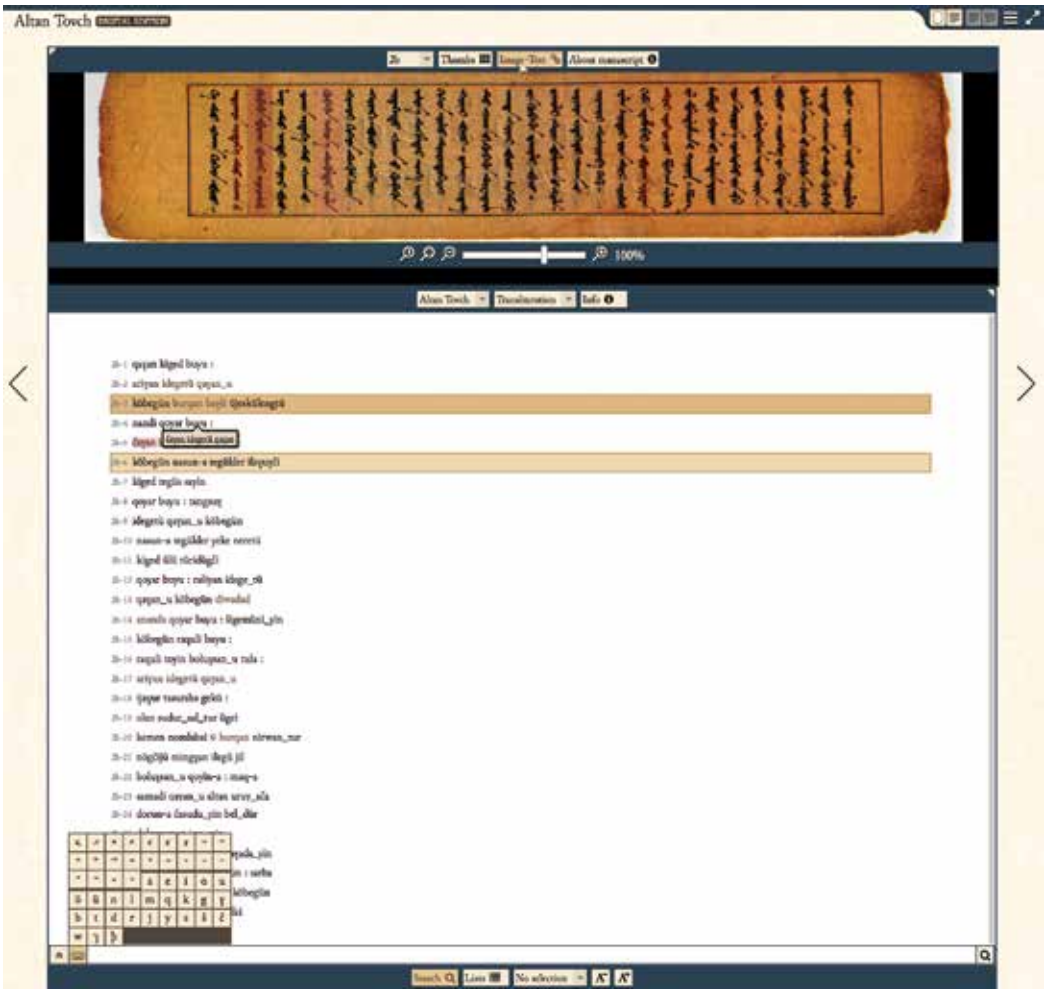

Figure 4. A digital edition with image-to-text link, a virtual keyboard and personal names' highlights in transliteration.

link a column in a manuscript folio image to the corresponding text and highlight them in all edition levels. As shown in **Figure 5**, all the named entities are listed as a full list with hyperlinks to the folios that appear certain named entity.

In addition, we made the following customizations in EVT to make it suitable for Mongolian manuscripts in traditional Mongolian script.

3.2.1.1. Parallel-text editions with transliteration

The proposed prototype can present scanned image-based editions with two edition levels: (1) diplomatic interpretative and (2) transliteration. Transliteration is helpful for those who are not familiar with a script of a certain language but understands that language. Transliteration in Latin letters of Mongolian historical documents is popular among scholars.

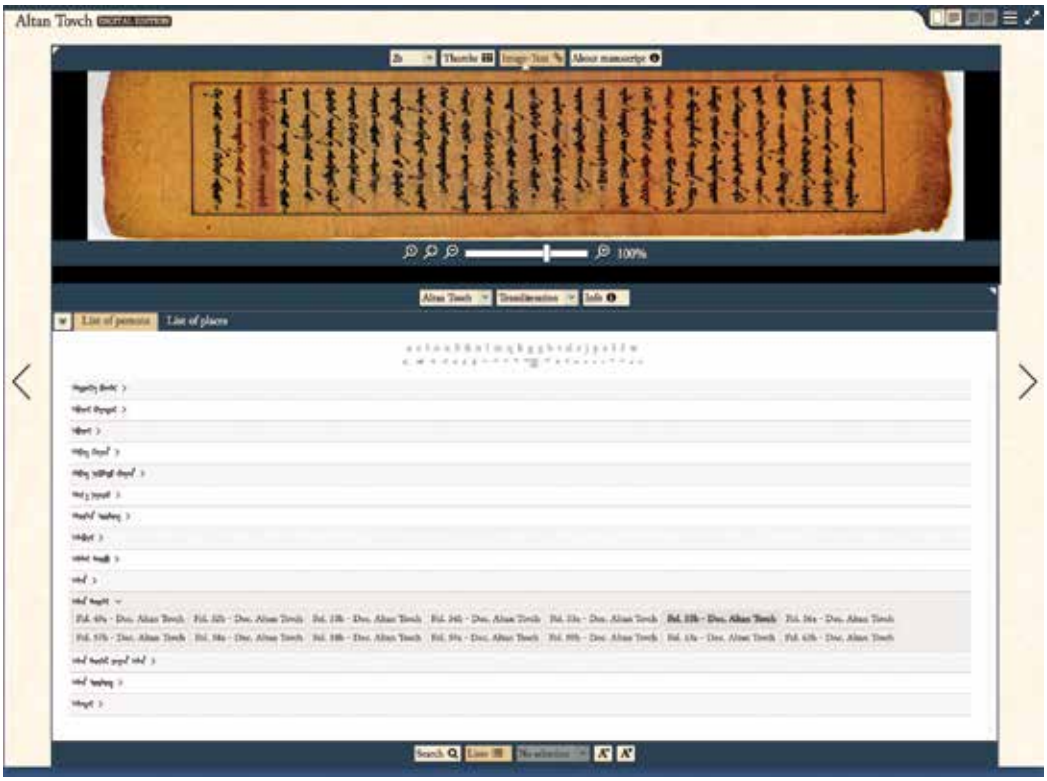

Figure 5. A list of named entities with hyperlinks to the folios of a Mongolian manuscript.

There is a limited recommendation to encode transliterations in TEI. Soualah and Hassoun [18] proposed to implement transliteration by using a specific model, which uses the [18] element with the *@xml:lang*, *@target* and *@type* attributes. However, we consider transliteration as a separate edition and use it as parallel-text editions as shown in **Figure 6**.

3.2.1.2. Supporting the traditional Mongolian script

A unique feature of traditional Mongolian script is displaying vertically, from top to bottom, in columns advancing from left to right. Due to poor support for traditional Mongolian script at the EVT, we customized it to display the scanned images at the top and the corresponding text in traditional Mongolian script below with the direction top to bottom and left to right. We also set to display text in traditional Mongolian script on the left, and the corresponding transliteration in Latin letters on the right that can be used to compare them. Additionally, as shown in **Figures 4** and **6**, we added a simple virtual keyboard composed of 22 traditional Mongolian letters and their corresponding Latin letters to help users to input a Mongolian keyword to benefit free-text search and keyword highlighting.

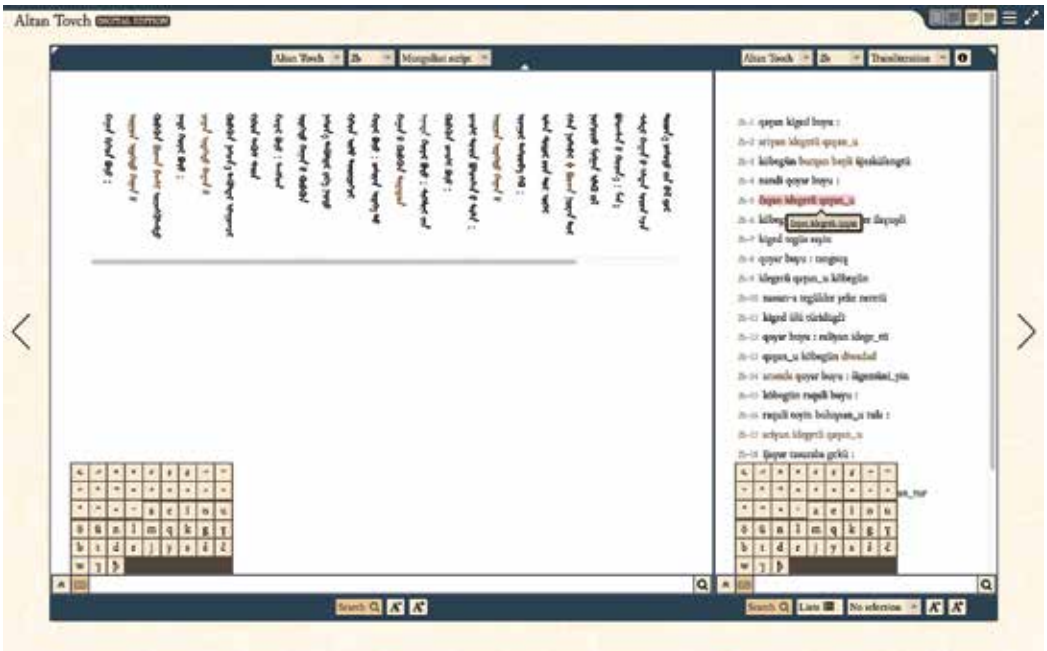

Figure 6. Parallel-text editions with personal names' highlights and virtual keyboards.

3.3. Applying and extending the proposed method to across languages

This section discusses (1) how the existing cross-language information retrieval techniques can be utilized in the proposed prototype system and (2) how the proposed approach can be applied to other languages in order to provide cross-lingual and cross-chronological information access to multilingual historical documents.

3.3.1. Adopting cross-language and cross-chronological information retrieval techniques in historical documents

There has been little research in information retrieval techniques for historical documents, and almost none of the breakthroughs in research in information retrieval and information access have aimed at retrieving information in the native language from ancient, cross-chronological and/or cross-script foreign language documents. Few approaches that could be considered a cross-chronological information retrieval have been proposed, and there has been little research in information retrieval techniques for historical documents. Ernst-Gerlach and Fuhr focused on modern and archaic German and developed a retrieval method that considers the spelling differences and variations over time [19]. Koolen et al. considered the spelling and pronunciation differences between ancient and modern Dutch [20], while Gotscharek et al. [21] and Hauser et al. [22] considered the spelling differences and variations between

modern and archaic German. Pilz et al. considered spelling variations of English and German historical texts [23]. In general, the main challenge for historical European languages like Dutch, English and German is the spelling variants.

Furthermore, Kimura and Maeda proposed a retrieval method that considers not only language differences over time but also cultural and time differences in modern and archaic Japanese [24]. Tripathi developed a retrieval system that considers the differences in various scripts and writing systems of Brahmic (Indic) and proposed a method to retrieve Sanskrit documents written in Sanskrit script or Brahmic families' scripts, using scripts such as Devanagari, Kannada, Telugu and Bengali [25]. To cope with cross-chronological and cross-script Mongolian documents, Khaltarkhuu and Maeda proposed a retrieval technique that is capable of searching traditional Mongolian script documents using modern Mongolian query [26–28].

We improved Khaltarkhuu and Maeda's grammatical-rule-based approach [26–28] and proposed an 'ancient-to-modern information retrieval' method [7, 29] by adding a dictionary-based query translation technique in order to consider cross-chronological differences in the writing systems of the ancient and modern Mongolian languages for accessing cross-chronological and cross-script ancient Mongolian documents by using a query in modern Mongolian in Cyrillic. To boost the quality of the translation, the 'ancient-to-modern information retrieval' approach [7, 29] matches query terms to words in a dictionary. If no exact match is found, the grammatical-rule-based approach [26–28] is used. In other words, the grammatical-rule-based query translation approach is used for inflected words, words with ancient spellings or grammar or the words missing from the dictionary. For the word sense disambiguation, in case if there are words which have multiple candidates, we choose the most frequent words. In our approach, we merge spelling variants of ancient Mongolian words.

We have already integrated the 'ancient-to-modern information retrieval' method in the TMSDL, and it can be easily applied to our digital edition for accessing ancient Mongolian historical collections written in traditional Mongolian script.

3.3.2. Applying the proposed approach to other languages

We have been demonstrating a facility for cross-language searching between English and Japanese for enabling English-speaking users to search Ukiyo-e databases available in Japanese by using English queries [30–32]. Such a feature is very useful for users, since the Ukiyo-e databases in Japanese institutions are mostly available in Japanese, so that users who do not understand Japanese may not find the desired information. Ukiyo-e, a Japanese traditional woodblock printing, is known worldwide as one of the fine arts of the Edo period (1603–1868). The texts of Ukiyo-e databases contain archaic Japanese words which reflect the Japanese language of the Edo period.

Like the 'ancient-to-modern information retrieval', a dictionary-based query translation approach is adopted by utilizing a domain-specific dictionary, which contains the terms related to Japanese arts and cultures. The proposed feature works well with a variety of keywords (i.e., no full sentences) that may include the personal names, specific terms such as 'Geisha', traditional Japanese female entertainers; 'Fuji', Mount Fuji, the highest mountain in

http://dx.doi.org/10.5772/intechopen.72421

Japan; and 'Sumo', Japanese traditional wrestling. For instance, if the search query submitted by the user is a name of the Ukiyo-e artist, i.e., 'Utagawa Hiroshige', then the query 'Utagawa Hiroshige' is translated into Japanese as '歌川広重' and sent to Japanese databases.

We are conducting further research to generalize the proposed method to other historical documents in various languages. We also believe that the proposed prototype could be applied to other historical documents in Todo, Manchu and Sibe, which are the derivative scripts of traditional Mongolian.

4. Summary and future directions

In this chapter, we have described our research to achieve cross-lingual and cross-chronological information access to ancient Mongolian historical materials. More specifically, we have introduced methods for providing information access that cuts across different historical periods and dialects.

We introduced an information extraction method for digitized ancient Mongolian historical manuscripts of the 13–16th century in Sections 3. The proposed information extraction method for ancient Mongolian historical documents performs computerized massive analysis on Mongolian historical documents. It can reduce traditional labour-intensive manual analysis on Mongolian historical text significantly. Named entities such as historical figures and places of ancient Mongolia that are difficult for manual examination are recognized from historical manuscripts.

The extracted results are utilized for building a digital edition of an ancient Mongolian historical document and made available through a web-based system.[1] We also believe the TEI-encoded digital edition that reflects the ancient Mongolian manuscripts would help scholars conducting research in the ancient history for digging hidden knowledge of the Middle Ages of Mongolia in ancient Mongolian historical documents that is not available in modern-language documents. Furthermore, explicitly encoded digital text enables users to search and browse ancient Mongolian manuscript using the named entities' visualization, i.e., it allows not only retrieving information but also analysing and visualizing the contents of the information. We also hope digital editions along with the scanned images would recreate the experience of encountering the original manuscripts. Its information visualization feature of ancient Mongolian texts and a TMSDL's feature that can retrieve ancient manuscripts written in traditional Mongolian script using a query in modern Mongolian (Cyrillic) would help researchers who are interested in using digital representations of ancient historical manuscripts as scholarly tools by using a modern language. Such a feature is very useful, since the needs of humanities researchers are diverse and might require access to information in ancient languages, rather than searching and browsing limited collections in modern languages. Indeed Mongolian ancient documents are mostly available in ancient scripts and dialects, so users who do not understand ancient Mongolian may not find the desired information.

[1] http://www.dl.is.ritsumei.ac.jp/AltanTovch/

Finally, the proposed prototype could be applied to other documents in Todo, Manchu and Sibe, which are the derivative scripts of traditional Mongolian. The systems introduced in this chapter are targeted primarily at researchers in the humanities field. Nevertheless, these systems are expected to be useful to users other than researchers, in the sense that they open up new possibilities for acquiring the kinds of information that cannot be found solely in modern documents available on the web.

Author details

Biligsaikhan Batjargal

Address all correspondence to: biligsaikhan@gmail.com

Research Organization of Science and Technology, Ritsumeikan University, Japan

References

[1] Shagdarsüren Ts. Study of Mongolian Scripts (Graphic Study or Grammatology). 2nd ed. Ulaanbaatar: Urlakh Erdem Khevleliin Gazar: Ulaanbaatar: Centre for Mongol Studies, National University of Mongolia; 2001. 299 pp

[2] The Unicode Consortium. Chapter 13: South and Central Asia-II, Other Modern Scripts. The Unicode Standard, Version 10.0.0 [Internet]. 2017 June 20 [Updated: 2017 June 20]. Available from: http://www.unicode.org/versions/Unicode10.0.0/ch13.pdf

[3] The Unicode Consortium. Chapter 14: South and Central Asia-III, Ancient Scripts. The Unicode Standard, Version 10.0.0 [Internet]. 2017 June 20 [Updated: 2017 June 20]. Available from: http://www.unicode.org/versions/Unicode10.0.0/ch14.pdf

[4] Sečenbaγatur, Qasgerel, Tuyaγ-a, ǰirannige B, Ying ǰe U. Mongγul kelen-ü nutuγ-un ayalγun-u sinǰilel-ün uduridqal. Kökeqota: Öbür mongγul-un arad-un keblel-ün qoriy-a; 2005

[5] Svantesson J, Tsendina A, Karlsson A, Franzén V. The Phonology of Mongolian. New York: Oxford University Press; 2005. p. 336

[6] Bira Sh. The Golden Summary which Relates Briefly the Deeds of Civil Governing Established by Ancient emperors (The Mongol Chronicle of the 17th Century). Ulaanbaatar, Mongolia: Mongolia: Ulsiin hevleliin gazar; 1990. 222 pp

[7] Batjargal B, Khaltarkhuu G, Kimura F, Maeda A. Developing a digital library of historical records in traditional Mongolian script. International Journal of Digital Library Systems. 2012;**3**(1):33-53. DOI: 10.4018/jdls.2012010103

[8] Finkel JR, Grenager T, Manning CD. Incorporating non-local information into information extraction systems by Gibbs sampling. In: Proceedings of the 43rd Annual Meeting

on Association for Computational Linguistics (ACL '05); June 25-30, 2005; Ann Arbor, Michigan. Stroudsburg, PA, USA: Association for Computational Linguistics; 2005. pp. 363-370. DOI: 10.3115/1219840.1219885

[9] Choimaa Sh. Qad-un ündüsün quriyangγui altan tobči (Textological Study). vol. 1 ed. Ulaanbaatar: Urlakh Erdem Khevleliin Gazar: Ulaanbaatar: Centre for Mongol Studies, National University of Mongolia; 2002. 276 pp

[10] Ramshaw LA, Marcus MP. Text chunking using transformation-based learning. In: Armstrong S, Church K, Isabelle P, Manzi S, Tzoukermann E, Yarowsky D, editors. Natural Language Processing Using Very Large Corpora. Dordrecht: Springer Netherlands; 1999. pp. 157-176. DOI: 10.1007/978-94-017-2390-9_10

[11] Asahara M, Matsumoto Y. Japanese named entity extraction with redundant morphological analysis. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1; May 27-June 01, 2003; Edmonton, Canada. Stroudsburg, PA, USA: Association for Computational Linguistics; 2003. pp. 8-15. DOI: 10.3115/1073445.1073447

[12] Yoshimura M, Kimura F, Maeda A. Personal name extraction from ancient Japanese texts. In: Proceedings of the Exploration, Navigation and Retrieval of Information in Cultural Heritage ENRICH 2013 Workshop; 1 August 2013; Dublin, Ireland. New York, NY, USA: ACM; 2013. pp. 31-34

[13] Chinggaltai, editor. A grammar of the Mongol language. Revised ed. New York: Frederick Ungar Publishing Co; 1963. 173 pp

[14] Batjargal B, Khaltarkhuu G, Kimura F, Maeda A. An approach to named entity extraction from Mongolian Historical Documents. In: Proceedings of the 2015 International Conference on Culture and Computing; 17-19 October; Kyoto, Japan. IEEE Computer Society; 2015. pp. 205-206. DOI: 10.1109/Culture.and.Computing.2015.41

[15] Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A library for large linear classification. The Journal of Machine Learning Research. 2008;**9**:1871-1874

[16] Del Turco RR, Buomprisco G, Pietro CD, Kenny J, Masotti R, Pugliese J. Edition visualization technology: A simple tool to visualize TEI-based digital editions. Journal of the Text Encoding Initiative [Online]. 2014;**8**. DOI: 10.4000/jtei.1077

[17] Batjargal B, Khaltarkhuu G, Kimura F, Maeda A. Applying text encoding initiative guidelines to a historical record in traditional Mongolian script. In: Proceedings of the 2013 International Conference on Culture and Computing; 16-18 September; Kyoto, Japan. IEEE Computer Society; 2013. p. 141-142. DOI: 10.1109/CultureComputing.2013.36

[18] Soualah MO, Hassoun MA. TEI P5 manuscript description adaptation for cataloguing digitized Arabic manuscripts. Journal of the Text Encoding Initiative [Online]. 2012;**2**:DOI: 10.4000/jtei.398

[19] Ernst-Gerlach A, Fuhr N. Retrieval in text collections with historic spelling using linguistic and spelling variants. In: Rasmussen E, editor. Proceedings of the 7th ACM/IEEE-CS

Joint Conference on Digital Libraries; June 18-23; Vancouver, BC, Canada. USA/New York City: ACM; 2007. p. 333-341. DOI: 10.1145/1255175.1255242

[20] Koolen M, Adriaans F, Kamps J, Rijke M. A cross-language approach to Historic Document retrieval. In: Lalmas M, MacFarlane A, Rüger S. M, Tombros A,Tsikrika T, Yavlinsky A, editors. Proceedings of the 28th European conference on Advances in Information Retrieval ECIR. Heidelberg: Springer-Verlag Berlin; 2006. pp. 407-419. DOI: 10.1007/11735106_36

[21] Gotscharek A, Reffle U, Ringlstetter C, Schulz K, Neumann A. Towards information retrieval on historical document collections: The role of matching procedures and special lexica. International Journal on Document Analysis and Recognition (IJDAR). 2011;**14**(2):159-171. DOI: 10.1007/s10032-010-0132-6

[22] Hauser A, Heller M, Leiss E, Schulz K, Wanzeck C. Information access to Historical Documents from the early new high German Period. In: Proceedings of the IJCAI-2007 Workshop on Analytics for Noisy Unstructured Text Data; January 8; Hyderabad, India. 2007. p. 147-154

[23] Pilz T, Ernst-Gerlach A, Kempken S, Rayson P, Archer D. The identification of spelling variants in English and German historical texts: Manual or automatic? Literary and Linguistic Computing. 2008;**23**(1):65-72. DOI: 10.1093/llc/fqm044

[24] Kimura F, Maeda A. An approach to information access and knowledge discovery from Historical Documents. In: Conference Abstracts of the Digital Humanities 2009; June 22-25; College Park, Maryland. 2009. pp. 359-361

[25] Tripathi A. Saraswati: Cross-lingual Sanskrit digital library. Library Hi Tech News. 2009;**26**(10):1-5. DOI: 10.1108/07419050911022252

[26] Khaltarkhuu G, Maeda A. Retrieval technique with the modern Mongolian query on traditional Mongolian text. In: Sugimoto Sh, Hunter J, Rauber A, Morishima A, editors. Digital Libraries: Achievements, Challenges and Opportunities, 9th International Conference on Asian Digital Libraries, ICADL 2006; November 27-30; Kyoto, Japan. Heidelberg: Springer, Berlin; 2006. p. 478-481. DOI: 10.1007/11931584_5

[27] Khaltarkhuu G, Maeda A. Building a digital library of traditional Mongolian Historical Documents. In: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries; June 18-23; Vancouver, BC, Canada. NY, USA: ACM New York; 2007. p. 483. DOI: 10.1145/1255175.1255282

[28] Khaltarkhuu G, Maeda A. Developing a Traditional Mongolian Script Digital Library. In: Buchanan G, Masoodian M, Cunningham SJ, editors. Digital Libraries: Universal and Ubiquitous Access to Information, 11th International Conference on Asian Digital Libraries, ICADL 2008; 2-5 December; Bali, Indonesia. Heidelberg: Springer, Berlin; 2008. pp. 41-50. DOI: 10.1007/978-3-540-89533-6_5

[29] Batjargal B, Khaltarkhuu G, Kimura F, Maeda A. Ancient-to-modern information retrieval for digital collections of Traditional Mongolian Script. In: Chowdhury GG, Koo C, Hunter J, editors. The Role of Digital Libraries in a Time of Global Change,

12th International Conference on Asia-Pacific Digital Libraries, ICADL 2010; June 21-25; Gold Coast, Australia. Heidelberg: Springer, Berlin; 2010. p. 25-28. DOI: 10.1007/978-3-642-13654-2_4

[30] Batjargal B, Kimura F, Maeda A. Approach to Cross-language Retrieval for Japanese Traditional Fine Art: Ukiyo-e Database. In: Lalmas A, Jose J, Rauber A, Sebastiani F, Frommholz I, editors. Proceedings of the 14th European Conference on Research and Advanced Technology for Digital Libraries (ECDL2010); September 6-10; Glasgow, UK. Heidelberg: Springer, Berlin; 2010. p. 518-521. DOI: 10.1007/978-3-642-15464-5_71

[31] Batjargal B, Kimura F, Maeda A. Realizing bilingual and parallel access to Ukiyo-e databases in the World. In: Proceedings of the International Conference on Culture and Computing (Culture and Computing 2011); October 20-22; Kyoto, Japan. IEEE; 2011. pp. 165-166. DOI: 10.1109/Culture-Computing.2011.48

[32] Batjargal B, Maeda A, Akama R. Providing bilingual access to multiple Japanese humanities databases: Text retrieval using English and Japanese queries. In: Hsiang J, editor. Digital Humanities: Between Past, Present, and Future. Taipei, Taiwan: National Taiwan University Press; 2016. pp. 351-367

Chapter 9

Innovative Multilingual CAPTCHA Based on Handwritten Characteristics

Maha Hamad Aldosari

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72599

Abstract

Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) is a kind of test which is commonly used by different websites on the Internet to differentiate between humans and automated bots. Most websites require users to pass the CAPTCHA before signing up or filling out most forms. CAPTCHA today is even used on some mobile applications to provide a higher security level that can protect websites and mobile applications against malicious attacks by automated bots and spammers. The technique essentially relies on employing the human recognition ability, which is not available in automated bots or machines, through leveraging the handwriting characteristics in designing CAPTCHA. The novelty of the technique proposed in this work is that it adopts handwritten characters of four different languages (English, Arabic, Spanish, and French) to generate handwritten multilingual CAPTCHA text. The technique was duly tested and the initial experiments' results for the technique have shown a promising security level that each of the techniques would provide.

Keywords: CAPTCHA, handwritten CAPTCHA, web security, optical character recognition (OCR)

1. Introduction

Web applications have increased rapidly and become a daily necessity for most people [1]. Creating an email account, using social networking sites, and accessing websites are examples of day-to-day activities for Internet users. The fast evolution of the Internet means that the security aspect is being threatened [2]. The number of bots (robot) programs that attack websites has increased. These bots can bring down the site and cause a significant amount of damage. These attacks can take many forms such as DDoS attacks, viruses, worms, and many other malicious devices. They are also considered as the primary reason for email spam [3].

Therefore, it is obvious that stopping such bots by means of a reliable Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) is inevitable. More so, in a multilingual world, multilingual CAPTCHAs are indispensable.

Completely Automated Public Turing Test to Tell Computers and Humans Apart () is considered one of the most common techniques that can be used to distinguish between humans and artificial agents (or bots). For time being, the exponential growth of free web services has led to the misuse of automated bots and spam [4], which has resulted in serious security issues in web services. Using CAPTCHA in its various types has proven to be effective in protecting websites, and the services they provide, from any harm caused by bots' attacks [1].

2. CAPTCHA technique based on handwriting

This technique adopts the handwritten text in the CAPTCHA images and applies a unique feature (separating handwritten characters). This feature can help in differentiating it from any previous handwritten CAPTCHA techniques, and prospectively enhances security level. Moreover, the CAPTCHA's text combines different text languages beside the default language (English) which makes it a multilingual CAPTCHA. The secondary language is selected from a set of languages (French, Spanish, and Arabic) based on the user's region. The main reason for providing multilingual CAPTCHA is that other OCR programs in other languages have not reached the professionalism level of the English OCR yet, and to expand the CAPTCHA usage scope to be used worldwide [1].

At the beginning, different handwritten characters were collected from 100 volunteers; each volunteer wrote the alphabet characters of the 4 adopted languages for the research, each using their own handwriting style. The handwritten characters were classified and stored in a database. These characters were used to synthesize random words that generate the CAPTCHA text, and users should recognize the words in order to pass the CAPTCHA. Furthermore, for the sake of adding a proper security level that will protect the website services from bots' attacks, some distortion methods are applied on each handwritten character separately at the generation process to increase the difficulty for bots to break the CAPTCHA, besides the handwritten characteristics that are fairly resistant for such bots to break.

In summary, this technique goes through two main phases as part of its generating process: the first phase is data gathering and preparation, and the second phase is CAPTCHA implementation with some steps included in each phase.

2.1. Data gathering and preparation

This phase goes through six steps. They are as follows:

- The first step is characters' samples creation. In this step, samples for each character in the four different adopted languages (English, Arabic, Spanish, and French) that will be used in the CAPTCHA text are made [1].

- The second step is samples distribution. In this step, the CAPTCHA characters' samples were distributed to 100 volunteers; each volunteer wrote the samples' characters of the 4 adopted languages by their own handwriting style [1]. As a result, we had a total of 100 different instances with different handwriting styles for each character in each language. So, we ended up with almost more than 10,700 samples' characters that need to be stored on a database in order to be used later in the CAPTCHA implementation phase.
- The third step is transforming samples into digital format. Here, all the collected samples were scanned and stored in digital formats (images).
- The fourth step is sorting data. In this step, we sorted out the collected data into four languages. So, at this point, we have 4 sets of images, each set belongs to one language, and there are 100 different images for each character in each language. Moreover, 4 tables were created on the database to store the images that will be used later to generate the CAPTCHA [1].
- The fifth step is classifying the worldwide countries into categories according to the spoken language there. The countries go with one of the four adopted languages; however, the rest of the countries where their main spoken language is not one from the four adopted languages, we classify them as English-speaking country. After that we stored the countries list with their matched languages on the database [1].
- The sixth step is identifying a list of inappropriate words in each language and storing it on the database as well. **Figure 1** summarizes the data gathering and preparation phase steps.

2.2. Algorithm technique

Figure 2 shows an abstract view of the technique process.

2.3. Handwriting characteristics

Choosing and utilizing the handwriting in designing new CAPTCHA technique was not decided randomly with any logical reasons. On the contrary, it was chosen after a quite long search and study of what characteristics the handwriting has, and how it could be utilized in security field.

Nevertheless, the handwriting in general has some characteristics that can only be utilized by humans. Due to the human brain's superior ability, the brain can analyze and recognize unclear handwritten characters and digits; it also can recognize various different

Figure 1. Steps of the data gathering and preparation phase.

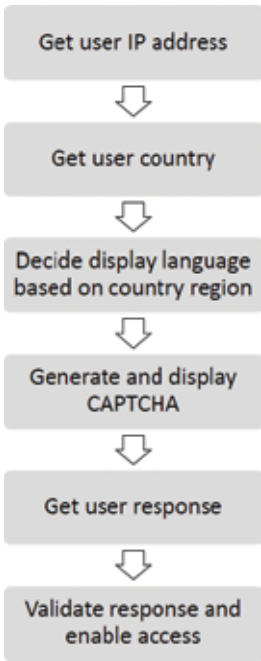

Figure 2. Abstract view of the technique process.

handwriting styles written by different people. Moreover, the human brain has the privilege of using its experience to figure out uncompleted characters or uncompleted words that have missing letters. It even can read the Arabic words written without any dots on its letters, because the words' shapes can be enough for the brain to figure out the words, unlike OCR machines which mostly cannot recognize the words if they are not complete or without dots in Arabic words case.

Overall, this confirms the human capability in utilizing the handwriting characteristics, which cannot be found in any OCR machine, and it encourages us to go through this CAPTCHA technique which is based on handwriting.

3. Technique implementation

The generation process of this technique starts by getting the user's IP address. Then, it gets the country's name where the user is located at the time of accessing which is obtained using the IP-API service. Consequently, a country language will be retrieved from the database using the country name, where a list of countries is sorted and classified into one of the adopted languages (Arabic, English, French, and Spanish). A list for each of the adopted language was created which contains the countries which speak the specified language. Hence, the countries classification is done based on the official spoken language in each country. However, if

http://dx.doi.org/10.5772/intechopen.72599

the country's name is not on any of the four countries' lists, then English will be the default language to use (English).

In addition, the user's website default language will be determined and compared to the retrieved country's language; if they are different, then the website default language will be used.

The following flowchart (**Figure 3**) illustrates the whole process of the first step in this CAPTCHA technique which decides the CAPTCHA language to be displayed to the user.

Furthermore, after the language has been decided on, the CAPTCHA generation process will move on to the next step which is choosing the CAPTCHA word length. The word length is chosen randomly from five to eight characters. Next, the word construction process will start by selecting the handwritten characters and distorting them separately. However, this step will be done little bit differently if the previously decided language is Arabic; the following flowchart (**Figure 4**) clarifies the CAPTCHA word construction process in detail.

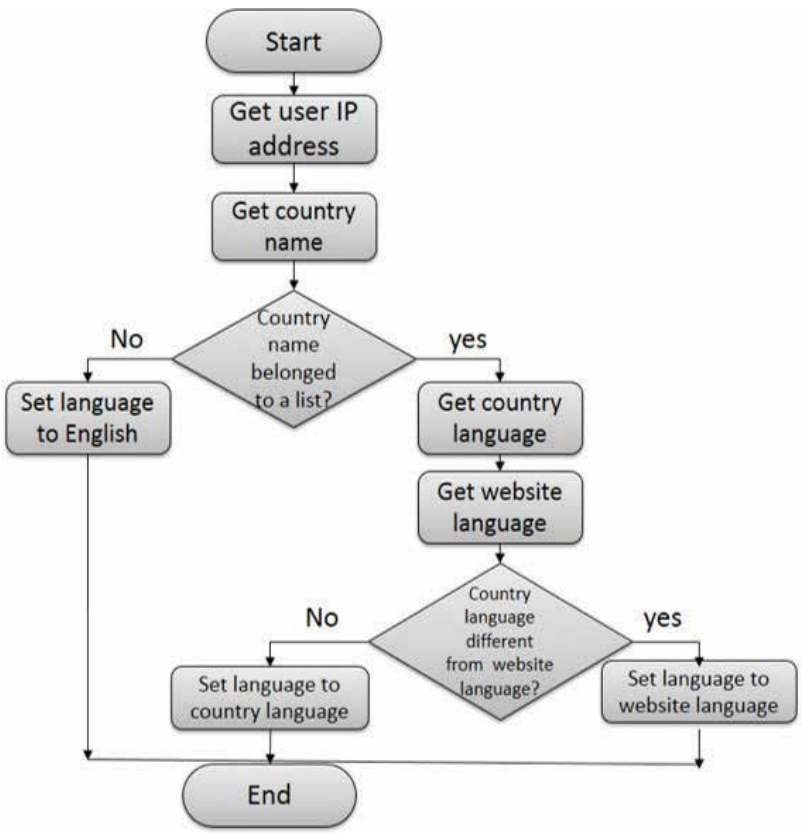

Figure 3. Deciding the CAPTCHA language process.

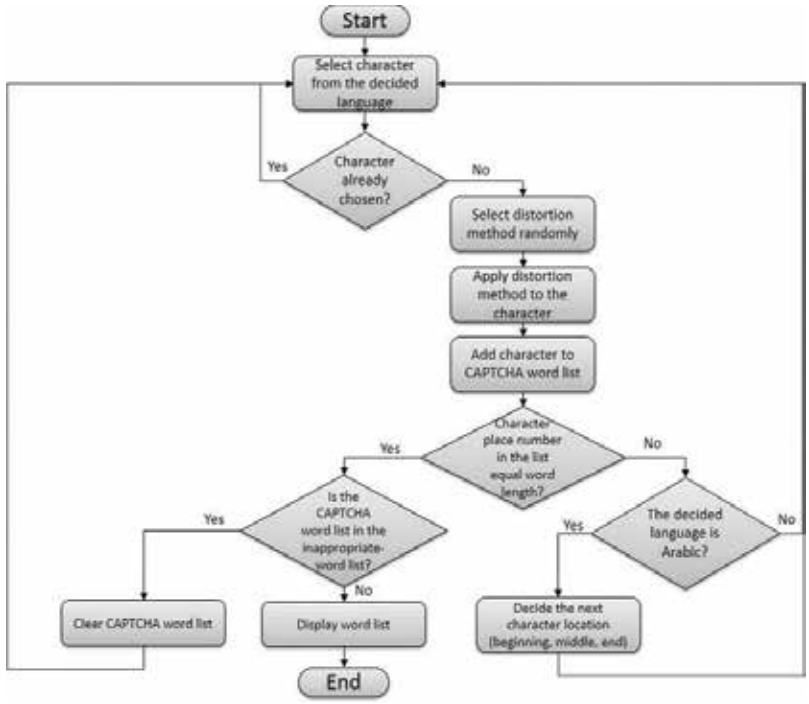

Figure 4. CAPTCHA word construction process.

Figure 5. English CAPTCHA "M u J F R t Q".

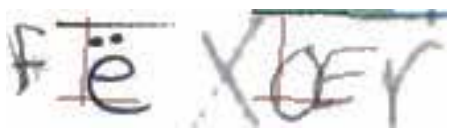

Figure 6. French CAPTCHA "F ë x Œ r".

Figure 7. Spanish CAPTCHA "X b CH y N R w".

As shown above, when the CAPTCHA word is generated, it is displayed to the user in one of adopted languages (Arabic, English, French, and Spanish). **Figures 5–8** show examples of the handwritten CAPTCHA technique with each adopted language.

http://dx.doi.org/10.5772/intechopen.72599

Figure 8. Arabic CAPTCHA "بوكشن".

4. Experiment techniques

In the conducted experiments, six different OCRs were used to test the technical performance of the proposed CAPTCHA techniques. The used OCRs have a good review from some technical experts and they provide good results when used to extract the regular text.

Moreover, other methods were used to test the usability, such as surveys and local web pages to get the users' responses and analyze them from different perspectives.

4.1. First OCR

The first OCR used in the experiments is an application called Free OCR. This application utilizes the most recent version of the Tesseract OCR engine (v3.01), which can ensure a reliable level of text-extracting accuracy. Tesseract is an open-source OCR engine maintained by Google. It offers support for different languages, with a level of accuracy potentially reaching 98% [1, 5].

4.2. Second OCR

Capture2Text is the second technique used in our experiments. It is an open-source OCR tool, like the first OCR; it uses the Tesseract engine introduced by Google to capture the written text in images and then copies it to the clipboard.

4.3. Third OCR

The third OCR used is a free online OCR called i2OCR. It is available in the following link: http://www.i2ocr.com/. This online OCR supports various recognition languages; it also has the ability to extract text from various columns in the images.

4.4. Fourth OCR

FreeOCR is the fourth OCR tool used in the experiments. As the name suggests, it is available online as a free service, which is available in the following link: http://www.free-ocr.com/. Moreover, the extraction process speed for this OCR site is considered fast in comparison with other online OCRs, and it produces the extracted text fairly quickly [6].

4.5. Fifth OCR

The fifth OCR we adopted in the experiments is an online OCR software called OnlineOCR. This OCR software is available in the following link: http://www.onlineocr.net/. Additionally, this

OCR software supports 46 recognition languages and it is able to extract texts in any of these languages. It also can detect text written in more than one language in the same image or document.

4.6. Sixth OCR

NewOCR is a free online OCR service that we used as the sixth technique in our experiments. The NewOCR service is available in the following link: https://www.newocr.com/. This online service supports more than 100 recognition languages and different fonts supports. In addition, the NewOCR service works using Tesseract OCR engine which is considered the best accurate OCR engine available at this time. It also supports the low-resolution images and can extract the text written in these images.

5. Technical performance testing

First of all, we started with the technical performance testing to test the technical aspects of the proposed handwritten CAPTCHA. A total of 500 different CAPTCHA images were generated for each language. Each of the 500 CAPTCHA images of the first 3 languages (English, French, and Spanish) were tested on 6 different OCRs, while the 500 Arabic CAPTCHA images were tested on the second and sixth OCRs. **Table 1** illustrates the testing results for the six different OCRs on each adopted language.

As shown in **Table 1**, the testing results were divided into four patterns [1]:

1. Correctly recognized pattern: it is when all the text in the CAPTCHA image has been correctly recognized.
2. Partially correctly recognized pattern: it is when the OCR has recognized three or more characters in the CAPTCHA text.
3. Incorrectly recognized pattern: it is when no characters have been correctly recognized in the CAPTCHA text.
4. No text pattern found: it is when the OCR was not able to recognize the text or any character in the CAPTCHA image.

In the English CAPTCHA images, the six OCRs have failed to recognize the full text in 99% of the images, while only 1% was correctly recognized. Nevertheless, the 99% includes 6% partially correctly recognized patterns, 46% no text found in the CAPTCHA image, and 47% totally incorrect text recognition [1].

Moreover, the other languages testing outcomes resulted in a lower recognition percentage compared to the English one. In Spanish language case, all the six OCRs have failed to correctly recognize 99.97% of the Spanish CAPTCHA images; this 99.97% includes 4.23% partially

http://dx.doi.org/10.5772/intechopen.72599

Pattern		Correctly recognized	Partially correctly recognized	Incorrectly recognized	No text found
First OCR	English	5	54	431	10
	Spanish	1	16	475	8
	French	0	2	486	12
Second OCR	English	0	30	390	80
	Spanish	0	24	403	73
	French	0	18	398	84
	Arabic	0	3	391	106
Third OCR	English	0	65	250	185
	Spanish	0	47	277	176
	French	0	23	201	276
Fourth OCR	English	0	40	150	310
	Spanish	0	37	84	379
	French	0	29	63	408
Fifth OCR	English	0	0	115	385
	Spanish	0	0	81	419
	French	0	0	97	403
Sixth OCR	English	0	5	85	410
	Spanish	0	3	118	379
	French	0	1	68	431
	Arabic	0	0	73	427

Table 1. The testing results for the six different OCRs on each adopted language.

correctly recognized, 47.8% no text found in the CAPTCHA image, and 47.9% totally incorrect text recognition. Likewise, in the French language case, the six OCRs used in the experiments did not succeed in correctly recognizing any of the French CAPTCHA images, while only [7] 0.4% of the whole French CAPTCHA images were partially correctly recognized, and 53.8% of the images resulted in no text found, and the remained 43.8% of the images were incorrectly recognized.

Similarly, the experiments result for the Arabic language shows that the two used OCRs have failed to correctly recognize any of the Arabic CAPTCHA images. However, the two used OCRs were able to partially correctly recognize only 0.3% from the whole Arabic CAPTCHA images, whereas 53.3% of the images resulted in no text found and 46.4% of the images were incorrectly recognized.

6. Usability testing

In order to infer the usability of the proposed CAPTCHA, two users' acceptance tests have been conducted on this technique. The first test was aimed at taking a large number of responses from different users, while the second test was aimed at testing a large number of the produced CAPTCHA images.

6.1. First usability test

The first test targeted 100 users through an online survey that aimed at understanding how users will interpret five different CAPTCHA images. These five images were chosen on the basis of different aspects to study users' responses regarding characters' distortions and unclear handwriting styles (see Appendix B for CAPTCHA images used in the survey).

Table 2 illustrates the results of the conducted survey and shows the answers' patterns for each CAPTCHA image used in the survey.

According to the results shown in **Table 2**, 82% of the users were able to correctly recognize the CAPTCHA characters of the first image, the remaining 18% failed and were confused between characters, and noise, and distortion.

As for the second and third images, they were recognized by 85 and 75% of the users, respectively, while the rest of the users were confounded by one character, due to the warping distortion method applied on that character.

Moreover, users succeed in correctly recognizing the fourth CAPTCHA image with a percentage of 74%, while 61% of them correctly recognized the fourth and fifth CAPTCHA images. However, the remain percentages of the users who failed to correctly recognize the last two CAPTCHA images have failed because of the unclear handwriting style which was selected on purpose to reflect the worst cases that could be produced from the collected database.

In general, it must be mentioned that the partially correctly recognized pattern indicates that the user misinterpreted three characters or fewer from the CAPTCHA word, otherwise it will be considered that the user has incorrectly recognized the whole CAPTCHA word.

Additionally, the average time taken by each participant to solve the survey was 2 minutes and 38 seconds.

Pattern	First image	Second image	Third image	Fourth image	Fifth image
Correctly recognized	82	85	75	74	61
Partially correctly recognized	18	15	25	26	39
Incorrectly recognized	0	0	0	0	0

Table 2. Results of the survey conducted for the first technique.

6.2. Second usability test

The second usability test was carried out using the implemented web pages to give users a real experience similar to solving a real CAPTCHA on any website. Therefore, as mentioned earlier, the implemented web pages produce CAPTCHA images at the real time and responds to users as soon as they click on the "validate" button to inform them either they solve the CAPTCHA correctly or not.

Consequently, the test was conducted on 20 users with five CAPTCHA images average for each user; each user viewed the CAPTCHA on the web page and solved five different CAPTCHAs, and in the meanwhile the users' answers and the web page responses to their answers were recorded.

Correspondingly, the testing results showed that 92% of the tested images were correctly recognized by users, while in the remaining 8% of the images fewer than three characters of each image were misinterpreted by the users.

However, the percentages of the results have proven a fair usability percentage, which also could be further improved with little adjustments on the distortion methods and on the collected handwritten characters database [1].

7. Conclusion

The rapid evolution of web and mobile applications turn these applications into an important part in people's daily life, where people rely on them to accomplish most of their activities. On the other hand, all the rapid improvement in these applications comes with a rapid increase in the number of malicious bots and applications that threatens the security of web and mobile applications.

Therefore, we introduced a new CAPTCHA technique that utilizes handwriting styles and we have put it through several experiments to adjust, improve, and test the technique while trying to reflect every needed adjustment to the technique immediately.

The introduced technique is a novel handwritten CAPTCHA, which basically relies on employing the handwriting characteristics that can only be interpreted by humans while being comparatively hard for OCRs to recognize. The proposed approach adopts four different languages (English, Arabic, Spanish, and French); each language has its own handwritten characters used in synthesizing the CAPTCHA text.

Moreover, few testing experiments have been conducted on the proposed CAPTCHA to test its robustness as well as the level of security it provides. The experiments were done using six different OCRs on 500 different CAPTCHA samples. Nevertheless, the results of the experiments manifest the significant benefits of utilizing handwriting characteristics with CAPTCHA samples [1].

Author details

Maha Hamad Aldosari

Address all correspondence to: aldosari.maha@gmail.com

King Saud University, Saudi Arabia

References

[1] Aldosari MH, Al-Daraiseh AA. Strong multilingual CAPTCHA based on handwritten characters. In: 2016 7th International Conference on Information and Communication Systems (ICICS). 2016. pp. 239-245

[2] ur Rizwan R. Survey on captcha systems. Journal of Global Research in Computer Science. 2012;**3**:54-58

[3] Gummadi R, Balakrishnan H, Maniatis P, Ratnasamy S. Not-a-Bot: Improving service availability in the face of botnet attacks. In: NSDI. 2009. pp. 307-320

[4] Mehrnejad M, Ghaemi Bafghi A, Harati A. SEIMCHA: A new semantic image CAPTCHA using geometric transformations. ISeCure. 2012;**4**:63-76

[5] Patel C, Patel A, Patel D. Optical character recognition by open source OCR tool tesseract: A case study. International Journal of Computer Applications. 2012;**55**:50-56

[6] Kishore A. (2015, December 2nd). 5 Free Online OCR Services Tested and Reviewed. Retrieved from https://www.online-tech-tips.com/cool-websites/convert-image-to-text-using-free-online-ocr-software/

[7] Yan J, El Ahmad AS. Usability of CAPTCHAs or usability issues in CAPTCHA design. In: Proceedings of the 4th Symposium on Usable Privacy and Security. 2008. pp. 44-52

Edited by Beban Sammy Chumbow

Research in the area of bilingualism and multilingualism invariably produces fascinating insights. In the Europe of yesteryears, the paradigm of one nation one language was dominant and fashionable as a nation-building ideology that multilingualism was considered a curse, a demon that had to be exorcised. Today, the avalanche of empirical evidence of research findings has established multilingualism and pluralism as an ideal for national development.

The nine chapters of this book provide further elucidations of the issue of benefits of bilingualism and multilingualism and also provide original research findings on developments in the areas of psychological dimensions of bilingualism and bilingualism in information retrieval systems.

The book by its illuminating description and insightful analysis of issues of bilingualism will be of significant interest to scholars, researchers, and all concerned with bilingualism and multilingualism from whatever perspective.

Published in London, UK

IntechOpen

ISBN 978-1-83881-332-1